

In vivo Molecular Imaging of Epithelial Pre-Cancers based on Dynamic Optical Scattering Modeling

George Papoutsoglou



Thesis submitted for the degree of Doctor of Philosophy

Technical University of Crete

School of Electronic & Computer Engineering

Electronics Laboratory

Optoelectronics Group

Committee: Assoc. Prof. Balas Konstantinos, supervisor
Prof. Zervakis Michalis, advisor
Prof. Ntziachristos Vasilis, advisor
Prof. Papageorgiou Markos, member
Prof. Kalaitzakis Konstantinos, member
Prof. Stavrakakis Georgios, member
Assoc. Prof Theodoropoulos Panagiotis, member

March 2014

Περίληψη

Στα πλαίσια της διατριβής αυτής παρουσιάζουμε μία καινοτόμο βιοφωτονική μέθοδο μοριακής απεικόνισης για την εκτίμηση και χαρτογράφηση των λειτουργικών και δομικών παραμέτρων του επιθηλίου του τραχήλου της μήτρας κατά τη διάρκεια εξέλιξης της νεοπλασίας. Οι εκτιμήσεις βασίστηκαν σε πειραματικά δεδομένα, που λαμβάνονται μέσω δυναμικής, ενισχυμένης-αντίθεσης, οπτικής απεικόνισης του τραχήλου μήτρας, *in vivo*. Τα δυναμικά χαρακτηριστικά του μετρούμενου οπτικού σήματος διέπονται από τα φαινόμενα επιθηλιακής μεταφοράς του βιοδείκτη. Βάσει αυτών αναπτύχθηκε ένα διαμερισματικό, φαρμακοκινητικό μοντέλο του νεοπλαστικού επιθηλίου του τραχήλου που προβλέπει τα δυναμικά οπτικά αποτελέσματα σε όλους τους δυνατούς συνδυασμούς τιμών των παραμέτρων του. Εννέα δομικές και λειτουργικές βιολογικές παράμετροι του ιστού εντοπίστηκε να συσχετίζονται με την ανάπτυξη νεοπλασίας οι οποίες εμφανίζονται στα μετρούμενα δεδομένα με περίπλοκο τρόπο. Έτσι, χρησιμοποιήσαμε μεθόδους καθολικής ανάλυσης ευαισθησίας με σκοπό τον εντοπισμό του υποσυνόλου εκείνου των παραμέτρων εισόδου που αποτελούν τους πιο καθοριστικούς παράγοντες της εξόδου του μοντέλου. Για πρώτη φορά αποδείξαμε ότι είναι δυνατόν να εκτιμηθεί, από δεδομένα δυναμικής οπτικής απεικόνισης μετρούμενα *in vivo*, οι ακόλουθες συναφείς με τη νεοπλασία παράμετροι: ο αριθμός των νεοπλαστικών στρωμάτων, οι διαστάσεις του εξωκυττάριου χώρου, η λειτουργικότητα των σφιχτών δεσμών και το εξωκυττάριο pH. Τεχνικές καθολικής βελτιστοποίησης, έδειξαν ότι οι εκτιμήσεις της μεθόδου μας είναι επαρκείς ως προς την ορθότητα και την ακρίβειά τους. Ιδιαίτερα, ο αλγόριθμος της Διαφορικής Εξέλιξης συνέκλινε στο σύνολο των τεσσάρων, πιο ευπροσδιόριστων, παραμέτρων με σφάλμα περίπου 1%. Κατόπιν τούτου, δείχνουμε ότι οι εκτιμώμενες, σε δύο εκατομμύρια εικονοστοιχεία, τιμές των τεσσάρων αυτών παραμέτρων είναι αρκετά συνεπής με τις πληροφορίες που παρέχονται στη βιβλιογραφία. Τα αποτελέσματά μας είναι μοναδικά υπό την έννοια ότι χάρτες των λειτουργικών και μικροδομικών παραμέτρων μπορούν να υπολογίζονται και εμφανίζονται μαζί για πρώτη φορά, μεγιστοποιώντας κατά συνέπεια τη διαγνωστική πληροφορία.

Η ποσοτική και η ποιοτική προσέγγιση αυτών των πληροφοριών είναι ανέφικτη από άλλες επεμβατικές και μη επεμβατικές μεθόδους σήμερα. Τα ευρήματα της διατριβής αυτής προτείνουν ανεπιφύλακτα ότι η μέθοδός μας μπορεί να βελτιώσει την κατανόηση των μηχανισμών ανάπτυξης της νεοπλασίας καθώς και της φυσιολογίας των μεταστατικών όγκων. Ως φυσικό επακόλουθο, αυτή η μελέτη μπορεί να αποτελέσει ένα πολύτιμο διαγνωστικό εργαλείο που θα διευκολύνει την ανάπτυξη και την αξιολόγηση νέων θεραπειών του καρκίνου.



Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Η παρούσα έρευνα έχει συγχρηματοδοτηθεί από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο - ΕΚΤ) και από εθνικούς πόρους μέσω του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» του Εθνικού Στρατηγικού Πλαισίου Αναφοράς (ΕΣΠΑ) – Ερευνητικό Χρηματοδοτούμενο Έργο: Ηράκλειτος II. Επένδυση στην κοινωνία της γνώσης μέσω του Ευρωπαϊκού Κοινωνικού Ταμείου.

Abstract

We present a novel biophotonic method and imaging modality for estimating and mapping neoplasia-specific functional and structural parameters of the cervical precancerous epithelium. Estimations were based on experimental data obtained from dynamic contrast-enhanced optical imaging of cervix, *in vivo*. The dynamic characteristics of the measured optical signal are governed by the epithelial transport effects of the biomarker. A compartmental, pharmacokinetic, model of the cervical neoplastic epithelium has been developed, which predicts the dynamic optical effects in all possible parameter value combinations. Nine biological parameters, both structural and functional, have been identified to be potentially correlated with the neoplasia growth and to be manifested to the measured data in a convoluted manner. We have performed Global Sensitivity Analysis for the purpose of identifying the subset of the input parameters that are the key determinants of the model's output. We have for the first time shown that it is possible to estimate, from *in vivo* measured dynamic optical data, the following neoplasia related parameters: number of neoplastic layers, extracellular space dimensions, functionality of tight junctions and extracellular pH. Global optimization techniques showed that the estimations of our method are of adequate accuracy and precision. Particularly, the Differential Evolution algorithm converged to the set of the four, most identifiable, parameters with an error of roughly 1%. We show that the estimated, in two millions of pixels, values of the four parameters are quite consistent with information provided in the literature. Our results are unique in the sense that for the first time functional and microstructural parameter maps can be estimated and displayed together, thus maximizing the diagnostic information. The quantity and the quality of this information are unattainable by other invasive and non invasive methods. The findings of this thesis suggest strongly that our method can improve our understanding of the neoplasia development mechanisms and of tumor growth and metastasis physiology. Corollary, it may become a valuable diagnostic tool that will also facilitate the development and evaluation of new cancer therapies.



Co-financed by Greece and the European Union

This research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: Heracleitus II. Investing in knowledge society through the European Social Fund.

Acknowledgments

Mathematical modelling of biological procedures is an interdisciplinary task where the participating sciences work in the most creative manner. However, the lack of scientific information over certain issues is, of course, a major drawback for this quest that can potentially lead to unrealistic results. Perhaps, the most efficient way to overcome such problems is to keep it stupid, simple.

First of all, I would like to acknowledge my supervisor, Professor Costas Balas for the assignment of this project almost 9 years ago when I started as a Master student, for his continuous support and guidance, the outmost patience during the endless arguing and for his ideas that in essence paved the way for the completion of this Thesis. I will never forget him saying: "*This is the project of my life*". I strongly hope that our efforts over the years were worthy of his expectations.

In addition, I would like to thank both Professors Michalis Zervakis and Vasilis Ntziachristos for accepting to participate in the advisory committee of this Thesis. Equally important is the contribution of Professor Giorgos Stavrakakis, member of the committee, who pinpointed our initial steps for solving the parameter estimation problem. Special thanks should also be directed to all the members of the committee for their time and patience in reading and commenting on this Thesis.

Thanks to Antonis Potirakis, whose master thesis has been the basis of my research, helping me to gain insight about the overall problem. Many thanks should be also directed to the Optoelectronics Group friends and fellow colleagues Thanasis Tsapras, Giorgos Epitropou and Vassillis Kavvadias, for encouraging and supporting me during the development of this Thesis. Furthermore, I would also like to thank the undergraduate students that have actively participated from the beginning of this Thesis; that is Antigoni Anastasopoulou, Maria Stamatiadou and Theodoros-Marios Giakoumakis. Their contributions and assistance in completing several tasks of this project has been decisive.

Finally, I want to thank my best friends Alexandros Psarropoulos, Aggelos Antonopoulos and Nikos Makris for their psychological support especially in times of utter desperation. Most importantly, I would like to express my gratitude to Anna Verganelaki for watching over, tolerating and caring about me. Above all, I want to thank my family to whom I owe my principles and values, whose love and support during my academic years has been endless.

Table of Contents

Preface	xiii
Outline of the following chapters.....	xvii
1 Modeling of metabolic systems.....	1
1.1 Introduction.....	1
1.2 Modeling purposes.....	1
1.3 Theoretical basis	2
1.4 Model building cycle	3
1.4.1 Conceptualization.....	3
1.4.2 Mathematical Realization.....	3
1.4.3 Solution	4
1.5 Mathematical Representation.....	5
1.5.1 Lumped versus Distributed Models	5
1.5.2 Deterministic versus stochastic models	6
1.5.3 Linear versus non-linear models	6
1.6 Compartmental Models	7
1.6.1 Functional schemas of material flux	8
1.6.2 Strictly compartmental and control system models	10
1.6.3 Steady state and perturbation schemes	10
1.6.4 Non – compartmental analysis	11
1.7 Model Identification	12
1.7.1 Structural (a priori) identifiability.....	13
1.7.2 Global ranking of the parameters.....	18
1.7.3 Parameter estimation/identification	19
1.7.4 Experimental Designs.....	26
1.8 Model Validation	27
1.8.1 Validation criteria.....	27
1.8.2 Validation within and after the modeling process	28
1.8.3 Validation metrics	28
1.9 Summary.....	30
2 Bio-Optical Background for Model Development	32
2.1 Introduction.....	32
2.2 Physiology of epithelial tissues.....	32
2.2.1 Integration of cells into tissues	33
2.2.2 Types of epithelial tissues	34
2.2.3 Epithelial tissue activity.....	35
2.2.4 Preservation of homeostasis.....	41

2.3	Epithelia and cancer.....	44
2.3.1	pH regulation as the primary cancer-specific functional signature.....	44
2.3.2	Structural signatures of carcinogenesis.....	45
2.4	Cancer of the cervical epithelium.....	46
2.4.1	Morphology of normal and abnormal cervical epithelia.....	48
2.4.2	Physiology of the cervix during CIN development.....	49
2.4.3	Probing altered scattering properties of the tissue.....	50
2.4.4	Biophysical and biochemical phenomena followed by the application of the contrast agent.....	51
2.4.5	Interpretation of the dynamic scattering characteristics of the optical phenomenon: correlation with the kinetics of the biomarker.....	53
2.4.6	CIN assessment using the DCE-OI apparatus.....	55
2.5	Summary.....	57
3	<i>In silico</i> modelling of bio-optical dynamics in epithelial neoplasia.....	58
3.1	Introduction.....	58
3.2	The conceptual model.....	58
3.3	Mathematical implementation.....	60
3.3.1	Modeling the passive (free) diffusion.....	61
3.3.2	Modeling the active and passive regulation mechanisms upon acidification in tumors.....	62
3.3.3	Integration of the functional description into the differential equation system	64
3.4	System parameters.....	67
3.4.1	Structural parameters.....	67
3.4.2	Functional parameters.....	69
3.5	Qualitative validation.....	71
3.6	Summary.....	73
4	Sensitivity analysis of the developed model.....	75
4.1	Introduction.....	75
4.2	Particularities of the developed model.....	75
4.3	Sensitivity analysis as a model assessment tool.....	76
4.3.1	Local vs global sensitivity analysis.....	77
4.4	Theoretical principles of global sensitivity analysis.....	78
4.4.1	The probabilistic framework of GSA.....	79
4.4.2	GSA methods.....	80
4.4.3	Implications of variance-based GSA in identifiability testing.....	89
4.5	Application of global sensitivity analysis to the proposed model.....	90

4.5.1	Details of the employed methodology.....	90
4.5.2	Curve normalization.....	92
4.5.3	Results.....	94
4.6	Model restructuring.....	102
4.6.1	Comparison with the earlier calculations.....	104
4.7	Summary.....	104
5	Solution of the inverse problem.....	107
5.1	Introduction.....	107
5.2	Statement of the parameter estimation problem.....	107
5.3	Global optimization methods.....	108
5.3.1	Control Random Search (CRS).....	109
5.3.2	Shuffled Complex Evolution (SCE).....	112
5.3.3	Differential Evolution (DE).....	114
5.3.4	Simulated Annealing (SA).....	116
5.3.5	Particle Swarm Optimization (PSO).....	119
5.3.6	Mixed Integer Distributed Ant Colony Optimization (MIDACO).....	121
5.3.7	Evolutionary Strategy with Stochastic Ranking (SRES).....	124
5.3.8	DIRECT.....	125
5.4	Performing of the parameter estimation.....	127
5.4.1	The in silico experiment setup.....	127
5.4.2	Measuring the performance of the GO methods and the overall problem convergence	128
5.4.3	Maintaining the algorithms' best possible performance and consistency in the results	131
5.5	Parameter estimation results.....	135
5.5.1	Fitness level results.....	136
5.5.2	Parameter estimation errors – Theoretical analyses' validation.....	142
5.5.3	Accuracy of the GO methods.....	145
5.5.4	Statistical convergence of the inverse problem – the best performing algorithms	148
5.5.5	Interpretation of the performances of the GO methods.....	150
5.6	Summary.....	151
6	In vivo mapping of neoplasia related tissue parameters.....	153
6.1	Introduction.....	153
6.2	Clinical data.....	153
6.2.1	Systematic and random error.....	155
6.3	Fitting of the clinical data.....	156

6.4	Estimation and mapping of the neoplasia-related tissue parameters	159
6.4.1	The fitting procedure	159
6.4.2	High grade results	159
6.4.3	Low grade results	164
6.5	Summary	167
7	Perspectives	168
7.1	Thesis overview	168
7.2	Future work	171
7.2.1	Time complexity improvements	171
7.2.2	Applications	171
7.2.3	Feasibility study: intraoperative staging of tumors	173
7.3	Conclusion	174
8	References	175

List of figures

Figure 1-1	The “well-mixed” assumption. (a) The concentration depends on where the sample is taken from the compartment ($s_1 \neq s_2$). (b) The concentration does not depend on where the sample is taken from the compartment ($s_1 = s_2$)	6
Figure 1-2	Most common approximate schemas to describe functional dependence of the material fluxes: (a) linear kinetics and (b) Michaelis-Menten kinetics.	9
Figure 2-1	Elementary components of animal cells.....	33
Figure 2-2	Classification of epithelia.....	34
Figure 2-3	Primary epithelial junction arrangement.....	35
Figure 2-4	Transport pathways in epithelial tissues	36
Figure 2-5	Secretion and absorption can be either through microvillis (a), glands (b) or directly through the membrane (c)	36
Figure 2-6	Composition of the cell's membrane	37
Figure 2-7	The dissolve-diffuse mechanism of passive membrane diffusion	37
Figure 2-8	Transport proteins	38
Figure 2-9	Diffusion schemes through the membrane's lipid matrix.....	38
Figure 2-10	Epithelial tight junctions.....	39
Figure 2-11	The gap junction plaques consisting of several gap junction channels.....	40
Figure 2-12	Intrinsic buffering power dependence on pH_i . A) Trace, in the absence of CO_2/HCO_3^- . A weak acid (NH_4^+) was added extracellularly and then it was removed stepwise in the absence of Na^+ in order to inhibit active extrusion mechanisms. B) The resulting β_i was calculated using the expression $\beta_i = \Delta[NH_4^+]_i / \delta pH$, where $\Delta[NH_4^+]_i$ was calculated using the Hendersson-Hasselbalch equation [88].	43
Figure 2-13	Extrinsic, intrinsic and total buffering in ventricular myocytes. A) The CO_2/HCO_3^- pair (extrinsic buffering) has been evaluated in Cl^- free solution in order to avoid extrusion of bicarbonate through active mechanisms. B) The total and intrinsic buffer capacities of the same experimental procedure [88].....	43
Figure 2-14	The pH_{is} dependence of the acid extrusion rate in giant barnacle muscle fibers. The rate of pH_{is} recovery from an acid load has been determined over a range of pH_{is} values, multiplied by the sum of intrinsic and extrinsic buffering powers i.e. Flux = rate of change of $pH_{is} \cdot \beta_t$ [84]. B) The course of efflux rate when pH_{is} continues to drop at constant pH_{es} (~7.4). C) The dependency of the efflux rate on the pH_{es} while the pH_{is} is kept constant (~6.5) [89].	43
Figure 2-15.	The regulation of pH_{is} in tumors. After the production of protons, passive (buffering) and active extrusion (V-atpases - primary active - and the NHE and NBC exchangers - secondary active) mechanisms are initiated in order to maintain pH_{is} at normal levels. The result is to drop the pH_{es} due to the abnormal blood perfusion [182].	44
Figure 2-16	The uterine cervix.....	47
Figure 2-17	Classification of cervical intraepithelial neoplasia	47
Figure 2-18	Cervical epithelial layers	48
Figure 2-19	Photon random walks in normal and atypical, AA treated, areas of the epithelial tissue. Protein denaturation, fixation and cross-linking provoked by week acids change the size of cell scatterers, especially the nucleoproteins. This phenomenon decreases light scattering in small forward-scattering angles while significantly increases the light scattering of cells in the large angles. Particularly, an increase in the magnitude in the refractive index of the nucleus ($n_{nucleus}$) has been shown.	54

Figure 2-20	The DCE-OI system. (a) After calibration the patient is placed in the lithotomy position and the vaginal speculum is inserted, thus enabling the visualization of the cervix. (b) Crude representation of image capturing. (c) From the captured image stack diffuse reflectance curves are calculated for every image pixel, expressing the temporal characteristics of the AW phenomenon. From the corresponding integrals a pseudocolor map is generated and overlaid onto the real-time displayed color image of the cervix.	56
Figure 2-21	Intensity of the backscattered light as measured in CIN tissues.....	57
Figure 3-1	(a) Schematic representation of the cervical stratified epithelium; (b) detail of (a) showing a cell column with the basic ES and IS compartments.	59
Figure 3-2	Paracellular and transmembrane fluxes of acetic acid and its hydrogen and acetate ions.	59
Figure 3-3	Typical model predictions. The simulations have been performed using randomly selected sets of parameter values.	67
Figure 3-4	Structural composition of a two-compartment layer as proposed in our model. In this figure the height, length and depth of the compartments and diffusing membranes are included.....	69
Figure 3-5	(a) The intracellular hydrogen ion concentration kinetics as predicted by the model for three ncrs values; (b) typical experimental DR vs. Time curves, corresponding to CIN I, II, III conditions, histologically confirmed; (c) the max DR vs. NCR scatter plot	73
Figure 4-1	Potential input distributions. A) normal, b) uniform, c)triangular (symmetric) and d) discrete (bimodal).....	79
Figure 4-2	a) Input/Output scatterplot that reveals no specific pattern in the relationship. B) A clear linear pattern in the input/output relationship [165].....	80
Figure 4-3	Example of a scatterplot of a model factor (X_i) and the relative output (Y). A linear regressor has been fitted to the data and the following regression coefficients were found: $b_{x_i} = 0.0038$ and $b_0 = 0.035$	81
Figure 4-4	(Left) Cutting the scatterplot into slices and calculating the average output $E(Y X_i)$. (Right) The variance of the expected values $V(E(Y X_i))$ [165].....	85
Figure 4-5	The total order sensitivity index plotted as a function of time for the entire set of model parameters. The time range corresponds to the measurement period of the simulated dynamic optical effect	95
Figure 4-6	The first order (dark gray) and the total order (light gray) sensitivity index estimates in (first order) descending sequence for eight model input parameters after discarding K_v	95
Figure 4-7	The peak estimability ratio er_{max} for the remaining eight model parameters. The time points (in s after the application of acetic acid solution) at which the max ER values are achieved appear at the top of the histogram bars.	96
Figure 4-8	The correlation matrix for the eight parameters in all possible pairs. Parameter pairs (ph_{es} , ph_{is}), (b , θ_{es}) and (a , θ_{is}) display (with rounding) unity correlation coefficient.	
	97	
Figure 4-9	The correlation matrix of the five highly ranked parameters (factor prioritization) after fixing the insignificant parameters (factor fixing). No unity correlation coefficient has calculated for all the pairs of different parameters. In both cases the correlation coefficients can be visualized with the aid of gray shade pallet. Darker shades correspond to high coefficient values.	97

Figure 4-10	Normalized experimental Diffuse Reflectance (DR) vs. Time curves, 10 corresponding to CIN I (low grade (LG)) and 20 to high grade (HG) neoplasia (10 CIN II and 10 CIN III).	99
Figure 4-11	The model's outputs that best fit the average experimental curve of each particular CIN class	99
Figure 4-12	Box and whisker plots of the estimated parameter values for the highest ranked parameters according to the GSA outcomes. The statistical analysis has been realized for each CIN class separately. Crosses indicate the outliers in the estimated values sets.	100
Figure 4-13	a) The new correlation matrix for the full (eight) parameter case. Notice that although the collinearity of the pairs (ϕ_{es}, ϕ_{is}) , (b, β_{es}) and (a, θ_{is}) has been dropped they still have the highest ones among all the pairs in the matrix. B) The new correlation matrix for the reduced (four) parameter case. After fixing parameter a and reducing the model's complexity, the identifiability of the highest ranked parameters, which are now four, has improved significantly. In both cases the same palette has been used as in Figure 4-8 and Figure 4-9.	105
Figure 5-1	Flowchart of the crswch algorithm	111
Figure 5-2	Flowchart of the SCE algorithm	113
Figure 5-3	Flowchart of the DE method	116
Figure 5-4	Flowchart of the tdasa method	118
Figure 5-5	Flowchart of the PSO method	120
Figure 5-6	Flowchart of the MIDACO method	123
Figure 5-7	Flowchart of the SRES method	125
Figure 5-8	Flowchart of the gblsolve procedure	126
Figure 5-9	The FOS functions of the full parameter set (a) and of the four most influential parameters (b) after the restructuring described in paragraph 4.6. The dashed, bold line indicates their sum. Clearly, more than 90% of the output's variance is explained by the parameters in both cases.	130
Figure 5-10	The 30 model output signals which have been created and were used as experimental ones. (a) The outputs for the full parameter set case. (b) The outputs for the reduced parameter set case.	137
Figure 5-11	The fitting result of each GO method in terms of curve convergence when estimation of the full parameter set is realized. On top of each sub-figure the algorithm is indicated. Each dashed blue line displays the PEX curve and the clouds of red lines their corresponding fits.	138
Figure 5-12	The fitting result of each GO method in terms of curve convergence when estimation of the reduced parameter set is realized. The same notation as in Figure 5-11 has been employed.	139
Figure 5-13	Loglog plot of the efficiency with respect to the time lapsed after the initiation of the fitting procedure of each GO method. Each algorithm is represented by separate color lines (see legend).	141
Figure 5-14	Parameter prediction error when the full parameter set is estimated. The error bars have been computed using the bootstrap method where 1000 bootstrap samples were used as input to the Matlab's function <i>bootci()</i> . Each algorithm is represented with a separate color bar (see legend).	142
Figure 5-15	The prediction error of the parameter values when the reduced parameter set is estimated. The same notation as in Figure 5-14 has been employed.	143
Figure 5-16	(a) The average estimation error of each parameter over all GO methods employed and over all curve fits for the full parameter PEX dataset. (b) The respective	

average estimation error for the reduced parameter PEX dataset. On top of the bars the mean average error value is indicated..... 144

Figure 5-17 Boxplots of the distributions of the prediction quality of the employed optimizers when the values of the more complex (a) and of the reduced (b) parameter set are estimated. It is clear that the reduction of the parameter complexity has significantly improved the accuracy in the predictions to all algorithms except tdasa. The circles within the boxes indicate the mean values..... 146

Figure 5-18 The overall performance of each algorithm in terms of mean estimation error over all PEX curves. For calculating the confidence of the mean values 1000 bootstrap samples with replacement have been employed. It is clear that crswch performs best in the full parameter case (a) and in the reduced parameter set case (b). Conversely, the performance of tdasa is the highest in both cases..... 148

Figure 6-1 Color images corresponding to LG and HG cervical epithelia which have been measured according to the procedure in paragraph 2.4.6. These images refer to the doctor's observations. The top row refers to the high grade cases and the bottom to the low grade ones. The numbered circles within the images point out the areas from which the biopsy samples (BS) were obtained. 154

Figure 6-2 The preparatory procedure of the experimental data. (a) Raw experimental curves. (b) Normalized experimental curves. (c) The smoothed (circles, diamonds) and the biexponential fitted curves. Notice that the initial raw intensity is different and that the features are noisy (a). Then, the preparatory step alleviated the unwanted effects bringing no modification on the features of either signal (c). 155

Figure 6-3 Twelve of the collected NDR curves. The scatter lines correspond to the smoothed normalized curve; and the dashed lines to the model's prediction. The average R^2 for the LG cases was 0.89 and for the HG ones 0.96..... 156

Figure 6-4 The distributions of the estimated model parameter values calculated by the crswch global optimizer. Data refer to both clinical sets (LG/HG). The x's mark the mean values. On top of the axes the parameter that the data refers to is indicated..... 157

Figure 6-5 (a), (b) The pseudocolor maps of N , b , ph_{es} and g_{ij} neoplasia-related parameters of the HG1 and HG2 cases of cervical epithelium, respectively. The circles denote the areas from where biopsies have been collected. Color-coding of the maps corresponds to various parameter value ranges. 160

Figure 6-6 (a), (b) The pseudocolor maps of N , b , ph_{es} and g_{ij} neoplasia-related parameters of the LG1 and LG2 cases of cervical epithelium, respectively. The same notation as in Figure 6-5 has been employed. 164

List of tables

Table 3-1	List of the biological parameters embedded on the developed model.....	68
Table 3-2	Geometric properties of cervical cells adopted from [113]. The gray shades designate the atypical cells conforming to the histological grading discussed in paragraph 2.4.1.	69
Table 4-1	Value ranges of model parameters employed in GSA studies. The assumed types of distributions used for parameter sampling in the context are also provided. ..	92
Table 4-2	Width of ES associated with each cell layer (adopted from [185], table 7-7-2). The gray shades illustrate the ES width that corresponds to the occupancy of the epithelium by neoplastic cells.	104
Table 5-1	Value intervals of the biological parameters when the full and the reduced parameter sets are estimated.	134
Table 5-2	The user specified parameters of the employed optimizers and their corresponding values as they were appointed for our calculations.....	134
Table 5-3	The average objective value (J) of each global optimizer after fitting the 30 PEX curves corresponding to the full and reduced parameter sets. The probability of each method to terminate due to either $cr.(1)$, $cr.(2)$ or $cr.(3)$ criterion is also indicated.	140
Table 5-4	Results of the Students t -test for significance level (threshold probability) $a=0.05$ for both estimation error thresholds when the reduced parameter set has been estimated. The hypothesis is rejected if the statement P -value $> a$ holds.	149
Table 5-5	Results of the Holm-Sidak test for significance level $a=0.05$. The hypothesis is rejected if the statement P -value $< a/i$ holds. Crswch is the reference method with $i = 0$	150
Table 6-1	The result of the histological assessment corresponding to the cases displayed in Figure 6-1.....	154

PREFACE

Functional and structural remodeling of human tissues is a hallmark of developing neoplasia. Currently, first line screening of the imposed abnormalities is performed by visual inspection. The detection of the abnormal patterns, however, is almost impossible for the unaided human eye, especially at an early, curable stage and mainly relies on the doctor's acuity. Furthermore, invasive biopsies followed by cytological or histopathological analysis remains the gold standard for diagnosis and surveillance for most cancers. Yet, these conventional diagnostic tools have their own limitations such as: (a) difficulty in distinguishing between benign and malignant lesions; (b) difficulty in determining the margins of lesions leading to high biopsy sampling error rate; (c) can suffer from high inter- and intra-observer disagreement, especially the histopathological analyses; and also (d) tissue biopsies can be invasive, painful and, most importantly, stressful for the patients.

To overcome these difficulties, emerging optical imaging technologies exploit the endogenous and/or exogenous contrast of abnormal tissue areas in order to assist physicians and researchers in visualizing the potential malformations, *in vivo*. Endogenous contrast comes from the differing intrinsic properties of living tissues during the development of lesions, as well as the intrinsic luminance under certain radiation stimulation of many substances including proteins, metabolites etc. Exogenous contrast comes from the growing collection of compatible chemical contrast agents which are topically, orally, or intravenously administered and are specifically targeted to activatable biochemical cancer-specific traits. For this latter case, the scientific interest on building biomarker diagnostic optical imaging tests has been increased remarkably, nowadays, and their potency as an alternative screening tool are under intensive investigation [1]–[5]. As a result the optical imaging modality has been evolving rapidly and has demonstrated several unique advantages over magnetic resonance imaging, positron emission tomography, X-ray computed tomography etc., such as: high temporal (picosecond) and spatial (submicron) resolutions and high sensitivity (single molecule level) [6], [7]. These advantages are of great significance and have comprised the basis for the development of advanced diagnostic and pharmaceutical screening methods and technologies.

A cardinal reason for attracting this tremendous attention has been the fact the integration of contrast agents and optical imaging technologies can be of very low cost and can often be implemented in a simple bench-top configuration requiring no shielding, and minimal operator training. For example, developing countries are seeking for alternative cost-effective solutions for diagnosis and screening, because they cannot afford to implement and maintain cytology-based primary screening methods. Recently, the results from a large clinical study investigating the performance of the Visual Inspection with Acetic acid (VIA), have been published [8]. According to this publication, approximately 50,000 women have been enrolled in the study and a reduction in cervical cancer incidence by 30% has been observed in the group examined with VIA, compared with the control group attending the established diagnostic chain (cytology, colposcopy, histology).

Despite this remarkable success, the problems of subjectivity, low specificity and lack of reproducibility remain the major restricting factor for the clinical adoption of such methodologies. In fact, clinical assessments must, still, be based on the phenomenological/empirical comparison between optical measurements and the cytological or histological analysis, both referring to the same tissue point. One principal reason for this is the fact that the induced optical effects (perceived as color changes) are treated only in a static, qualitative fashion disregarding their temporal characteristics. As a rule, these effects emanate from both the biodistribution pathways and the reversible pharmacodynamic responses in which the contrast agents are involved. As such, the measurable dynamic optical signals are inherently very informative carrying a rich palette of cancer-specific tissue characteristics. It is reasonable, hence, to suggest that if the correlation between the biomarker kinetics and the macroscopically measured dynamics signals is mathematically established, the estimation of the functional and microstructural signatures of the abnormal tissue can be possible. It follows that the probing, modeling, parametric analysis and mapping of the optical signals can be instrumental for assessing the structure and function of several neoplastic tissues. If this is successful, it holds the promise to provide not only more sensitive imaging modalities, but also quantitative descriptions of several cancers that will alleviate the problems of subjectivity and specificity and also to eradicate the need of unnecessary biopsy samples. There is a clear necessity, therefore, for merging together the phenomenological approaches, which are still dominating the screening of cancers, and computational modeling or “executable biology” approaches.

To meet this goal, Dynamic Contrast Enhanced (DCE) imaging has emerged over the past decade as a versatile technique for diagnosis, grading, staging, new drug development and evaluation of treatment effectiveness [9]. In general, DCE imaging techniques rely on tracking the effects of the contrast agent(s) on organs, tissues, cells, proteins etc. Then, understanding how various disease types and states are expressed to these dynamic effects, DCE imaging comprises the basis for the development of a series of novel, non-invasive, real-time diagnostic methods and technologies. Indicatively, under magnetic resonance tomography, the concept of DCE imaging has already been established as a non-surgical tool for the study of tissue perfusion kinetics, *in vivo*, and has shown an impressive prognostic and predictive capacity in cancer radiology [10]–[12]. Thereupon, DCE imaging is progressively implemented to almost all biomedical imaging modalities other than magnetic resonance including, Optical Imaging, X-Ray imaging, Ultrasound Imaging, etc., by developing and utilizing high-affinity labeled tracers. Physical modeling, mathematical/computational techniques and real-time data processing platforms have been instrumental to these developments.

Contributing to these efforts C. Balas *et al.* developed in the late nineties a novel dynamic, contrast-enhanced optical imaging (DCE-OI) platform able to measure, quantitatively, the optical signals that the biomarker-tissue interaction generates [13], [14]. Particularly, a DCE-OI system was developed that measures and maps the evanescent backscattering signals which are generated after the application of a contrast agent on the exterior of the neoplastic epithelial tissues of the cervix, the larynx and the skin [15]. The method employs the acetic acid (AA) dilute solution (3-5%) as a biomarker and allows for the spatio-temporal recording of the induced biological events and processes, *in vivo*, non-invasively and in real time. In two international clinical trials, enrolling hundreds of patients; it was proved to be very efficient, demonstrating an improvement of more than 63% in diagnostic sensitivity over Papanicolaou test and colposcopy [16].

Due to the fact that DCE imaging reflects the kinetics of diffusible, high-affinity labeled tracers in tissues it allows for the model-based quantitative assessment of several biological parameters that are determining the features of the experimental data. In addition, since DCE-OI measures these data in every spatial location, it enables the mapping of the estimated parameters. Building on such grounds, we have interpreted the agent-uptake pharmacokinetics and identified a list of biological parameters that determine them on both molecular and macroscopic scale in order to obtain new insights on the state of neoplasia, its physiology and its progression [17]. For this, a rationally designed mechanistic computational model was developed (compartmental), which exactly simulates the metabolic pathways that are followed by the biomarker in the epithelial tissue. Taking normal and pathologic epithelium architecture into account, the developed model encapsulates the abnormal part of the tissue and partitions it into a stack of functionally and structurally identical cells. Each cell, in turn, corresponds to an entity that comprises of two interconnected compartments of interaction. Functional dependences between compartments correspond to the diffusion of the biomarker as a molecule and of its ionization counterparts. In the end, a non-linear algorithm that includes a set of scalable differential equations is programmed in Matlab™ and Fortran 90 that has as an input the biological parameter value ranges and outputs the concentration temporal changes of the solutes under study.

Qualitative validation of model predictions with biopsy readings have shown that a clear correlation between the dynamic scattering characteristics of the AW effect and the structural and functional characteristics of the epithelium exists [18]. Then, since the biomarker uptake kinetics determine the dynamics of the biological target's optical activation, they establish the link between optical and biological parameters and responses. Hence, the fitting of the developed pharmacokinetic model predictions to the experimental data, may allow for the estimation of a set of neoplasia-related biological parameters, comprising the solution of the inverse problem.

In a general sense, the parameter estimation problem is the identification of a system from experiments and studies of its behavior. Hence, for a dynamic biological system to be identifiable the developed model must be capable of emulating the available input-output data through its proposed structure [19], [20]. In other words, the developed model can be regarded as predictive, if the available experimental data can be reproduced through its artificial parameters and structure. Proving this, according to the field of computational systems biology, involves the study of the identifiability and estimability of the model parameters [21]. In our case, the fact that the developed model is highly non-linear and cannot be represented analytically limits our options. Thus, towards verifying the uniqueness of the inverse problem's solution we have proposed and performed consecutively three *in silico* analyses; namely, global sensitivity analysis (GSA), parameter identifiability analysis and parameter estimability analysis [22]. Through this procedure we expect to find an answer to the question whether the unknown system parameters can be uniquely estimated, at least in theory.

On the basis of this roadmap, we were able to decide which of the pertinent model parameters can potentially be estimated from available input/output data and which are impossible to assess. Particularly, our analysis has concluded that unique estimation is possible by reducing the model's parameter complexity to the subset of those parameters that are most influential to the dynamics of the output. More specifically, we have identified a set of parameters that fulfill the following criteria: a) they are the key determinants of the line-shape of the model's output that fits the experimental dynamic optical data and b) they display no collinearities and minimum interdependency with each other. These findings have allowed us

to introduce a new methodology for estimating concurrently and non-invasively a set of biological parameters with proven correlation with neoplasia growth, and hence of undisputed diagnostic value. These parameters regard both structural and functional properties of the neoplastic tissues, namely the number of neoplastic layers, the size of the extracellular space, the extracellular pH and the tissue's porosity. Under the same context, a first, high level validation of the model's structure, calibration and dimensionality reduction has been realized by attempting to estimate the selected parameter set through the comparison of experimental results with model predictions. This comparison was not regarded as a solution to the inverse problem in the strict sense. For the first time, nevertheless, it has been shown that it is, in principle, possible for the parameter estimation problem to convergence to unique solutions.

Even though our implementation enhances the well-posedness of our parameter estimation problem, it does not prove a deterministic relation between the solution and the parameter values. In fact, a prerequisite for well-posedness of the parameter estimation problem is the uniqueness and reproducibility of the candidate solutions [23]. Towards establishing this one should state the inverse problem as the minimization of a cost function that calculates the goodness of the fit between the model output and a given experimental data set, subject to the dynamics of the system. Here, the constructed optimization problem is a mixed-integer nonlinear (MINLP) problem that pertains to no analytical solution. Under such context, we have identified global optimization (GO) methods as the most efficient tool for addressing the issue of verifying whether our method has the capacity of in providing a substantially unique parameters' value combination, when a certain experimental curve is fitted.

The first step of our presented analysis involves the generation of a set of "pseudoexperimental" (PEX) data by collecting the model's responses to a known range of values of all parameters. These PEX curves have been appropriately selected to correspond to all clinical cases/classes, with each case/class mimicking the experimental data obtained by cases with different neoplasia grades. Next, the PEX curves were fitted until an acceptable level of goodness of fit was reached. At that point a new set of parameters (candidate solution) was collected and the procedure is repeated many times. Finally, the accuracy and reproducibility of model-predicted parameter values is accessed by comparing them with the known PEX values.

For implementing this approach we have employed and tested a series of global estimation techniques. These methods benefit from the fact that they are easy to implement and not critically dependent on a priori information about the objective function. Algorithm selection was driven by the diversity of each method due to the fact that GO techniques give no assurance that will reach to the optimal solution. Under this framework, our capability to predict, in practice uniquely and reproducibly, the full and/or the reduced set of the proposed parameters has been assessed by comparing eight, equally constrained, state-of-the-art GO algorithms [24]. Specifically, constrained, direct search, point-to-point and population based, global optimization algorithms namely the gblSolve (DIRECT); the thermodynamic adaptive simulated annealing (tdASA); the Mixed Integer Distributed Ant Colony Optimization (MIDACO); the controlled random search with competing heuristics (CRSwCH); the differential evolution (DE); the shuffled complex evolution (SCE); the particle swarm optimization (PSO) and; the Evolutionary Strategy with Stochastic Ranking (SRES) have been comparatively evaluated with the purpose of identifying the best performing algorithm in estimating parameter values in the vicinity of the PEX curve input values, with adequate reproducibility.

Our results have demonstrated that model predictions may converge adequately to the PEX input parameters with less than 1% prediction error. This occurs more predominantly when the reduced set of parameters is used, in agreement with GSA results. In addition, the DE and the CRSwCH algorithms have been statistically superior to the rest of the GO methods employed here. To the best of our knowledge such setting entails an innovative approach for addressing the problem of estimating tissue biological parameters of high relevance to cancer progression. In fact, the introduction of global optimization, computational biology and DCE-OI approaches has never been used for facilitating the interpretation of dynamic bio-optical effects and processes and for enabling the mining of valuable information in complex biological systems.

The fact that we have been able to verify that the solution of the inverse problem (a certain set of biological parameter values) can be substantially unique allowed us to exercise our methodology to actual experimental data. For this, we have selected clinical results from 2 patients with low grade (LG) and 2 with high grade (HG) lesions, biopsy confirmed. As a step forward, we established the correlation between the highest ranked model parameters and the progresses of neoplasia from LG (CIN I) to HG (CIN II/III). Our findings suggest that our predictions concerning the fluctuation of the model parameters from LG to HG cases largely conforms to the biology of the tissue during the progression of neoplasia. Particularly, our predictions for the number of dysplastic layers, which is the golden standard for histology classification, have been in total agreement with the doctors' assessments. Similarly, the interpretations of the predictions regarding the rest of the parameters are compatible with the finding of other groups for both HG and LG neoplastic epithelia.

Finally, we present in this thesis, for the first time, our results from the pixel-by-pixel estimation of biological parameters from DCE-OI of a tissue, measured *in vivo* [25]. We furthermore show that the estimated, in two millions of pixels, values of the four parameters are quite consistent with information provided in the literature. Our conclusions suggest strongly that our method can improve our understanding of the neoplasia development mechanisms and of tumor growth and metastasis physiology. Corollary, it may become a valuable diagnostic tool that will also facilitate the development and evaluation of new cancer therapies. Potentially, this has the power to become the next paradigm shift in optical biopsy and cancer screening.

OUTLINE OF THE FOLLOWING CHAPTERS

Chapter 1 involves the task of building a rigorous computational model of a biological system. This involves a lot of intermediate, possibly repetitive, steps. In principle one goes from conceptualization to mathematical realization and finally to solution. Following this course, several types of models can be developed. For dynamic, metabolic processes, which are our interest, compartmental models are the mostly suitable. Model identification and validation then includes several separate studies. These basic tools for building a computational model of a biological system are described in chapter 1 and will reflect the framework of the rest of this thesis.

Chapter 2 follows the course suggested in chapter 1 where we present the theoretical basis of the developed model. This involves the building of cells into epithelial tissues and how this is disrupted during neoplasia development. Next, the biophysical signatures of neoplasia are described and how these are involved in the optical activation of a chemical contrast agent. Finally, the technical specifications of the developed DCE-OI apparatus are described with the aid of which the malformations are clinically probed.

Chapter 3 discusses the mathematical implementation of the theoretical discussion of chapter 2. Here, the building blocks of the compartmental model are developed and the structuring of the differential equation system suitable to describe the kinetics of the biomarker. In addition, the biological parameters of the metabolic system under study are thoroughly discussed. Then, the proposed model's output is presented. Lastly, the model predictions are qualitatively validated by clinically collected experimental data.

Chapter 4 describes the sensitivity of the model's output to the proposed biological parameters. Here, we discuss the reasons to reduce the parameter complexity of the model in order to be able to predict with accuracy and reproducibility the clinical status of the tissue. To this end, global sensitivity is employed and from its outcomes the identifiability and estimability of the parameters are theoretically assessed. Again, the proposed reduction is qualitatively validated with experimental data.

Chapter 5 involves the answering of the question whether the proposed model structure allows for the unique estimation of the unknown system parameters when experimental, noise-free, data become available. For this, eight separate global optimization methods are employed and the statistics of their parameter predictions are computed. Our results validate our initial assumption that parameter reduction increases the solvability and uniqueness of solutions of the inverse problem.

Chapter 6 describes the application of our methodology to measure from *in vivo* images of clinically collected, biopsy confirmed, data the characteristics of neoplastic tissues. Particularly, we exploit this novel biophotonic and imaging modality for estimating and mapping a set of neoplasia-related biological parameters, from dynamic optical data (DCE-OI), *in vivo*. Confirmation of model based prediction with literature is extensively discussed.

Finally, chapter 7 summarizes the most important points made throughout this thesis. It presents the additional work that needs to be performed in order for our innovative method to become a reliable diagnostic tool. Potential applications have also been discussed here.

1 MODELING OF METABOLIC SYSTEMS

1.1 INTRODUCTION

Many processes of the real world are so complex that physical experimentation can oftentimes be too time consuming, too expensive or even impossible to reproduce them. As a result, scientists have turned to mathematical models in order to interrogate many of the compound and dynamic scientific phenomena around us. By definition, mathematical models are convenient approximations of real systems that correspond to a set of mathematical relations between measurable system quantities. The relations in the model are projections of fundamental physical laws that, essentially, describe the behavior of the real system to a degree adequate for the purposes for which the model has been originally intended and under all the circumstances for which the model is deemed to be valid. The basic guidelines to follow for the purpose of establishing a rigorous, *in silico*, mathematical representation of the metabolic system which is responsible for the macroscopically observed dynamic optical signal during biomarker-tissue interaction, are outlined below [26], [27].

1.2 MODELING PURPOSES

The framework onto which a mathematical model is constructed is essentially dependent on the purposes for which the model is required i.e. for *descriptive*, *predictive* or *explanatory* purposes. For descriptive purposes, a mathematical model intends to establish rigorous mathematical expressions that provide a comprehensible and objective representation of the experimental data. For predictive uses, a model serves as a mean to simulate the system's responses upon external stimuli. Otherwise, explanatory models satisfy the need of understanding the internal structure of the system; specifically, for understanding how its constituents depend on and interact with each other, for guiding experimental research by investigating a hypothesis and/or for measuring inaccessible system quantities, from observable ones.

Alternatively, in the physiological sciences and medicine the modeling purposes can be narrowed down to the following fields of application and objectives:

1. Identification of system structure in terms of the nature of its components, that is organs, reactions, and transport mechanisms that comprise it, and the manner of their interconnection. The formulation of alternative structures of model enables different hypotheses regarding the nature of specific physiological or biochemical mechanisms to be evaluated.
2. Estimation of internal parameters for diagnostic purposes in terms of using fixed mathematical models in estimating internal inaccessible parameters of physiological interest from input/output data. Such parameter estimation finds

application in physiological and pharmacokinetic studies, and in clinical medicine.

3. Predictive models in terms of identifying the response of the physiological organism or a specific organ to the effect of a drug in order to maintain an optimal therapeutic level. Such models can be used as potential aids in patient management.
4. In physiological research and clinical practice in terms of serving a specific teaching function. The development of models that can readily be used to simulate a wide variety of effects constitutes a significant aid in the teaching of many aspects of physiology, clinical medicine and pharmacokinetics.

Ultimately, mathematical models in the life sciences aim to provide lucid explanations of the behavior of biological systems both in health and disease, assisting in the identification of therapeutic targets with the promise of improving, on the long run, our well-being.

1.3 THEORETICAL BASIS

The basis of a mathematical model is formulated on current knowledge about the system under investigation. The more the area of research in which the system belongs to develops the more the available theories and data in the field. Nevertheless, the amount of, *a priori* knowledge included for the basis of a model should be as much as is consistent with the model's purpose. For mathematical modeling of metabolic systems, this basis entails fundamental biochemical and biophysical laws, theories and experimental data. According to the features of the background knowledge available, models can be characterized as either empirical, or theoretical or a combination of both.

Empirical models assume no *a priori* knowledge other than the nature of the variables of the system and the existence of causal relationships between them. Derivation of such models is realized by tuning the system's independent variables or inputs and observing the responses of the respective dependent variables or outputs. In this way, input/output relationships are empirically determined and without making any theoretical assumption about their form, simple curve equations are fitted to the data. Due to the fact that no assumptions are made concerning the internal features of the system the model is called a black-box or alternatively a curve fitting model. It follows that the power of the model is basically descriptive only while its predictive power is restricted. Therefore, with such a model, it is only possible to determine how the system would respond to stimuli or changes occurring within the range for which experimental observations exist.

Theoretical models, on the other hand, are based on validated *a priori* knowledge. They use information regarding the structure and function of the system, the basic biochemical and biophysical laws applicable, and quantitative data about dimensions, material properties, process rates, etc. In addition to providing a description, they can be used to predict the behavior of the system under circumstances that have not been directly tested or cannot be tested. In addition, theoretical models are capable of explaining the relations between the various features of the system structure and behavior in terms of laws established for a wide spectrum of phenomena and systems. Their power is limited only by the extent to which they are valid. Therefore, they always need to be tested for validity on the system.

Often, it is rare to have adequate detailed knowledge for the derivation of fully theoretical models. Thence, a combination of theoretical and empirical knowledge is also used when modeling a metabolic system, where, certain parts or processes of the system may be

modeled empirically, whereas others are modeled theoretically. Empirical-theoretical models may be called gray-box, lying between the black-box empirical models and the wholly theoretical model that assumes that the system is fully transparent with all its internal structure and function known. In their predictive and explanatory power, the gray-box models are intermediate between wholly empirical models and validated theoretical models.

1.4 MODEL BUILDING CYCLE

If the procedure of converting a system into a mathematical model is considered as an entity of elements then, according to the systems theory, the building of sound computer models of dynamic systems is formally regarded as a cycle between these elements. For this, three distinct procedural components can be identified; the model conceptualization, the mathematical realization of that concept, and the assessment of the variables of interest through the solution of the model. At any stage in the cycle, it may be necessary to obtain additional data. Due to the fact that the final model will have either unknown or uncertain structure or parameters, this model is used to gather new data, restarting the cycle.

1.4.1 Conceptualization

At the beginning of the cycle, the definition of the purpose of the model and some underlying assumptions, which are necessary in order for the model to be tractable, should be realized. On the basis of validated a priori physiological knowledge, conceptual assumptions attempt to create a functional description of the relevant processes and their interconnection. Generally, these involve lumping together different components (e.g. cells or organs) into one single entity for the purpose of modeling (aggregation); neglecting features of the system, to the degree to which only important aspects of a system are considered into the model (abstraction); approximating or treating structures or behavior of a system that are difficult to be described by more simple idealized ones (idealization).

Notably, however, there are several essential suppositions that cannot always hold for biological systems. For example, the system inputs may not be easily and freely perturbed to reveal a richer ensemble of dynamic responses. In addition, these responses, in turn, may not always be able to be accurately measured. Furthermore, the experiments carried out for modeling purposes cannot always be iterated several times without significant extra 'costs'. Therefore, preliminary conceptual assumptions should be kept under careful consideration when the modeling of a biological system is intended [28], [29].

1.4.2 Mathematical Realization

Having produced a suitable conceptual model, the next step is to choose and evaluate the appropriate mathematical framework upon which the model is built. Initially, this involves the determination of the biophysical and biochemical laws that describe best the internals of the conceptual system i.e. functions, relations and/or physico-chemical processes between variables of interest. Then, mathematical realization is allowed to be performed and the appropriate equations to be constructed. In bio-models these equations are usually differential and associate with the mass or concentration of a substance in a specific location of the system, for example in a cell or in an organ, and with the interconnection of the several locations of the system. These equations constitute the initial model structure.

In addition, prior to performing any solution procedure the collection of information regarding the system parameters' values is necessary. Usually, parameter values are available as tabulated numerical values in the literature. In the case of unprecedented experimental

measurements two practical options can be suited. One is to adapt the model parameter values through repetitive discrepant simulations until they are consistent with most of the heterogeneous datasets available. This also helps to gain confidence in the predictions of the model in the following steps. The second option is to design a pilot experiment in order to acquire experimental data from the system. In this latter case, the system is appropriately perturbed and from the resultant responses parameter values are either directly measured or are estimated using established models from the literature.

1.4.3 Solution

The solvability of the model is an important consideration in its formulation. In the case where the structure of the model have been known *a priori* model solution simply involves obtaining the required explicit relations between variables and parameters, by computer implementation. However, this is rarely the case in bio-modeling. In fact, it is common that the system's structure is not known a priori. Primarily, it is the conceptual preliminary assumptions concerning the structure of the model and/or its parameters that give rise to uncertainties. It would, therefore, be unwise to solve the model directly. Instead, identification of the model's cogency from input/output data should be carried out.

Inferring whether model predictions adequately agree with the observed data requires the experimental determination of the model, considering its structure and/or its parameter values. Cardinally, this is performed through the estimation of the proposed parameters by minimizing some function of the error between the experimental and model responses. If the estimation process is successful then, the early, working version of the model must be validated with new experiments. In most cases, however, parameter estimation reveals certain design caveats. This means that the formulated model cannot be tractable either for solution or for gathering of new data. Probably, this would arise, if there is a substantial mismatch between the complexity of the postulated model and the paucity of data. Under these circumstances, the postulated model should be restructured through the reduction of its complexity and/or new experiments have to be designed.

For model reduction/restructuring there are two approaches. First, is to start with a comprehensive model followed by subsequent systematic model reduction soundly based on current physiological knowledge. In other words, return to the conceptual stage and redefine the entire acknowledged framework upon which the model was initially formulated. Naturally, the result should be a simpler conceptual model that will not, however, lack in physiological realism. This ensures that all necessary modeling assumptions are made clear and explicit. Second is to reduce the model's complexity by reconfiguring its parameters. For example, sensitivity analysis of the model output(s) can designate the influence of each of the postulated parameters to the measured variables. This allows the rejection of the unimportant/unnecessary parameters ensuring, up to a point, that the relevant ones could be estimated from the experimental data. In any case, new experimental designs may always be requested to plan in order to generate optimally informative data for parameter estimation. This involves finding the optimal way to perturb the system, which of its components to measure and at which time points to sample (model-based experiments).

The reduction of the model's complexity and the design of new experiments are two complementary rather than mutually exclusive operations. Consequently, these processes are repeated iteratively until the validation step is considered satisfactory. This iterative cycle of prediction and experimental validation progressively strengthens the predictive power of the model.

1.5 MATHEMATICAL REPRESENTATION

As indicated earlier, physiological processes exhibit complex dynamics. In addition, the characteristics of the data available for system identification and model validation limit the accessible approaches that could be adopted in modeling metabolic systems, along with purpose and theory. Naturally, therefore, models of metabolic systems have covered most of the spectrum of approaches typically available to the control engineer. Hence, within the literature there are examples of lumped and distributed models, deterministic and stochastic models and linear and nonlinear models of metabolic processes.

1.5.1 Lumped versus Distributed Models

As defined by the semantics of the words, the distributed models treat the contents of metabolic systems as autonomous entities while the lumped models integrate them into a single homogenous entity. As a rule, lumping distributed effects together is allowed only if the representation satisfactorily describes the dynamics of a metabolic system. For this purpose, the notion of compartment is introduced which constitutes an idealized store of a substance.

In lumped models, active material within a compartment is considered well-mixed and kinetically homogenous. Well-mixed means distribution uniformity i.e. the same concentration yields are obtained by any two samples taken from anywhere in a compartment (Figure 1-1). Thus, the concept of well-mixed relates to uniformity of information contained in a single compartment. Kinetic homogeneity, on the other hand, means that all particles have the same chances to leave the compartment. Since the reasons for leaving a compartment in metabolic systems are related to transport and utilization, kinetic homogeneity implies that all particles in the compartment have the same probability of leaving due to one of these events.

Mathematically, these conceptual assumptions lead to ordinary differential equations which for modeling a metabolic system are usually of the first order. This set of first order differential equations are better known as *state variable* equations, where the mass of metabolite/substance in each of the compartments is a state variable. State variable equations can be expressed compactly by the use of vector notation:

$$\dot{\mathbf{x}} = \mathbf{f}[\mathbf{x}(t, \mathbf{p}), \mathbf{u}(t), t; \mathbf{p}], \quad \mathbf{x}_0 = \mathbf{x}(t_0, \mathbf{p}), \quad (1.1)$$

$$\mathbf{y} = \mathbf{g}[\mathbf{x}(t, \mathbf{p}); \mathbf{p}], \quad t_0 \leq t \leq t_{\max}, \quad (1.2)$$

where \mathbf{x} is the vector of n_x state variables (e.g. mass or concentration), the dot ($\dot{\cdot}$) represents the time derivative $d\mathbf{x}/dt$; \mathbf{u} denotes the n_u inputs vector, \mathbf{y} is the vector of n_y observed outputs or measurement variables; $t \in [t_0, t_{\max}]$ denotes the observation interval; \mathbf{p} is the vector of n_p unknown constant parameters (e.g. transport parameters); \mathbf{f} defines the general model structure based on *a priori* physicochemical knowledge relating the state variable vector \mathbf{x} to the input vector \mathbf{u} where the values of the relationships are specified by the parameters \mathbf{p} ; \mathbf{g} defines the measurement process, that is it relates the output vector \mathbf{y} with to the state variable vector \mathbf{x} where, again, the specific relationships are defined by the appropriate parameters in the vector \mathbf{p} . If the parameters \mathbf{p} are constant, the system is said to be *time invariant*. The mathematical representation (1.1), (1.2) is a fairly comprehensive one that allows both the dynamics of the metabolic control systems (e.g. hormones) as well as the kinetics of individual substances (e.g. drugs) in the organism to be described.

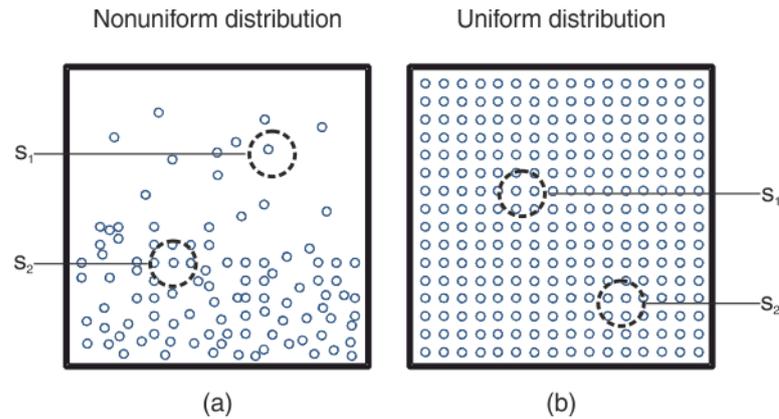


Figure 1-1 The “well-mixed” assumption. (a) The concentration depends on where the sample is taken from the compartment ($s_1 \neq s_2$). (b) The concentration does not depend on where the sample is taken from the compartment ($s_1 = s_2$)

In many cases the representation of metabolic systems by lumped parameter compartmental models is adequate. This assumes that there is a homogeneous distribution of material within a distribution space. In distribution models no such assumptions of homogeneity are made. This leads to mathematical formulation in terms of partial differential equations. A general introductory framework for modeling the kinetics of metabolic processes in distributed systems is given in [30].

1.5.2 Deterministic versus stochastic models

The majority of mathematical models formulated to describe metabolic processes are deterministic. Actually, deterministic models and input/output identification experiments have been and are widely and successfully used to provide a quantitative description many metabolic systems. However, in cellular and subcellular levels where random effects occur, a purely deterministic model is fairly adequate.

In order to include probabilistic effects that evolve over time, stochastic processes are incorporated into lumped compartmental models of metabolic systems. Intrinsic stochasticity can be integrated either into the state variables corresponding to the various compartments or into the transfer rate parameters of substances within the system. The two approaches may also be combined yielding a general framework for the stochastic modeling of compartmental metabolic systems.

1.5.3 Linear versus non-linear models

Fortunately, in many instances, the adoption of a linear time-invariant dynamic model for a metabolic system is completely adequate. In such linear models the state variables and their derivatives appear in linear combination only, and as a consequence the superposition theorem applies. Thus, the total response to several inputs is the sum of the responses to the individual inputs. Linear models are adequate when the intrinsic dynamics of the metabolic system are essentially linear, even when the applied test signals result in large perturbations of system variables.

In such linear models the equations (1) and (2) are converted to

$$\dot{\mathbf{x}} = \mathbf{A}(\mathbf{p})\mathbf{x}(t) + \mathbf{B}(\mathbf{p})\mathbf{u}(t), \quad \mathbf{x}_0 = \mathbf{x}(t_0, \mathbf{p}), \quad (1.3)$$

$$\mathbf{y} = \mathbf{C}(\mathbf{p})\mathbf{x}(t), \quad t_0 \leq t \leq t_{\max}, \quad (1.4)$$

where \mathbf{A} , \mathbf{B} and \mathbf{C} are constant coefficient matrices of suitable dimensions.

The linear model is also adequate in cases where linearization can be carried out. For example, when only small or negligible perturbation about the operation point occurs, then the dynamic equations are linear even though the intrinsic dynamics of such a system is nonlinear.

In contrast small perturbation about the operation point, frequently, reveals little information concerning the behavior of the system. To analyze the behavior of the system away from its steady state it is necessary large perturbation signals to be applied. In such cases linear models are generally invalid unless the nature of the system is linear. So as nonlinear models to be accurately solved complicated numerical techniques are required.

1.6 COMPARTMENTAL MODELS

As already indicated, the main goal of a mathematical model in biosciences should be to create, to the extent feasible, a simplified representation of the system under consideration. In the context of metabolic processes and tracer kinetics, lumped deterministic models are particularly apposite. This occurs because this form of models allows the clustering of the spatially distributed similar information of a biological system in a topology of discrete compartments of identical behavior. Consequently, the reduction of an, initially, complex physiologic system into a finite, tractable, number of entities and pathways is manageable, through the notion of compartment.

By definition, if a substance is present in a biological system in several distinguishable locations, then all the substance in a particular form or all the substance in a particular location, or all the substance in a particular form and location are said to constitute a compartment [26]. Regularly, a compartmental model consists of a finite number of compartments, the number of which depends on both the system being studied and the richness of the experimental configuration. Each compartment “communicates” with adjacent ones through specific interconnections. These interconnections represent material flux which may depict, in turn, transport from one location to another, chemical transformation, or both.

In order to account for the amount of a substance in a compartmental model, the model itself must fulfill some conservation law. Accordingly, the amount of substance that can be stored in a compartment is equal to the difference between what flows in and what flows out of the compartment. The general form of the mass balance equation used to describe this function is as follows. Let Q_i be the quantity of substance in compartment i that is interchanged with j other compartments in the system at rates R_i , then:

$$\frac{dQ_i}{dt} = R_{i0} + \sum_{\substack{j=1 \\ j \neq i}}^n R_{ij} - \sum_{\substack{j=1 \\ j \neq i}}^n R_{ji} - R_{0i}, \quad i = 1, 2, \dots, n, \quad (1.5)$$

where the sum of R_{ij} 's denotes the sum of all inward fluxes of Q from j compartments to compartment i , the sum of R_{ji} 's is the sum of all outward fluxes of Q from compartment i to j compartments, n is the number of compartments and R_{i0} and R_{0i} is the flux of material into compartment i from external environment and vice versa, respectively. Here, it must be not-

ed that the mentioned fluxes cannot be negative. Also, if Q is divided by V , the volume of the compartment, then the chemical concentration of material is calculated.

It should be emphasized, that flows in and out of a compartment cannot be treated as independent components because it is highly possible that they will be somehow interlinked. For example, the amount of material that enters or exits a compartment may depend on the compartment's volume. Similarly, in multi-compartmental models, material flow into a compartment may be dependent on its outflow from another, or several, compartments. Mathematically, this means that the state-variables of the model may not only depend on what flows in and out, but also on each other and on the state of the system as a whole. These concepts are briefly described in the following sub-sections.

1.6.1 Functional schemas of material flux

After setting the suitable mass balance equation(s), the functional dependences of the fluxes that are embedded on each compartment are specified. As we have already mentioned, these fluxes are regulated only through the interaction of chemical reaction, storage, and transport. The nature of the dependences along with their values can be derived from preliminary knowledge or from parameter estimation. Hence, these functional dependences may be markedly difficult to establish mathematically. For example, material kinetics embedded in the functionality of a compartment is naturally non-uniformly distributed in several parts of the organism which, in addition, have complex geometries. Therefore, only approximate representations are incorporated in compartmental models. Essentially, the approximate schemas available to describe functional dependence of the material fluxes can be narrowed down to two; linear and non-linear schemas.

1.6.1.1 Linear dependence

In the simplest case, the rate of transfer can be governed by a linear relationship i.e.:

$$R_{ij} = k_{ij}Q_j \quad (1.6)$$

where k_{ij} is a constant that defines the fractional rate of transfer of material into compartment i from compartment j (Figure 1-2(a)). A principal mechanism that can be described by equation (1.6) is that of a first-order chemical reaction of one reactant. Specifically, let $A_o \rightarrow \text{products}$ represent the elimination reaction of substance A , then:

$$\frac{dC_A}{dt} = -kC_A^o \quad (1.7)$$

denotes the elimination flux of substance A from the compartment, where $C = Q/V$ is the concentration of material in a compartment of constant volume V and the exponent o is the order of the reaction which is equal to the unit in first-order chemical reactions.

Another principal mechanism that equation (1.6) is employed is material diffusion. Theoretically, diffusion of a substance, takes place from regions where its chemical concentration is high to areas of low concentration. Though, the diffusion process is expressed through concentrations, densities and their derivatives a finite difference approximation is regularly employed. For a simple, two-compartmental model, where the locations of different concentration are separate compartments of equal volume, this reads:

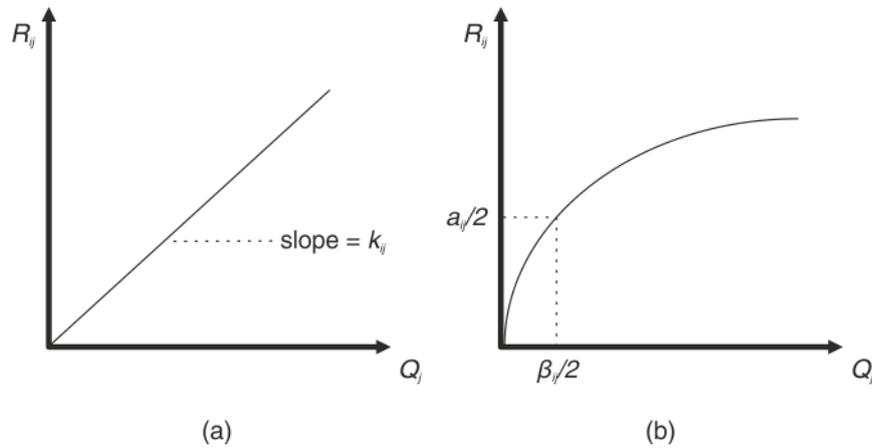


Figure 1-2 Most common approximate schemas to describe functional dependence of the material fluxes: (a) linear kinetics and (b) Michaelis-Menten kinetics.

$$\begin{aligned}
 \text{for compartment 1: } & \frac{dQ_1}{dt} = k_{21}(C_2 - C_1) \quad \text{and} \\
 \text{for compartment 2: } & \frac{dQ_2}{dt} = k_{12}(C_1 - C_2).
 \end{aligned} \tag{1.8}$$

These equations are a special case of the Fick's Law of diffusion that is discussed in more detail in chapter 4.

1.6.1.2 Non-linear dependence

One commonly occurring nonlinear type of functional dependence of material flux can be specified by Michaelis–Menten dynamics:

$$R_{ij} = \frac{a_{ij}Q_j}{\beta_{ij} + Q_j} \tag{1.9}$$

where a_{ij} is the saturation value of flux R_{ij} and β_{ij} is the value of Q_j at which R_{ij} is equal to half its maximal value (Figure 1-2(b)). This type of dynamics has been initially used for approximating the saturation rate of enzymatic reaction mechanisms. Nevertheless, Michaelis–Menten kinetics has been extensively employed in the description of several saturating fluxes such as ionic fluxes in biological channels, cell pumping mechanisms etc. [31]–[35].

Another mechanism whose flux is described by a nonlinear schema is that of a chemical reaction of several substances. According to the mass action law, the rate of reaction of a substance undergoing a chemical transformation in a constant volume is proportional to its concentration (equation (1.7)) and, if several substances take part in the reaction, the rate is proportional to the product of the concentrations. Specifically, let: $A + B \rightarrow C$ denote a first-order irreversible reaction of two reactants and one product, then the production rate of quantity C in the respective compartment is equal to the flux:

$$\frac{dC_C}{dt} = kC_A C_B, \tag{1.10}$$

where k is the rate constant of production of substance C .

Generally, there are several other nonlinear relations that can be employed for modeling metabolic systems and tracer kinetics e.g. sigmoidal, Langmuir, etc. The most suitable one always depends on the detail that the modeler wishes to use for describing the system's dynamics.

1.6.2 Strictly compartmental and control system models

A class of lumped parameter deterministic models that do not allow active control of the material flux that enters and/or exits a compartment is the class of *strictly compartmental models*. In this modeling sub-class inter-compartmental fluxes are assumed to depend, either linearly or nonlinearly, only on the amount of material in the source compartment. That is:

$$R_{ij} = R_{ij}(Q_j) \quad (1.11)$$

On the other hand, the compartmental models discussed thus far involve no control. It is however widely accepted that every biological process is susceptible to some sort of control. This means that model properties such as geometrical (volume), physical (diffusion rate) or chemical (production rate) may change, in time, under the influence of extrinsic or intrinsic variables applied on the system quantities. In this case, the flux is always a nonlinear function of variables:

$$R_{ij} = R_{ij}(Q_j; Q_a, Q_b, \dots) \quad (1.12)$$

where Q_a, Q_b, \dots specify the control variables regulating the flux R_{ij} of substance Q_j . These control variables may represent complexities such as stochastic phenomena, distributed effects, etc. and are, again, obtained based on preliminary knowledge or parameter estimation.

One class of models that is widely used in interpreting the behavior of control systems commonly occurring in metabolism is the one that is deterministic and lumped. Such models can be used to represent the dynamics of metabolic substrates and their regulating enzymes, the dynamics of hormones, and the interactions of hormones, enzymes and metabolic substrates. Thus, they may involve the use of the class of strictly compartmental models to describe some of their subsystems.

1.6.3 Steady state and perturbation schemes

Whatever type of compartmental model is employed, it is always possible to assume that dynamic equilibrium holds between flows in a compartment. In fact, many metabolic systems exhibit a constant steady state. At steady state, the quantities in each compartment remain at essentially constant values with the entering and exiting fluxes being essentially equal. From equation (1.5) this reads:

$$\frac{dQ_i}{dt} = 0 \Leftrightarrow R_{i0} + \sum_{\substack{j=1 \\ j \neq i}}^n R_{ij} = \sum_{\substack{j=1 \\ j \neq i}}^n R_{ji} + R_{0i}, \quad i = 1, 2, \dots, n \quad (1.13)$$

The system of equation (1.13) is consistent when there are enough measurements of observable variables from which all the parameters can be uniquely estimated. It is however common that the functional schemas that are used to describe the flow of a substance in a compartment, are insufficient to provide a full description of the dynamics of the system, re-

ardless of their degree of linearity. Hence, complementary information must be acquired in order to determine the functional dependences between fluxes and materials.

To obtain the best possible characterization of the dynamics of the system in practice, appropriate test input signals are employed and the respective output data are measured through observable variables. To this purpose, the system is perturbed through a tracer experiment. This concerns the application of the same substance (tracer) as the one in the compartment of unknown dynamics (tracee). According to equation (5) the mass balance equations for the tracer, assuming perfect mixing in the compartment, are of the form:

$$\frac{dq_i}{dt} = r_{i0} + \sum_{\substack{j=1 \\ j \neq i}}^n r_{ij} - \sum_{\substack{j=1 \\ j \neq i}}^n r_{ji} - r_{0i}, \quad i = 1, 2, \dots, n, \quad (1.14)$$

where q is the tracer material (similar to Q that is resident in compartment i) and r denotes the flux rate of the tracer in the system. Assuming steady-state of the system at all times the flux of the tracer can be written as:

$$r_{ij} = \frac{R_{ijs}}{Q_{is}} q_i(t) = k_{ij} q_i(t), \quad (1.15)$$

where R_{ijs} and Q_{is} are the steady-state values of R_{ij} and Q_i respectively. Then, since during the experiment the course of the tracer kinetics is recorded, the relationship of R and Q can be approximated by fitting equation (1.15) to the measured data. Obviously, the best case scenario is when R and Q i.e. the system is always at steady state. However, if these are time varying and/or controlled by other processes this procedure cannot readily apply. To do this under such circumstances, steady state is again assumed but information can only be acquired for those compartments where the tracer can reach.

Another way to obtain information of the functional dependences is through small-signal perturbation. This procedure assumes that the unknown fluxes can be expanded in Taylor series for about their steady state values. Then, considering only the linear term of the expansion the derivative terms $dR_{ij}/dQ_j|_{\text{steady-state}}$ can be used in order to obtain an approximation of the functional dependences.

In any case, a single tracer experiment cannot guarantee that the general form of the functional schema of R_{ij} can be identified. It is possible under certain assumptions but usually, only through repetition of tracer experiments can this be provided with confidence.

1.6.4 Non – compartmental analysis

As already discussed, the mathematic representation of metabolic systems by strictly compartmental and control models depends on the *a priori* available theoretical knowledge. These types of models are powerful investigative tools which have been successfully used to estimate physical and chemical quantities not directly available to measurement. However, these modeling approaches are prone to identification and validation problems due to several reasons; either because of improper aggregations of complex physicochemical processes, inadequate theoretical knowledge, inaccurate definition of the model structure or insufficient data availability. In effect, empirical or empirical-theoretical approaches, better known as non-compartmental ones, are employed, instead.

Typically, the only difference between compartmental and non-compartmental analysis of a system is the way with which the inaccessible portion of the system is described. Thus, in non-compartmental approaches the only compartments taken into account are those that are directly connected to observable variables. This means that fewer assumptions are necessary than in compartmental models. For this, models of this type give emphasis on behavioral rather than structural/mechanistic properties of the system.

There are three properties of the metabolic systems that can be employed in non-compartmental analysis. First, is the constant total clearance/elimination rate of a substance from the system which is equivalent to equation (1.7). Second, is the concept of dose proportionality, where the amount of this substance is related to input rate of a substance by the convolution integral equation:

$$Q(t) = (h * R_a) = \int_0^t R_a(\tau) h(t - \tau) d\tau, \quad (1.16)$$

where, h is the unit impulse response of the system after the instantaneous input of the substance and R_a is the rate of appearance the substance. Third, is the existence of a terminal monoexponential phase in the substance's systemic concentration time course:

$$Q(t) = Q_z e^{-\lambda_z t}, \quad t > t_z, \quad (1.17)$$

where Q_z is the intercept value of the amount of substance Q , λ_z is a rate constant and t_z is the threshold time after which the quantity of the substance can be described by equation (1.17).

Using these properties, several quantities can be estimated; e.g. the rates of production, utilization and clearance, residence times, volumes of distribution, etc. It essential to keep in mind, though, that not all metabolic systems adhere to the properties mentioned above because their use is, still, an approximation.

1.7 MODEL IDENTIFICATION

Aforementioned, the structure of the model and/or its parameters is highly possible to be uncertain, regardless being linear or nonlinear. Structural uncertainties concerning the structure of the model and its parameters may rise as the result of questionable conceptual assumptions, incomplete background biophysical and biochemical knowledge, or insufficient observable data. Therefore, having successfully postulated a suitable model structure the question that should arise is the following: "*To what extent can the unknown model parameters be uniquely estimated from the available information encapsulated in the data using this uncertain model structure?*"

To answer this question, techniques concerned to challenge the predictive power of mathematical models of dynamical systems, used in the field of system identification, are employed. The purpose of these techniques is to infer whether the model parameter values can be accurately and uniquely reproduced from the proposed model structure using measured data. Recall, here, that the model structure can originate from background theoretical knowledge or from empirical or empirico-theoretical approaches. The former approaches search to define rigorous mathematical representations of the system based on physico-chemical laws, while the latter attempt to find some structure, not necessarily mirroring the underlying processes, that fits the experimental data adequately. Then, two general forms of

mathematical realizations can be derived onto which system identification from experimental data can be applied. The first form, which is derived on the basis of background knowledge approaches, is similar to the nonlinear state-space form of equation (1.1) i.e.:

$$\dot{\mathbf{x}}(t, \mathbf{p}) = \mathbf{f}[\mathbf{x}(t, \mathbf{p}), u(t), t; \mathbf{p}], \quad \mathbf{x}_0 = \mathbf{x}(t_0, \mathbf{p}), \quad (1.18)$$

$$\mathbf{y}(t, \mathbf{p}) = \mathbf{g}[\mathbf{x}(t, \mathbf{p}); \mathbf{p}], \quad (1.19)$$

$$\mathbf{h}[\mathbf{x}(t, \mathbf{p}), \mathbf{u}(t, \mathbf{p}), \mathbf{p}] \leq 0, \quad (1.20)$$

where \mathbf{h} represents all the equality and inequality constraints related with the state variables (\mathbf{x}), the inputs of the system (\mathbf{u}) and the model parameters (\mathbf{p}). The second form is an unspecified that relates the input/output data through a behavioral model:

$$\mathbf{y}(t, \mathbf{p}) = \mathbf{w}[\mathbf{u}(t); \mathbf{p}], \quad (1.21)$$

where \mathbf{w} is a functional that relates mathematically the measured output (\mathbf{y}) to the system input(s) and the undetermined postulated model parameters.

Having decided a suitable model structure that utilizes either of the general forms mentioned above, two distinct analyses must then be performed according to control engineering theory. First, is to investigate the solvability of the model, under ideal conditions of noise-free data and perfect model structure. This analysis is better known as theoretical or *a priori* or structural identifiability analysis. Second, is to further investigate under real experimental conditions the extent to which the postulated parameters can be uniquely reproduced. This second analysis is better known as practical or *a posteriori* identifiability analysis. Prior to, during and after the performance of these analyses, it is common that new data may be needed. For this, new experiments may have to be designed. This complementary step is known as optimal experimental design. Its purpose is to provide optimal experimental information regarding the modeled system that will increase the confidence of the identification results. In the next sections a brief description of these concepts is given.

1.7.1 Structural (a priori) identifiability

According to the state-space description given above, the *a priori* identifiability concerns the existence of a unique solution for \mathbf{p} under an idealized environment where the data are noise-free and the model structure is perfect and complete. It is essential to stress here that structural identifiability is related only to the model structure and possibly to the type of excitation utilized and, therefore, is independent of the values of the model parameters.

In order to explain the general structure of the identifiability problem, we assume that every time an experiment is realized, a set of initial conditions \mathbf{x}_0^* , an excitation signal \mathbf{u}^* and some observations \mathbf{y}^* are specified. Then, the model response to these experimental conditions constitutes a solution of the structural equation (similar to equation (1.18)):

$$\mathbf{F}[\mathbf{x}, \dot{\mathbf{x}}, \dots, \mathbf{u}^*, t; \mathbf{p}] = 0, \quad \mathbf{x}_0^* = \mathbf{x}(t_0, \mathbf{p}), \quad (1.22)$$

where \mathbf{F} may be a function of derivatives. It is possible, however, that this response cannot enclose information for all of the postulated model parameters, thus equation (16) reads:

$$\mathbf{y}^* = \mathbf{G}[\mathbf{x}, \mathbf{u}^*, t; \Phi], \quad (1.23)$$

where, Φ is a vector of the parameters that can actually appear in the experimental observations. Suppose now that three types of parameters may be generally distinguished: the structural (\mathbf{p}_s), the experimental (\mathbf{p}_e) and the observational (Φ) ones [19]. The structural parameters are the model parameters that inherent the structural equations. As said, not all of the structural parameters appear in the observations. The experimental parameters, on the other hand, correspond to those parameters that the conditions of the experiment i.e. the initial conditions and the excitation signal, impose on the equation system. Cardinally, the excitation input and the initial conditions are fully specified in biological experiments and, for this, the experimental parameters are usually not considered. However, for completeness we must state that the set which is created by joining the vectors \mathbf{p}_s and \mathbf{p}_e is, by definition, the vector \mathbf{p} i.e. $\mathbf{p} = [\mathbf{p}_s^T \mathbf{p}_e^T]^T$.

Finally, the observational parameters ($\Phi = [\varphi_1 \varphi_2 \dots \varphi_k]^T$) denote the uniquely determinable parameters that are actually obtained from the equations of the observations (equation (1.19)). This does not necessarily imply that $\Phi \in \mathbf{p}$. In fact, these observational parameters are algebraic functions of the structural and experimental parameters that appear on the observations i.e. $\Phi = \Phi(\mathbf{p})$. Hence, from every input/output experiment a particular value is provided which means that $\Phi^* = \Phi(\mathbf{p}^*)$. Again, we have to note that, in most cases, not all of the presupposed parameters are passed on the observations i.e. $\mathbf{p}^* \subseteq \mathbf{p}$. Therefore, the identifiability problem can be also stated as: “*Are the structural and experimental (if any) parameters uniquely determined by the observational parameters?*”

Under such framework, a parameter p_i is considered *a priori*:

- *globally identifiable* if and only if, for almost any the system:

$$\mathbf{y}(\Phi(\mathbf{p}), t) = \mathbf{y}(\Phi^*, t) \quad (1.24)$$

has the only one solution $\mathbf{p} = \mathbf{p}^*$;

- *locally (or nonuniquely) identifiable* if and only if, for almost any $\mathbf{p}^* \in \mathbf{p}$ the system (1.21) has for p_i more than one, but a finite number of solutions;
- *nonidentifiable* if and only if, for almost any $\mathbf{p}^* \in \mathbf{p}$ the system (1.21) has for p_i an infinite number of solutions.

Thus, to study the *a priori* global identifiability of \mathbf{p} , there is the need to solve the system of (possible) nonlinear algebraic equations

$$\Phi^* = \Phi(\mathbf{p}) \quad (1.25)$$

where Φ^* are known. Equation (1.22) can also be referred as the *exhaustive summary* of the model [28]. If some equality constraints (equation (1.20)) are present, these need to be added to the algebraic equations to form the exhaustive summary.

Obviously, the theoretical analysis of *a priori* identifiability requires the algebraic solution of equation (1.25) i.e. the factorization of the model's output (equation(1.19)) until an analytical input/output relation has been extracted. Finding, however, an exact symbolic solution for equation (1.25) is almost impossible for complex models in biosciences. For this, there are several alternative methods that explore the identifiability of model structures which can be categorized into those methods which are applied to the linear class of models and those which are applied to the non-linear ones. Structural identifiability for linear models has

been extensively addressed and several global methods have been documented and automated [36]. Asserting, however, *a priori* identifiability to nonlinear models is a difficult task that is proportional to the model's complexity. For this, several sophisticated mathematical approaches have been proposed based on expanding the observations equation (equation (1.19)). Nevertheless, global identifiability is usually a very difficult mathematical problem to solve and accurate solution can only be obtained only for some type of models, usually linear ones. For this, local identifiability methods have been argued for use instead [37]. In the following sub-sections some of the methods of structural model identifiability, which are mostly used in literature, are briefly described.

1.7.1.1 Solving the structural (*a priori*) identifiability of linear, deterministic models

For the linear, time invariant class models which are commonly encountered in computational biosciences, recall equations (1.3) and (1.4) which for convenience are repeated here:

$$\dot{\mathbf{x}} = \mathbf{A}(\mathbf{p})\mathbf{x} + \mathbf{B}(\mathbf{p})\mathbf{u}, \quad \mathbf{x}_0 = \mathbf{x}(t_0, \mathbf{p}), \quad (1.26)$$

$$\mathbf{y} = \mathbf{C}(\mathbf{p})\mathbf{x}, \quad (1.27)$$

where \mathbf{A} , \mathbf{B} and \mathbf{C} are stationary matrices, which are functions of the parameter vector \mathbf{p} . In general, \mathbf{B} and \mathbf{C} are known *a priori* although the latter may sometimes include unknown parameters. Therefore, the goal according to equation (1.25) is to find if the nonzero elements of \mathbf{A} can be uniquely identified by the observations in \mathbf{y} . Three of the most employed methods that deal with the extraction of this type of inferences, namely the transfer function or impulse response method, the similarity transformation method and the modal matrix and eigenvalues method, are described next. More details about these methods and others can be found in [19], [38].

The transfer function or impulse response method relies on the determination of an error-free impulse response of the system under identification. Using the Laplace transform of equations (1.26) and of (1.27) and accepting zero initial state for convenience ($\mathbf{x}_0 = \mathbf{0}$), the observation vector \mathbf{y} becomes, after some matrix manipulation:

$$\mathbf{Y} = \mathbf{H}(s, \mathbf{p})\mathbf{u}, \quad \mathbf{H}(s, \mathbf{p}) = \mathbf{C}(\mathbf{p})(s\mathbf{I} - \mathbf{A}(\mathbf{p}))^{-1} \mathbf{B}(\mathbf{p}), \quad \mathbf{x}_0 = \mathbf{0}, \quad (1.28)$$

where \mathbf{H} is the transfer function of the experiment, s is the Laplace variable and \mathbf{I} the identity matrix. Each element of the transfer function \mathbf{H}_{ij} corresponds to the Laplace transform of the response in the measured observable variable at port i , $\mathbf{y}_i(t, \mathbf{p})$, to a unit impulse test input at port j , $\mathbf{u}_j(t) = \delta(t)$, where δ is the Dirac delta function. In other words, each element of \mathbf{H}_{ij} , corresponds to an experiment performed on the system, between input j and output i . Then, the problem becomes to determine whether the nonzero elements of \mathbf{A} can be uniquely identified from the coefficients of the numerator and denominator polynomials of the transfer matrix $\mathbf{H}_{ij}(s, \mathbf{p})$ that act as the known observational parameters in Φ . If a unique solution for a parameter p_i exists, then this parameter is globally identified; if a finite number of solutions exist, p_i is not globally identified, but locally; and if an infinite number of solutions exist, parameter p_i is unidentifiable through the proposed model structure.

If the solutions are difficult to obtain using the transfer function, the similarity transformation method is another approach to the solution of the identifiability problem. In this method matrix \mathbf{A} is suitably transformed via similarity transformation i.e. $\mathbf{A} \rightarrow \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$, where \mathbf{P} is a non-singular matrix. Then, the Laplace transform of the transformed model becomes:

$$\mathbf{Y} = \mathbf{C}(\mathbf{p})\mathbf{P}^{-1}(\mathbf{s}\mathbf{I} - \mathbf{A}(\mathbf{p}))^{-1}\mathbf{P}\mathbf{B}(\mathbf{p}), \quad \mathbf{x}_0 = \mathbf{0}. \quad (1.29)$$

The transformation $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ must satisfy all the model constraints and also for the input-output response to remain unchanged the relations $\mathbf{C}\mathbf{P}^{-1}=\mathbf{C}$ and $\mathbf{P}\mathbf{B}=\mathbf{B}$ must be fulfilled. If after applying these conditions to determine the elements of \mathbf{P} , $\mathbf{P}=\mathbf{I}$ is found, then the model is globally identifiable.

Similar to the similarity transformation is another well-known method of global identification of model parameters that utilizes the modal matrix and eigenvalues of \mathbf{A} . Specifically; let all the elements of \mathbf{A} be able to be mapped to independent components of eigenvectors and eigenvalues. Then, since $\mathbf{A}=\mathbf{M}\mathbf{\Lambda}\mathbf{M}^{-1}$, where \mathbf{M} is the modal matrix and $\mathbf{\Lambda}$ the diagonal matrix of eigenvalues, the impulse response of the system can be written as:

$$\mathbf{y} = \mathbf{C}e^{\mathbf{M}\mathbf{\Lambda}\mathbf{M}^{-1}t}\mathbf{B} = \mathbf{C}\mathbf{M}e^{\mathbf{\Lambda}t}\mathbf{M}^{-1}\mathbf{B}. \quad (1.30)$$

Therefore, the problem reduces to specifying if the elements of \mathbf{M} and $\mathbf{\Lambda}$ can be uniquely identified from the observations \mathbf{y} subject to the constraint on the \mathbf{A}_{ij} elements.

1.7.1.2 Solving the structural (*a priori*) identifiability of non-linear, deterministic models

As said, for the nonlinear class of models, several sophisticated mathematical approaches have been proposed [39] (and references within). However, for inferring the structural identifiability there is no global method amenable to every nonlinear model. Limitations in applying each method may arise either from the size of the system i.e. the number of state variables, parameters and observables, or the size of its non-linearity. However, it is straightforward that the general principle behind any method that exploits the identifiability of a non-linear model is that of relaxing the model's non-linearities to the extent that its exhaustive summary (equation (1.25)) could then be better manipulated.

The Taylor series expansion method accepts that the observations are unique analytical functions of time which, in turn, produce infinite number of unique time derivatives. The idea, then, is to establish a system of algebraic equations on the parameters and check it for unique solutions. Specifically, provided that the state variables (\mathbf{x}), the inputs (\mathbf{u}), the outputs (\mathbf{y}) and the functions \mathbf{f} and \mathbf{g} have infinitely many derivatives, it is possible to represent the equation (1.19) as a Taylor series expansion around the initial state \mathbf{x}_0 :

$$y_i(t, \mathbf{p}) = y_i(t_0, \mathbf{p}) + t \dot{y}_i(t_0, \mathbf{p}) + \frac{t^2}{2!} \ddot{y}_i(t_0, \mathbf{p}) + \dots, \quad (1.31)$$

where $i = 1, 2, \dots, n_y$ denotes the number of observations in \mathbf{y} . Considering each of the time derivatives to be an observable variable in Φ , the exhaustive summary of the model according to equation (1.25) is:

$$\Phi^* = y_i^{(d.o.)}(t_0, \mathbf{p}^*), \quad (1.32)$$

where $d.o. = 0, 1, 2, \dots, d.o._{\max}$ is the derivative order and $d.o._{\max}$ is equal to the smallest positive integer, such that the symbolic computations give the solution of the parameters. The uniqueness of solution of equation (1.32) guarantees that the model is structurally identifiable. However, the popularity of the method is limited because equation (1.32) is, on many occasions, very complex to solve and because theory does not provide *a priori* the number of derivatives required for one in order to study the identifiability of a given model and experiment.

Several other methods on structural identifiability of non-linear models have been proposed. Indicatively, apart from the Taylor series expansion described earlier, usage of the conceptually similar generating series method has been also proposed. The method concerns the expansion of the observables in series and then taking the Lie derivatives of the latter. Then, the exhaustive summary includes the series coefficients and the Lie derivatives [39]. Its uniqueness means that the model is structurally globally identifiable. To visualize the results of the generating series method identifiability tableaux has been introduced. Another approach to the problem is the similarity transform which based in the local state isomorphism theorem seeks state variable transformations that leave invariant the stimuli-observables map and the structure of the system [40]. Finally, differential algebra approaches have been introduced [23]. The algorithm is based on replacing the stimuli-observables behavior by some polynomial or rational mapping by utilizing a differential algebra algorithm. Then, the Buchberger algorithm is used to solve the resulting set of nonlinear algebraic equations of the exhaustive summary. Probably, the algorithm with the most applicability domain that uses a differential algebra approach is the one proposed in. Based on the Ritt algorithm for differential algebra, The algorithm is capable of handling many features arising in biological system models, including zero initial conditions and time-varying parameters.

1.7.1.3 Structural (*a priori*) local identifiability

It is obvious from the discussion thus far, that the existing deterministic techniques for inferring (global) structural identifiability have limited applicability. However, the trend in scientific research is to push towards the creation of complicated, integrative, models that describe biological activities in detail. Unfortunately, the majority of these models are often so mathematically complex that no approach can be readily employed without radical alterations in the model's structure that may eventually alter the model's profile. Consequently, there is a practical need for structural identifiability approaches that provide some level of assurance that the developed complex models can be predictive. Therefore, instead of rigorous deterministic approaches for global identifiability, local ones have been argued for use despite their apparent limitations derived from their local nature [37].

One of the standard tools available for studying the local identifiability of a dynamic model is sensitivity analysis [19], [37], [41]. In particular, the calculation, with respect to time, of the sensitivity of the model's output to the variations of each of the postulated parameters leads to the construction of each of the parameters' output sensitivity functions. The latter, in turn, allow one to infer, on a local scale, whether the respective model parameters are identifiable. The method is as follows: first, the partial derivatives, $\partial x_i / \partial p_j$ which is the partial derivative of the i th state ($i=1 \dots n_x$) to the j th parameter ($j=1 \dots n_p$); $\partial \mathbf{f} / \partial \mathbf{x}$ which is the Jacobian matrix; and $\partial \mathbf{f} / \partial \mathbf{p}$ which is the parametric Jacobian matrix, are numerically calculated. Then, by assuming that the measured n_y responses in \mathbf{y}^* are a subset of the n_x states in \mathbf{x} ($\mathbf{y}^* = \mathbf{C}\mathbf{x}$) the sensitivity matrix is constructed i.e.

$$\dot{\mathbf{S}} = \frac{d}{dt} \frac{\partial \hat{\mathbf{y}}}{\partial \mathbf{p}} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \frac{\partial \mathbf{x}_i}{\partial p_j} + \frac{\partial \mathbf{f}}{\partial p_j}, \quad (1.33)$$

which is a $(n_y \times n_x)$ matrix. In theory, sensitivity analysis is practiced on static systems but its application on dynamic ones can be easily extended provided that steady state is assumed at all times. Hence, the dynamic sensitivity matrix will be naturally extracted from (30):

$$\dot{\mathbf{S}}_D = \begin{bmatrix} \left. \frac{\partial \hat{\mathbf{y}}}{\partial p_1} \right|_{t_1} & \dots & \left. \frac{\partial \hat{\mathbf{y}}}{\partial p_j} \right|_{t_1} \\ \vdots & \ddots & \vdots \\ \left. \frac{\partial \hat{\mathbf{y}}}{\partial p_1} \right|_{t_s} & \dots & \left. \frac{\partial \hat{\mathbf{y}}}{\partial p_j} \right|_{t_s} \end{bmatrix}. \quad (1.34)$$

Each column of \mathbf{S}_D corresponds to the sensitivity function of each of the n_p parameters. Finally, the correlation matrix \mathbf{M} of \mathbf{S}_D is calculated. The parameters with correlations with all other parameters between -1 and +1 are locally identifiable while, the parameters that have at least one correlation of exactly -1 or +1 with another parameter are unidentifiable. This means that the latter parameters influence the observables in exactly the same or exactly the opposite manner. The original parameter set, \mathbf{p} can be reduced to the identifiable parameter set, \mathbf{p}_I , of length l , by repeatedly calculating \mathbf{M} and removing one unidentifiable parameter each time, until no more unidentifiable parameters remain.

It is essential to note here that this method of *a priori* local identifiability is a numerical method and bases itself on the corresponding sensitivity functions. For this, the sensitivity matrix needs be calculated to a large enough number of time points. In addition and in order to stress this further, we repeat that such an analysis is performed locally and, thus, its outcomes do not concern the model as a whole, rather the nominal point at which they were calculated.

1.7.2 Global ranking of the parameters

By the end of the theoretical analyses of parameter structural identifiability two possibilities may arise. The first is that for the particular postulated model structure and experimental data, the model parameters are all uniquely identifiable. Then, parameter estimation can be readily performed, as it will be discussed in the next paragraph. The second possibility is that some of the parameters may have been found uniquely identifiable, some locally identifiable and/or some unidentifiable. In this case, restructuring of the model is an option. However, model rejection and reconstruction may be inefficient and/or require the execution of complex/unfeasible experiments. All the same, another, suitable, option is to reduce the model's complexity by reconfiguring its parameters through their global ranking.

Theoretically, parameter ranking designates which parameters are the most relevant to the model's output(s) and is usually used as a complementary analysis to structural identifiability [39], [42], [43]. In order to quantify the dependence of the system's behavior on the parameters that affect the process dynamics, sensitivity analysis is, again, employed. Specifically, there are two types of sensitivity analysis, local and global sensitivity analysis.

Local parametric sensitivity measures were described in the previous subsection. There, it was specifically stressed that local sensitivities can produce inferences of local sys-

tem behavior and under no circumstances can these readily refer to the dynamics of the system as a whole. In other words, if the ranking of the overall importance of the parameters is intended, calculating the sensitivity of the model dynamics for a single set of nominal parameter values is poor. Furthermore, the chances are slim that the optimal parameter values for calculating the local sensitivities are known *a priori*. Therefore, in this context where we seek the overall ranking of the parameters, a large number of local sensitivities have to be calculated from large vectors of parameter values. Hence, optimal parameter values must be selected in such a way that the local sensitivity calculations address judiciously the entire parameter space. The Morris method and the derivative based global sensitivity measures are the most famous techniques that utilize local sensitivity measures for the ranking of parameters [44]. For this, sampling approaches like quasi-Monte Carlo and Latin Hypercubes are usually employed.

It is understandable that as the number of parameters and the complexity of the model increase, local sensitivity analysis methods become inapplicable, despite the fact that parameter sampling can provide some level of assurance that the parameter space will be adequately covered. In addition, it is highly probable that the boundaries of the parameter values are unknown or the model may not always be able to be represented in a strictly analytical form (black-box model). For those cases where local sensitivities are impossible to be assessed, global sensitivity analysis techniques can be employed instead.

Global sensitivity analysis is a robust alternative for parameter ranking and the detection of the unimportant model parameters. Techniques of global sensitivity analysis are able to evaluate the effect of a parameter while all other parameters are varied simultaneously. In this way, they are able to explore the entire parameter space, accounting both for the individual influence of each of the parameters and for their ambiguity without depending on the stipulation of a nominal point. Cardinally, global sensitivity analysis is practiced on static systems but its application on dynamic ones can be easily implemented provided that steady state is assumed at discrete times. The main advantage of global sensitivity analysis methods against the classical, local, ones, that are employed in identifiability problems, is that they treat each model as a 'black box' and the quality of the sensitivity inferences does not rely on any prior knowledge over the designed system such as function analyticity, linearity and/or additivity. The most widely used global methods for parameter ranking are the FAST, extended FAST and the method of Sobol' [45]–[47]

1.7.3 Parameter estimation/identification

Having concluded an adequate structural identifiability of the model and its parameters, parameter estimation can then be possible. Theoretically, the parameter estimation problem is well posed when global structural identifiability has been confirmed for all parameters. Then, actual experiments are performed in order to provide real data for the estimation of the physiological system parameters.

1.7.3.1 Test input signals

An essential step towards parameter estimation is to carefully select the input signals for the excitation of the system under study and the collection of optimal information. Desirably, the effect of the test signals on the experimental data due to constraints on the perturbation should be minimal.

In physiology a system is excited for data collection through either an injection or an infusion. The signals that resemble such stimuli and are commonly applied as inputs to bio-

models are the impulsive (injection) and the step (infusion) signals. The transient response of the system to these inputs is regarded as the system's output and contains the information used for parameter estimation. However, imperfections in the test signal(s) is highly possible and may result in responses that are inappropriate for interpretation. For example, the impulse may be inadequate to excite certain nodes of the system, too large and drive the system into a nonlinear mode of operation or the step signal may be too long and excessive. In addition, material introduction to an organism has to be performed only in non-restricted sites. Hence, there may be no guarantee that the stimuli will reach the target and produce the appropriate response.

The frequency response, on the other hand, by harmonic test signals may be a convenient alternative to the transient testing. Specifically, the system is excited with sinusoidally varying signals and the amplification or attenuation and phase shift of the oscillations of the output frequency components is observed. Though special instrumentation is required for this type of system stimulation it may offer some advantages in terms of accuracy and convenience of identification. Especially, when noise is a considerable characteristic of the system the frequency response testing has the merit of effectively filtering that noise because it allows one to select only the frequency component in the output that corresponds to the frequency of the sine wave applied as the input. Another signal testing able to alleviate the effects of errors is that of random signal testing. This is realized by applying a random test signal and then cross-correlate input and output. The cross-correlation function corresponds to the impulse response of the system.

1.7.3.2 Sources of error

Uncertainties in the parameter estimation process may arise due to the existence of several sources of error. Typically, these include: imperfections in the test signal; disturbance or bias to the system variables; model error due to inappropriate or false model structure (parameter local identifiability/unidentifiability); and errors during the measurements. Considerable difficulties arise when attempting to deal simultaneously with more than one source of error. Usually, however, the errors are additive and thus are independently examined [48].

With regard to imperfections in the test signals, these are either explicitly known prior to collecting the data, in which case this information can be directly taken into account, or can often be neglected. Disturbances or biases on the observables can either be neglected if their effect on the data is assumed minimum or be passed on as an additional noise factor embedded to the measurements. Model errors or false model structures can be extremely crucial. Usually the assessment of model inappropriateness is difficult to track and becomes apparent only during the validation step. For this, several alternative model structures are employed and compared against their quality in estimating the proposed model parameters. The structure with the most accurate parameter predictions is selected as the most appropriate to represent the system.

On the other hand, errors in the measurement of the outputs are always present and, for this, it is the most attended problem in parameter estimation. As said before, errors are considered in most cases additive and may enclose several types of error that may propagate to the measured output signal. So, let \mathbf{z} denote the observed outputs collected at discrete time points $\mathbf{t} = [t_0 \ t_1 \ \dots \ t_{\max}]$ and \mathbf{e} the respective measurement noise, then considering the state space notation of equations (1.1) and (1.2) we have:

$$\mathbf{z}(t, \mathbf{p}) = \mathbf{y}(t, \mathbf{p}) + \mathbf{e}(t), \quad t = [t_0 \ t_1 \ \dots \ t_{\max}]. \quad (1.35)$$

Apparently, for boosting the quality of parameter estimates it is highly desirable that a statistical description of the error is known prior to the parameter estimation. In some instances the error can be conveniently characterized by a Gaussian or a Poisson distribution. In other cases, it may only be possible to assume that the errors are uncorrelated. The properness of the statistical information is mirrored on the goodness of the results of the parameter estimation procedure. This means that the better the knowledge about the nature of the noise on the observables the more the confidence of accepting or rejecting the proposed model.

1.7.3.3 Parameter estimation/identification procedures

Then, selection of the best possible mathematical representation of the inverse model in is the next important step. The criterion for this selection is the optimization of a scalar cost function, $J(\mathbf{p})$, with respect to the model parameters [28].

In its most general form, the cost function is based on the probability theory. Specifically, suppose that the observed variables, \mathbf{z}_i , where $i=[1, \dots, n_y]$, in equation (32) are independent and identically distributed random variables conditioned on the set of parameters, \mathbf{p} , with probability (mass) functions $p(\mathbf{z}_i|\mathbf{p})$. Then, the probability of observing the measured data, say \mathbf{z}^* , under the condition that a parameter set of values, say \mathbf{p}^* , are incorporated, is numerically given by the joint probability function:

$$p(\mathbf{z}^*|\mathbf{p}^*) = \prod_{i=1}^{n_y} \prod_{t=t_0}^{t_{\max}} p(\mathbf{z}_i^*(t)|\mathbf{p}^*). \quad (1.36)$$

The product of individual probability functions in equation (1.36) is only a function of the unknown parameter vector \mathbf{p} , often referred to as the likelihood function, $L(\mathbf{p})$. Through the likelihood function one is allowed to include in the design of the cost the available information on the nature of the noise acting on the system. Then, $L(\mathbf{p})$ can be used either to generate estimators or as a key ingredient in Bayesian inference [26], [28], [29].

Theoretically, an estimator corresponds to any set of rules that when followed, they assign numerical values to the unknown parameters. This set of rules is mathematically defined as a function which, based on the input/output data, yields estimates on the parameter values. In order to yield good estimates from the estimation process, an estimator should possess a number of properties. These include: unbiasedness, minimum mean square error (variance), efficiency and consistency. First, bias is a measure of accuracy for the estimator and denotes the difference between the mean estimate and the true value. In other words, it represents how much is the distribution of the estimated values clustered around the true value. Naturally then, the smaller the bias, the more accurate the estimation process is. Second, the mean square error is a measure of precision and uncertainty for the estimate. It denotes the 'distance' between the estimated and the true value. Therefore, the smaller the variance, the smaller the uncertainty and the more precise the estimation process is. Desirably, the variance of the estimates should be the smallest that can theoretically be achieved. In this case, third, the estimator is efficient. The lowest value for the variance is better known as the Cramér-Rao lower bound [28]. Finally, consistency is the property according to which the estimated value tends towards the true one, as the sample of estimates grows to infinity (asymptotic consistency).

Naturally then, the optimal values for \mathbf{p} depend on the chosen cost function and the properties of the estimator. For this, various approaches have been employed in order to build suitable estimation procedures which can be distinguished in two categories, namely the Bayesian methodologies and the frequentist methodologies.

In a Bayesian framework, each parameter in \mathbf{p} is treated as a random variable that follows a specific distribution often called the prior distribution. The prior distribution, or simply prior, encloses the uncertainty of the parameter to be estimated based on the knowledge available before data is acquired. Usually, priors express either current belief (informative priors) or vague information (non-informative prior) about a parameter. In order to find a degree of belief regarding the true value of \mathbf{p} after observing the data, Bayes' theorem is employed:

$$p(\mathbf{p}|\mathbf{z}) = p(\mathbf{p}) \cdot p(\mathbf{z}|\mathbf{p}) / p(\mathbf{z}) \Rightarrow p(\mathbf{p}|\mathbf{z}) \propto p(\mathbf{p}) \cdot L(\mathbf{p}). \quad (1.37)$$

Bayes' theorem, uses the priors and the joint probability distribution of the observation (likelihood function) to find the posterior distribution for \mathbf{p} , $p(\mathbf{p}|\mathbf{z})$. In other words, $L(\mathbf{p})$ is used to modify the prior beliefs $p(\mathbf{p})$, with the updated knowledge summarized in a posterior density [49]. Once equation (1.37) is constructed, parameter estimation can be processed using several point estimators such as the maximum *a posteriori* estimator, the minimum risk estimator, etc. In general, analytic expressions of the posterior distribution may be unavailable or intractable. Hence, numerical simulation techniques that include sophisticated sampling methodologies such as the Markov chain Monte Carlo method are usually employed. In addition, we note that, in the denominator in equation(1.37), the probability $p(\mathbf{z})$ of the data is found by integrating the likelihood over the priors:

$$p(\mathbf{z}) = \int p(\mathbf{z}|\mathbf{p}) p(\mathbf{p}) d\mathbf{p}. \quad (1.38)$$

This term is known as the marginal likelihood of the data. It does not depend on the parameters and thus does not influence the posterior distribution (proportionality relation in equation(1.37)). Once the posterior distribution has been constructed, quantitation of parameter identification is then realizable. Clearly, since the parameter values follow the posterior distribution, the employment of statistical hypothesis tests is straightforward. Therefore, parameter inferences are assessed by determining from the calculated posterior probability density/mass, a confidence interval at a specified significance level [50], [51].

The Bayesian methodology that was briefly described here is purely probabilistic. It requires the knowledge of the parameter priors in addition to the probability functions $p(\mathbf{z}_i|\mathbf{p})$ assumed for the data. Due to the complexity of dynamic systems, knowledge retrieval can be mathematically difficult or even impossible if previous experiments are not available. Especially in the field of parameter identification of dynamic bio-models prior determination may not be sufficiently traced and/or is often impractical. In effect, and contrary to Bayesian methodologies, the frequentist paradigm opts for a purely data-based approach.

Theoretically, in a frequentist approach, one searches for the value of the parameter vector \mathbf{p} that gives the highest probability (likelihood) to the observed data. Starting from the likelihood function an appropriate model for the error(s) on the system is presumed and the likelihood function in equation (1.36) is set as a cost function $J(\mathbf{p})$. Then, the maximum of the latter will be the optimal solution to the parameter identification problem. Hence, the optimization problem becomes a problem of finding the maximum of $J(\mathbf{p})$. Such approaches are

better known as maximum-likelihood methods and can be found in many variants in the literature [1]. Maximum likelihood asks what value of \mathbf{p} makes the observed data most probable. For convenience, the majority of these variants employ independent additive Gaussian noise with constant variance for each measurement. Upon this assumption, the maximum-likelihood principle yields the respective cost function:

$$J(\mathbf{p}) = \frac{1}{\prod_{i=1}^{n_y} (\sqrt{2\pi}\sigma_i)^{t_{\max}}} e^{-\sum_{i=1}^{n_y} \sum_{t=t_0}^{t_{\max}} \left(\frac{z_{i,j}^* - z_i(t, \mathbf{p})}{\sqrt{2}\sigma_i} \right)^2}, \quad (1.39)$$

where σ_i is the variance of the error in the measurement i . Clearly, in order to find the maximum of a cost function like the one in equation (1.39) partial differentiation is required. For this, whenever differentiation is difficult the logarithm of the likelihood function is usually employed because the logarithm function is monotonically increasing and does not change the location of the optimum. Nevertheless, it is not always the case that the cost function is differentiable or not always possible to use calculus, iterative, methods to determine this maximum. In such cases, surface and contour plots can be employed in order to designate the regions of possible extremes.

The most intuitive and straightforward frequentist approach for parameter estimation is that of minimizing the sum of square errors. Supposing that in equation (1.39) the noise is uncorrelated (white) with zero mean and constant variance then, the optimization problem reduces to minimizing only the exponent factor, i.e.:

$$J(\mathbf{p}) = \sum_{i=1}^{n_y} \sum_{t=t_0}^{t_{\max}} \left(\frac{z_i^*(t) - z_i(t, \mathbf{p})}{\sigma_i} \right)^2 = \sum_{i=1}^{n_y} RSS_i(\mathbf{p}), \quad (1.40)$$

where $RSS(\mathbf{p})$ is called the residual sum of squares or sum of square errors since the difference between the measured, observed, variables, $\mathbf{z}^*(t)$, and the simulation of the state variables, $\mathbf{z}(t, \mathbf{p})$, can be considered as the error for each time t . It is essential to say here that equation (1.40) can be arrived at independently of any statistical consideration, even if we have shown that its use has been motivated by information on the nature of the noise acting on the system [28]. Usually, the solution of this optimization problem is found by minimizing the $RSS(\mathbf{p})$ through a linear or nonlinear least-squares procedure, known as regression, depending on whether the residuals are linear in the parameters or not. Linear least-square solutions are unique, analytic solutions and are always possible if the number of data points is more than the number of postulated parameters. Nonlinear least-square solutions, on the other hand, are approximate no closed-form solutions. Arriving to a solution in a nonlinear least-squares problem can be realized either through linearization/convexification or by iterative refinement of the cost function. In the latter case, the system is iteratively approximated by a linearized/convex form of the function until convergence is achieved. Both linear and nonlinear least-squares problems are convex problems where a global minimum, which can be deterministically found or approximated by computing the partial derivatives (Hessian matrix) of the cost function, always exists. However, the calculation of derivatives may not always be an option since there may be non-differentiable terms in the systems dynamics. In this case, the global minimum can be found by direct search optimization methods [52]. Direct optimization methods are unconstrained step-wise methods that initially guess a feasible parameter set as optimum solution and then they systematically “descent” towards the glob-

al optimum by direct simulation of the model's state-space function in an iterative fashion. The descent can be either deterministic or stochastic.

As a last note we have to say that several possible pitfalls and difficulties may occur during the parameter estimation procedure in dynamic systems. For example, if the optimization problem is multimodal (typical for dynamical systems), many local optimal exist and, thus, convergence to local solutions is most probable. In addition, the geometry of the objective may be very flat in the neighborhood of the solution or the model may be over-determined leading to many solution vectors. For these reasons, advanced frequentist algorithms have been developed that apply a multi-start procedure according to which many initial feasible sets of parameter values are employed as initial guesses. Then, the descent is simultaneously allowed from several different sides with the expectation of overcoming solutions of local optimality.

1.7.3.4 Practical (a posteriori) identification

Theoretical identifiability has already been described as a procedure that designates the capability of the developed model to allow for unique predictions from measured data assuming ideal conditions in which the system is noise-free and the model's structure is perfect and complete. In practice, these assumptions never hold and there is always some type of error embedded on the system. In effect, practical identifiability analysis is a natural, quantitative, extension of the theoretical one, where the difference lays in that it seeks for uniqueness and accuracy in the model predictions under real experimental conditions where noise is commonly present. In other words, practical identifiability can be regarded as an analysis that provides the goodness of fit between model response and experimental data.

Especially in the context of biological systems, a major issue for model identification is the lack of practical identifiability [53]. Provided that theoretical identifiability holds, practical identifiability relates to the amount and the quality of the experimental data. It can be crudely distinguished between *a priori* and *a posteriori* analysis. Practical *a priori* identifiability analysis anticipates the expected uncertainty of the parameters in order to compare possible experimental designs without performing any experiment. To this end, the maximum expected noise is employed on simulated data and the accuracy of the parameter estimates is calculated. Practical, *a posteriori*, identifiability analysis, on the other hand, assesses the confidence region of the parameter estimates from the available experimental information. Information about the noise, in this case, can be extracted from the data by statistical calculations either by repeating the experiments and then manipulating the acquired data or using the residuals between model predictions and the experimental measurements [28].

Primarily, the uncertainty of the parameters is assessed from the variance of the estimates in two ways: through the Cramér-Rao lower bound or through simulation using the bootstrap approach. In theory, the quality of an estimate for \mathbf{p} is given by its mean square error between its expected, true, value and the estimated one:

$$\sigma_{\mathbf{p}}^2 = E \left[\left(\mathbf{p}_0 - E[\mathbf{p}^*] \right)^2 \right], \quad (1.41)$$

where $E[.]$ evaluates the mean and \mathbf{p}_0 denotes the true value of \mathbf{p} . Then, the Cramér-Rao inequality designates the lower bound for the variance in equation(1.41):

$$E\left[\left(\mathbf{p}_0 - E[\mathbf{p}^*]\right)^2\right] \geq E\left[\left(\frac{\partial \log L(\mathbf{p})}{\partial \mathbf{p}}\right)^T \frac{\partial \log L(\mathbf{p})}{\partial \mathbf{p}}\right]_{\mathbf{p}=\mathbf{p}^*}^{-1} = \mathbf{FIM}(\mathbf{p})^{-1}. \quad (1.42)$$

The matrix that is calculated from the expectation on the right hand side of the inequality in equation (1.42) is known as the Fisher Information Matrix (FIM) and is regarded as the covariance matrix of the parameter estimates. For the calculation of the FIM the linearization of the model's output for small parameter perturbations is required. These perturbations are described by the local parameter sensitivities which account for the local behavior of the system dynamics upon an infinitesimal change of the parameter values. In the case of independent, Gaussian noise we have previously seen in equation (1.35) that the dynamic sensitivity matrix of the model's output can be constructed by calculating the partial derivatives at the discrete time points of measurements. Then, the FIM is given by the formula:

$$\mathbf{FIM}(\mathbf{p}) = \sum_{n_y} \mathbf{S}^T \mathbf{C}^{-1} \mathbf{S}, \quad (1.43)$$

where \mathbf{C} is the variance-covariance matrix of the measurements. The diagonal elements of the inverse of the FIM designate the uncertainty (*a priori*) of each of the model parameters:

$$\sigma_{p_j}^2 \geq (\mathbf{FIM}^{-1})_{jj}. \quad (1.44)$$

Hence, the confidence (*a posteriori*) for each parameter can be defined by:

$$p_j^* - \sigma_{p_j}^* \cdot t_{a/2, df} < p_j^* < p_j^* + \sigma_{p_j}^* \cdot t_{1-a/2, df}, \quad (1.45)$$

where $t_{1-a/2, df}$ is the value of the student's t distribution at the user selected confidence level $(1-a)100\%$ and df degrees of freedom which is the difference between the number of data points in a measurement, and the number of postulated parameters. In addition, since the FIM is the inverse of the lower bound of the parameter estimates covariance matrix renders possible the calculation of the parameter correlation matrix. This allows one to quantify the interrelationships between all possible pairs of parameters, designate their compensations to the model's output and identify possible practical un-identifiabilities [37].

There are two major drawbacks in the approach just described: the variance value provided by equation (1.45) is only a lower bound and the estimated confidence interval is centered on this value. This happens because local sensitivities are a linear approximation to the local behavior of the model's output. In reality, this is not always accurate, especially when the model is highly nonlinear, and therefore a bootstrap or jackknife approach is employed instead.

In general, the bootstrap methods are simulation data-based methods which rely on Monte Carlo sampling in order to quantify the expected uncertainty of the parameters and/or the confidence region [50]. The idea behind the method is to collect as many as possible parameter estimates for a given experimental error. Because in reality it is not possible to exactly replicate the experiment for a hundred times or more this is performed by computer simulation. From the repetitive application of the parameter estimation to a large number of replicate datasets, a cloud of parameter estimates is collected and recorded in a matrix. Since this cloud encloses all the possible parameter estimates for a given experimental

noise, it renders possible the construction of the distribution of values for each uncertain parameter. Then, from the analysis of the respective histograms the mean values, the maximum expected uncertainties (*a priori*) and the confidence intervals (*a posteriori*) of the parameters is possible [54]. Furthermore, assuming that the cloud of solutions corresponds to, or is fully contained in, a hyper-ellipsoid, principal component analysis can be applied to a selected range of confidence. In this way information on hyper-ellipsoid eccentricity (correlation between parameters) and pseudo-volume (accuracy of the parameters) can be acquired [55].

Under the Bayesian framework we have previously discussed that parameter estimation accuracy and confidence intervals can be calculated from sampling from the posterior distribution. This means that the quantification of the quality of the estimates is directly related to the quality of the priors used and the likelihood function of the observations that was constructed from the experimental data.

Finally, contour plots of the cost function values $J(\mathbf{p})$ by pairs of parameters is a direct way to perform practical identifiability for either of the two analyses mentioned previously [56]. These graphical mappings may reveal the existence of correlations between the parameters, the lack of identifiability for some parameters if they extend to infinity or the presence of sub-optimal solutions. However, as the number of parameters increases and the complexity of the model grows this method can become quite inefficient.

1.7.4 Experimental Designs

If the proposed dynamic bio-model has proven practical identifiability then, validation of the model predictions can be carried out. However, for those cases where there may be a lack in estimation accuracy in some (or all) of the parameters or the data may be inappropriate for utilization, further experiments should be planned and conducted.

Whether increased accuracy of the parameter estimates is sought or satisfaction of some physiological/clinical constraints, the primal variables that are considered to influence the design of the new experiments can be narrowed down to the following:

- the form of the test signal for system excitation, \mathbf{u} ,
- the duration of the measurement, t_{\max} ,
- the number of samples in a measurement, n_t ,
- the sampling schedule \mathbf{S} , i.e. the values of $t_0, t_1, \dots, t_{\max}$ and
- the variance of the measurement errors \mathbf{C} .

As it can be easily understood the actual performance of several experiments for the parsing of all of the potential combinations of experimental conditions is impractical. Instead, systematic *in silico* approaches can be employed. These regard the building of a new dynamic optimization problem where the aforementioned experimental variables are treated as parameters using the cost function:

$$J = J(\mathbf{u}, t_{\max}, n_t, \mathbf{S}, \mathbf{C}). \quad (1.46)$$

In the previous section we indicated that the *a priori* practical identifiability can serve as the means to evaluate several experimental designs without actually performing them. Accordingly, we saw that the FIM is a measure of the information content of the data and its inverse provides an estimate of the accuracy of the parameter estimates. Hence, optimal

experimental designs can be formulated as the optimization of a scalar measure of the FIM with regard to the design variables mentioned above.

Specifically, given a desired set of new experiments, the actual values of design variables is sought which maximizes or minimizes a certain FIM-based criterion subject to the constraints of the state space model, the bounds for the control variables, etc. Regarding the criteria employed, several alternatives exist all of which are related to the eigenvalues of the FIM and thus related to the shape and size of the associated hyper-ellipsoids. Ultimately, the E -optimality and the D -optimality criteria have been proved to be the best employed [55]. The former corresponds to the maximization of the minimum eigenvalue of the FIM and is mainly applied to cases where the parameters are highly correlated or the sensitivities with respect to the parameters are highly uneven. The latter corresponds to the maximization of the determinant of the FIM and is successfully applied to all other cases where the E -optimality criterion fails.

Bottom line, the goal of this procedure is the production of minimum-variance, unbiased parameter estimates. Therefore, the confidence intervals obtained from these pseudo-experiments for the parameters, will enable a decision to be made on the need to perform further experiments to improve the accuracy of the parameter estimates and, thus, the predictive capabilities of the model.

1.8 MODEL VALIDATION

Identification does not necessarily imply that the model is accurate and complete and it is usually the case that model validation may still be needed. Validity is a multi-resolution concept reflecting the purpose, the current theory, the proposed structure, and the experimental data, along with other relevant knowledge. Thereupon, validation may be regarded as an integral component of the modeling process, being firmly embedded within all its stages rather than merely being an activity carried out once model identification has been completed [57].

1.8.1 Validation criteria

Validation criteria primarily involve the model's assessment in order to check whether it is well-founded and fulfills its purpose. Although these requirements can change when new theories and data become available, the validation criteria for a model to be deemed valid remain the same. In any case, they can be roughly divided into internal and external ones.

Internal criteria refer to judging the intrinsic specifications of the model to be developed. Specifically, these criteria include: consistency and algorithmic validity. Model consistency corresponds to the idea of contradictions, meaning that the model should not include any logical, mathematical or conceptual inconsistency. For example, the mathematical realization must be consistent with the conceptual model. This means that the physical and chemical concepts relevant to the system should be properly described in mathematical terms. Algorithmic validity, on the other hand, corresponds to the determinism of the model's output and solution. It refers to the desirable compatibility and convergence of the latter two in terms of possible stochastic elements and/or numerical approximations of differential and algebraic equations.

In contrast, the external criteria refer to judging the compliance of the model's predictions with its purpose, available theory and experimental data. These criteria include: empirical, theoretical, pragmatic and heuristic validity. Empirical and theoretical validities refer to

the model's consistency with the available data and theory or models, respectively. Both of these criteria should be applied at any level within the modeling process in a recursive manner. Although this may sometimes be extremely difficult to apply at all levels, empirical validity is in general of primary importance. Aforementioned, if satisfaction of the appropriate criteria is failed, it is necessary to re-address the model's structure and/or in some cases design new experiments. Pragmatic validity refers to the effectiveness of the model in terms of satisfying its objectives. Usually for this, a measure of effectiveness is defined, which is then used to determine whether the objectives have been adequately met. Finally, heuristic validity refers to the potential of the model to provide with solid, scientific explanations about the system under study which may enable the testing of new hypotheses, discovery of new concepts, etc.

1.8.2 Validation within and after the modeling process

Validity considerations emerge right from the start of the modeling process and, as said, the criteria in terms of which validity is assessed reflect the several stages of the modeling process. Some of the validity criteria that were presented above may be applicable either during or after the formulation of the model, or in both cases. Therefore, a dynamic model of physiology systems may involve considerations of both internal and external validity at all stages of its formulation.

For instance, algorithmic validity concerns may arise, within the modeling process, at the stage of model solution or simulation. In addition, theoretical validity is highly considered in conceptual modeling and mathematical realization stages. Accordingly, the pragmatic validity which the model must meet together with any considerations of heuristic value follows the stage at which the model's purpose is defined.

In contrast, at the end of the modeling process empirical validity considerations are mostly concerned. Whether the model is required for descriptive, predictive or explanatory reasons the empirical validity can vary. Specifically, if the model is designed merely to fulfill a descriptive function, criteria of empirical validity are satisfied if the model simply reproduces the available experimental data i.e. the output responses that are produced using the proposed input signals as stimuli. On the other hand, if the model is intended to be predictive, more stringent requirements for its empirical validity must be fulfilled. In this case, the model has to reproduce the input/output experiments not only for available data but also for interpolated and extrapolated data. In other words, the predictive model has to be capable of producing credible inferences for situations outside the limits of the observed experimental data. Finally, if the model is built in order to fulfill an explanatory purpose, it must then embrace the requirements of both descriptive and predictive models and also incorporate valid structure and parameters. In other words, it is essential for the mathematical and/or conceptual structure of the model to be consistent with the actual physical and chemical phenomena occurring in the system under study. The explanatory model, incorporating appropriate levels of structural and parametric isomorphism, has its empirical validity examined by testing against other available data or by designing new experiments.

1.8.3 Validation metrics

The validation strategy essentially involves two issues; the assessment of any behavior mismatch between the one predicted by the model and the one of the actual physiological system under study; and the examination of the plausibility of the model's behavior.

Mismatch assessment or goodness of the fit can be realized through the comparison of either the overall pattern of the dynamic responses or particular features embedded on the dynamic responses. In order to quantify the discrepancy in the overall pattern of the dynamic responses one has to measure the goodness of the fit of the dynamic model's output to the response of the physiological system as indicated in the data collected from the experiments. Usually, the residual sum of squares (*RSS*) corresponding to the final parameter estimates of the model, \mathbf{p}^* , is employed as the metric:

$$RSS = \sum_{i=1}^{n_y} \sum_{t=t_0}^{t_{\max}} \left(\frac{z_i^*(t) - z_i(t, \mathbf{p}^*)}{\sigma_i} \right)^2. \quad (1.47)$$

On the other hand, it is common for the dynamics of the experimental data to include several features both simple and complex ones such as the rise and fall times, the peak value, the time to reach the peak value, possible oscillatory behavior, etc. It is therefore possible to infer on the accuracy of the developed model by exploiting the occurrence of such distinct features. This can be done qualitatively by seeking whether the features have been adequately reproduced by the model's output. However, it may be useful to quantify the occurrences. For this, a metric for the distance between the features as predicted by the model and those exhibited in available experimental test data, is set. For instance, say that \mathbf{x}_D and \mathbf{x}_M be the vectors that denote the corresponding features on the data and on the model predictions, then the adequacy of the model in terms of the features mimicked can be described by a 'figure of merit' (*A*) such as:

$$A = \frac{1}{\left(1 + \left(\frac{1}{N} \right) \sum_{i=1}^N \omega_i \frac{x_M(i) - x_D(i)}{x_D(i)} \right)}, \quad A \in [0, 1], \quad (1.48)$$

where the ratio in the sum that appears in the denominator represents the fractional error of the *i*-th model feature ($i=1 \dots N$) and ω_i are the weighting factors that may possibly apply. A figure of merit such as this provides a useful tool in the validation of a single model and also as a comparison metric when deciding between two or more competing models. In the case where the data are small in number, detecting possible inconsistencies such as outliers or systematic deviations between experimental data and predicted model response can be realized through empirical plots of the residuals against time.

Statistically, one can quantify the goodness of the fit from the manipulation of the errors calculated during the study of the mismatch between the model's state variables and the observed experimental data. Following on from the, already mentioned, residual sum of errors – *SSR_i* – statistical hypothesis tests may be compiled for every state variable. For example, an *F*-test can be performed between the standard deviations of the the residual errors in the measured observation variables and the standard deviations of the respective measurement noise. Supposing that these two standard deviations are χ^2 distributed variables, the confidence interval for the model's accuracy is given by:

$$F_{a/2, df_1, df_2} < SSR(i) < F_{1-a/2, df_1, df_2}, \quad (47)$$

where *F* is the value of the *F* distribution at the, user selected, confidence level (1-*a*)100% and *df₁* and *df₂* are the degrees of freedom of observation variables and of the parameter

estimates, respectively. The degrees of freedom in the observation variables correspond to the number of available experimental replicates from which the standard deviation has been calculated while the degrees of freedom in the parameter estimates correspond to the difference between the number of data points in a measurement, and the number of postulated parameters.

Similarly, if the model has proven unique identifiability the statistical determination of the confidence interval of the parameter estimates can be a reliable illustration of the model's validity. Often, the parameter precision is used as representative metric. For this, the coefficient of variation (*CV*) (or fractional standard deviation (*FSD*)) is utilized:

$$CV(p_i^*) = FSD(p_i^*) = \left(\sqrt{\sigma(p_i^*)} / p_i^* \right) \times 100. \quad (46)$$

Finally, model validity can turn to examining the plausibility of each particular parameter as well as the overall plausibility of the model. Specifically, if the parameters to be estimated correspond to some specific physical or physiological properties, then their estimated values must accord with the known values of such properties. In addition, the estimates should also accord to validated data independent of those used in the fitting process. If reconciling model predictions with any new data fails and/or reveals modeling defects, the model must be deemed invalid and rejected and a new model structure and/or new experiments must be planned and implemented. Special care however must be given here because the judgment of the plausibility and, in turn, the validity of the model by the estimated values of locally identifiable parameters can be misleading.

1.9 SUMMARY

The growth of computer science in past few decades has allowed complex mathematical models to be developed and evaluated on many disciplines. Especially, in the biomedical sciences where there has also been a boost in available data over the past decades, there has been a tremendous increase in the interest of scientists in computational modeling. Developed models found in literature span several spatial and temporal scales from molecular to organ systems. Most of these models are compartmental and/or non-compartmental approaches that simulate metabolic, pharmacokinetic and physiological systems.

Generally, building such computer representations to be predictive can be regarded as a cyclic, iterative process. Presuming that the purpose of the model and some preliminary knowledge are defined the cycle begins by putting forward a trial model. Then, model identification is performed. Primarily this entails the study of the theoretical identifiability of the model parameters. This is a mathematical investigation performed *a priori* for evaluating whether the model can, under ideal conditions, uniquely reproduce the system. However, this is a very difficult task that is proportional to the model's complexity. A complementary analysis to infer whether the postulated parameters of the model can be theoretically measured through the proposed structure is sensitivity analysis. Once the study for the theoretical identifiability completes the model is examined to infer whether its predictions adequately agree with the observed data. In most cases, this requires the estimation of the model parameters from the available measurements, through the minimization of a likelihood function. This procedure is better known as practical model identification.

Usually, model identification, either theoretical or practical, reveals certain design caveats. Then, once these analyses complete the model's design can be reformulated and/or

the construction of other experiments can be planned. These processes are repeated iteratively, until the parameters' predictions are improved and the estimation errors are minimized to a satisfactory level. Last but of primary importance, validation is an integral component of the modeling process that can be carried out at any stage of the model building cycle.

Many of these concepts will be investigated in the following chapters with the purpose of developing a predictive model of the biomarker-tissue interaction during the development of epithelial neoplasias. Our ultimate goal is to develop a novel imaging modality for the early detection of precancerous abnormalities in epithelial tissues.

2 BIO-OPTICAL BACKGROUND FOR MODEL DEVELOPMENT

2.1 INTRODUCTION

Neoplasia is fundamentally a broad group of diseases involving tissue growth regulation failure. Neoplastic cells, in particular, are known to divide and grow uncontrollably forming either benign or malignant tumors and possibly invading nearby parts of the body through the lymphatic system. In order for a normal cell to transform into a precancerous one the genes which regulate its growth and differentiation must be altered. Then, due to the fact that epithelial tissues constantly undergo several regeneration activities they sustain the highest risk of developing neoplastic lesions. In fact, 85 out of every 100 cancers are epithelial. Despite this, the causes of epithelial tissue malformations are diverse, complex, and only partially understood. Hence, detection of the functional and structural alterations that signify them is of primary therapeutic importance. To this direction, chemical contrast agents, affine to cancer-specific molecular, cellular, or tissue components, have been designed and clinically tested. The corresponding outcomes have led to the development of advanced medical devices which base themselves on the measuring and quantification of the induced optical signals.

Along these lines, in this chapter, we describe the physiology of human epithelia and the metabolic pathways that preserve their homeostasis. We introduce the key changes brought about by neoplasia progression and present, according to literature, the cancer-specific transitions in the case of cervical cancer. Based on these findings, we identify and interpret the metabolic activities that are responsible for the evanescent optical phenomena observed macroscopically when acetic acid is used as a biomarker of cervical precancerous lesions. Finally, we describe, briefly, a DCE-OI system that has been developed capable of measuring the characteristics of this phenomenon during the time-lapsed after the application of the agent.

2.2 PHYSIOLOGY OF EPITHELIAL TISSUES

All animal cells can be classified as being components of just five main classes of tissue: epithelial tissue, connective tissue, muscular tissue, nervous tissue, and the blood. Epithelia are a diverse group of tissues which, with rare exceptions, cover or line all exposed body surfaces such as organs, cavities and tubes. They protect and mediate absorption, secretion and communication; abilities that are essential for the physiological operation of the body's biological mechanisms. In the following paragraphs the morphology and function of this type of tissues are briefly described.

2.2.1 Integration of cells into tissues

The cell is the cornerstone of every living organism. Structures called organelles house cell functions producing nutrients, preserving the homeostasis and removing hazardous elements. Foremost important organelle is the nucleus, controlling and designating cell's activities. Proteins serve as the intermediate of every cell function by working as enzymes catalyzing reactions, as structures and regulators, as molecular machines for power and movement and organize the cell's defense and signaling. Their unique conformation is essential for their identification and differentiation and is produced by chemical reactions (e.g. due to the hydrophilic and hydrophobic attribute of chemical combinations), violation of the systems' entropy (e.g. temperature) or other physical conditions (e.g. electrical, changes in pH) surrounding the protein. All these are structured on the cytoskeleton and are enclosed by a plasma membrane protecting the inner environment from unwanted external stimuli. Figure 2-1 depicts the primary components of an animal cell along with their basic functions.

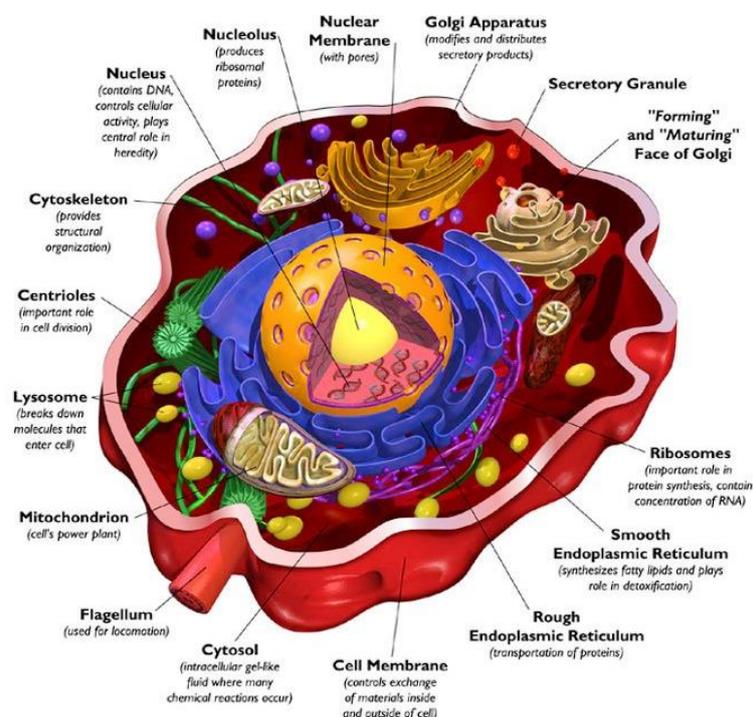


Figure 2-1 Elementary components of animal cells

In order to create complex multicellular entities such as tissues, organs, and appendages, cells must be properly arranged in precise patterns of staggering complexity. For this, progenitor cells differentiate into distinct "types" that have characteristic compositions, structures, and functions. The assembly of cells of a given type into distinct tissues and the organization of the latter into organs are determined by molecular interactions at the cellular level and would not be possible without the temporally and spatially regulated expression¹ of a wide array of adhesive molecules. As it will be described in more detail in the following paragraphs, cells in tissues can adhere directly to one another through specialized integral membrane proteins called cell-adhesion molecules that often cluster into specialized cell junctions. In addition, cells in animal tissues also adhere indirectly through the binding of adhesion receptors in the plasma membrane to components of the surrounding extracellular matrix. These two basic types of interactions not only allow cells to aggregate into distinct

¹ *Protein expression*: the ability of a protein to operate properly, according to its physiology

tissues but also provide a means for the bidirectional transfer of information between the exterior and the interior compartment.

Similar, in type, cells aggregate into a tissue in order to cooperatively perform a common function. This coordinated functioning of many types of cells within tissues, as well as of multiple specialized tissues, permits the organism as a whole to move, metabolize, reproduce, and carry out other essential activities. In addition, many cells exhibit distinct functional and/or structural asymmetries, a property often called polarity. From such polarized cells arise asymmetric, polarized tissues such as the lining of the intestines and structures like hands and hearts.

2.2.2 Types of epithelial tissues

Epithelia comprise of firmly coherent cells with minimal interval. One surface of the epithelium is exposed to the lumen and the other usually rests on a basement membrane (basal lamina) as depicted in Figure 2-2a. The latter, is a thin fibrous network of collagen and other extracellular matrix components that supports all epithelia and connects them to the underlying connective tissue. In addition, epithelial tissues are usually polarized keeping basal and apical cellular surfaces structurally and functionally different exhibiting distinctive characteristics. Thus, epithelia function as interfaces between different biological compartments. As such, they mediate a wide range of activities such as selective diffusion, absorption and/or secretion, physical protection and containment.

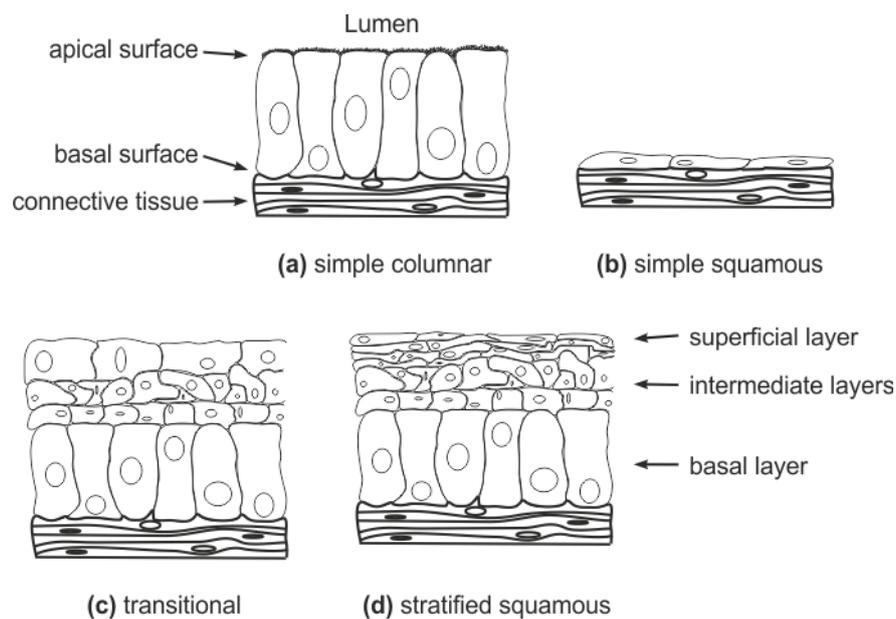


Figure 2-2 Classification of epithelia.

Generally, epithelial tissues are distinguished by their cellular shape, number of cell layers, cilia coating (microvilli - natural membrane extensions that promote absorption and secretion of nutrients) and keratin containment. Accordingly, they can be narrowed down into four distinct categories (Figure 2-2):

- *Simple columnar* epithelia (Figure 2-2a) that consist of elongated cells, including mucus-secreting cells (in the lining of the stomach and cervical tract) and absorptive cells (in the lining of the small intestine).

- *Simple squamous* epithelia (Figure 2-2b), composed of thin cells, line the blood vessels (endothelial cells/endothelium) and many body cavities.
- *Transitional* epithelia (Figure 2-2c), composed of several layers of cells with different shapes, line certain cavities subject to expansion and contraction (e.g., the urinary bladder).
- *Stratified squamous* (nonkeratinized) epithelia (Figure 2-2d) line surfaces such as the mouth and vagina; these linings resist abrasion and generally do not participate in the absorption or secretion of materials into or out of the cavity.

2.2.3 Epithelial tissue activity

As said, the integration of cells into tissues relies on the adherence of cells between one another and to the surrounding extracellular matrix. In this way, epithelia are able to regulate the movement of ions and small molecules to the tissues underneath.

Adherence is established using certain types of protein structures called junctions. Prime junctions are the desmosomes, the tight and the gap junctions (Figure 2-3). Tight junctions and desmosomes assemble harmonically the cytoskeletons of nearby cells. The former clasp the apical surfaces of cells while the latter connect the rest of the cells' bodies. The gap junctions, on the other hand, are responsible for the interconnection of cells. They exchange nutrients and second messengers through neighboring cytoplasms promoting intercellular communication [58].

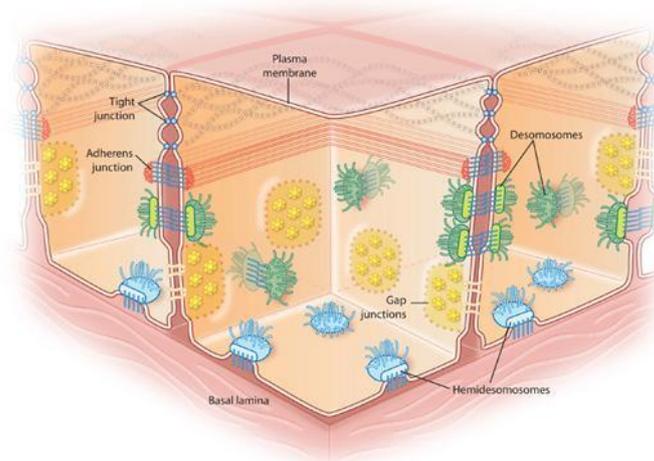


Figure 2-3 Primary epithelial junction arrangement

In general, these conformations play a critical role in controlling the movement of ions and small molecules in epithelia. Cardinaly, transport of molecules in tissues can be either active or passive. Passive transport allows solute diffusion down its chemical concentration gradient spending no metabolic energy. It is achieved through random thermal motion of suspended or dissolved molecules moving from an area of higher to lower concentration. On the other hand, active transport is the mode that is not only independent, but also operates against extreme concentration gradients; for this, it is fuelled by ATP² hydrolysis. Therefore, during physiological circumstances, molecules penetrating epithelial barriers may diffuse, either actively or passively, via the cell's membrane (transcellular), the tight junctions (para-

² Adenosine 5'-triphosphate (ATP) is a multifunctional nucleotide that is most important in intracellular energy transfer.

cellular) or the gap junctions (intercellular) before they are consumed or canalized in the vasculature underneath (Figure 2-4).

2.2.3.1 Secretion-Absorption: The transcellular pathway

Transcellular is the main absorptive pathway for epithelia that allows solutes to traffic directly through the membrane into the cell's cytoplasm. Secretion and absorption of nutrients is carried through microvillis, glands or direct transmembrane communication as shown in Figure 2-5. Microvillis are leaf-like formation lined over the outmost cell layers of most epithelia acting as a toll bridge between the lumen and the tissue. Glands are cells specialized for secretion of a specific substance or substances. In epithelia, glands are usually embedded, occurring as isolated cells containing the secretion, or as layer of cells surrounding interstitial space (secretory cavity or canal) into which secretion is discharged. In the absence of microvillis it is facilitated by the cell's apical surface and all intra-extra lumen transport is managed through the uppermost portion of the plasma membrane [59].

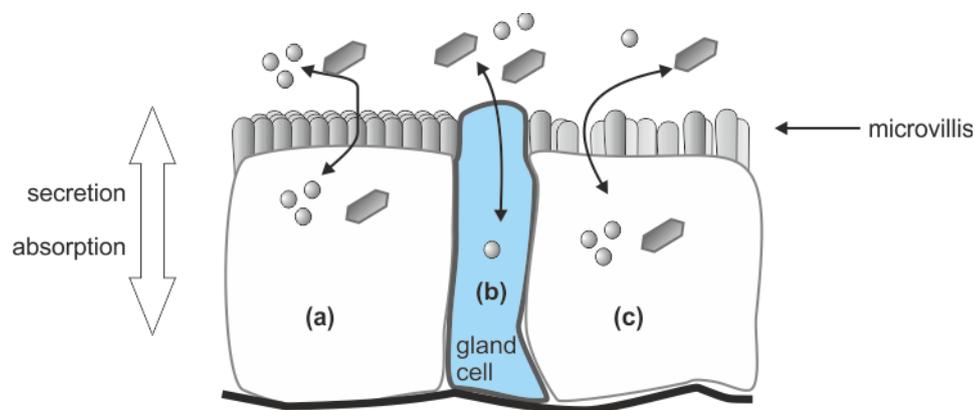


Figure 2-5 Secretion and absorption can be either through microvillis (a), glands (b) or directly through the membrane (c)

The plasma membrane encloses the cell. It functions as a physical barrier, regulates the exchange of materials with the environment, supports the communication between the cell and its environment and conduces to structural support. The cellular membrane is comprised of cholesterol, proteins and phospholipids which are positioned in a bilayeral mode as depicted in Figure 2-6. The distribution of lipids is unequal across the bilayer and, for this; their composition influences the physical properties of membranes.

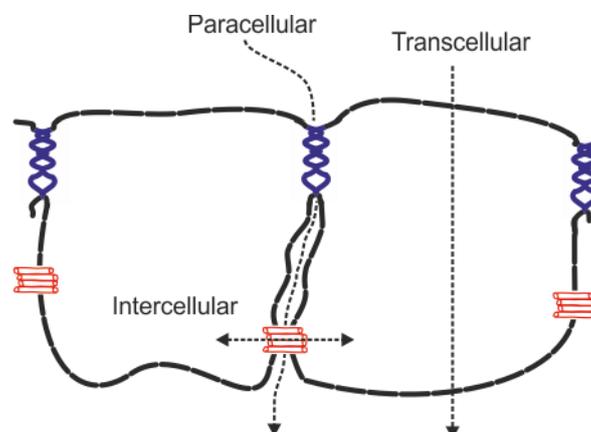


Figure 2-4 Transport pathways in epithelial tissues

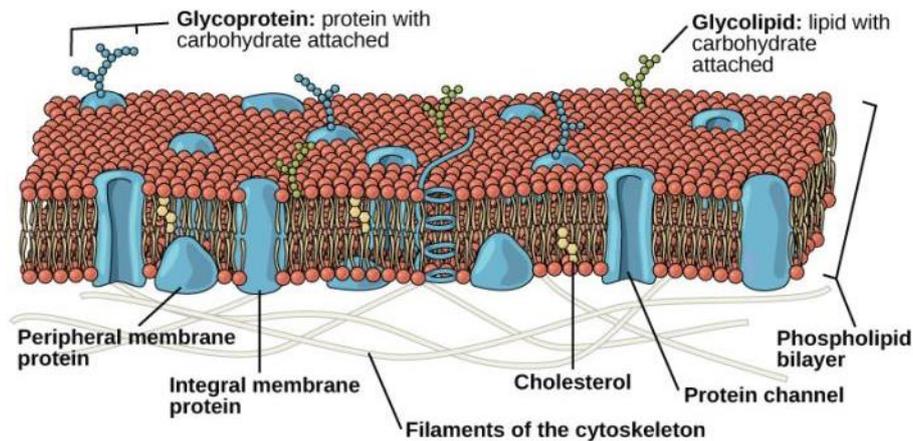


Figure 2-6 Composition of the cell's membrane

Diffusion of nutrients through the transcellular pathway has been classified above into simple passive transport and protein mediated transport. In simple passive diffusion solutes traverse solely and passively through the lipoidal matrix. As a rule, this is allowed only for a small amount of molecules such as gases and small uncharged solutes (O_2 , CO_2 , acetic acid, etc.) and is characterized by two terms, namely the membrane's *fluidity* and the solute's *permeability*. The former refers to the ease of motion of the membrane's constituents and the latter refers to the ease of motion of solutes through the membrane. The first to suggest that only the undissociated molecules are permeant through a lipid bilayer was Overton, in 1895 [60]. Today the prevailing model that describes simple passive diffusion through the cell's membrane is the solubility – diffusion model as corrected for the misbalanced distribution of phospholipids across the liquid bilayer membrane [61]. This model relates a given solute's permeability coefficient to its ability to dissolve into and then diffuse across the lipoidal membrane interior (Figure 2-7) [62].

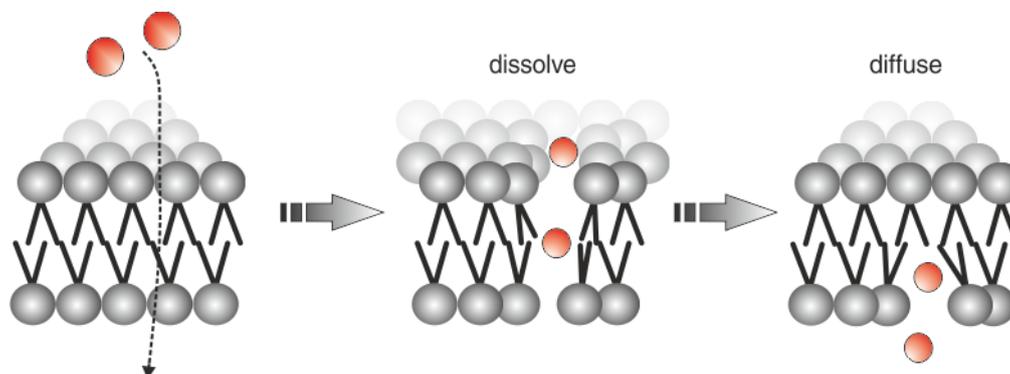


Figure 2-7 The dissolve-diffuse mechanism of passive membrane diffusion

Protein mediated transport is the supplement mechanism of cells that allows the trafficking between the intracellular (IS) and the extracellular space (ES) for charged solutes and macromolecules. Morphologically, the transport proteins protrude the lipoidal matrix acting as integral membrane entities and appear either to traverse the membrane or be part of a structure which does. Functionally, on the other hand, all protein-mediated transported molecules demonstrate specificity, competition and saturation. For this, transport proteins are found to allow either the unidirectional transport of solutes (uniports), the transport of one solute upon simultaneous carriage of another in the same direction (symports) or the car-

riage of one solute into the cell upon simultaneous transport of another out of the cell (antiports) (Figure 2-8).

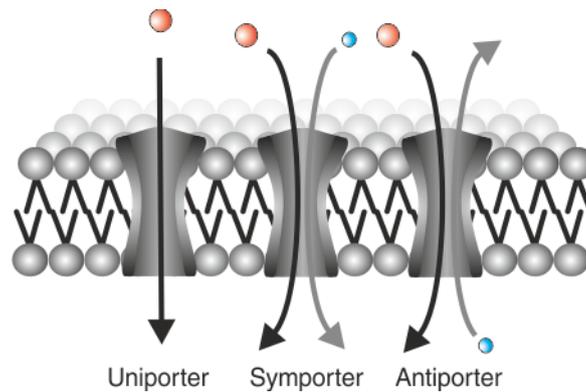


Figure 2-8 Transport proteins

These intrinsic membrane proteins, mediate transport of a solute either passively or actively (Figure 2-9). Protein mediated, passive membrane transport can be either facilitated or carrier mediated. Facilitated diffusion is the movement of molecules that depends on their

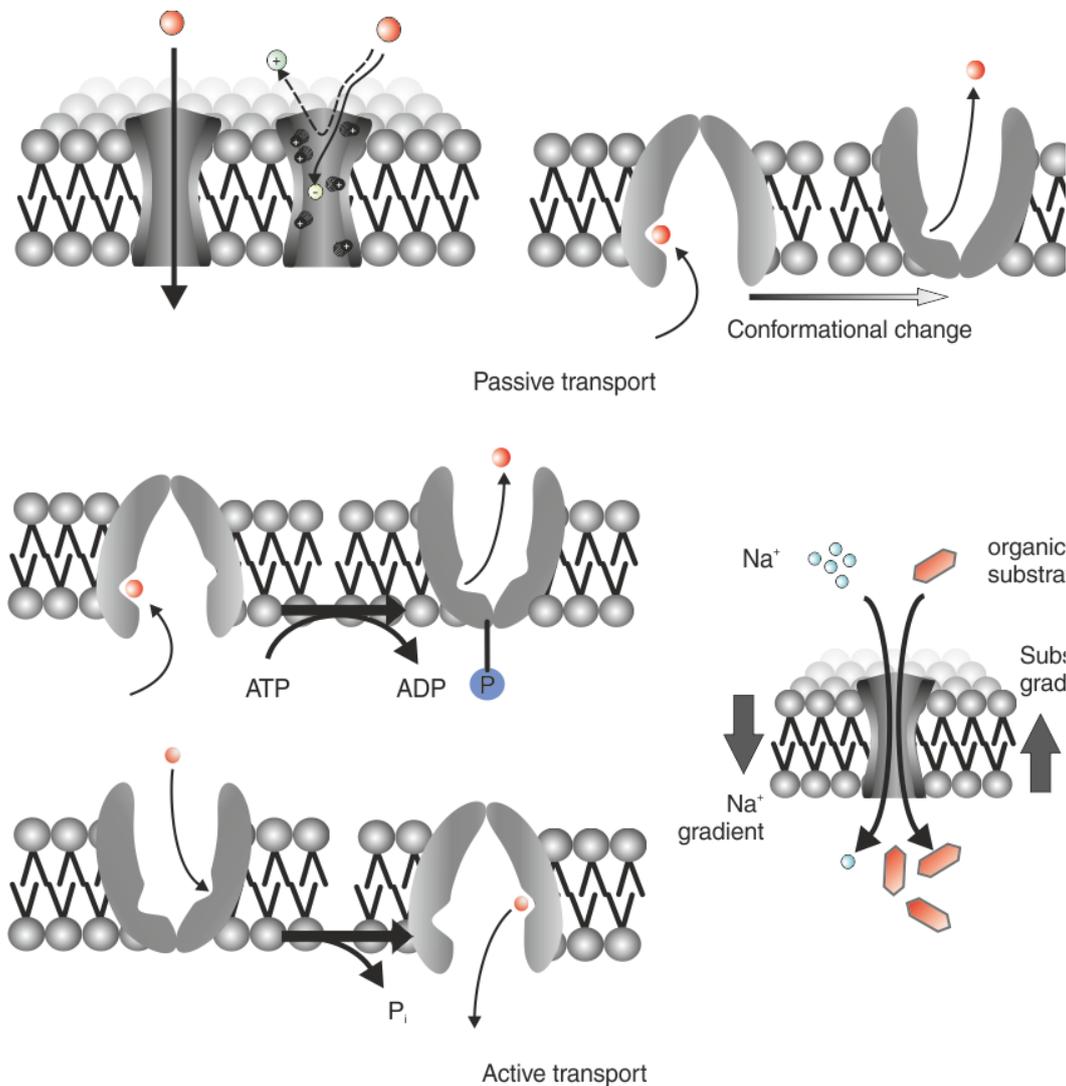


Figure 2-9 Diffusion schemes through the membrane's lipid matrix

electrochemical gradient between the extracellular and intracellular areas (e.g. ion channels). Selectivity for these channels is determined by the pore's size and/or charges lining the inside of the channel. The carrier-mediated is a diffusion mode that uses a carrier molecule or protein (permease) to bind with the transported solute, assisting in its transport across cell membranes without being permanently altered in the process (i.e. ions bind specifically and reversibly). On the other hand, active transport is the energy-dependent carriage that utilizes the ATP hydrolysis or the potential energy stored in an electrochemical concentration gradient. Its primary role is to keep the cell under physiological conditions i.e. the preservation of cytoplasmic composition in ions such as H^+ , Na^+ , Cl^- , K^+ and Ca^{+2} and the diffusion of macromolecules that cannot diffuse passively. Actively transported molecules move against their concentration gradient through a channel protein and are blocked by specific inhibitors. Indicative examples are the $Na^+-K^+-ATPase$ antiport pump that maintains gradient of Na^+ and K^+ across the cell membrane and is important for cell volume control, the $H^+-ATPase$ uniport pump, the sodium dependent Na^+-H^+ symporter, the non-sodium dependent antiporter $HCO_3^-Cl^-$, etc.

2.2.3.2 Protection: The paracellular pathway

The paracellular is the pathway delimited by the tight junctions and the lipid membranes of adjacent epithelial cells of the latter interspace (Figure 2-10). Tight junctions are complexes bound to the cytoskeleton that consist of several protein families with the protein family of claudins (CLDNs) being their primary structural and functional component. Tight junctions perform two vital functions: 1) they prevent the passage of molecules and ions through the interstitium and 2) they block the movement of integral membrane proteins between the apical and basolateral surfaces of the cell. Beside the desmosomes, tight junctions are responsible for the maintenance of tissue's architecture and cohesiveness that open and close under certain physiological stimuli [63], [64]. Therefore, the affiliation administered, acts as a gate which controls the diffusion along the paracellular pathway and also as a fence that produces a morphological boundary for lipids, keeping the tissue polarized [65], [66].

Structurally, tight junctions comprise of a network of anastomosing strands, constructing small regions analogous to pores 4.9-7 Å in diameter (sometimes 15 Å or even more)

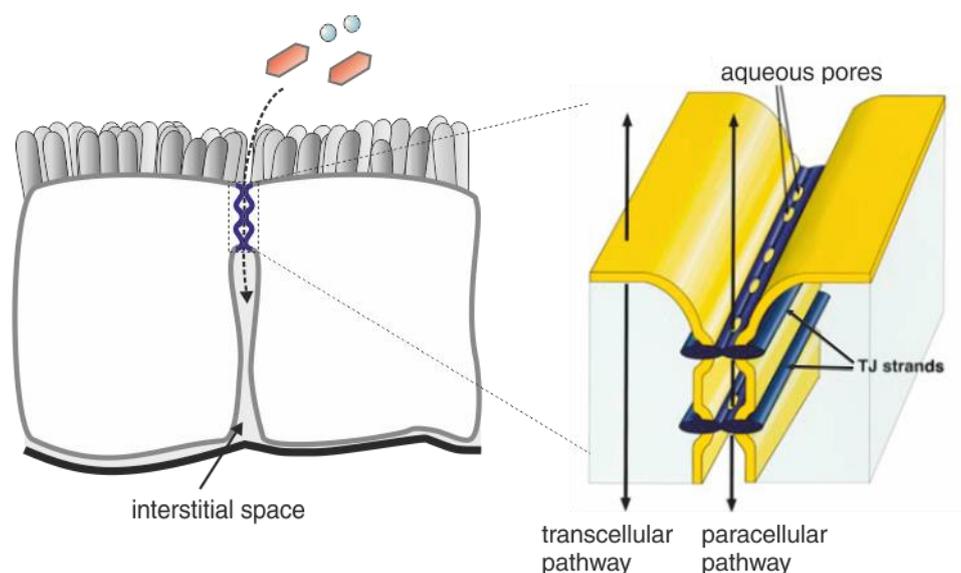


Figure 2-10 Epithelial tight junctions

[67]–[70] (Figure 2-10). These pores are oriented parallel to the plasma membrane, showing no tortuosity, arranged, however, in an irregular network. Functionally, these pores allow for the paracellular transport of both charged and non-charged species. Such transport can only be passive and results from diffusion, electrodiffusion or osmosis down the gradients created by transcellular mechanisms and/or extrinsically produced gradients. In general, tight junctions have been found to share the biophysical properties of ion channels, including size and charge selectivity, dependence on ion concentration, pH, extracellular Ca^{+2} , as well as the presence of ionic competition. However, the selectivity of the paracellular barrier varies from one epithelial tissue to another and different stimuli or manipulations can result in opposite effects on the permeability of ions and tracers.

The tight junction pores under physiological conditions are cation (+) selective but not gated. In general, resistance is proportional to pore and solute size (mostly), number of strands (junctional tightness) and the probability of finding a pore in open or closed state [64], [69], [71]. However, resistance may vary substantially along the spectrum of the so-called tight to leaky epithelia. In addition, experiments have revealed that the tight junction strands may be broken and annealed according to environmental stimuli [63], [72]. Reasonably then, the pore's diameter and, in turn, the tight junction barrier is thought to have a dynamic profile.

2.2.3.3 Communication-Differentiation-Growth: The intercellular pathway

The intercellular pathway accommodates the direct communication between the cytoplasm of neighboring cell interiors via the gap junctions which allows them to share vital resources, exchange information, grow and differentiate. It supports the integration of the metabolic activities of all cells in a tissue, known as metabolic cooperation, by allowing the direct passage of ions and small molecules from the cytosol of one cell to that of another [58], [73], [74].

Structurally, the gap junction channels are leakproof formations composed of a family of transmembrane proteins called connexin. They are located perpendicular to the membrane's axis, allowing structural continuity between adjacent cells numbering approximately 10^6 channels units (Figure 2-11) [58], [75]–[77]. Structural analysis of a gap junction revealed an 80-100Å in length protein junctional unit between adjacent cytoplasm with a 14-16Å hydro-

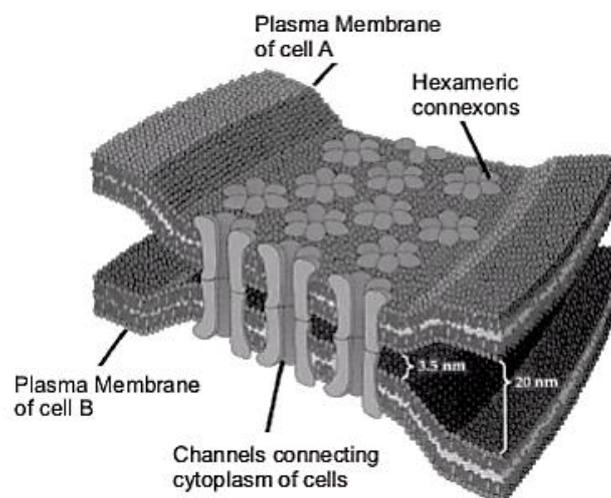


Figure 2-11 The gap junction plaques consisting of several gap junction channels

philic pore diameter. The size selectivity of these junctional units (in MW) ranges from 0,3 to approximately 1-1,5 KDa solutes and movement is granted through chemical or electrical potentials, allowing passive movement of ions and hydrophilic molecules [78], [79].

The nature of solute diffusion through the gap junction complex may be only passive, similarly to the paracellular one. Whereas the selectivity of the gap junctions does not fluctuate significantly for small molecules, quantitative data have shown that the permeability of large molecules is proportional not only to the solute and junction sizes, but also to the charges of either diffusion participants and the expression of proteins found therein [80]–[82].

The structural and functional continuity of adjacent cells facilitated by gap junctions is uncoupled by the ionic composition of the cytoplasm, the transjunctional voltage between the interconnected cytoplasms and phosphorylation. Gap junctions, just like ion channels are gated i.e. they can undergo a reversible conformational change that closes or opens the channel in response to changes in the cell. Sensitivity emanates from intracellular calcium and pH. The relation of gap junction permeability is linear and almost spontaneous with calcium concentration permutation. In contrast, junctional conductance is indirectly related with pH, where its dependence exhibits a time-scale hysteresis. In vitro, however, a cell with prolonged low pH, i.e. malfunction in pH regulation mechanisms, has been shown to quit its metabolic cooperation [83]. Probably, acute acidification provokes the gap junctions' closure working as a protective mechanism, in order to avoid the possible spread of the potential impairment.

2.2.4 Preservation of homeostasis

Of utmost importance in the maintenance of tissue's natural condition is the coordinated regulation of the pH of the intracellular and extracellular space ($\text{pH}_{\text{IS/ES}}$). Specifically, under physiological circumstances, a cell maintains its pH_{IS} slightly basic (7.2 – 7.6 pH units). This is realized either through passive regulatory mechanisms or active ones [84]. The former disseminate H^+ through: a) chemical buffer pairs, b) metabolic processes, c) dissemination to neighboring cells (metabolic cooperation) and d) diffusion into proton-craving cellular compartments such as the mitochondria, golgi etc. The latter, as said previously, are protein-based mechanisms that pump ionic molecules from one biologic compartment to another based on the intracellular and the extracellular pH levels and according to cell's capacity in free energy. The provision of the necessary nutrients for the maintenance of these physiological processes is ensured by the surrounding capillaries through the extracellular (interstitial) space. Under ideal diffusive coupling between cells and capillaries, the pH_{ES} is maintained basic, roughly higher than the intracellular one and the only ways that this can diverge, under normal circumstances is when the cell-capillary distance is increased (e.g. as a result of poor blood perfusion), the excreted acid/base load is elevated or when the local buffering capacity is altered [85]. Next, we will describe in more detail the mechanisms responsible for the maintenance of pH_{IS} and pH_{ES} under normal circumstances.

2.2.4.1 Passive Mechanisms – Buffering Power

The buffering power represents the capacity of the cell to blunt changes imposed to the intracellular pH. It is mathematical defined as the amount of strong base (or strong acid) that must be added, to raise (or lower) the pH by a certain amount i.e. $\beta = d[\text{B}]/d\text{pH}$ (or $d[\text{A}]/d\text{pH}$) (M/pH units) and is segregated in intrinsic (β_{IS}) and extrinsic (β_{ES}) one. The former represents the stationary cytoplasmic mechanisms i.e. native chemical and metabolic processes and also the sequestering of protons to organelles while, the latter, represents those chemical moie-

ties that, apart from absorbing protons, can transient between the ES and IS. As the acid extrusion mechanisms are categorized into short and long term ones, due to the time interval needed before completing their task, likewise, the buffering modalities can be sub-categorized into fast and slow processes. As such, the intrinsic mechanisms, that are chemical reactions and occur instantaneously, will belong to the former subcategory and the extrinsic buffers, that are mediated by transmembrane diffusion which is many-fold slower, even for gases such as CO₂ [86], will belong to the latter.

Upon acidification, protons titrate the intrinsic (closed) buffers which are rapid processes and proportional to the kinetics of the corresponding chemical reactions. These are, a) $H^+ + B^- \leftrightarrow HB$, for chemical buffers, b) $R^n + qH^+ \leftrightarrow P^{n+q}$, where R is the reactant, P is the product, n is the valence and q is the H⁺ stoichiometry (q>0), for biochemical buffers and c) $Organelle() + H^+ \leftrightarrow organelle(H^+)$, for organelles [87]. These processes can absorb the load to the extent that is allowed by their intracellular concentration (a), metabolic activity (b) and the amount that is needed for their internal processes (c), respectively. Intuitively then, we can infer that the intrinsic buffering depends clearly on the status of these mechanisms i.e. the pH_{IS} level that exists *a priori* (Figure 2-12).

If the cell was a closed system then β_{IS} would account for all the buffering power. However, the cell is an open system with molecules diffusing inwards and outwards constantly, according to physicochemical gradients (Figure 2-13). The biochemistry of open buffers is similar to (a), with the exception that their HB proportion can traverse the cellular membrane and as such, it can be considered as a slower buffering mechanism compared to the intrinsic one. The most significant open buffer is the CO₂/HCO₃⁻ pair. If the cell was a closed system as described before then, the capacity of the CO₂/HCO₃⁻ buffer pair would be miniscule but, since CO₂ is a product of aerobic respiration, the main energy provider, and can switch instantaneously between the IS, the ES and the vasculature, it accounts for most of the buffering in mammalian cells [87]. The buffer is formed through the hydration of CO₂ into bicarbonate according to the reaction: $CO_2 + H_2O \leftrightarrow H_2CO_3 \leftrightarrow HCO_3^- + H^+$. The forward reaction i.e. the CO₂ hydration, is inherently slow compared to its dehydration (~8-fold) and thus, in many biological compartments this process is catalyzed by the carbonic anhydrase enzymes [90], [91]. Hence, $\beta_a (= \beta_{CO_2})$ also plays a significant, additive, role to the total capacity of the cell being dependent, however, on pH_{ES}, concentration gradients and transmembrane diffusivity apart from pH_{IS} [84], [87], [92].

2.2.4.2 Active Mechanisms – Pumps and Exchangers

Active acid extrusion is accommodated by the H⁺ ATPases (V-ATPase) (primary active) and the Na⁺/H⁺ (NHE) and Na⁺-driven Cl⁻/HCO₃⁻ (NBC) exchangers (secondary active) [87]. The former, on the one hand, is an electrogenic system that transports two H⁺ ions extracellularly in order to create an electric gradient and relies on the availability of intracellular free energy. The latter, on the other hand, extrude protons, due to ionic concentration gradients; either directly or indirectly. Specifically, the NHE exchanges intracellular H⁺ for extracellular Na⁺ and the NBC exchanges Cl⁻ for Na⁺ while two bicarbonate ions neutralize equal amount of intracellular protons; thus, this mechanism is considered as a recirculation mechanism for the CO₂/HCO₃⁻ open buffer.

Unlike buffering, these mechanisms are kinetically slow and their rates of extrusion are proportional to the pH_{ES} and inversely proportional to pH_{IS} (Figure 2-14). Typically, for constant pH_{IS}, the exchangers are almost inactive when pH_{ES} is low (~6), while they are mostly

active when pH_{ES} is high (7-7.4). In addition, for both exchangers, pH_{ES} appears to play a limiting role inhibiting their dynamics [86], [87] and as it has been shown, their activity takes several hours to return to normal when pH_{ES} is instantly altered [93] which means that pH_{ES} has a “chronic”, rather than instant, effect on the status of these systems.

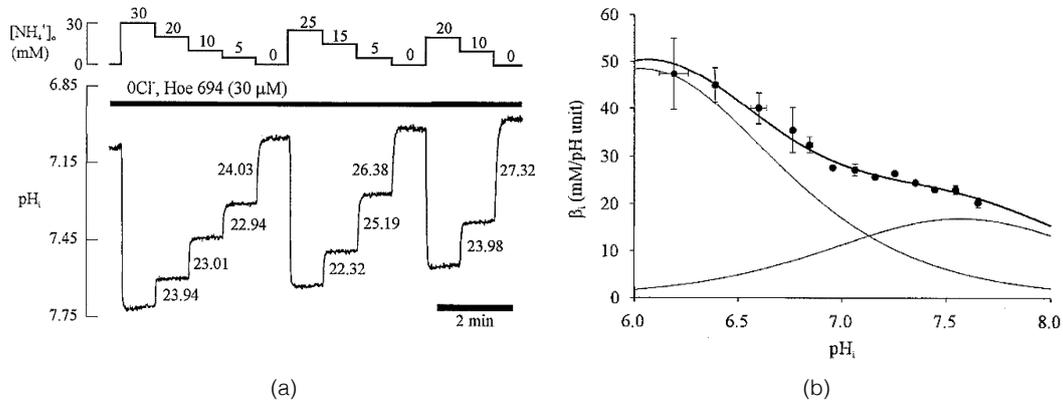


Figure 2-12 Intrinsic buffering power dependence on pH_{iS} . a) Trace, in the absence of CO_2/HCO_3^- . A weak acid (NH_4^+) was added extracellularly and then it was removed stepwise in the absence of Na^+ in order to inhibit active extrusion mechanisms. b) The resulting β_i was calculated using the expression $\beta_i = \Delta[NH_4^+]_i / \Delta pH_i$, where $\Delta[NH_4^+]_i$ was calculated using the Hendersson-Hasselbalch equation [88].

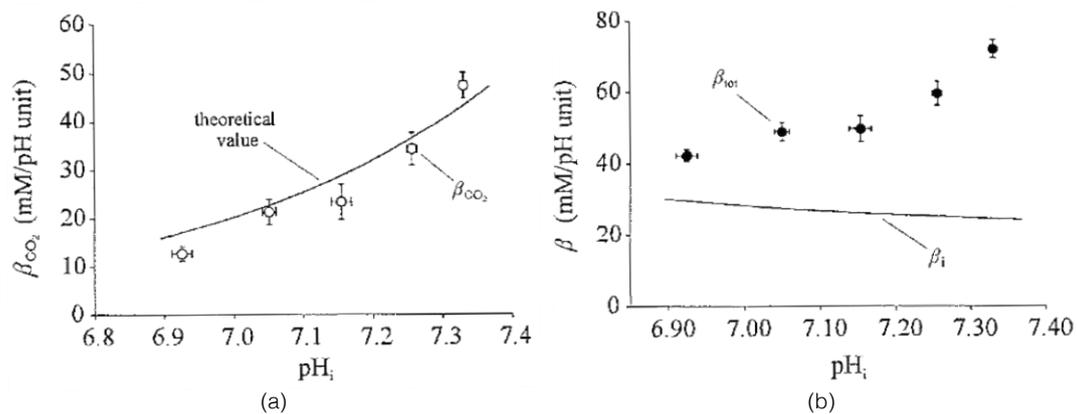


Figure 2-13 Extrinsic, intrinsic and total buffering in ventricular myocytes. a) The CO_2/HCO_3^- pair (extrinsic buffering) has been evaluated in Cl^- free solution in order to avoid extrusion of bicarbonate through active mechanisms. b) The total and intrinsic buffer capacities of the same experimental procedure [88].

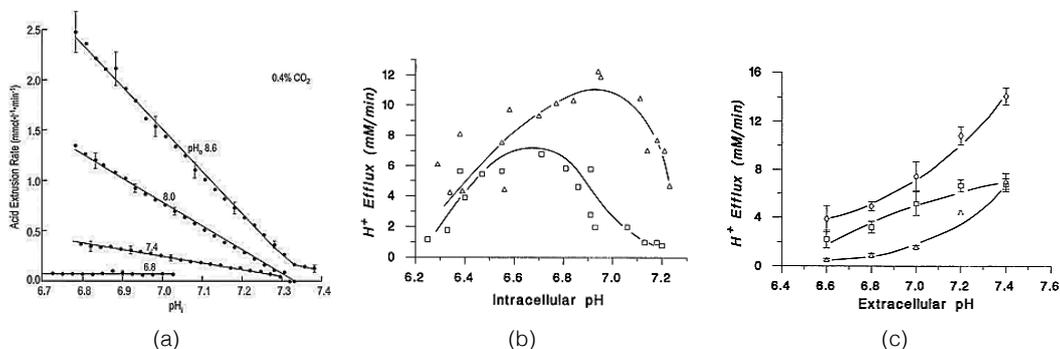


Figure 2-14 The pH_{iS} dependence of the acid extrusion rate in giant barnacle muscle fibers. The rate of pH_{iS} recovery from an acid load has been determined over a range of pH_{iS} values, multiplied by the sum of intrinsic and extrinsic buffering powers i.e. flux = rate of change of $pH_{iS} \cdot \beta_T$ [84]. b) The course

of efflux rate when pH_{IS} continues to drop at constant pH_{ES} (~7.4). c) The dependency of the efflux rate on the pH_{ES} while the pH_{IS} is kept constant (~6.5) [89].

On the other hand, for constant pH_{ES} , the exchangers are inactive when pH_{IS} is high, while they are mostly active when pH_{IS} is low (6.4-6.8) [94], [95]. Interestingly, as seen in (Figure 2-14b) when pH_{IS} drops even more (~6.2) its relation with the extrusion flux reverses and acid extrusion ceases [89]. Finally, the threshold of operation for the NHE and NBC is when pH_{IS} is approximately 7 and 7.4, respectively and after that, the maximal extrusion rate is usually described with a Hill model as a function of pH_{IS} , according to the mechanism's stoichiometry [89], [93], [96].

2.3 EPITHELIA AND CANCER

As said in the introduction, the epithelia are tissues that naturally undergo repetitive regenerative activities which renders them the most favorable body sites in sustaining a developing cancers. In medicine, a neoplasm or tumor is considered the result of an abnormal pattern of growth or division of cells that creates an abnormal mass of tissue. The molecular origins of the tumor's survival mechanism cannot yet be totally understood. However, according to the phenotype of a tumor, proliferative neoplastic cells remodel the epithelium both structurally and functionally as a result of the loss of proper protein gene transcription and aberrant localization of proteins. The most important cancer-specific functional signature that preserves the tumor homeostasis is the ability of neoplastic cells to regulate the internal pH levels.

2.3.1 pH regulation as the primary cancer-specific functional signature

The clockwork setup described earlier concerning the maintenance of tissue homeostasis is compromised during cancer development. Typically, tumors are known for their high metabolic rate and the abnormal blood perfusion [85]. Cells, which under physiologic conditions would maintain their homeostasis and metabolic rates through the, high ATP yield, aerobic respiration, are deprived from necessary nutrients such as the O_2 and, thus, hypoxic conditions (O_2 partial pressure $\text{pO}_2 \sim 1-0.1\%$) are created [97], [98]. These circumstances, compels them to switch to anaerobic glycolysis for energy production, that occurs in the cytoplasm rather than in the mitochondria, and converts glucose into metabolic energy and the waste-product lactic acid (Warburg Effect) [99]. The fact that this mechanism is less efficient and the fact that cellular energy needs for the sustainment of the tumor are elevated, in-

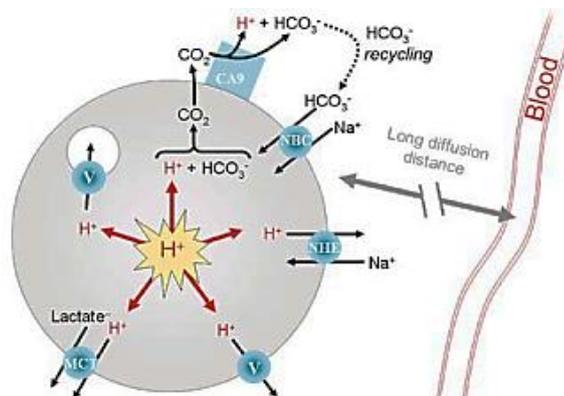


Figure 2-15. The regulation of pH_{IS} in tumors. After the production of protons, passive (buffering) and active extrusion (V-ATPases - primary active - and the NHE and NBC exchangers - secondary active) mechanisms are initiated in order to maintain pH_{IS} at normal levels. The result is to drop the pH_{ES} due to the abnormal blood perfusion [182].

creases the glycolytic flux [100]. Then, logically, one would expect that the abrupt increase in lactic acid production would be threatening cell's homeostasis. However, cancer cells manage to prevent glycolytic protons from being cytotoxic ($\text{pH}_{\text{IS}} \sim 7-7.4$) by driving lactic acid to the interstitium, through the elevation of the kinetics of direct and indirect clearance mechanisms, consequently causing tissue acidification ($\text{pH}_{\text{ES}} \sim 6.4-6.9$) [89], [94], [95], [101].

As the maintenance of pH_{IS} at normal levels is of great importance for tumor cells to survive, their ability to tolerate acidic conditions is of equal importance by the time this tissue acidification is considered to favor tumor expansion and has been linked with invasion and metastasis [97], [102]. Early investigations have shown that lactic acid was solely responsible for the acidification. However, another governing source of extracellular acidity in tumors has been proposed that principally precedes lactic acid, the CO_2 [90], [91]. CO_2 is abundant in tissues in the form of $\text{CO}_2/\text{HCO}_3^-$ buffer pair. Upon acidification from lactic acid the $\text{CO}_2/\text{HCO}_3^-$ pair is titrated yielding CO_2 intracellularly that is readily extruded due to the outward pCO_2 gradient (description below). There, the existence of the, recently explored, carbonic anhydrase (CA) enzyme, that catalyzes the CO_2 hydration yields back the acid that was buffered intracellularly. The latter is imprisoned extracellularly due to the deficient waste-product vascular clearance causing, again, tissue acidosis. Besides, if CO_2 was to remain in the cell and hydrate, it would cause cytoplasmic acidosis that would have led to its apoptotic death. The only way that CO_2 can re-enter the cytoplasm is through active mechanisms (Figure 2-15) in the form of bicarbonate ion. Therefore, this recirculation phenomenon is dual, it keeps pH_{IS} while lowering pH_{ES} and has been recently proposed as a target of many anti-tumor strategies [91], [103].

The ability of the cancer cells to maintain their pH_{IS} close to normal values and to proliferate, in turn, in an acidic extracellular environment has given rise to the so-called "acid-mediated tumor invasion model" [99]. This model provides a simple mechanism linking the reduced pH_{ES} with the ability of tumor cells to form invasive cancers. Particularly, it has been shown experimentally that H^+ flow to peritumoral normal tissue provokes normal cell necrosis or apoptosis and extracellular matrix degradation. Tumor cells develop resistance to acid-induced toxicity during carcinogenesis, which permits them to invade the damaged normal tissue.

2.3.2 Structural signatures of carcinogenesis

Beside the maintenance of the regulation of the pH_{IS} in order for the cancer cells to proliferate, the intercellular control of cell growth, especially that mediated by connexins and cell adhesion molecules like the claudins, appears to play a critical role in carcinogenesis. As a matter of fact, for cancer cells to grow without constraint from surrounding normal cells, it may indeed be advantageous to avoid communicating with them by creating a mechanism to block heterologous gap junction intercellular communication. Furthermore, it is widely accepted, that the expression level of proteins that mediate adhesion at the tight junctions and desmosomes directly associates with the degree of differentiation of cancer cells [104]. These functional misbehaviors that lead to structural tissue alteration are briefly considered next.

Aforementioned, the gap junctions are intercellular plasma membrane domains enriched in channels that allow direct exchange of ions and small molecules between adjacent cells. Connexins, the building blocks of the channel, are responsible for regulating cellular growth and differentiation. Persistent perturbation mechanisms ranging from loss of connexin gene transcription to aberrant trafficking of connexin proteins of the gap junction complex

can cause chronic effects (*e.g.*, cancer) [105]. In fact, cancer cells usually have downregulated levels of gap junctions, and several lines of evidence suggest that loss of gap junctional intercellular communication is an important step in carcinogenesis. The inhibition of the intercellular communication through the gap junctions is a common fact among tumor promoters. Furthermore, the loss of homologous gap junction intercellular communication among cancer cells themselves has been shown to contribute in the increase of their heterogeneity and those cells with more malignant phenotypes may gradually dominate the tumor cell population during the progression process [106]. In support of this hypothesis studies have shown that the connexins remain reduced quantitatively in the tumor progression stage and may be qualitatively altered in metastasis since connections are made between the primary tumor cells and foreign host cells at the secondary metastatic site.

Along the same lines, tumor promoters target the dysfunction of the tissue adhesion structures in order for the cancer cells to expand unattached to the surrounding normal tissue. This disruption of the state of the tight junction and desmosome adhesion, which are responsible for maintaining the coherent morphology of an epithelium, plays a major role in malignant transformation and metastasis [104], [107]. For example, perturbation in the CLDNs, *i.e.* the proteins that comprise the tight junction, has been associated with several cancers indicating altered polarity, abnormal cellular organization and loss of differentiation [108]. In fact, a common feature of carcinogenesis is the dysregulation and aberrant localization of these proteins [104]. The former clearly suggests that, at precancer stages, there is an increase in the number of tight junction complexes which, in turn, increases the density of the tight junction pores and, for this reason; it has been associated with increased tissue permeability [109]. The latter, reveals that the tight junctions are remodeled in various locations promoting an overall junction dysfunction. These facts lead to the loosening of the junctions, increasing the possibility of metastasis. Over and above, the extracellular pH drop explained earlier, increases the permeability of paracellular route for non-charged solutes and changes the ion flux profile of the barrier [67]. Therefore, there is an obvious correlation between protein downregulation and the junction inactivation both of which play a major role in malignant transformation and metastasis.

Particular interesting attempts have been made to adjoin regular tight junctional behavior and gap junctions [110]. Argument on the matter is brought on because of the homeostatic purpose of gap junctions and their maintenance feature of general tissue's functions. It has been shown that the gap junction pathway inactivation influences the tight junction complex accordingly. On top, as alluded to above, gap junctions are unexpressed or aberrantly localized in cervical cancer cells [111], [112]. Thus, paracellular permeability is indirectly influenced by the lack of intercellular communication and metabolic cooperation in the cervical dysplastic epithelium.

2.4 CANCER OF THE CERVICAL EPITHELIUM

The cervix is the portion of the uterus attached to the top of the vagina (Figure 2-16). Generally, the cervix is covered by two epithelial tissue types, namely: the columnar epithelium and the stratified squamous epithelium. The former is a single layer cell structure typically found in the cervical canal while the latter consists of several cell layers with cell morphology transitionally converted from cuboidal to squamous cells at the apical surface. Somewhere on the cervix the two cell types meet at a place called Squamo-Columnar Junction (SCJ). During the woman's life span, columnar epithelium is converted progressively to

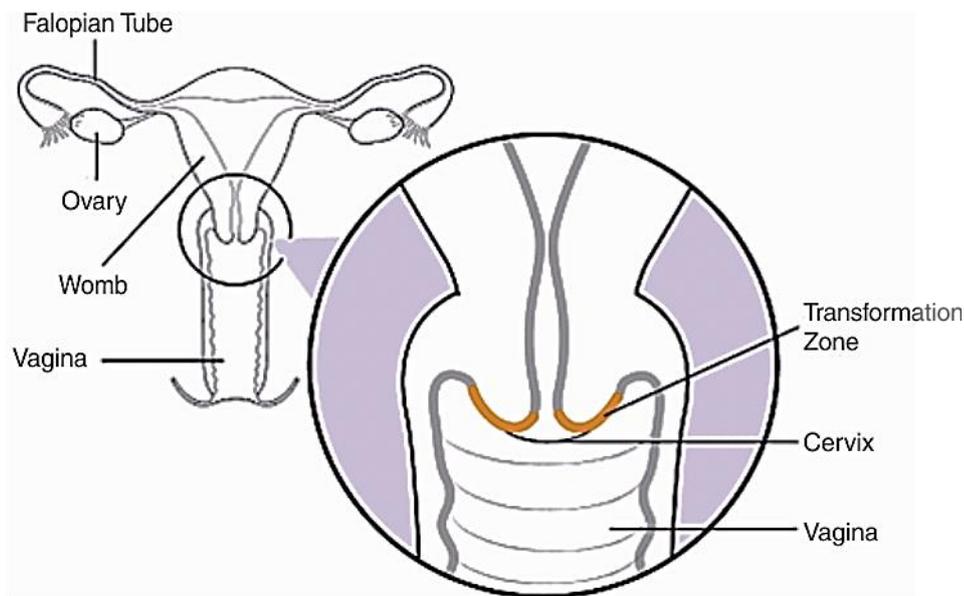


Figure 2-16 The uterine cervix
 squamous and for this reason SCJ is often called Transformation Zone (TZ). CIN tends to form in the vicinity of the TZ.

As indicated in the introduction, the cervical cancer is a major health care issue. It arises as the rather rare end-point of a series of stages, which begin with a sexually transmitted high risk HPV virus sub-type infection of the cervical epithelium. The infection may progress slowly to the immediate precursor of invasive squamous carcinoma, which is known as Cervical Intraepithelial Neoplasia (CIN) grade I, II, III, Microinvasive and Invasive (Figure 2-17). Grade II or more severe lesions are characterized as High Grade Squamous Intraepithelial Lesion (HSIL), while grade I or less severe lesions are characterized as Low Grade (LSIL). In the next paragraphs the physiology of the cervical epithelium in normal and precancer stages is described.

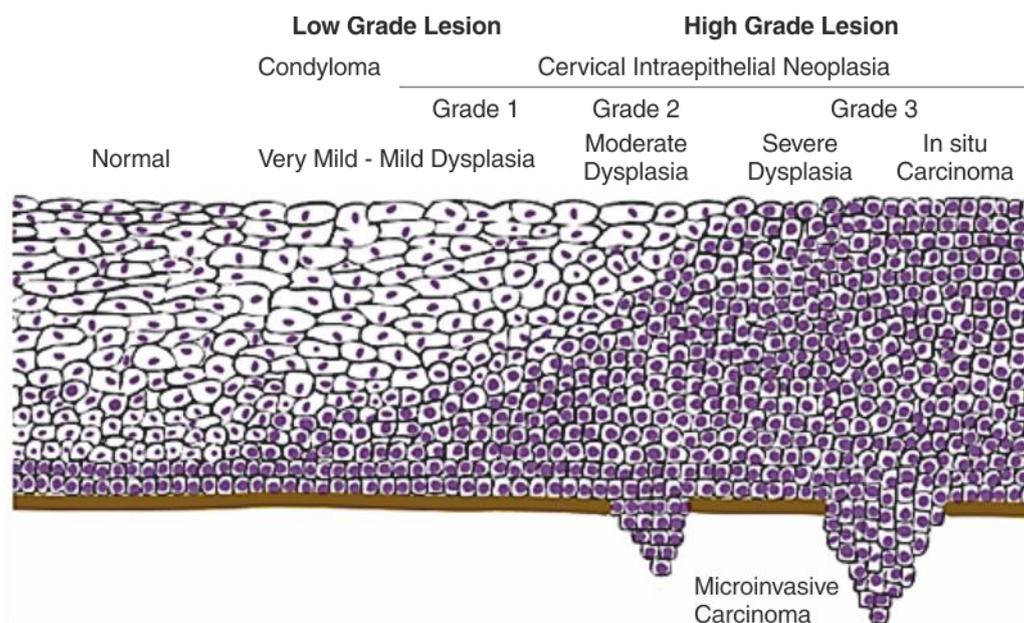


Figure 2-17 Classification of cervical intraepithelial neoplasia

2.4.1 Morphology of normal and abnormal cervical epithelia

The building of the cervical epithelium structure originates in the cuboidal cells that lay at the base of the tissue (basal layer). These basal cells are metabolically active and one of their activities is to proliferate. Initially, the descendant cells are also cuboidal but, as they mature and propagate towards the upper layers of the tissue they progressively enlarge and flatten. A snapshot of a physiological cervical epithelium cross-section depicts a diverse group of cells that are progressively transformed from cuboidal to squamous, starting from the bottom, up. Hence, the cellular stratification creates several layers which can be narrowed down to the following: basal, parabasal, intermediate and superficial layer (Figure 2-18).

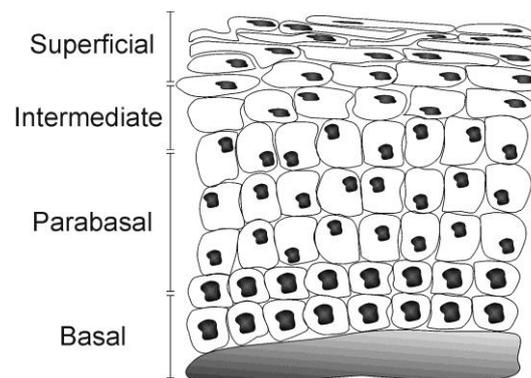


Figure 2-18 Cervical epithelial layers

Such stratified setup is altered dramatically during carcinogenesis which starts at the base of the epithelium. In general, CIN is a continuous spectrum of change, whereby the epithelium may become progressively more abnormal in appearance, or neoplastic. This occurs because the metabolically active basal cells switch genetically and despite the fact that they maintain their ability to propagate, they yield atypical cells, similar to themselves, that do not have the ability to grow and to adhere with each other. Under these circumstances, abnormal tissue portions of evenly shaped cells of enlarged extracellular spaces are generated which, depending on the severity of the disease, can span the entire epithelium. Thereupon, it is common to observe macroscopically that the cellular volume has been reduced, while the volume of the nucleus has increased, resulting in an increase of both nuclear density (Nuclear-to-Cytoplasmic Ratio - NCR) and extracellular space [113].

Figure 2-17 illustrates the basic structure of the cervical epithelium together with the structural changes provoked during the progress of the disease. Viral genomes, that manifest high-risk HPV types, are maintained as episomes in the basal layer, with viral gene expression being tightly controlled as the infected cells move towards the epithelial surface [114]. Common histological grading correlates CIN with the proportion of the total epithelial thickness affected by neoplasia. For instance, tissue diagnosed as typical CIN I will be normal throughout the upper two-thirds of its thickness, but enlarged nuclei may be visible in the lower third. Correspondingly, in a typical CIN II case the lower two-thirds of the epithelium may appear abnormal, whereas in the case of the CIN III, abnormal nuclei and cell structures may be present throughout the entire cross-section of the epithelium. Nevertheless, there are no quantitative descriptors for the CIN grades and, thus, histological diagnosis is highly subjective.

2.4.2 Physiology of the cervix during CIN development

Consistent with the general epithelial structure under physiological circumstances, cervical cells are firmly adjoined via desmosomes and tight junctions while their intercommunication channel is facilitated by the gap junctions. The TZ of the cervical epithelium, where CIN develops, contains no gland cells and thus, secretion and absorption are managed by the transcellular and paracellular pathway alone. Vaginal mucus secretion is delivered mainly from the tissue of the endocervix. Withal, nutrient absorption is carried through both the paracellular route and the direct transmembrane communication of the apical cell layers with the cervical lumen. On the other hand, intercellular communication through the gap junctions is always operational and diffusion is feasible in all cell layers. During the development of neoplasia, however, the proteins that construct these complex, multitasking, structures become dysfunctional.

2.4.2.1 Loss of intercellular communication

In the cervical epithelium there are three main types of connexins present, Cx26, Cx30 and mainly Cx43 [78], [112], [115]. Immunofluorescence staining and phosphorylation methods revealed the localization of connexins inside the cervical epithelium layers and also that they exhibit chemical gating with acidification [116]–[119].

Recent studies have shown a direct relation between gap junctions, connexins, cervical neoplasia and the HPV virus [78], [112], [115]. Comparing normal and neoplastic specimens, authors argued that cervical carcinoma is starkly related to the gap junctional pathway deficiency and connexin aberrant localization during carcinogenesis and tumor growth. In [112], authors correlated normal and CIN affected epithelium using connexin phosphorylation. Gap junction evanescence was also governed to be the offspring of HPV virus in both [78] and [112] and that the situation is due to reduced expression rather than localization in cytoplasm. Finally, in [115] aberrant localization of connexins was specifically examined for various epithelial cancer models. As the authors suggested, gap junctional intracellular communication can be found either among cancer (even deficiently) or normal cells. However, irregular cytoplasmic localization of connexins is due to cancer stage (possibly by cadherin inexpression) which is a general phenomenon of carcinogenesis.

According to these findings, connexins may play a role in promoting carcinogenesis by the time tumor, in contrast with normal, cells communicate in an extremely difficult manner. In conclusion, gap junctions control the intercellular cervical epithelial transport pathway, integrating metabolic activities of all cells in the tissue. However, a remarkable reduction of the cell-to-cell cross talking and metabolic cooperation is associated with the development of neoplasia.

2.4.2.2 The coherence of the cervical tissue in neoplasia

In a recent study, the composition of the endocervix (columnar epithelium) and of the ectocervix (stratified epithelium) regarding its adhesion junctions has been established [120]. Specifically, the distribution of the tight junction structures has been investigated by electron microscopy, and permeability assessment was made possible by tracking the penetration of the fluorescent immunoglobulin G (IgG). The results indicate that the endocervical cells are joined by tight junctions while the most apical layers of the ectocervical stratified squamous epithelium do not contain classical cell-cell adhesions. In contrast, molecules characteristic of exclusionary junctions are detected three to four cellular layers below the luminal surface of the ectocervical stratified squamous epithelium and extend to the basement membrane.

The most robust junctions of the ectocervix occur just above the basal layer in contact with the basement membrane. In addition, adheren junctions (desmosomes) are particularly abundant throughout the whole body of the tissue. Summing up, the natural integrity of the junctions progressively lessens as epithelial cells are pushed toward the apical surface, where they become cornified, lose all cellular contacts, and are sloughed into the lumen.

During carcinogenesis the architecture of the cervical squamous epithelium is gradually altered starting from the basal cell layers and extending to the superficial ones. As a result, the functionality of the tight junctions decreases while the desmosomes ease off during CIN [64], [66], [107], [113], [121]. The reason for this is the atypical expression of the adhesion proteins (claudins and cadherin) which tends to increase, contrary to common cancer phenotype [122]. Specifically, during CIN development, the CLDNs found in the TJ complexes of the cervical epithelium (CLDN's 1, 2, 4, and 7) are up-regulated and aberrantly localized. Up-regulation suggests that, at precancer stages, there should be an increase in the number of TJ complexes and, in turn, a decrease of solute permeabilities. Aberrant localization, however, reveals that the TJs are remodeled in various locations promoting an overall junction dysfunction. One possible explanation to such controversy may be ascribed to an adaptation response of the cell to preserve barrier function [123]. Even so, it is still an open subject whether the solute permeabilities increase or decrease due to CIN. Nonetheless, under physiological circumstances, the overexpression of these proteins has been reported to increase paracellular flux of non-charged solutes [109], [124], [125]. Although neoplastic tissues do not conform to physiological ones it is highly probable that an increase to solute permeabilities should also apply in atypical lesions of the cervix. For this, it is generally accepted that the cervical epithelium becomes "leaky" letting the free exchange of ions and solutes. Thereupon, the extracellular environment becomes spacious, probably allowing greater diffusivity to ions and molecules.

2.4.3 Probing altered scattering properties of the tissue

We have already mentioned that chemical substances, known as optical biomarkers, are used in order to increase the acuity of physicians. Generally, imaging techniques are limited by the inherently weak optical signals of endogenous chromophores and fluorophores and the subtle spectral differences of normal or diseased biological samples. In order to punch up these weak optical signals contrast agents have been developed that probe molecular specific signatures [5]. These exogenous labels may be an antibody, a photoprotein, inorganic nanoparticles, organic substances or fluorescent semiconductor nanocrystals (Quantum Dots). Their concentration and/or spectral properties alter by the biological processes under investigation after linking with high selectivity to specific targets. That is the case; diagnostic procedures have become more accurate improving the early clinical diagnosis of fatal diseases such as cancer.

In the cervical cancer's case, topical application of acetic acid (AA) solution 3-5% is routinely used as a contrast agent for more than 70 years in order to highlight abnormal areas. The agent-tissue interaction generates an optical signal, which is perceived as transient tissue whitening. Clinical evidence supports that the degree and duration of the latter is associated with the lesion's grade and the phenomenon is known as acetowhitening (AW) effect. In practice, the method dictates the application of acetic acid to the cervix exterior and then visualizing the biochemical reaction on it. The phenomenon can be observed under incandescent light with or without magnification producing a low intensity chemiluminescent

light. This chemiluminescent light allows the physician to have a subjective interpretation of the epithelial condition *in vivo*.

Typically, with high-grade CIN, there will be an almost instant response within about 50-60 seconds and the acetowhitening will appear markedly white as a greater part of the epithelium consists of atypical cells. The effect is then slowly reversed and finally disappears within about 40 seconds. With low-grade CIN, the onset of white is commonly delayed because the acid must penetrate into the lower half of the epithelium where the atypical cells are situated [126].

Next, based on our previous discussion concerning the metabolic activities of neoplastic tissues, we attempt to interpret the kinetics of AA and correlate them with the transient phenomenon.

2.4.4 Biophysical and biochemical phenomena followed by the application of the contrast agent

After the topical application of acetic acid (AA) dilute solution onto the surface of the cervical epithelium, the acid molecules will be partially disassociated to hydrogen (H^+) and acetate (Ac^-) ions according to the chemical reversible dissociation equation:



Being a weak electrolyte, the amount of AA disassociation depends on its dissociation constant ($pK_a^{37C}=5.1$) and on the pH of the buffered solution it is diluted in i.e. on the pH_{ES} and on the pH_{IS} of the cervix and its cells. As said, the normal level of pH_{ES} is roughly equal to 7.3, while in the majority of tumors pH_{ES} is between 6.4 and 6.9. At the same time, at both normal and neoplastic epithelia the pH_{IS} is maintained at neutral levels (7.3-7.1) [60-62, 80-84]. It follows that a pH gradient builds up between the IS and the ES in tumors which significantly favors the intracellular accumulation of weak acids. Subsequently, because the cell membrane is impermeable to charged species, the latter that are produced by the acid's dissociation will be trapped inside the cell and interact with the environment (protein denaturation, etc.). This phenomenon is called "ion trapping" and in pharmacology such cancer specific signature is exploited for the design of anti-tumor treatment procedures [32], [127].

To prove this mathematically, the Henderson and Hasselbalch equation for buffered solutions can be manipulated to give the ratio between the intracellular and extracellular concentrations (C_{IS} , C_{ES}):

$$C_{IS}/C_{ES} = \left(1 + 10^{pH_{IS} - pK_a}\right) / \left(1 + 10^{pH_{ES} - pK_a}\right) \quad (2.2)$$

By substituting the above mentioned pH values to equation (2.2), it yields that the ratio C_{IS}/C_{ES} is approximately 1.6 for normal epithelium while 2.5 for the neoplastic tissue. This clearly indicates that a greater amount of AA is accumulated into the IS of the neoplastic cells. In addition, a careful analysis of the epithelial biophysics suggests that the previously discussed differences in the intracellular-to-extracellular volume ratio, between normal and neoplastic epithelium, should be an additional factor determining the inwards flux of the un-ionized AA. The structure of the normal cervical squamous epithelium dictates that the IS at the utmost layers, gradually aggrandizes in volume, while the extracellular space falls away [113]. As a result, the hydraulic pressure applied from the intracellular space overcomes the pressure of the extracellular space, thus minimizing the inward passive diffusion flux in nor-

mal epithelial cells. In contrast, in neoplastic epithelial cells, the aforementioned volume ratio is reversed, further favoring the inwards flux of AA. Summing up, the general consensus is that the AA molecules have an intrinsic tendency to accumulate in the intracellular space of neoplastic cells.

Under these circumstances, let us assume the administration of AA to a buffered solution such as the layered epithelial tissue is, with potential neoplasia of any CIN stage. Upon application, the more likely scenario for the AA molecules is to encounter the normal cell layers from where they could potentially be transmitted across the epithelium following either the transcellular and/or the paracellular path. However, most of the initial AA will dissociate instantly into hydrogen (H^+) and acetate (Ac^-) creating a bi-ionic system $H - Ac$ (the valences + and - will be dropped henceforth). Being in this bi-ionic form, AA is denied an entry to a cell. Consequently, it will diffuse mainly through the paracellular route i.e. the tight junction ionic channels. Earlier we discussed that the superficial layer is devoid of tight junctions which means that, essentially, AA will penetrate mostly unobstructed through the upper portion of the tissue. In the case where some amounts of AA molecules penetrate into normal cells, it is reasonably expected that they will be disseminated to the neighboring normal cells through the well-functioning gap junctions, thus enabling metabolic cooperation. If the AA influx continues and the acidic load exceeds a certain level, the gap junctions are closing in order to protect the neighboring cells from this high acidic load that could irreversibly damage them [58], [79], [82], [116].

By reaching an area sealed with tight junctions, the concentration of the diffusing AA bi-ionic system starts to build up. However, the accumulation of H^+ ions causes the abrupt fall of the local pH_{Es} levels. This reaction acts as stimulation for the tight junctions to break the barrier and allow AA to further penetrate the tissue until it reaches the neoplastic area. In support of this hypothesis the amount of AA molecules that intrude the intracellular space of normal, tightly joined cells also lowers the local pH. In fact, it has been found that the intracellular acidification produces cytoskeletal contraction in normal epithelial cells which, in turn, opens the tight junction barriers, creating macromolecular gaps between them [63], [128].

Beside these mechanisms, the effective tissue penetration areas exposed to the AA can also increase due to several naturally occurring conditions. These include: crypt openings which are better known as “cuffed gland openings”; different tissue borders such as the transformation zone; immature metaplasia; and already compromised cell junctions [129]. Therefore, the effective penetration area exposed to the AA can be quite large.

Then, upon reaching the acidic environment of the extracellular space of a neoplastic cell, the bi-ionic system partially recombines and a considerable fraction of the AA molecules diffuses passively to the intracellular space with high selectivity. There, due to the almost neutral pH of the cell's interior, the intruded AA molecules readily disassociate into Ac and hydrogen H ions, anew. As a result the intracellular pH drops acutely, according to the ion trapping mechanism described above. In the absence of gap junctions and, in turn, metabolic cooperation this acidification is encountered solely by triggering the intrinsic cellular pH regulating mechanisms. Then, in order to consume the generated protons, the cell employs several short and long-term mechanisms (buffering and ion pumping) in order to gradually reverse this phenomenon.

The AA molecules and ions that did not manage to pass the neoplastic cell's membrane together with the ones that are extruded by the cell continue to diffuse along the ab-

normal tissue cross-section which means that these processes are repeated in the underlying neoplastic epithelial layers. The phenomenon is favored from the fact that, since the tight junctions are leaky and dysregulated in neoplasia, the porosity of the neoplastic portion of the tissue is increased which, in turn, allows for the rapid spread of the AA molecules. Eventually, the epithelium restores its original state when AA is consumed or drained through the stroma to the blood circulation and to the lymphatic system.

2.4.5 Interpretation of the dynamic scattering characteristics of the optical phenomenon: correlation with the kinetics of the biomarker

The cervical epithelium appears transparent in the visible spectrum because normal cells and, most importantly, their nuclei act as Mie-scattering particles, promoting the forward scattering of photons [130], [131]. For this reason, a typical optical examination of the tissue with white light clearly depicts the underlying connective tissue. Particularly, it has been found that it is mostly the constituents of the stroma that contribute to the overall tissue appearance i.e. the collagen fibers that backscatter incident photons and the hemoglobin in the vasculature that significantly absorbs the incident light [132]–[134].

Upon application of the contrast agent the atypical areas of the epithelium turn white or opaque and it is then quite easy to distinguish them from normal epithelium which appears pink. Confocal and phase contrast microscopy have both demonstrated that this phenomenon is predominantly developed in the nucleus of the cell [135], [136]. Furthermore, experiments presented in [126] have shown that dicarboxylic acids such as lactic acid is not producing AW phenomenon. This explains why although lactic acid is produced constantly in neoplastic cells, no native AW effect is observed. The same study demonstrated that sodium acetate did not produce any AW effect either. Both observations indicate that AW effect is not associated with the acetate ions. It is therefore reasonable to suggest that the AW effect is associated with either the intracellular concentration of the unionized AA or with the H concentration.

It is, however, well known that the fluctuation of pH_{IS} can provoke conformational changes in cellular components. In fact, a slight decrease of pH_{IS} can influence two possible mechanisms: (1) the increase in the cytokeratin polymerization of the cytoplasm of epithelial cells which, in turn, alters the thickness of their filamentous architecture; and (2) the provocation of deacetylation of the histones that package and order the DNA, which results in a dense chromatin structure [137]. Both effects of AA-induced intracellular pH drop are changing the size of cell scatterers which decreases light scattering in small forward-scattering angles while significantly increases the scattering of cells in the large angles [135], [136]. Particularly, this phenomenon has been shown to result in an increase in the magnitude and frequency of fluctuations in the local refractive index [130]. Consequently, the backscattered photon increase accrues to small-sized-AA-induced cellular components in the cytoplasm and the nucleus. In other words, the predominant effect of acetic acid is an alteration of the refractive-index structure of the nucleus. Therefore, protein denaturation, fixation and cross-linking are responsible for causing the local variation in the refractive index and consequently an alteration in the light scattering characteristics. Obviously, this can explain the alterations in the scattering properties of the abnormal epithelium, observed as AW phenomenon. Figure 2-19 illustrates these notions.

The determination of the actual nucleic components comprising the biological origins of the AW phenomenon lies outside the scope of this thesis and it is a subject of our on-going research in the field. Nevertheless, the reversible nature of the AW effect indicates that the

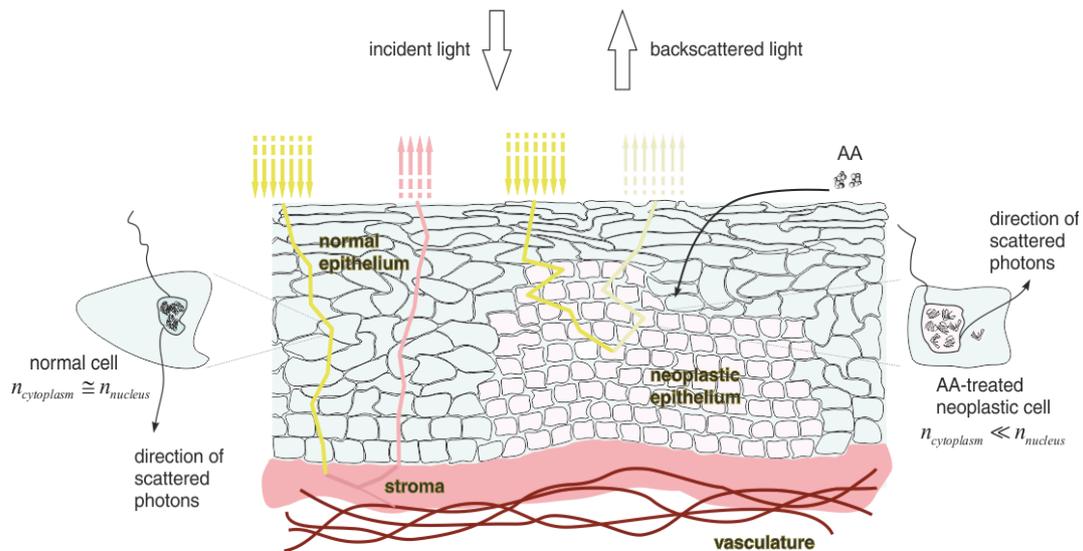


Figure 2-19 Photon random walks in normal and atypical, AA treated, areas of the epithelial tissue. Protein denaturation, fixation and cross-linking provoked by weak acids change the size of cell scatterers, especially the nucleoproteins. This phenomenon decreases light scattering in small forward-scattering angles while significantly increases the light scattering of cells in the large angles. Particularly, an increase in the magnitude in the refractive index of the nucleus (n_{nucleus}) has been shown.

associated biological processes should also be reversible after AA removal from the cell. It is well known, however, that reversible biological processes, such as protein denaturation, fixation and cross-linking provoked by weak acids are very fast following the concentration changes of the latter. For example, proteins restore their native structure spontaneously and within a ns/ μ s time scale after denaturant's removal [138]. Then, provided that the AW effect lasts for more than 4min, we can reasonably adopt that the phenomenon's dynamics is not determined by the intrinsic time characteristics of the conformational changes; rather it reflects the dynamic characteristics of the change causing factor, which is the intracellular proton concentration dynamics. This is a very important assumption by which we are allowed to study the correlation between the agent's pharmacodynamics, the tissue characteristics and the observed dynamic signals. Then, subject to the amount of correlation of the former two with the macroscopic light scattering properties of the tissue, the quantitative assessment of the AW phenomenon could provide a means for the *in vivo* assessment of functional alterations associated with the development of the neoplasia. Moreover, it is expected that the intracellular AA concentration dynamics at each epithelial layer will be affected by the cell packing and the nuclear density of the entire epithelium.

The measuring and modeling of the macroscopic AW dynamic scattering characteristics can render possible the assessment of both microstructure and functionality of the cervical epithelium, *in vivo*. It is expected that besides the intracellular proton concentration, the AW scattering characteristics are also affected by the concentration of nuclear components that are subjected to conformational changes in the presence of H^+ . Such a component may be the nucleoprotein, the concentration of which is increasing with the neoplasia grade. Nevertheless, for the rest of this thesis we will assume that the AW kinetics is largely determined by the pH_{IS} , while the concentration of nucleic AW scattering centers is expected to affect the static scattering intensity. Then, the developed model's output –the concentration of H_{IS}^- will be compared with the experimental data and the validity of the aforementioned assumptions will be assessed and discussed.

For completeness, it must be noted that this acetowhite appearance is not unique for neoplasia and can be seen on other occasions of increased presence of nucleoproteins e.g. during metaplasia, in columnar epithelia, during healing and at the presence of virus or viral products [126], [139]. Lastly, swelled tissues or epithelia with increased keratin filament proteins have also been shown to turn white by acetic acid [139], [140].

2.4.6 CIN assessment using the DCE-OI apparatus

In the introduction we have indicated that attempting to improve the accuracy of the clinical diagnosis of CIN, the dynamic optical characteristics of the AW effect have been measured *in vivo* [15], [141]. A brief description of the technical characteristics of the device and of the data collection procedure is provided next.

2.4.6.1 Specifications

The DCE-OI device integrates an optical imaging head supported by a mechanical basis, which includes weight counter balancing and space translating mechanisms. The basis allows for the easy manual positioning of the optical imaging head to obtain a sharp image of the tissue and includes mechanical and electromagnetic brakes that are manually activated through a trigger, once the optimum field-of-view has been achieved. In order to maintain approximately the same field-of-view throughout the entire examination procedure, the optical imaging head is detachably connected through a mechanical shaft attached to a vaginal speculum. Once the latter has been inserted into the vagina, its blades are opened to enable the visualization and the imaging of the cervix of the uterus. Tissue imaging is performed with the aid of a 35 mm lens, coupled with a 1024X768, 8bit/channel color CCD camera. The camera is interfaced with a dual core microprocessor computer unit, through a fire-wire (IEEE-1394) cable. The optical imaging head further comprises a white Light Emission Diode (LED), coupled with light collimating and focusing optics for uniform tissue illumination. A pair of polarizers, one disposed at the imaging and one at the illumination pathways having their polarization axes perpendicular with each other, is employed for the cutting-off of the surface reflection originating from the air-tissue interface and of the glare originating from the mucus. The optical head has been configured to capture images from a 23mm X 20mm tissue area, including the entire transformation zone of the cervix. Assuming an average 35 μ m cell diameter and given the aforementioned sensor's spatial resolution, it can be easily calculated that the back-scattered light intensity recorded by a given pixel, corresponds approximately to the area occupied by a single cell. Technical evaluation shows that the sensor has a substantially linear response across its entire dynamic range. In order to ensure device independent and reproducible imaging, systems calibration is performed using a Ba_2SO_4 calibration plate, with unity reflectance across the visible spectral range.

2.4.6.2 Data collection

After the calibration procedure has been successfully completed, the patient is placed in the lithotomy position and the vaginal speculum is inserted, thus enabling the visualization of the cervix (Figure 2-20(a)). This is performed according to the typical procedures followed in routine colposcopy. Then, the initiation of the image capturing procedure is triggered by an AA applicator in order to ensure synchronization between the acquisition scheduling and the phenomenon. The AA applicator is a nozzle connected with a syringe, which has been pre-filled with a 2 ml AA solution 3% (0.5M). A reference image is captured before the application of the AA solution, which is followed by an automatic image capturing of images during the evolution of the AW phenomenon, for 4 minutes (Figure 2-20(b)). The video stream is post-processed for image registration, in order to compensate for both image translations

a) clinical setup (www.dysismedical.com)



b) observations



c) pixel-by-pixel recordings

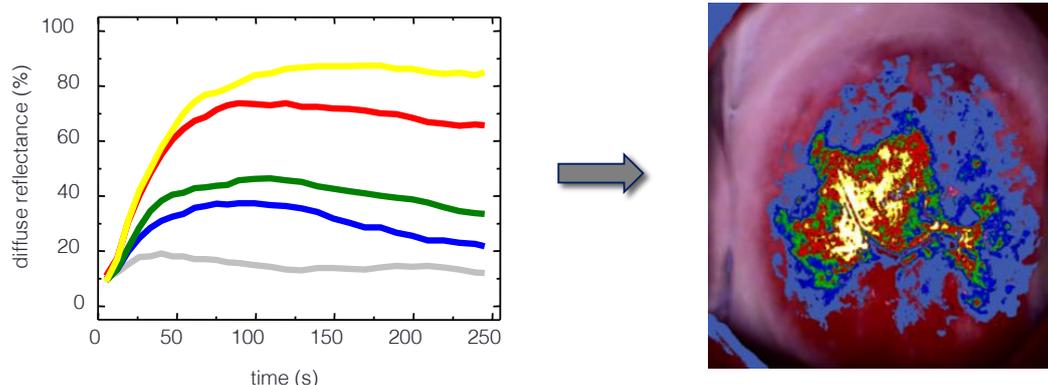


Figure 2-20 The DCE-OI system. (a) After calibration the patient is placed in the lithotomy position and the vaginal speculum is inserted, thus enabling the visualization of the cervix. (b) Crude representation of image capturing. (c) From the captured image stack diffuse reflectance curves are calculated for every image pixel, expressing the temporal characteristics of the AW phenomenon. From the corresponding integrals a pseudocolor map is generated and overlaid onto the real-time displayed color image of the cervix.

and deformations, provoked by tissue contractions and patient's micromovements, respectively.

From the captured image stack Diffuse Reflectance (DR) curves are calculated for every image pixel, expressing the temporal characteristics of the AW phenomenon (Figure 2-20 (c), Figure 2-21). From the obtained curves a series of curve features are then calculated such as time integral, DR peak value (DR_{max}) etc. A pseudocolor map is then generated, with different colors representing different feature values, which is overlaid onto the real-time displayed color image of the cervix. This allows for the direct comparison of the feature values, corresponding to the biopsied tissue area with the histological results, thus, enabling their direct comparison with histological classification. Based on this procedure, the evaluation of the diagnostic value of the derived dynamic optical parameters is enabled.

As already said in the introduction, this device has been clinically evaluated in a multi-site clinical trial enrolling 380 women with abnormal pap-test. The results have demonstrated a high correlation between the computed temporal characteristics and neoplasia grades

[16]. Nevertheless, these macroscopic features suffer from the fact that they cannot provide any information about the status of the epithelium. In fact, the diagnostic indices are still based on biopsy samples. Even so, the fact that this DCE-OI procedure captures the kinetics of the diffusible AA may allow for the model-based estimation of the biological changes that are determining the features of the experimental data. The development of a predictive model of the biomarker's uptake kinetics is the main goal of this thesis and will be described in the following chapter.

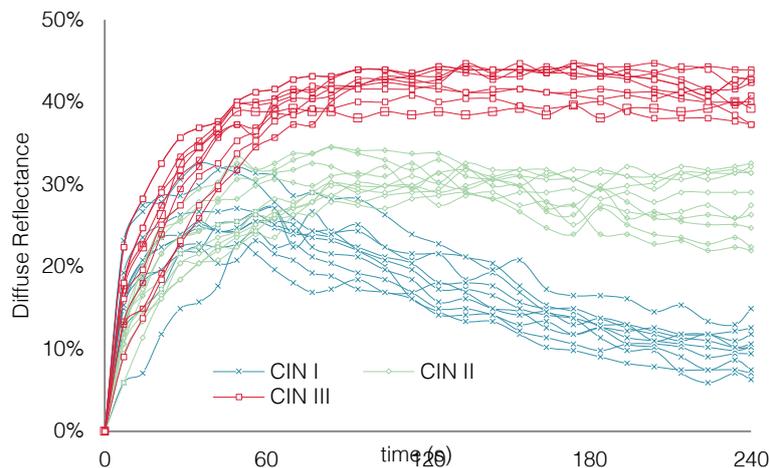


Figure 2-21 Intensity of the backscattered light as measured in CIN tissues

2.5 SUMMARY

From the biological information presented thus far it should be obvious that neoplasia induces several structural and functional alterations in tissues. Distinctively, a pH imbalance rises between the intracellular and extracellular space, primarily, due to the elevated production of lactic acid by cancer cells. For this reason, while cytoplasmic pH is maintained neutral, the extracellular is lowered. This situation allows for the selective accumulation of weak acids intracellularly, where they extensively ionize. This phenomenon favors the biomarker molecules to infiltrate the neoplastic tissues and provoke transient optical effects that relate to the metabolic response of the atypical tissue to the perturbation. For the rest of this thesis, the solidity of this claim will be investigated having as an example the neoplastic tissue of the cervical epithelium.

Along these lines, we presented in this chapter a theoretical framework concerning the AA kinetics in the neoplastic tissue of the cervix. Furthermore, we presented a reasonable link between the metabolic processes and the dynamic, macroscopic phenomenon provoked by the biomarker. In the next chapter, we will describe the mathematical realization of this novel conceptualization that will permit the creation of a predictive model of the tissue metabolic processes. This will allow the interpretation of the phenomenological observations and, potentially, become a novel screening methodology of epithelial pre-cancers.

RELEVANT PUBLICATIONS

- [1] C. Balas, G. Papoutsoglou, and A. Potirakis, "In vivo molecular imaging of cervical neoplasia using acetic acid as biomarker," *IEEE J. Sel. Top. Quantum Electron.*, vol. 14, no. 1, pp. 29–42, 2008

3 *IN SILICO* MODELLING OF BIO-OPTICAL DYNAMICS IN EPITHELIAL NEO- PLASIA

3.1 INTRODUCTION

According to the model building cycle described in the chapter 2, the building of a tractable model requires the definition of its purpose, of the appropriate theoretical background and of a set of necessary conceptual assumptions. Along these lines, in the previous chapter, we presented a suitable theoretical basis on which we can interpret the pharmacokinetics and pharmacodynamics of a biomarker applied onto the exterior of a neoplastic epithelial tissue. Specifically, we have expounded our conceptual assumptions regarding the acetic acid (AA) behavior after its application on a cervical tissue that has developed a pre-cancerous lesion and identified the metabolic origins of the resulting bio-optical phenomenon (acetowhitening–AW). Based on this information, it is possible to develop a tissue-level predictive model by simulating, *in silico*, the biomarker-tissue interaction.

The next step, then, is to choose and evaluate the appropriate mathematical framework upon which the model is built. Additionally, the collection of information regarding the values of the system parameters is also necessary that will allow the performance of problem solution procedures. In this chapter we present all the necessary tools that we have employed in order to put forward a mathematical, *in silico*, representation of the AA–tissue interaction that will render possible the study and the correlation of the measured AW dynamic signals with the underlying biology.

3.2 THE CONCEPTUAL MODEL

The fundamental goal of this thesis is to create a predictive model based on the biomarker–tissue interactions that will allow the foresight of the biological status of neoplastic epithelia, on a macroscopic scale. To this end, the physiological knowledge concerning the creation of the *in vivo* measured dynamic optical signal was described at the second half of the previous chapter. Here, we continue and conclude our conceptualization; that is, we develop the compartmental framework and create the appropriate functional description of the relevant processes and their interconnection.

To begin with, one may consider the cervical epithelium as a naturally occurring, multi-layer assemblage of well-differentiated cells. At precancer stages atypical cells are generated at the basal layer of the epithelium and progressively proliferate towards the superficial

one. From a morphological perspective, these cells are downsized, undifferentiated entities of enlarged nuclei and widened extracellular space. Functionally, on the other hand, neoplastic cells are able to maintain their cytoplasmic pH at physiological levels at the expense of extracellular acidification. At the same time, they lose their polarization and ability to crosstalk with one another because the adhesive junctions relax and the gap junctions cease operating.

Keeping in mind that the abnormal cells are functionally and structurally homogenous and that the respective abnormal cell layers increase with neoplasia progression, we have

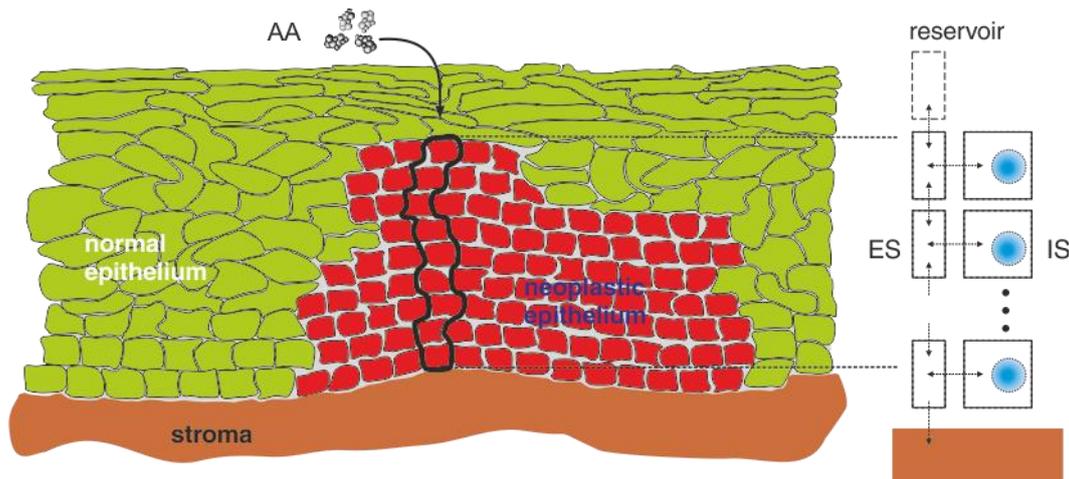


Figure 3-1 (a) Schematic representation of the cervical stratified epithelium; (b) detail of (a) showing a cell column with the basic ES and IS compartments.

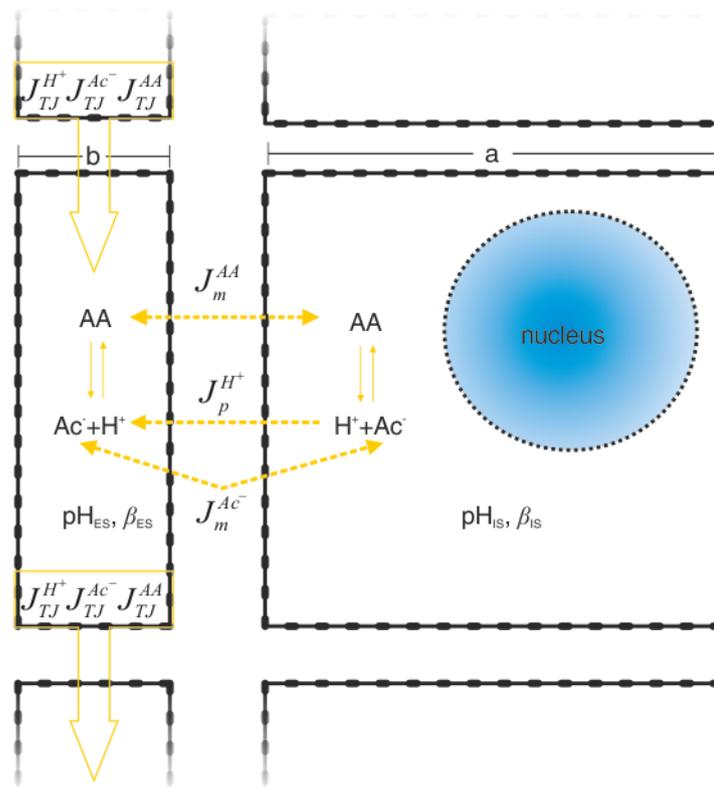


Figure 3-2 Paracellular and transmembrane fluxes of acetic acid and its hydrogen and acetate ions.

partitioned the neoplastic portion of the epithelium into a stack of identical cell elements. The exact number of cell elements in the stack has been designed to be flexible and arise from the number of cell layers appointed to a given CIN grade. In turn, each of these virtual cell layers has been modeled as an entity that comprises of two separate compartments of interaction: the intracellular space (IS) and the extracellular space (ES). For convenience, the IS and ES have been assumed to have cubical and rectangular cuboid shapes, respectively, while both have been considered to be buffered, well-stirred and kinetically homogeneous. Figure 3-1, illustrates the topology of our tissue-level modeling.

As implied this adaptable compartmental framework has been assumed to represent only the atypical portion of the tissue. This limited level of modeling has been adopted because the colposcopically observed normal cervical tissue does not acetowhite, according to clinical practice [136]. It follows that the respective normal cells do not participate, or may participate very little in the production of the AW phenomenon. As a result, our compartmental design has been adopted to encapsulate only the abnormal part of the cervical epithelium. To complement for the existence of the normal tissue portion a compartment that acts as a repository from where the biomarker infuses to the tissue has been added above of the developed framework.

Then, as in all epithelia, diffusion can be realized through the transcellular, the paracellular and the intercellular paths. As indicated, however, in paragraph 2.4.2.1, the deficiency of cancer cells to communicate with each other is a common characteristic of neoplasia. Therefore, since we model only the abnormal part of the tissue, the functional description of compartmental interconnection has been designed to include only the transcellular and paracellular routes. Thereupon and being consistent to our, *in silico*, tissue-level architecture, the AA and its ionic counterparts diffuse passively from the reservoir compartment to the ES of the first neoplastic cell layer through the paracellular route. From there, they diffuse passively, again, from the ES to the IS through the cell's membrane and from the first to the latter layer(s) of the abnormal epithelium through the pores of the, extracellular, deficient tight junctions. Both transmembrane and paracellular passive fluxes are driven by concentration and potential gradients, obeying to biochemical and biophysical laws that will be described shortly. The model also embodies two dynamic intrinsic processes, the AA ion buffering, occurring in both IS and ES spaces, and the active transmembrane pumping processes, which contribute in restoring the original intracellular pH (pH_{IS}) though the extrusion of H ions from the IS to the ES. Figure 3-2 displays the functional description of solute transfer that has been integrated between the tissue compartments.

3.3 MATHEMATICAL IMPLEMENTATION

The simplified version of the tissue topology that was described above reduces the complex physiologic system of the cervix into a finite, tractable, number of manageable entities, through the notion of compartment (paragraph 1.5). In theory, the functional interconnection between adjacent compartments is represented by material flux. The latter, in turn, may depict transport from one location to another, chemical transformation, or both. Here, due to the fact that all cell layers are structurally and functionally identical, material kinetics can be narrowed down to the following:

1. Passive (free) diffusion paracellular flux at the tight junction pores (J_{TJ}) which has been illustrated in (paragraph 2.2.3.1). It corresponds to the interconnection between the ES compartments i.e. consecutive layers, including the repository one.

2. Passive (free) diffusion transcellular flux at the cell membrane (J_m) between the ES and the IS compartment within a cell layer.
3. Active flux facilitated by the cell's pumping mechanisms (J_p) which have been illustrated in (paragraph 2.2.4.2).
4. Chemical transformation flux that accounts for the acetic acid elimination and recombination due to its ionization (J_{ion}) and the chemical buffering mechanisms of the ES (β_{ES}) and the IS (β_{IS}) compartments, respectively.

The mathematical details of the biophysical and biochemical laws that fundamentally describe these fluxes are described in the following subsections.

3.3.1 Modeling the passive (free) diffusion

Theoretically, when a solute (S) diffuses between two compartments of the same solvent that are separated by a permeable membrane, the corresponding flux can be written very simply, in a form proposed by Teorell [142]: *Flux = Mobility × Concentration × Driving Force*. In mathematical terms this translates into:

$$J^{(S)} = U^{(S)} C^{(S)} \left(-\frac{d\mu^{(S)}}{dx} \right), \quad (3.1)$$

where J is the flux, U is the mobility of solute that depends on the barrier; the solvent; the solute; and the temperature, C is the concentration (will be also denoted with brackets ([.]) henceforth) and $d\mu/dx$ represents the force applied in the direction of x that depends on the thermodynamics of the system and also on the concentration and the potential gradients between the two compartments. By reasonably assuming that the temperature remains constant and that the pressure is negligible, integration of the driving force differential term over the membrane's thickness can be possible. Finally, if, in addition, the concentration is assumed to remain constant inside the membrane, then equation (3.1) yields to the Fick's first law formula:

$$J_{I \rightarrow II}^{(S)} = P^{(S)} \left([S]^I - [S]^{II} \right), \quad (3.2)$$

where **I** and **II** denote the diffusion compartments and $P^{(S)}$ is the membrane specific permeability of the solute S .

Fick's law formula can describe the passive diffusion of uncharged solutes where the driving force refers to the concentration gradient, alone. However, if the solute is charged, the $\mu^{(S)}$ term above has two constituents each of which represents the chemical (concentration) and the electrostatic (electrical) potentials, respectively. Then, the Teorell equation yields to the Nerst–Planck equation of electrodiffusion:

$$J^{(S)} = U^{(S)} [S] \left[\overbrace{\frac{RT}{C^{(S)}} \frac{d[S]}{dx}}^{\text{chemical}} - \overbrace{z_s F \frac{d\Psi}{dx}}^{\text{electrostatic}} \right], \quad (3.3)$$

where R is the gas constant, T is the temperature, F is the Faraday's constant, z is the valence of the solute and Ψ is the electrical field that exists between the two boundaries of the

membrane. One special case of the last equation that is very popular is found by assuming that the electric field remains constant inside the membrane. Then, equation (3.3) yields to the well-known Goldman–Hodgkin–Katz (GHK) equation:

$$J_{I \rightarrow II}^{(S)} = \frac{z_S F V_m}{RT(1 - e^{-z_S F V_m / RT})} P^{(S)} \left([S]^I - [S]^{II} e^{-z_S F V_m / RT} \right), \quad (3.4)$$

where V_m is the transmembrane voltage.

Based on the description of the biophysical rules of diffusion just presented, we can reasonably distinguish solute fluxes into uncharged and charged ones according to the state of the diffusing solute. Then, if the solute is uncharged, the Fick's Law (equation (3.2)) should be used to describe its flux. Respectively, if the solute is charged the GHK formalism is the most appropriate except when the transmembrane voltage is zero in which case the Fick's Law should be used instead. In our case we have considered as solutes the AA molecule which is uncharged, and its ionization counterparts Ac and H which are negatively and positively, unitary charged species, respectively. For these solutes, diffusion has been considered to occur via the transcellular path, through the cell membrane and via the paracellular path, through the tight junction pores. Therefore the matching that we have adopted between solutes, pathways and respective flux models is as follows.

Phenomenologically, we know that the cell membrane is impermeable to small ions such as H and that a transmembrane potential is maintained by active mechanisms. These mean that only AA and Ac are allowed to passively cross into the cell's interior and vice versa. Hence, the Fick's Law and the GHK formalism have been utilized to describe their respective diffusion flux (J_m). Equivalently, we know that the tight junction pores behave as ion channels. This means that the paracellular diffusion flux (J_{TJ}) of all solutes is passive (free) through these channels and subject to its resting potential. Since this potential can influence the diffusion of charged solutes i.e. of Ac and of H, the GHK equation has been employed as the appropriate mathematical realization of the respective flux. Respectively, the Fick's Law has been employed in order to denote the paracellular kinetics of AA.

3.3.2 Modeling the active and passive regulation mechanisms upon acidification in tumors

Even if the pH_{IS} regulatory system is largely unaffected by neoplasia, this does not mean its response upon an external acidification would be quantitatively equal. In fact, it is argued that a disorganization exists in buffering capacity; in both the IS and ES where, in the latter case it is diminished [97]; in the free energy either in ATP or NADH form, which is lessened almost 3-fold; and also in the energy stored in ion concentration gradients, which is altered, especially the $\Delta G[\text{Na}^+]_{ES/IS}$ that is elevated [98]. Therefore, the response of a tumor cell to an external acidic load is expected to be less efficient.

Nevertheless, prior to an acidic load, intrinsic and extrinsic buffers will be in dynamic equilibrium in both intracellular and extracellular compartments and the rate with which protons are added will equal the rate of protons extruded. Then, if a weak acid such as AA is applied externally it would perturb the H concentration in both the IS and ES according to the equation (for the IS):

$$\frac{d[\text{H}]_{IS}}{dt} = \frac{-2.303[\text{H}]_{IS}}{\beta_i} \left[P^{(\text{CO}_2)} ([\text{CO}_2]_{ES} - [\text{CO}_2]_{IS}) + k_f [\text{CO}_2] - k_r [\text{H}][\text{HCO}_3^-] + \frac{dQ_{IS}^{(H)}}{dt} \right], \quad (3.5)$$

where the first term of the function in large brackets depicts the diffusion of CO_2 between the ES and IS according to Fick's law (P is the permeability), the second term is the $\text{CO}_2/\text{HCO}_3^-$ buffer titration where k_f and k_r represent the forward and reverse reaction rates, respectively and $(dQ_H)^{\text{IS}}/dt$ defines the rate with which protons are added intracellularly from the rest of the biological processes i.e. the net effect of weak acid's diffusion and the concurrent active extrusion (through proton pumps). Here, we have to stress that β_{IS} is a, time independent, function of pH_{IS} because the intrinsic buffers are instantly titrated and their capacity increases as long as pH_{IS} drops (Figure 2-12b).

Considering the kinetics of the $\text{CO}_2/\text{HCO}_3^-$ pair in tumors, the complexity of equation (3.5) can be reduced. To this end, assume that the ES is, anecdotally, a reservoir of CO_2 and that no buildup of CO_2 occurs in tumor cells, because of their switch to glycolysis. Then, it would have been reasonably expected that an inward pCO_2 gradient exists. However, this is unrealistic because an indirect source of CO_2 exists, namely the $\text{CO}_2/\text{HCO}_3^-$ buffer pair. Biochemically, this buffer absorbs protons produced by glycolysis and extrudes them extracellularly in the form of CO_2 , thus, compensating for the slow hydration of intracellular CO_2 , even in the presence of carbonic anhydrase enzymes [90]. Consequently, to avoid cytoplasmic acidification, tumor cells sustain an outward pCO_2 and a stable $[\text{CO}_2]_{\text{IS}}$ and the amount of protons consumed exactly equals the amount of HCO_3^- that disappears. It follows that the capacity of the buffer is limited only by the availability of HCO_3^- ; whose intracellular concentration is sustained high through the elevated activity of ionic exchangers e.g. the NBC. In other words, we can describe the extrinsic buffering through the kinetics of the bicarbonate solely. Then, upon acute acidification since k_r is small the preexisting bicarbonate will behave as an intrinsic buffer and the NBC exchanger as a transient one. Under such circumstances, we can abolish the first two terms in the brackets of equation (3.5) and lump the bicarbonate's capacity in ($\beta_{\text{IS}} = \beta_i + \beta_{\text{HCO}_3^-}$) and the transient effect in the flux of the NBC exchanger (in Q), respectively. Thus, it yields:

$$\frac{d[\text{H}]^{\text{IS}}}{dt} = -\frac{2.303}{\beta_{\text{IS}}} [\text{H}]^{\text{IS}} \frac{(dQ_H)^{\text{IS}}}{dt} . \quad (3.6)$$

On the other hand, active extrusion depends on several biological factors. Even so, all the mechanisms that were described in paragraph 2.2.4.1 work in synergy and independently which means that the efflux is, in general, an additive function of pH_{IS} . As such, let J_p account for the efflux of all the active extrusion mechanisms and K_p define its peak value. In addition, let, for simplicity, a linear model to be a reasonable fit of the extrusion rates instead of a Hill model (Figure 2-14a). Then, the bell shaped curves in Figure 2-14b that depict the individual activity of the NBC and NHE can be lumped mathematically, along with the V-ATPase operation, using a triangular function i.e.:

$$J_p(\text{pH}_{\text{IS}}) = \begin{cases} \frac{K_p}{\text{pH}_{\text{IS},K_p} - 7.4} \text{pH}_{\text{IS}}, & \text{pH}_{\text{IS}} > \text{pH}_{\text{IS},K_p} \\ K_p, & \text{pH}_{\text{IS}} = \text{pH}_{\text{IS},K_p} \\ \frac{K_p}{\text{pH}_{\text{IS},K_p} - 6} \text{pH}_{\text{IS}}, & \text{pH}_{\text{IS}} < \text{pH}_{\text{IS},K_p} \end{cases} , \quad (3.7)$$

where $\text{pH}_{\text{IS}} \in [6, 7.2]$ which is consistent with the cell's acidic stress tolerance provided that the IS in tumors cells is maintained in physiological levels, $\text{pH}_{\text{IS}, K_p}$ is the pH_{IS} value where the extrusion mechanisms operation peaks (6.6–6.8) and K_p is the peak value. Next, since pH_{ES} has a “chronic” effect on the efflux rate, after an acute acidification the time course of pH_{ES} drop is not expected to alter the pH_{IS} regulation kinetics. Still, the pH_{ES} level that preexisted would have already modulated the peak rate of acid extrusion. Hence, we can express K_p as an exponential function of pH_{ES} (Figure 2-14c) when the pH_{IS} remains constant i.e. prior to acidification:

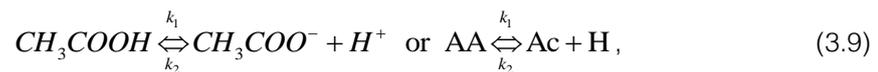
$$K_p = \theta_1 \cdot \text{pH}_{\text{ES}}^{\theta_2} + \theta_3, \quad (3.8)$$

where $\text{pH}_{\text{ES}} \in [6, 7]$ which is consistent with the pH_{ES} values found in tumors. Therefore, equations (3.7) and (3.8) can be used in order to imprint the relation of the efflux rate with the pH_{IS} and pH_{ES} level in tumors. After fitting equation (3.8) to data taken from [89] we found that parameters θ_1 , θ_2 and θ_3 are 1.1×10^{-12} , 9.616 and -3.1×10^{-5} , respectively ($R^2 > 0.98$).

3.3.3 Integration of the functional description into the differential equation system

Having presented the mathematical formalism above, the natural next step towards building a sound model is to appropriately incorporate the available biophysical and biochemical rules into a single system of mathematic relations. Recall that the IS and ES compartments have been adopted to follow the same functionality regardless the number of the cell layers that will be eventually integrated. Then, our description for the coupled mathematical system shown below comprises of the equations that describe the kinetics in a single compartment, say compartment i . This equation subsystem is continuously added to the total differential system until the number of neoplastic layers appointed to a simulation is reached.

To begin with, we elaborate on the ionization of AA in a compartment. Particularly, as it was indicated in paragraph 2.4.4 AA dissociates according to the chemical equation:



where k_1 and k_2 are the forward and backward dissociation rates. According to the mass action law, the AA ionization rate is:

$$J_{\text{ion}} = -k_1 C_{\text{AA}}(t) + k_2 C_{\text{Ac}}(t) \cdot C_{\text{H}}(t), \quad (3.10)$$

where $C(t)$ is the solute concentration at a given time. Because of the acute nature of AA ionization we consider that AA is constantly at chemical equilibrium. Thus, in steady state ($J_{\text{ion}}=0$) the following relationship holds:

$$K_a = k_1/k_2 = [\text{Ac}] \cdot [\text{H}]/[\text{AA}], \quad (3.11)$$

where K_a is the AA dissociation constant. Supposing now that $[\text{TA}] = [\text{AA}] + [\text{Ac}]$ is the total concentration of AA in both its ionized and unionized form then, the undissociated AA that remains in a compartment, is equal to:

$$C_{AA}(t) = \frac{C_{AA}(t)}{C_{AA}(t) + C_{Ac}(t)} \cdot C_{TA}(t) = \frac{C_{H^+}(t)}{C_{H^+}(t) + K_a} \cdot C_{TA}(t) = \xi \cdot C_{TA}(t), \quad (3.12)$$

where ξ denotes the fraction of the total AA (TA) that remains undissociated. Then, starting from the IS compartment where the AW phenomenon occurs and following Boron and De Weer [143], the [TA] in the IS is described by the flux of intracellularly diffusing AA molecules and the Ac's electro-diffusion that combines with free H ions of the cytoplasm in order to produce AA. Then, the [TA] time rate of change is:

$$\dot{[TA]}_{IS}^i = \rho_{IS} (J_m^{AA_i} + J_m^{Ac_i}), \quad (3.13)$$

where the dot ($\dot{\cdot}$) denotes the concentration time derivative and ρ_{IS} is the cell's membrane area/IS volume ratio (A_m/V_{IS}). Equally, the proton load fluctuates due to diffusing AA and Ac, pump extrusion rate and buffering. Particularly, from equation (3.12), we can readily deduce that the [H] time rate of change increases by a fraction $(1-\xi)$ of the entering AA and decreases by a fraction ξ of the electrodiffusive Ac. Even if the former is straightforward considering the dissociation of the diffusing AA into Ac and H ions, the latter is more complex. Primarily, the Ac ions intrude the IS and form AA consuming, according to stoichiometry, equal concentration of protons. The fresh AA molecules, however, ionize yielding back a fraction $(1-\xi)$ of the H ions. In the end, protons are decreased only by a fraction ξ of electrodiffusive Ac. Thus, the proton load imposed on the IS will be:

$$\dot{[H]}_{IS}^i = -\ln 10^{[H]_{IS}^i} \rho_{IS} (\beta_{IS})^{-1} \left((1-\xi_{IS}^i) J_m^{AA_i} - \xi_{IS}^i J_m^{Ac_i} - J_p^{H_i} \right), \quad (3.14)$$

where the term that employs the natural logarithm is the pH to H concentration conversion factor equal to $2.303[H]_{IS}$ as indicated in equation (3.5).

Next, according to Keifer and Roos [144] the definition of ξ can be refined if we recognize that AA has a self-burning effect, i.e., the decrease in pH resulting from a given addition of AA will be less if TA is already present in the solution. In other words, the pH shift in a compartment that AA provokes influences its own ionization. Therefore, the ξ and $(1-\xi)$ terms in equation (3.14) are respectively transformed into:

$$\omega_{IS}^i = \frac{K_a}{[H]_{IS}^i + K_a + 2.3(\beta_{IS})^{-1} K_a \xi_{IS}^i [TA]_{IS}^i}, \quad (3.15)$$

$$\varphi_{IS}^i = \frac{[H]_{IS}^i}{[H]_{IS}^i + K_a + 2.3(\beta_{IS})^{-1} K_a \xi_{IS}^i [TA]_{IS}^i}, \quad (3.16)$$

where ω and φ are the refined definitions. We should note here that these terms are dynamic and depend on the concentrations of AA and proton at any time.

Working similarly to the IS, there are nine fluxes contributing to the [TA] in the ES compartment. These are the fluxes of AA, Ac and H between, the ES and the IS through the cell membrane, the ones between the previous ($i-1$) and the current layer (i) and the ones between the current and the next layer ($i+1$) through the tight junctions. Then, we can write:

$$[\dot{\text{TA}}]_{\text{ES}}^i = \rho_{\text{ES}} (J_m^{\text{AA}_i} + J_m^{\text{Ac}_i}) + \rho_{\text{TJ}} g_{\text{TJ}} (J_{\text{TJ}}^{\text{AA}_{i-1}} - J_{\text{TJ}}^{\text{AA}_{i+1}} + J_{\text{TJ}}^{\text{Ac}_{i-1}} - J_{\text{TJ}}^{\text{Ac}_{i+1}}), \quad (3.17)$$

where ρ_{ES} is the cell's membrane area/ES volume ratio (A_m/V_{ES}) and ρ_{TJ} is the tight junction diffusion area/ES volume ratio ($A_{\text{TJ}}/V_{\text{ES}}$) and g_{TJ} refers to the alterations in the porosity of the tissue due to the altered geometry of the TJ complexes during neoplasia development. Again, using the adoption of Keifer and Roos, but now for the extracellular concentrations, and similarly with equations (3.14)-(3.16) we can write the concentration change rate of the H ions in the ES:

$$[\dot{\text{H}}]_{\text{ES}}^i = -\ln 10^{[\text{H}]_{\text{ES}}^i} (\beta_{\text{ES}})^{-1} \left\{ \rho_{\text{ES}} (\varphi_{\text{ES}}^i J_m^{\text{AA}_i} - \omega_{\text{ES}}^i J_m^{\text{Ac}_i} - J_p^{\text{H}_i}) + \rho_{\text{TJ}} g_{\text{TJ}} \left[\varphi_{\text{ES}}^i (J_{\text{TJ}}^{\text{AA}_{i-1}} - J_{\text{TJ}}^{\text{AA}_{i+1}}) + \omega_{\text{ES}}^i (J_{\text{TJ}}^{\text{Ac}_{i-1}} - J_{\text{TJ}}^{\text{Ac}_{i+1}}) \right] \right\}. \quad (3.18)$$

It should be noted here that for the reservoir compartment equations (3.17) and (3.18) are abolished and for the last layer the outgoing paracellular flux (J_v) is replaced by the $\mathbf{K}_v \cdot C_s$ term, where \mathbf{K}_v is the permeability at the boundary between the epithelium and the stroma and C_s corresponds to the concentration of either AA, Ac or H.

To summarize, equations (3.13) to (3.18) reflect the pharmacodynamics of AA which we attribute to a single neoplastic layer. To form, therefore, the complete equation system this set of differential equations is repeatedly coupled following the number of neoplastic layers. Then, according to our discussion in paragraph 2.4.5 the dynamic characteristics of the measured evanescent optical effects are largely attributed to the temporal changes in the intracellular [H]. For this reason, the time-course of the excess [H] that accumulates inside the neoplastic cells of the entire tissue cross-section has been selected to comprise the model's output (y^M):

$$y^M(t) = \sum_{i=1}^N \left\{ [\text{H}]_{\text{IS}}^i(t) - [\text{H}]_{\text{IS}}^i(0) \right\}. \quad (3.19)$$

From the mathematical realization that was just discussed one can deduce that the developed model is a deterministic, non-linear, algorithm that includes a scalable (due to the \mathbf{N} parameter) system of coupled differential equations. As such, it is impossible to find an analytic solution to the differential system. Hence, a computer program has been written that allows for the calculation of the numerical solution to the differential equation system.

Particularly, a series of Matlab™ scripts have been developed. Then, the *ode15s()* function has been utilized for the numerical integrations that employs, for stiff problems, an iterative Runge-Kutta formula. As input to the model, a set of numerical values of the parameters included in Table 3-1 is employed. As initial conditions for the integrations we include: the $[\text{TA}]_{\text{ES}}$ in the reservoir layer (0.5M), similar to the quantity applied on the tissue during a clinical test; the $[\text{TA}]$ that preexists in the rest of the tissue, which we have assumed to be zero (0); and the proton concentration of both the IS and ES according to their respective pH which we provide as input parameter values.

Using the aforementioned numerical integrator a numerical solution is calculated for 240 seconds after the initiation of the phenomenon. The time frame that we allow the integrations to occur is the same as in clinical practice. Then, after the solution is acquired, the intracellular TA and H concentration vs. time functions in both the IS and the ES and for every

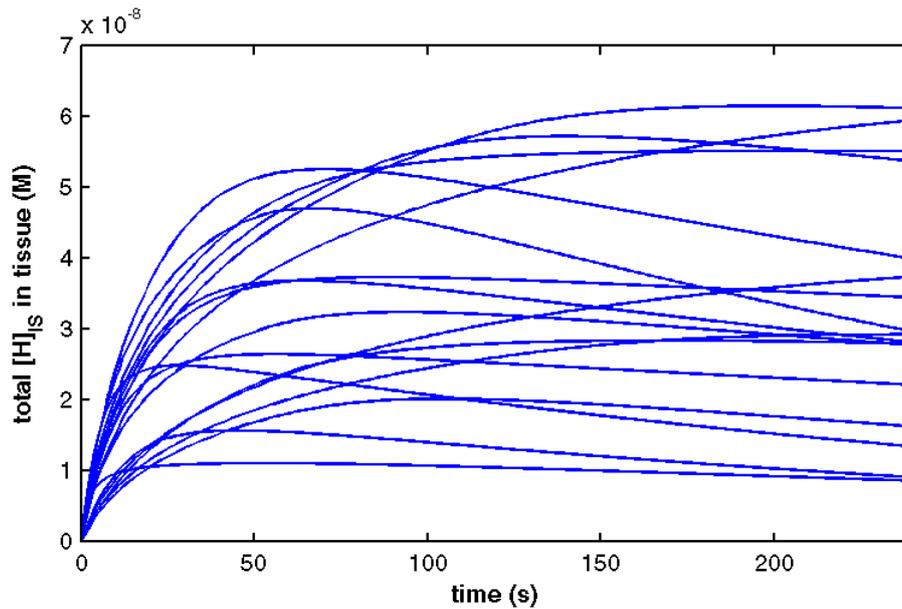


Figure 3-3 Typical model predictions. The simulations have been performed using randomly selected sets of parameter values.

layer can be plotted. As said, however, the H_{IS} concentration vs. time function is used as the model's output. Figure 3-3 illustrates typical model predictions according to equation (3.19) for several different numbers of neoplastic layers i.e. stages of neoplasia.

3.4 SYSTEM PARAMETERS

The, per layer, differential equation system presented for the kinetics of AA in the neoplastic tissue can be described by equations (3.13)-(3.18). These equations contain several biological model parameters that influence the course of the AW phenomenon. Model parameters can be roughly categorized into the structural and functional ones. Table 3-1 shows these model parameters. The structural parameters are the ones that define the virtual tissue's geometrical properties. The functional are the ones that define the solute behavior. Some of them, as we will discuss in the next chapter, are going to be used as input to the model when its identification is intended and some will be kept constant throughout this procedure.

3.4.1 Structural parameters

According to the structural alterations that neoplasia imposes the biological parameters in which we have been interested are: the number of neoplastic layers (N), the volumes of the IS and of the ES (V_{IS} , V_{ES}) and the diffusion areas of the membrane and of the tight junctions (A_m , A_{ij}).

3.4.1.1 The number of neoplastic layers

In clinical practice the number of neoplastic layers on a biopsy sample is the strongest indicator of neoplasia severity. Here, it is core parameter in the model and represents the size of the cell stack in our tissue topology. For the definition of the exact number of cells attributed to a CIN stage we have found that the average cervical epithelial depth is $360 \mu\text{m}$ [113]. Supposing that all sections (basal, parabasal, intermediate, and superficial) cover equal areas, it follows that the height of each one is $360/4=90 \mu\text{m}$. The total number of epithelial layers for each CIN grade can, then, be estimated by the following relationship:

Table 3-1 List of the biological parameters embedded on the developed model

	Parameter	Epexegesis
Structural	a^3	volume (in μm^3) of the IS – V_{IS}
	$a^2 \times b$	volume (in μm^3) of the ES – V_{ES}
	a^2	area (in μm^2) of the cell membrane through which diffusion occurs – A_m
	$a \times b$	area (in μm^2) of the tight junctions through which diffusion occurs – A_{TJ}
	N	number of epithelial neoplastic layers
Functional	β_{IS}	IS intrinsic buffering capacity
	β_{ES}	ES intrinsic buffering capacity
	pH_{IS}	IS pH
	pH_{ES}	ES pH
	$\text{pH}_{IS,Kp}$	pH level in which the efflux rate of the active pH regulation (pump activity) maximizes
	g_{TJ}	TJ porosity fluctuation due to the alteration of its geometrical properties
	K_V	solute diffusion rate at the tissue-basement membrane interface
	$P^{(S)}$	solute permeability through the TJ complexes and the cell's membrane
	ΔV_{TJ} , ΔV_m	voltage gradient at the TJ complexes and at the cell's membrane

$$N = \sum_{i=1}^4 \frac{90}{h_i}, \quad (3.20)$$

where N is the total number of layers and h_i is the cell height for each epithelial section. Thus, the maximum number of dysplastic layers at the squamous cervical epithelium, in different CIN grades, can be estimated using (3.20) and cell morphological values obtained from Table 3-2 as follows:

$$\begin{aligned} \text{CIN}_I &\propto 90/13 \cong 6, \\ \text{CIN}_{II} &\propto 90/13 + 90/15 \cong 12 \text{ and} \\ \text{CIN}_{III} &\propto 90/13 + 90/16 + 90/16 + 90/14 \cong 24 \end{aligned}$$

These values determine the number of the two-compartment differential systems that compose the epithelial model for each CIN grade. Our model simulations for these values indicate that, roughly, no more than 10-15 layers are needed to describe sufficiently the AW phenomenon.

3.4.1.2 Geometrical properties of the diffusion pathways

As said, several groups observed that during CIN the tissue tends to homogenize in structure developing evenly shaped cells. For this reason, the IS and ES compartments have been considered to be cubical and rectangular cuboid, respectively. In effect, we have limited, for simplicity, the transmembrane diffusion to occur in only one side of the cellular “cube” i.e. the one adjacent to the paracellular pathway. Now, let a be the length of the IS and b the corresponding of the ES. Then, compartments' volumes must be $V_{IS}=a^3$ and $V_{ES}=a^2 \cdot b$ while, $A_{TJ}=a \cdot b$ and $A_m=a^2$ are the tight junction and transmembrane areas, respectively. Using this formalism and some algebraic manipulations we deduce that the area-to-volume ratios included in our equation system reduce to: $\rho_{ij} = 1/a$ $\rho_{ES} = 1/b$ and $\rho_{IS} = 1/a$, again. According to literature, the nominal range for the values of the a and b parameters are 10-20 μm and

Table 3-2 Geometric properties of cervical cells adopted from [113]. The gray shades designate the atypical cells conforming to the histological grading discussed in paragraph 2.4.1.

Tissue section	Tissue type	Length (μm)	Height (μm)	NCR
<i>Superficial</i>	<i>Normal</i>	66 ± 9.2	14 ± 4.8	0.01
<i>Intermediate</i>	<i>Normal</i>	52 ± 10	14 ± 4.7	0.01
<i>Parabasal</i>	<i>Normal</i>	31 ± 6.0	17 ± 5.1	0.03
<i>Basal</i>	<i>Normal</i>	16 ± 3.2	15 ± 5.0	0.19
<i>Superficial</i>	<i>CIN I</i>	46 ± 10	14 ± 1.3	0.01
<i>Intermediate</i>	<i>CIN I</i>	41 ± 9.4	14 ± 2.4	0.01
<i>Parabasal</i>	<i>CIN I</i>	25 ± 5.0	16 ± 3.8	0.04
<i>Basal</i>	<i>CIN I</i>	12 ± 4.2	13 ± 2.2	0.21
<i>Superficial</i>	<i>CIN II</i>	38 ± 7.1	15 ± 2.3	0.02
<i>Intermediate</i>	<i>CIN II</i>	25 ± 7.1	17 ± 4.1	0.04
<i>Parabasal</i>	<i>CIN II</i>	18 ± 5.0	15 ± 2.5	0.17
<i>Basal</i>	<i>CIN II</i>	12 ± 1.6	13 ± 1.7	0.33
<i>Superficial</i>	<i>CIN III</i>	20 ± 5.5	13 ± 3.2	0.14
<i>Intermediate</i>	<i>CIN III</i>	18 ± 5.9	16 ± 1.8	0.15
<i>Parabasal</i>	<i>CIN III</i>	15 ± 3.6	16 ± 1.5	0.20
<i>Basal</i>	<i>CIN III</i>	13 ± 2.3	14 ± 2.4	0.28

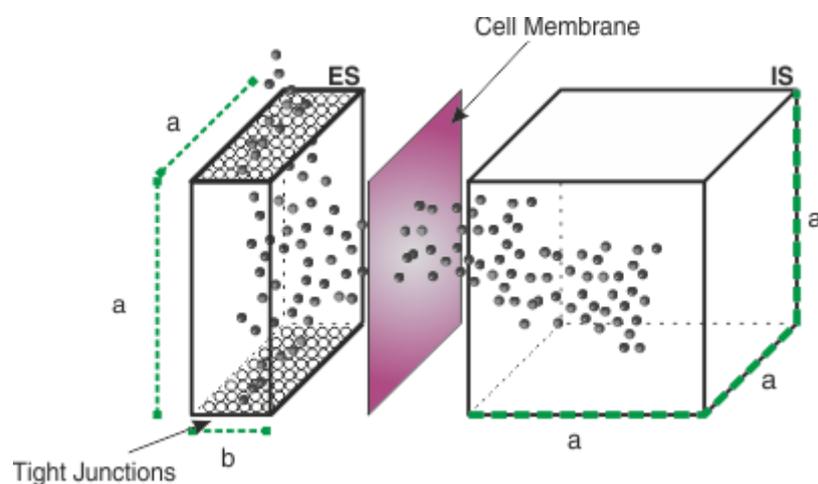


Figure 3-4 Structural composition of a two-compartment layer as proposed in our model. In this figure the height, length and depth of the compartments and diffusing membranes are included

0.1-0.8 μm , respectively [113], [145]. Figure 3-4 displays the proposed structural composition of a two-compartment layer.

3.4.2 Functional parameters

The functional parameters that are involved in the pharmacodynamics and pharmacokinetics of the AA in the neoplastic cervical tissue and that have included in our model are: the ones that are linked with the pH regulation mechanisms i.e. the pH value where the pumping rate maximizes ($\text{pH}_{i,Kp}$), the pH ranges and buffering capacities of the IS and ES (pH_{IS} , pH_{ES} and β_{ES} , β_{IS}); the solute permeabilities ($P^{(\text{AA})}$, $P^{(\text{Ac})}$ and $P^{(\text{H})}$) for the paracellular

(TJ) and transcellular pathways (m); the tight junction porosity (g_{TJ}), the transmembrane voltages (ΔV_m , ΔV_{TJ}); and the rate of vascular drain (K_V).

3.4.2.1 pH regulation parameters

We have already said in paragraph 2.3.1 that cancer cells manage to maintain their internal pH at normal levels at the expense of ES acidification. Accordingly, we have adopted the pH_{IS} to range between 7 and 7.4 and the respective pH_{ES} to range between 6 and 7 [88], [89], [93], [94], [101]. Equally, we adopted that the $pH_{IS,Kp}$ is, on average, 6.8. On the other hand, the buffering capacity has been shown to be fairly constant in the range of $pH_{IS} \sim 6.4$ -7.2 [88], [146]. Then, being consistent with measurements of the total buffering capacity of tumor cells (bladder, breast) we assumed that β_{IS} lays between 30 and 50mM (the negative sign has been dropped) [93]. At the same time the extracellular buffering power in tumors has been measured to be 28mM and is thought to be constant through all the tissue regardless of the distance from the capillaries [98].

3.4.2.2 Transcellular, paracellular and vascular permeabilities

Prior to determining the appropriate permeability values for the solutes at a neoplastic cervical epithelium we have made two reasonable assumptions. First, we assumed that the transcellular permeabilities remain unaffected by cancer since there is no sound alteration in the composition of the cell's membrane during carcinogenesis. On the other hand, second, we assumed that the paracellular permeabilities are influenced by cancer although, as indicated in paragraph 2.4.2.2, there is no solid literature evidence that justifies whether the solute permeabilities increase or decrease due to CIN progression. Only the fact that the over-expression of the TJ proteins is the reason to transform under normal circumstances an epithelium from tight to leaky. Then, by the time similar conditions appear in neoplastic regions it is reasonable to assume that the diffusivity of solutes does change.

Under these circumstances, we have no reason than to adopt that the transcellular permeabilities are constant regardless of the neoplasia stage. Accordingly, we have adopted that the cell membrane permeability is, approximately, between $38\mu m/s$ and $66\mu m/s$ for the AA and $67\mu m/s$ for the Ac [88], [147], [148]. In contrast, no reference that quantifies, *per se*, the paracellular permeabilities in the cervical epithelium during CIN progression has come to our attention. Therefore, the AA and Ac permeabilities in the TJs were approximated with the values of $0.97\mu m/s$ and $1.1\mu m/s$, respectively, with reference to Mannitol's permeability, which has been found to be $0.22\mu m/s$ in normal cervical epithelium cultures [149]. In addition, the respective diffusion of H has been considered to be, unobstructed by the channel pores, due to its unitary molecular weight. However, due to the ES buffering, channel resistance, and race conditions with the other molecules, it was given the value $25\mu m/s$ which is in accordance with the H mobility in buffered solutions. Then, the apparent fluctuation of the paracellular permeabilities was attributed to the altered geometrical properties of the TJ pores. According to literature the geometrical length of a pore can range between 0.1 and 10 times its initial value [150]. This means that the respective solute permeability should vary within the same range. On this basis, the parameter g_{TJ} has been introduced as a multiplicative factor to the aforementioned nominal TJ permeability values.

Lastly, as much as the vascular drain rate is concerned there is no indication in literature specifically for the AA and its ionization products. Therefore, we have treated all the solutes equally by applying a range of values between 10^{-6} and $10^{-7}m/s$ conforming to literature concerning other solutes [151], [152].

3.4.2.3 Electrical properties

Theoretically, the main tissue electrolytes are K^+ , Na^+ and Cl^- . These ions, under normal circumstances sustain a constant potential between the IS and ES in order for a cell to keep its homeostasis. As we have already indicated, the neoplastic cell adapts to its acidic environment by maintaining its pH regulatory mechanisms increased. This however, does not imply that it happens in expense of the rest of the ion regulatory mechanisms. Therefore, in our calculations, we kept the ES-IS transmembrane voltage constant and equal to the resting potential of normal cells i.e. $\Delta V_m = 67.9\text{mV}$ [142].

On the other hand, there is little to no indication in literature over the value of translayer voltage (ΔV_{TJ}) during neoplasia. Therefore, we have adopted that the following situation holds. Initially, we know that the tight junction pores of the paracellular route act as ion channels keeping the tissue polarized. The latter means that there is a constant electrical gradient between the upper and the lower areas in a tissue layer. However, this tissue organization is dismantled by cancer. Apart from the increased acidity, the suppression of tissue polarity originates, primarily, from the malfunctioning tight junctions. Thereupon, since during neoplasia the passing through the tight junctions becomes easier and faulty, in terms of tight junction pore permittivity, we can reasonably assume that all extracellular neoplastic compartments contain the same concentrations of ionic species, at electrodiffusive equilibrium. This means that, no trans-layer potential difference should exist prior to the application of AA. Next, we know from literature that after the application of AA the perturbation in the tissue impedance provoked by its ionization, is negligible [153]. With this in mind, we can adopt that the diffusion of the bi-ionic system $H - Ac$ creates no net current within the tissue and, thus, the zero trans-layer potential is maintained and after the AA application. If the current that is induced by the diffusion of the bi-ionic system $H - Ac$ is zero then, the individual fluxes of H and Ac ions must be equal (neglecting that other mobile ions may resident the extracellular space). In this case, equation (3.4) can be manipulated to predict the trans-layer potential [154]:

$$\Delta V_{TJ} = \frac{RT}{F} \ln \frac{P_{TJ}^{(H)} [H]_{ES}^{i-1} + P_{TJ}^{(Ac)} [Ac]_{ES}^i}{P_{TJ}^{(H)} [H]_{ES}^i + P_{TJ}^{(Ac)} [Ac]_{ES}^{i-1}}, \quad (3.21)$$

We should note here that this potential is dynamic as it follows the kinetics of AA throughout the epithelium.

3.5 QUALITATIVE VALIDATION

The mathematical representation of the biomarker-tissue interaction presented above is an enhanced and more detailed version of the initial model that we have developed and presented in [17]. For more information, for that model the interested reader is referred to [155]. This early version of the model has been qualitatively validated based on selective results obtained from clinical feasibility tests performed before the phase II clinical trial, as a preparatory phase [156]. Particularly, the maximum value of the dynamic DR signals measured *in vivo*, the slope before and the slope after this value have been correlated with the NCR measured in histological samples obtained from the same site of the tissue. In general, the NCR is used in clinical practice as a basic criterion for identifying neoplastic cells and for grading CIN. Therefore, using histology as reference a first rough validation was realized in which the model predicted increase of the NCR (interpreted as an increment in cell packing) results in an increase of the peak value of the DR vs. time curves of the AW effect.

Following the same validation procedure we have compared the updated model predictions with the measured dynamic optical signals and the biopsy samples collected from the respective abnormal areas [18], [157]. This time, however, a quantitative evaluation has also been incorporated beside the visual cross-comparison of the features of the dynamic curves as predicted by the model and as measured in the clinic. Particularly, the correlation of the max DR vs. NCR had been quantitatively assessed, from the histological slides. The reason for this extra evaluation is that we want to quantitatively indicate that the profiles of the DR signals are correlated with the underlying biological microstructural alterations of the epithelium, based on which biopsy assessments are made. The procedure for this was as follows. First, 103 biopsy samples were obtained from a subgroup of 74 women that participated in the aforementioned clinical trial. Then, the max DR values were recorded in selected tissue sites using the DCE-OI device and tissue samples were obtained from the same sites, using the pseudocolor map of the effects temporal features as a guide. Finally, from the histological slides the NCR was calculated as the sum of the area occupied by the neoplastic epithelial nuclei divided by the overall area of the epithelial tissue, using image analysis tools [113]. Using these, the max DR vs. NCR graph was plotted and the correlation coefficient was calculated.

Figure 3-5a illustrates the intracellular hydrogen ion concentration kinetics as predicted by the model for three N values ($5\alpha\text{NCR}=0.1$, $10\alpha\text{NCR}=0.2$, $15\alpha\text{NCR}=0.1$). These values correspond roughly to CIN I, II, III, conditions. It is clearly seen that the max DR values increase by increasing the N values. Figure 3-5b illustrates typical experimental DR vs. time curves, corresponding to CIN I, II, III conditions, histologically confirmed. The similarities of the curve profiles with the ones displayed in Figure 3-5a indicate that the AW dynamic optical characteristics are determined by the intracellular hydrogen ion concentration kinetics in the total epithelium. It is also seen that the higher the CIN grade the greater the max DR value. This is because the neoplastic growth is associated with an increased number of nuclei that become optically active after the application of AA, which increases the intensity of the backscattered light. Figure 3-5c illustrates the max DR vs. NCR scatter plot. The line in the graph represents the linear regression curve. The correlation coefficient was found to be 0.74, indicating that optical data measured *in vivo* can provide information for microstructural features related with the neoplasia grade.

These findings can be considered as a qualitative validation of our assumptions thus far. Furthermore, they show that a clear correlation between the dynamic scattering characteristics of the AW effect and the structural and functional characteristics of the epithelium exists. It is strongly suggested, therefore, that the model-based fitting of the experimental data can, in principle, enable the calculation of dynamic optical parameters that correlate well with histological features used to grade biopsy samples.

Apart from the microstructural information, this fitting approach can also provide information about the functional alterations accompanying the neoplastic growth. It is obvious that in contrast to what happens with the morphological features, the functional ones can only be assessed in live tissues. This could comprise additional and complementary diagnostic information that could further improve the accuracy of the *in vivo* diagnosis. In addition, this renders possible the minimization or even elimination of the need of biopsies since similar information can be obtained *in vivo* and without the risk for biopsy sampling errors, which is common in the standard diagnostic chain.

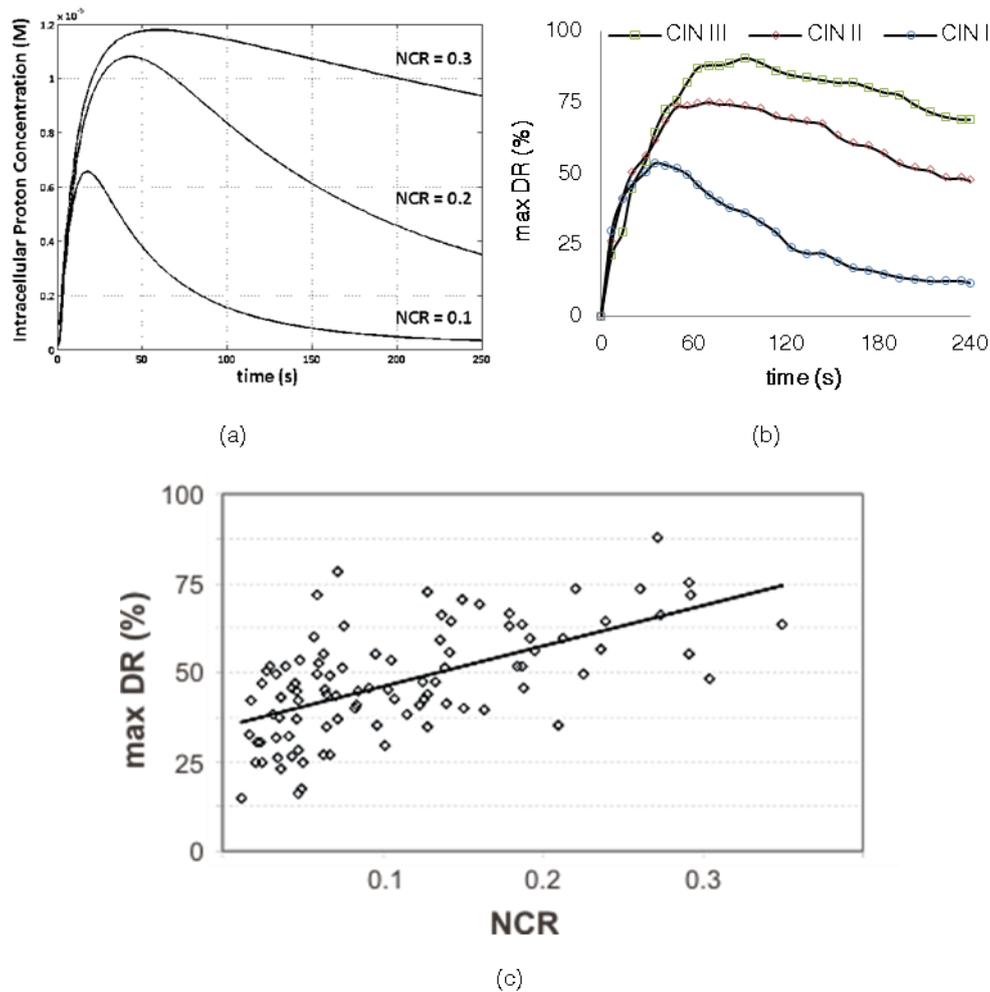


Figure 3-5 (a) The intracellular hydrogen ion concentration kinetics as predicted by the model for three NCRs values; (b) typical experimental DR vs. time curves, corresponding to CIN I, II, III conditions, histologically confirmed; (c) the max DR vs. NCR scatter plot

3.6 SUMMARY

During neoplasia development a series of alterations in the epithelial tissue's structure and functionality occur. Particularly, the number of abnormal cell layers increases. In addition, the size of the cells is reduced while their nucleus size increases resulting in an increase of the intercellular space and of the nuclear-to-cytoplasmic ratio, respectively. On the other hand, functionally the extracellular acidity is increased, the tissue is dismantled and the metabolic cooperation between the cells is canceled. Under such circumstances the externally administrated AA molecules penetrate with high selectivity the neoplastic cells. Upon entering the cell AA is readily dissociated and generated hydrogen ions provoke a reversible denaturation and crosslinking of nuclear proteins, changing the index of refraction. This alters the scattering properties of the epithelium, converting it from transparent to opaque, which becomes evident as AW effect. Next, the ions are extruded to the extracellular space with the aid of the cell pumps. These processes are repeated in the underlying neoplastic epithelial layers, "lighting-up" their cells.

On the basis of this conceptual framework, a novel compartmental/pharmacokinetic model of the epithelial transport phenomena has been developed in this chapter. Using the total $[H]_{IS}$ as output, our simulations have correlated, for the first time, the dynamic optical

parameters expressing the temporal characteristics of the AW effect with the histological parameters of biopsy samples obtained from the same tissue sites. Particularly, it has been found that the max DR values correlate with the NCR, thus making its approximation possible, *in vivo*, and for every pixel area. The method has been tested in a large clinical trial, where a remarkable improvement in diagnostic sensitivity has been demonstrated. Therefore it comprises a promising alternative optical biopsy method for the non invasive diagnosis and screening of cervical neoplasia.

This highlights the potential of the developed method and imaging technology in replacing traditional endoscopes and in improving the endoscopic performance in terms of both diagnostic and biopsy sampling accuracy. In addition, by providing user-independent, quantitative information for neoplasia grade *in vivo*, it holds the promise to become a more reliable and cost-effective alternative screening tool for the early detection and prevention of cervical neoplasia.

RELEVANT PUBLICATIONS

- [1] C. Balas, G. Papoutsoglou, and A. Potirakis, "In vivo molecular imaging of cervical neoplasia using acetic acid as biomarker," *IEEE J. Sel. Top. Quantum Electron.*, vol. 14, no. 1, pp. 29–42, 2008.
- [2] C. Balas, G. Papoutsoglou, C. Loukas, Y. Skiadas, C. Pappas, D. Haidopoulos, E. Diakomanolis, and W. P. Soutter, "In vivo assessment of microstructural and functional alterations in cervical neoplasia," in *Conference proceedings: European Conferences on Biomedical Optics (ECBO)*, 2009.

4 SENSITIVITY ANALYSIS OF THE DEVELOPED MODEL

4.1 INTRODUCTION

Having created a suitable simulation procedure based on a strong biological background it is tempting to address the inverse problem (*in vivo* parameter estimation), directly. However, as indicated in the first chapter, there is always a high possibility that the model's structure and/or its parameters are uncertain. In addition, model parameters may be mathematically entangled and/or correlated with each other. Therefore, the question whether the unknown system parameters can be uniquely estimated from the available experimental data and using the proposed model's structure should be theoretically addressed, before any estimation attempt is made. This problem is a parameter identification problem that encloses two important steps, namely parameter identifiability and parameter estimability.

To theoretically address these, in the current chapter, we utilize the outcomes of global sensitivity analysis. Then, based on the acquired results we systematically reduce the parameter complexity of our proposed model. This is expected to improve the posedness of the inverse problem. From this procedure, we can reasonably assume that uniqueness in the estimations can be adequately achieved and be practically realized, as we will describe in the following chapter.

4.2 PARTICULARITIES OF THE DEVELOPED MODEL

Based on system's identification theory, the solution of the inverse problem may be substantially unique if the model parameters are globally identifiable under an idealized environment of noise- and error- free model structure and data. In principle, determining global identifiability requires the extraction of an analytical relation between the parameters and the measured output(s) through algebraic manipulation and transformation of the model equations. As indicated in paragraph 1.7.1, however, this theoretical model assessment can only be obtained for some type of models, usually linear ones. In case the model is too complex to manipulate mathematically, sensitivity analysis procedures can be followed instead (paragraph 1.7.1.3). Therefore, selecting the most suitable course of action largely relies on the mathematical manageability of the designed model.

To investigate the latter, the model has to be, first, represented in an appropriate form. Here, according to state space formalism indicated in equations (1.18) - (1.21) the proposed model can be written as:

$$\dot{\mathbf{x}}_{m \times 1}(t, \mathbf{p}) = \mathbf{f}_{m \times 1}[\mathbf{x}_{m \times 1}(t, \mathbf{p}), u(t) = 0.5, t; \mathbf{p}], \quad \mathbf{x}_{m \times 1}(0) = \mathbf{x}_{m \times 1}(t=0, [pH_{IS}, pH_{ES}]), \quad (4.1)$$

$$y^M(t, \mathbf{p}) = g[\mathbf{x}_{N \times 1}^{[H]}(t, \mathbf{p}); \mathbf{p}], \quad (4.2)$$

$$y^M \geq 0, \quad t \in [0, 240], \quad \mathbf{p} \in [\mathbf{p}_L, \mathbf{p}_U], \quad (4.3)$$

where the x_i ($i=1 \dots m$) states are the [TA] and [H] as described by equations (3.13), (3.14), (3.17) and (3.18), $m=2+4N$ is the row size of the state space variable \mathbf{x} and \mathbf{p}_L , \mathbf{p}_U are the lower and upper bounds of the parameters values.

From equations (4.1)-(4.3) it is obvious that the state space model is allowed to vary in size in accordance with the value of parameter N ; the number of neoplastic layers. This means that the representation of the model's equation system has no closed-form due to its adaptable/flexible structure in terms of abnormal cell layer integration. Recalling our earlier analysis, the number of layers has been defined so as to specify the height of the cell stack, proportionally to the severity of neoplasia. Mathematically, this means that the number of layers is not just an independent integer parameter; in fact, it defines the number of differential equations that are integrated in the system. More specifically, each time the number of layers are set to increase or decrease by one, a four differential equations system (equations (3.13), (3.14), (3.17) and (3.18)) is automatically added to or deducted from the final differential equation system of the model, respectively. Simultaneously, the model's differential equation system is too complex to have a general analytical solution. This can be easily evidenced by a close inspection of the terms that are integrated in the differential equations. Particularly, strong nonlinearity is introduced by: the existence in the equation system of the ω and φ terms which dynamically designate the degree of ionization of AA (equations (3.15) and (3.16)); the trans-layer voltage (equation (3.21)) which also influences dynamically the paracellular flux of ionic species; the existence of the [H] conversion factors which designate the IS and ES buffering procedures (equations (3.14) and (3.18)); and the complex piecewise linear function of pH_{IS} and pH_{ES} that designates dynamically the proton pumping rate (equations (3.7) and (3.8)).

Therefore, the developed model is a multi-resolution design that cannot be represented analytically while its complexity does not allow for the model equations to be factorable or anyhow algebraically manipulative.

4.3 SENSITIVITY ANALYSIS AS A MODEL ASSESSMENT TOOL

The recognition of such particularities leads to the conclusion that handling the global identifiability question with classical treatments is impossible here. Effectively, only sensitivity analysis procedures may assert, with some level of assurance, that the solution of our inverse problem can provide, theoretically, unique parameter estimations. In general sensitivity analysis techniques can provide valuable insight into model characteristics. Hence, it has been a common practice to employ sensitivity analysis as a tool for model assessment. In biomedical engineering, in particular, sensitivity analysis has been successfully employed at evaluating the effects of parameter uncertainty on model outcomes; for analyzing model ro-

bustness; for reducing the model's complexity; for adjusting the model's structure; for parameter identification; and for the design of experiments for model testing [27], [158].

Sensitivity analysis procedures commonly include the calculation of dependency measures of the model outputs to the parameters that modulate them. Theoretically, the derivative at a “base point” of an output versus an input parameter ($\partial p_i / \partial y$) is considered as the mathematical definition of a sensitivity measure. To assess, therefore, the identifiability of the model as a whole, one primarily computes the partial derivatives in terms of all the postulated parameters i.e. the Jacobian matrix of the model's output. Next, to capture the maximum of the system dynamics the Jacobian matrix has to be calculated in the maximum possible time instances of the real experiment. As seen in paragraph 1.7.1.3, the matrix that is deduced from this procedure comprises the system's output sensitivity matrix (equation (1.34)). Each column of the sensitivity matrix denotes the sensitivity function of each of the input parameters. Then, the calculation of the amount of correlation between any two columns of the sensitivity matrix reveals the amount of the respective parameter identifiability.

In general, the output sensitivity functions denote the dynamic influence of a presupposed model parameter to the measured output [41]. Still, even if this is sufficient for marking a model parameter theoretically identifiable, there is no guarantee that it is also sufficient for the parameter to be estimable [19]. Poor estimability comes from high level of interactions/correlations between the pertinent identifiable parameters. Such feature cannot be readily imprinted on the sensitivity function. For this reason, several criteria are utilized which, based on the Fisher information matrix, determine the parameters' estimability [39]. In addition, another methodology is to perform principal component analysis on the sensitivity matrix to get a clear idea of how many parameters should be estimated and which those are [159]–[161]. As indicated in [162], a principal component analysis is a useful post processing tool that can extract at no additional cost relevant features of the dynamic mechanism and can assist in the setup of the parameter estimation step.

Regardless of how one can draw firm conclusions regarding the identifiability and estimability, there is a caveat concerning the model parameters; that is as their number increases the parameter estimation problem becomes intractable. Consequently, it is of outmost importance to identify which parameters are irrelevant and may be fixed to a nominal value, without significant loss of information on model behavior. Within this context, we have seen earlier that sensitivity analysis is a method able to quantify the dependence of the system's behavior on the parameters which affect the process dynamics. For example, if an infinitesimal change in a parameter results in relatively large changes in the model predictions then, it is said that the model predictions are sensitive to this parameter. This means that this parameter is important and has to be determined very precisely to achieve an accurate model. On the contrary, parameters to which the process dynamics exhibit a small sensitivity they do not need to be measured rigorously. Therefore, to fight this curse of dimensionality of the space of model parameters the sensitivity measures are useful as model reduction/simplification tools.

4.3.1 Local vs global sensitivity analysis

Earlier, we have shown that sensitivity analysis methods are the only option in order to infer whether the parameter estimates can be, in theory, substantially unique. Along these lines, we have presented a roadmap for the assessment of identifiability and estimability measures which is commonly employed. The procedures described therein rely on the derivatives at “base” points which, in turn, can be thought as a sensitivity measure. This kind of

sensitivity analysis, however, is only of local scale. In fact, it pertains only to a particular point in the parameter space. Consequently, although the general sense is that the partial derivatives reveal substantial information over the system's behavior, they can do so only for that specific point for which they are computed. This also means that local sensitivities cannot provide for a rigorous exploration of the entire k -dimensional input space.

For linear systems these would matter very little since a measure at a point away from the base can be found by linear extrapolation. On the other hand, this is not the regular case for nonlinear systems especially when some parameters may vary extensively and in an interactive manner and/or when an analytical model representation does not exist e.g. a computer program. In other words, classical (local) methods provide local inferences that can be generalized to capture the full spectrum of the model space only under specific conditions of linearity and analyticity of the model. Therefore, application of local methods for the study of the identifiability and estimability in our model is most probable to be misleading.

Since the efficacy of local identifiability methods clearly depends on the model's complexity, it is more appropriate to explore the sensitivity of the model output to its parameters and the possibilities of non-linear effects, in a probabilistic context. On such grounds, global sensitivity analysis (GSA) has been presented as a robust alternative, detecting non influential parameters and their interactions from simultaneous parameter variations of arbitrary magnitudes. In general, although local sensitivity analysis is highly appreciated in identifiability practices this is rarely the case for its global counterpart; even if strong correlations between identifiability and GSA methods have been shown [43], [163]. The main advantage of GSA methods against the local ones, that are employed in identifiability problems, is that they treat each model as a 'black box' and the quality of a sensitivity inference does not rely on any prior knowledge over the designed system such as function analyticity, linearity and/or additivity. An additional advantage of global sensitivity analysis is that the sensitivity estimates of individual parameters are evaluated while varying all other parameters as well. In this way, the relative variability of each parameter is taken into account, thus revealing any existing interactions.

Summing up, we have discussed in paragraph 4.2 that the developed model as is it structured it cannot be solved analytically. Hence, a sensitivity analysis method has to be employed in order to elucidate whether it is, in principle, possible to estimate separately the full set or a subset of model parameters by fitting the experimental data with the model's output. To investigate this local and global analyses exist in literature. However, our predictions contain convoluted information for all the neoplasia-related structural and functional features that determine the AA epithelial transport effects. Over and above, the model equation system has no closed form. It is straightforward, therefore, to employ GSA for exploiting our theoretical capability to predict the biological parameters from the dynamic optical characteristics, measured *in vivo*. Next, the theoretical preliminaries of GSA are described.

4.4 THEORETICAL PRINCIPLES OF GLOBAL SENSITIVITY ANALYSIS

GSA can be roughly regarded as a tool that allows for the assessment and interpretation of the way with which the uncertainty of the model inputs is transferred to its designed output. For this, it has been traditionally employed as a tool for rendering the uncertainty and robustness of dynamic system models. In addition, global sensitivity measures usually serve as model reduction principles before parameter estimation is performed. To do this, GSA methods possess two unique properties [164]. First, is the property of multidimensional averaging, according to which the sensitivity of each factor is evaluated by varying all other

factors as well. Second, is the property of inclusion of influence of scale and shape, which refers to the fact that each factor's sensitivity includes the effect of the factor's probability density function shape and size. The characteristic steps that entitle GSA methods to systematically address, through these features, the influence of each of the postulated parameters to the modeled output is shortly described next.

4.4.1 The probabilistic framework of GSA

Typically, the global sensitivity methods are based on the sampling of a sufficiently large set of input factor values in order to evaluate the uncertainty in the outcome. As such, they run on a framework in which the input factors are exploited in a probabilistic context. Thereupon, suppose $Y=f(\mathbf{X},\mathbf{u})$ describes a deterministic model, where Y is the model output, $\mathbf{X}=(X_1,X_2,\dots,X_k)$ is a vector of uncertain model inputs independent to each other (k is the dimensionality of the space) and \mathbf{u} is a vector of known (fixed) model inputs. In addition, let each of these model inputs to be a random variable that follows some distribution (Figure 4-1). Then, the general goal of GSA is to apportion the model's output uncertainty to the input parameter variations that are causing this uncertainty.

The first step towards this direction is to determine the distribution of the input factors in \mathbf{X} . In general, there is no standard procedure to determine the distribution and which will be best way depends largely on the input factor itself. Basically, the distribution may be assessed, without excluding the one over the other, by means of scientific literature investigation; empirical knowledge; controlled experiments; or expert judgment. If no specific distribution can be extracted a triangular distributions is commonly employed using the lowest, the highest and the most probable value for a factor.



Figure 4-1 Potential input distributions. a) normal, b) uniform, c)triangular (symmetric) and d) discrete (bimodal)

The second step entails the definition of the sampling procedure according to which data for the simulations are chosen. This is of great importance in GSA and essentially determines the successfulness of the procedure. Especially in biological modeling, where the entire uncertainty range of parameters is usually of particular importance and model parameters can vary within large intervals (two or three decades) depending on their physiological meaning. Along these lines it is often the case that when a population of test values is randomly chosen, it exhibits clusters and gaps i.e. regions with several values close together or regions of no values at all. As a result, not all of the feasible parameter space is represented in the simulations. Effectively they are, in a way, biased since some regions are over-sampled and others are not sampled at all. In fact, random values tend to be poorly conditioned and mean values estimated with random samples have an uncertainty that diminishes slowly as $1/\sqrt{n_s}$, where n_s is the number of randomly chosen samples. On top, such an issue is even more evident when several factors appear in the modeling procedure forming a multidimensional space. To surpass these limitations one can assign factorial designs, latin hypercubes and multivariate stratified sampling methods [165]. These, however, are inadequate when designs have many parameters and require a large number of successive simulations. For this reason, a mathematical measure called *discrepancy* which characterizes the

lumpiness of a sequence of points has been introduced. Correspondingly, stratified low discrepancy point sequences that guarantee that the values are evenly distributed and cover the most of the feasible interval for each factor at all times are now commonly utilized. Such a sampling procedure is most suitable for statistical analyses as GSA and is called *quasi-random sampling* [166]–[168].

4.4.2 GSA methods

Following the previous discussion, the last step is to choose the most appropriate between the available methods in order to compute the desirable model inferences. These can be obtained through either scatterplots or correlation or (standardized) regression coefficient measurements or variance-based methods which are the most common tools of GSA [165]. Below the basic concepts of these methods are given with emphasis on the variance-based ones.

4.4.2.1 Scatterplots

Roughly speaking scatterplots can be considered as convenient representations that summarize the relationship between two independent variables X_i and X_j . Under GSA, they represent a simple and informative way to visually depict input/output relationships in order to infer the relative importance of variables.

To create scatterplots Monte Carlo simulations are employed. The procedure is as follows. First, the distribution of each input factor (X_k) is defined. Then, from each distribution, equally large number of values (n_s) are randomly sampled creating a vector for each factor. Next, the vectors are fused together into matrix form (\mathbf{M}) of n_s rows and k columns. Using each row as an input vector for the Monte Carlo simulations of the model, the output is computed. Finally, each column (k) of \mathbf{M} which represents the sampled values of a factor is plotted against the calculated model's output to create the scatterplots. As shown in Figure 4-2, the shape and pattern of the scatterplot may reveal the relative importance of the factors.

4.4.2.2 Standardized regression coefficients

A major drawback in scatterplot statistical inferences is that the outcome is neither quantitative nor in condensed format. For this reason, either correlation measurements (Pearson's or Spearman's) or standardized regression coefficients can be used, alternatively. In the latter case, the created scatterplots are fitted with a linear or nonlinear regressor

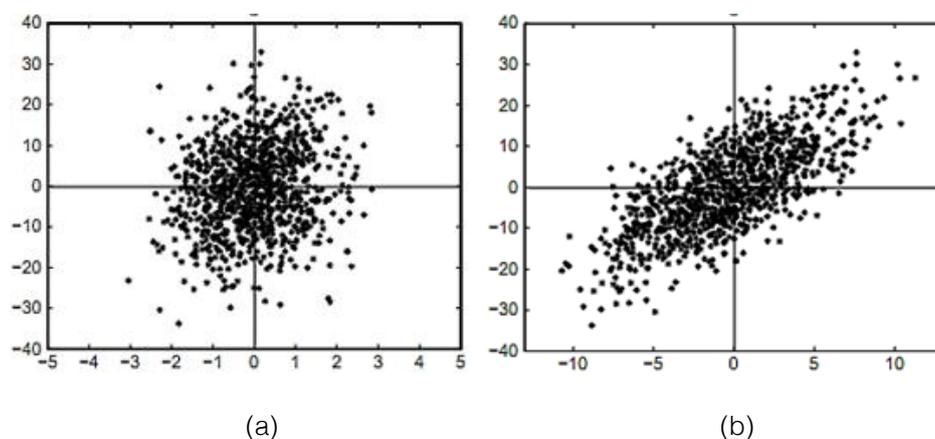


Figure 4-2 a) Input/Output scatterplot that reveals no specific pattern in the relationship. b) A clear linear pattern in the input/output relationship [165].

formula and the regression coefficients are used to extract statistical rankings of factors (Figure 4-3). For example, consider the linear regression for a mathematical model with a k -dimensional input: $Y(i) = b_0 + \sum_{j=1}^k b_{X_j} X_j^{(i)}$, where b_0 and b_{X_j} are determined using least squares fitting. Then, the standardized regression coefficients $\beta_{X_i} = b_{X_i} \sigma_{X_i} / \sigma_Y$, where σ is the standard deviation, may be used as a measure of the model's output sensitivity to the input factor X_j .

Even if scatterplots tend to have lower uncertainties and can reveal outliers, the use of regression for GSA may lead to inconsistent results. In principle, regression statistics using either a linear or a nonlinear formula assume a linear input/output dependence and, thus, the robustness of the results depends clearly on the validity of such prediction. For linear dependences this is not a major issue. For nonlinear, however, the linear approximation introduces bias and one must be very careful during the process. A way to resolve such an issue is either by linear transformation of the problem or by piecewise linear fitting (segmented regression). For convenience, we can also test the linearity of our model by summing up the calculated standardized regression coefficients which should be equal to the output variance if the model is additive:

$$\sum_i (\beta_{X_i})^2 = 1 = \sum_i (b_{X_i} \sigma_{X_i} / \sigma_Y)^2 \Leftrightarrow \sum_i (b_{X_i} \sigma_{X_i})^2 = \sigma_Y^2 = V(Y), \quad (4.4)$$

where $V(Y)$ denotes the variance of the measured output. In most cases however, this sum is less than one and equals to the fraction of linearity of the model. From another point of view, the sum of the standardized regression coefficients equals to the fraction of variance explained by the chosen regression model. For this reason, it is also known as the coefficient of determination (R^2) of the regression model. Being less or equal to unity it expresses the

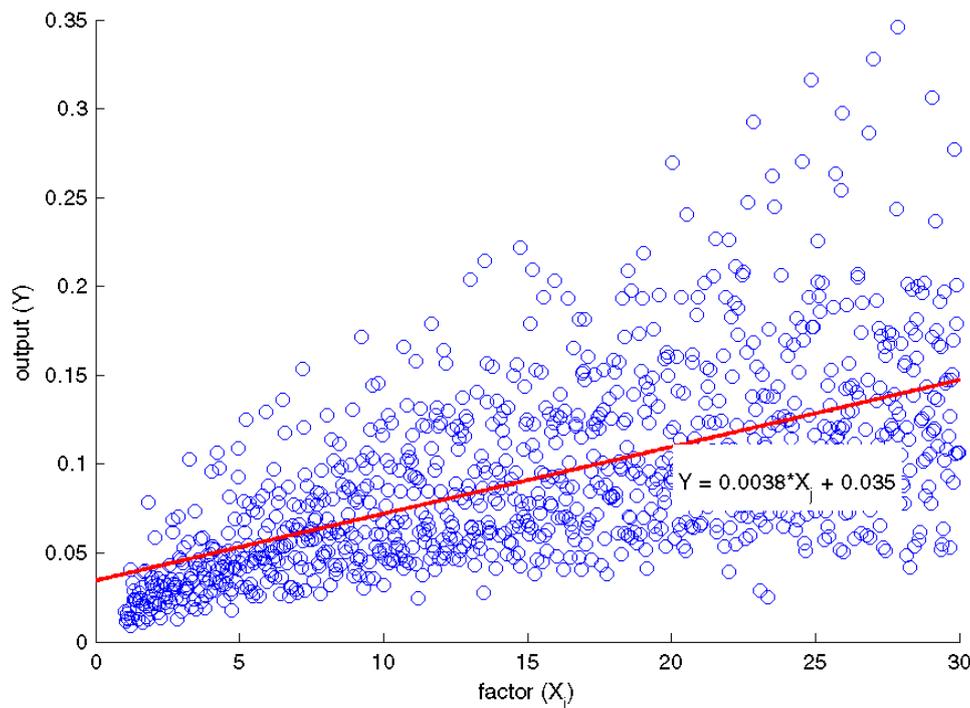


Figure 4-3 Example of a scatterplot of a model factor (X_i) and the relative output (Y). A linear regressor has been fitted to the data and the following regression coefficients were found: $b_{X_i} = 0.0038$ and $b_0 = 0.035$.

manner with which the output variance is decomposed. For instance, if the calculated result according to the linear regression described here is close to unity ($R^2 > 0.9$), then our model is fairly linear. More specifically, more than 90% of the output variance is attributed to the linear factor relationships that came up from the regression formula while the rest to nonlinear ones e.g. higher-order terms, interactions. Therefore, correlation measurements and regression coefficients may produce a sensitivity measure, in an aggregate manner. However, should nonlinearities or interactions between factors exist (low R^2), the measured statistics are not always reliable.

4.4.2.3 Variance-based methods

A different approach to the measurement of the total output variance $V(\mathbf{Y})$, other than the regression coefficients, has been demonstrated by decomposing the output into summands of increasing dimensionality. First, let the $\mathbf{Y} = f(\mathbf{X})$ be defined on the k -dimensional unit hypercube $\mathbf{I}^k = \{ \mathbf{X} : 0 \leq X_i \leq 1, i=1 \dots k \}$ and be square integrable. Without loss of generality, let also each input X_i have nonnull range of variation or uncertainty and be uniformly distributed over a unit interval $[0, 1]$ which also means that $\text{prob}(X_i) = 1$, where prob denotes probability. Then, Sobol' has shown that $f(\mathbf{X})$ can be decomposed into square integrable terms $f_{ij \dots k} = f(X_i, X_j, \dots, X_k)$ over \mathbf{I}^k using a High Dimensional Model Representation (HDMR) [168]–[170]:

$$f(\mathbf{X}) = f_0 + \sum_{i=1}^k f_i(X_i) + \sum_i \sum_{j>i} f_{ij}(X_i, X_j) + \dots + f_{1,2 \dots k}(X_1, X_2, \dots, X_k) \quad (4.5)$$

This decomposition is also known as the Hoeffding decomposition [171]. In order such representation to be unique f_0 must be a constant and each summand in the HDMR be chosen to have a zero mean, such that:

$$\int_{\mathbf{I}} f_{1,2 \dots s}(X_1, X_2, \dots, X_s) dX_\gamma = 0, \text{ for } s \leq k \text{ and } \gamma = 1, \dots, s \quad (4.6)$$

Then, it follows that f_0 represents the mean value of the output \mathbf{Y} (also remember that $\text{prob}(X_i) = 1$):

$$f_0 = \int_{\mathbf{I}} f(\mathbf{X}) d\mathbf{X} = E(\mathbf{Y}), \quad (4.7)$$

where $E(\cdot)$ is the expectation operator, and any pair terms are orthogonal:

$$\int_{\mathbf{I}} f_{i_1 \dots i_{s_1}} f_{i_1 \dots i_{s_2}} d\mathbf{X} = 0, \text{ for } (i_1, \dots, i_{s_1}) \neq (i_1, \dots, i_{s_2}). \quad (4.8)$$

By integration of (4.5) over all variables except X_i we find that each one-dimensional term f_i of the expansion is:

$$f_i(X_i) = \int_{\mathbf{I}} f(\mathbf{X}) \prod_{\gamma \neq i} dX_\gamma - f_0 = E(\mathbf{Y} | X_i) - E(\mathbf{Y}) \quad (4.9)$$

Equally, by integration of (4.5) over all variables but X_i and X_j we find that each two-dimensional term f_{ij} is:

$$\begin{aligned}
f_{ij}(\mathbf{X}_i, \mathbf{X}_j) &= \int_{\mathbf{I}} f(\mathbf{X}) \prod_{\gamma \neq i, j} d\mathbf{X}_\gamma - f_i(\mathbf{X}_i) - f_j(\mathbf{X}_j) - f_0 = \\
&= E(\mathbf{Y} | \mathbf{X}_i, \mathbf{X}_j) - E(\mathbf{Y} | \mathbf{X}_i) - E(\mathbf{Y} | \mathbf{X}_j) - E(\mathbf{Y})
\end{aligned} \tag{4.10}$$

By continuing this procedure all the higher-order terms of the expansion can be depicted as conditional expectations of their indices. Finally, by using the properties (4.7) and (4.8), if we square equation (4.5) and integrate over \mathbf{I}^k we get:

$$\int_{\mathbf{I}} f^2(\mathbf{X}) d\mathbf{X} - f_0^2 = \sum_{s=1}^k \sum_{i_1 < \dots < i_s} \int_{\mathbf{I}} f_{i_1 \dots i_s}^2 \prod_{\gamma} d\mathbf{X}_\gamma \tag{4.11}$$

Clearly, on the left of equation (4.11) we have the definition of the output variance $V(\mathbf{Y})$ and on the right its decomposition into conditional variances. In succession, it follows that :

$$V(\mathbf{Y}) = \sum_{i=1}^k V_i + \sum_i \sum_{j>i} V_{ij} + \sum_i \sum_{j>i} \sum_{k>j} V_{ijk} + \dots + V_{1,2,\dots,k}, \tag{4.12}$$

where $V_{ij\dots k}$ are the conditional variances of the expected output given k factors, either independently (first-order terms) or by interaction (higher-order terms). In result, equation (4.12) a 'model-free' decomposition of the output variance and equation (4.5) becomes an ANOVA (Analysis Of VAriance) representation of $f(\mathbf{X})$.

Now that we have achieved to decompose $V(\mathbf{Y})$, let's examine the conditional variances in the ANOVA-HDMR representation just described. In general, a conditional variance $V_{\mathbf{X}_{-i}} = V(\mathbf{Y} | \mathbf{X}_i = \mathbf{X}_i^*)$ corresponds to the output variance when \mathbf{X}_i is fixed to a certain value \mathbf{X}_i^* . Initially, it reasons that this statistic is smaller than the unconditional variance i.e. $V_{\mathbf{X}_{-i}} < V(\mathbf{Y})$. In addition, if the value of $V_{\mathbf{X}_{-i}}$ is large and close to $V(\mathbf{Y})$ means that the influence of \mathbf{X}_i to the output variance is small. This is true considering that $V_{\mathbf{X}_{-i}}$ represents the output variance due to all factors other than \mathbf{X}_i . For this, \mathbf{X}_i will be fairly unimportant. The opposite would hold if $V_{\mathbf{X}_{-i}}$ was small. In result, we would infer that conditional variances can be conceived as relative importance measures of the input factors. However, for this approach there are two important caveats. First, since the variance is measured for a fixed value of \mathbf{X}_i , it is biased by the position \mathbf{x}^* and, second, a conditional variance greater than the unconditional one is possible to appear e.g. when interactions exist. To diminish such issues, the expected value (average) of the conditional variances mentioned above is used i.e. $E_{\mathbf{X}_i}(V_{\mathbf{X}_{-i}}(\mathbf{Y} | \mathbf{X}_i = \mathbf{x}^*))$. This statistic is calculated over many fixed \mathbf{x}^* values of input factor \mathbf{X}_i and is always less than the unconditional variance. It holds from $V(\cdot) = E(\cdot)^2 - E^2(\cdot)$ that:

$$V_{\mathbf{X}_i} \left(E_{\mathbf{X}_{-i}}(\mathbf{Y} | \mathbf{X}_i = \mathbf{X}^*) \right) + E_{\mathbf{X}_i} \left(V_{\mathbf{X}_{-i}}(\mathbf{Y} | \mathbf{X}_i = \mathbf{X}^*) \right) = V(\mathbf{Y}). \tag{4.13}$$

The interpretation of this formula is that the output variance equals to the output variance due to \mathbf{X}_i alone (first term left – called the *main effect*) plus the expectation of the output variance due to all the other factors, both independently and by interaction with \mathbf{X}_i (second term left – called the *residual effect*). We can, then, deduce that a factor \mathbf{X}_i is important when its main effect is large or when its residual is small.

Additionally, the measure:

$$S_i = \frac{V_{X_i} \left(E_{\mathbf{X}_{-i}} \left(Y | \mathbf{X}_i = \mathbf{X}_i^* \right) \right)}{V(Y)} = \frac{V_i}{V(Y)} \quad (4.14)$$

is called the *first-order sensitivity* (FOS) effect of \mathbf{X}_i on \mathbf{Y} . It represents, quantitatively, the relative influence of the factor \mathbf{X}_i to the output variance independent of interactions. It is always between 0 and 1 and is commonly used for ranking the model factors, otherwise called the “*Factor Prioritization*” setting. The properties of this statistic are important in sensitivity analysis. First, by establishing the priority among inputs reduces output variance since an input with high value of its first order effect has good chance to be estimated when observations of \mathbf{Y} will become available (identifiable/estimable). Second, is that $\sum S_i$ is always less or equal than unity (≤ 1). At the extreme case of being equal to unity, the model is additive and the output variance is sensitive to each input factor independently. Finally, by the time $\sum S_i$ is considered a measure of model additivity, the result $1 - \sum(S_i)$ is indicative of the presence of interactions between model factors.

Analogously, to find higher than first order sensitivity (FOS) indices, calculation of the conditional variances of more than one factor is needed. For instance, the second-order sensitivity effect of $\mathbf{X}_i, \mathbf{X}_j$ on \mathbf{Y} , with $i \neq j$ is by definition:

$$S_{ij} = \frac{V_{ij}}{V(Y)}, \quad (4.15)$$

where V_{ij} is the variance from the interaction of \mathbf{X}_i and \mathbf{X}_j alone i.e.:

$$V_{ij} = V_{X_{ij}} \left(E_{\mathbf{X}_{-ij}} \left(Y | \mathbf{X}_i = \mathbf{X}_i^*, \mathbf{X}_j = \mathbf{X}_j^* \right) \right) - V_i - V_j \quad (4.16)$$

Altogether, the influence of a factor \mathbf{X}_i on \mathbf{Y} both independently and through interactions may be assessed by analogy to equation (4.13):

$$V_{\mathbf{X}_{-i}} \left(E_{X_i} \left(Y | \mathbf{X}_{-i} = \mathbf{X}_{-i}^* \right) \right) + E_{\mathbf{X}_{-i}} \left(V_{X_i} \left(Y | \mathbf{X}_{-i} = \mathbf{X}_{-i}^* \right) \right) = V(Y). \quad (4.17)$$

Here, the main effect represents the output variance due to all non- \mathbf{X}_i factors (\mathbf{X}_{-i}) and the residual equals to the expected variance of \mathbf{X}_i both individually and by interaction. This latter term is called the *total-order effect* of \mathbf{X}_i on \mathbf{Y} and the measure:

$$S_{Ti} = \frac{E_{X_i} \left(V_{\mathbf{X}_{-i}} \left(Y | \mathbf{X}_{-i} = \mathbf{X}_{-i}^* \right) \right)}{V(Y)} = 1 - \frac{V_{X_i} \left(E_{\mathbf{X}_{-i}} \left(Y | \mathbf{X}_{-i} = \mathbf{X}_{-i}^* \right) \right)}{V(Y)} \quad (4.18)$$

is called the total order sensitivity (TOS) index of factor \mathbf{X}_i . This term sums up the overall influence of a factor \mathbf{X}_i on the output variance and comprises of all the sensitivity terms of all the orders that include the factor \mathbf{X}_i :

$$S_{Ti} = S_i + S_{ij} + \dots + S_{ijk}, \text{ where } i < j < \dots < k. \quad (4.19)$$

If the total sensitivity index for a factor \mathbf{X}_i is small it is implied that its influence on the output variance, either independently or by interaction, is negligible. Thence, it can be fixed to any value \mathbf{x}^* without affecting the output variance. Obviously, if only the FOS index had been used for such inference, then it would have been in error because a factor can have

nominal influence individually and a significant one through interactions. Hence, the most important property of the total effect term is that it measures the overall effect of an input on the output variance and thereupon, can ascertain which of the factors are unimportant in our analysis and can be fixed in order to simplify our model. The latter is also known as “*Factor Fixing*” setting of sensitivity analysis.

Eventually, returning back to equation (4.12) and dividing with the output variance, it yields that:

$$\sum_i S_i + \sum_i \sum_{j>i} S_{ij} + \sum_i \sum_{j>i} \sum_{l>j} S_{ijl} + \dots + S_{123\dots k} = 1. \quad (4.20)$$

Equation (4.20), as well as all the others presented thus far is valid provided that model inputs are independent. Usually, indices up to second order are easily calculated. However, higher-order sensitivities are difficult and computationally expensive to measure since we need to calculate $2^k - 1$ sensitivity terms. Nevertheless, with the knowledge presented thus far, we can reveal the whole sensitivity pattern of the output, even for a nonadditive model.

4.4.2.3.1 Using scatterplots to measure the conditional variances

One way of calculating the conditional variances is from scatterplots. As mentioned before, scatterplots depict the relationship of the output Y while varying one factor X_i (Figure 4-2 and Figure 4-3). In order to calculate the FOS index one has to find the variance of the expected X_i value i.e. $V(E(Y|X_i))$. We begin by cutting the scatterplot in slices and for each slice the expected value of Y is acquired. This corresponds to keeping X_i constant while averaging over the rest of factors \mathbf{X}_{-i} , which, in turn, equals to $E(Y|X_i)$. Then, the limit of this while the slices become extremely thin equals to $V(E(Y|X_i))$ (Figure 4-4).

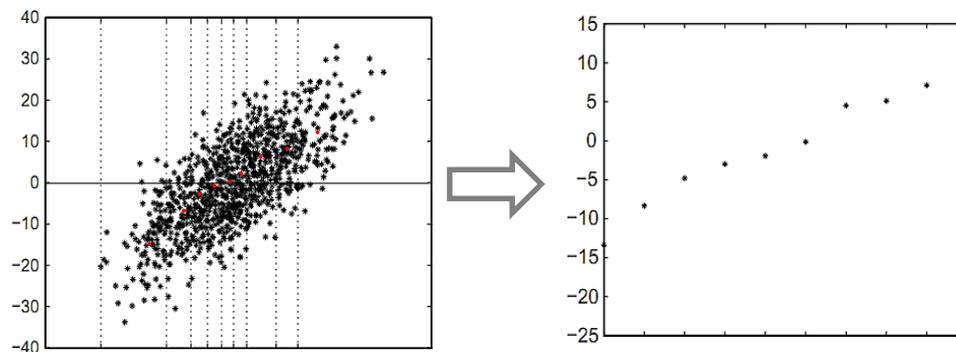


Figure 4-4 (Left) Cutting the scatterplot into slices and calculating the average output $E(Y|X_i)$. (Right) The variance of the expected values $V(E(Y|X_i))$ [165].

This technique, however, is prone to outliers. As a result, the calculated conditional variances are oftentimes inaccurate and, subsequently, the results become unreliable. Furthermore, higher order indices that represent interactions (if they exist) are extremely difficult to trace when no analytical representation exists especially above the third-order. In addition, such process is inefficient and costly because it requires the creation of 2^k scatterplots and calculation of equal number of sensitivity terms. Therefore, scatterplots are commonly used for preliminary screening of the output sensitivity and small scale models [172], [173].

4.4.2.3.2 Using Fourier transform to measure the conditional variances

Due to the fact that these limitation of the use of scatterplots, an alternative way to calculating the conditional variances is by use of the Fourier transform of the model described

by $\mathbf{Y}=f(\mathbf{X})$. In general, Fourier transform methods consider the same model scheme as mentioned initially but, in the frequency domain i.e. $y=f(\mathbf{X}(s))$. Theoretically, application of the ergodic theorem as demonstrated by Weyl in order to compute the variance of a function f yields:

$$V(\mathbf{Y}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f^2(x(s)) ds - \hat{f}_0^2, \text{ where } \hat{f}_0 = \int_{-\pi}^{\pi} f(x(s)) ds \quad (4.21)$$

Equation (4.21) is similar to equation (4.11) if the latter, which is a multidimensional integral given as sum, is expressed into one-dimensional integral form. Then, using the Parseval's theorem the above relationship can be expressed in Fourier series as:

$$V(\mathbf{Y}) = 2 \sum_{j=1}^{\infty} (A_j^2 + B_j^2) \quad (4.22)$$

where A and B are the Fourier coefficients of the cosine and sine series respectively and j the number of samples $j=1 \dots n$. Then, in order to acquire the conditional output variance for the calculation of the FOS index of a factor \mathbf{X}_i , we need to assign from the sum in equation (4.21) only the frequency of \mathbf{X}_i 's and its corresponding harmonics:

$$S_i = \frac{2 \sum_{j=1}^{\infty} (A_{j \cdot fs_i}^2 + B_{j \cdot fs_i}^2)}{2 \sum_{j=1}^{\infty} (A_j^2 + B_j^2)} \quad (4.23)$$

where $i=1 \dots k$ and fs_i are linearly independent integer frequencies. The sampling procedure for a k dimensional input space is created considering a space-filling curve of \mathbf{N}_{sample} points, according to the ergodic theorem mentioned earlier. Each point represents a transformed value of the \mathbf{X}_i factor using a periodic function $\mathbf{X}_i(s_j) = \mathbf{G}_i(\sin fs_i s_j)$, where s_j is a randomly chosen value between $(-\pi, \pi)$ and \mathbf{G}_i is a function chosen according the probability density functions of factor \mathbf{X}_i . During the calculations of the Fourier spectrum special care should be taken when choosing the frequencies in order to avoid interferences. For this reason, only a specific optimal set of frequencies is used which is *a priori* found from algorithms in literature and are free of interferences up to the 4th order [165], [174].

The algorithm just described is called FAST (Fourier Amplitude Sensitivity Test) and, as implied, it is able to calculate only the first order sensitivities. This is true since we excluded the interactions when we chose our frequency set in equation (4.23). However, the original algorithm has been extended in order to calculate the total sensitivity indices as well (eFAST - extended FAST) [175]. The latter is accomplished by calculating, at the same time, not only the optimal set of frequencies but also their complementary ones. According to the respective bibliography eFAST is an accurate and computationally efficient, in terms of the required number of model evaluations, algorithm. However, it is limited from the fact that when dealing with discrete inputs its efficiency is weakened considerably [46]. Since in our model we include an integer input factor (number of neoplastic layers) we cannot readily use this method for calculating the global sensitivities. Therefore, an equally efficient, if not better, method that is known for its ease in the algorithmic implementation and clarity in the derivation of the higher order indices is described next.

4.4.2.3.3 The Monte Carlo method

Theoretically, multidimensional integrals may also be estimated using Monte Carlo simulation if the sample size N_{sample} tends to infinity. Hence, it is straightforward that the conditional variances required for the assessment of sensitivity indices, of any order, can be calculated via Monte Carlo simulations. Equation (4.7), would then be:

$$E(Y) = f_0 = \int_{\mathbf{I}} f(\mathbf{X}) d\mathbf{X} \cong \frac{1}{N_{\text{sample}}} \sum_{m=1}^N f(\mathbf{x}_m) = \hat{f}_0 \quad (4.24)$$

and equation (4.12) of the output variance $V(Y)$ would also be:

$$V(Y) = \int_{\mathbf{I}} f^2(\mathbf{X}) d\mathbf{X} - f_0^2 \cong \frac{1}{N_{\text{sample}}} \sum_{m=1}^N f^2(\mathbf{x}_m) - \hat{f}_0^2 \quad (4.25)$$

Similarly, any one-factor conditional variance V_i is found if we square and integrate equation (4.9):

$$\begin{aligned} V_i &= \int_0^1 f_i^2(X_i) dX_i = \int_0^1 \left[\int_{\mathbf{I}} f(\mathbf{X}) \prod_{p \neq i} dX_p - f_0 \right]^2 dX_i = \\ &= \int_0^1 \left[\int_{\mathbf{I}} f(\mathbf{X}_p, X_i) \prod_p dX_p \times \int_{\mathbf{I}} (\mathbf{X}_p, X_i) \prod_p dX_p \right] dX_i - 2f_0 \int_0^1 \left[\int_{\mathbf{I}} f(\mathbf{X}) \prod_{p \neq i} dX_p \right] dX_i + f_0^2 = \\ &= \int_0^1 \left[\int_{\mathbf{I}} f(\mathbf{u}, X_i) d\mathbf{u} \times \int_{\mathbf{I}} f(\mathbf{v}, X_i) d\mathbf{v} \right] dX_i - f_0^2 = \int_{\mathbf{I}} f(\mathbf{u}, X_i) f(\mathbf{v}, X_i) d\mathbf{u} d\mathbf{v} dX_i - f_0^2 \cong \\ &\cong \frac{1}{N_{\text{sample}}} \sum_{m=1}^{N_{\text{sample}}} f(\mathbf{u}_m, X_{im}) f(\mathbf{v}_m, X_{im}) - \hat{f}_0^2, \end{aligned} \quad (4.26)$$

where \mathbf{u} and \mathbf{v} belong to the \mathbf{I}^{k-1} subspace. Equally, for higher-order terms the variance is estimated as:

$$V_{ij\dots k} \cong \frac{1}{N_{\text{sample}}} \sum_{m=1}^{N_{\text{sample}}} f(\mathbf{r}_m, X_{im}, X_{jm}, \dots, X_{km}) f(\mathbf{s}_m, X_{im}, X_{jm}, \dots, X_{km}) - \hat{f}_0^2 \quad (4.27)$$

where \mathbf{r} and \mathbf{s} belong to the \mathbf{I}^{k-s} subspace.

Equation (4.27) is generic and shows how the variance $V_{ij\dots k}$ that corresponds to a sensitivity term of any order may be estimated by Monte Carlo simulation. To calculate each variance [42], [171] proposed the generation of two matrices, say \mathbf{A} and \mathbf{B} , of N_{sample} rows and k columns. Then, one should take samples from all factors but X_i (X_{-i}) and store them in these matrices. The values of \mathbf{A} correspond to the first term of the product in the summand in (4.27) and the values of \mathbf{B} to the second term of this product. We now introduce matrices $\mathbf{A}_{\mathbf{B}}^{(i)}$ and $\mathbf{B}_{\mathbf{A}}^{(i)}$ where all columns are from \mathbf{A} and \mathbf{B} , respectively, except the i -th column which is from \mathbf{B} and \mathbf{A} , again respectively. Then, S_i can be computed from either couple of matrices, \mathbf{A} , $\mathbf{B}_{\mathbf{A}}^{(i)}$ or \mathbf{B} , $\mathbf{A}_{\mathbf{B}}^{(i)}$, e.g. :

$$V_{X_i}(E_{X_{-i}}(Y|X_i)) = \frac{1}{N_{\text{sample}}} \sum_{j=1}^{N_{\text{sample}}} f(\mathbf{A})_j f(\mathbf{B}_{\mathbf{A}}^{(i)})_j - f_0^2, \quad (4.28)$$

where $(\mathbf{A})_j$ denotes the j -th row of matrix \mathbf{A} and

$$f_0^2 = \frac{1}{N_{\text{sample}}} \left(\sum_{j=1}^{N_{\text{sample}}} f(\mathbf{A})_j \right)^2 \quad (4.29)$$

is the mean. Similarly, the computation of S_{Ti} can be realized from:

$$V_{\mathbf{X}_{-i}}(E_{\mathbf{X}_i}(Y|\mathbf{X}_{-i})) = \frac{1}{N_{\text{sample}}} \sum_{j=1}^{N_{\text{sample}}} f(\mathbf{A})_j f(\mathbf{A}_{\mathbf{B}}^{(i)})_j - f_0^2 \quad (4.30)$$

Note here that the output variance $V(Y)$ can be readily calculated from equation (4.25) using as \mathbf{x}_m either the \mathbf{A} or the \mathbf{B} matrix.

Several numerical improvements of estimators (4.28) and (4.30) have been proposed in literature [171]. For the purposes of this thesis the following have been employed:

$$V_{\mathbf{X}_i}(E_{\mathbf{X}_{-i}}(Y|\mathbf{X}_i)) = \frac{1}{N_{\text{sample}}} \sum_{j=1}^{N_{\text{sample}}} f(\mathbf{B})_j \left(f(\mathbf{A}_{\mathbf{B}}^{(i)})_j - f(\mathbf{A})_j \right), \quad (4.31)$$

for the FOS computations and

$$E_{\mathbf{X}_{-i}}(V_{\mathbf{X}_i}(Y|\mathbf{X}_{-i})) = \frac{1}{2N_{\text{sample}}} \sum_{j=1}^{N_{\text{sample}}} \left(f(\mathbf{A})_j - f(\mathbf{A}_{\mathbf{B}}^{(i)})_j \right)^2, \quad (4.32)$$

for the TOS computations.

Here, it is essential to note that the re-sample values i.e. the random numbers chosen for creating \mathbf{B} must always be identical in all estimations of any order. Otherwise, we introduce noise in the estimations that will affect the ranking of the input factors. To avoid this we generate a matrix of random numbers for the simulations beforehand. Since we need N_{sample} simulations and two sets of sampling values (for sampling (\mathbf{A}) and re-sampling (\mathbf{B})) the dimensions of the random number matrix will be $(N_{\text{sample}}, 2k)$. For instance, in equations (4.26) and (4.27) \mathbf{u}_m , \mathbf{r}_m and $\mathbf{X}_{i,j,\dots,km}$ values will be calculated from the first k columns while \mathbf{v}_m and \mathbf{s}_m values of \mathbf{B} from the last k ones. As for the experimental design of the random generated values, quasi-Monte Carlo algorithms of stratified sampling outperform crude Monte Carlo. More specifically, the statistic convergence of the former is $n^{-1+\epsilon}$, where ϵ is arbitrary small, in comparison to the latter which is $n^{-1/2}$ that needs a lot more evaluations (N) and, thus, converges slower. Commonly, the low discrepancy LP_{τ} number sequences (also known as Sobol' sequence) are used when the number of input factors is less than 51 [168].

Computation of variances using Monte Carlo methods, however, introduces accuracy loss. Indeed, the error in the estimations is inversely proportional to the number of simulations which is the reason why sometimes such methods are unreliable when simulations are time consuming. For this, Sobol' introduced the "probable error" as an estimate of variability. Under the assumption of a standard normal distribution, the confidence interval would then be 0.6745 times the standard error. However, the standard distribution assumption is valid when a large sample is also considered. That is the case, the use of bootstrap confidence intervals have been suggested as more reliable. In general, bootstrap confidence is robust here by the time we sample from a set of independent equally distributed data. Basically, a random number of values is chosen from the initial sample and from this subsample a new set of values is created *with replacement* i.e. some of the values can be repeated in the new

set. Then, sensitivities are re-estimated and the result is called bootstrap estimate. When a large number of such estimates has been acquired ($boot > 1000$) we calculate their standard error. Thence, the sensitivity's confidence interval would be ± 1.96 times the standard error of the bootstrap estimate. The latter is reliable if the bootstrap estimates are normally distributed, which is valid if $boot$ is quite large. Otherwise, if the distribution is skewed we can estimate the confidence through the percentiles of the bootstrap distribution. Concluding, the bootstrap procedure may also be applied for quasi-random sample without any further errors.

4.4.3 Implications of variance-based GSA in identifiability testing

From the above discussion we can easily deduce that the main purpose of GSA methods, especially the variance-based ones, is to explore how the input uncertainty is passed on the modeled output. In this way, the most significant factors or variables affecting the model predictions are identified. Then, a natural extension of this, in terms of identifiability, is the idea that if a parameter influences the output a lot, it also means that when the output is measured its dynamic features will occur due to the existence of this particular parameter. Effectively, if the output is insensitive to a parameter this parameter may be generally acknowledged as being non-identifiable.

Since our focus is on dynamic systems, the computations of the FOS and TOS indices proceed from equations (4.31) and (4.32) for all the experimental instances of time. Based on this procedure the respective FOS and TOS functions are formulated. It is straightforward therefore, that the FOS function may be used to provide some insight on the identifiability of input parameters, especially the highest ranked ones. However, there is no solid justification that the high ranking of a parameter implies its identifiability because, as seen, its interactions may also be significant. In general, the non-identifiability of a parameter may originate from structural properties of the model; from value constraints of the input parameters; from extreme experimental constraints on the input signals; and/or from the measurement time instants. Thus, sensitive parameters could also be non-identifiable under certain circumstances. Additionally, the TOS function may be used to point out the unessential parameters.

On the same page, the links between global identifiability and global sensitivity techniques in dynamic models has not been fully explored. As a matter of fact, the only solid connections between parameter non-identifiability and output sensitivity can be narrowed down to the following [176]. First, the nullity (or inferiority to an empirical threshold) of a TOS implies a non-identifiable parameter. Second, the non-injectivity of the model output and subsequently of the TOS functions with respect to the parameter p_i also leads to the non-identifiability conclusion. Third, the linear dependence with respect to time between parameters implies both the non-identifiability of the parameters in question and the colinearity of their total sensitivity functions. The converse, however, is not necessarily true. This means that collinear analysis of TOS functions is not always valid. Therefore conclusions about parameter non-identifiability drawn from the analysis of global sensitivity functions must be carefully analyzed.

As a last note it should be stressed that GSA is an *a posteriori* analysis viz. its inferences are calculated from model simulations. This means that the credibility of the FOS and TOS indexes largely depends on the accuracy and completeness of the model's structure and the probability density function assumed for each input factor. Therefore, to sum up, GSA inferences should be treated with caution when model identifiability is addressed, either theoretically or practically.

4.5 APPLICATION OF GLOBAL SENSITIVITY ANALYSIS TO THE PROPOSED MODEL

For the purpose of setting the grounds for the solution of this parameter estimation problem we performed three essential *a priori* (theoretical) analyses under an idealized, *in silico*, environment assuming that the data are noise-free and that the model structure is perfect and complete. These analyses are: a) GSA; b) parameter identifiability; and c) parameter estimability. As indicated earlier, GSA provides an estimate of model's output sensitivity to its input parameter variations. For the purposes of this thesis parameter identifiability uses GSA outcomes as an input and designates a certain parameter as identifiable when its sensitivity function is neither null nor collinear with another parameter's sensitivity function. Finally, we performed parameter estimability analysis using again GSA outcomes as an input in order to define the level of interactions and/or correlations between the identifiable parameter sets.

On the basis of these analyses, the biological parameters of the model have been ranked with the purpose of reducing its complexity and to improve the posedness of the inverse problem. More specifically, parameters that display high sensitivity, no collinearities and minimum interdependency with other parameters are considered as the key determinants of the bio-optical dynamic characteristics. Parameters that are not fulfilling these criteria can be considered as non-identifiable and can be kept constant at around their nominal values.

4.5.1 Details of the employed methodology

As explained earlier, GSA methods are better suited in cases where the model is a highly nonlinear algorithm and especially in cases where it cannot be expressed with a single analytical function. This is exactly the case of our model, where the number of neoplastic layers parameter (M) is not just an independent variable of the model's differential system; it also defines the number of equations in the differential equation system. Particularly variance-based GSA methods are considered as model independent approaches, indicating their suitability for the problem addressed in this thesis.

We have already described that there exist two distinct settings in variance-based GSA, namely the factor prioritization and the factor fixing. In brief, the purpose of factor prioritization is to designate, quantitatively, the participation of each input factor to the output variance so that the parameters are ranked according to their importance. The highest ranked parameters are expected to be dominant in the formation of the model's output line shape, which, in principle, maximizes the possibility for their identification. Contrariwise, the purpose of factor fixing is to define the unimportant parameters that have little or no relevance with the output variance and can be fixed to their nominal values.

4.5.1.1 The sensitivity analysis

Thereupon, we have performed our GSA calculations based on the Sobol's method as described in paragraph 4.4.2.3.3 where the variances in (4.12) are considered as multidimensional integrals and their estimates can be computed using a quasi-Monte Carlo algorithm. The sampling procedure was conducted utilizing low discrepancy LP_r number sequences due to their optimal space filling properties and fast statistical convergence. Then, supposing that the values of the calculated sensitivities are normally distributed, the 95% bootstrap confidence interval has also been calculated in all measurements since Monte Carlo estimates impinge accuracy losses [170].

4.5.1.2 The identifiability analysis

We have repetitively indicated that the general purpose of identifiability analysis is to designate which of the pertinent model parameters can potentially be acquired from available input/output data and which will be impossible to assess [28]. Formally, this can only be realized by defining explicitly the relationship between the model parameters and the measured output. However, for computational models, such as the one described in this thesis, this is infeasible in most cases [177]. Therefore, GSA has been utilized for investigating, numerically, how the couplings between the model parameters create the profiles of the measured output [178]. From eq. (4.19), let $S_{T_i|t}$ denote the TOS of parameter i at time t , where $i=1\dots k$, and using the notation presented in eq. (1.34) let

$$\mathbf{S}_D = \begin{bmatrix} S_{T_1|t_1} & \cdots & S_{T_k|t_1} \\ \vdots & \ddots & \vdots \\ S_{T_1|t_s} & \cdots & S_{T_k|t_s} \end{bmatrix} \quad (4.33)$$

be the dynamic global sensitivity matrix of the model's output to each of the j postulated parameters at discrete times $t_1 \dots t_s$. Then, from the equations in (4.33) the parameter correlation matrix \mathbf{C} can be calculated with the aid of Pearson's linear correlation coefficients, which provide the amount of correlation between any two different columns. The parameters with correlations with all other parameters between -1 and +1 will be identifiable while, the parameters that have at least one correlation of exactly -1 or +1 with another parameter will be unidentifiable. In other words, parameters that influence the model's output in exactly the same or exactly the opposite way cannot be identified.

As indicated in paragraph 4.4.3 the links between global identifiability and global sensitivity techniques in dynamic models have not yet been fully explored. To this end, the conclusions about parameter non-identifiability drawn from the above analysis must be treated with caution. To this end, our results will be backed up with qualitative experimental data in the following paragraphs and with quantitative ones in the next chapter.

4.5.1.3 The estimability analysis

The possibility for a meaningful estimation of a model parameter increases when this parameter affects the model output directly rather than indirectly, through its interactions with the other parameters [44]. One way to express this quantitatively is to define estimability as the ratio of the FOS index of a certain parameter S_i over the sum of the TOSs of the entire parameter set. As discussed earlier, sensitivity measures are expected to be functions of time, which is also expected to apply to the estimability index or ratio (ER). This index is defined as:

$$ER(t) = S_i(t) / \sum_{j=1}^k S_{T_j}(t). \quad (4.34)$$

We can easily conclude that the task of estimating an identifiable parameter simplifies when its FOS index is highly ranked and when it affects the model's output directly, which is the case when the ER index approaches its maximum value.

4.5.1.4 Parameter values and sampling distributions of input parameters

Parameter value ranges and nominal values were obtained according to paragraph 3.4 and further literature research. More specifically, we collected several values for each parameter until a histogram could be made possible for the purpose of approximating the corresponding cumulative distribution. Generally, the trend was a beta distribution which could be approximated with a uniform one due to its small curvature. The parameters: a , β_{ES} , pH_{ES} and pH_{IS} fall under this category, while the parameter β_{IS} appeared to have a “pure” uniform distribution. Parameters b , K_V and g_{TJ} , even if their lowest and highest values have been indicated, no clear definition of a nominal value was found to be available. For this reason, a triangular distribution has been appointed using as the most probable value the middle of each parameter’s value range. Finally, the distribution of the *layer* parameter has been chosen to be uniform between its nominal bounds. Table 4-1 summarizes the bounds of the values and respective distributions that were appointed for each particular parameter for performing GSA.

Table 4-1 Value ranges of model parameters employed in GSA studies. The assumed types of distributions used for parameter sampling in the context are also provided.

Parameter	Range	Distribution	Reference
a	10-20 μm	uniform	[113]
b	0.1-0.8 μm	triangular	[113]
K_V	0.1-1 M/s	triangular	[151]
β_{ES}	10-30 mM	uniform	[84], [98]
β_{IS}	30-50 mM	uniform	[84]
N	1-15	-	[113]
pH_{ES}	6-7	uniform	[90], [179]
pH_{IS}	7-7.4	uniform	[90], [179]
g_{TJ}	0.01-1	log-triangular	[150]

4.5.1.5 Experimental data

Apart from the aforementioned analysis experimental data were also used for a preliminary validation of the model’s structure, of the GSA and of the parameter reduction and estimation processes. The details of the experimental procedure have been described in paragraph 2.4.6. In this study 30 DR vs. time curves were used, in total, corresponding to 10 CIN I (low grade (LG)) and 20 CIN II, III (high grade (HG)) cervical neoplasias (biopsy confirmed). As discussed in paragraphs 2.4.5 and 3.5 the intracellular [H] time variations, occurring after the application of AA solution, are directly correlated with dynamic DR signals that are measured *in vivo*. For the purpose of comparing and fitting optical experimental data with model predictions, both DR and [H] curves have been normalized. The normalization procedures have been as follows.

4.5.2 Curve normalization

Earlier we pointed out the operating principles of the developed DCE-OI apparatus. Under that description, a reference image captures the intensity of the backscattered light before the application of the biomarker on the tissue. As a rule, this intensity is largely determined by the spectral characteristics of the nonabsorbed and back scattered photons from the vascular plexus. Because of the irregular distribution of this biologic material it follows that prior to the application of the biomarker every image pixel is given a separate in-

tensity value. This means that after the application of the AA the dynamic range of the phenomenon's evolution will be biased i.e. different from one tissue location to another. To compare, therefore, and fit the optical experimental data with model predictions the image intensities must be properly scaled.

To this end, the intensities of each dynamic curve need, first, to be normalized to their respective initial values. Let, the initial intensity be I_0 and I_t the intensity at time t , then the normalized value of the intensity will be:

$$I_{norm} = \frac{I_t - I_0}{I_0}, \quad (4.35)$$

where I_{norm} is the normalized intensity value. Similarly, the dynamic range of each curve should also be reconfigured. Hence, if the dynamic range of the intensity in the raw images is 0-255 then, the normalized curve would extent from 0 to $(I_{max} - I_0)/I_0 = I_{max,new}$. In succession, since the $I_{max,new}$ is different from one curve to another rescaling of the intensity timeseries must be performed using the general rescaling formula:

$$I_{new} = C * \left(1 - \frac{I_{old} - A}{B - A} \right) + D * \left(\frac{I_{old} - A}{B - A} \right), \quad (4.36)$$

where A, B, C and D denote that we scale from [A,B] to [C,D] and I_{old} and I_{new} denote the pre- and post- scale values, respectively. In our case $A=0$, $B=I_{max,new}$, $C=0$ and $D=1$. However, since $A=C=0$ and $D=1$ the above formula reduces to:

$$I_{norm,scaled} = \frac{I_{norm}}{I_{max,new}} = \frac{\frac{I_t - I_0}{I_0}}{\frac{I_0}{255 - I_0}} = \frac{I_t - I_0}{255 - I_0}, \quad (4.37)$$

where $I_{norm,scaled}$ denotes the normalized intensity scaled in [0,1]. Finally, if we examine the last formula we see that this is a classic standardization/normalization formula:

$$I_{[0,1]} = \frac{I_t - I_0}{I_{t_{max}} - I_0}, \quad (4.38)$$

which can be interpreted as a systematic bias correction where the normalized intensity is found by the subtraction by a measure of location and division by a measure of scale.

As assumed, the photons that backscatter from each neoplastic layer add up and formulate the dynamic signals the normalization of which has just been described. Then, in a similar way, the [H] concentration in each layer should also be normalized. For this, we first need to define the initial and the maximum [H], respectively. The former is easy since it is equal to the initial intracellular [H] value ($10^{-pH(0)}$) that we apply in the beginning of a model simulation and lays between 10^{-7} - $10^{-7.4}$. The latter, however, has not been that trivial.

Basically, in order to find the maximum [H], we need to find, in literature, the lowest pH value that an epithelial cancer cell is capable of maintaining after acute acidification. However, every experimental reference relative to killing cancer cells by exposing them to an acidic stress evaluates cells' survival after incubation for several minutes, hours or days. As

it is easily understood, this situation is incompatible to the question we want to answer (acute acidification). Naturally then, our bibliographic focus had been tuned on experiments that exploit the dynamics of the intracellular acidification at shorter time scales. Commonly, such experiments are the ones that investigate the active pumping efficiency of tumor cells. Unfortunately, a weak point here is that the acidification dynamics depend on the concentration of the acid (usually NH_4Cl or CO_2) and on the concentration of the artificial buffers that are used during the experimental procedures. As a result, our main concern, which is the extent of acidification, may be found to vary from one reference to another. For this, the lowest achievable pH value could not be clarified explicitly. Still, however, it was indicated that *in vivo* intracellular pH can lay, regularly, between 6.5 and 7.2 pH units and therefore experiments target acute acidifications inside this region [89], [180].

The only reference closest to answering this is described in [181]. There the acidification of breast and bladder carcinoma cells by weak acids under acidic extracellular environments is examined. In short, what this reference indicates is that: a) weak acids cannot be cytotoxic *in vivo* when nominal concentrations are applied due to the blood flow and the high intracellular buffering capacity and b) acidification by monocarboxylic acids (such as AA), *in vitro*, resulted in the reduction of intracellular pH by ~ 0.45 units per minute. Then, adoption of (b) means that a single epithelial cell allows its pH to drop, at the very most, by ~ 0.45 units in one minute after acute acidification (*in vivo* it should be lower). On such basis, the intracellular pH drop, *in vivo*, is has been heuristically adopted to lay between 0.3 and 0.4 units for a single epithelial cell. Thereupon, the formula for normalizing the intracellular [H] would be:

$$[H]_{\text{IS}}^i(t)_{\text{norm}} = \frac{[H]_{\text{IS}}^i(t) - 10^{-\text{pH}_{\text{IS}}^i(0)}}{10^{-\text{pH}_{\text{IS}}^i(0)+\lambda} - 10^{-\text{pH}_{\text{IS}}^i(0)}}, \quad (4.39)$$

where λ is a value between 0.3 and 0.4. Then, the model's output will comprise the sum of the normalized [H] that accumulates inside the neoplastic cells of the entire tissue cross-section.

4.5.3 Results

For the purposes of this investigation the quasi-Monte Carlo algorithm of GSA was implemented and executed in Mathwork's MatlabTM environment and twenty thousand parameter sets were used for estimating the FOS and TOS indices. Since the bio-optical processes under investigation are dynamic in nature, we have calculated these indices, for all model input variables, as a function of the time lapsed after the application of the biomarker. In addition, the calculations for the identifiability and estimability analyses have also been performed in the same environment using in-house developed codes.

4.5.3.1 FOS and TOS indices of the model parameters

Figure 4-5 illustrates the total order sensitivity of the model's output to its input parameters' variations. The TOS of the K_v parameter has been found to be zero for all time points and for this reason it has been omitted from subsequent analysis as not fulfilling the non-null TOS criterion. Figure 4-6 depicts the average FOS (dark gray) and TOS (light gray) index estimates for all time points in (FOS) descending sequence for the rest eight model input parameters. As it can be seen in fig. 3 the average TOS indices of parameters: neoplastic layers (N), the IS (a) and ES (b) sizes, tight junction pore function g_{TJ} and extracellular acidity pH_{ES} vary in the range of 0.12-0.39, while the average TOS value range for ion buffering pa-

rameters β_{ES} , β_{IS} , and intracellular acidity pH_{IS} is 0.02-0.04. It is therefore evident that the GSA parameter ranking has resulted in two distinguishable parameter groups. From another viewpoint, discrimination of these two parameter classes can be based on the fact that the average FOS index of the lower ranked member of the first group (g_{TJ}) has a greater value than the TOS of all the members of the second (lower ranked) group.

Figure 4-7 captures the peak ER values obtained at the corresponding time points. The time points (in s) that correspond to max ER values are shown at the top of the histogram bars. ER index comprises a second parameter ranking option. As discussed earlier, ER informs on whether a parameter affects the model's output directly or indirectly. As it can be seen in Figure 4-7, the ER ranking is more or less similar to the TOS ranking of Figure 4-6. The only difference is that g_{TJ} parameter is ranked first in the ER ranking, something that can be attributed to the fact that the FOS index is quite high at around the first time instances of the simulated dynamic phenomenon (see Figure 4-5). Hence, it is evident that the same set of parameters is better suited for estimation.

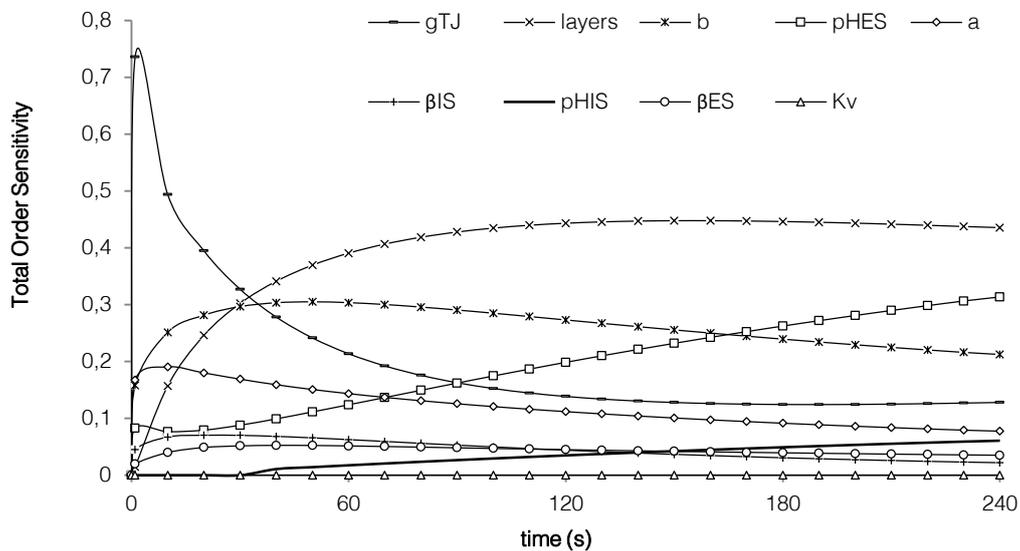


Figure 4-5 The total order sensitivity index plotted as a function of time for the entire set of model parameters. The time range corresponds to the measurement period of the simulated dynamic optical effect

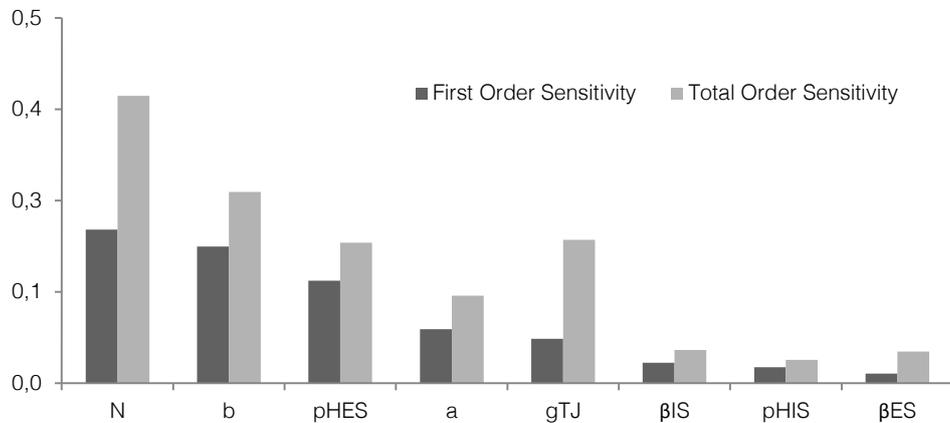


Figure 4-6 The first order (dark gray) and the total order (light gray) sensitivity index estimates in (first order) descending sequence for eight model input parameters after discarding Kv .

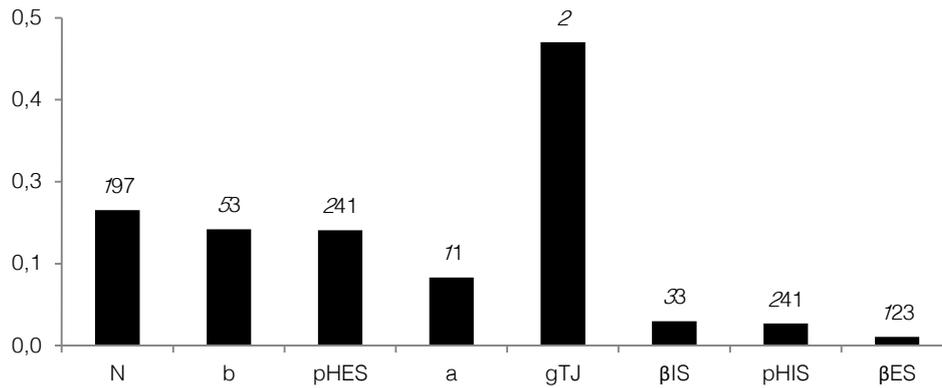


Figure 4-7 The peak estimability ratio ER_{max} for the remaining eight model parameters. The time points (in s after the application of acetic acid solution) at which the max ER values are achieved appear at the top of the histogram bars.

4.5.3.2 Identifiability and estimability of the tissue's structural and functional parameters

On the basis of the data of Figure 4-6 and Figure 4-7, the three lower ranked parameters can be potentially discarded as negligibly affecting and substantially indirectly (low FOS, TOS and ER values respectively) the model's output. This model's parameter reduction would have a clear benefit to the reduction of the problem's complexity/dimensionality, when attempting to estimate model parameters from simulated experimental data. However, before concluding on the set of parameters suitable to be discarded, identifiability analysis was performed. As discussed in paragraph 4.5.1 this analysis outputs as non-identifiable the parameters that are highly correlated by having the same or exactly the opposite effect on the model's output. Quantitatively, the condition that should apply for the case of non-identifiable parameters is the Pearson's linear correlation coefficient, calculated for any two different parameters, to have at least one correlation coefficient value equal with either -1 or +1, per parameter pair.

Figure 4-8 depicts the correlation matrix for the eight parameters in all possible pairs. The correlation coefficients are displayed numerically and their values can be easily visualized with the aid of gray shade pallet. Darker shades correspond to high coefficient values. As it can be seen the parameters: layers (N) and g_{TJ} display, in all pairs with the other parameters, correlation coefficients smaller than 0.95 and therefore they are, in principle, identifiable in the presence of all the other parameters. Parameter pairs displaying correlation coefficients greater or equal to 0.95 are assumed to be (by rounding to one digit) equal to one. This implies that the parameter pairs (pH_{ES} , pH_{IS}), (b , β_{ES}) and (a , β_{IS}) cannot be disentangled or, in other words, they cannot be identified separately. This is consistent with Figure 4-5, where, as it can be seen, the corresponding curve profiles of each pair are very similar, with their main differences to be the magnitude of the sensitivity values. Naturally, the origins of this unity correlation go back to the construction of the model and particularly to the formulation of the epithelial transport phenomena. The unity correlation coefficient between the buffering power β of the intra- and extra- cellular spaces and the sizes (a , b) of these spaces indicate that their effect to the model's output is expressed with a similar time function. This is consistent with the biology and the function of the epithelial tissue since a possible change (increase) in the space size parameters (a , b) results in a reduction of ion concentration within the same space, a result that can be equivalently obtained by increasing its ion buffering power. Similar logic applies to the (pH_{ES} , pH_{IS}) pair where the same effect to the model's

output can be obtained by either perturbing directly pH_{IS} or indirectly through pH_{ES} (recall passive diffusion effects described in paragraph 3.4.2.1).

These observations render clear that the parameters β_{IS} , β_{ES} , and pH_{IS} , do not pass the combined high sensitivity and identifiability criteria and for this reason their initial values will be kept constant at around their nominal values. It is not expected this reduction of the dimensionality of the problem to have any significant negative impact to the functional and structural information that can be gained. This is because, to the best of our knowledge, there is no reported evidence to support that β_{IS} , β_{ES} , and pH_{IS} , are changing with neoplasia progress [182]. On the contrary, the correlation between the parameters layers (N), the sizes of the IS (a) and ES (b), g_{TJ} and pH_{ES} with neoplasia growth is well established.

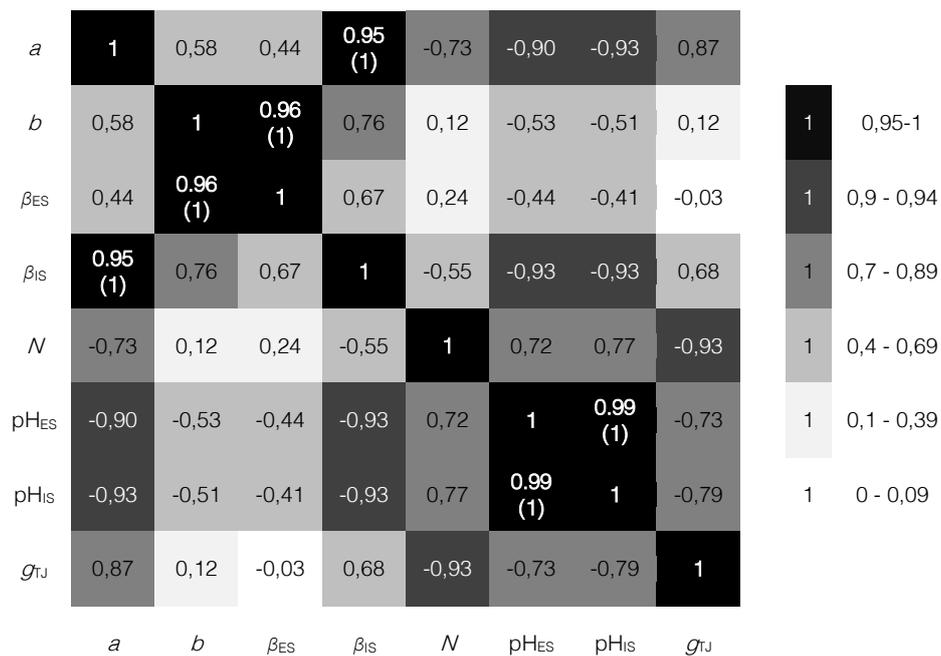


Figure 4-8 The correlation matrix for the eight parameters in all possible pairs. Parameter pairs (pH_{ES} , pH_{IS}), (b , β_{ES}) and (a , β_{IS}) display (with rounding) unity correlation coefficient.

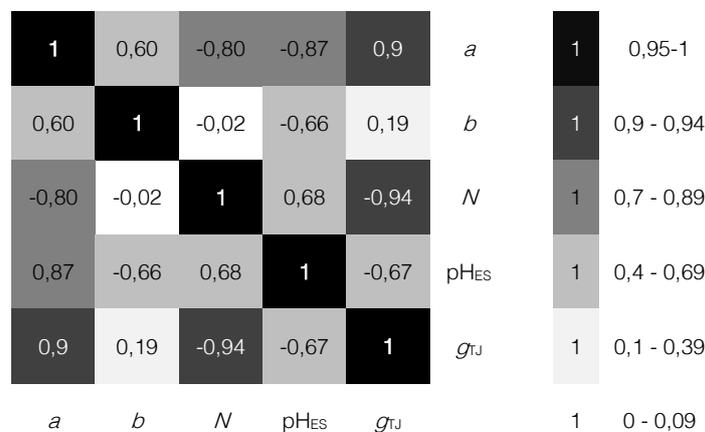


Figure 4-9 The correlation matrix of the five highly ranked parameters (factor prioritization) after fixing the insignificant parameters (factor fixing). No unity correlation coefficient has calculated for all the pairs of different parameters. In both cases the correlation coefficients can be visualized with the aid of gray shade pallet. Darker shades correspond to high coefficient values.

Figure 4-9 illustrates the correlation matrix recalculated for only the highly ranked parameters listed in the last sentence. Apparently, all correlation coefficient values, estimated for all possible parameter pairs are different to either 1 or -1. This, in combination with the fact that this parameter subset is relatively highly ranked from the sensitivity and estimability analyses perspective, indicates that all the members of this subset pass the combined sensitivity, estimability and identifiability criteria.

Concluding, the initial nine input parameters of the model have been reduced to five by discarding parameter K_V as displaying null sensitivity function and the parameters β_{IS} , β_{ES} , and pH_{IS} , as not fulfilling the combined non-unity correlation coefficient, displaying also low ER and sensitivity indices. The remaining five, highly ranked, parameters can therefore be regarded as the key determinants of the model's output.

4.5.3.3 Validation of the results

It is reasonably expected the reduction of the problem's dimensionality, by fixing insignificant parameters, to enhance the well-posedness of our parameter estimation problem. It is, however, understood that to show the well-posedness of a problem, we need to prove existence, uniqueness, and continuous dependence of the solution to the parameter values. Nevertheless, it is possible to perform a first, high level validation of the model's structure, calibration and dimensionality reduction by attempting to estimate the selected parameter set through the comparison of experimental results with model predictions. This comparison cannot be regarded as a solution to the inverse problem in the strict sense. In the next chapter more efficient and mathematically rigorous methods for solving this problem by employing global optimization approaches are presented. In any case, the dimensionality reduction, presented below, is essential for facilitating the convergence of the problem to unique solutions.

With these constraints in mind we have used normalized DR (NDR) vs. time curves obtained from *in vivo* examinations of women with cervical neoplasia of different grades. Figure 4-10 illustrates 30 NDR vs. time experimental curves, 10 corresponding to CIN I (LG) and 20 to HG neoplasia (10 CIN II and 10 CIN III). Within the context of GSA analysis, 20,000 normalized $[H^+]$ vs. time curves are generated by the model, corresponding to a wide variety of randomly selected input parameter combinations. This comprises an intermediate step for performing the Monte Carlo calculations (see paragraph 4.4.2.3.3). The generated by the model curves were used for fitting the experimental data with the purpose of performing a first estimation the neoplasia-related parameters for each particular CIN class. Figure 4-11 illustrates typical model-predicted curves representative for each CIN class. It is reasonably expected that, in principle, a certain model curve that fits the experimental data would not necessarily correspond to a unique set of parameter combination. This source of uncertainty is preliminary addressed by fitting each particular experimental curve with ten, best fitting model curves. In all cases, a satisfactory fitting accuracy ($R^2 > 93\%$) was obtained. With this approach, one hundred best fitting model curves were assigned to each CIN class, totaling to a set of three hundred curves.

Figure 4-12 illustrates the box and whisker plots statistical analysis applied to the extracted three hundred values per parameter. These diagrams provide a convenient graphical representation of the dispersion, skewness, and outliers in a single given data sample, but also enable a visual comparison of different data samples. The top and bottom edge of the box in a boxplot represent the 25th and 75th percentiles of the sample, respectively; in consequence, the box height corresponds to the interquartile range (IQR). The line in the

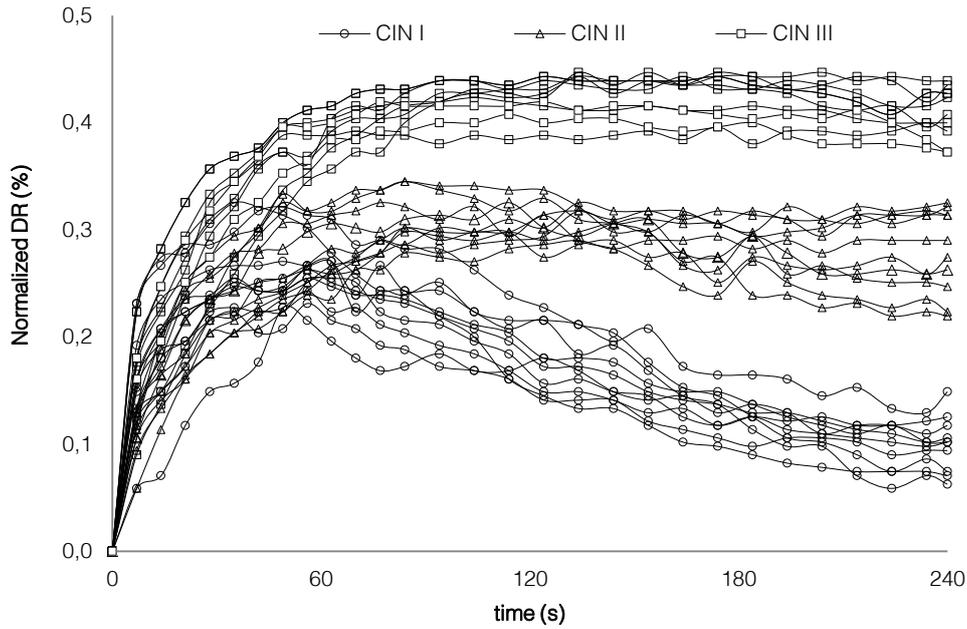


Figure 4-10 Normalized experimental Diffuse Reflectance (DR) vs. time curves, 10 corresponding to CIN I (low grade (LG)) and 20 to high grade (HG) neoplasia (10 CIN II and 10 CIN III).

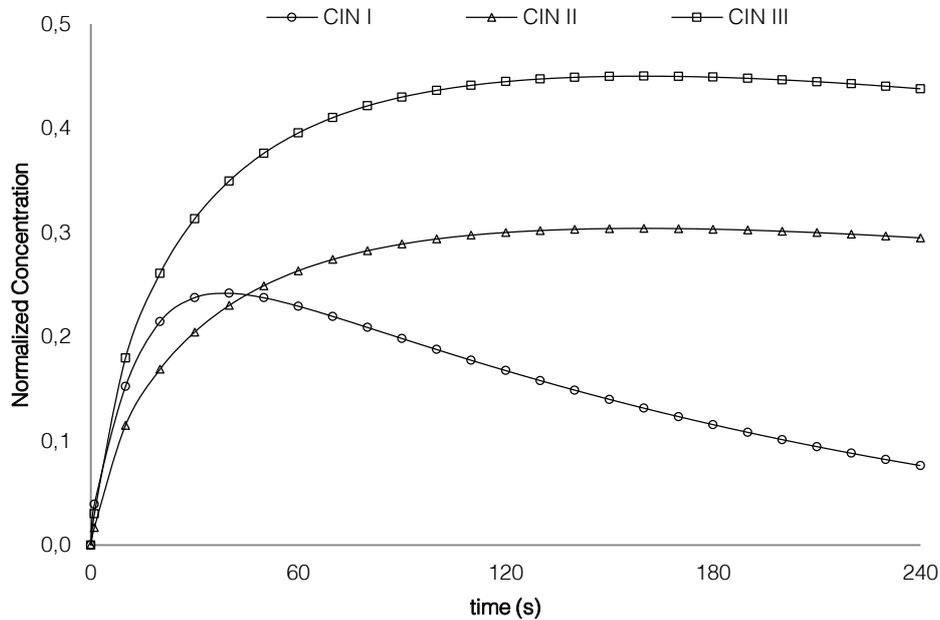


Figure 4-11 The model's outputs that best fit the average experimental curve of each particular CIN class

middle of the box corresponds to the sample median. The sample mean is represented with a circle. The “whiskers”, i.e., the two lines extending above and below the box, represent the sample range. The maximal length of the whiskers is set to $1.5 \times \text{IQR}$. Data points above and below the whiskers' end points correspond to the sample outliers and are represented with “+” markers. Finally, the notches of the boxes represent the variability of the median in the sample. The width of a notch is computed so that boxplots whose notches do not overlap

have medians that are significantly different at the 0.05 significance level, assuming a normal data distribution.

The analysis displayed in Figure 4-12 refers solely to the highly GSA ranked parameters and has been performed for each CIN class separately. These results are unique in the sense that from a dynamic diffuse reflectance curve computed, non-invasively, for a single pixel's area, a list of quantitative parameters can be derived, having direct relation with neo-

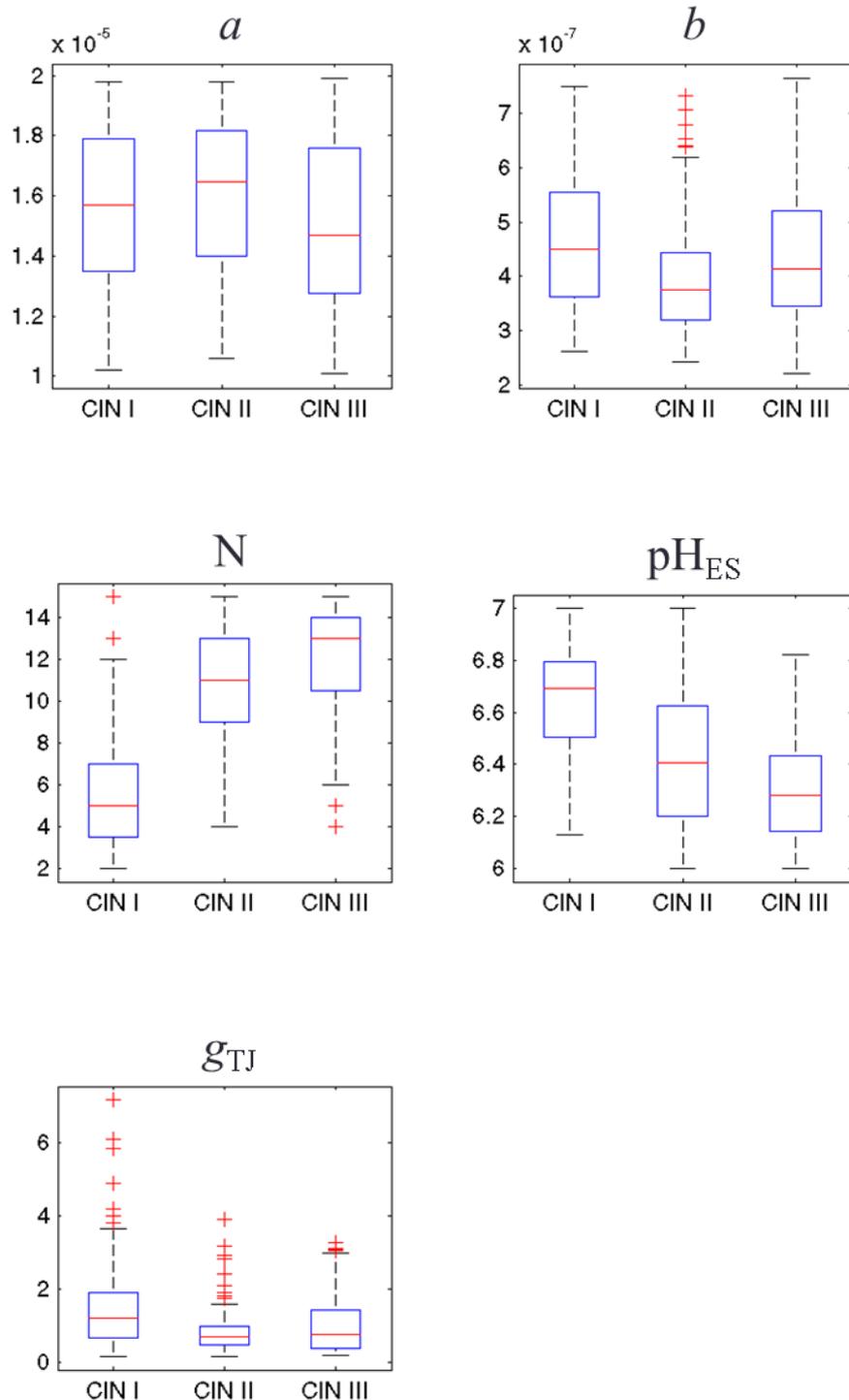


Figure 4-12 Box and whisker plots of the estimated parameter values for the highest ranked parameters according to the GSA outcomes. The statistical analysis has been realized for each CIN class sepa-

plasia growth. Particularly, we have already established that the number of neoplastic layers N and the extracellular/intracellular dimensions (a , b respectively) are structural parameters that are changing with neoplasia progress [113]. Histological grading is based mainly on the qualitative assessment of these features, in (dead) biopsy samples. The parameters g_{TJ} and pH_{ES} are affecting the functionality of the epithelium and also are closely related with neoplasia growth [99], [104]. These functional parameters can only be assessed in living tissues and in fact, to the best of our knowledge, no other method has been presented so far, enabling their assessment in living epithelia.

As it can be seen in Figure 4-12 the number of dysplastic layers increases as neoplasia progresses from LG (CIN I) to HG (CIN II/III) in a statistically significant manner. A similar trend seems to apply in the case of pH_{ES} , where differences between CIN I and CIN III appear to be statistically significant and in full agreement with the physiology of the disease [179]. Moreover and again in accordance with CIN's biology, the median of parameter g_{TJ} seems to follow a decreasing trend as the neoplasia develops [104]. However the box percentiles are overlapping indicating no pure statistical significance in the predicted values. The same applies to the a , b parameters, since there is no clear trend of their median values in relation with the status of the disease. This would be due to the fact that the changes of these biological characteristics occurring during the progress of the disease lay below the detection level of our method. We reasonably expect that by employing global optimization methods, the signal-to-noise ratio will be improved, thus further reducing the detection cut-off for these parameters.

It is essential to stress here the fact that the diagnostic importance of the pH_{ES} parameter has only recently been identified and appreciated. According to the so-called "acid-mediated tumor invasion model", there is a strong link of the reduced pH_{ES} with the ability of tumor cells to form invasive cancers [99]. Particularly, it has been shown experimentally that H flow to peritumoral normal tissue provokes extracellular matrix degradation and normal cell necrosis or apoptosis. At the same time, tumor cells develop resistance to acid-induced toxicity during carcinogenesis. This permits them to spread and invade the damaged normal tissue promoting metastasis. Furthermore, promotion of cancer progression and invasion has also been shown to occur due to alterations in the tight junction complexes [104]. Particularly, the link between neoplasia and tight junction functional status lies with claudins, which are the structural and functional components of the tight junctions that modulate the diffusion of molecules across the paracellular route. The dysfunction of these proteins has been associated with several cancers indicating altered tissue polarity, abnormal cellular organization and loss of differentiation. These have been shown to strongly contribute to the barrier's leakiness and allow progression and invasion for both cervical and several other epithelial tumors [122].

The fact that these parameters are the key determinants of measured/modeled dynamic optical characteristics indicates that they can be estimated concurrently and for every spatial point of the tissue through noninvasive optical measurements. It follows that the *in vivo* assessment and mapping of these parameters would comprise a valuable tool for more accurate grading, and more importantly for estimating the risk for metastasis on a personalized basis. Ideally, diagnosis and prediction can be performed at the same consultation and non-invasively. Apart from that, the method holds the promise to contributing to the better understanding of the mechanisms involved in neoplasia growth and to the development and evaluation of targeted and personalized therapies.

4.6 MODEL RESTRUCTURING

The aforementioned study has proven that the systematic reduction of the model using the combined results of the sensitivity, identifiability and estimability analysis is theoretically possible. Then, as indicated, the initial set of nine input parameters can be reduced to a set of five key determinant ones; namely, the sizes of the IS and ES (a , b), the extracellular pH (pH_{ES}), the number of neoplastic layers (N) and the TJ geometrical dimensions (g_{TJ}). Even so, the parameter estimation results illustrated by the box-and-whisker plots in Figure 4-12 indicated that the changes of the parameters a , g_{TJ} and b during the progress of the disease are poorly detected. Chances are that these parameters will still be difficult to estimate even if more efficient and mathematically rigorous methods for solving the inverse problem are employed. To reduce to a minimum the probability of this and further improve the posedness of the inverse problem some essential modifications concerning these model parameters have been performed.

To begin with, there are three reasons for a to be regarded as a no relevant parameter and be fixed to a nominal value. First, among the highest ranked parameters, a is the lowest ranked according to the results in Figure 4-6; second, it has the smallest estimability ratio according to the results in Figure 4-7; and third, but foremost, it is the least identifiable parameter according to the correlation matrix depicted in Figure 4-8. Particularly, as far as the latter is concerned, if we average the absolute correlation values of each parameter over all the pairs that it participates, then the average correlation of a is the highest one and approximately 0.8. On top, the correlation of a with N , pH_{ES} and g_{TJ} is more than 0.8 indicating a negative effect on the identifiability of these parameters. Therefore, all things converge to the fixing of a even if, in principle, its correlations are below the cut-off value (0.95). Besides, the statement that the IS of an abnormal cell alters during neoplasia is theoretically weak. To explain this further, imagine that the tissue cells are divided into normal and abnormal ones. Macroscopically, during neoplasia development, the diverse normal cells are displaced and, eventually, subdued by the smaller, homogeneous abnormal ones, allowing larger intercellular spaces to be created [113], [183]. Looking this from a normal cell's viewpoint it means that their population decreases, their volume shrinks and the space between them starts to increase. On the other hand, from an abnormal cell's point of view this means that, their population increases, their volumes remain the same and the space between them keeps increasing during carcinogenesis. With these in mind, since our model emulates only the abnormal cell group, then we should have allowed from the beginning only the parameters concerning the number of neoplastic cell layers (N) and the width of the ES (b) to alter and leave the size of the IS (a) to a nominal value between 10 and 20 μm . Consequently, if this theoretical analysis is regarded as a fourth reason for fixing parameter a , we can set our parameter estimation problem to capture one less structural parameter without loss of generality.

Next, due to the fact that the correlation of g_{TJ} and N is very close to unity a different but also relative to g_{TJ} description of the solutes' permeabilities and their variation during carcinogenesis has been put forward. This new description is based on a model of paracellular diffusion presented in [184] which mathematically defines the paracellular permeability of a solute in epithelial physiological monolayers as a function of its nominal chemical characteristics and of the paracellular channels' characteristics. Specifically, the following formula is used:

$$P_{TJ} = \left(\frac{\varepsilon}{\delta}\right) \cdot D_{aq} \cdot F\left(\frac{r}{R}\right), \quad (4.40)$$

where (ε/δ) is the effective porosity-pathlength ratio of the channel, D_{aq} is its aqueous diffusion coefficient, F is the Renkin hydrodynamic sieving function³ for cylindrical water channels, r is the hydrodynamic radius of the solute and R is the effective pore radius of the paracellular channel. About the terms appearing on the right-hand side of equation (4.40) we know that D_{aq} of any solute can be readily calculated from the Stokes-Einstein relation of kinetic theory⁴ and that $F(r/R)$ can be calculated provided that the pore radius is known. In fact, the only unknown term is the effective porosity-pathlength ratio of the TJ complex. However, we can approximate the value of the latter if we employ, again, the paracellular permeability value of Mannitol in a normal cervical epithelium (0.22 μ m/s). The procedure is as follows. Initially, we find that Mannitol's D_{aq} is $D_{Mannitol}=9.3\times 10^{-6}$ cm²/s. Next, assuming that the radius of a TJ pore in a normal cervical epithelium is approximately 7 \AA the sieving function value $F(r_{Mannitol}/R=7\text{\AA})$ can be readily calculated [113], [149]. Finally, we can approximate the unknown (ε/δ) ratio specifically for the normal cervical epithelium by dividing $P_{Mannitol}$ with the value of the calculated product $D_{Mannitol}\times F(r_{Mannitol}/R=7\text{\AA})$. Having this term calculated, we can return to equation (4.40) and determine the respective permeabilities of the model solutes in the normal cervical epithelium. Our results indicate that the permeabilities of AA and Ac are approximately 1.06 μ m/s and 1.07 μ m/s, respectively and of H approximately 44 μ m/s. Then, for this second description the fluctuation of these paracellular permeabilities due to carcinogenesis will be attributed, as before, to the altered geometrical properties of the TJ pores. Keeping in mind that the tight junctions ease off during neoplasia development, it is reasonable to assume that the solute permeabilities should also increase [122]. This means that the $(\varepsilon/\delta)\times F(r/R)$ term of equation (4.40) increases probably through an increase to the effective pore radius. In [184], the maximum possible pore radius attributed to leaky epithelia has been calculated and was round about 18 \AA . Thereupon, we can approximate the increment of g_{TJ} between tight and leaky epithelia through algebraic manipulation of equation (4.40). Our results indicate that the value g_{TJ} should range between 7 and 30cm⁻¹ for the AA and Ac ($r_{AA}\cong r_{Ac}$) and correspondingly between 45 and 55cm⁻¹ for the H. If these values of g_{TJ} are multiplied with the D_{aq} of AA, Ac and H we can assess the approximate values of their permeabilities during carcinogenesis. On such basis, the parameter g_{TJ} will represent several leakiness scenarios which can be parameterized using the fact that $R\in[7\text{\AA}, 18\text{\AA}]$.

Lastly, in an attempt to decrease the overlapping of the estimated values of b shown in the box-and-whisker-plots of Figure 4-12 we have truncated the range of its size based on tabulated values acquired from literature. Specifically, Table 4-2 depicts the average extracellular width at every epithelial section calculated from histological slides according to [185]. Thereupon, let in a CIN I epithelium 1/3 of its volume be occupied by neoplastic cells, in a CIN II epithelium 2/3 of its volume be occupied by neoplastic cells and in a CIN III epithelium almost the entire tissue volume be occupied by neoplastic cells. In addition, recall that our model has been constructed to capture only the abnormal cells of the epithelium and that neoplasia develops from the bottom of the epithelium towards the upper layers. Then, according to Table 4-2 the initial values of b can be truncated and allowed to range between 0.4 μ m and 0.8 μ m (indicated in gray shades). The latter is further supported from the fact that

³ $F(r/R) \approx 1 - 2.1044(r/R) + 2.089(r/R)^3 - 0.948(r/R)^5$

⁴ The aqueous diffusion follows the Einstein relation of kinetic theory: $D=\mu_w k_B T$, where μ_w is the mobility of the solute in water, k_B is the Boltzmann's constant and T is the temperature.

other groups have found that the volume of the ES increases approximately 6 times during carcinogenesis i.e. $\sqrt[3]{6} \cong 1.82 \cong 0.8/0.4$ (included in [1]). Therefore, without loss of generality this new range of values for b has been adopted.

Table 4-2 Width of ES associated with each cell layer (adopted from [185], table 7-7-2). The gray shades illustrate the ES width that corresponds to the occupancy of the epithelium by neoplastic cells.

Pathological Type	Extracellular width in layer (μm)			
	Basal	Parabasal	Intermediate	Superficial
Normal	0.4	0.2	0.1	0.05
CIN I	0.6	0.4	0.2	0.1
CIN II	0.6	0.4	0.4	0.2
CIN III	0.8	0.6	0.6	0.4

4.6.1 Comparison with the earlier calculations

By asserting these changes to the full model and reconfiguring the correlation matrix we get the result depicted in Figure 4-13. We should note that the samples for calculating these correlations have been drawn, for all parameters, from uniform distributions within their new (for b and g_{TJ}) and previously reported range of values.

As it is clearly seen the impact on the identifiability of the parameters of the full parameter set has been, in principle, positive. Particularly, one can notice that the collinearity of the previously reported parameter pairs (pH_{ES} , pH_{IS}), (b , β_{ES}) and (a , β_{IS}) has been slightly dropped. However, their correlation values remain among the five highest ones between all parameter pairs. This implies that the separate identification of these parameters will still be difficult. What is of interest though is the significant increase of the correlation between the parameter pairs (g_{TJ} , \mathbf{N}) and (g_{TJ} , a). Especially the latter will be now impossible to disentangle due to the almost total correlated behavior of the respective parameters. This mathematically validates our earlier proposition that parameter a should be fixed due to its high collinearity with the rest of the parameters in the proposed set. As far as the correlation value of the (g_{TJ} , \mathbf{N}) parameter pair is concerned, we can see the correlation remained unchanged with previously (0.93). Even so, this correlation is significantly lower (0.83) in the reduced parameter set case Figure 4-13 when compared to the results shown in Figure 4-13 where it remained the same (0.94). Probably, the individual interaction of parameter \mathbf{N} and g_{TJ} with parameter a has been indirectly influencing the correlation between the former two. In general, after the fixing of parameter a and the reduction of the model's complexity, as before, the identifiability of the highest ranked parameters, which are now four, has improved significantly. This means that the proposed restructuring was successful in improving the posedness of the inverse problem which we will attempt to solve in the next chapter.

4.7 SUMMARY

In this chapter, we reported the first *in silico* simulation and sensitivity analysis of dynamic bio-optical signals obtained from *in vivo* optical molecular imaging of cervical neoplastic epithelium. In general, sensitivity analysis results can be used to establish a ranking of importance for the parameters, allowing to fix unessential parameters and to prioritize the most influential ones and facilitating the identification task. Within this context, GSA provides sensitivity inferences for complex, highly nonlinear systems that are much more informative and robust than their local counterparts which are limited by model's linearity and analyticity.

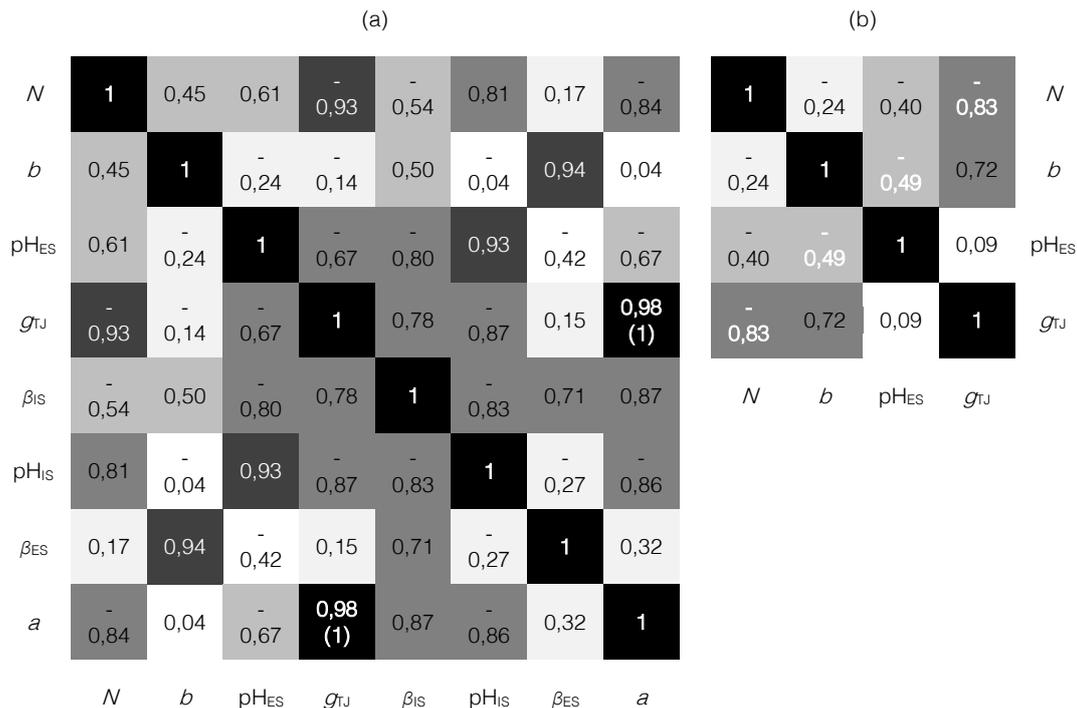


Figure 4-13 a) The new correlation matrix for the full (eight) parameter case. Notice that although the collinearity of the pairs (pH_{ES} , pH_{IS}), (b , β_{ES}) and (a , β_{IS}) has been dropped they still have the highest ones among all the pairs in the matrix. b) The new correlation matrix for the reduced (four) parameter case. After fixing parameter a and reducing the model's complexity, the identifiability of the highest ranked parameters, which are now four, has improved significantly. In both cases the same palette has been used as in Figure 4-8 and Figure 4-9.

Thereupon, towards verifying the uniqueness of the inverse problem's solution we have performed consecutively three *in silico* analyses; namely, global sensitivity analysis, parameter identifiability analysis and parameter estimability analysis. On the basis of these analyses, we were able to decide which of the pertinent model parameters can potentially be estimated from available input/output data and which are impossible to assess. Specifically, we have identified a set of parameters that fulfill the following criteria: a) they are the key determinants of the line-shape of the model's output that fits the experimental dynamic optical data and b) they display no collinearities and minimum interdependency with each other. The initial nine input parameters of the model have been reduced to four by discarding parameter K_V as displaying null sensitivity function and the parameters β_{IS} , β_{ES} , and pH_{IS} , as not fulfilling the combined non-unity correlation coefficient, displaying also low estimability ratio and sensitivity indices. Additionally, the IS linear dimension parameter (a) did not fulfill the collinearity criterion. On the basis of both sensitivity and estimability analyses we have identified that the key determinants (in descending ranking order) are: the number of neoplastic layers (N), the size of the ES (b), the extracellular pH (pH_{ES}) and the tissue's porosity (g_{TJ}).

These findings allowed us to introduce a new, noninvasive method for estimating concurrently and non-invasively a set of biological parameters with proven correlation with neoplasia growth, and hence of undisputed diagnostic value. This is obviously very essential since valuable knowledge will be gained on how functional and structural biological factors are independently or in combination altering during the neoplasia growth, as well as for developing and evaluating novel diagnostic and treatment modalities.

RELEVANT PUBLICATIONS

- [1] Papoutsoglou G., Anastasopoulou A., Stavrakakis G., Scouter P., Balas C., “In Vivo Dynamic Imaging, in Silico Modeling and Global Sensitivity Analysis for the Study and the Diagnosis of Epithelial Neoplasia.”, 33rd Annual International IEEE EMBS Conference, August 30 - September 3, Boston Marriott Copley Place, Boston, MA, USA, 2011.
- [2] G. S. Papoutsoglou and C. Balas, “Estimation of neoplasia-related biological parameters through modeling and sensitivity analysis of optical molecular imaging data.”, IEEE Transactions Bio-Medical Eng., vol. 60, no. 5, pp. 1241–9, May 2013.

5 SOLUTION OF THE INVERSE PROBLEM

5.1 INTRODUCTION

Thus far we have described the pathophysiology of the biological system under study and presented its mathematical realization by putting forward a working model. Then, towards solving the system identification problem the question whether the proposed model structure allows for the unique estimation of the unknown system parameters when experimental data become available has been discussed. Our theoretical results on the latter, indicate that the inverse problem should be adequately solved when the parameter complexity of the model is narrowed down to the most influential input parameters.

Even though these theoretical studies concerning the model inputs are expected to enhance the well-posedness of our parameter estimation problem, they do not prove existence, uniqueness, and continuous dependence of the solution to the parameter values. To establish this requires the practical confirmation of the *a priori* analyses by efficient and mathematically rigorous estimation methods. The realization of this will be described in the current chapter.

5.2 STATEMENT OF THE PARAMETER ESTIMATION PROBLEM

According to paragraph 1.7.3.3 the first step towards solving the parameter estimation problem is to state the inverse problem as the minimization of a cost function that calculates the goodness of the fit between the model output and a given experimental data set, subject to the dynamics of the system [186]. Thereupon, our optimization problem has been formulated as follows:

$$\begin{aligned}
 &\text{minimize} && J = \left\| y^{\text{exp}}(\mathbf{p}, t) - y^{\text{M}}(\bar{\mathbf{p}}, t) \right\|, && (5.1) \\
 &\text{subject to} && f\left(\frac{d\mathbf{x}}{dt}, \mathbf{x}, \bar{\mathbf{p}}, \mathbf{u}, t\right) = 0 \\
 &&& \mathbf{x}(t_0) = \mathbf{x}_0 \\
 &&& t \in [0, 240] \\
 &&& \mathbf{p}_L \leq \bar{\mathbf{p}} \leq \mathbf{p}_U \\
 &&& \mathbf{u} = \mathbf{u}_0
 \end{aligned}$$

where J is the cost function that utilizes the Euclidean norm (length) between experimental (**exp**) and model (**M**) predicted data, t is the time in seconds, \mathbf{p} and $\bar{\mathbf{p}}$ are the experimental and estimated parameter sets, respectively, f describes the dynamic model, \mathbf{x} is the differential state variables (as in eq. (1) – (4)), \mathbf{p}_L , \mathbf{p}_U are the lower and upper bounds of the parameters values, respectively, that act as inequality constraints, $\mathbf{u}=\mathbf{u}_0=0.5M$ is the initial [AA] that

is applied on the tissue which is always constant and, thus, not included in the minimization process and \mathbf{x}_0 is a vector of the initial state values i.e. the initial [TA] and [H] values applied at each layer both extracellularly and intracellularly.

From a strict mathematical viewpoint, the formulation just described is that of a nonlinear, mixed integer programming problem (MINLP) with differential-algebraic constraints (DAEs). In most cases, finding an optimal solution for problems of this type is a very difficult, or sometimes impossible, task. This happens because such problems are generally non-convex and unless some prior knowledge about the region of the global optimum is known, *a priori*, their solution will most likely be of local optimality. For tackling these kind of problems global optimization (GO) strategies are commonly employed.

5.3 GLOBAL OPTIMIZATION METHODS

The general idea behind GO of MINLP problems is to scan the entire feasible parameter space in order to find regions of potential global optima and then by using different combinatorial patterns to reach the global solution. Under this framework, either deterministic or stochastic strategies can be employed [187]. Deterministic strategies involve the transformation of the problem in the form of a surrogate representation that, in turn, can be solved deterministically and in polynomial time. For this, the definition of the optimization problem as an explicit, closed-form, mathematical representation is prerequisite. The stochastic strategies, on the other hand, do not assimilate this formulation rather; based on probabilistic approaches, they treat the problem as a 'black box'. Regardless, both strategies lack of strong theoretical guarantees to converge, efficiently, to the global optimum since this type of problems are generally NP-hard problems. Specifically, the deterministic methods even though they provide some level of assurance, they do so only for some class of problems and not in general [186]. Then again, stochastic methods give no guaranty of reaching the global optimum rather they can only designate the region close to the global optimum with relative efficiency. It is evident, therefore, that there exist a multitude of alternative solution methodologies that can potentially suit to our parameter estimation problem.

From the structure of the proposed estimation problem two distinct features can be discerned which theoretically designate the class of the most appropriate GO methodology for solving our minimization problem. As indicated in paragraph 4.2, the differential equations are non-factorable due to the existence of the dynamic ionization constants φ and ω in equations (3.15) and (3.16) which introduces high levels of nonlinearity in the equation system. Second, the representation of the model's equation system has no closed-form due to its adaptable/flexible structure in terms of abnormal cell layer integration. Recalling our conceptual model (paragraph 3.2), the number of layers has been defined so as to specify the height of the cell stack, proportionally to the severity of neoplasia. Mathematically, this means that the number of layers is not just an independent integer parameter; in fact, it defines the number of differential equations that are integrated in the system. More specifically, each time the number of layers are set to increase or decrease by one, a four differential equations system (eq. (1)-(4)) is automatically added to or deducted from the final differential equation system of the model, respectively. Therefore, the developed model is a nonlinear multi-resolution design that cannot be represented analytically. For this, direct-search (derivative-free) GO methods are mostly suitable.

In literature numerous alternative algorithms can be found which provide solutions for our nonlinear parameter estimation task [187]. As said, however, no GO method can guarantee optimality. Often, the set of solutions found by several dissimilar methods is utilized in

order to decide the degree of convergence that is actually achieved [186]. Therefore, based on their published performance, a series of available, state-of-the-art global optimizers that can treat models as a 'black box' have been selected in order to evaluate the uniqueness of the solution to our parameter estimation, in practice.

Specifically, we employed the gblSolve (DIRECT), the thermodynamic adaptive simulated annealing (tdASA), the Mixed Integer Distributed Ant Colony Optimization (MIDACO), the controlled random search with competing heuristics (CRSwCH), the differential evolution (DE), the shuffled complex evolution (SCE), the particle swarm optimization (PSO) and the Evolutionary Strategy with Stochastic Ranking (SRES), respectively. The first belongs to the deterministic family of algorithms while the latter to the stochastic one. For this choice someone would argue that it is biased towards the stochastic GO methodologies. However, with these methods we attempt to cover most of the diverging universe of GO methods that handle MINLP black-box optimization problems in which the stochastic approaches prevail over the deterministic ones not only in number but also in scientific application. Probably, a future work will be to apply hybrid deterministic/stochastic methods that utilize model alternates for finding the optimal solution of this parameter estimation problem, such as higher order polynomials, kringing, etc. [177], [188], [189]. Next, a description of the basis of each of the employed methods is provided. As a rule, a uniform probability distribution is assumed for all random decisions unless otherwise stated.

5.3.1 Control Random Search (CRS)

The CRS algorithm is a random search procedure effective in the identification of the global optimum in both constrained and unconstrained objective functions with multiple extrema. It is a direct method and is used regardless of the existence of continuous variables. The core algorithm mainly focuses on the full search of the k -dimensional parameter space, rather than on the speed of the fitting procedure. Thus, it is often used in cooperation with other, faster, converging GO algorithms for further fitting their results.

The CRS algorithm initially entails the definition of a search domain Ω that is specified within the limits of the k variables. Over Ω a predefined population, say \mathbf{P} , of parameter sets, called points, is randomly chosen, consistent with any constraints, and evaluated according to the objective function. The remaining of the algorithm is an iterative optimization of this population. The procedure is as follows. First, the worst point in terms of high objective value is found. Second, a new trial point is created with the purpose of replacing this worst one. In the original CRS algorithm of this latter procedure the point reflection was employed as proposed by Nelder and Mead [190]. Specifically, $k+1$ points are randomly chosen from the population, thus creating a sub-population known as the simplex (\mathbf{S} : $\mathbf{S} \subset \mathbf{P}$). From this $k+1$ dimensional simplex an arbitrary point ($\mathbf{P}^{(arb)}$) is extracted while the centroid (\mathbf{G}) of the remaining k points is calculated. Then, the new trial point is defined as the image (reflection) point of $\mathbf{P}^{(arb)}$ with respect to the centroid, \mathbf{G} :

$$\mathbf{P}^{(trial)} = 2 * \mathbf{G} - \mathbf{P}^{(arb)}, \quad (5.2)$$

where $\mathbf{P}^{(trial)}$ is the next trial point, \mathbf{G} is the centroid and $\mathbf{P}^{(arb)}$ is the arbitrary point chosen from the simplex of the $k+1$ randomly chosen sets. If this new $\mathbf{P}^{(trial)}$ set of parameters does not violate any constraint, then its objective value is obtained and compared against the worst one. If it is lower, the new point $\mathbf{P}^{(trial)}$ replaces the worst one in the population. Otherwise the procedure is repeated until a better point is actually generated.

5.3.1.1 CRS with Competing Heuristics (CRSwCH)

Historically, the CRS algorithm was first implemented by W. L. Price in 1976 and since then several variants have been developed [191], [192]. Most of the algorithm alternatives propose a different heuristic mechanism for the creation of $P^{(trial)}$, either by a different reflection schemes or by a completely different procedure. Generally, these modifications are mathematical, non-deterministic, rules purposed to produce a new estimated point $P^{(trial)}$ inside the feasible search space. These CRS algorithms are simple and objective and possess several properties that are desirable for effective and efficient GO. However, they also have certain weaknesses. First, they treat each region of the sampled area in a nonpreferential manner, and convergence can be slow. Second, they place their major emphasis on the best point in the current sample, so that the search can easily become biased toward the region of a local minimum. Third, the CRS strategy of always replacing the current worst point in the population by each newly generated point causes the population to shrink into a small region very quickly. This increases the possibility that the population will begin to contain repeated points, therefore becoming degenerate and causing the search to terminate prematurely. In order to minimize this possibility, the sample size has to be sufficiently large. Under these circumstances, the CRS with Competing Heuristics (CRSwCH) version presented in [192] by Josef Tvrđik has been exploited for our purposes.

In general, the CRSwCH is an adaptive search algorithm that belongs to the stochastic, population-based family of GO strategies. It integrates four well-known modifications of the original CRS algorithm that have been separately used in solving, with great success, several GO problems. Three of the heuristics are based on a randomized reflection in the simplex. Particularly, a new trial point $P^{(trial)}$ is generated from the simplex using the relation:

$$P^{(trial)} = G + rand_{[s, \alpha-s]} \times (G - P^{(worst)}), \quad (5.3)$$

where $P^{(worst)}$ denotes the point with the highest objective value in S . The multiplication factor $rand_{[s, \alpha-s]}$ is a random variable distributed uniformly in $[s, \alpha-s]$, where α ($\alpha > 0$) and s ($0 < s < \alpha/2$) are input parameters. One of the employed heuristics (1st reflection scheme) utilizes the $\alpha=2$ and $s=0.5$ values while a second one (2nd reflection scheme) uses the $\alpha=5$ and $s=1.5$ values respectively. In the third heuristic (3rd reflection scheme) the $P^{(worst)}$ in equation (5.3) is replaced by $P^{(best)}$ i.e. the point in P with the minimal objective value. The rest of the k points in the simplex are randomly chosen as before. As input parameters for this heuristic the $\alpha=2$ and $s=0.5$ values are used again.

The fourth heuristic does not employ a point reflection scheme; rather it is based on differential evolution (DE). Although we will extensively elaborate on the DE algorithm in the next paragraphs we briefly state here that, contrary to the previous heuristics, this one generates the new trial point in a genetic-like fashion. Specifically, a temporary point $P^{(temp)}$ is constructed according to

$$P^{(temp)} = r_1 + F(r_2 - r_3), \quad (5.4)$$

where r_1, r_2 and r_3 are three random points taken from \mathbf{P} . The scaling factor $F > 0$ is an input parameter according to:

$$F = \begin{cases} \max \left(F_{\min}, 1 - \frac{|obj_{\max}|}{|obj_{\min}|} \right) & \text{if } \frac{|obj_{\max}|}{|obj_{\min}|} < 1 \\ \max \left(F_{\min}, 1 - \frac{|obj_{\min}|}{|obj_{\max}|} \right) & \text{otherwise,} \end{cases} \quad (5.5)$$

where obj_{\min} and obj_{\max} are the minimum and maximum objective values in \mathbf{P} and $F \in [F_{\min}, 1)$. Based on this temporary point, the elements of the new trial point $P^{(trial)}$ are generated by the crossover of $P^{(temp)}$ and of another, random, point ($P^{(random)}$) taken in \mathbf{P} according to:

$$P_i^{(trial)} = \begin{cases} P_i^{(temp)} & \text{if } rand_{(0,1)}^{(i)} \leq Cr \text{ or } i = l \\ P_i^{(random)} & \text{if } rand_{(0,1)}^{(i)} > Cr \text{ and } i \neq l, \end{cases} \quad (5.6)$$

where $i = \{1, 2, \dots, k\}$ – recall that k reflects the number of model parameters – and l is a ran-

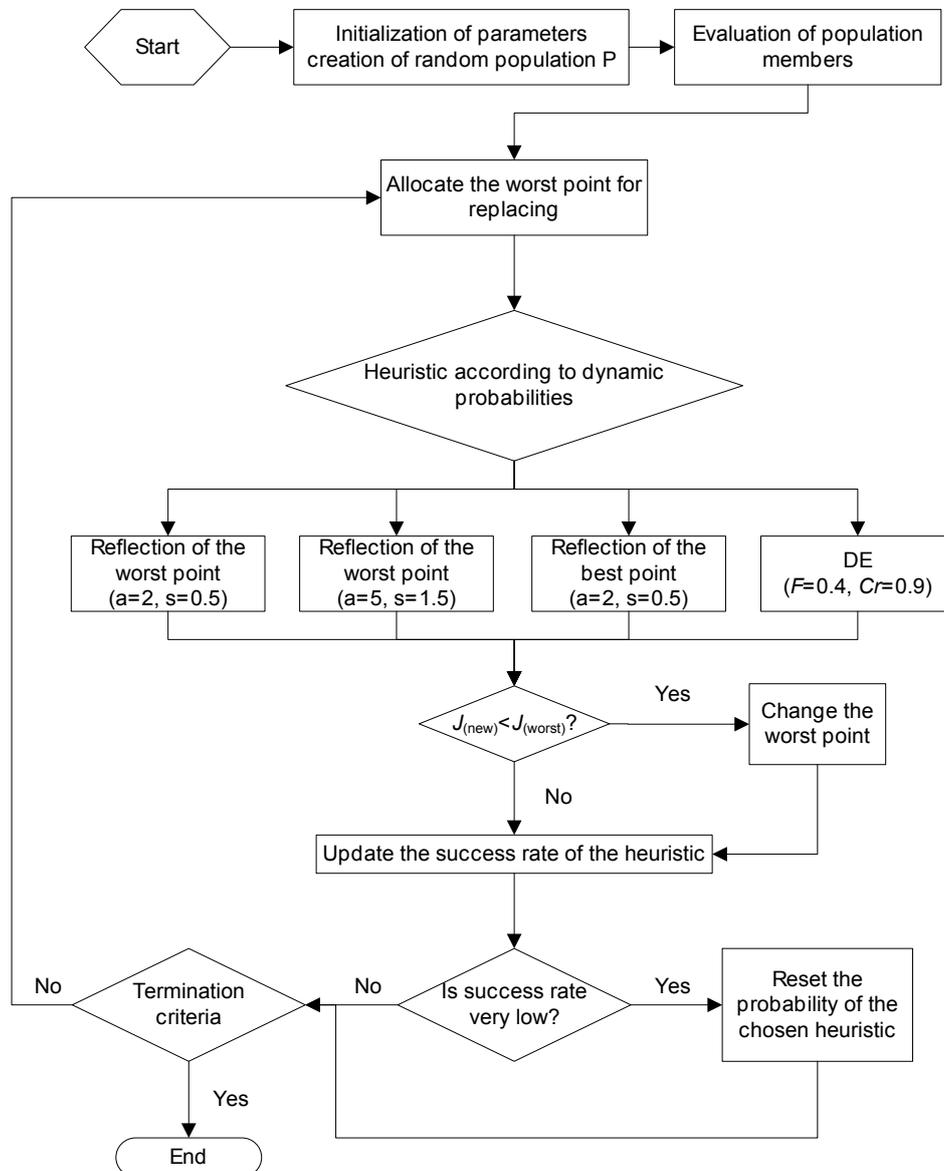


Figure 5-1 Flowchart of the CRSwCH algorithm

domly chosen integer from $\{1, 2, \dots, k\}$ and $Cr \in [0, 1]$ is an input parameter that influences the number of elements exchanged during crossover. Within this context, the fourth heuristic uses as input for F and Cr the 0.4 and 0.9 values, respectively.

The choice between these four heuristics is performed dynamically by using a probability function that is updated after a successful iteration of the algorithm. Specifically, starting with an initial equal probability (0.25) for every heuristic, the algorithm randomly selects one of them in order to designate the first trial point. If the selected heuristic mechanism leads to the improvement of the current population its probability to be selected in the next iteration is then increased:

$$prob_{heuristic} = \frac{\#successes_{heuristic} + c}{\sum_{j=1}^{\#heuristics} (\#successes_j + c)}, \quad (5.7)$$

where the $\#successes_{heuristic}$ is the number of successful improvements currently performed and $c > 0$ is a constant. Therefore, the probability can be simply evaluated by frequency of success. In order to avoid the degeneration of evolutionary process the probability value is reset to its starting value when it decreases below a user provided given lower limit. This procedure is repeated until the convergence criteria are met or a trial point capable of replacing the best in the population cannot be found. Figure 5-1 illustrates the workflow of the CRSwCH algorithm.

5.3.2 Shuffled Complex Evolution (SCE)

The Shuffled Complex Evolution (SCE) algorithm involves the best of the clustering methods and the nature inspired algorithms of systematic and competitive evolution. Particularly it is based on a synthesis of four concepts that have proved successful for GO: (a) combination of probabilistic and deterministic approaches; (b) clustering; (c) systematic evolution of a complex of points spanning the space, in the direction of global improvement; and (d) competitive evolution [193]. The SCE algorithm requires no derivative information and, for this, has been used with great success in several parameter estimation problems [194].

The most important concept of the SCE algorithm is (a) i.e. the simultaneous use of deterministic and random approaches. The former allows the algorithm to make effective use of response surface information to guide the search, while the inclusion of the latter helps to make the algorithm flexible and robust. Next, through the concept of clustering, SCE attempts to ensure that the search is focused on the most promising regions of the parameter space. In addition, the use of a systematic complex evolution strategy ensures that the search is relatively robust and is guided by the structure of the objective function. This allows SCE to be able to cope very well with rough, insensitive, and highly nonconvex objective function surfaces and is relatively unaffected by the small local minima that are encountered enroute to the global solution. Finally, the implementation of the concept of a competitive evolution strategy helps in improving the efficiency of the global convergence.

5.3.2.1 The SCE method

A starting point for the SCE method has been the CRS procedure presented previously. Thereupon, the algorithm begins with randomly sampling a series of points that span the entire feasible input space Ω thus, creating the initial population \mathbf{P} . Choosing a large enough number of points ensures that most of the solution regions with potential local extrema will be

included during the process. For each of the sampled members, the objective function is evaluated and accordingly, the population is sorted in ascending order. After this procedure, the ordered population is partitioned into a finite number of complexes that are allowed to evolve individually, with use of a competitive complex evolution (CCE) algorithm. In general, what the CCE algorithm does is to replace the 'bad' members of a complex (\mathbf{C}), and hence of \mathbf{P} , through a finite amount of evolutions within \mathbf{C} which are controlled by some user-defined input parameters. When the evolution of each \mathbf{C} completes, the old complexes are replaced with the new ones and then, the population is shuffled through resorting of the respective objective function values in ascending order. Then, a new round of evolution begins that continues until the user defined evolution steps are reached or if the convergence criteria have been met.

Regarding the CCE procedure, its first step is to randomly create a subset ($\mathbf{S}: \mathbf{S} \subset \mathbf{C} \subset \mathbf{P}$) of the members of the \mathbf{C} being processed according to a trapezoid distribution. This distribution is defined in such a way so that the best evaluated members of \mathbf{C} have larger proba-

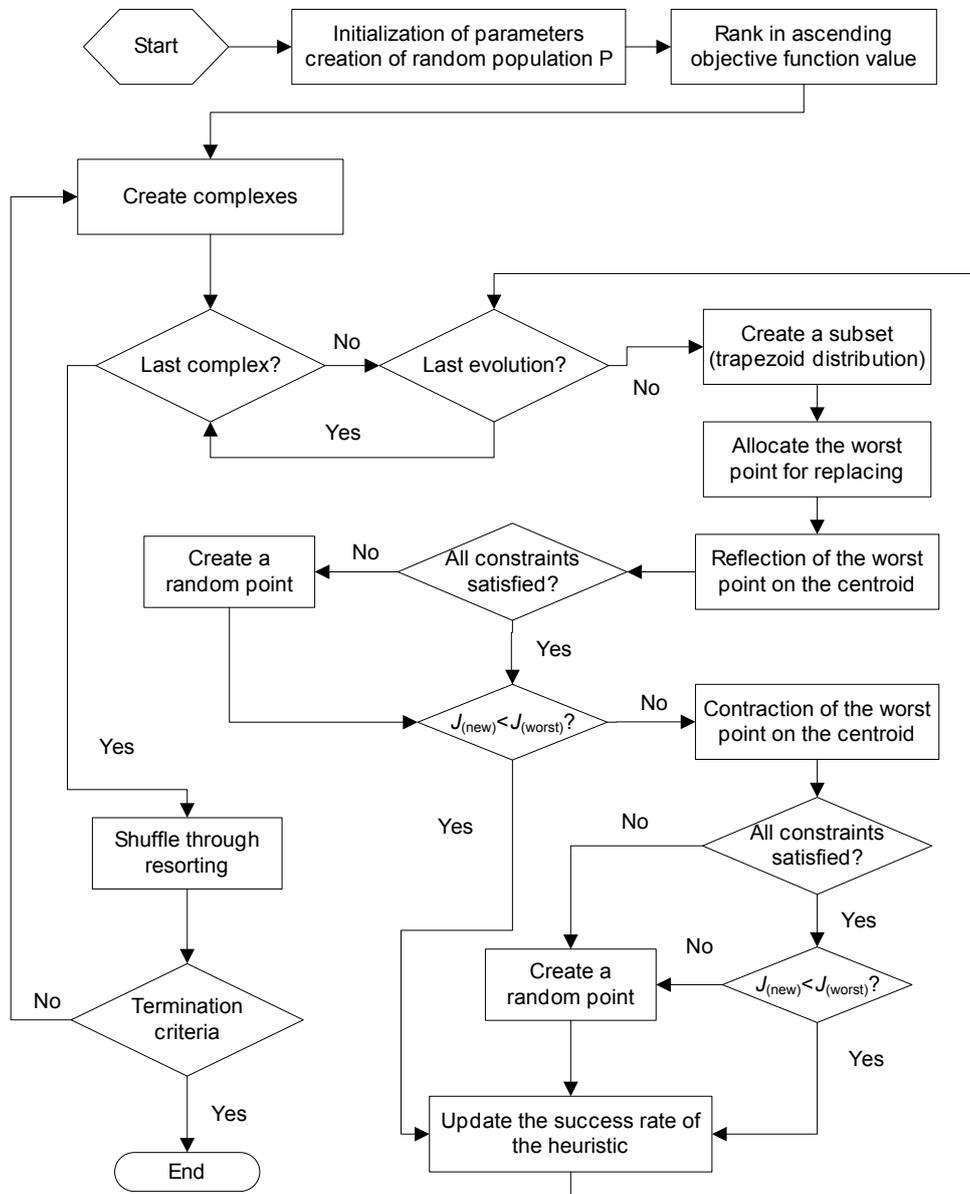


Figure 5-2 Flowchart of the SCE algorithm

bility to be chosen and the worst members have the least probability. Once \mathbf{S} is created, the next step is to identify its worst member ($P^{(worst)}$) and calculate the “centroid” (G) of the rest of its members. Based on this, a reflection of $P^{(worst)}$ against G is attempted and a trial point ($P^{(trial)}$) is generated, similarly to the original CRS algorithm described previously:

$$P^{(trial)} = 2 * G - P^{(worst)}, \quad (5.8)$$

If the resulting $P^{(trial)}$ does not follow the constraints, a random point ($P^{(random)}$) is generated instead (point *mutation*). Then, the objective value of either $P^{(trial)}$ or $P^{(random)}$ is compared against the one of $P^{(worst)}$. If the new objective value is lower, the respective new point replaces the worst one. If not, a point *contraction* is attempted according to which another, new, trial point is generated by calculating the center between the “centroid” and the worst member i.e.:

$$P^{(trial)} = \frac{G + P^{(worst)}}{2}. \quad (5.9)$$

Then again, $P^{(trial)}$ is compared against $P^{(worst)}$ and replaces it only if the former is better than the latter in terms of objective function value. In case this also fails another point mutation is performed and the resulting $P^{(random)}$ replaces $P^{(worst)}$, this time regardless of its objective function value. Such refinement of \mathbf{S} is continuously repeated according to the predefined number of iterations given as input. Once these iterations end, the next step is to replace the refined members of \mathbf{S} back to \mathbf{C} and restart the same procedure for a new \mathbf{S} . Again, the number of \mathbf{S} -related refinements of \mathbf{C} is repeated until it reaches the predefined number provided by an input parameter.

5.3.3 Differential Evolution (DE)

Differential Evolution (DE) is arguably one of the most powerful stochastic real-parameter, population-based algorithms that has attracted a lot of interest over the past 15 years [195]. Primarily, DE belongs to the class of evolutionary methods and also integrates some essentials from the adaptive random search methods like the CRS mentioned above. Particularly, unlike traditional evolutionary algorithms, DE employs difference of the parameter vectors to explore the objective function landscape. In this respect, it owes a lot to its two ancestors namely—the Nelder-Mead algorithm, and the CRS algorithm, which also rely heavily on the difference vectors to perturb the current trial solutions.

The superiority of DE in comparison to other evolutionary strategies relies on several factors. First, compared to most other evolutionary strategies, DE is much more simple and straightforward to implement in fact, its main body takes four to five lines to code. Second, DE exhibits much better gross performance in terms of accuracy, convergence speed, and robustness in comparison with several others established methods. Third, the number of tuning parameters in DE is very few making it easy to devise in order to improve the algorithm’s performance. Fourth, the space complexity of DE is low allowing practitioners to extend DE for handling large scale and expensive optimization problems. For these reasons, the popularity of DE among researchers has been boosting since 1996 when it was first introduced by Storn and Price [196]. Since then, DE has triggered a plethora of research groups to propose several alternatives and hybrids as it is readily evident from the respective bibliography.

5.3.3.1 The core of the DE procedure

In the beginning of the algorithm, a random initial population \mathbf{P} is preferably created by uniformly sampling parameter sets, over their natural bounds. Each member of the population is then evaluated according to the objective function and the best member, in terms of low objective value, is identified. This is the first *generation* of candidate solutions i.e. the first generation of feasible parameter sets which is next optimized through an iterative procedure. In every iteration all the members (candidate solutions) that comprise a generation undergo through three processing stages i.e. *mutation*, *crossover* and *selection*. For these, three randomly chosen vectors participate: the *parent* vector called the *target*, a *mutant* vector known as the *donor* and the *offspring* vector that results from the combination of the donor and the target and is called the *trial* vector.

At mutation stage, the donor vector is created by reflection of three randomly chosen, mutually exclusive and different from the target, parameter vectors, as in equation (5.4). Specifically, the difference of any two of these random vectors is scaled by a scalar F and the scaled result is added to the third one, thus creating the donor vector ($\mathbf{P}^{(donor)}$).

$$\mathbf{P}^{(donor)} = r_1 + F(r_2 - r_3), \quad (5.10)$$

where r_1, r_2, r_3 are the random vectors and F is a real and constant factor $\in [0, 2]$ which controls the amplification of the differential variation ($r_1 - r_3$).

At the crossover stage, in order to enhance the diversity of the next generation, the donor vector exchanges its constituents with the target vector in order to form a trial vector ($\mathbf{P}^{(trial)}$). For this, there are two crossover methods, namely the exponential and the binomial crossover. In exponential crossover, two integers from the interval $[1, k]$ are randomly chosen. The first integer, say L_1 , acts as a starting point in the target vector, from where the crossover or exchange of components with the donor vector starts. The second integer, say L_2 , denotes the number of components the donor vector actually contributes to the target vector. To extract the latter an iterative procedure is employed according to the following pseudo-code:

- 1: initialize, $L_2 = 0$
- 2: increase, $L_2 = L_2 + 1$
- 3: keep increasing L_2 as long as $L_2 \leq k$ and a randomly chosen value in the $[0, 1)$ interval is less than or equal to the crossover rate constant (Cr)

Then, crossover occurs according to:

$$P_i^{(trial)} = \begin{cases} P_i^{(donor)} & \text{for } i = L_1 \bmod k, (L_1 + 1) \bmod k, \dots, (L_1 + L_2 - 1) \bmod k \\ P_i^{(target)} & \text{for all other } i \in [1, k] \end{cases}, \quad (5.11)$$

where $-\bmod-$ denotes the modulo operation. In binomial crossover, the same procedure is followed according to which the trial point is acquired by

$$P_i^{(trial)} = \begin{cases} P_i^{(donor)} & \text{if } rand_{(0,1)}^{(i)} \leq Cr \quad \text{or} \quad i = l \\ P_i^{(target)} & \text{if } rand_{(0,1)}^{(i)} > Cr \quad \text{and} \quad i \neq l. \end{cases} \quad (5.12)(11)$$

At selection stage, in order to select if the target or the trial survives to the next generation, we compare the objective values of both vectors. Therefore, the trial replaces the target

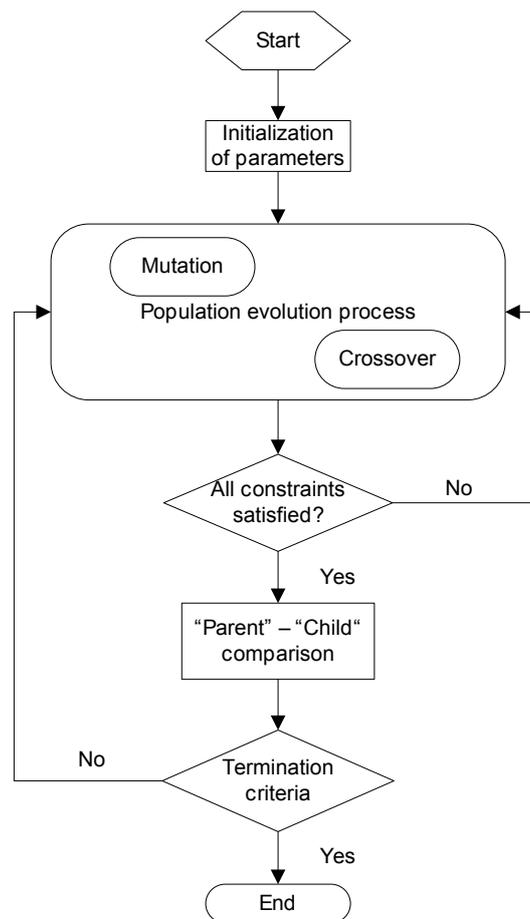


Figure 5-3 Flowchart of the DE method

only if its objective is lower using the greedy criterion. Finally, these procedures continue iteratively until the maximum of generations are created or the fitness criteria are met.

5.3.3.2 Variants of DE

In order to classify the plethora of DE variants, the notation: $DE/\zeta_1/\zeta_2/\zeta_3$ has been introduced, where $\zeta_1, \zeta_2, \zeta_3$ are strings that designate:

ζ_1 : specifies the vector to be mutated, which can be “rand” (a randomly chosen population vector), “best” (the vector of lowest cost from the current population) and target-to-best (all the vectors are attracted toward the same best position found so far by the entire population, thereby converging faster toward the same point).

ζ_2 : is the number of difference vectors used.

ζ_3 : denotes the crossover scheme i.e. “exp” when exponential and “bin” when binomial experiments are used at the crossover stage. We should note here that several other schemes have been proposed in literature. For this, the interested reader is referred to [195].

5.3.4 Simulated Annealing (SA)

Generally, the Simulated Annealing (SA) techniques are powerful evolutionary optimization procedures successfully used mainly for combinatorial optimization but also for continuous ones. Algorithms based on SA employ a stochastic generation of solution vectors by imitating the physical process of crystallizing a solid into a minimum free energy state (annealing) by heating and then by slowly cooling it. During the cooling process transitions are

accepted to occur from a low to a high energy level through a Boltzmann probability distribution (Metropolis criterion). This allows SA methods to be randomized towards locating near optimum solutions, at the expense, however, of increased computation burden and low rigourousness [197].

Based on this procedure the general SA optimization algorithm begins by defining the annealing schedule. After the formulation of the objective function, this includes: (i) defining an initial temperature, which is calculated by random sampling a finite number of feasible input coordinates from a uniform distribution; (ii) setting the number of candidate solutions (configurations) that will be stochastically generated at each temperature level; (iii) choosing the method to decrease the temperature; (iv) fixing the acceptance/rejection criteria of a new configuration at each temperature level; and (v) designating the termination criteria.

Therefore, SA procedures include a two-loop structure whereby the inner loop controls the alternative configurations and the outer one the temperature decrease. Theoretically, if the inner loop is allowed to continue indefinitely, at some point the system will reach equilibrium and hence global convergence. Throughout this way, a time-homogeneous Markov chain will have then been generated which asymptotically reaches equilibrium under ergodicity conditions. In practice though only a finite number of configurations are generated at each temperature level. Hence, global convergence may not be always reached.

5.3.4.1 Thermodynamic, adaptive SA (tdASA)

It is common for the optimization algorithms for nonconvex NLP problems to lead to sub-optimal solutions as non-convexities may cut off the global optimum from the current search region. Adaptive random search methods therefore have been traditionally drawing particular interest especially for NLP and MINLP problems. On such grounds, several SA variants and SA hybrid algorithms have been developed to tackle these demanding problems. Such is the thermodynamic, adaptive SA (tdASA) in which the temperature is free to evolve and its value is continuously updated from the variation of state functions as the internal energy and entropy [198], [199]. On such basis a tdASA algorithm has been employed in which the temperature schedule control and random step selection are automatically adjusted according to the algorithm's progression. While a SA algorithm is traditionally built around a single initial temperature it is easily adjusted into a population-based method. For time efficiency purposes, therefore, a population scheme has been also adopted so that during an iteration of the algorithm, several solution candidates can be perturbed.

First, let a random population of solution candidates, called *system configurations*, has been randomly generated. Then, each configuration is perturbed by employing the following Boltzmann generating scheme:

$$P^{(trial)} = P^{(prev)} + \sqrt{T} \times rand_{[0,1]}, \quad (5.13)$$

where $P^{(trial)}$ is the new (trial) configuration, $P^{(prev)}$ is the previous configuration and T is the current temperature. If the resulting $P^{(trial)}$ is infeasible according to the parameter value constraints, a random configuration is created instead. Next, the objective function value, called *energy state*, of the new individual is evaluated and compared against the one of $P^{(prev)}$. If the new configuration possesses a lower energy it is readily accepted and replaces the previous one. If the energy is higher, then the Metropolis criterion is employed as in all SA algorithms. This criterion controls the acceptability of the randomly created trial configurations according to their energy states. Particularly, a system configuration is accepted if its

energy state is higher than the existing one. Otherwise, and by analogy with the Boltzmann distribution for energy states at thermal equilibrium, an energy state will be accepted as an improved state according to the Boltzmann probability of its difference with the resting energy state:

$$P(\Delta E) = \exp\left(-\frac{E^{j+1} - E^j}{K_B T^j}\right), \quad (5.14)$$

where E^j and E^{j+1} are the resting and trial energy state, respectively, T^j is the current temperature level and K_B is the Boltzmann's constant. Note here that at high temperatures this probability is close to one i.e. most energy transitions are permissible and as the system temperature decreases the probability of equation (5.14) approaches zero. Thereupon, prior to initiating the tdASA algorithm the temperature is set by the user to a very high value.

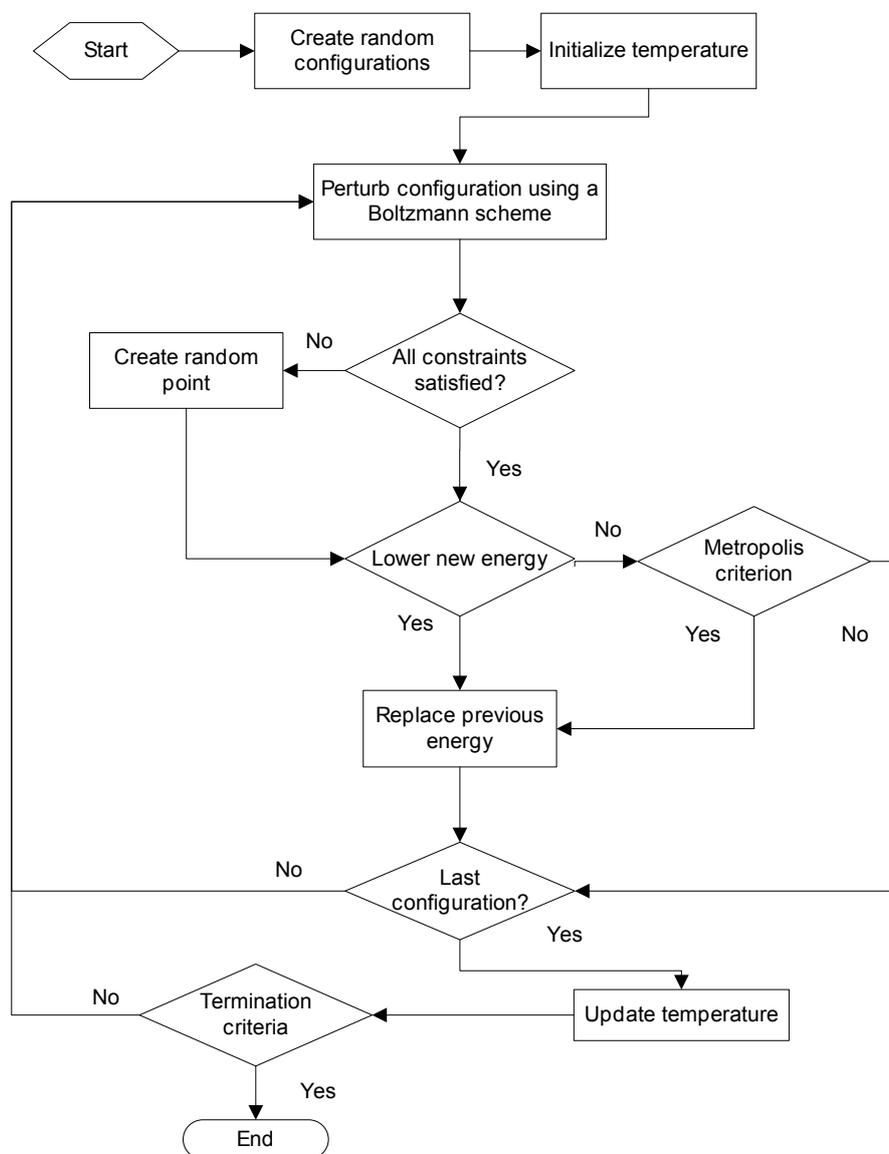


Figure 5-4 Flowchart of the tdASA method

Next, the cooling schedule that has been employed is an exponentially decreasing function of the annealing time t , defined according to the system:

$$T^{j+1} = T^j e^{-c_1/k}, \quad (5.15)$$

$$c = c_1 e^{-c_2/k}, \quad (5.16)$$

$$c_1 = -\ln\left[0.5(10^{-5} + E_{best})\right], \quad (5.17)$$

$$c_2 = -\ln\left[0.5(k + 2, E_{worst})\right], \quad (5.18)$$

where E_{best} and E_{worst} denote the energy states of the best and the worst configurations. As the algorithm progresses and E_{best} improves approaching 0, the annealing procedure slows down. As said the estimations are made based on the laws of thermodynamics (equation (5.13)). This means that the current temperature of the algorithm has an active role to the estimation made. For example, when a good estimation is produced we would like the algorithm to remain at the current temperature for a longer time. Then, according to the relation in (5.17), when a better configuration is generated, the value of c_1 decreases. Subsequently the value of c decreases and then the temperature annealing slows down through (5.15). Conversely, when a bad configuration is produced we would like the algorithm to anneal faster. Then, according to (5.18), when the estimations worsen, the value of c_2 increases. Subsequently the value of c increases and then the temperature annealing runs faster through (5.15).

The algorithm proceeds with reducing the temperature by use of this cooling schedule. Once the procedure is concluded for all of the new configurations, the termination criteria are evaluated. If they are met, the algorithm terminates offering the best configuration as a solution; otherwise it proceeds with scheduling a new iteration.

5.3.5 Particle Swarm Optimization (PSO)

PSO is an algorithm based on a metaphor of social interaction that has been extensively employed in solving MINLP problems [187]. It is a stochastic population-based optimization technique developed by Eberhart and Kennedy in 1995 [200]. In general, the PSO algorithm searches the input space by adjusting the trajectories of the trial parameter sets, called “particles”, as they “move” in a multidimensional fashion. Particle movement is drawn toward the positions of their own previous best performance and the best previous performance of their neighbors in the swarm. This social experience obtained by social interaction with the neighboring particles allows PSO to evidently share similarities with evolutionary algorithms. However, contrary to the principles of evolutionary optimization, PSO does not employ selection, crossover or mutation operations. Rather, the main driving force of the swarm is the social interaction implicitly encoded in the social network structure.

Empirical evidence has accumulated that the PSO algorithm is a useful tool for optimization [201]. In addition, PSO is simple in concept having few parameters to adjust and being easy to implement. Over the past several years, PSO had been successfully applied in many research and application areas where it has been demonstrated that PSO can obtain better results in a faster and cheaper way compared with other stochastic methods [202].

5.3.5.1 The PSO strategy

According to the swarm adaptation strategy, a swarm of particles (population of parameter sets) can better itself by stochastically following the individual with the best previous

performance found and the individuals with best previous performances in their neighborhoods. At the start of the algorithm, and if not provided, a randomly positioned population of particles \mathbf{P} is created. The members \mathbf{P} are evaluated by the objective function and the best member of the population is identified. Next, a velocity is attributed to each of the positional coordinates of each particle according to two sources of information. The first one, called cognitive component, reflects the experimental knowledge of the particle, which is the member's best previous position and the best position in the \mathbf{P} so far. The second one, called social component, reflects the local knowledge of the search space obtained from the particle's neighborhood. More specifically, the velocity of each dimension of a particle evolves through a, randomly weighted, sum of differences i.e. the difference between the best point found so far ($P^{(own\ best)}$) and the individual's current position; and the difference between the best point in the neighborhood of each individual ($P^{(best\ neighbor)}$) and the individual's current position. This can be depicted as:

$$P^{j+1} = P^j + u^{j+1}, \quad (5.19)$$

where $j, j+1$ are the current and next movement instance and u the velocity attributed to each of the positional coordinate where

$$u^{j+1} = \kappa_1 u^j + \kappa_2 \times rand_{[0,1]}(P^{(own\ best),j} - P^j) + \kappa_3 \times rand_{[0,1]}(P^{(best\ neighbor),j} - P^j), \quad (5.20)$$

where κ_1 is the inertia weight employed to control the impact of the previous history of velocities on the current velocity, thus to influence the trade-off between global (wide-ranging) and local (nearby) exploration abilities of the particles and κ_2 and κ_3 are the cognitive and social parameters of the PSO algorithm. Then the members are moved according to the calculated velocity, while making certain that no constraints are violated, and the new positions are

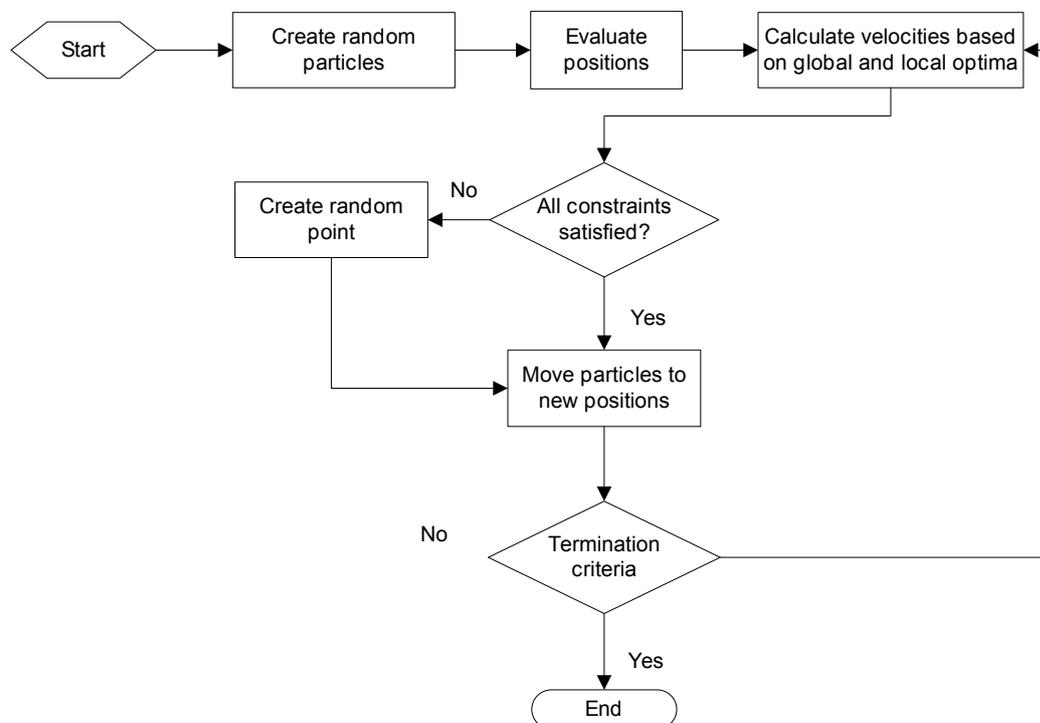


Figure 5-5 Flowchart of the PSO method

evaluated once more.

Equation (5.20) shows that the PSO algorithm can enforce a steady improvement in solution quality, accepting only moves that improve fitness. This allows the oscillation of the particle's trajectory around potentially better particles. In time, the swarm is expected to converge to the area where the best solution candidate particle exists [202]. Finally, if any termination criteria is met the algorithm halts otherwise a new iteration begins anew by calculating each members new velocity.

5.3.6 Mixed Integer Distributed Ant Colony Optimization (MIDACO)

The MIDACO GO algorithm is a black-box point-to-point optimizer which has managed to outperform many established MINLP solvers [203]. It is based on a mixed integer extension of the Ant Colony Optimization (ACO) metaheuristic proposed by [204]. Originally the ACO metaheuristic was considered only for combinatorial optimization problems. Nevertheless, several extensions of the ACO metaheuristic for continuous search domains can also be found in the literature. However, extensions for mixed integer search domains are very rare. The MIDACO algorithm included in this study is a copyrighted algorithm, distributed freely only for academic purposes.

In general, the ACO methods mimic the foraging behavior of biological ants whose goal, in real life, is to find the shortest path possible to a source of food. If an ant succeeds in finding that source then, it will return back to the nest. Throughout its return, it lays down a chemical pheromone that will attract other ants to follow it. If the path is regularly updated with pheromone then it will remain attractive, otherwise it will evaporate. In time, the shortest paths are expected to survive.

The original ACO methodologies conventionally employ pheromone tables in order to obtain the optimal path in terms of length or some other cost-resource which is requested by the problem. These tables are completely connected Graphs $Gr(vert, conn)$, where $vert$ is the set of vertices of solution components, which every ant chooses incrementally to create a path, and $conn$ is the set of connections in a path. Then, probabilistic decisions are made for the ants according to the pheromone values within the pheromone table. Keeping an updated table of pheromone values from the information gained on the search domain, an ACO metaheuristic may potentially provide very good and hopefully global optimal solutions. However, when one wants to optimize a mixed integer problem, the biological visualization of ants choosing their way through a graph-like search domain does not hold any longer. Hence, the MIDACO algorithm being an extended version of this original ACO metaheuristic attempts, successfully, to surpass this limitation without any major conceptual change.

5.3.6.1 Details of the MIDACO procedure

In contrast to the original ACO methodologies which conventionally employ pheromone tables, MIDACO uses a pheromone guided discretized probability density function for the discrete search domain [205]. This approach allows an intuitive handling of integer variables besides continuous ones within the same algorithm framework. Then, based on a continuous ACO methodology the optimizer is extended in order to handle discrete variables. This is done by computing a Gaussian probability density function (PDF) as a weighted sum of several one-dimensional Gaussian functions. Consequently, candidate solutions are incrementally constructed with regard to multi-kernel Gaussian probability density functions which are created for every dimension of the search space.

In particular, it is principally accepted that any function P with the property

$$\int_{-\infty}^{+\infty} P(x) dx = 1, \quad (5.21)$$

can act as a PDF. Then, since the Gaussian function has some clear advantages like ease of implementation and fast sampling times, it is used for every dimension k of the original search domain:

$$P^k(x) = \sum_{l=1}^{krn} w_l^k \frac{1}{\sigma_l^k \sqrt{2\pi}} e^{-\frac{(x-\mu_l^k)^2}{2(\sigma_l^k)^2}}, \quad (5.22)$$

where the triplet $(w, \sigma, \mu)^k$ characterizes the PDF of every dimension k and krn is the number of kernels i.e. the number of Gauss functions used within $P^k(x)$. Within this context, w is a weight factor, σ is the kernel's standard deviation and μ the kernel's mean value. These triplets, since they fully characterize the PDF and, in turn, guide the sampled solution candidates, they are called pheromones within the MIDACO sense.

In order to update the pheromone values a solution archive (SAr) of size krn is kept that saves the best solution candidates in terms of the triplet $(w, \sigma, \mu)^k$. In particular, the weights w which designate the importance of an ant are calculated according to:

$$w_l^k = \frac{krn - l + 1}{\sum_{j=1}^{krn} j}. \quad (5.23)$$

With this fixed distribution of the weights, the SAr is sorted in ascending order. This means that the solutions with a low index refer to the best solution found and vice versa. Then, every time a new ant (solution) is created within a generation its attraction to the solutions stored in SAr is evaluated starting from the best solution. If the attraction of the new ant is found to be better than, say the j -th, one, then the new ant's attraction is stored in the j -th position in the SAr while the rest of the attractions are shifted one place. Effectively, the worst attraction will be rejected from the SAr i.e. evaporate. If the new solution has no better attraction than any of the already ones saved in the SAr, the latter remains unaffected.

Updating the SAr with this procedure leads automatically to a positive pheromone update. Note that a negative pheromone update (evaporation) is indirectly performed when dropping the worst solution of the SAr. Thereupon, the σ is calculated according to the solutions stored in the SAr as:

$$\sigma_l^k = \frac{dis_{\max}(k) - dis_{\min}(k)}{\#generation}, \quad (5.24)$$

where dis_{\max} and dis_{\min} denote, respectively, the maximal and minimal absolute distance of the krn solutions which is extracted by calculating all the differences between the solution values of a single component. On the other hand, μ values are given directly from the solution values of a single component i.e.

$$\mu_l^k = SAr_l^k. \quad (5.25)$$

Therefore, a new ant is created from the following procedure. First, a mean value (μ_{new}) is randomly selected between the ones stored in the SAR for all the components of the new ant. This is performed in respect to the weight of each solution i.e. the solution with the highest rank also has the highest probability to be selected. Then, second, using the μ_{new} and the deviation σ for all the components defined by equation (5.24) k new values are sampled. Since, these values comprise potential solutions to the problem; these are evaluated according to the objective function.

In order to maintain the same framework as described above a discretization of the continuous random numbers (sampled by the PDF) is performed in order to handle integer variables. However, there is a chance for an integer dimension k to share the same value in SAR. If this happens, the deviation will be zero and then no further progress in the integer component will be possible. For this reason, and only for the integer parameter case equation (5.24) is slightly altered as follows:

$$\sigma_i^k = \max \left\{ \frac{dis_{\max}(k) - dis_{\min}(k)}{\#generation}, \frac{1}{\#generation}, \frac{\sqrt{\#int} - 1}{2\sqrt{\#int}} \right\}, \quad (25)$$

where $\#int$ denotes that number of integer parameters.

We note here that the SAR also contains the corresponding violation of the constraints and the penalty function value for every solution. However, since there are no constraints other than the parameter value bounds in our problem we will not describe the procedure

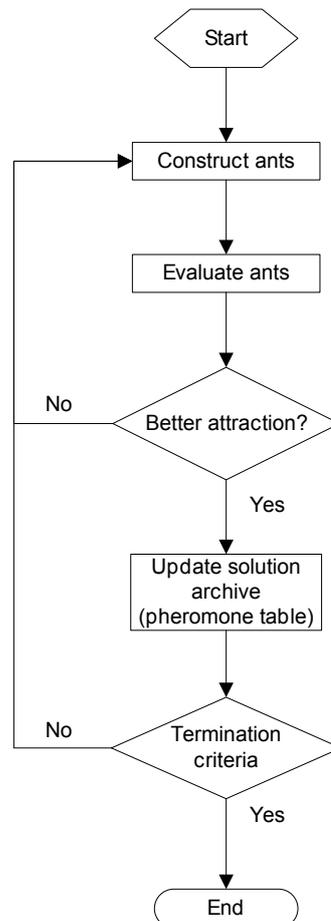


Figure 5-6 Flowchart of the MIDACO method

here, for brevity. Figure 5-6 illustrates a crude flow diagram of the MIDACO optimizer.

5.3.7 Evolutionary Strategy with Stochastic Ranking (SRES)

Generally speaking, there is an increasing amount of literature which strongly indicates that pure genetic algorithms are inferior to evolutionary strategies, especially for continuous problems. In fact, in [186] after comparing a number of stochastic methods, evolutionary programming was found to be the best performing algorithm for solving the inverse problem of a dynamic metabolic pathway. Under these circumstances the SRES algorithm is one of the most competitive evolutionary, GO methods available according to [186]. For this reason, it has been employed in parameter estimation problems in biochemistry with great success.

The main advantage of the SRES is that it has the ability to adjust the balance between the objective and penalty functions automatically during the evolutionary search. Penalty functions are real-valued functions commonly added as a weighted sum to the objective function in order to penalize constraint violations of the respective trial points. The influence of each penalty function is controlled by a penalty coefficient. Although the penalty function method may work quite well for some problems; the decision for an optimal (or near-optimal) value for these coefficients turns out to be a difficult optimization problem itself.

In a general sense, all a penalty method tries to do is to obtain the right balance between objective and penalty functions so that the search moves toward the optimum in the feasible space, not just toward the optimum in the combined feasible and infeasible space [206]. The most promising approach to tackle different constrained optimization problems has been to adjust the penalty coefficients dynamically and automatically by an evolutionary algorithm itself. This is exactly what the SRES algorithm performs i.e. to adjust directly and explicitly the balance between objective-dominated and penalty-dominated function optimization search. To do this, a (κ, λ) evolution strategy and a stochastic bubble-sort algorithm are utilized for ranking individuals according to their objective and penalty values. The procedure is as follows.

5.3.7.1 The procedure of the SRES algorithm

In the beginning of the method λ random vectors of parameters are selected according to a uniform probability distribution over the entire search space. Then, the initial “mean step size” ($st^{(0)}$) is calculated for every vector as the weighted distance:

$$st^{(0)} = (p_{\max} - p_{\min}) / \sqrt{k} . \quad (5.26)$$

Each distance is then stochastically ranked by bubble sort.

As said, the penalty coefficient in a penalty function is hard to determine and requires *a priori* knowledge about a problem. For this reason, in SRES the ranking procedure utilizes the idea of dominance according to which the strength of each individual is first quantified by probabilistically balancing between the objective and penalty function values and then, ranked accordingly. The initial ranking is always created at random. But afterwards, the bubble sort swaps between individuals are controlled by the penalty function values of adjacent individuals; by the objective function values of individuals; and by a probability that is uniformly sampled between 0 and 1 at each time. That is, given any pair of two adjacent individuals, the probability of comparing them (in order to determine which one is fitter) according to the objective function is 1 if both individuals are feasible; otherwise, it is equal to the sampled probability. Figure 5-7 briefly illustrates this notion.

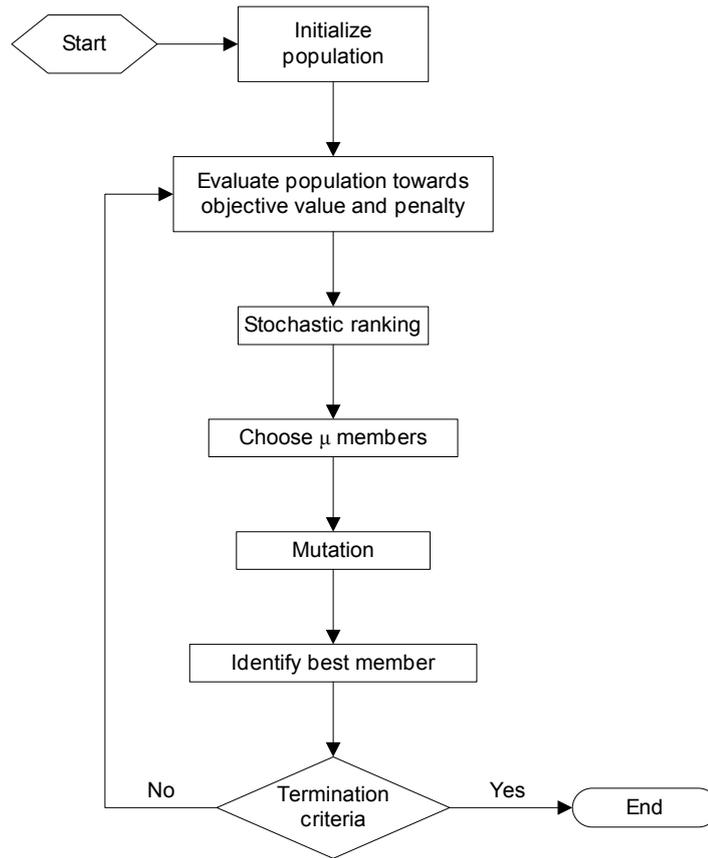


Figure 5-7 Flowchart of the SRES method

Next, κ individuals are selected for generating λ new vectors (strategy parameters) of the next generation (offsprings). The truncation level is set at $\kappa/\lambda \approx 1/7$. Generation of the offsprings is performed according to a log-normal, self-adaptive update rule (mutation stage) using the “mean step sizes”:

$$st_{\kappa,k}^{(gen+1)} = st_{\kappa,k}^{(gen)} \exp\left(\tau \cdot rand_{[0,1]} + \tau \cdot rand_{[0,1]}^k\right), \quad (5.27)$$

where gen denotes the current generation, and τ, τ' are called “learning rates” and are equal to $1/\sqrt{(2\sqrt{k})}$ and $1/\sqrt{(2k)}$, respectively. The superscript k in $rand_{[0,1]}^k$ indicates that the random number is generated anew for each value of the k parameters. Prior to mutation stage, a global intermediate recombination stage between two parents is implemented as:

$$st_{\lambda,k}^{(gen)} = \left(st_{\kappa,k}^{(gen)} + st_{\kappa',k}^{(gen)}\right)/2, \quad (5.28)$$

where κ' is selected at random from the κ individuals ($\kappa' \neq \kappa$).

When an offspring is generated outside the bound constraints, the mutation of the objective variable will be retried until the variable is within nominal bounds. In any other case, the object variable is left in its original state within the parameter bounds. Finally, these algorithmic procedures are halted when the predefined number of generations is reached.

5.3.8 DIRECT

The DIRECT, is an algorithm for finding the global minimum of a multivariate function subject to simple bounds, using no derivative information. It belongs to the deterministic,

point-to-point class of GO algorithms. It is a modification of the standard Lipschitzian approach without the need specify a Lipschitz constant. Roughly speaking, Lipschitzian approaches attempt to place a lower bound on the objective function in any closed interval whose endpoints have been evaluated. For this reason, the value of the Lipschitz constant is usually large because it must be equal to or exceed the maximum rate of change of the objective function. As a result, these methods place a high emphasis on global search, which leads to slow convergence. In DIRECT this constant is viewed as a simple weighting parameter that indicates how much emphasis to place on global versus local search. Hence instead of using a constant value, exhaustive simultaneous searches are carried out using all possible values of constants from zero to infinity. Once the global part of the algorithm finds the basin of convergence of the optimum, the local part of the algorithm quickly and automatically exploits it. The clear advantage of such adoption is that it accounts for the fast convergence of the algorithm. For this reason, DIRECT is one of the core optimizers of a well-known and well-established optimization suite for research and teaching in optimization and has been employed to provide solution to many challenging problems [207].

5.3.8.1 The DIRECT procedure and its variant, the gblSolve

The DIRECT algorithm, according to its authors is guaranteed to converge to the global optimal function value if the objective function is continuous or at least continuous in the neighborhood of a global optimum. The heart of this algorithm bases on the idea of iteratively dividing the k -dimensional parameter space in smaller parts which are then, compared to one another.

At the beginning of the procedure, the initial search space is transformed to that of a

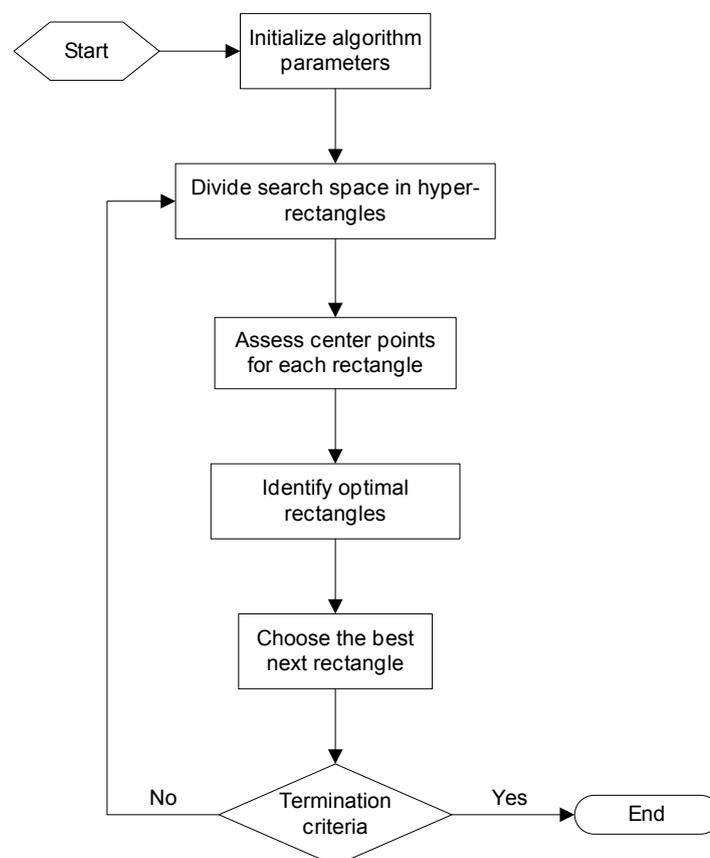


Figure 5-8 Flowchart of the gblSolve procedure

unit hypercube. Then, the algorithm initiates with evaluating the objective function at the center of the cube instead of doing it at the vertices as standard Lipschitzian approaches would do. This is an advantage especially when dealing with problems in higher dimensions which cannot ensure good performance in such spaces. Then, the hypercube is divided into smaller sub-hyper-rectangles from the center of which samples are drawn, anew. Instead of using a Lipschitz constant for determining which cube to sample next, DIRECT chooses each time a set of the best potential optimal sub-hypercubes i.e. the ones with the lowest objective value. For this choice, an algorithm known as Graham's scan is employed in order to find the optimal convex hull [208]. This procedure is performed iteratively until the convergence criteria are met or the maximum number of iterations has been reached.

For our purposes, we have utilized a variant of the original DIRECT algorithm, known as *gblSolve*. The main difference between these two is that the latter, instead of using a tree structure for storing the information of each rectangle, straightforward matrix/index-vector technique is used. In effect, *gblSolve* has been shown to have slightly better results than the original DIRECT [207].

5.4 PERFORMING OF THE PARAMETER ESTIMATION

Up to now, we have argued that the proposed inverse problem may be, in theory, solvable if the parameter complexity is reduced. Still, our capability to predict the model parameters within adequate accuracy and reproducibility remains to be challenged in practice. For this purpose, we stated in the beginning of this chapter the parameter estimation problem as the minimization of the function of the error between the experimental and model responses. Afterwards, it has been concluded that solution to this minimization problem can only be provided by derivative-free, GO algorithms. Since none of these methods is able to posit solution optimality, we have selected a representative set of algorithms giving particular attention to their diversity in terms of ensuring that the minimization is handled through different systematic patterns. Then, in order to infer whether the solution of the proposed inverse problem converges with adequacy and precision, a systematic evaluation of its solvability has to be specified. The latter is described next.

5.4.1 The *in silico* experiment setup

First and foremost, it is essential to collect a representative ensemble of experimental responses whose corresponding physiological parameter values are known in advance. The feasibility of this, however, is limited for most model parameters because the *in situ* characteristics of living tissues are significantly modified or become obsolete during their assessment by traditional sampling and analysis procedures. Consequently, we have generated 30 pseudoexperimental (PEX) data responses by simulation from a known range of parameter values. In order to represent all the clinical stages of neoplasia and concurrently capture the full spectrum of the model's output space; these data were selected as follows. First, ten PEX responses were empirically adopted based on the features of the experimental dynamic optical data. Next, another ten PEX curves were developed from an equidistributed, space filling sequence of ten parameter value sets. The rest of the PEX curves were randomly created from a stochastic sample, uniformly distributed within the domain of the parameter values. Then, accepting that the model's structure is accurate and complete, this ensemble of parameter sets were considered as nominal representatives of the, *in vivo*, functional and structural parameters of the abnormal tissue and the corresponding PEX curves as the true dynamic experimental responses measured after the application of the biomarker.

To better assess whether the convergence of the inverse problem to unique solutions practically increases by reducing the parameter dimensionality, this procedure has been followed in order to create two PEX data ensembles, one for the full and one for the reduced set of parameters. Of course, all of these simulated data represent exact results, that is, devoid of measurement noise. While this may seem somewhat artificial, the omission of noise was intentional in order to see how the methods perform under ideal conditions and because a preparatory smoothing step of the actual experimental signals will be performed before the fitting that significantly reduces if not eliminates noise [209].

We should stress, once more, that the number of layers defines the number of differential equations that are integrated in each simulation. Hence, if the value of this parameter is different then, the respective PEX curve will be the result of the simulation of a separate equation system. From a mathematical perspective this means that the signals included in the PEX dataset are realizations of separate mathematical models. It follows, that the estimation of the parameters of each PEX signal is by itself a separate optimization problem. In other words, our model is an integration of a subset of tissue models the number of which depends on the number of cell layers employed. Therefore, the optimization problem discussed here not only refers to the estimation of a set of biological parameters but also to the finding of the most suitable mathematical model that explains the observed data.

5.4.2 Measuring the performance of the GO methods and the overall problem convergence

It is generally accepted that the performance of a GO algorithm can be defined by two measures. The first measure regards the quality of the inverse solution and is naturally expressed by the cost function or the estimation error when the parameter values are known beforehand. The second measure regards the algorithm's speed, time or search effort, and is expressed by the number of fitness evaluations, CPU time, wall-clock time, etc. Then, in most, if not in all, GO problems the performance metrics used are based on variations and combinations of these two measures. In our case, the parameter values of each of the fitted signals are known *a priori* because they have been artificially generated by simulations of the model's differential system. This means that beside the cost function value, we can measure the error in the parameter estimations as well. Therefore, given a maximum running time (in minutes), algorithm performance can be expressed as the best fitness value in terms of Euclidean distance at termination and/or as the smallest estimation error in terms of accuracy/reproducibility at termination.

Calculation of the performance in terms of Euclidean distance between PEX and estimated datasets is straightforward using the cost function values. On the other hand, the accuracy/reproducibility of model predicted parameter values can be assessed by comparing them with the known PEX values. Obviously, for the deterministic GO algorithms (DIRECT) a single fit is sufficient in order to define this difference. In contrast, for the stochastic group of algorithms, multiple fits on the same signal are necessary in order to get a good approximation of the estimation errors. Specifically, for each of the 30 PEX curves and for each of the 8 algorithms the parameters' normalized root mean squared deviation (NRMSD) has been employed as an accuracy/reproducibility statistic:

$$NRMSD(p)_{al}^{PEX} = \frac{RMSD(p)_{al}^{PEX}}{(p_{\max} - p_{\min})} = \frac{\sqrt{\sum_{i=1}^{rn} (p_{i,al}^{PEX} - \hat{p}_{al}^{PEX})^2}}{rn \cdot (p_{\max} - p_{\min})}, \quad (5.29)$$

where p_{max} , p_{min} denote the maximum and minimum parameter values.

From this statistic, the performance of each GO algorithm can be readily obtained by aggregating over the number of parameters and PEX datasets. A common approach for this is to calculate the mean over all parameters in order to define the average accuracy/reproducibility of the predicted parameter values and then, the mean over all PEX signals. However, the mean as a statistic is prone to outliers, especially when observing large variances. Therefore, in order to quantify the performance of each GO method we shall perform two separate aggregations.

Aggregation over all parameter deviations will be based on a weighted average because the parameters do not influence equally the model's output. In fact, according to the FOS index calculated and presented in our GSA results the sensitivity of the model's output to each input parameter varies considerably. Furthermore, we have indicated in paragraph 4.4.2.3 that the higher the first order effect of a parameter is; the more are the chances of it being estimated when observations of the output become available. In other words, when fitting of the model's output is intended the estimation error is expected to be large if the FOS index of the respective parameter is small and vice versa. This inherent issue implies that the mean over the estimated parameter NRMSDs will not be a fair indication of the performance of an algorithm for a PEX signal fit. Therefore, the accuracy according to which a GO algorithm estimates an input parameter set, for any curve fit, has been calculated as the weighted average:

$$error_{al}^{PEX} = \sum_{i=1}^k \left\{ wg_i \times NRMSD(p_i)_{al}^{PEX} \right\}, \quad (5.30)$$

where $wg_i \in (0,1)$ is the weight value and $NRMSD(p_i)$ refers to the value of the calculated NRMSD for each parameter. To account for the wg 's we have utilized the mean FOS value of each parameter because it quantitatively depicts which is the amount of the output variance that is attributed to each individual parameter. In our case, the sum of all wg 's for both the full and the reduced parameter set has been found to approximate the value of 0.92 (Figure 5-9). This means that 92% of the output variance is attributed to the parameters alone while the rest 8% to interactions between them. Reasonably therefore, the parameter estimation accuracy for each curve is weighted according to the corresponding parameter's average first order sensitivity value.

Let now, the aggregation over all of the calculated errors with which a GO algorithm estimates the input parameters of each PEX signal denote the method's performance. In addition, let each of these errors be a random variable. Then, according to the Central Limit Theorem the distribution of the mean of these accuracies:

$$\mu_{al} = \frac{1}{30} \sum_{PEX=1}^{30} error_{al}^{PEX}, \quad (5.31)$$

will be approximately normal provided that the sampled PEX dataset is sufficiently large. Empirically, if the sample size is greater than 30 the Central Limit Theorem approximation is adequate [50]. Thereupon, since the generated PEX dataset comprises of 30 model outputs which aspire to capture the full spectrum of the possible output signals that can be simulated, the performance will approximately, be:

$$Performance_{at} \sim N(\mu_{at}, s_{at}^2), \quad (5.32)$$

where, s^2 is the variance of the calculated accuracies.

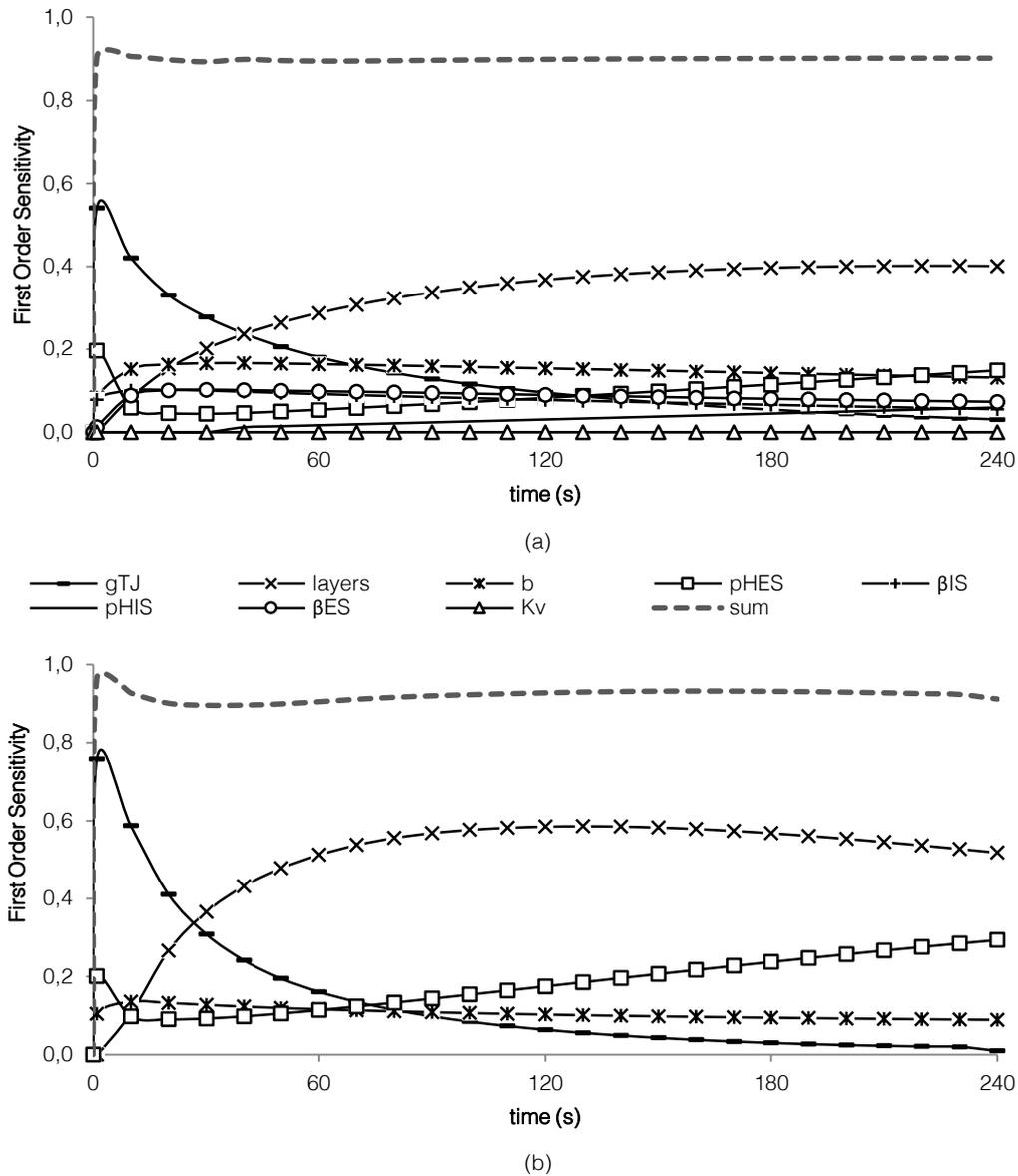


Figure 5-9 The FOS functions of the full parameter set (a) and of the four most influential parameters (b) after the restructuring described in paragraph 4.6. The dashed, bold line indicates their sum. Clearly, more than 90% of the output's variance is explained by the parameters in both cases.

Even if normality has been assumed, doubts may appear in case the population of accuracies contains outliers. As indicated above each PEX curve formulates a separate optimization problem. Hence, deviations from normality are very likely to occur in the results. In effect, the confidence in the population of the results of each algorithm won't be accurate if it is extracted from the respective variance. In this case, to define the confidence intervals, nonparametric bootstrap can be performed. In general, nonparametric bootstrap is a procedure that allows one to define the confidence of the mean of a sample when the distribution from which it came from is unknown. This is done by sampling with replacement from the data itself. Particularly, given the original sample of 30 accuracies for a GO method, a large number of random, new samples (>1000) of size 30 are drawn, with replacement, from this

original sample. Sampling with replacement means that each of the 30 accuracies can be sampled several times in each new sample. Then, the variance of the means of these new samples, called the bootstraps, will approximate the variance of the algorithm's performance.

After quantifying the performance of each GO algorithm, we can say that the solution to the inverse problem converges to unique solutions if these performances are statistically under a certain threshold (e.g. 5 or 10%). This can be evaluated with a statistical test (z-test, Student's t, etc.) for every algorithm. The hypothesis H_0 for the test is as follows: "*is the algorithm's mean estimation error (μ_{al}) lower than a threshold (e.g. 5 or 10%)?*". Furthermore, using μ_{al} we can define the best working GO method through statistical significance testing between the optimization methods. Specifically, using the post-hoc multiple comparison Holm-Sidak test [210] we can rank, in ascending μ_{al} value order, the optimization methods and assign a score sc_i for $i = 0, \dots, n_{GO}-1$, where n_{GO} is the number of GO methods and i is the appropriate rank ($i = 0$ corresponds to the best ranked method). Then, we calculate the values:

$$z_i = \frac{sc_0 - sc_i}{\sqrt{\frac{n_{GO}(n_{GO} + 1)}{6n_{runs}}}}, \quad (5.33)$$

where n_{runs} is the number of considered PEX test problems for each method. Finally, we calculate the cumulative normal distribution values p_i corresponding to z_i and compare them with the corresponding α/i values where α is the significance level, set to 0.05 in our case. We report the z_i , p_i , and α/i values in a table, where each row corresponds to one of the methods. The best ranking method, used as reference method, is excluded from the table. Based on these values, we can make a decision about the null-hypothesis that H_0 : "*there is no significant difference in the estimation error between this (the best ranking) and any other chosen optimization method*".

Finally, visualizations like boxplots and accuracy versus time plots and accuracy of hypothesis tests are most suitable to be employed in order to identify the overall convergence and robustness of each GO algorithm's predictions.

5.4.3 Maintaining the algorithms' best possible performance and consistency in the results

To retain the best possible performance of each optimizer and also consistency in the results, two procedures have been considered and implemented. The first is to specifically set the internal parameters of each GO method so as to maximize their best possible performance. The second has been to produce the estimation results maintaining the same, *in silico*, conditions for all calculations.

5.4.3.1 Tuning of the GO algorithms' search mechanism

The GO methods described in the text above, especially the stochastic ones include a series of *design parameters* that can guide and essentially modify their search capabilities for the global optimum and, consecutively, their performance. Usually, in order to obtain the best possible performance on a given problem, one should consider the task of specific tuning of the design parameter setting for the optimization method used [211], [212]. It is important therefore, to configure each evolutionary algorithm by appropriate choosing such optimal design parameter values. Depending on the algorithm the number of these design parameters, needed to be specified, can vary substantially.

In general, the problem of finding the best possible configuration of the GO methods is similar to the problem of tuning the experimental design in order for the model to be predictive (paragraph 1.7.4). The difference here is that we will not prepare any real experiment in order to acquire the best possible experimental data for the validation of our estimations. Rather we want to identify the best possible *in silico* prerequisites in order to give each optimization method the most of its chances to converge to the global optimum, which we assume to exist and is feasible to reach.

Usually, in order to do so, trial and error is the safest option. Hence, sampling methods are used for testing several possible configurations and their respective performances. As mentioned in paragraph 4.4.1 the choice of the best sampling procedure is very important as it may reduce the search effort by decreasing the number of investigated design parameter settings. In any case, the search for the optimal configuration of design parameters is an optimization task in itself, which can be very computationally expensive. Therefore, although this procedure should have been followed for all the proposed algorithms we chose to do so only for those ones whose literature reports were scarce or obscure on solving optimization problems like the one presented here. In other words, no systematic approach for tuning the algorithm design parameters has been performed for all of the proposed GO algorithms. In fact, for all of the algorithms we have decided to employ literature reported design parameter values. In addition, for those GO methods where appropriate design parameter values were vague, we have heuristically assigned the most appropriate ones through sampling within their acceptable margins.

Under these circumstances, we have first performed an extensive research in the literature for finding the optimal design parameter values for each of the GO methods employed. Particularly, for the CRSwCH algorithm the unique user supplied parameters of algorithm as indicated in 5.3.1.1 are the threshold value for the probability of choosing a heuristic in the next iteration is reset to its initial one and the constant $c > 0$ that satisfies equation (5.7). In [192] the recommended, default, values for these parameters in the case of non-linear problems are $thr = 0.04$ for the threshold and 0.5 for the constant.

For the SCE algorithm the unique, user defined, parameters according to the descriptions above are the number of complexes used during the search (n_{gs}), the number of evolutions that a complex will undergo before it is shuffled (n_{evol}) and the number of refinements (n_{ref}) that each subset will undergo. According to [193] and heuristic experimentation we have concluded that for our inverse problem we should use as much complexes as our parameter complexity is i.e. k . As far as the number of evolutions is concerned these should be as much as $2*k+1$ and the refinements of each subset be equal to 2.

For the DE the design parameters that can be identified are the population size, the mutation scale factor (F) and the crossover probability (Cr). The group that conceived DE have indicated that a reasonable value for the population members could be chosen between $5-k$ and $10-k$ (recall that k is the parameter dimension of the problem). Then, F can be sampled from the interval $[0, 2]$ while Cr from the interval $[0, 1]$. As also described in [196], the design parameter F , which provides the mutation scale factor for the DE, is agreed that it should be optimally chosen within the interval $[0.4, 1]$ as other values are only occasionally effective. It is also noted that the higher the value of F the more premature the convergence of the population and its degeneration is. Hence, as is suggested in [213], the values of F suitable for our problem should be within the interval $[0.9, 1]$. On the other hand, the design parameter Cr , controls how many parameters in expectation are changed in a population member.

This means that if a low value is chosen (say 0.1) then, each generation will tend to be orthogonal to the previous one. In this case the population will converge slowly. Instead if a high value (say 0.9) is regarded most of the directions of the mutant vector will be inherited by the next generation. In this case the will converge fast. Therefore, in [193] it is stated that Cr should lie in (0, 0.2) when the function is separable, while in (0.9, 1) when the function's parameters are dependent. Here, since our parameters are expected to be entangled the interval [0.7, 0.9] should be used.

For the PSO method, the unique design parameters that can be specified by the user are the inertial weight (κ_1), the cognitive (κ_2) and the social parameters (κ_3) reported in paragraph 5.3.5.1. As claimed in [214] these parameters are not critical but influence the speed of convergence and the avoidance of local minima entrapment of the algorithm. One special problem that can arise is the possible explosion of the swarm's velocity as particles' velocities and positional coordinates careen toward infinity due to the random weighting. Therefore, the implementation of properly defined constriction coefficients can prevent this explosion [201]. Indicatively, in [215] it is, suggested that a default value lower or equal to 2 for κ_2 and κ_3 should be applied. Here, we have used the values the 2.8 for the κ_2 and 1.3 for the κ_3 .

As indicated in 5.3.6 the MIDACO algorithm includes several design parameters which are: the *Ants*, which is the amount of ants/iterates within a generation, the *Oracle* which is used by the penalty function of the algorithm but will not be considered in our case, the *Kernel* which defines the kernel size in a generation used by the Gauss distributions used and the *Character* that defines the ants' character. As indicated in [203] the *Ants* and *Kernel* must be used in combination and are intended to make MIDACO more efficient on specific problems. After employing the sampling procedure described earlier we concluded in the use of the values of 136 and 12, respectively. As far as the Character parameter is concerned, this has been given the value 3 that is de facto used for MINLP problems. Apart from these parameters there are two other ones, namely the *Autostop* and the *Quickstart*. The former defines the number of consecutive unsuccessful improvements after which the algorithm should restart the search. The greater the *Autostop*, the higher the chance of global optimality, but also the longer the optimization runs. Any integer greater or equal to zero can be selected for this parameter, giving the user the freedom to compromise between global optimality and running speed. In our case the value 10 has been considered as the optimal one. The latter lets the user to specify the quality of the starting population. The higher the value implies the better quality of the first estimation and the closer to that first estimated solution the algorithm searches for the global optimum. Since however, our initial population is randomly created we assigned the 0 value to this parameter, which equals to the absence of a quality evaluation of the starting point.

Finally, the design parameters of the SRES algorithm are: the λ , which is the number of offsprings generated in every generation that lays within the interval [100,200], κ , which is the number of parents $\kappa/\lambda \approx 1:7$, pf , which is the pressure of fitness that is valued within the range [0,0.5], $varphi$, which defines the expected rate of convergence and Gen , which is the maximum number of generations. Taking into consideration the successful use of a certain parameter set of SRES algorithm in [186] plus the author's comments on the free licensed Matlab™ script of the SRES algorithm and also the results from our sampling procedure, we attributed the following values: $\lambda=150$, $\kappa=20$, $pf=0.45$, $varphi=1$ and $G=500$.

The unique user supplied parameters of the DIRECT algorithm is only the epsilon, which is the global/local weight parameter. As noted in [207], [208] the epsilon value should

Table 5-2 The user specified parameters of the employed optimizers and their corresponding values as they were appointed for our calculations

Algorithm	Parameters				
CRSwCH	$w_0 = 0.5$	$thr = 0.04$	$F = 0.4$	$Cr = 0.9$	
SCE	$n_{evol} = 1$	$n_{ref} = 2$	$ngs = 4$ or 8 depending on parameter dimension		
DE	$F = 1$	$Cr = 0.8$	strategy: DE/rand/1/bin		
PSO	$\kappa_1 = 1$	$\kappa_2 = 2.8$	$\kappa_3 = 1.3$		
tdASA	-				
SRES	$\lambda = 150$	$\mu = 20$	$pf = 0.45$	$varfi = 1$	$G = 500$
MIDACO	Ants = 136	character = 3	krn = 12	Autostop = 10	SEED, QSTART, Oracle = 0
DIRECT	$\varepsilon = 10^{-4}$				

be 10^{-4} . Lastly, from the above description the tdASA algorithm does not require any user specified parameter. Table 5-2 shows the values of the design parameters as were selected for the purposes of this thesis.

5.4.3.2 Rules of fair comparison

After tuning the optimization parameters a certain set of rules of fairness have to be decided in order for the statistical comparison of the GO algorithms to be on the same basis. These rules include the use of the same (beside the cost function and PEX curves):

- (i) upper and lower bounds for the parameter values,
- (ii) sampling distributions,
- (iii) size of the initial population of search points,
- (iv) termination criteria i.e. time or search effort, convergence threshold, etc. and
- (v) computation power (cpu speed, parallelism, etc).

Table 5-1 summarizes the upper and lower bounds of the parameter values that were employed for the current study when the full and the reduced parameter sets are estimated. According to the discussion in paragraph 4.6 the parameter that defines the size of the IS has been kept constant. Specifically, the value $a=15\mu\text{m}$ was employed since this value has already been used by another groups for modeling the size of an abnormal cell in the cervical epithelium [216]. Additionally, we have performed a heuristic calibration of the model parameter values. Particularly, the maximum number of epithelial layers (N) has been truncated to 10 layers. The reason for doing this is that our preliminary tests performed after the reasonable restructuring; there is no significant change in the profiles of the simulated output

Table 5-1 Value intervals of the biological parameters when the full and the reduced parameter sets are estimated.

Parameter	Full parameter case	Reduced Parameter case
a	15 μm	15 μm
b	0.4-0.8 μM	0.4-0.8 μM
K_V	0.1-1	1
β_{ES}	10-30 mM	20 mM
β_{IS}	30-50	40 mM
N	1-10	1 - 10
pH_{ES}	6-7	6 - 7
pH_{IS}	7-7.4	7.2
g_{TJ}	7-18 \AA	7-18 \AA

when more than 10 layers are used as an input value for N . Biologically, this means that the possibility of the AA to reach and “light-up” more than 10 epithelial cells is low which is a biochemically reasonable assumption. Therefore, the new range of values has been adopted here, which will be between 1 and 10.

Samples for each of the model parameters have been drawn uniformly within the respective value intervals. Then, the initial sample/population, through which most of the GO algorithms start their computations, has been set to be randomly created. The size of the initial population equal for all GO method has been set to 60 parameter sets, thus providing the same starting position for all the population-based methods employed.

In addition, each algorithm has been set to halt once any of the following three criteria is met:

- cr.(1)* If a minimum curve fitness level fitting was achieved. A curve is considered to be fitted if the Euclidian distance of the experimental data from the mathematically produced data was at maximum 0.001 units. We have deliberately chosen this level to be substantially low in order to avoid the early stops of an algorithm.
- cr.(2)* If the population of the algorithm degenerates that much that the mean distance of the members from the geometrical mean of the population falls under a user provided threshold. This criterion corresponds to the convergence to a local optimum. This threshold was set to 1×10^{-3} .
- cr.(3)* If the appointed maximum running time was reached or exceeded. The running time concerns one trial of a given algorithm for each experiment. It has been set to 2,5 minutes.

Finally, the described benchmark of the GO methods has been performed using Intel Core i7 Processor running at 2.66GHz per core (quad-core with Hyper-threading Technology) with 8GB RAM memory under Matlab™'s 2012a environment. Then, for assessing the accuracy and reproducibility of the solutions an in-house software developed in Matlab™ was used. In principle, however, Matlab™ programs are significantly slow. To minimize this effect, parallel processing has been implemented in order to speed up the computations. Particularly, the *matlabpool* function of Matlab™'s parallel toolbox has been used in order to specify the number of workers available for the computations. Then, in order to be profiled in parallel the iterative procedures of each GO algorithm have to be able to be sliced. This primarily holds when the data produced during consecutive loops are independent to each other. Hence, for those algorithms where it was computationally feasible and efficient, parallelization has been performed. For those algorithms where parallelization was infeasible we let the 30 estimation repeats for each PEX to be performed in parallel. The latter case was utilized for the DIRECT algorithm due to its deterministic nature and for the CRSwCH method where during a loop the improvement of a single population member is realized.

Then, provided that all these similarity requirements are met the convergence of the parameter estimation problem can be assessed and compared.

5.5 PARAMETER ESTIMATION RESULTS

To assess the solution to our parameter estimation problem the collection of the appropriate experimental data is needed. According to the procedure described in paragraph 5.4.1 we generated these data artificially by simulation of the model (PEX data). Figure 5-10

depicts the 30 model output signals which have been created and were used as experimental ones when the full and the reduced set of parameters are estimated, respectively. It is clear that no two PEX datasets coincide and that in both cases they cover with adequate uniformity the experimental data space.

After generating these artificial data the minimizations were readily performed. Specifically, for every PEX fit a cost function was defined using equation (5.1) and the respective optimization problem was formulated. Then, using every GO method the objective function was allowed to minimize until the acceptable level of the goodness of the fit between the PEX and estimated dataset was reached. At that point, the set of estimated parameters was collected and, if multiple fits were necessary, the procedure was repeated 30 times. In total, $30 \times 30 \times 7 + 30 = 6330$ optimization runs were performed by the stochastic algorithms methods and the deterministic method i.e. 211 optimal solutions have been calculated for each PEX curve after the employment of all of the global optimizers described above.

In order to conduct this *in silico* experiment efficiently we developed software in order for consecutive algorithmic trials to run automatically. The program was developed in Matlab™ environment which allowed for the immediate processing of the results and production of statistical figures. However, the most costly part of the problem i.e. the solution of the differential system's dynamics has been written as a Fortran 90 module, callable from Matlab, using the LSODA as the initial value solver. As a result, we were able to retain all the advantages of the Matlab environment (e.g. for visualization and post-processing) while keeping a satisfactory computational efficiency. In addition, in their canonical form, apart from MIDACO, the scripts of the GO algorithms presented were only capable of handling continuous variables. For maintaining the robustness of these optimizers we have configured them in order to be capable of handling integer parameters i.e. the number of neoplastic layers parameter (N). Then, for those algorithms that would not accept integers, conversion of the real value to an integer value by truncation (rounding to the nearest integer) has been realized. Even though, more sophisticated approaches for handling the integer parameters in combinatorial problems have been argued for use in literature, this particular has been chosen because of its simplicity and ease of implementation [202].

5.5.1 Fitness level results

Figure 5-11 and Figure 5-12 show the performances of each GO method in terms of fitness level. The bold, dashed blue lines denote the PEX dynamic signals and the red clouds of continuous lines their respective fits.

By observing these clouds of fitted curves we can see that the majority of the algorithms appear to have a weaker performance when fitting of the full, more complex, parameter set is realized (Figure 5-11). However, a significant amount of algorithms have managed to fit several PEX curves quite satisfactorily in either case. In fact, what is soundly noticeable is that the fitness levels of the CRSwCH are remarkably good on all the optimization problems, regardless of the parameter set estimated. Similarly, the clouds of fitted curves of the DIRECT, the DE, the SCE and the PSO imply their good performance although for some of the curves, especially the ones regarding the full parameter set, the respective fits appear to be more scattered around the PEX dynamic signals. This is more evident in the majority of the fits of the SRES and the MIDACO optimizers which potentially implies their poor performance in solving our inverse problem. Lastly, the tdASA algorithm fails to fit the majority of the PEX data.

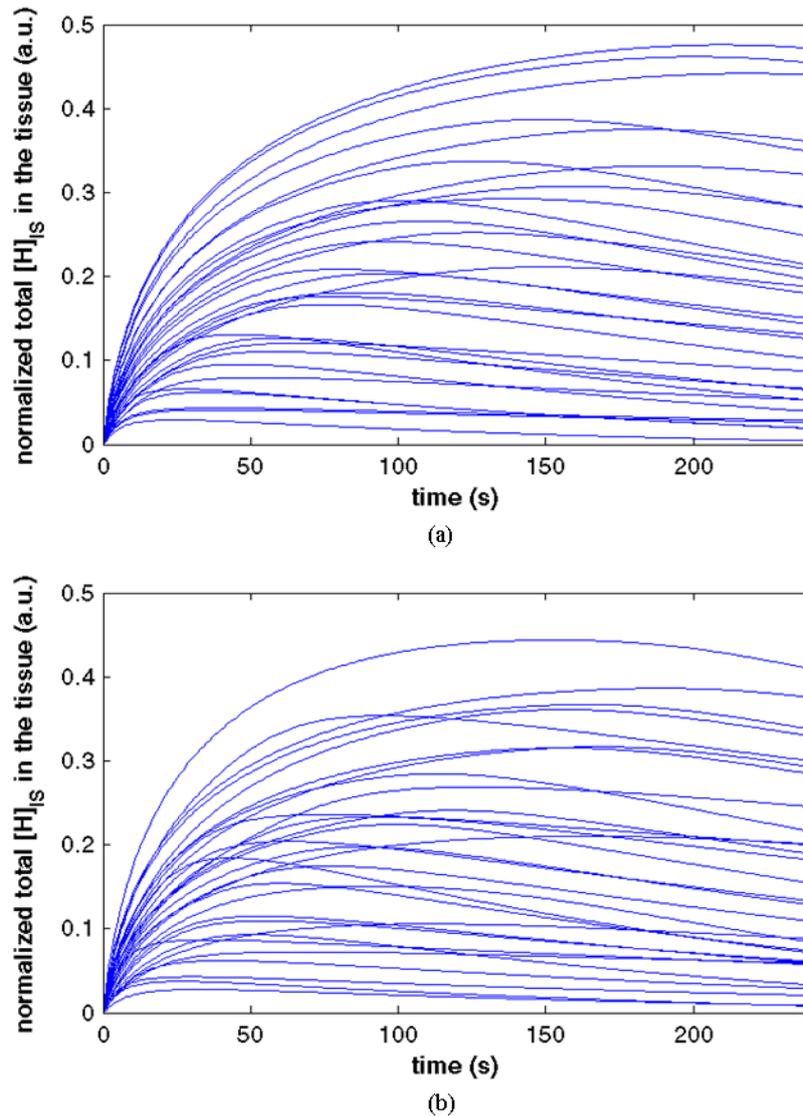


Figure 5-10 The 30 model output signals which have been created and were used as experimental ones. (a) The outputs for the full parameter set case. (b) The outputs for the reduced parameter set case.

The discrepancy among the performances of any GO technique in the fitting of the same PEX curve is a clear indication that the respective objective function includes several local minima in which the optimizers can be trapped; especially when the more complex dataset is fitted. This has been expected, because on the one hand the formulated objective functions are inherently multimodal and, on the other hand, the employed optimizers give no guaranty of reaching the global optimum rather they can only designate the region close to the global optimum. Furthermore, the discrepancy among the performances of the same GO technique upon fitting different PEX curves also implies that the fitting of each PEX curve is by itself a separate optimization problem. From a mathematical perspective this has also been expected because the number of layers defines the number of differential equations that are integrated in each simulation. As said, if the value of this parameter is different then, the respective PEX curve will be the result of the simulation of a separate equation system. Therefore, if an algorithm performs well in fitting a PEX curve it does not necessarily imply that it will perform equally well to any other fit. This also justifies our choice to create such a diverse group of PEX curves that attempts to cover most of the model's output space.

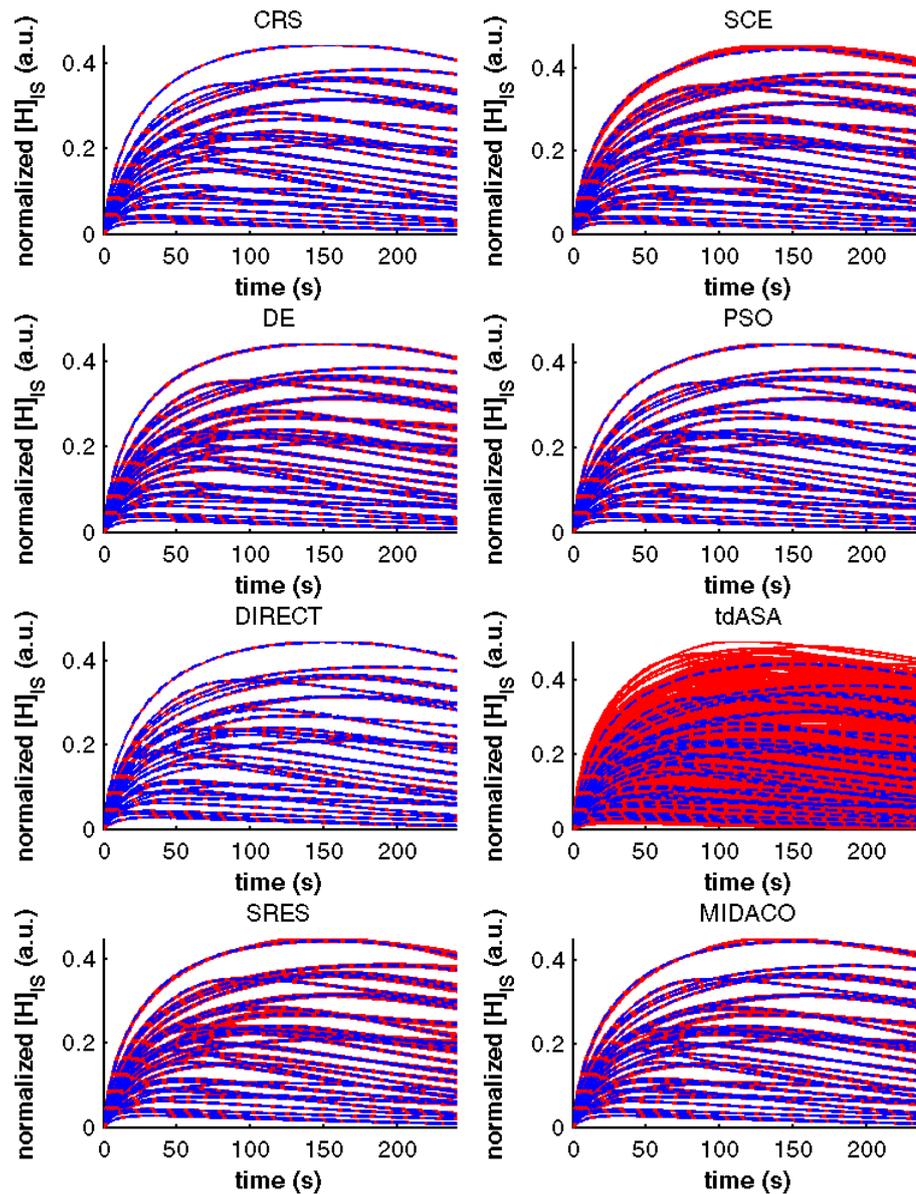


Figure 5-11 The fitting result of each GO method in terms of curve convergence when estimation of the full parameter set is realized. On top of each sub-figure the algorithm is the algorithm is indicated. Each dashed blue line displays the PEX curve and the clouds of red lines their corresponding

In order to numerically depict the illustrated fitness levels, the second column in Table 5-3 shows the average Euclidean distance over all PEX problems and over all runs that each GO method has accomplished after termination through either *cr.(1)*, *cr.(2)* or *cr.(3)* criterion in the full parameter set case. The dotted line is used to separate the stochastic optimizers from the deterministic one. According to these calculations we can see that the average objective value of CRSwCH is the lowest one (gray-shaded value) and is nearly equal to the threshold value (0.001). This verifies our earlier observation that this algorithm should be the one with the highest curve fitting capability. Then, the ranking is as follows. First, is the DE method whose average distance is little more than twice the distance of the CRSwCH method's. Second, is the MIDACO whose average J is three times that of CRSwCH. Then, the PSO, the SRES and the SCE methods follow in average ascending distance values which are however very close to the one of the MIDACO method. After that, the DIRECT method has

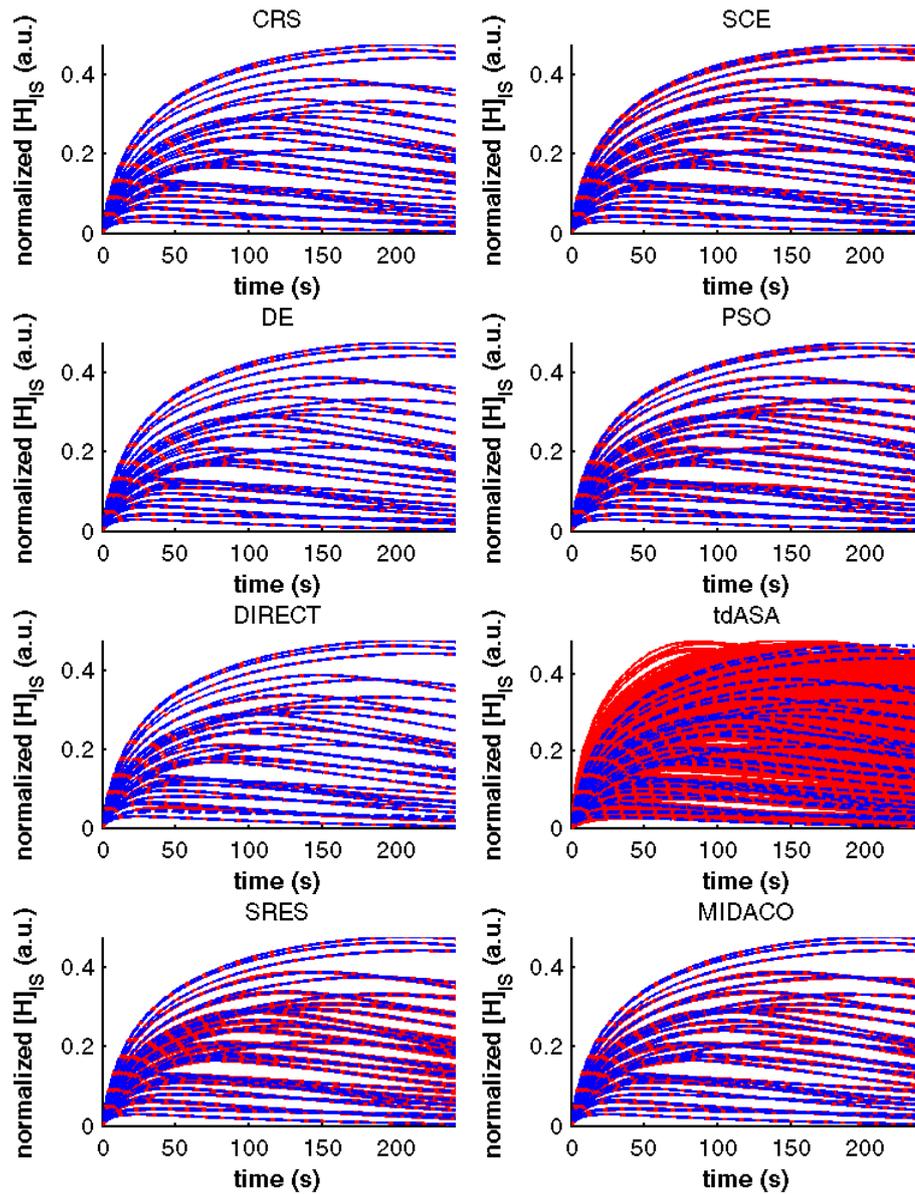


Figure 5-12 The fitting result of each GO method in terms of curve convergence when estimation of the reduced parameter set is realized. The same notation as in Figure 5-11 has been employed

almost six times the objective value of the best performing algorithm while the tdASA has shown the worst fitting performance in terms of average objective function value.

The next three columns indicate how many times each algorithm was forced to end the optimization due to either termination criterion. The values represent the respective probability measured calculated as a percentage over the number of runs. What we can notice from these three columns is that all the stochastic algorithms terminate their search most probably due to *cr.(2)*, then *cr.(3)* and lastly due to *cr.(1)*. Recall that an algorithm may terminate its search due to: *cr.(1)*, when the distance between a PEX and a fitted curve is lower that the threshold value; *cr.(2)*, when the population of potential solutions degenerates; and *cr.(3)*, when the maximum optimization time was exceeded. Then, in conjunction with the values of the first column and the not-so-good curve convergence depicted in Figure 5-11, these results imply that the stochastic optimizers should have been trapped into iterating around some locally optimal solution. This verifies our earlier observation on the discrepancy among

the fitness levels when referring to the objective functions of the more complex parameter case. Another interesting observation here is that the probability of an algorithm to terminate because the maximum running time was insufficient is larger than the probability to terminate because the minimum curve fitness level was achieved, in most cases. Perhaps, if some of the optimizers were allowed to perform their minimizations on a longer time scale they will improve. However, this does not mean that the algorithm will avoid its entrapment on a locally optimum solution.

Table 5-3 The average objective value (J) of each global optimizer after fitting the 30 PEX curves corresponding to the full and reduced parameter sets. The probability of each method to terminate due to either $cr.(1)$, $cr.(2)$ or $cr.(3)$ criterion is also indicated.

Algorithm	Full parameter set				Reduced parameter set			
	J	$cr.(1)$	$cr.(2)$	$cr.(3)$	J	$cr.(1)$	$cr.(2)$	$cr.(3)$
CRSwCH	1,6E-3	33%	54%	14%	7,7E-4	67%	28%	5%
SCE	5,6E-3	4%	83%	13%	1,8E-3	62%	32%	6%
DE	3,7E-3	10%	56%	34%	5,9E-4	82%	16%	2%
PSO	5,0E-3	10%	52%	38%	2,6E-3	62%	32%	6%
tdASA	1,1E-2	0%	92%	8%	1,1E-2	0%	93%	7%
SRES	5,2E-3	11%	64%	25%	5,1E-3	25%	52%	23%
MIDACO	4,7E-3	7%	77%	15%	4,7E-3	37%	53%	10%
DIRECT	9,7E-3	0%	-	100%	9,7E-3	23%	-	77%

This latter result is totally reversed in the reduced parameter set case (columns six to nine) in which the probability to terminate because the maximum running time was insufficient was minimal for most cases. In fact, the reason to terminate has shifted towards the $cr.(1)$ criterion for the stochastic optimizers except for the MIDACO and the SRES whose probabilities to terminate their search for the globally optimal solution was found almost the same as before. Similarly the DIRECT method retained its average objective value while for some cases, the algorithm had terminated due to the convergence between the PEX and predicted curves contrary to before. As expected, the average objective value of the tdASA method was the only one that, by comparison to the respective value in the full parameter set case, was unchanged. This means that that this technique cannot fit many of the PEX curves and hence it will probably be unsuitable for estimating the reduced parameter set. In contrast and similarly to the full parameter set case, we can notice that the average J of the CRSwCH method has been, again, adequate although the lowest average objective value belongs to the DE method. In both cases, the average J has been substantially below the threshold value indicating an extremely good curve convergence. Equally, as far as the SCE and PSO methods are concerned, their average objective values were found 3 and 2 times lower than before, respectively. Probably, the substantial improvements in these fitness levels of the last four methods, implies that their respective performances on the estimation of the reduced parameter set will be equally improved. However, these are only speculations because at this point we study only the curve convergences and not the respective parameter convergences. In the following paragraphs where we will describe the respective parameter estimation results, rigorous inferences will then be made.

As a last illustration of our curve fitting results, Figure 5-13 shows in logarithmic scale the efficiency of each GO method in terms of fitness level with respect to the time lapsed after the initiation of the fitting procedure. Each point of the lines drawn refers to the average of the Euclidean distance over all repetitive runs and all PEX data as it was recorded every 3

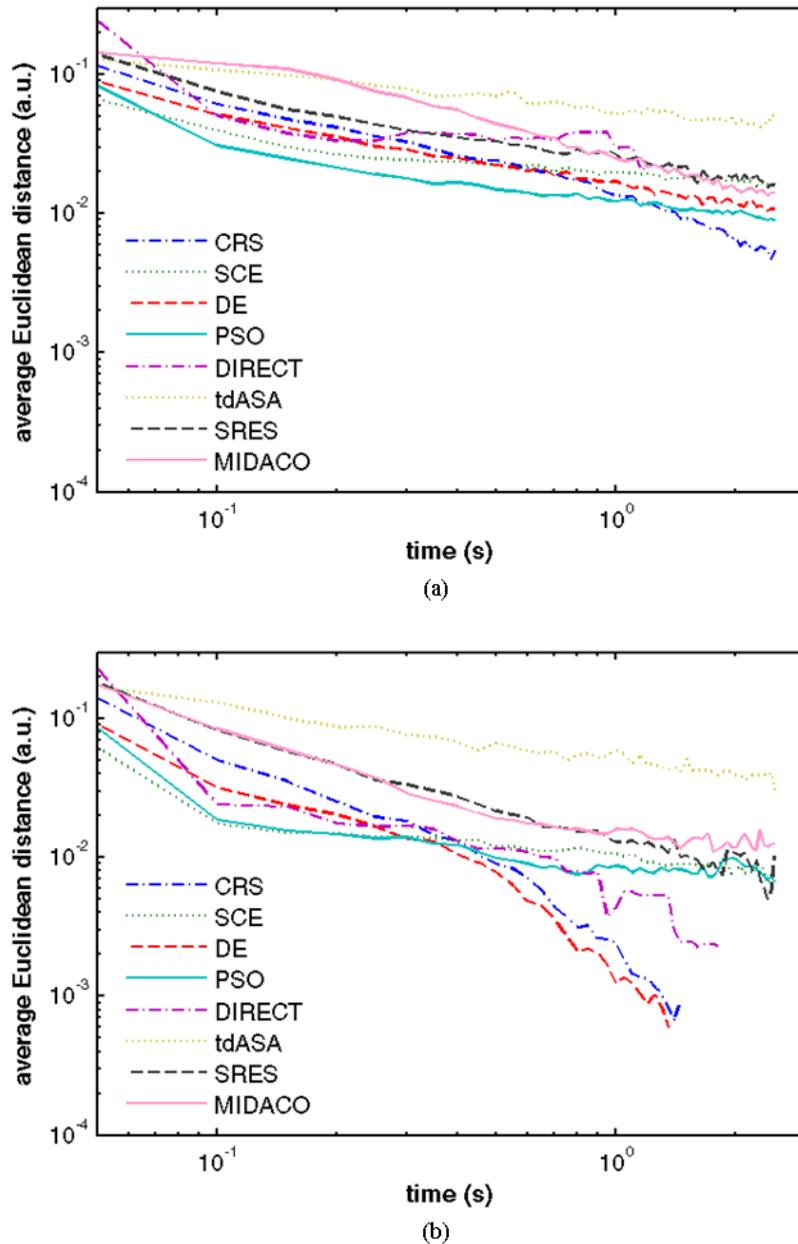


Figure 5-13 Loglog plot of the efficiency with respect to the time lapsed after the initiation of the fitting procedure of each GO method. Each algorithm is represented by separate color lines (see legend).

seconds. Recall that the maximum time is 2.5 minutes. Also note that the average values do not take into account the fact that in some of the PEX curve fit cases the optimizer may have terminated i.e. the lines drawn depict the worst case scenario.

Then, according to the illustrations for the full parameter, more complex, case (Figure 5-13a) we can notice that the efficiency is significantly lower for the CRSwCH followed by the DE, the SCE and the PSO method. This result concurs with the ranking of the GO methods according to the values of the second column in Table 5-3. Other than that, what is of interest here is that all the algorithms expend, on average, the maximum running time of 2.5 minutes. This links to the aforementioned observation that if they were allowed to perform their minimizations on a longer time scale then they will probably perform better. In addition, as before, the tdASA was the worst performing algorithm. Similarly, according to (Figure 5-13b) the

worst performing algorithm in terms of objective function value in the case of estimating the reduced parameter set was, again, the tdASA method. After that, the fitting efficiencies of the MIDACO, the SCE, the PSO and SRES were quite similar. Interestingly, the DIRECT method seems to have a quite adequate fitting efficiency although the average objective value indicated in column six of Table 5-3 implied otherwise. Finally, as expected, the CRSwCH and the DE methods appear to be the most efficient curve fitting algorithms. In fact, they managed to fit any PEX curve in only a minute, in the worst case.

5.5.2 Parameter estimation errors – Theoretical analyses' validation

Thus far we demonstrated that the performance and efficiency of most of the employed GO methods in fitting the model's output to PEX observations is more than satisfactory. In no case, however, may this result prove that the parameter values of the respective fitted predictions are also accurate and reproducible. In fact, we have already mentioned that the optimization problems formulated by each PEX curve probably contain several locally optimal solutions. The latter can be further illustrated by reference to the results of the previous chapter where the mathematical entanglement and/or correlation of the input parameters were assessed. Therefore, in order to extract rigorous conclusions on the solvability of the inverse problem the respective parameter estimation errors have been investigated.

Towards this direction, the barplots in Figure 5-14 and Figure 5-15 capture the errors in the predictions of the parameter values when the full and the reduced set were estimated, respectively. In both graphs, the horizontal axis indicates the estimated input parameter and each of the bars above its corresponding mean NRMSE over all PEX fits. The latter has been

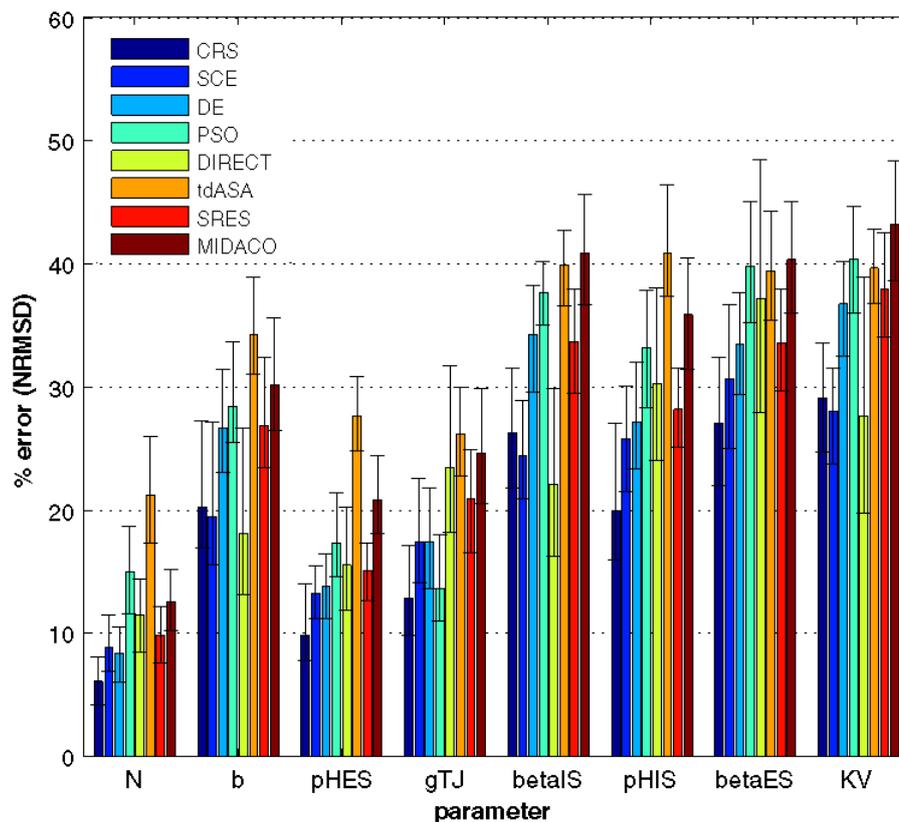


Figure 5-14 Parameter prediction error when the full parameter set is estimated. The error bars have been computed using the bootstrap method where 1000 bootstrap samples were used as input to the Matlab's function *bootc()*. Each algorithm is represented with a separate color bar (see legend).

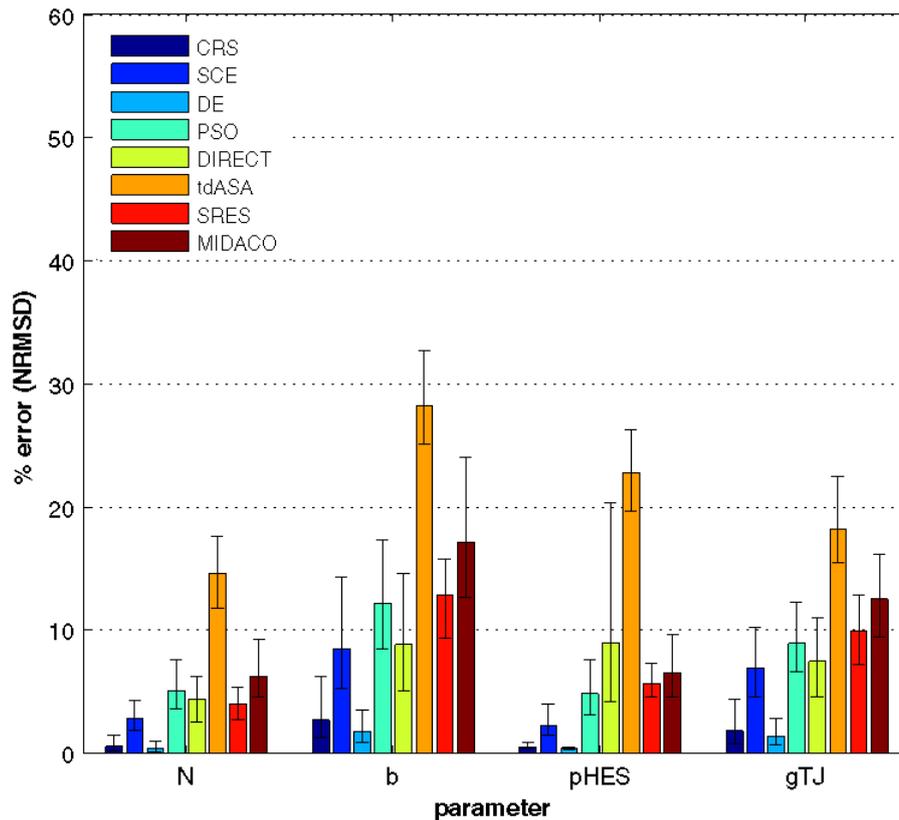


Figure 5-15 The prediction error of the parameter values when the reduced parameter set is estimated. The same notation as in Figure 5-14 has been employed.

calculated using equation (5.29) for each of the employed optimizers. The confidence intervals denoted by the error lines have been determined according the bootstrap method. Then, as far as the more complex case is concerned (Figure 5-14), it is clear that the good curve fit performances illustrated earlier in Figure 5-11 did not pass on the respective prediction performance of the parameter values. In fact, the estimation error values are significantly high even for the most efficient algorithm according to Figure 5-14, the CRSwCH (navy blue colorbars). Furthermore, the error margins defined by the 95% confidence intervals indicate that the maximum uncertainty for every prediction is at least twice the mean value. This is also significant considering that the respective mean values are already high. In any case, these results indicate that the best performing algorithm for estimating the full parameter set appears to be the CRSwCH while the worst one is the MIDACO algorithm. In addition, parameter N is consistently the one with the least estimation error regardless the employed GO technique.

This latter observation also appears to apply to the results illustrated in Figure 5-15 which concern the parameter prediction errors in the case of the reduced set. Actually, there is a significant improvement in the prediction of all of the parameters in this case which means that the dimensionality reduction improved the solvability of the parameter estimation problem. This improvement is particularly impressive when the DE method is employed where the predictions appear to be highly accurate for all four of the most influential parameters of the model. The latter indicates, for the first time, that that our inverse problem can be, in principle, solvable.

On the same basis, Figure 5-16 shows the barplots of the aggregate estimation error of each parameter over all GO methods employed and over all curve fits for the two PEX da-

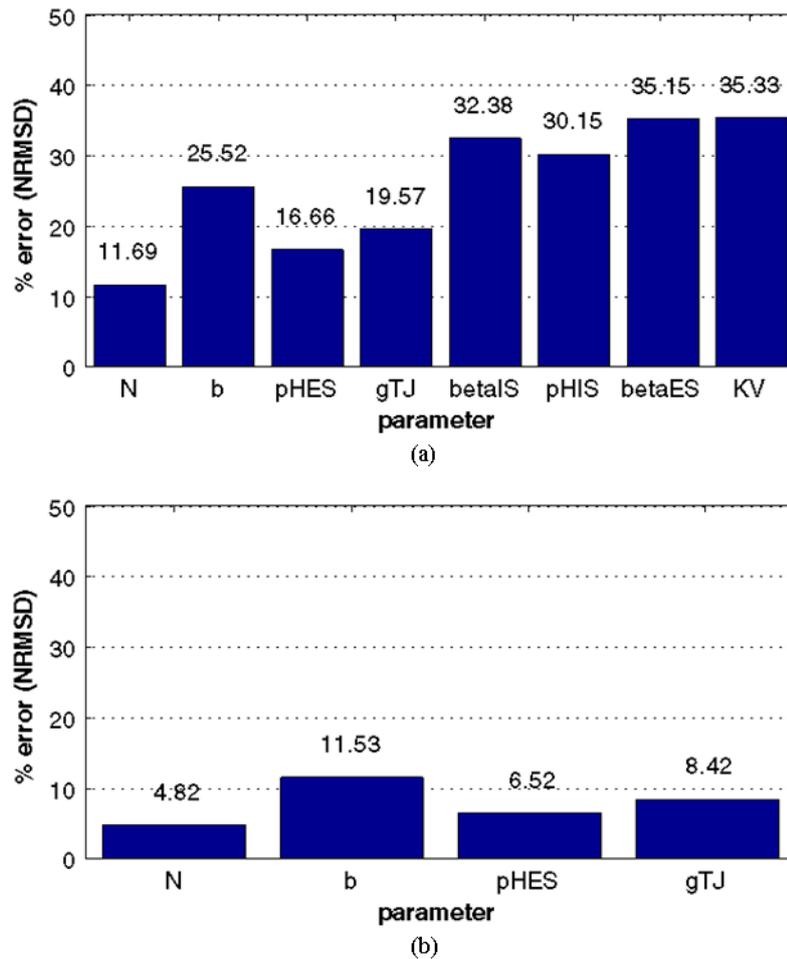


Figure 5-16 (a) The average estimation error of each parameter over all GO methods employed and over all curve fits for the full parameter PEX dataset. (b) The respective average estimation error for the reduced parameter PEX dataset. On top of the bars the mean average error value is indicated.

tasets. Again, in both figures the horizontal axis indicates the estimated input parameter but now the bar above it corresponds to the overall NRMSE. The numbers on top of each bar denote the marginal values of each box.

According to these marginal values we can readily verify our earlier assumption that the parameter N is the best estimated one. This is consistent with the results of the analyses presented in the previous chapter. Particularly, the fact that N is better predicted than any other parameter in the set is in total agreement with the estimability ratio of N calculated in the previous chapter (paragraph 4.5.3.2) where we have shown that N is theoretically the most estimable parameter. Next, N is followed, in ascending estimation error order, by the extracellular pH (pH_{ES}), the tissue's fluidity (g_{TJ}) and the size of the extracellular space (b) in both the full and the reduced set case. This is in partly agreement with the theoretical estimability calculations where b had an equal with pH_{ES} followed by g_{TJ} and, in practice; we have found that b is the least estimable of these three parameters. Still, this is not unexpected because we have to keep in mind that according to the restructuring proposed in paragraph 4.6 the range of values of b has been truncated. Obviously, this has lowered its influence on the model's output and, hence, its estimability. Even so, what is interesting to note here is that the two parameter groups that are distinguishable are the same ones that resulted from the GSA parameter ranking in Figure 4-6 (without parameter a). Clearly, this is

not a coincidence. Rather, it is a validation of our earlier statement that the higher the FOS effect of a parameter is; the more are the chances of it being estimated when observations of the output become available. Then, as far as the lowest ranked parameters are concerned, their respective estimation accuracy seen in Figure 5-16a is unacceptable. In addition, no clear distinction about their ranking can be made although their mean values indicate the same trend as their respective FOS indexes have in Figure 4-6. This proves that, in practice, these parameters are unidentifiable and thus cannot be estimated through the proposed model structure. This justifies, in turn, our proposition for these parameters to be fixed to a nominal value in order for the inverse problem to be solvable. Along these lines, the overall NRMSDs of \mathbf{N} and \mathbf{pH}_{ES} are shown quite insignificant in Figure 5-16b while the error in the predictions of g_{TJ} and \mathbf{b} has also been found low. In fact if we focus only on the results of CRSwCH and DE we can reasonably assume that the predictions for these parameters can be substantially unique when the reduced parameter set is estimated. These findings clearly demonstrate, therefore, that the switch from the full set to the reduced, most sensitive, set according to GSA greatly enhances our ability to uniquely predict the neoplasia relevant biologic parameters when actual clinical data are available.

5.5.3 Accuracy of the GO methods

In paragraph 5.4.2 we indicated that the performance of the optimization process can be determined by aggregation of the calculated accuracies, first, over all parameters and, second, over all PEX problems. Along these lines, we have computed using equation (5.30) the weighted average of the NRMSD in the parameter estimates. To form a complete picture of the prediction quality of the optimizers and, additionally, compare their respective performances in the best possible way box-and-whisker diagrams (boxplots) have also been constructed. Since the objective functions of some PEX curves may serve as separate optimization problem these graphical summaries are the best possible ways to represent the performance distributions of each of the employed GO technique. Thereupon, Figure 5-17 displays these boxplots corresponding to the full and the reduced set, respectively. The horizontal axis in both graphs denotes the algorithm and the boxes above the distributions the respective average parameter estimation errors. The circles within the boxes indicate the mean values.

Generally, the performance medians illustrated in Figure 5-17a are unacceptably large and show significant variability indicating a small confidence in the results. As one would expect, though, the CRSwCH has been performing better in terms of low prediction error. Conversely, the SCE method is the most robust one in terms of error variability in its predictions. For the rest of the algorithms employed DE shows the largest variability although its mean performance is quite low. Similar results to the DE are shown by the DIRECT method but with a lower variability in its performance. The same applies between the SRES and the PSO methods where the variability of the former is larger. Lastly, the MIDACO and the PSO methods show the highest prediction errors and, in addition, large variability.

The high variance and evident outliers in the values confirm that some of the PEX signals lead to more difficult optimization tasks. Recall here that the employed optimizers cannot guaranty global optimality and, additionally each PEX curve is by itself a separate optimization problem. Consequently, one can reasonably anticipate that some algorithms will have significant trouble to repeatedly reach a satisfactory low parameter estimation error. This has already been implied earlier where Figure 5-11 and Figure 5-12 clearly indicate that some algorithms are persistently unable to reach satisfactory fitness levels. As a rule, the

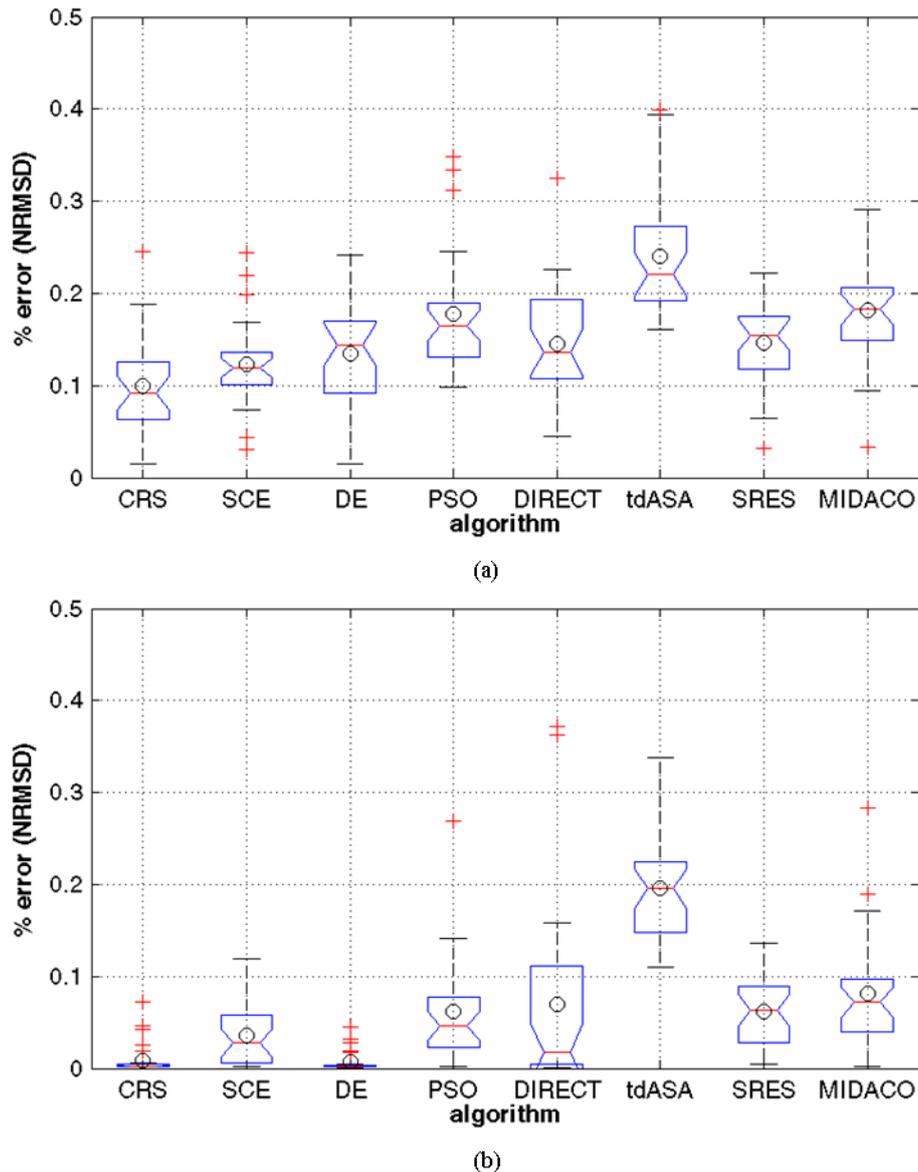


Figure 5-17 Boxplots of the distributions of the prediction quality of the employed optimizers when the values of the more complex (a) and of the reduced (b) parameter set are estimated. It is clear that the reduction of the parameter complexity has significantly improved the accuracy in the predictions to all algorithms except tdASA. The circles within the boxes indicate

inconsistency in the values comes from either the existence of locally optimum solutions within which the algorithms are trapped or the fact that the optimization time has been insufficient. Yet, our results strongly suggest that the latter is not as likely to occur (see Table 5-3) because in most cases the reason for an algorithm to stop its search has been the hitch on a locally optimal solution (full parameter case) or the reaching of an objective value below the employed threshold (reduced parameter case). It follows that the apparent variability in the results stems, in principle, from the inherent inability of the GO algorithms to reach the global optimum when several local minima appear in the objective functions. Besides, as dictated by the theoretical analyses described in the previous chapter the correlation for some parameter pairs are strong and, hence, their identifiability has already been expected to be poor. This is clearly indicated from the fact that even if the discrepancy in the fitness levels has been found low for some methods (CRSwCH, DE, SCE) the respective variability in their

predictions is substantially high. Therefore, the fixing of the most unimportant parameters to a nominal value is undoubtedly needed.

Figure 5-17b summarizes the quality of the predictions that corresponds to the estimation of the reduced parameter set. Here, the quality of the mean and median prediction errors shows substantial improvement over the complete set case while the respective variability shows a moderate improvement for most of the algorithms. Evidently, the reduction of the parameter complexity according to the analysis described in the previous chapter leads us to conclude that the unique solution of the inverse problem can be possible. Especially, the CRSwCH and the DE methods show great robustness and, additionally, are performing more than satisfactory in terms of both mean and median values. Essentially, their performance indicates that these optimizers are capable of predicting with extreme accuracy and reproducibility all the input parameters of the reduced parameter set. Next, the performance of SCE compares favorably to the one of the PSO, and the SRES optimizers although there is no statistical significance between them. The DIRECT method even if it shows a very low median value the respective variance is quite large. Similarly, although the mean and median performance of the SRES and MIDACO methods are appreciably good their robustness is quite poor. Lastly, the tdASA method shows the most variability even and the worst average performance among the employed optimizers.

Next, and in order to assess the overall performance of the employed algorithms and designate the best performing one aggregation of the NRMSDs over all PEX problems has been performed. As already indicated the mean over each population of computed average NRMSD is the appropriate statistic here since we have fitted a considerable amount of PEX data in order for the mean estimation error and its variance to be, approximately, normally distributed. However, as expected, several outliers appear in the calculated average errors. Also for some of the algorithms their respective population is considerably skewed especially when the reduced set is estimated (skewness can be seen from the difference between the mean and the median). These demonstrate a departure from normality for some of the methods. Hence, the calculated standard deviations will differ from their actual values. For this reason, the non-parametric bootstrap method has been employed in order to compute the confidence.

The barplots in Figure 5-18 show the overall estimation errors (*performance*) of each algorithm according to equation (5.31), for both parameter estimation datasets. The 95% confidence intervals have also been calculated and plotted as error bars. The number on top of each error bar indicates the mean value. Predominantly, one can easily deduce that these two graphs confirm all our previous observations. First, is that if the trends in the performances are compared, we can directly infer, once more, that when the reduced set is estimated the estimation error drops significantly. Specifically, the maximum possible error drops from 35% when the 8 parameter set is fitted to 16% when the 4 parameter set is fitted. In contrast, the minimum possible error drops from almost 10% when the 8 parameter set is fitted to only 0.9% when the 4 parameter set is fitted. Second, is that the CRSwCH, the hybrid CRS and DE method, is undeniably the best optimizer that can be utilized to estimate the parameters in both cases (1% and 15% maximum error, respectively). In fact, the 1% error obtained from the NRMSD in the parameter estimates proves that the unique and repeatable solution of the inverse problem is achievable. Third, is that the next in ranking is the DE method for the reduced set case and the SCE method for the full parameter case and the reverse thereafter. The fact that the aggregate error of these two algorithms is extremely low when the four most influential parameters are estimated, demonstrates the statistical signifi-

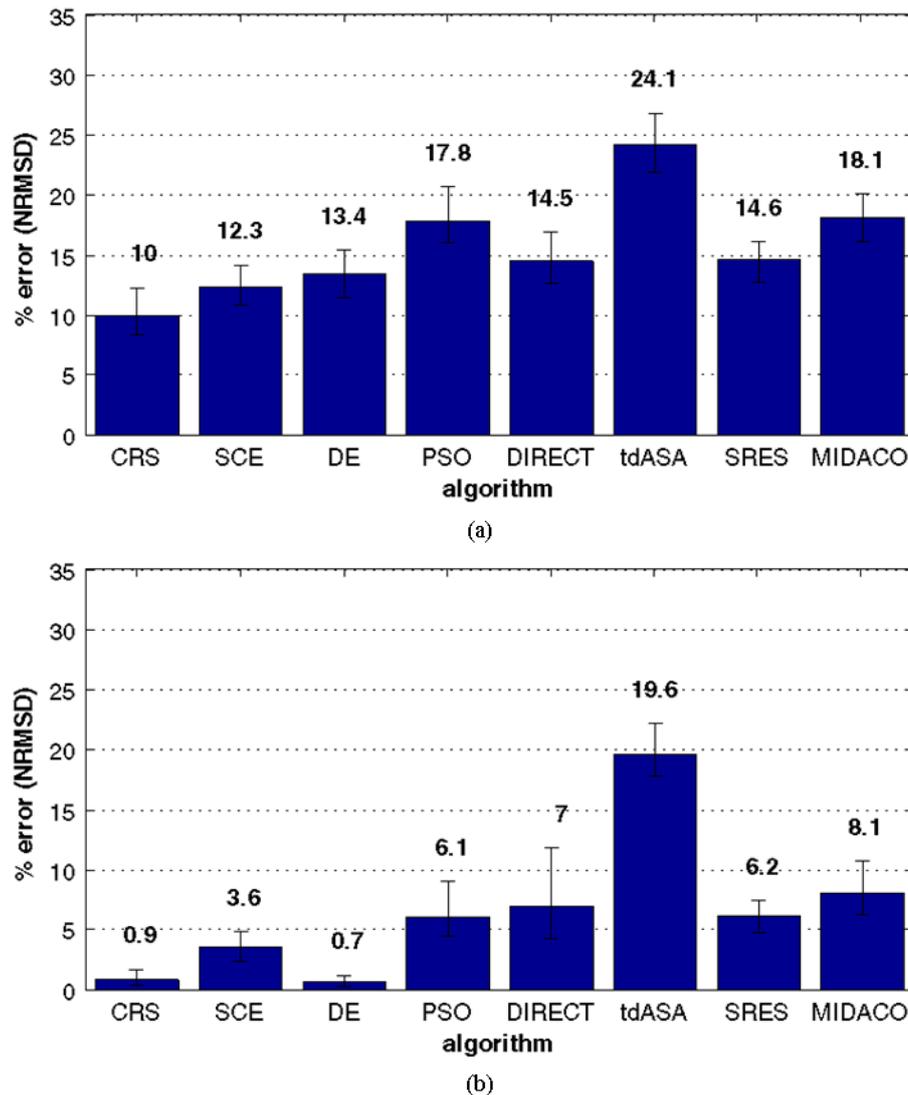


Figure 5-18 The overall performance of each algorithm in terms of mean estimation error over all PEX curves. For calculating the confidence of the mean values 1000 bootstrap samples with replacement have been employed. It is clear that CRSwCH performs best in the full parameter case (a) and in the reduced parameter set case (b). Conversely, the performance of tdASA is the highest in both cases.

cance of the successfulness of the CRSwCH method i.e. that it is unlikely for its predictions to have occurred by chance.

5.5.4 Statistical convergence of the inverse problem – the best performing algorithms

Thus far, the interpretation of illustrations such as boxplots and barplots of the performance of each GO method strongly suggested that the inverse problem may converge to unique solutions. Then, a common practice to evaluate whether this particular observation is consistent, are statistical hypothesis tests. Roughly speaking, hypotheses are statements about population parameters like expected value and variance. For this, there are two competing claims between which we have a choice; the null hypothesis, denoted as H_0 , against the alternative hypothesis, denoted as H_1 . In most situations, the null hypothesis says that the effect indicated by the sample is due only to random variation between the sample and the population [50]. Conversely, the alternate hypothesis says that the effect indicated by the sample is real, in that it accurately represents the whole population.

In our case, we wish to put forward a hypothesis test for making certain that the inverse problem may converge to unique solutions. Mathematically the validity of this hypothesis entails the assessment of the plausibility that the mean accuracy of each GO method is statistically lower than a threshold. Thereupon, and based on the populations of errors calculated using equation (5.30) the null hypothesis will be as follows:

H_0 : "can a GO method predict the parameters with less than a certain estimation error?"

To investigate this hypothesis a one-tail, unpaired Student's t -test has been conducted for every GO algorithm. The significance level was set to 0.05. As a preparatory step, the outliers in the populations of accuracies have been dropped [54]. Then, t -tests have been performed on the remaining populations. In addition, we have heuristically put to test the validity of a 5% and of an 8.5% estimation error. Table 5-4, shows the t -test result of every algorithm. Note that only the estimation error values of the reduced set are evaluated because the respective of the full parameter set has been found unacceptably large.

Table 5-4 Results of the Students t -test for significance level (threshold probability) $\alpha=0.05$ for both estimation error thresholds when the reduced parameter set has been estimated. The hypothesis is rejected if the statement $P\text{-value} > \alpha$ holds.

Algorithm	$\mu_{al} \leq 5\%$		$\mu_{al} \leq 8.5\%$	
	P-value	Hypothesis	P-value	Hypothesis
CRSwCH	$2,8 \cdot 10^{-42}$	accepted	$4,7 \cdot 10^{-48}$	accepted
SCE	0,02	accepted	$1,3 \cdot 10^{-08}$	accepted
DE	$5,7 \cdot 10^{-43}$	accepted	$1,0 \cdot 10^{-48}$	accepted
PSO	0,69	rejected	0,0002	accepted
DIRECT	0,46	rejected	0,0007	accepted
tdASA	1	rejected	1	rejected
SRES	0,95	rejected	0,0011	accepted
MIDACO	0,98	rejected	0,046	accepted

According to these statistical results one can readily observe that the hypothesis that the calculated mean estimation error is less than 5% is plausible when either the DE or the CRSwCH or the SCE optimizers are employed. This markedly shows that there is a 95% chance that the predictions of each of these methods for the values of the four most important parameters will be highly accurate when the estimation of the latter is intended. In fact, our results indicate that the H_0 is satisfied even at the 1% estimation error for the DE and CRSwCH optimization methods. Essentially, therefore, the predictions of these algorithms for the reduced set of parameter values designate that the convergence of the inverse problem to unique and reproducible solutions is statistically possible. To support this claim, the hypothesis that the estimation error is less than 8.5% has been accepted for all algorithms, except tdASA. Consequently, the convergence hypothesis of the inverse problem is statistically robust at an acceptable prediction error. It follows that the structural and functional, cancer-specific microscopic parameters can be substantially assessed in a rigorous and reproducible manner.

Towards determining the best algorithm on a rigorous statistical basis a Holm-Sidak post hoc significance test has been applied for between-algorithm comparisons. The hypothesis for this test was the following:

H_0 : "there is no significant difference in the estimation error between the best ranking and any other chosen optimization method"

According to the description in paragraph 5.4.2 and our calculated μ_{ai} values, the best ranking algorithm has been the DE optimizer ($i=0$). Afterwards, in ascending μ_{ai} value order, are: the CRSwCH; the SCE; the PSO; the SRES; the DIRECT; the MIDACO; and, lastly, the tdASA optimizer. Next, Table 5-5 summarizes the outcome of the Holm–Sidak test performed on the mean error values obtained by parameter estimation with the eight GO methods from the artificial data. Specifically, it reports the P -value and the α/i values where each row corresponds to significance test between the DE and every other GO method.

Table 5-5 Results of the Holm-Sidak test for significance level $\alpha=0.05$. The hypothesis is rejected if the statement $P\text{-value} < \alpha/i$ holds. CRSwCH is the reference method with $i = 0$.

i	Test	P -value	α/i	Hypothesis
7	DE – tdASA	0.0000	0.0073	<i>rejected</i>
6	DE – MIDACO	0.0000	0.0085	<i>rejected</i>
5	DE – DIRECT	0.0000	0.0102	<i>rejected</i>
4	DE – SRES	0.0001	0.0127	<i>rejected</i>
3	DE – PSO	0.0001	0.0170	<i>rejected</i>
2	DE – SCE	0.0364	0.0253	fail to reject
1	DE – CRSwCH	0.8641	-	accepted

Using DE as reference the application of this pairwise multiple comparisons test has shown that this method's predictions have genuine statistical significance when compared to all others but the CRSwCH and the SCE. Particularly, there is little statistical significance between the results of the DE and the SCE method. This shows that although the mean errors between these two have been different, their variation has been partially overlapping. Conversely, between the prediction accuracies of DE and CRSwCH there is no statistical significance. This shows that the results from both methods can be equally employed to determine the solution to the inverse problem.

5.5.5 Interpretation of the performances of the GO methods

Since we are dealing with stochastic methods there can be no standard justification for the superior efficiency of one over the other. As already said, the performance of an algorithm is problem specific and since none of these methods can guarantee optimality, it is only the distinct features of each method which, when compared to the rest of the employed ones, may provide, to some extent, a satisfactory explanation. Here, the CRSwCH, and the DE methods clearly perform better in fitting either PEX curve set than any other method. A strong reason for this could be that both algorithms rely heavily on the difference of parameter vectors in order to perturb the trial solutions. The simplicity of their internal procedures due to this fact may be another reason that gave them an advantage over the rest of the stochastic methods that employ more elaborate/complex, time consuming search mechanisms. In addition, another reason may be the fact that both algorithms employ several different mechanisms in order to evolve the potential solution candidates. For example, the CRSwCH method integrates three different versions of the original CRS algorithm and a DE variant. It stochastically chooses between these heuristics in order to generate the next potential solution. Similarly, the DE method is a mixture of evolutionary optimization, the Nelder-Mead algorithm and the CRS algorithm. It stochastically chooses between the *trial* and the *target* vector in order to appoint a new population member. Perhaps, the adoption of different mechanisms for the search of the global optimum by these algorithms is the reason for them to behave with so much better efficiency.

Furthermore, the fact that the employed optimizers, with the exception of the DIRECT, are stochastic the repeatability of this cannot be foreseen. Obviously, however, at least in the reduced parameter set case, DE and CRSwCH are significantly superior to the other methods. One strong reason to explain the accuracy and robustness of these two best performing methods is their hybrid nature; as it did with their superior efficiency which was explained in the previous paragraph. In fact, it is generally accepted that the inclusion and adaptive switching between different mechanisms in a single algorithm allows the adaptation to the conditions of the problem which, in turn, improves both efficiency and robustness. For example the SCE method that persistently performs well is also a hybrid technique that includes the essences of both the adaptive search methods and the evolutionary strategies. It adopts reflection, contraction and/or mutation in order to optimize each member of the created population. We should also remind here that all three algorithms employ difference vectors which speedup and simplify their internal procedures.

As far as the rest of the employed techniques are concerned, the poor efficiency and robustness they have shown may originate from various reasons. For example, although the PSO method is also utilizing difference vectors in order to evolve its search, perhaps, its weakness has been that the search is influenced by the points' neighborhoods which can potentially drive the population to local solutions way off the desired global one. Next, although the SRES strategy resembles the SCE in terms of employing a population sorting mechanism at each generation evolution step, it shows an increase in mean and median average estimation deviations. A strong reason for this is the fact that the SRES method evolves its potential solutions in a more complex fashion i.e. using a log-normal, self-adaptive update rule at mutation stage as indicated in paragraph 5.3.7.1, which may have slow down the search procedure. Similarly, the MIDACO strategy utilizes another purely stochastic strategy where several Gaussian kernels are employed iteratively in order to sample the next potential optimal solution candidates. On the other hand, the performance of the tdASA technique has been moderate to low. This is somewhat expected because, traditionally SA based algorithm do not perform well in problems that involve continuous parameters [186]. Finally, the average deviation of the DIRECT method although it was expected to be poor it was not, surprisingly, the worst method employed in the reduced parameter set case. Probably, its good performance emanates from the low parameter dimensionality of the inverse problem.

Summing up, our findings for the statistical convergence of predictions in the previous paragraph indicate that the assessment of the reduced parameter set's values may substantially convergence to unique solutions. To this end, the performances of the DE and the CRSwCH methods have been found superior to the rest that were employed. Essentially, then the computation of the four most important biological parameters, is possible with highly accuracy and reproducibility.

5.6 SUMMARY

The goal of this chapter was to examine the solvability of the inverse problem through the fitting of the model predictions to experimentally collected signals. Normally this would have been performed by fitting the model predictions to dynamic optical signals. However, the actual biological background of the optical signals cannot be measured *in vivo*. Hence, a set of experimental signals has been artificially generated through simulation of the model's differential system. Then, fitting of these pseudo-experimental (PEX) signals has been performed by employing several of global optimization (GO) methods. The fact that the parame-

ter values of each of these signals were known beforehand allowed for the calculation of the respective estimation errors. In this way, the assessment of whether the model parameter values can be predicted in a unique and reproducible manner through fitting of the model's output to the artificial dataset was rendered possible. Finally, using statistical analysis we verified whether the solution of the inverse problem (a certain set of biological parameter values) can be substantially unique under ideal, noise free, conditions. On average, the deviation of the predicted values from the reference ones has been found to be as low as 1% suggesting a high degree of convergence of an almost unique set of parameters' for a given experimental curve.

Therefore, we have presented an innovative approach for addressing the problem of estimating tissue biological parameters of high relevance to cancer progression. The presented method introduces global optimization, computational biology and DCE-OI approaches. The result of merging together these emerging sciences facilitates the interpretation of dynamic bio-optical effects and processes and enables the mining of valuable information in complex biological systems. The demonstrated results suggest strongly that the method is capable of estimating, with adequate accuracy, both structural and functional characteristics associated with the growth of epithelial neoplasia.

RELEVANT PUBLICATIONS

- [1] Papoutsoglou G., Stamatiadou M., Stavrakakis G., Balas C., "*In vivo Estimation of Functional and Structural Characteristics in Epithelial Neoplasia*", Optical Molecular Probes, Imaging and Drug Delivery (OMP) - Advances in Instrumentation or Algorithms I (OMA) Monterey, California April 4, 2011
- [2] G. Papoutsoglou, M. Stamatiadou, and C. Balas, "*In silico Modeling and Global Optimization of Dynamic Bio-optical Processes for Probing, in vivo, Biological Features of Neoplasia*," in 2012 International Conference on Biomedical Engineering and Biotechnology, 2012, pp. 332-335.

6 IN VIVO MAPPING OF NEOPLASIA RELATED TISSUE PARAMETERS

6.1 INTRODUCTION

Earlier we saw that the model's output can be largely matched with the dynamics of the generated PEX signals by the majority of the employed optimizers. Still, the corresponding accuracy (NRMSD) of the parameter predictions has not been satisfactory when the full set of the neoplasia correlated parameters are estimated. To tackle this issue the theoretical sensitivity, identifiability and estimability analyses dictated that a systematic reduction of the parameter complexity should enhance the well-posedness of our parameter estimation problem. On such basis, the fixing of the unessential biological parameters to a nominal value allowed the remarkable improvement of the prediction quality of every global optimization method that has been employed. Particularly, the Differential Evolution and the Controlled Random Search with Competing Heuristics optimizers have been performing more than satisfactory. Furthermore, statistical analysis has demonstrated that the assessment of unique and reproducible estimates can be possible.

The latter finding proves that the inverse problem is, in principle, solvable and, hence, the actual status of the biological parameters can be measured. To do so, fitting of the model's output to the *in vivo* collected DCE-OI experimental data should be performed. Then, the pixel-by-pixel estimation of the acquired parameter values renders possible the projection of their spatial distribution onto the clinical images. The resulting mappings are expected to improve the assessments of disease progression. Therefore, in this chapter, we exploit the potentials of this procedure to become the next paradigm shift in cancer screening.

6.2 CLINICAL DATA

Let us remind first that the DCE-OI technology described in paragraph 2.4.6 has been clinically tested in two international clinical trials where hundreds of women enrolled after being referred to the colposcopy clinic as having an abnormal pap-test. For each of these patients biopsy samples were obtained from tissue areas corresponding to the presence of AW effect. Then, the biopsy samples were submitted for histological assessment. The selective results that will be presented below have been obtained from 2 patients with LG and 2 with HG lesions, biopsy confirmed. This is done because this particular LG and HG pathology discrimination is more popular and commonly used in clinical practice. Both patients signed an informed consent and the measurements have been approved by the hospitals' ethics committee.

Figure 6-1 displays the color images corresponding to the LG and HG cervical epithelia cases chosen. These are static images that reflect the observation of the cervical tissue from doctor's position. The top row refers to the high grade cases and the bottom to the low grade ones. The numbered circles within the images point out the areas from which the biopsy samples (BS) were obtained. Table 6-1 shows the corresponding histological results that have been assessed by two separate doctors.

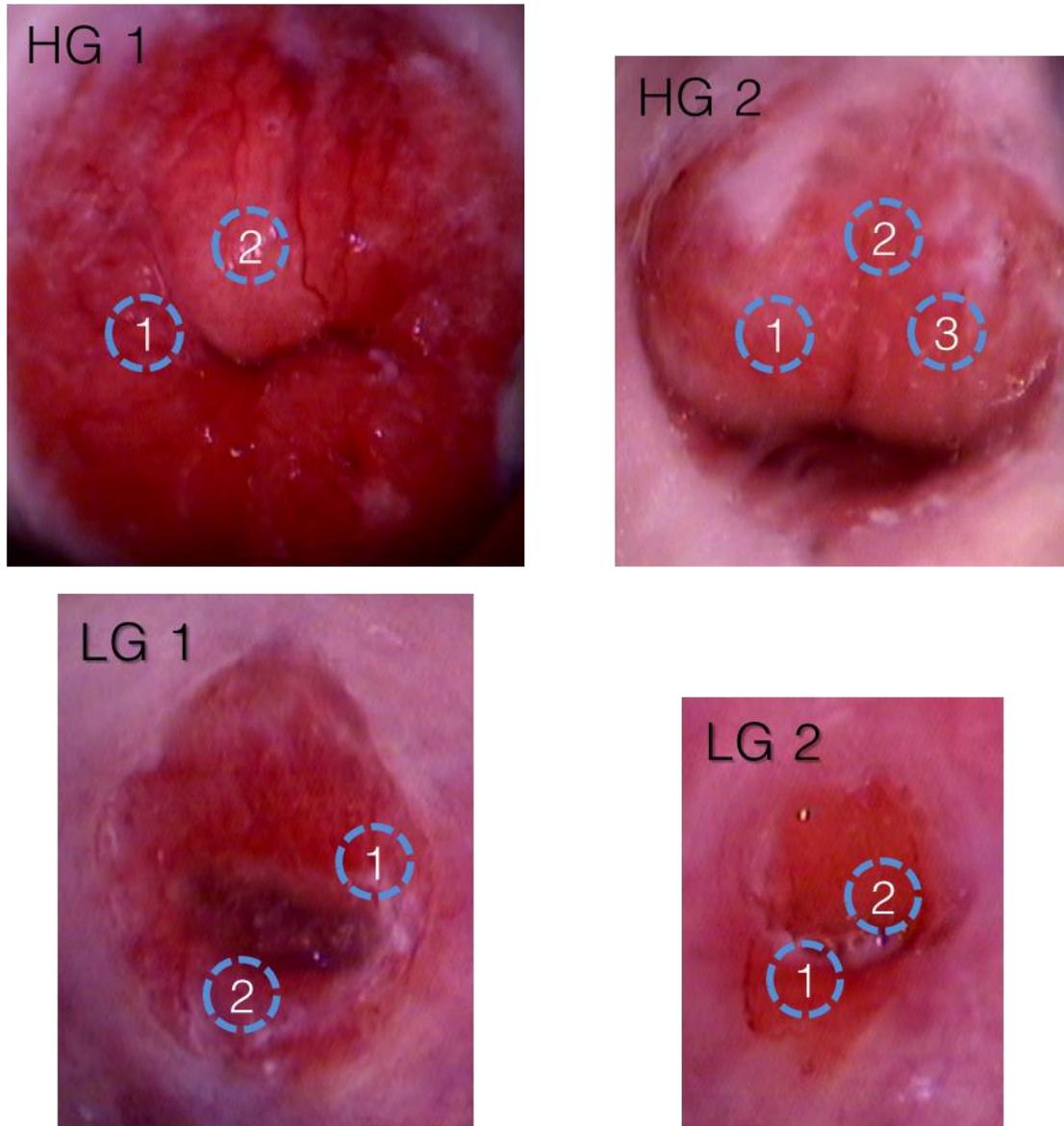


Figure 6-1 Color images corresponding to LG and HG cervical epithelia which have been measured according to the procedure in paragraph 2.4.6. These images refer to the doctor's observations. The top row refers to the high grade cases and the bottom to the low grade ones. The numbered circles within the images point out the areas from which the biopsy samples (BS) were obtained.

Table 6-1 The result of the histological assessment corresponding to the cases displayed in Figure 6-1.

Sample	Low Grade		High Grade	
	<i>LG1</i>	<i>LG2</i>	<i>HG1</i>	<i>HG2</i>
<i>BS1</i>	CIN I	CIN I	CIN III	CIN III
<i>BS2</i>	CIN I	CIN I	CIN II	CIN III
<i>BS3</i>	-	-	-	CIN III

6.2.1 Systematic and random error

So far our concern has been focused on the existence of a unique solution for the parameter set under an idealized environment where the data are noise-free and the model structure is perfect and complete. As indicated in paragraph 1.7.3.4 the practical identification is a natural, quantitative, extension of the theoretical one, where the difference lays in that it seeks for uniqueness and accuracy in the model predictions under real experimental conditions where noise is commonly present. Then, assuming that the developed model is correct, it is common for experimental data to be subjected to error due to disturbances or bias to the system variables and/or errors during the measurements.

Assuming that the noise in the measurements is additive, identically distributed, uncorrelated and normally distributed, data smoothing is the best known reliable solution that may significantly reduce it [209]. On the hand, as indicated in paragraph 1.7.3.2 the disturbances or biases on the observables can either be neglected if their effect on the data is assumed minimum or be passed on as an additional noise factor embedded to the measurements. Under such context, a double reading of the normalization described in paragraph 4.5.2 re-

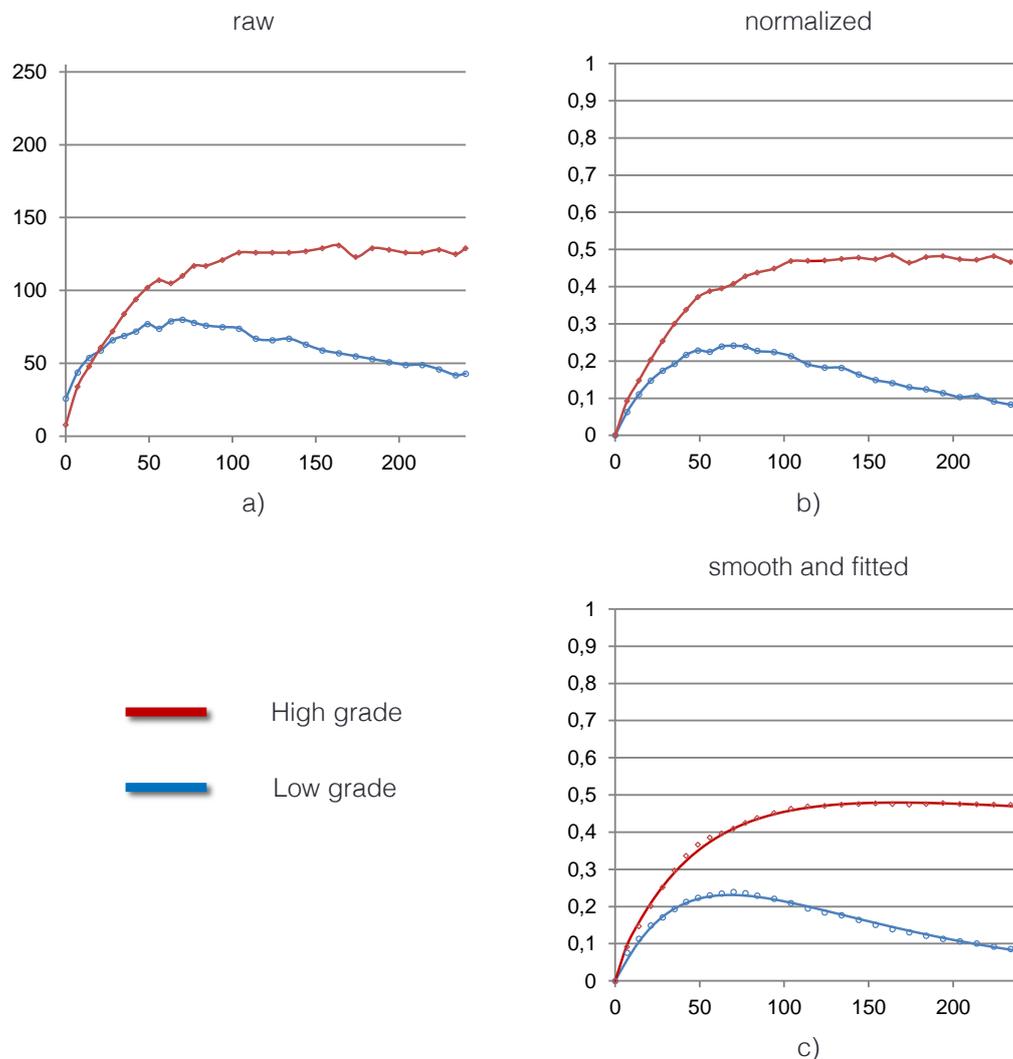


Figure 6-2 The preparatory procedure of the experimental data. (a) Raw experimental curves. (b) Normalized experimental curves. (c) The smoothed (circles, diamonds) and the biexponential fitted curves. Notice that the initial raw intensity is different and that the features are noisy (a). Then, the preparatory step alleviated the unwanted effects bringing no modification on the features of either signal (c).

veals that its effect on the experimental data has been twofold. Primarily, it permits the fitting between the model's predictions with experimental data. At the same time, however, it makes the data from different tissue areas and patients comparable without worrying that the revealed curve features are due to systematic variation. This means that it largely excludes the system disturbances from the measurements. On the basis of such dual interpretation we can reasonably claim that the normalization procedure alleviates the experimental observations from unwanted effects such as the bias.

To demonstrate these claims Figure 6-2 illustrate two typical low grade (blue) and high grade (red) raw signals, taken from the above indicated biopsy sites, before normalization is performed (left); after the normalization is performed (middle); and after the data smoothing (right). For the latter, after smoothing with a moving average filter, a biexponential function has been fitted to every experimental signal. Clearly, the preparatory step brought no modification on the features of either signal. This allows us to assume that the bias in the transformed signals is negligible and the respective error has been considerably eliminated. Hence, the wealth of biological information that is disseminated through the experimental data should have been well preserved and, therefore, prediction of the underlying neoplasia specific parameters based on the transformed values is rendered meaningful.

6.3 FITTING OF THE CLINICAL DATA

Our first attempt to fit the clinical data was described in paragraph 4.5.3.3 where a set of 30 NDR vs. time curves have been employed in order to validate the proposed dimensionality reduction. For this, a large pool of simulated model outputs was allocated and one hundred best fitting model curves from that pool were assigned, by comparison, to each CIN class. The statistics of this qualitative study showed that neoplasia and the median fluctua-

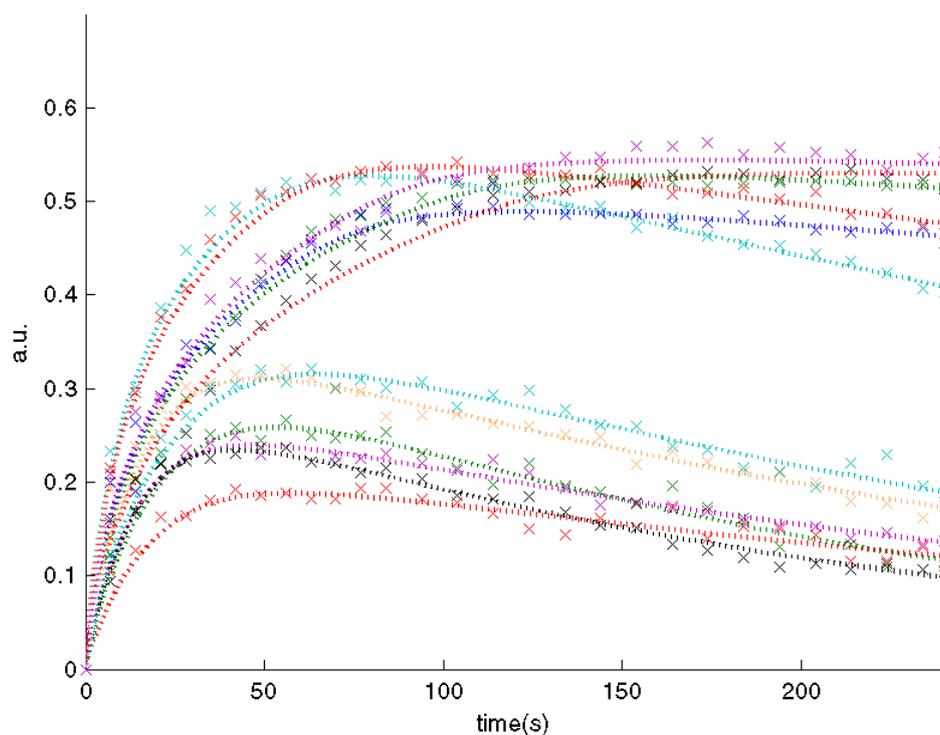


Figure 6-3 Twelve of the collected NDR curves. The scatter lines correspond to the smoothed normalized curve; and the dashed lines to the model's prediction. The average R^2 for the LG cases was 0.89 and for the HG ones 0.96.

tion of the estimated parameter values progress in a significant manner. However, the prediction uncertainty was also significant indicating that more efficient and mathematically rigorous methods for solving this problem are needed. Then, in chapter 5 we have demonstrated that the inverse problem can be solvable with high precision using several GO methods. This has been the first validation of the proposed methodology. Here, to obtain a solid picture of the correlation between the highest ranked model parameters and the progresses of neoplasia from LG (CIN I) to HG (CIN II/III) the fitting of a set of NDR vs. time curves has been repeated with the aid of a global optimizer.

Towards this direction, we have chosen several representative DR curves from each biopsy point of the cases mentioned earlier. In total, forty (40) dynamic signals were collected. Figure 6-3 displays twelve of the collected NDR curves (blue circles). Each of these timeseries data has been prepared according to the procedure described earlier and then model predictions were fitted to the transformed data using the CRSwCH optimizer.

In Figure 6-3 the smoothed analogues (scatter x-marker lines) and the respective model predictions (dotted profiles) for the subset of the selected LG and HG cases are also shown. Our results indicate that the average coefficient of determination (R^2) for the LG cases were 0.89 and for the HG ones 0.96. This shows that the model has adequately predicted the dy-

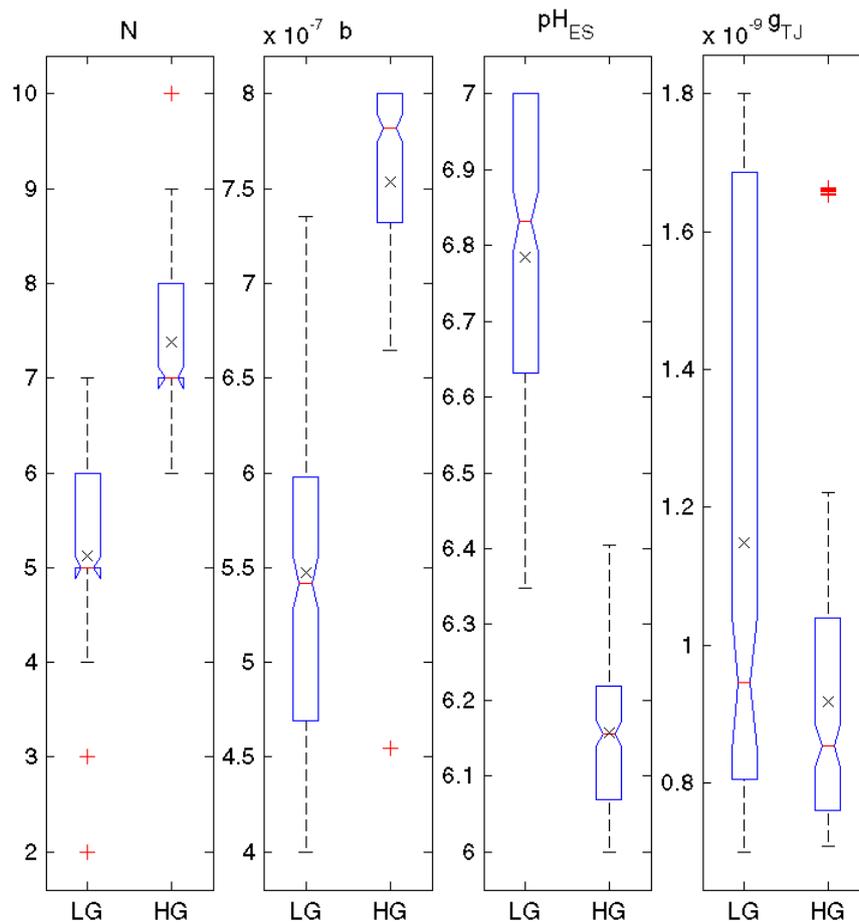


Figure 6-4 The distributions of the estimated model parameter values calculated by the CRSwCH global optimizer. Data refer to both clinical sets (LG/HG). The x's mark the mean values. On top of the axes the parameter that the data refers to is indicated.

namics of each case. Then, Figure 6-4 summarizes the distributions of the model parameter values that were calculated by the CRSwCH optimizer referring to both clinical datasets (LG/HG). The x's mark the mean values. As it can be seen, the number of dysplastic layers that are estimated for both the LG and HG biopsied lesions is fully consistent with the histological classification and are lying within the nominal value ranges found in the literature [113]. In fact, their average number increases by almost 50% between the LG and the HG cases and the difference between the medians of the two groups is statistically significant. This finding proves the correlation of neoplasia progression and the parameter that expresses the number of neoplastic layers according to our developed model. Subsequently, it is strongly suggested that neoplasia progression can be, in principle, predicted from the estimated values of the structural parameters such as the dysplastic cell packing. It is worth mentioning, again, that the number of the dysplastic epithelial cell layers is until today visually estimated in selected biopsy points and it is the basic criterion for grading the lesions. It is therefore evident that the non-invasive method presented in this thesis can comprise a novel optical biopsy scheme generating grading maps of the lesions.

As far as the increase of the size of the ES (b) is concerned, Figure 6-4 illustrates a moderate to high correlation with neoplasia grading. The overlapping in the estimated values of b has also been shown in Figure 4-12. There, no pure statistical significance in the predicted values has been computed. Here, although the overlapping reappears the statistical difference between the medians is significant. This shows that the restructuring performed combined with the use of rigorous optimization algorithms has improved the detection cut-off of this parameter. More importantly, however, this demonstrates the consistency of our predictions with literature [113]. Potentially, the mapping of this structural parameter can be utilized, along with N , to designate abnormal tissue areas.

On the other hand, the fact that our approach is based on live tissue imaging makes, in principle, the assessment of functional characteristics possible, in contrast to histology, which uses dead tissue samples. To this end, as it can be seen in Figure 6-4, the estimated pH_{ES} values for the LG cases are significantly higher than the ones estimated for the HG cases. Again, this is consistent with nominal value ranges [182]. As we have already stressed in paragraph 4.5.3.3, the ability of the method to estimate pH_{ES} concurrently with structural features in epithelial tissues gains exceptional importance in view of the so-called "acid-mediated tumor invasion model". As a reminder, this model suggests that in order for the tumor cells to invade the surrounding tissues, they have to create a toxic microenvironment for the normal cells. Primarily, this is realized by use of their unregulated glycolytic rates to their advantage by substantially reducing intratumoral pH_{ES} . Then, according to our results the tendency of HG cells to reach out to neighboring tissues is clearly evident. Hence, the described method could potentially estimate and map pH_{ES} values *in vivo* as an additional criterion for grading neoplasia. More importantly, mapping of pH_{ES} values will eventually provide a means of assessing the risk for invasion in epithelial neoplasia.

Lastly, the altered tissue porosity due to the tight junctions reshaping seems to follow a decreasing trend as the neoplasia develops. According to the confidence of the median values defined by the notches, there is a statistical difference between the LG and HG cases. This is consistent to our initial assumptions. However, the box percentiles highly overlap indicating no pure statistical significance in the difference of the predicted values. This was also indicated in Figure 5-17 where we assumed that the changes of this biological characteristic should lay below the detection level of our method. Even so, in the previous chapter, we have shown that global optimization methods can predict the value of this parameter with

high precision. It follows that this parameter is probably little correlated with the progress of the disease even though it is highly ranked according to our GSA results. The obscure relation between the tissue porosity and the CIN progression has been aforementioned in paragraphs 2.4.2.2 and 3.4.2.2. In fact we have already indicated that there is no solid literature evidence that justifies whether the solute permeabilities increase or decrease due to CIN progression. It is probable that our initial assumption may have been incorrect. However, the fact that the overlapping in the predicted values happens to occur also in the close related structural parameter that refers to the size of the extracellular space (b), is very interesting. Perhaps, the lumping of these two parameters into one hyper-parameter will provide a better understanding to the biophysical status of the paracellular route during CIN development.

6.4 ESTIMATION AND MAPPING OF THE NEOPLASIA-RELATED TISSUE PARAMETERS

The results just described have determined that the model parameters concerning the number of neoplastic layers (N), the size of the extracellular space (b), the extracellular pH (pH_{ES}) and the tight junction functionality (g_{TJ}) are correlated with clinically detected cervical neoplasia of different grades. This clearly shows the capability of our novel methodology to estimate and map both structural and functional, neoplasia-correlated parameters. On such basis, we claim that the biological parameters that can be estimated through this method from the DCE-OI and the derived DR vs. time curves are realistic and reflect the actual status of the structure and functionality of the tissue.

6.4.1 The fitting procedure

Under these circumstances fitting of all the NDR curves of each clinical case and mapping of the calculated parameter estimates depicted in Figure 6-1 can be readily performed. As proved, the DE and CRSwCH have been the most accurate and precise algorithms for estimating the parameter values of the PEX data. However, their curve fitting efficiency (Figure 5-13b) demonstrated that their nominal running times for each trial have close to two minutes. Then, according to the most ambitious scenario the optimizer should be allowed to search for the best possible fit at least for 1.5 minutes. It follows that the time that it would take to complete the procedure for a 250×250 pixel area of a single case would approximate 65 days which is unrealistic to perform.

To surpass this limitation and compute the mappings of the experimental data efficiently a simple subsampling operation followed by value interpolation has been employed instead. The procedure has been as follows. In the beginning, a set of 1000 pixel coordinates were collected from each case. Coordinate selection has been performed by low discrepancy sampling of pixels in the horizontal and vertical axis of the image using the Sobol's number sequences discussed in paragraph 4.4. Next, each of the acquired timeseries data has been normalized and then model predictions were fitted to the transformed data using the DE optimizer. The sets of parameter values that were estimated from this fitting were reinstated to their corresponding image coordinates thus providing four separate sparse matrices. Then, as far as the parameter values of the intermediate, not sampled, coordinates is concerned linear interpolation has been employed.

6.4.2 High grade results

As just described, each curve fit of the NDR profiles with the model's predictions provides four biological parameter matrices. Based on these tabulated values parameter mappings have been formulated and overlaid onto the clinical green channel image of the cervix

for visual inspection. The pseudocolor images resulting from this procedure are given below. Figure 6-5a and Figure 6-5b display the four maps expressing the biological parameter value ranges and their spatial distribution, calculated for every pixel of the images referring to the HG 1 and HG 2 cases. As a footnote, in each figure, the black circles denote the tissue areas from where the biopsy samples have been acquired. In addition, the parameter mapping that each of the four pseudocolor images refers to is indicated in the lower left corners. The top rows depict the spatial distribution of the structural parameters (N , b) and the bottom rows depict the spatial distribution of the functional parameters (pH_{ES} , g_{TJ}). Lastly, the color-coding of the maps, which is given below and above these images, corresponds to various parameter values. Notice that the color palette for the pH_{ES} parameter has been reversed in order to be consistent with its reverse correlation with disease progression. We should also notice that the pixels with values for parameter N above 10 (layers) have been coded with the same color.

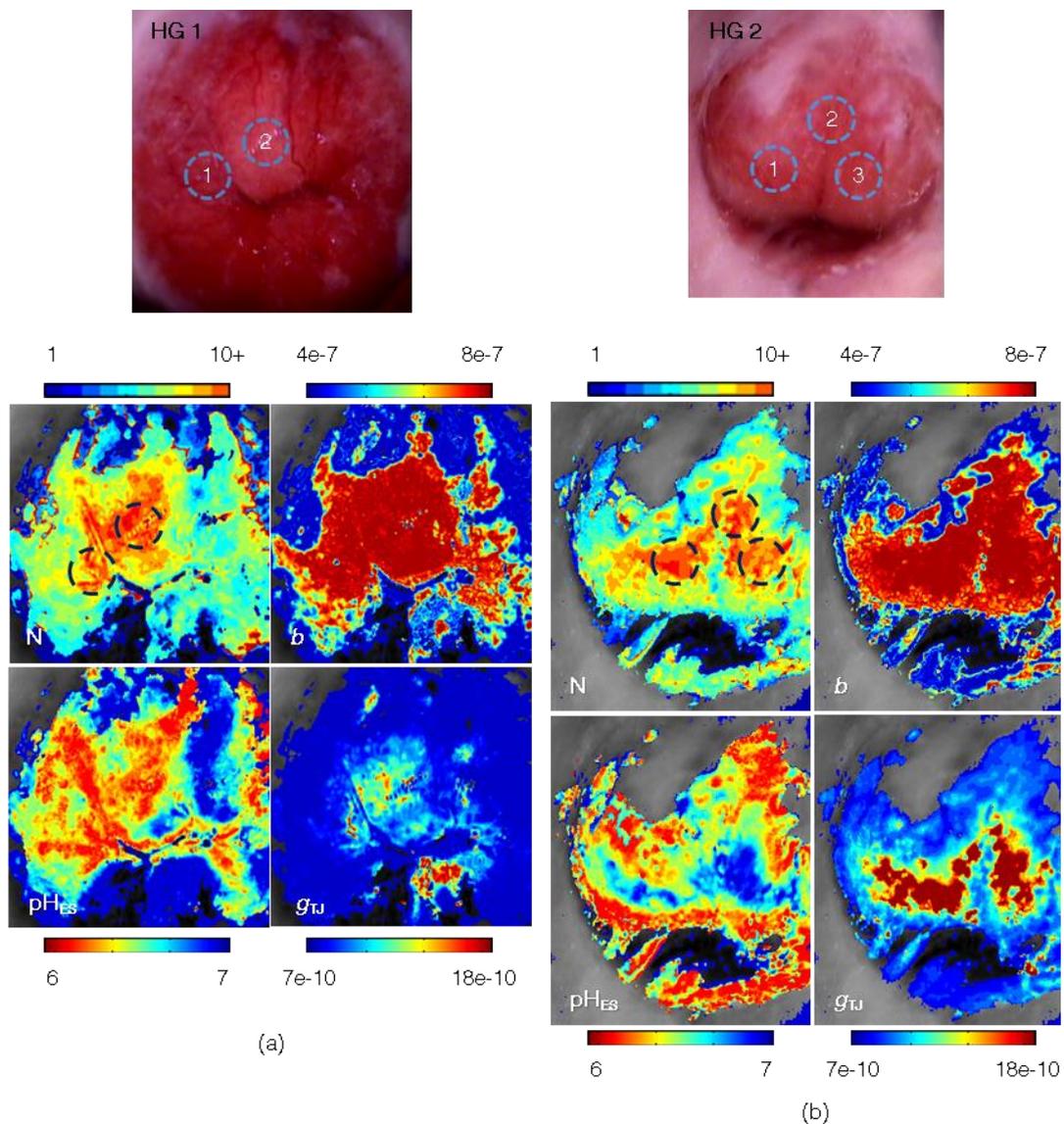


Figure 6-5 (a), (b) The pseudocolor maps of N , b , pH_{ES} and g_{TJ} neoplasia-related parameters of the HG1 and HG2 cases of cervical epithelium, respectively. The circles denote the areas from where biopsies have been collected. Color-coding of the maps corresponds to various parameter value ranges.

Figure 6-5a.N and Figure 6-5b.N depict the spatial distribution of the structural parameter expressing the number of neoplastic epithelial layers. It is clear in both figures that this parameter takes the maximum values (8-10+) at the points where the biopsies had been taken and histology results suggested high occupation of the epithelium by neoplastic layers. As discussed several times previously, it has been established that the number of neoplastic layers are increasing with the neoplasia growth [113]. Reasonably then, this can be considered as a confirmation of the validity of our results.

On such ground, one can easily notice that the mappings of the layer number parameter delineate in both HG cases the areas at risk and subsequently the size of the lesion. The size of the lesion is also known to be related to the grade of dysplasia in histology and studies have shown that it has an independent predictive clinical value. Particularly, it has been shown that if the size of the lesion is used as an additional criterion in colposcopy assessments, the accuracy of the examination is improved and fewer lesions are underestimated concerning the grade of dysplasia [217]. In another study, it was shown that in women having cervical smears with severe dyskaryosis (potentially malignant cells with abnormal nuclei) the size of the lesions were larger than among those with mild dyskaryosis [218]. Furthermore, the mean size of HG lesions containing parts of microinvasion has been shown to be 7-fold greater than that for severe dyskaryosis without invasion and 100-fold greater than for mild dyskaryosis [219]. Recall here that the AW phenomenon occurs in irregular cell nuclei and specifically due to the reversible coagulation of the epithelial and stromal cytokeratins. That said, since our predictions are based on the acetowhite dynamics we have just demonstrated the direct relation between the sizes of the acetowhite patches of the tissue, the volume of the nuclei and the predicted number of abnormal cells. On the other hand, from the color-coding one can determine the margins of the atypical cervical epithelium. Clinically, atypical cervical epithelia are graded according to a number of features, which include sharpness, conformation and thickness of the border and the presence of internal margins in a lesion [220]. Particularly it is common for the margins in HG areas to exhibit distinct raised edges that may also be located within a larger low-grade lesion. This is exactly shown in Figure 6-5a.N and Figure 6-5b.N where the HG areas, coded in red, are situated within lower lesions coded in yellow. The latter are surrounded by even lower lesions (green-yellow) which in turn are situated within an area of atypical, probably mild dyskaryotic cells (light blue). Clearly, this shows the potentials of our novel methodology to discriminate within the group of HG lesions.

Next, Figure 6-5a.b and Figure 6-5b.b depict the spatial distribution of another key structural parameter, which is also assessed histologically as having high predictive value. Recall here that parameter (*b*) expresses the size of the extracellular space, which is also known to increase with the neoplasia progress [113]. As it can be seen in the maps of parameter (*b*), accumulation of large extracellular spaces ($\sim 0.8\mu\text{m}$) exist in the vicinity of the biopsy confirmed high grade points. Additionally, high *b*-parameter areas are co-localized with the high *N*-parameter areas, something that can be considered as a verification of the consistency of our findings with the actual biology of neoplasia growth.

Similarly to the mappings of the layer parameter the respective ones of the ES size parameter clearly delineate the tissue area within which the AW phenomenon occurs. This anatomical zone is the squamocolumnar junction (known as the transformation zone) where CIN and invasive cervical carcinoma commonly arise, and, therefore, is a major focal point of the colposcopic examination. If it is only partially seen, the examination is deemed to be inadequate or unsatisfactory for the purpose of ruling out CIN and invasive carcinoma. Hence, this

parameter can be used to outline macroscopically the tissue areas which have the potentials to develop a malformation. In addition, one can notice fine cobblestone patterns surrounded by a red margin and red, continuous patches in the color-coding of the b -parameter areas. These demarcation patterns resemble to the two atypical vascular patterns found within abnormal epithelium, either separately or in combination, which are better known as punctuation and mosaicism [221]. Although AA staining is expected to obscure the underlying intraepithelial atypical vessels, it is often the case that these vessels may appear as an indication of more severe lesions or invasive cancers [220]–[223]. According to the histology of the cervix, punctuation consists of dilated and often twisted, irregularly terminating vessels while mosaic pattern comprise of capillaries arranged in a honeycomb like configuration. An important factor in assessment of mosaic pattern or punctuation is the intercapillary distance. This refers to the distance of two adjacent vessels or to the diameter of the mosaic-like vessels. During the increase of the pre-invasive or malignant nature of the lesion these capillary strands become wider, coarser and more irregular [220], [221]. In fact, the combination of all these factors, sometimes in addition to avascular whiteness appearing within the epithelium, indicate early invasion and as malignant cells proliferate these avascular areas get larger [222], [223]. Another possible explanation for these patterns is that they may designate the paracellular pathways through which AA diffuses. Similarly to the mosaic pattern just described, the thin, well separated web may designate a fine diffusion around the cells of the epithelium. Conversely, the coarse red patches seen in Figure 6-5a. b and Figure 6-5b. b may demonstrate a situation of undetermined or bulk diffusion of the AA into the tissue. In either case, the fact that these atypical patterns appear in the mappings of the b -parameter in both figures further illustrate the consistency between the characteristics of neoplasia growth and our predictions.

As it can be seen in the mapping of the N -parameter the number of dysplastic layers is fully consistent with the histological classification and are laying within the nominal value-ranges found in the literature. It is therefore evident that the method presented in this thesis may comprise a non-invasive a novel optical biopsy method. Principally, the fact that our approach is based on live tissue imaging makes the assessment of functional characteristics possible, in contrast to histology, which uses dead tissue samples.

Referring to the functional parameters, then, our consistency claim applies also to the findings illustrated in Figure 6-5a. pH_{ES} and Figure 6-5b. pH_{ES} where areas with high, close to normal, pH_{ES} values are color-coded with blue and areas with low pH_{ES} , values with red. Particularly in Figure 6-5a. pH_{ES} and Figure 6-5b. pH_{ES} it is distinctly shown that lower pH_{ES} are co-located with areas of high N and b values depicting the lower acidity of the extracellular space, which is in full agreement with the finding of other studies [98], [151], [182]. Recall here that tumor cells in order to proliferate elevate unregularly their glycolytic rates. This results in an unstable acidic microenvironment at the interface with the normal tissue. Due to the fact that normal tissue lacks the adaptive traits of tumors e.g. the increased H^+ transport against concentration gradients and mutations in acid-induced apoptotic pathways, it is vulnerable to acid-mediated toxicity. This includes cell necrosis and apoptosis, and degradation of the extracellular matrix. It follows that hypoxia and acidosis play a key role in the progression of tumor from in situ to invasive cancer [97]. It is reasonable, therefore, to assume that the pseudocolor landscapes of this parameter may also show the underlying extensiveness of the malignancy, the phenotypic properties of its propagation to the surrounding tissues and subsequently its growth from primary to metastatic cancer.

What is noticeable in Figure 6-5a. pH_{ES} and Figure 6-5b. pH_{ES} is that at the posterior lips of the cervix low pH_{ES} values have been predicted (~ 6) while the N -parameter is also low (~ 4). It follows that the optical signals from these areas are described by small amplitudes and prolonged dynamics. Although this may seem as an inconsistency in the results there are three possible reasons of it happening. First, is that the vicinity of the posterior lips is commonly occupied by single layered, columnar epithelial cells. Malignancies may also develop in the areas of columnar cells. However, our model does not conform to this type of tissue. Even so, accepting that in these areas are single-layered and thus only single numbers of abnormal cell layers are able to develop then AA will rapidly diffuse towards the basement of the tissue. There it will stagnate until it is properly drained to the underlying vasculature. This, however, maintains its presence in the ES of the cells and, in turn, prolongs the intracellular acidification. Consequently, the fact that columnar structured cells appear at the lips of the cervical epithelium is highly probable to impose this inconsistency. The second explanation to this phenomenon may be the existence of a small ulceration which also acetowhites and is known to be the result of nodular or polypoid lesions [217]. Third and last is that the optical signal from the tissue in these pixels may be entangled with the optical signal *due to the foam* from the chemical agent that is common to appear during colposcopy. A fine example of this happening is more apparent in Figure 6-5b. pH_{ES} where two irregular stripes of large pH_{ES} values extend between the 7 and the 8 o'clock positions of the cervix. To validate this further a close examination of the color image of the HG 2 case displayed in Figure 6-1 above shows clearly that these white stripes do not correspond to an optical signal that carries information about the biomarker-tissue interactions. This shows that, the optical signal at the periphery of the phenomenon has great chances to be severely amplified by external mechanisms which, in turn, can disorient our results.

Finally, Figure 6-5a. g_{TJ} and Figure 6-5b. g_{TJ} show the mappings of the permittivity of the tissue to the biomarker. It is generally accepted that, in a high-grade CIN there will be an almost instant response to the application of AA. This is clearly illustrated in Figure 6-5b. g_{TJ} where the values of g_{TJ} maximize ($> 15\text{\AA}$) in the areas of the biopsy points. Conversely, the permittivity of the tissue in the HG 1 case (Figure 6-5a. g_{TJ}) is moderate ($10\text{-}11\text{\AA}$) throughout the most of the tissue's area. Probably the grade of the lesion is not as high as the N -mapping has indicated. In fact, consulting the histology outcomes shown in Table 6-1, we notice that the HG 1 case included both severe (CIN III) and a moderate (CIN II) dysplasia. It is therefore possible that this parameter can be able to distinguish neoplasia among the high grade cases which is extremely difficult for practitioners to classify. Even so, we should keep in mind that this parameter has not shown a strong statistical correlation with neoplasia progression according to the fitting of the NDR curves with the global optimization method. Hence, it should be used with caution when inferences on the severity of the disease are made.

In general, CIN carcinogenesis disrupts the state of the tissue adhesion structures, which has been associated with increased tissue permeability [122]. This leads to the loosening of these particular junctions, transforming the tissue from tight to leaky increasing the possibility of metastasis. Particularly, according to the "acid-mediated tumor invasion model" mentioned earlier, the H^+ flow to peritumoral normal tissue provokes normal cell necrosis or apoptosis and extracellular matrix degradation [98]. Because, the tumor cells are capable of resisting to the toxicity induced from this flow, they are able to invade the damaged normal tissue. This allows them to spread, and eventually form invasive cancers. Therefore, the

combination of both contours corresponding to the functional parameters suggests strongly that our method can provide a better insight to the neoplasia growth and tumor metastasis.

6.4.3 Low grade results

Figure 6-6a and Figure 6-6b display the spatial distributions of the value ranges of the four biological parameters, calculated for every pixel of the images referring to the LG 1 and LG 2 cases. For the presentation of these mappings the previous formalism has been employed. Then, Figure 6-6a.N and Figure 6-6b.N denote the pseudocolor maps of the structural parameter expressing the number of neoplastic epithelial layers. As seen, there is no sign of severe malformations rather a landscape of moderate cell packing (4-5 layers). Only in Figure 6-6a.N a higher cell packing is illustrated (6-7 layers) at the 6 o'clock position that refers to the point where one of the biopsies had been taken and histology results suggested a CIN I epithelium. This is not totally unexpected because according to Figure 6-4 some overlapping does appear in the distributions of the estimated values of N when the selected NDR curves of the LG and HG cases were fitted with the global optimization technique. Even so, the comparison between the previous mappings of the layer parameter and the current ones clearly demonstrates that significantly risky lesions do not appear in either LG case. Additionally, no distinct sharp edges or significantly large tissue patches appear in the maps

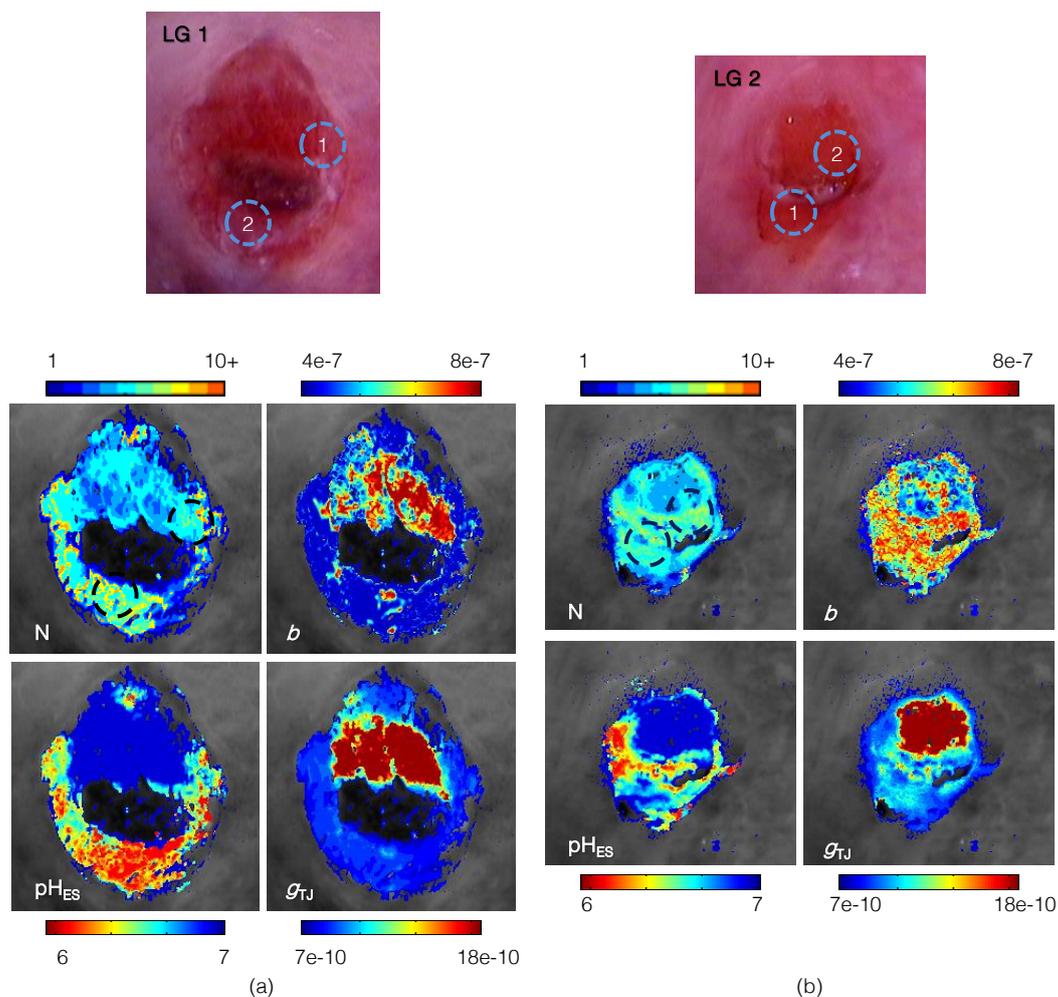


Figure 6-6 (a), (b) The pseudocolor maps of N , b , pH_{ES} and g_{TJ} neoplasia-related parameters of the LG1 and LG2 cases of cervical epithelium, respectively. The same notation as in Figure 6-5 has been employed.

as they have in the HG images above. These findings exemplify the consistency in our results. Furthermore, by the time the number of neoplastic layers is until today the basic criterion for grading a lesion they prove that there is a direct correlation of our predictions with the actual biology of neoplasia growth.

A possible reason for the moderate cell packing seen in the LG tissue may be a mild dyskaryosis. As already indicated, dyskaryotic cells have malignant potential. Nevertheless, most cases of mild dyskaryosis are considered LG by physicians and, for this, will typically be followed up and referred. The risk becomes increasingly high only with severe dyskaryosis which is probably not the case here. On the other hand, the margins of low-grade lesions are usually described as irregular, geographic, and indistinct [224]. This undetermined pattern is clearly illustrated in the *N*-color maps of both LG figures. Additionally no sharp, symmetrical edges appear in the topographies. Therefore, we have no reason other than to support that the dull and indifferent color mappings in Figure 6-6a.*N* and Figure 6-6b.*N* indicate lesions of low risk.

Figure 6-6a.*b* and Figure 6-6b.*b* reflect the predicted spatial distribution of the relative structural parameter that corresponds to the size of the ES. As before, the highest *b*-parameter areas (0.7-0.8 μm) are co-localized with the highest *N*-parameter values (5-6) predicted for the biopsied tissue areas. Notably, at the 6 o'clock position of the LG 1 image (Figure 6-6a.*b*) the only patch of very high ES values exists exactly where the biopsy sample had been taken and histology results suggested a CIN I epithelium. This shows the sensitivity of our novel method to point out potential abnormal areas. Furthermore, and unlike the previous HG cases, regions of low and very low ES (0.4-0.5 μm) within the acetowhite epithelium also become apparent here. This shows the great detail of these mappings with which can represent the status of the epithelium and the potential specificity of the method. On the other hand and predominately in the LG 2 example few distinct demarcation patterns appear that may imply the existence of atypical vessels according to our earlier discussion. However, it must be noted that punctuation and mosaic pattern may also be found in normal epithelium (e.g. in inflammatory tissues) but are usually not excessive. The fact that the patterns here are of several color codes, significantly smaller than in the HG examples, irregular and rather indistinct it is more likely that they represent a benign process (metaplasia) or LG CIN. Therefore, all these findings manifest that our results concerning the structural parameters are coherent and in agreement with the progression of cancer in the tissue.

Quite interestingly, however, some enlarged ES (>0.6 μm) appears in these LG mappings even at areas where lower *N*-parameter values (3-4) have been predicted. Fine examples of this are seen at the 6 o'clock position of the LG 1 case and at the 5 o'clock position of the LG 2 case. Such condition is assumed to hold, at least mathematically, because the distributions of *b*-parameter estimates for the LG and HG NDR curves are partially overlapping (Figure 6-4). According to literature the structure of the intercellular space of the ectocervix may be modified by several pathogens [120]. These pathogens predominately disrupt the adhesion molecules (TJs) of the cells in order to infiltrate into the tissue. It is reasonable therefore that enlargement of the size of the ES may originate from this mechanism. Another possible mechanism is that the stratified squamous epithelium of the cervix is aglandular and if fluid is secreted or absorbed for any physiological reason this should be realized through the ES. To facilitate this transport the epithelium the TJ functionality is subsided, temporarily. Then, the volume of the ES enlarges and the physiological epithelium becomes "leaky" [225]. This means natural fluid secretion or absorption may be another reason that explains the increased values in the *b*-mappings at possibly normal areas of the LG images.

Figure 6-6a. pH_{ES} and Figure 6-6b. pH_{ES} display the predictions of our method concerning the extracellular pH of the tissue for the LG 1 and LG 2 clinical examples, respectively. By the descriptions thus far one would expect the color-coding of this functional parameter to be fainter in the LG cases. Instead, we see that low pH_{ES} values close to 6 have also been computed here. One reasonable explanation of this would be the existence of a columnar epithelium which was discussed earlier. Another explanation for this inconsistency could be the similarity of the features passed on the dynamic signals by two different biophysical mechanisms. As a reminder, a low pH_{ES} value implies that more AA molecules enter the cell provoking the acidification of the cytoplasm and at the same time fewer protons can be cleared by the pumping mechanisms. This induces prolonged acidification that passes onto the temporal characteristics of the measured signals as persistent whitening. At the same time, it is generally accepted that in LG neoplastic tissues, the onset of acetowhitening is delayed because the acid must penetrate into the lower half of the epithelium where the atypical cells are situated [226]. This also passes on the dynamics of the measured signals as persistent whitening which resembles to that of the prolonged acidification, but it is of significantly lower intensity. Based on this similarity, therefore, our predictions of low pH_{ES} values in LG areas are reasonable but exaggerate. Then, the key to discriminate between the HG and the LG cases based on pH_{ES} values is in the consultation of both N -parameter and pH_{ES} -parameter contours. Thus, a more thorough look at the pH_{ES} results in the LG examples would read that low pH_{ES} values (<6.5) are co-locating with the low-to-moderate N values. Hence each of these cases refers to a LG lesion. Probably the lumping of the biomarker-normal tissue interaction in the reservoir compartment may address, to some extent, the disentanglement of the aforementioned biophysical mechanisms.

Finally, Figure 6-6a. g_{TJ} and Figure 6-6b. g_{TJ} illustrate the spatial distribution of values predicted for the tissue's permittivity to AA. As seen in these images, there are roughly two heterogeneous regions of tissue, the upper part of the ectocervix that shows significant absorption of AA and the respective lower part that shows low and moderate absorption. As discussed earlier, however, the correlation of the g_{TJ} with neoplasia progression is mathematically weak and still unclear to us. This means that reliance on this parameter as a sole indicator of a HG or LG area can be misleading. To improve, therefore, our insight to the progression of the abnormalities from LG to HG, predictions for the tissue's diffusivity should be combined with those for the rest of the biological parameters.

On such basis one can notice that the pseudocolor maps of the functional parameters have somewhat complementary nature here. Specifically, the biopsy locations in both LG 1 and LG 2 examples have been found to correspond to low tissue porosity and low extracellular pH. At the same time they are occupied by a moderate number of abnormal layers. This translates to a slow diffusion of AA at the presence of a small amount of cells with atypical functionality. Clearly, this is consistent with our earlier assumption that the excessively low pH_{ES} values predicted for those areas have been due to the delayed transition of AA to the lower half of the LG epithelium. It is therefore evident that, the low porosity values (blue colors) of the tissue in combination with the lower pH_{ES} values (red colors) may demonstrate the proliferation of atypical cells in LG lesions.

Under these circumstances and according to our earlier descriptions if we accept that the functional parameters reflect the proliferation of neoplastic cells, it is interesting to note the propagation of CIN to the surrounding tissue and subsequently its growth in both LG examples. Particularly, the abnormality in the LG 1 case (Figure 6-6a. pH_{ES}) seems to have initi-

ated at the lower parts of the cervix and then gradually expanded towards the upper parts. Likewise, the abnormality in LG 2 case (Figure 6-6b.pH_{ES}) appears to have been extending from the 9 o'clock position towards the center of the cervix. Additionally, no severe neoplasia has been grown yet to either case as indicated by low to moderate number of neoplastic layers (4-6). The latter is supported by the histology inferences where both biopsy samples although they indicated neoplasia it has been of low-grade (CIN I) (Table 6-1). Therefore these findings demonstrate that the mappings of the functional parameters can picture the underlying functional behavior at LG lesions and how these develop into HG ones. Since the latter is impossible to assess with standard screening techniques, our novel methodology holds the promise to become an essential diagnostic tool of neoplasia at early stages.

6.5 SUMMARY

Based on the enhancement of the posedness of the inverse problem and our global optimization results we have presented in this chapter a novel methodology for optical biopsies. Specifically, we have first utilized a series of DCE-OI clinical measurements in order to validate the correlation between the highest ranked model parameters and the progresses of neoplasia from LG (CIN I) to HG (CIN II/III). Fitting of the experimental curve profiles with model predictions has been realized using the CRSwCH method that is extremely robust and accurate for our inverse problem according to the results of the previous chapter. Our findings suggest that our predictions concerning the fluctuation of the model parameters from LG to HG cases largely conforms to the biology of the cervix during the progression of neoplasia. This has proven that our model can be, in principle, predictive. Next, by employing a simple fitting operation (subsampling followed by linear interpolation) the spatial distribution of the parameter predictions has been realized. The pseudocolor mappings that resulted from this procedure revealed great consistency with histology assessments. Particularly, our predictions for the number of dysplastic layers, which is the golden standard for histology classification, have been in total agreement with the doctors' assessments. Similarly, the interpretations of the predictions regarding the rest of the parameters are compatible with the finding of other groups for both HG and LG neoplastic epithelia.

In essence, by the time there has been no obvious contradiction between our results and literature, we can reasonably claim that our model is largely predictive and thus our novel methodology can be exploited as a new optical biopsy technique. It therefore holds the promise to become the next paradigm shift in cancer screening.

RELEVANT PUBLICATIONS

- [1] G. Papoutsoglou, T.-M. Giakoumakis, and C. Balas, "Dynamic contrast enhanced optical imaging of cervix, *in vivo*: A paradigm for mapping neoplasia-related parameters.," in Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference, 2013, vol. 2013, pp. 3479–82

7 PERSPECTIVES

7.1 THESIS OVERVIEW

Epithelial neoplasia is the precursor of invasive carcinoma. Current diagnostic chain involves the tissue's visual inspection and subsequent biopsy sampling if abnormal screening results occur and finally treatment for the precancerous abnormalities. To aid the early detection of the abnormalities optical biomarkers or contrast agents are employed. These chemical compounds have the unique property of generating measurable dynamic optical signals when they interact with biomolecular targets and processes of diagnostic importance. In live cell or tissue imaging the dynamics of the target's optical activation is regulated by the uptake kinetics of the biomarker. This implies that the quantitative monitoring and modeling of the generated dynamic optical signal can be used for assessing the biomarker's uptake kinetics, which can, in turn, provide valuable information for both functional and structural characteristics and for the associated abnormalities. A basic prerequisite for achieving this is to establish a pharmacokinetic model that accounts for all tissue structural characteristics and also for all biomarker transport phenomena in the living tissue. The second step is to fit the experimental data with the model's output characteristics and estimate tissue structural and functional characteristics by attempting to solve the inverse problem.

On such grounds, cancer of the cervical epithelium is a major health care issue. Recent estimates indicate that 493,000 new cases occur every year while 274,000 women die of cervical cancer annually. The immediate precursor of invasive squamous carcinoma is the CIN, grade I, II, III and micro invasive. Grade II or more severe lesions are classified as high-grade squamous intraepithelial lesions. Topical application of acetic acid solution 3-5% is routinely used as a contrast agent for more than 70 years in order to highlight abnormal areas. The latter is perceived as transient tissue whitening (AW effect). The sensitivity of colposcopy in detecting high grade lesions is limited to 56%, indicating that in almost half of the cases the lesions are misdiagnosed and may progress to invasive cancer.

Attempting to improve the accuracy of the clinical diagnosis of CIN, the dynamic optical characteristics of the AW effect have been measured *in vivo* using a DCE-OI apparatus. Then, in this thesis, we have presented a pharmacokinetic model of the dynamic, metabolic responses of the tissue to the perturbation that the biomarker imposes. The development has been based on the documented cancer-specific structural and functional alterations of the cervical tissue described in chapter 2. Particularly, due to the higher extracellular acidity met in neoplastic epithelia, the topically applied acetic acid remains considerably undissociated in the ES of the neoplastic epithelium selectively and, as such, can penetrate passively into the cell membrane of the neoplastic cells, with high selectivity. Due to the almost neutral pH in the IS, the intruded AA molecules are disassociated and the generated ions stimulate conformational changes in the nuclear proteins. This, in turn, provokes local changes to the index of refraction and determines the macroscopically observed dynamic optical signals. Next, and because the cervical epithelium is stratified, these processes are repeated in the underlying neoplastic epithelial layers increasing the intensity of the DR signal. Finally, the

tissue restores its original light scattering characteristics when the biomarker has been consumed or drained into the stroma.

Based on this conceptualization a mathematical, *in silico*, representation of the tissue has been developed in chapter 3. Particularly, the developed model partitions the neoplastic part of the cervical epithelium into a stack of abnormal cell layers, which is delimited by a reservoir layer –a repository that supplies the bio-marker. Each of these layers is modeled with two compartments: the IS and the ES compartments. AA diffuses passively from the ES to the IS through the cell's membrane and from the upper to the lower layer(s) of the epithelium through the tight junctions. The model embodies also two dynamic intrinsic processes: a) the AA ion buffering, occurring in both IS and ES spaces; and b) the active transmembrane pumping processes, which contribute in restoring the original pH_{IS} through the extrusion of AA ions from the IS to the ES. Under this context, a non-linear, mixed-integer algorithm that includes a system of scalable, coupled differential equations governs the deterministic model that we have developed. The sum of per layer calculated intracellular [H] provides the total epithelium [H] for any time point after the application of the AA. The fact that the conformational changes respond to pH changes almost instantaneously (nano/microsecond time scale) suggests strongly that the dynamic optical signals are determined basically by the [H] vs. time function. Thereupon, using the total [H] as the model's output our simulations showed for that there is a high correlation between model predictions and histological parameters of biopsy samples. Particularly, it has been found that the max DR values correlate with the packing of atypical cell layers, thus making the approximation of the latter possible, *in vivo*, and for every pixel area. Additionally, the fact that this model parameter is consistent with histology showed that through the described model, information for the functional alterations can potentially be provided, which is essential for complementing diagnosis and for improving its accuracy. Therefore, our model may comprise a promising alternative optical biopsy method for the non-invasive diagnosis and screening of cervical neoplasia.

The observed correlation between the dynamic scattering characteristics of the AW effect and the structural and functional characteristics of the epithelium enables the assessment of the latter, *in vivo*, and noninvasively. However, according to the model building walkthrough described in chapter 1, in order to be predictive a metabolic model has to be identifiable. The study of the latter was described in chapter 4. Initially, we observed that the developed model is nonlinear in the parameters. This means that the model's output consists of mathematically entangled information regarding the status of the tissue. Likely, when the solution of the inverse problem is intended, the outcome of the fitting procedure will be of poor quality. It has been essential, therefore, to investigate the *a priori* capacity of the solution to this parameter estimation problem to designate, at least theoretically, a set or subset of the model parameters in a unique and reproducible manner. Towards this direction and for the purpose of calibrating the posedness of our parameter estimation problem, we proposed and performed consecutively three *in silico* analyses; namely, GSA, parameter identifiability analysis and parameter estimability analysis. On the basis of these analyses, we were able to decide which of the pertinent model parameters can potentially be estimated from available input/output data and which are impossible to assess. Specifically, we have identified a set of parameters that fulfill the following criteria: a) they are the key determinants of the line-shape of the model's output that fits the experimental dynamic optical data and b) they display no collinearities and minimum interdependency with each other. Then, by combining GSA and theoretical identifiability and estimability theory we have shown for the first time that the curve profiles of dynamic backscattering signals are mainly determined by two structural and two functional parameters. These are the number of dysplastic layers (\mathbf{N}), the

size of the ES (b), the pH_{ES} and the fluidity of the tissue (g_{TJ}). To reach this finding the initial nine input parameters of the model were systematically reduced to these four most identifiable and estimable ones. This facilitates the solution of the inverse problem, namely the concurrent estimation of all these parameters from *in vivo* measurements and for every spatial point. These parameters are of utmost importance in clinical practice where N and b are the key parameters onto which routine histological grading of biopsies is based; pH_{ES} is strongly linked with the ability of tumor cells to form invasive cancers; and g_{TJ} represents the tight junction porosity changes due to malfunction at precancers. It is essential to stress here, that pH_{ES} and g_{TJ} are functional parameters that cannot be assessed *in vivo* by any medical procedure known thus far. Such achievements hold the promise to provide new insights in the study of epithelial neoplasia and novel noninvasive diagnostic methods and devices, both contributing to cancer prevention.

Towards this direction, the estimation of the model parameters has been attempted in chapter 5 using several GO methods. Specifically, the inverse problem was expressed as the estimation of the values of a plausible set of parameters through the fitting of the normalized DR experimental data (NDR) with the normalized output of the model. Our approach for solving this problem involved a fitting scheme that utilizes the Euclidean distance between simulated and experimental data as an objective function. This function, which describes the discrepancy of the measured and the predicted data, was iteratively minimized and the model parameter values that correspond to this minimization are expected to be the solution of the inverse problem. As it is well understood, the solution of the inverse problem is meaningful only if the employed optimizer converges adequately and estimates the parameters in a reproducible manner. We have employed and tested, from that perspective, a series of global estimation techniques that randomly search to converge to the global extrema under a recursive, evolutionary strategy. These methods benefit from the fact that they are easy to implement and not critically dependent on a priori information about the objective function. Under this framework, constrained, direct search, point-to-point and population based, global optimization algorithms have been comparatively evaluated. Statistical analysis revealed that the solution of the inverse problem is possible in a unique and reproducible manner if the DE and the CRSwCH methods are employed. Consequently, we have presented an innovative approach for addressing the problem of estimating tissue parameters of high relevance to the disease, from optical data measured *in vivo* and non-invasively. This facilitates the interpretation of dynamic bio-optical effects and processes and enables the mining of valuable information in complex biological systems.

Being able to measure the epithelium's characteristics associated with the growth of malformations from dynamic optical signals with adequate accuracy, allows for the pixel-by-pixel estimation of such biological signatures from DCE-OI of a tissue, *in vivo*. Then, in chapter 6 we described the first attempt for estimating and mapping the set of neoplasia-related biological parameters, from optical data generated by DCE-OI, *in vivo*. On the basis of the described analysis we presented, then, mappings of these biological parameters overlaid onto the clinical color image of the cervix. Our findings show that the estimated values of the four parameters are consistent with information provided in the literature. These results are unique in the sense that for the first time functional and microstructural parameter maps can be estimated and displayed together, thus maximizing the diagnostic information. The quantity and the quality of this information are unattainable by other invasive and non-invasive methods. It follows that the method described in this thesis can comprise a novel optical biopsy technique and a valuable research tool contributing to improve our understanding of

neoplasia growth and tumor metastasis. Corollary, it may become a valuable diagnostic tool that will also facilitate the development and evaluation of new cancer therapies.

7.2 FUTURE WORK

Although there is always room for some improvements the results presented in this thesis are extremely encouraging. As repeatedly said, the novel methodology which we developed and described here has the power to become the basis for several imaging modalities in which screening of several human diseases, especially cancer is involved.

7.2.1 Time complexity improvements

Prior to the clinical application of our proposed procedure there is a point that should be addressed which entails the improvement of the time complexity of the pixel-by-pixel based fitting. The best candidate for achieving this goal is through a simple table lookup operation has been employed instead. The procedure could be as follows. First, a large set of parameter values can be allocated capturing the full spectrum of their ranges. For each of these values, model simulations should then be performed and the corresponding normalized outputs be recorded in a lookup table. Then, instead of fitting each NDR curve by employing an optimization algorithm, these tabulated values may be utilized. Although the convergence of these pre-prepared model outputs to the transformed NDR curves is not expected to be the highest possible, each fit can be considered unique in a sense that, the model's prediction can reproducibly converge to the same dynamic output using the optimal GO method. In other words, since the optimal optimizer can reproduce a normalized model output with high accuracy and precision, this will apply in the case of a transformed signal assuming that the same normalized output is its best fit.

Another possible approach can be the creation of an approximate, analytic model whose inverse solution can be acquired by non-linear least squares, in polynomial time. Such model could be created using the HDNR component functions described in paragraph 4.4.2.3 and orthonormal polynomials [169]. Another way to reduce the fitting time is to investigate whether other algorithms, probably hybrid ones, could be employed/created that can perform equally or better than the DE and the CRSwCH but in a significantly smaller time frame. Kinging-based approaches and/or the combination of global and local, deterministic solvers are two possible options [37], [189]. Lastly, we can always provide warm start estimates to the GO algorithms in order for their search procedures to be, *a priori*, close to the possible global solutions. These estimates can be readily supplied by the use of a small size lookup table similar to the one that was described in paragraph 6.3. Therefore, the fitting procedure can be accelerated using several different approaches. Once this is realized it will allow the construction of pathology maps in real-time. Potentially, this may lead to the development of a rapid diagnostic tool that can be readily applied to clinical practice for the screening of the disease.

7.2.2 Applications

It is known that the study of biological functions is greatly aided by multi-scale methods that enable the coupling and simulation of models which span several spatial and temporal scales [227], [228]. Since our modeling approach of the atypical cells and their environment in the epithelium is a multi-scale representation, it can be utilized in order to better understand the cancer's biology and the complex mechanisms that follow it at several scales. For example, using our prediction on the pH_{ES} we could be able to assess the hypoxic conditions of the neoplastic tissue. These conditions are also known to be a functional signature of

the disease. Although this is still debatable, the irregular supply of O_2 from the blood and deprivation of lymphatic drainage is characteristic of tumor sites [183].

Another example is that our method permits the development of a tissue scattering model that may transmute the biophysical parameters into diffuse reflected photon intensity. This will link specifically the tissue's dynamic optical characteristics to its structural and functional conditions allowing us to address its heterogeneity from a biophotonic, macroscopic point of view. Besides, since our innovative approach can be used to provide additional measures of the temporal signatures of biomarkers, then, compared to simple fluorescence intensity measurements it can potentially provide improved detection and quantitation of these chemicals. As such, it can enable the derivation of anatomical information that could also allow post hoc correction of measured data via more realistic and constrained optical modeling to infer the true depth, location and concentration of the optical labels (e.g. fluorescent) [229].

There is currently a significant progress in developing affinity molecular probes that are selectively accumulating in malignant, tumor-relative surroundings by binding to overexpressed proteins or through other uptake mechanisms. At the moment, active research pursues two major reporting strategies that can be loosely grouped into non-targeting and targeting agents [237]. Non-targeting optical contrast agents are circulation oriented fluorochromes with a high molar extinction coefficient, good quantum yield, and low non-specific binding to non-cancerous tissue. At present, the most promising non-targeting agents approved by the Food and Drug Administration (FDA) for clinical applications are the 5-aminolevulinic acid (5-ALA) and the indocyanin green (ICG) dyes [238]. The former is a pro-drug that is metabolized intracellularly to form the fluorescent photoporphyrin IX which selectively accumulates in tumor cells as well as epithelial cells. The latter is a fluorescent contrast agent that is 98% protein-bound *in vivo* and behaves like a macromolecule in circulation. Targeting molecular probes typically involve the conjugation of a fluorescent dye to tumor-targeting biomolecules such as monoclonal antibodies, high affinity peptide ligands, epidermal growth factor receptors (EGFR and HER2) or metabolic characteristics such as glucose transporters (GLUTs) [239]–[242]. This approach has been successfully used in nuclear imaging, where radiolabeled biomolecules have been shown to detect human cancer noninvasively. Replacement of the radionuclide with a fluorescent dye has become a viable approach in optical imaging. Fluorescent dye families that are excitable in the NIR wavelength window such as the Cyanines, the Alexas, the DyLights, the IRDyes, the Quantum Dots and so forth have been extensively used and functionalized for conjugation to various targeting molecules in many preclinical studies [237], [243], [244]. However, only fluorescein is an FDA approved contrast agent that has already been used for in human use (FITC – folate conjugation) [239]. An even more important achievement is the development of fast-reacting tumor-specific molecular probes, such as the activatable molecular probe gGlu-HMRG [245], which can be aerosolized activatable molecular probes for topical application in the surgical field. Diagnostic optical signals are generated within a very short time (10 seconds), which makes these types of probes very suitable for intra-operative use. However the translation of these fast-acting activatable molecular probes to humans will take an unknown time course for receiving regulatory approval.

On the other hand, there is a number of regulatory approved markers that have been used in humans for long time for highlighting tissue abnormalities, including neoplasia, inflammation etc. The most commonly used are: Lugol's iodine (stains to glycogen of normal squamous epithelium), Methylene blue (stain Intestinal type epithelium), Toluidine blue (affin-

ity for mucopolysaccharides, i.e., goblet cells, intestinal metaplasia, Indigo carmine (stain Guelrud Intestinal Metaplasia Patterns), Crystal violet (highlights dysplastic mucosal patterns), acetic acid (highlights neoplastic epithelium), [246]. In the near infrared range, the most widely used contrast agent is indocyanine green (ICG), which has been used for deep tissue imaging, including lymphatic nodes [244].

Such plethora of contrast agents clearly indicates the numerous tissue and disease specific applications that our developed methodology can be employed. Indicatively, our method can be used to explore the biodistribution and pharmacokinetic properties of the contrast agents, in general, and elucidate their pathways and mechanisms of action. It can also be used to simply improve imaging of targeted contrast agents [230], [231]. Furthermore, it can allow for the imaging of the agent prior to and during its complete clearance from the body. Then, since areas where uptake is occurring should be simple to delineate from normal or wash-out regions, our optical imaging platform can enhance the specificity and improve the estimates of the agent uptake kinetics. Similarly, effects of autofluorescence can be removed in a fashion similar to conventional multi-spectral unmixing, in that autofluorescence would remain constant throughout imaging and therefore be a different temporal component to the dynamically changing contrast agent [4], [232]. The latter if adjoined by digitally processed tissue images with supervised or unsupervised classification and segmentation algorithms; it would delineate the lesion and increase the specificity of the proposed method [233], [234]. As a matter of fact, the sensitivity and specificity of our method should be systematically addressed by computing the pathology maps of many more clinical data. Fortunately, there is a plethora of clinical data in our possession that can be used for this purpose. The ultimate goal of such study would be the quantitative assessment of the clinical value of our methodology and the movement of our novel concepts from the bench-top to the bedside in the form of imaging device development and feasibility studies.

7.2.3 Feasibility study: intraoperative staging of tumors

A major scientific and clinical challenge and ambitious feasibility target of the described methodology is to develop a novel technology for improving the accuracy and the workflow in the intraoperative staging of tumors. Particularly, surgery remains the primary treatment paradigm for most solid tumors. Today's surgeons are exceptionally skilled in the art of open and minimally invasive surgeries with good patient outcomes. Intraoperative consultation examinations are often required by surgeons from pathologists for immediate important decisions regarding the optimal extent of surgery. Surgeons particularly want to know whether a lesion is malignant or not, or whether it has been totally removed. In an open surgery environment these critical answers are required to be given within 10-15 minutes. The surgeon often modifies his/her surgical plan based on the intraoperative consultation from pathologist. If rapid analysis of the tumors cannot be undertaken intraoperatively, it is most probable that a second operation will be required later which will be technically more difficult, delay the start of adjuvant radiotherapy and chemotherapy, have additional cost implications and may have an adverse psychological impact.

Successful and accurate completion of lymphatic mapping and lymph node biopsy procedures involves the integration of several technical components that cross multiple disciplines. These include the injection of the sentinel node localization agents, lymphoscintigraphy, surgical lymphatic node harvesting, and finally, careful cytologic and/or histologic examination of the node. Intraoperative evaluation of the resected tissue can be done using imprint cytology, frozen section analysis, scrape cytology, rapid cytokeratin immunostaining

etc. All these procedures are time consuming and are characterized by limited accuracy in staging the tumor during surgery. Indicatively, in the case of the axillary sentinel lymph nodes in breast cancers the methods' sensitivities in detecting macrometastases and micrometastases can vary between 45 and 85% and between 13 and 75%, respectively [235]. This performance is very discouraging provided that in most cases the presence of one or more positive lymph nodes is an indication for adjuvant chemotherapy, meaning that their detection is crucial. These difficulties in the intraoperative disease staging have led to recommendations for a minimum number of lymph nodes to be examined in resection specimens to limit under-staging [236]. Therefore, a reliable and robust 'single-setting' strategy is clearly needed for the intra-operative tumor detection and staging that will allow surgical resection to be tailored to the stage of the disease.

Particularly, real-time novel image guidance systems can facilitate intraoperative assessment of surgical margins and the detection of small positive nodules that are not visible to the unaided human eye. To enhance tumor-to-normal tissue contrast for intraoperative assessment of the lesion's type and margins, these systems rely on contrast agents that stain tumors selectively. Thus, a combination of the imaging methodology described in this thesis and tumor-selective dye or fluorescent molecular probes can facilitate the identification of tumors and the assessment of surgical margins with high sensitivity and specificity, and define with high accuracy extent of resection. Under such context our novel methodology can provide the means to develop technologies for improving the intraoperative detection threshold, the diagnostic and staging accuracy, while, at the same time improving the related workflow. Potentially this will become the next paradigm shift in cancer surgery.

7.3 CONCLUSION

In this thesis we have attempted to develop an innovative optical biopsy approach that will revolutionize the current clinical practice. Towards this direction we have shown that the systematic study of the tissue metabolic processes and their optical activation by molecular probes can provide a macroscopic assessment of cancer in a rigorous manner. Roughly speaking, DCE-OI can be performed either *in vivo* or *ex vivo* and/or *in vitro* for the purpose of examining and grading of atypical tissue sites. For this purpose there are, nowadays, several optical probes used by physicians and researchers specifically designed to mark biological targets which, in turn, are affine to an abnormality of the human body. It follows that we can apply the described procedures to a series of precancerous lesions. Hence, the universality of our methodology is apparent.

In addition, mapping of small molecular alterations taking place early in aetiopathogenesis may suffice for visualization, potentially allowing for much earlier detection of premalignant lesions or inflammation. At the same time, tissues can be analyzed for specific molecular features, with potential implications on the choice of targeted therapies. This can also offer a decisive insight into the molecular blueprints of tissues during an ongoing examination. Under such context, this novel technique has the potentials to become a platform which is expected to improve the specificity of the standard endoscopic imaging, intraoperative imaging, etc. and make surgical guiding more accurate and reliable.

Even so, there is a long way ahead of us and several open subjects need our attention before any of this is clinically employed. However, it is certain that we have established a pioneering procedure that holds the promise to become a paradigm in cancer assessment.

8 REFERENCES

- [1] C. Balas, "Review of biomedical optical imaging—a powerful, non-invasive, non-ionizing technology for improving in vivo diagnosis," *Meas. Sci. Technol.*, vol. 20, no. 10, 2009.
- [2] X. Huang, P. K. Jain, I. H. El-Sayed, and M. A. El-Sayed, "Gold nanoparticles: interesting optical properties and recent applications in cancer diagnostics and therapy.," *Nanomedicine (Lond)*, vol. 2, no. 5, pp. 681–93, Oct. 2007.
- [3] K. Sokolov, J. Galvan, A. Myakov, A. Lacy, R. Lotan, and R. Richards-Kortum, "Realistic three-dimensional epithelial tissue phantoms for biomedical optics," *J. Biomed. Opt.*, vol. 7, no. 1, pp. 148–156, Jan. 2002.
- [4] E. M. C. Hillman, C. B. Amoozegar, T. Wang, A. F. H. McCaslin, M. B. Bouchard, J. Mansfield, and R. M. Levenson, "In vivo optical imaging and dynamic contrast methods for biomedical research.," *Philos. Trans. A. Math. Phys. Eng. Sci.*, vol. 369, no. 1955, pp. 4620–43, Nov. 2011.
- [5] D. J. Bornhop, C. H. Contag, K. Licha, and C. J. Murphy, "Advances in contrast agents, reporters, and detection," *J. Biomed. Opt.*, vol. 6, no. 2, p. 106, 2001.
- [6] W. Du, Y. Wang, Q. Luo, and B. F. Liu, "Optical molecular imaging for systems biology: from molecule to organism," *Anal Bioanal Chem*, vol. 386, no. 3, pp. 444–457, 2006.
- [7] K. Sokolov, J. Aaron, B. Hsu, D. Nida, A. Gillenwater, M. Follen, C. MacAulay, K. Adler-Storthz, B. Korgel, M. Descour, R. Pasqualini, W. Arap, W. Lam, and R. Richards-Kortum, "Optical systems for in vivo molecular imaging of cancer.," *Technol. cancer ...*, vol. 2, no. 6, pp. 491–504, 2003.
- [8] R. Sankaranarayanan, P. O. Esmey, R. Rajkumar, R. Muwonge, R. Swaminathan, S. Shanthakumari, J.-M. Fayette, and J. Cherian, "Effect of visual screening on cervical cancer incidence and mortality in Tamil Nadu, India: a cluster-randomised trial," *Lancet*, vol. 370, pp. 398–406, 2007.
- [9] C. Yue, "The Promise of Dynamic Contrast-Enhanced Imaging in Radiation Therapy," *Semin. Radiat. Oncol.*, vol. 2, no. 21, pp. 147–156, 2011.
- [10] Z. M. Bhujwala, D. Artemov, P. Ballesteros, S. Cerdan, R. J. Gillies, and M. Solaiyappan, "Combined vascular and extracellular pH imaging of solid tumors.," *NMR Biomed.*, vol. 15, no. 2, pp. 114–9, Apr. 2002.
- [11] Y. Hama, Y. Koyama, P. L. Choyke, and H. Kobayashi, "Two-color in vivo dynamic contrast-enhanced pharmacokinetic imaging.," *J. Biomed. Opt.*, vol. 12, no. 3, p. 034016, 2007.
- [12] M. A. Rosen and M. D. Schnall, "Dynamic contrast-enhanced magnetic resonance imaging for assessing tumor vascularity and vascular effects of targeted therapies in renal cell carcinoma.," *Clin. Cancer Res.*, vol. 13, no. 2 Pt 2, pp. 770–776, Jan. 2007.
- [13] K. Balas, "Methods for characterizing tissues," WO 2008/0010372011.
- [14] C. J. Balas, G. C. Themelis, E. P. Prokopakis, I. Orfanudaki, E. Koumantakis, and E. S. Helidonis, "In vivo detection and staging of epithelial dysplasias and malignancies based on

- the quantitative assessment of acetic acid-tissue interaction kinetics," *J. Photochem. Photobiol. B-Biology*, vol. 53, no. 1–3, pp. 153–157, 1999.
- [15] C. Balas, A. Dimoka, I. Orfanudaki, and E. Koumantakis, "In vivo assessment of acetic acid–cervical tissue interaction using quantitative imaging of back-scattered light: its potential use for the in vivo cervical cancer detection grading and mapping," in *Optical Biopsies and Microscopic Techniques III*, 1998, vol. 3568, pp. 31–37.
- [16] W. P. Soutter, E. Diakomanolis, D. Lyons, S. Ghaem-Maghami, T. Ajala, D. Haidopoulos, D. Doumplis, C. Kalpaktsoglou, G. Sakellaropoulos, S. Soliman, K. Perryman, V. Hird, C. H. Buckley, K. Pavlakis, S. Markaki, R. Dina, V. Healy, and C. Balas, "Dynamic spectral imaging: improving colposcopy," *Clin Cancer Res*, vol. 15, no. 5, pp. 1814–1820, 2009.
- [17] C. Balas, G. Papoutsoglou, and A. Potirakis, "In vivo molecular imaging of cervical neoplasia using acetic acid as biomarker," *IEEE J. Sel. Top. Quantum Electron.*, vol. 14, no. 1, pp. 29–42, 2008.
- [18] C. Balas, G. Papoutsoglou, C. Loukas, Y. Skiadas, C. Pappas, D. Haidopoulos, E. Diakomanolis, and W. P. Soutter, "In vivo assessment of microstructural and functional alterations in cervical neoplasia," in *Conference proceedings: European Conferences on Biomedical Optics (ECBO)*, 2009.
- [19] J. A. Jacquez and P. Greif, "Numerical Parameter Identifiability and Estimability - Integrating Identifiability, Estimability, and Optimal Sampling Design," *Math. Biosci.*, vol. 77, no. 1–2, pp. 201–227, 1985.
- [20] T. Ideker, T. Galitski, and L. Hood, "A new approach to decoding life: systems biology.," *Annu. Rev. Genomics Hum. Genet.*, vol. 2, pp. 343–72, Jan. 2001.
- [21] J. Vanlier, C. a Tiemann, P. a J. Hilbers, and N. a W. van Riel, "Parameter uncertainty in biochemical models described by ordinary differential equations.," *Math. Biosci.*, Mar. 2013.
- [22] G. S. Papoutsoglou and C. Balas, "Estimation of neoplasia-related biological parameters through modeling and sensitivity analysis of optical molecular imaging data.," *IEEE Transactions Bio-Medical Eng.*, vol. 60, no. 5, pp. 1241–9, May 2013.
- [23] S. Audoly, G. Bellu, L. D'Angiò, M. P. Saccomani, and C. Cobelli, "Global identifiability of nonlinear models of biological systems.," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 1, pp. 55–65, Jan. 2001.
- [24] G. Papoutsoglou, M. Stamatidou, and C. Balas, "In silico Modeling and Global Optimization of Dynamic Bio-optical Processes for Probing, in vivo, Biological Features of Neoplasia," in *2012 International Conference on Biomedical Engineering and Biotechnology*, 2012, pp. 332–335.
- [25] G. Papoutsoglou, T.-M. Giakoumakis, and C. Balas, "Dynamic contrast enhanced optical imaging of cervix, in vivo: A paradigm for mapping neoplasia-related parameters.," in *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 2013, vol. 2013, pp. 3479–82.
- [26] E. R. Carson, C. Cobelli, and L. Finkelstein, *The Mathematical Modeling of Metabolic and Endocrine Systems: Model Formulation, Identification, and Validation*. New York: J. Wiley, 1983, p. 394.
- [27] N. a W. van Riel, "Dynamic modelling and analysis of biochemical networks: mechanism-based models and model-based experiments," *Brief. Bioinform.*, vol. 7, no. 4, pp. 364–374, Dec. 2006.

- [28] E. Walter and L. Pronzato, *Identification of parametric models from experimental data*. Berlin; New York; Paris: Springer ; Masson, 1997, p. xviii, 413 p.
- [29] L. Ljung, *System identification—Theory for the user*. Upper Saddle River, N.J.: Prentice Hall, 1999, p. 672.
- [30] K. H. Norwich, *Molecular dynamics in biosystems: The kinetics of tracers in intact organisms*. Oxford, England: Pergamn Press, 1977, p. 418.
- [31] A. Belič, “Modelling in systems biology, neurology and pharmacy,” *Math. Comput. Model. Dyn. Syst.*, vol. 15, no. 6, pp. 479–491, Nov. 2009.
- [32] W. Boron, “Transport of H⁺ and of ionic weak acids and bases,” *J. Membr. Biol.*, vol. 16, pp. 1–16, 1983.
- [33] R. K. Jain, “Transport of molecules across tumor vasculature.,” *Cancer Metastasis Rev.*, vol. 6, no. 4, pp. 559–93, Jan. 1987.
- [34] T. R. Shannon, F. Wang, J. Puglisi, C. Weber, and D. M. Bers, “A mathematical treatment of integrated Ca dynamics within the ventricular myocyte.,” *Biophys. J.*, vol. 87, no. 5, pp. 3351–71, Nov. 2004.
- [35] S. Kuyucak, O. S. Andersen, and S.-H. H. Chung, “Models of permeation in ion channels,” *REPORTS Prog. Phys.*, vol. 64, no. 11, pp. 1427–1472, Nov. 2001.
- [36] S. Audoly, L. D’Angiò, M. P. Saccomani, and C. Cobelli, “Global identifiability of linear compartmental models—a computer algebra algorithm.,” *IEEE Tansactions Bio-Medical Eng.*, vol. 45, no. 1, pp. 36–47, 1998.
- [37] M. Rodriguez-Fernandez, J. a Egea, and J. R. Banga, “Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems.,” *BMC Bioinformatics*, vol. 7, p. 483, Jan. 2006.
- [38] K. Godfrey, “The identifiability of parameters of models used in biomedicine,” *Math. Model.*, vol. 7, no. 9–12, pp. 1195–1214, Jan. 1986.
- [39] O.-T. Chis, J. R. Banga, and E. Balsa-Canto, “Structural identifiability of systems biology models: a critical comparison of methods.,” *PLoS One*, vol. 6, no. 11, p. e27755, Jan. 2011.
- [40] S. Vajda, K. R. Godfrey, and H. Rabitz, “Similarity transformation approach to identifiability analysis of nonlinear compartmental models.,” *Math. Biosci.*, vol. 93, no. 2, pp. 217–48, Apr. 1989.
- [41] D. Zak, G. Gonye, J. Schwaber, and F. Doyle, “Importance of input perturbations and stochastic gene expression in the reverse engineering of genetic regulatory networks: insights from an identifiability analysis of,” *Genome Res.*, pp. 2396–2405, 2003.
- [42] A. Saltelli, “Making best use of model evaluations to compute sensitivity indices,” *Comput. Phys. Commun.*, vol. 145, no. 2, pp. 280–297, May 2002.
- [43] A. Saltelli, S. Tarantola, and F. Campolongo, “Sensitivity analysis as an ingredient of modeling,” *Stat. Sci.*, vol. 15, no. 4, pp. 377–395, 2000.
- [44] A. Kiparissides, M. Koutinas, C. Kontoravdi, A. Mantalaris, and E. N. Pistikopoulos, “‘Closing the loop’ in biological systems modeling - From the in silico to the in vitro,” *Automatica*, vol. 47, no. 6, pp. 1147–1155, 2011.

- [45] S. Marino, I. B. Hogue, C. J. Ray, and D. E. Kirschner, "A methodology for performing global uncertainty and sensitivity analysis in systems biology.," *J. Theor. Biol.*, vol. 254, no. 1, pp. 178–96, Sep. 2008.
- [46] a. Kiparissides, S. S. Kucherenko, a. Mantalaris, and E. N. Pistikopoulos, "Global Sensitivity Analysis Challenges in Biological Systems Modeling," *Ind. Eng. Chem. Res.*, vol. 48, no. 15, pp. 7168–7180, Aug. 2009.
- [47] A. Saltelli, T. Andres, and T. Homma, "Sensitivity analysis of model output: an investigation of new techniques," *Comput. Stat. Data Anal.*, vol. 15, 1993.
- [48] C. Cobelli and A. Mari, "Validation of mathematical models of complex endocrine-metabolic systems. A case study on a model of glucose regulation," *Med. Biol. Eng. Comput.*, vol. 21, no. 4, pp. 390–399, Jul. 1983.
- [49] P. Congdon, *Applied Bayesian Modelling*, Wiley Seri. Wiley, 2003, p. 478.
- [50] W. Navidi, *Statistics for Engineers and Scientists*. New York: McGraw-Hill, 2010, p. 928.
- [51] D. C. Montgomery and G. C. Runger, *Applied Statistics and Probability for Engineers*. New York: John Wiley & Sons, Inc., 2011, p. 784.
- [52] T. G. Kolda, R. M. Lewis, and V. Torczon, "Optimization by Direct Search: New Perspectives on Some Classical and Modern Methods," *SIAM Rev.*, vol. 45, no. 3, pp. 385–482, Jan. 2003.
- [53] K. Jaqaman and G. Danuser, "Linking data to models: data regression.," *Nat. Rev. Mol. Cell Biol.*, vol. 7, no. 11, pp. 813–9, Nov. 2006.
- [54] M. Joshi, a. Kremling, and a. Seidel-Morgenstern, "Model based statistical analysis of adsorption equilibrium data," *Chem. Eng. Sci.*, vol. 61, no. 23, pp. 7805–7818, Dec. 2006.
- [55] E. Balsa-Canto, A. Alonso, and J. Banga, "Computational procedures for optimal experimental design in biological systems," *Syst. Biol. IET*, vol. 2, no. 4, pp. 163–172, 2008.
- [56] E. Balsa-Canto, A. a Alonso, and J. R. Banga, "An iterative identification procedure for dynamic modeling of biochemical networks.," *BMC Syst. Biol.*, vol. 4, p. 11, Jan. 2010.
- [57] C. Cobelli and E. Carson, *Introduction to modeling in physiology and medicine*. Amsterdam: Else, 2008, p. 337.
- [58] W. R. W. Loewenstein, "Junctional Intercellular Communication and the Control of Growth," *Biochim. Biophys. Acta*, vol. 560, no. 560, pp. 1–65, 1979.
- [59] H. Lodish, A. Berk, P. Matsudaira, C. A. Kaiser, M. Krieger, M. P. Scott, S. L. Zipursky, and J. Darnell, "Integrating Cells into Tissues," *Mol. Cell Biol. 5th ed.*, New York, W.H. Free. Co., vol. 6, pp. 197–231, 2004.
- [60] E. Overton, "Über die osmotischen eigenshaften der lambenden pflanzen und tierzellen," *Vierteljahrschr. Naturforsch. Ges. Zurich*, vol. 40, pp. 159–201, 1895.
- [61] T.-X. Xiang and B. D. Anderson, "Phospholipid surface density determines the partitioning and permeability of acetic acid in DMPC : Cholesterol Bylayers.," *J. Membr. Biol.*, vol. 148, pp. 157 – 167, 1995.
- [62] A. Finkelstein, "Water and nonelectrolyte permeability of lipid bilayer membranes," *J. Gen. Physiol.*, vol. 68, pp. 127–135, 1976.

- [63] H. Sasaki, C. Matsui, K. Furuse, Y. Mimori-Kiyosue, M. Furuse, and S. Tsukita, "Dynamic behavior of paired claudin strands within apposing plasma membranes," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 7, pp. 3971–3976, Apr. 2003.
- [64] C. M. Van Itallie and J. M. Anderson, "Claudins and epithelial paracellular transport.," *Annu. Rev. Physiol.*, vol. 68, pp. 403–29, Jan. 2006.
- [65] K. Matter and M. Balda, "Functional analysis of tight junctions," *Methods*, vol. 30, pp. 228–234, 2003.
- [66] M. Mori, N. Sawada, Y. Kokai, and M. Satoh, "Role of tight junctions in the occurrence of cancer invasion and metastasis," *Med. Electron Microsc.*, vol. 32, pp. 193–198, 1999.
- [67] V. W. Tang and D. A. Goodenough, "Paracellular ion channel at the tight junction.," *Biophys. J.*, vol. 84, no. 3, pp. 1660–73, Mar. 2003.
- [68] V. Wong and D. a Goodenough, "Paracellular Channels!," *Sci. New Ser.*, vol. 285, no. 5424, p. 62, Jul. 1999.
- [69] P. Claude, "Morphological Factors Influencing Transepithelial Permeability" A Model for the Resistance of the Zonula Occludens," *J. Membr. Biol.*, vol. 39, no. 2–3, pp. 219–232, Mar. 1978.
- [70] a Martínez-Palomo, D. Erij, and A. Martinez-Palomo, "Structure of tight junctions in epithelia with different permeability.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 72, no. 11, pp. 4487–91, Nov. 1975.
- [71] J. L. Madara, "Regulation of the movement of solutes across tight junctions.," *Annu. Rev. Physiol.*, vol. 60, pp. 143–59, Jan. 1998.
- [72] A. Nusrat, J. R. Turner, and J. L. Madara, "Molecular physiology and pathophysiology of tight junctions IV. Regulation of tight junctions by extracellular stimuli: nutrients, cytokines, and immune cells," *Am. J. Physiol. Liver Physiol.*, vol. 279, no. 5, pp. G851–G857, 2000.
- [73] R. Dermietzel, T. K. Hwang, and D. S. Spray, "Anatomy and Embryology Review article The gap junction family : structure , function and chemistry," *Anat. Embryol. (Berl.)*, vol. 182, no. 6, pp. 517–528, 1990.
- [74] W. H. Evans and P. E. M. Martin, "Gap junctions: structure and function (Review)," *Mol. Membr. Biol.*, vol. 19, no. 2, pp. 121–136, 2002.
- [75] M. V. L. Bennet, M. E. Spew, D. C. Spray, M. Bennett, and M. Spira, "Permeability of gap junctions between embryonic cells of Fundulus: A reevaluation," *Dev. Biol.*, vol. 125, no. 65, pp. 114–125, 1978.
- [76] L. S. Musil and D. a Goodenough, "Biochemical analysis of connexin43 intracellular transport, phosphorylation, and assembly into gap junctional plaques.," *J. Cell Biol.*, vol. 115, no. 5, pp. 1357–74, Dec. 1991.
- [77] L. S. Musil and D. a Goodenough, "Multisubunit assembly of an integral plasma membrane channel protein, gap junction connexin43, occurs after exit from the ER.," *Cell*, vol. 74, no. 6, pp. 1065–77, Sep. 1993.
- [78] T. Aasen, S. V Graham, M. Edward, and M. B. Hodgins, "The relationship between connexins, gap junctions, tissue architecture and tumor invasion, as studied in a novel in vitro model of HPV-16-associated cervical cancer progression," *Oncogene*, vol. 22, pp. 6025–6036, 2003.

- [79] P. Swietach and R. D. Vaughan-Jones, "Novel method for measuring junctional proton permeation in isolated ventricular myocyte cell pairs.," *Am. J. Physiol. Heart Circ. Physiol.*, vol. 287, no. 5, pp. H2352–63, Nov. 2004.
- [80] F. L. Cao, R. Eckert, C. Elfgang, J. M. Nitsche, S. A. Snyder, D. F. Hulser, K. Willecke, B. J. Nicholson, and D. F. H-ulser, "A quantitative analysis of connexin-specific permeability differences of gap junctions expressed in HeLa transfectants and *Xenopus* oocytes," *J. Cell Sci.*, vol. 111, pp. 31–43, Jan. 1998.
- [81] B. J. Nicholson, P. A. Weber, F. Cao, H. C. Chang, P. Lampe, and G. Goldberg, "The molecular basis of selective permeability of connexins is complex and includes both size and charge," *Brazilian J. Med. Biol. Res.*, vol. 33, no. 4, pp. 369–378, Apr. 2000.
- [82] P. a Weber, H.-C. Chang, K. E. Spaeth, J. M. Nitsche, and B. J. Nicholson, "The Permeability of Gap Junction Channels to Probes of Different Size Is Dependent on Connexin Composition and Permeant-Pore Affinities," *Biophys. J. Vol.*, vol. 87, no. 2, pp. 958–973, Aug. 2004.
- [83] C. Peracchia, "Increase in gap junction resistance with acidification in crayfish septate axons is closely related to changes in intracellular calcium but not hydrogen ion concentration," *J. Membr. Biol.*, vol. 92, no. 113, pp. 75–92, 1990.
- [84] A. Roos and W. F. Boron, "Intracellular pH," *Physiol. Rev.*, vol. 61, pp. 296–434, 1981.
- [85] R. K. Jain, "Transport of molecules, particles, and cells in solid tumors.," *Annu. Rev. Biomed. Eng.*, vol. 1, no. 96, pp. 241–63, Jan. 1999.
- [86] C. H. Leem and R. D. Vaughan-Jones, "Out-of-equilibrium pH transients in the guinea-pig ventricular myocyte.," *J. Physiol.*, vol. 509 (Pt 2, no. 2, pp. 471–85, Jun. 1998.
- [87] W. Boron, "Control of intracellular pH," *Kidney Physiol. Pathophysiol.*, no. 428, pp. 199–202, 1992.
- [88] C. H. Leem, D. Lagadic-Gossmann, and R. D. Vaughan-Jones, "Characterization of intracellular pH regulation in the guinea-pig ventricular myocyte.," *J. Physiol.*, vol. 517 (Pt 1, no. 1, pp. 159–80, May 1999.
- [89] J. M. Boyer, I. F. Tannock, and M. J. Boyer, "Regulation of intracellular pH in tumor cell lines: influence of microenvironmental conditions.," *Cancer Res.*, vol. 52, no. 16, pp. 4441–7, Aug. 1992.
- [90] P. Swietach, S. Patiar, C. T. Supuran, A. L. Harris, and R. D. Vaughan-Jones, "The role of carbonic anhydrase 9 in regulating extracellular and intracellular ph in three-dimensional tumor cell growths.," *J. Biol. Chem.*, vol. 284, no. 30, pp. 20299–310, Jul. 2009.
- [91] C. P. S. Potter and a L. Harris, "Diagnostic, prognostic and therapeutic implications of carbonic anhydrases in cancer.," *Br. J. Cancer*, vol. 89, no. 1, pp. 2–7, Jul. 2003.
- [92] M. Szatkowski, "The effect of extracellular weak acids and bases on the intracellular buffering power of snail neurones.," *J. Physiol.*, pp. 103–120, 1989.
- [93] M. J. Boyer, M. Barnard, D. W. Hedley, and I. F. Tannock, "Regulation of intracellular pH in subpopulations of cells derived from spheroids and solid tumours.," *Br. J. Cancer*, vol. 68, no. 5, pp. 890–7, Nov. 1993.
- [94] A. H. Lee and I. F. Tannock, "Heterogeneity of intracellular pH and of Mechanisms That Regulate Intracellular pH in Populations of Cultured Cells," *Cancer Res.*, vol. 58, pp. 1901–1908, 1998.

- [95] M. Stubbs, L. Rodrigues, F. A. Howe, J. Wang, K. Jeong, R. L. Veech, and J. I. Griffiths, "Metabolic Consequences of a Reversed pH Gradient in Rat Tumors Metabolic Consequences of a Reversed pH Gradient in Rat Tumors," *Cancer Res.*, vol. 54, pp. 4011–4016, 1994.
- [96] P. Besson, J. Gore, E. Vincent, C. Hoinard, P. Bougnoux, and J. Goré, "Inhibition of Na⁺/H⁺ Exchanger Activity by an Alkyl-Lysophospholipid Analogue in a Human Breast Cancer Cell Line," *Biochem. Pharmacol.*, vol. 51, no. 9, pp. 1153–1158, May 1996.
- [97] J. S. Fang, R. D. Gillies, and R. a Gatenby, "Adaptation to hypoxia and acidosis in carcinogenesis and tumor progression," *Semin. Cancer Biol.*, vol. 18, no. 5, pp. 330–337, Oct. 2008.
- [98] P. a Schornack and R. J. Gillies, "Contributions of cell metabolism and H⁺ diffusion to the acidic pH of tumors.," *Neoplasia*, vol. 5, no. 2, pp. 135–45, 2003.
- [99] R. A. Gatenby and E. T. Gawlinski, "The glycolytic phenotype in carcinogenesis and tumor invasion: insights through mathematical models.," *Cancer Res.*, vol. 63, no. 14, pp. 3847–54, Jul. 2003.
- [100] R. J. Gillies, I. Robey, and R. a Gatenby, "Causes and consequences of increased glucose metabolism of cancers.," *J. Nucl. Med.*, vol. 49 Suppl 2, no. 6, p. 24S–42S, Jun. 2008.
- [101] M. Yamagata and I. F. Tannock, "The chronic administration of drugs that inhibit the regulation of intracellular pH: in vitro and anti-tumor effects," *Br. J. Cancer*, vol. 73, pp. 1328–1334, 1996.
- [102] S. D. Webb, J. A. Sherratt, and R. G. Fish, "Mathematical Modelling of Tumour Acidity: Regulation of Intracellular pH," *J. theor. Biol.*, vol. 196, pp. 237–250, 1999.
- [103] L. G. Smyth, G. O'Hurley, A. O'Grady, J. M. Fitzpatrick, E. Kay, and R. W. G. Watson, "Carbonic anhydrase IX expression in prostate cancer.," *Prostate Cancer Prostatic Dis.*, vol. 13, no. 2, pp. 178–81, Jun. 2010.
- [104] I. J. Latorre, K. K. Frese, and R. T. Javier, "Tight Junction Proteins and Cancer," *Tight Junctions, Springer US ed.*, pp. 116–134, 2006.
- [105] H. Yamasaki, C. C.G.Naus, and C. C. Naus, "Role of connexin genes in growth control," *Carcinogenesis*, vol. 17, no. 6, pp. 1199–1213, Jun. 1996.
- [106] J. W. Holder, E. Elmore, and J. C. Barrett, "Gap junction function and cancer," *Cancer Res.*, vol. 53, no. 15, pp. 3475–85, 1993.
- [107] S. Hirohashi, "Inactivation of the E-cadherin-mediated cell adhesion system in human cancers," *Am. J. Pathol.*, vol. 153, no. 2, pp. 333–339, 1998.
- [108] M. Steinau, M. S. Rajeevan, D. R. Lee, M. T. Ruffin, I. R. Horowitz, L. C. Flowers, T. Tadros, G. Birdsong, M. Husain, D. C. Kmak, G. M. Longton, S. D. Vernon, and E. R. Unger, "Evaluation of RNA markers for early detection of cervical neoplasia in exfoliated cervical cells.," *Cancer Epidemiol. Biomarkers Prev.*, vol. 16, no. 2, pp. 295–301, Mar. 2007.
- [109] S. S. Oliveira, J. a Morgado-Díaz, and J. A. Morgado-Díaz, "Claudins: multifunctional players in epithelial tight junctions and their role in cancer.," *Cell. Mol. Life Sci.*, vol. 64, no. 1, pp. 17–28, Jan. 2007.
- [110] J. J. Eckert, A. McCallum, A. Mears, M. G. Rumsby, I. T. Cameron, and T. P. Fleming, "Relative contribution of cell contact pattern, specific PKC isoforms and gap junctional communication in tight junction assembly in the mouse early embryo," *Dev. Biol.*, vol. 288, no. 1, pp. 234–247, Dec. 2005.

- [111] T. Aasen, S. V Graham, M. Edward, and M. B. Hodgins, "Reduced expression of multiple gap junction proteins is a feature of cervical dysplasia," *Mol. Cancer*, vol. 4, no. 1, p. art. no.–31, Jan. 2005.
- [112] I. Steinhoff, K. Leykauf, U. Bleyl, M. Dürst, A. Alonso, D. Matthias, and A. Alonsob, "Phosphorylation of the gap junction protein Connexin43 in CIN III lesions and cervical carcinomas," *Cancer Lett.*, vol. 235, no. 2, pp. 291–297, Apr. 2006.
- [113] D. C. Walker, B. H. Brown, a D. Blackett, J. Tidy, and R. H. Smallwood, "A study of the morphological parameters of cervical squamous epithelium.," *Physiol. Meas.*, vol. 24, no. 1, pp. 121–35, Feb. 2003.
- [114] J. Doorbar, "Molecular biology of human papillomavirus infection and cervical cancer," *Clin. Sci.*, vol. 175, pp. 1110–1113, 2006.
- [115] M. Mesnil, S. Crespín, J.-L. J.-L. Avanzo, and M.-L. Zaidan-Dagli, "Defective gap junctional intercellular communication in the carcinogenic process," *Biochim. Biophys. Acta*, vol. 1719, no. 1–2, pp. 125 – 145, Dec. 2005.
- [116] C. Peracchia, "Chemical gating of gap junction channels Roles of calcium, pH and calmodulin," *Biochim. Biophys. Acta*, vol. 1662, no. 1–2, pp. 61–80, Mar. 2004.
- [117] J. F. Ek-Vitorín, G. Calero, G. E. Morley, W. Coombs, S. M. Taffet, M. Delmar, and J. F. Ek-Vitorin, "PH regulation of connexin43: molecular analysis of the gating particle.," *Biophys. J.*, vol. 71, no. 3, pp. 1273–84, Sep. 1996.
- [118] G. E. Morley, S. M. Taffet, and M. Delmar, "Intramolecular Interactions Mediate pH Regulation of Connexin43 Channels," *Biophys. J.*, vol. 70, pp. 1294–1302, 1996.
- [119] K. T. Davis, N. Prentice, V. L. Gay, and S. A. Murray, "Gap junction proteins and cell–cell communication in the three functional zones of the adrenal gland," *J. Endocrinol.*, vol. 173, no. 173, pp. 13–21, Apr. 2002.
- [120] C. D. Blaskewicz, J. Pudney, and D. J. Anderson, "Structure and function of intercellular junctions in human cervical and vaginal mucosal epithelia.," *Biol. Reprod.*, vol. 85, no. 1, pp. 97–104, Jul. 2011.
- [121] C. Tselepis, M. Chidgey, A. North, and D. Garrod, "Desmosomal adhesion inhibits invasive behavior," *Cell Biol.*, vol. 95, no. 14, pp. 8064–8069, Jul. 1998.
- [122] G. Sobel, C. Páska, I. Szabó, A. Kiss, A. Kádár, Z. Schaff, C. Paska, I. Szabo, and A. Kadar, "Increased expression of claudins in cervical squamous intraepithelial neoplasia and invasive carcinoma.," *Hum. Pathol.*, vol. 36, no. 2, pp. 162–9, Feb. 2005.
- [123] C. Carrilho, M. Alberto, L. Buane, and L. David, "Keratins 8,10,13 and 17 are useful markers in the diagnosis of human cervix carcinomas," *Hum. Pathol.*, vol. 35, no. 5, pp. 546 – 551, 2004.
- [124] C. M. Van Itallie, J. Holmes, A. Bridges, J. L. Gookin, M. R. Coccaro, W. Proctor, O. R. Colegio, and J. M. Anderson, "The density of small tight junction pores varies among cell types and is increased by expression of claudin-2.," *J. Cell Sci.*, vol. 121, no. Pt 3, pp. 298–305, Feb. 2008.
- [125] J. M. Anderson, "Molecular structure of tight junctions and their role in epithelial transport.," *News Physiol. Sci.*, vol. 16, pp. 126–30, Jun. 2001.
- [126] A. B. MacLean, A. Singer, and H. Critchley, "The use of acetic acid," *Low. Genit. tranct neoplasia, RCOG Press. London*, pp. 144–152, 2003.

- [127] L. E. Gerweck, S. Vijayappa, and S. Kozin, "Tumor pH controls the in vivo efficacy of weak acid and base chemotherapeutics.," *Mol. Cancer Ther.*, vol. 5, no. 5, pp. 1275–9, May 2006.
- [128] J. R. Turner, B. K. Rill, S. L. Carlson, D. Carnes, R. Kerner, R. J. Mrsny, and J. L. Madara, "Physiological regulation of epithelial tight junctions is associated with myosin light-chain phosphorylation," *Am. J. Physiol. Physiol.*, vol. 42, no. 4, pp. C1378–C1385, 1997.
- [129] H. B. Kaufman and D. M. Harper, "Magnification and chromoscopy with the acetic acid test.," *Endoscopy*, vol. 36, no. 8, pp. 748–50; author reply 751, Aug. 2004.
- [130] R. Drezek, A. Dunn, and R. R. Richards-Kortum, "Light scattering from cells: finite-difference time-domain simulations and goniometric measurements.," *Appl. Opt.*, vol. 38, no. 16, pp. 3651–61, Jun. 1999.
- [131] L. Perelman, V. Backman, M. Wallace, G. Zonios, R. Manoharan, A. Nusrat, S. Shields, M. Seiler, C. Lima, T. Hamano, I. Itzkan, J. Van Dam, J. Crawford, and M. Feld, "Observation of Periodic Fine Structure in Reflectance from Biological Tissue: A New Technique for Measuring Nuclear Size Distribution," *Phys. Rev. Lett.*, vol. 80, no. 3, pp. 627–630, Jan. 1998.
- [132] I. Georgakoudi, E. E. Sheets, M. G. Müller, V. Backman, C. P. Crum, K. Badizadegan, R. R. Dasari, and M. S. Feld, "Trimodal spectroscopy for the detection and characterization of cervical precancers in vivo," *Am. J. Obstet. Gynecol.*, vol. 186, no. 3, pp. 374–382, Mar. 2002.
- [133] J. R. Mourant, T. J. Bocklage, T. M. Powers, H. M. Greene, K. L. Bullock, L. R. Marr-Lyon, M. H. Dorin, A. G. Waxman, M. M. Zsemlye, and H. O. Smith, "In vivo light scattering measurements for detection of precancerous conditions of the cervix.," *Gynecol. Oncol.*, vol. 105, no. 2, pp. 439–45, May 2007.
- [134] S. K. Chang, D. Arifler, R. Drezek, M. Follen, and R. Richards-Kortum, "Analytical model to describe fluorescence spectra of normal and preneoplastic epithelial tissue: comparison with Monte Carlo simulations and clinical measurements.," *J. Biomed. Opt.*, vol. 9, no. 3, pp. 511–22.
- [135] R. A. Drezek, T. Collier, C. K. Brookner, A. Malpica, R. Lotan, R. R. Richards-Kortum, and M. Follen, "Laser scanning confocal microscopy of cervical tissue before and after application of acetic acid," *Am. J. Obstet. Gynecol.*, vol. 182, no. 5, pp. 1135–1139, 2000.
- [136] T. T. Wu, J. Y. Qu, T. H. Cheung, S. F. Yim, and Y. F. Wong, "Study of dynamic process of acetic acid induced-whitening in epithelial tissues at cellular level," *Opt. Express*, vol. 13, no. 13, pp. 4963–4973, 2005.
- [137] R. a Drezek, R. Richards-Kortum, M. a Brewer, M. S. Feld, C. Pitris, A. Ferenczy, M. L. Faupel, and M. Follen, "Optical imaging of the cervix," *Cancer*, vol. 98, no. 9, pp. 2015–2027, Nov. 2003.
- [138] a. V. V Finkelstein and O. V. V Galzitskaya, "Physics of protein folding," *Phys. Life Rev.*, vol. 1, no. 23–56, pp. 23–56, Apr. 2004.
- [139] P. Maddox, A. Szarewski, J. Dyson, and J. Cuzick, "Cytokeratin expression and acetowhite change in cervical epithelium. Journal of clinical pathology," *J. Clin. Pathol.*, vol. 47, no. 1, pp. 15–17, 1994.
- [140] F. Smedts, F. Ramaekers, and P. Vooijs, "The dynamics of keratin expression in malignant transformation of cervical epithelium: a review," *Obstet. Gynecol.*, vol. 82, no. 3, p. 465, 1993.
- [141] C. Balas, "A novel optical imaging method for the early detection, quantitative grading, and mapping of cancerous and precancerous lesions of cervix," *IEEE Trans Biomed Eng.*, vol. 48, no. 1, pp. 96–104, 2001.

- [142] M. H. Friedman, *Principles and Models of Biological Transport*. New York, NY: Springer New York, 2008.
- [143] W. F. Boron and P. De Weer, "Intracellular pH transients in squid giant axons caused by CO₂, NH₃, and metabolic inhibitors.," *J. Gen. Physiol.*, vol. 67, no. 1, pp. 91–112, Jan. 1976.
- [144] D. Keifer and A. Roos, "Membrane permeability to the molecular and ionic forms of DMO in barnacle muscle," *Am. J. Physiol.* ..., 1981.
- [145] D. C. Walker, B. H. Brown, D. R. Hose, and R. H. Smallwood, "Modelling the electrical impedivity of normal and premalignant cervical tissue," vol. 36, no. 19, 2000.
- [146] S. Grinstein, S. Cohen, and a Rothstein, "Cytoplasmic pH regulation in thymic lymphocytes by an amiloride-sensitive Na⁺/H⁺ antiport.," *J. Gen. Physiol.*, vol. 83, no. 3, pp. 341–69, Mar. 1984.
- [147] T.-X. X. Xiang and B. D. Anderson, "Permeability of acetic acid across gel and liquid-crystalline lipid bilayers conforms to free-surface-area theory.," *Biophys. J.*, vol. 72, no. 1, pp. 223–37, Jan. 1997.
- [148] A. Walter and J. Gutknecht, "Permeability of small nonelectrolytes through lipid bilayer membranes.," *J. Membr. Biol.*, vol. 90, no. 3, pp. 207–17, Jan. 1986.
- [149] G. I. Gorodeski, M. F. Romero, U. Hopfer, E. Rorke, W. H. Utian, and R. L. Eckert, "Human uterine cervical epithelial cells grown on permeable support--a new model for the study of differentiation.," *Differentiation*, vol. 56, no. 1–2, pp. 107–18, Apr. 1994.
- [150] L. Langbein, C. Grund, C. Kuhn, S. Praetzel, J. Kartenbeck, J. M. Brandner, I. Moll, and W. W. Franke, "Tight junctions and compositionally related junctional structures in mammalian stratified epithelia and cell cultures derived therefrom.," *Eur. J. Cell Biol.*, vol. 81, no. 8, pp. 419–35, Aug. 2002.
- [151] R. Jain, L. Gerlowski, and E. Cilento, "Extravascular transport in normal and tumor tissues," *Crit. Rev. Oncol.* ..., vol. 5, no. 2, 1986.
- [152] L. E. Gerlowski and R. K. Jain, "Microvascular permeability of normal and neoplastic tissues.," *Microvasc. Res.*, vol. 31, no. 3, pp. 288–305, May 1986.
- [153] L. Balasubramani, B. H. Brown, J. Healey, and J. a Tidy, "The detection of cervical intraepithelial neoplasia by electrical impedance spectroscopy: the effects of acetic acid and tissue homogeneity.," *Gynecol. Oncol.*, vol. 115, no. 2, pp. 267–71, Nov. 2009.
- [154] T. F. Weiss, "Ion Transport and Resting Potential," in *Cellular Biophysics-v. 1. Transport*, Cambridge, Massachusetts: The MIT Press, 1996, pp. 449–532.
- [155] G. Papoutsoglou, "Modelling of epithelial transport phenomena related with the acetowhitening optical characteristics: Potential for the in-vivo diagnosis of Cervical Neoplasia," Technical University of Crete, 2007.
- [156] P. Soutter, E. Diakomanolis, D. Lyons, S. G. Maghami, T. Ajala, D. Haidopoulos, D. Doublis, C. Kalpaktsoglou, G. Sakellaropoulos, S. Soliman, K. Perryman, V. Hird, C. Simos, Y. Skiadas, H. Kavagiou, and C. Balas, "No Title," in *British International Congress of Obstetrics & Gynaecology London*, 2007.
- [157] C. Balas, G. Papoutsoglou, E. Diakomanolis, D. Haidopoulos, and W. P. Soutter, "Quantitative Assessment of the AW phenomenon and determination of its biophysical origins as a means for in vivo assessing CIN-related micro-structural and functional changes," *J. Low. Genit. Tract Dis.*, vol. 14, no. 3, pp. 264–269, Jul. 2010.

- [158] M. Rodriguez-Fernandez and J. Banga, "Global sensitivity analysis of a biochemical pathway model," *2nd Int. Work. ...*, no. 0, pp. 233–242, 2009.
- [159] R. Li, M. a. Henson, and M. J. Kurtz, "Selection of model parameters for off-line parameter estimation," *Control Syst. Technol. IEEE ...*, vol. 12, no. 3, pp. 402–412, May 2004.
- [160] S. Vajda, P. Valko, and T. Turanyi, "Principal component analysis of kinetic models," *Int. J. Chem. Kinet.*, vol. 17, no. 1, pp. 55–81, Jan. 1985.
- [161] K. R. Godfrey and J. . Distefano, *Identification and System Parameter Estimation*. Oxford, England: Pergamon Press, 1985.
- [162] A. Saltelli, M. Ratto, S. Tarantola, and F. Campolongo, "Sensitivity analysis for chemical models.," *Chem. Rev.*, vol. 105, no. 7, pp. 2811–28, Jul. 2005.
- [163] S. Dobre, T. Bastogne, M. Barberi-Heyob, and A. Richard, "Identifiability and sensitivity analysis of a Photodynamic Therapy model," *Procedia - Soc. Behav. Sci.*, vol. 2, no. 6, pp. 7648–7649, Jan. 2010.
- [164] S. Wagner, "Global sensitivity analysis of predictor models in software engineering," *Predict. Model. Softw. Eng. 2007. ...*, 2007.
- [165] A. Saltelli, *Global sensitivity analysis: the primer*. Chichester: John Wiley, 2008, p. x, 292 p.
- [166] I. M. Sobol', "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates," *Math. Comput. Simul.*, vol. 55, no. 1–3, pp. 271–280, 2001.
- [167] I. Dimov and R. Georgieva, "Monte Carlo algorithms for evaluating Sobol' sensitivity indices," *Math. Comput. Simul.*, vol. 81, no. 3, pp. 506–514, Nov. 2010.
- [168] I. M. Sobol' and S. Kucherenko, "Global sensitivity indices for nonlinear mathematical models, Review," *Wilmott Mag*, no. 2000, pp. 2–7, 2005.
- [169] G. Li, S.-W. Wang, and H. Rabitz, "Practical Approaches To Construct RS-HDMR Component Functions," *J. Phys. Chem. A*, vol. 106, no. 37, pp. 8721–8733, Sep. 2002.
- [170] G. E. B. Archer, A. Saltelli, and I. M. Sobol, "Sensitivity measures, ANOVA-like techniques and the use of bootstrap," *J. Stat. Comput. Simul.*, vol. 58, no. 2, pp. 99–120, May 1997.
- [171] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola, "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index," *Comput. Phys. Commun.*, vol. 181, no. 2, pp. 259–270, 2010.
- [172] T. Homma and A. Saltelli, "Importance measures in global sensitivity analysis of nonlinear models," *Reliab. Eng. Syst. Saf.*, vol. 52, 1996.
- [173] J. Sacks, W. Welch, T. Mitchell, and H. Wynn, "Design and analysis of computer experiments," *Stat. Sci.*, vol. 4, no. 4, pp. 409–435, 1989.
- [174] R. Cukier, H. Levine, and K. Shuler, "Nonlinear sensitivity analysis of multiparameter model systems," *J. Comput. Phys.*, vol. 42, pp. 1–42, 1978.
- [175] A. Saltelli, S. Tarantola, and K. Chan, "A quantitative model-independent method for global sensitivity analysis of model output," *Technometrics*, 1999.
- [176] S. Dobre, T. Bastogne, and C. Profeta, "Limits of variance-based sensitivity analysis for non-identifiability testing in high dimensional dynamic models," *Automatica*, vol. 1, no. November 2011, pp. 2740–2749, 2012.

- [177] M. Rodriguez-Fernandez, P. Mendes, and J. R. Banga, "A hybrid approach for efficient and robust parameter estimation in biochemical pathways.," *Biosystems.*, vol. 83, no. 2–3, pp. 248–65, 2006.
- [178] Y. F. Chu and J. Hahn, "Quantitative optimal experimental design using global sensitivity analysis via quasi-linearization," *Ind. Eng. Chem. Res.*, vol. 49, no. 49, pp. 7782–7794, 2010.
- [179] R. J. Gillies, N. Raghunand, G. S. Karczmar, and Z. M. Bhujwala, "MRI of the tumor microenvironment.," *J. Magn. Reson. Imaging*, vol. 16, no. 4, pp. 430–50, Oct. 2002.
- [180] I. F. Tannock and D. Rotin, "Acid pH in tumors and its potential for therapeutic exploitation," *Cancer Res.*, vol. 49, pp. 4373–4384, 1989.
- [181] A. R. Karuri, E. Dobrowsky, and I. F. Tannock, "Selective Cellular Acidification and Toxicity of Weak Organic-Acids in an Acidic Microenvironment," *Br. J. Cancer*, vol. 68, no. 6, pp. 1080–1087, Dec. 1993.
- [182] P. Swietach, R. D. Vaughan-Jones, and A. L. Harris, "Regulation of tumor pH and the role of carbonic anhydrase 9.," *Cancer Metastasis Rev.*, vol. 26, no. 2, pp. 299–310, Jul. 2007.
- [183] R. a Gatenby, E. T. Gawlinski, A. F. Gmitro, B. Kaylor, and R. J. Gillies, "Acid-mediated tumor invasion: a multidisciplinary study.," *Cancer Res.*, vol. 66, no. 10, pp. 5216–23, May 2006.
- [184] A. Avdeef, "Leakiness and size exclusion of paracellular channels in cultured epithelial cell monolayers-interlaboratory comparison.," *Pharm. Res.*, vol. 27, no. 3, pp. 480–9, Mar. 2010.
- [185] D. Walker, "Modelling the electrical properties of cervical epithelium.," *PhD Thesis - Univ. Sheff.*, no. October, 2001.
- [186] C. G. Moles, P. Mendes, and J. R. Banga, "Parameter estimation in biochemical pathways: a comparison of global optimization methods.," *Genome Res.*, vol. 13, no. 11, pp. 2467–74, Nov. 2003.
- [187] C. a. Floudas and C. E. Gounaris, "A review of recent advances in global optimization," *J. Glob. Optim.*, vol. 45, no. 1, pp. 3–38, Aug. 2008.
- [188] C. a. Meyer, C. a. Floudas, and A. Neumaier, "Global Optimization with Nonfactorable Constraints," *Ind. Eng. Chem. Res.*, vol. 41, no. 25, pp. 6413–6424, Dec. 2002.
- [189] E. Davis and M. Ierapetritou, "A Kriging-Based Approach to MINLP Containing Black-Box Models and Noise," *Ind. Eng. Chem. Res.*, vol. 47, no. 16, pp. 6101–6125, Aug. 2008.
- [190] W. L. Price, "A controlled random search procedure for global optimization," *Comput. J.*, vol. 20, no. 4, pp. 367–370, 1976.
- [191] W. L. Price, "Global optimization by controlled random search," *J. Optim. Theory Appl.*, vol. 40, no. 3, pp. 333–348, 1983.
- [192] J. Tvrdik, "Generalized controlled random search and competing heuristics," in *10th International Conference on Soft Computing*, 2004, no. 1, pp. 228–233.
- [193] Q. Y. Duan, V. K. Gupta, and S. Sorooshian, "Shuffled Complex Evolution Approach for Effective and Efficient Global Minimization," *J. Optim. Theory Appl.*, vol. 76, no. 3, pp. 501–521, 1993.
- [194] P. Li, H. Tang, and Z. Wang, "Parameter Estimation Using a SCE Strategy," *Neural Inf. Process.*, pp. 107–113, 2009.

- [195] S. Das and P. N. Suganthan, "Differential Evolution : A Survey of the," *IEEE Trans. Evol. Comput.*, pp. 1–28, 2010.
- [196] R. Storn and K. Price, "Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces," *J. Glob. Optim.*, pp. 341–359, 1997.
- [197] M. F. Cardoso, R. L. Salcedo, and S. Feyo de Azevedo, "The simplex-simulated annealing approach to continuous non-linear optimization," *Comput. Chem. Eng.*, vol. 20, no. 9, pp. 1065–1080, Sep. 1996.
- [198] L. Ingber, "Very fast simulated re-annealing," *Math. Comput. Model.*, vol. 12, no. 8, pp. 967–973, Jan. 1989.
- [199] J. D. Vicente, J. Lanchares, and R. Hermida, "Annealing placement by thermodynamic combinatorial optimization," *ACM Trans. Des. Autom. Electron. Syst.*, vol. 9, no. 3, pp. 310–332, Jul. 2004.
- [200] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95 - International Conference on Neural Networks*, vol. 4, pp. 1942–1948.
- [201] M. Clerc and J. Kennedy, "The particle swarm - explosion, stability, and convergence in a multidimensional complex space," *IEEE Trans. Evol. Comput.*, vol. 6, no. 1, pp. 58–73, 2002.
- [202] L. Yiqing, Y. Xigang, and L. Yongjian, "An improved PSO algorithm for solving non-convex NLP/MINLP problems with equality constraints," *Comput. Chem. Eng.*, vol. 31, no. 3, pp. 153–162, Jan. 2007.
- [203] M. Schlüter, M. Gerdt, and J. Rückmann, "A numerical study of MIDACO on 100 MINLP benchmarks," *Optimization*, no. 1, 2012.
- [204] K. Socha and M. Dorigo, "Ant colony optimization for continuous domains," *Eur. J. Oper. Res.*, vol. 185, no. 3, pp. 1155–1173, Mar. 2008.
- [205] M. Schlüter, J. Egea, and J. Banga, "Extended ant colony optimization for non-convex mixed integer nonlinear programming," *Comput. Oper. Res.*, pp. 1–24, 2009.
- [206] T. P. Runarsson, "Stochastic ranking for constrained evolutionary optimization," *IEEE Trans. Evol. Comput.*, vol. 4, no. 3, pp. 284–294, 2000.
- [207] M. Bjorkman and K. Holmstrom, "Global optimization using the DIRECT algorithm in Matlab," *AMO - Adv. Model. Optim.*, vol. 1, no. 2, pp. 17–37, 1999.
- [208] D. Jones, C. Perttunen, and B. Stuckman, "Lipschitzian optimization without the Lipschitz constant," *J. Optim. Theory Appl.*, vol. 79, no. 1, 1993.
- [209] P. Polisetty, E. Voit, and E. Gatzke, "Identification of metabolic system parameters using global optimization methods," *Theor. Biol. ...*, 2006.
- [210] S. Holm, "A simple sequentially rejective multiple test procedure," *Scand. J. Stat.*, vol. 6, no. 2, pp. 65–70, 1979.
- [211] a. E. Eiben and S. K. Smit, "Parameter tuning for configuring and analyzing evolutionary algorithms," *Swarm Evol. Comput.*, vol. 1, no. 1, pp. 19–31, Mar. 2011.
- [212] K. Tashkova, P. Korošec, and J. Šilc, "Parameter estimation with bio-inspired meta-heuristic optimization: modeling the dynamics of endocytosis," *BMC Syst. Biol.*, vol. 5, no. 159, p. 26, 2011.

- [213] M. Pedersen, "Good parameters for differential evolution," 2010.
- [214] K. Parsopoulos and M. Vrahatis, "Recent approaches to global optimization problems through particle swarm optimization," *Nat. Comput.*, vol. 1, no. 2–3, pp. 235–306, 2002.
- [215] J. F. Schutte and A. A. Groenwold, "A Study of Global Optimization Using Particle Swarms," *J. Glob. Optim.*, vol. 31, no. 1, pp. 93–108, Jan. 2005.
- [216] D. C. Walker, R. H. Smallwood, A. Keshtar, B. A. Wilkinson, F. C. Hamdy, and J. A. Lee, "Modelling the electrical properties of bladder tissue - quantifying impedance changes due to inflammation and oedema," *Physiol. Meas.*, vol. 26, no. 3, pp. 251–268, 2005.
- [217] O. Kierkegaard, C. Byralsen, K. C. Hansen, K. H. Frandsen, and M. Frydenberg, "Association between colposcopic findings and histology in cervical lesions: the significance of the size of the lesion.," *Gynecol. Oncol.*, vol. 57, no. 1, pp. 66–71, Apr. 1995.
- [218] M. R. Jarmulowicz, D. Jenkins, S. E. Barton, A. L. Goodall, A. Hollingworth, and A. Singer, "Cytological status and lesion size: a further dimension in cervical intraepithelial neoplasia.," *Br. J. Obstet. Gynaecol.*, vol. 96, no. 9, pp. 1061–6, Sep. 1989.
- [219] P. Tidbury, A. Singer, and D. Jenkins, "CIN 3: the role of lesion size in invasion.," *Br. J. Obstet. Gynaecol.*, vol. 99, no. 7, pp. 583–6, Jul. 1992.
- [220] E. Burghardt, H. Pickel, and F. Girardi, *Colposcopy-cervical pathology: textbook and atlas*. New York: Thieme, 1998, p. 344.
- [221] C. Kärrberg, "Cervical dysplasia and cervical cancer in pregnancy: diagnosis and outcome," Goteborg, 2012.
- [222] P. Kolstad, "THE COLPOSCOPICAL DIAGNOSIS OF DYSPLASIA, CARCINOMA IN SITU, AND EARLY INVASIVE CANCER OF THE CERVIX.," *Acta Obstet. Gynecol. Scand.*, vol. 43, p. SUPPL 7:105–8, Jan. 1965.
- [223] P. Kolstad, "THE DEVELOPMENT OF THE VASCULAR BED IN TUMOURS AS SEEN IN SQUAMOUS-CELL CARCINOMA OF THE CERVIX UTERI.," *Br. J. Radiol.*, vol. 38, pp. 216–23, Mar. 1965.
- [224] D. G. Ferris, T. J. Cox, D. O'Connor, and C. V. Wright, *Modern colposcopy: textbook and atlas*. Dubuque: Kendall-Hunt Publishing Co., 2004, p. 712.
- [225] G. I. Gorodeski, M. F. Romero, U. Hopfer, E. Rorke, W. H. Utian, and R. L. Eckert, "The cultured human cervical epithelium: a new model for studying paracellular transport.," *J. Soc. Gynecol. Investig.*, vol. 3, no. 5, pp. 267–80, 1996.
- [226] A. Singer, J. Monaghan, and S. Quek, *Lower genital tract precancer: colposcopy, pathology and treatment*. Oxford, England: Wiley-Blackwell, 2000, p. 323.
- [227] J. O. Dada and P. Mendes, "Multi-scale modelling and simulation in systems biology.," *Integr. Biol. (Camb)*, vol. 3, no. 2, pp. 86–96, Mar. 2011.
- [228] H. Huber, G. Gomez Estrada, H. Dussmann, C. O'Connor, and M. Rehm, "Extending the Explanatory Power of Live Cell Imaging by computationally modelling the Execution of Apoptotic Cell Death ," in *Modern Research and Educational Topics in Microscopy*, vol. 1, A. Mendez-Vilas and J. Diaz, Eds. 2007, pp. 88–99.
- [229] K. Welsher, S. P. Sherlock, and H. Dai, "Deep-tissue anatomical imaging of mice using carbon nanotube fluorophores in the second near-infrared window.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, no. 22, pp. 8943–8, May 2011.

- [230] D. J. Cuccia, F. Bevilacqua, A. J. Durkin, S. Merritt, B. J. Tromberg, G. Gulsen, H. Yu, J. Wang, and O. Nalcioglu, "In vivo quantification of optical contrast agent dynamics in rat tumors by use of diffuse optical spectroscopy with magnetic resonance imaging coregistration.," *Appl. Opt.*, vol. 42, no. 16, pp. 2940–50, Jun. 2003.
- [231] M. Gurfinkel, A. B. Thompson, W. Ralston, T. L. Troy, A. L. Moore, T. A. Moore, J. D. Gust, D. Tatman, J. S. Reynolds, B. Muggenburg, K. Nikula, R. Pandey, R. H. Mayer, D. J. Hawrysz, and E. M. Sevick-Muraca, "Pharmacokinetics of ICG and HPPH-car for the detection of normal and tumor tissue using fluorescence, near-infrared reflectance imaging: a case study.," *Photochem. Photobiol.*, vol. 72, no. 1, pp. 94–102, Jul. 2000.
- [232] N. Kosaka, M. Mitsunaga, M. R. Longmire, P. L. Choyke, and H. Kobayashi, "Near infrared fluorescence-guided real-time endoscopic detection of peritoneal ovarian cancer nodules using intravenously injected indocyanine green.," *Int. J. cancer*, vol. 129, no. 7, pp. 1671–7, Oct. 2011.
- [233] P. Hsiung, J. Hardy, S. Friedland, R. Soetikno, D. CB, A. Wu, P. Sahbale, J. Crawford, A. Lowe, C. Contag, and T. Wang, "Detection of colonic dysplasia in vivo using a targeted fluorescent septapeptide and confocal microendoscopy," *Nat. Med.*, vol. 14, no. 4, pp. 454–458, 2012.
- [234] M. S. Hoetker and M. Goetz, "Molecular imaging in endoscopy," *United Eur. Gastroenterol. J.*, vol. 1, no. 2, pp. 84–92, Mar. 2013.
- [235] S. Krishnamurthy, F. Meric-Bernstam, A. Lucci, R. F. Hwang, H. M. Kuerer, G. Babiera, F. C. Ames, B. W. Feig, M. I. Ross, E. Singletary, K. K. Hunt, and I. Bedrosian, "A prospective study comparing touch imprint cytology, frozen section analysis, and rapid cytokeratin immunostain for intraoperative evaluation of axillary sentinel lymph nodes in breast cancer.," *Cancer*, vol. 115, no. 7, pp. 1555–62, Apr. 2009.
- [236] J. P. Tiernan, I. Ansari, N. a Hirst, P. a Millner, T. a Hughes, and D. G. Jayne, "Intra-operative tumour detection and staging in colorectal cancer surgery.," *Colorectal Dis.*, vol. 14, no. 9, pp. e510–20, Sep. 2012.
- [237] J. Rao, A. Dragulescu-Andrasi, and H. Yao, "Fluorescence imaging in vivo: recent advances.," *Curr. Opin. Biotechnol.*, vol. 18, no. 1, pp. 17–25, Feb. 2007.
- [238] C. Hadjipanayis, H. Jiang, D. Roberts, and L. Yang, "Current and future clinical applications for optical imaging of cancer: from intraoperative surgical guidance to cancer screening," *Semin. Oncol.*, vol. 38, no. 1, pp. 109–118, 2011.
- [239] G. M. van Dam, G. Themelis, L. M. A. Crane, N. J. Harlaar, R. G. Pleijhuis, W. Kelder, A. Sarantopoulos, J. S. de Jong, H. J. G. Arts, A. G. J. van der Zee, J. Bart, P. S. Low, and V. Ntziachristos, "Intraoperative tumor-specific fluorescence imaging in ovarian cancer by folate receptor- α targeting: first in-human results.," *Nat. Med.*, vol. 17, no. 10, pp. 1315–9, Oct. 2011.
- [240] K. Rosbach, M. Williams, A. Gillenwater, and R. Richards-Kortum, "Molecular Imaging of Multiple Biomarkers of Epithelial Neoplasia: Epidermal Growth Factor Receptor Expression and Metabolic Activity in Oral Mucosa," *Transl. Oncol.*, vol. 5, no. 3, pp. 160–171, 2012.
- [241] K. Sano, M. Mitsunaga, T. Nakajima, P. L. Choyke, and H. Kobayashi, "In vivo breast cancer characterization imaging using two monoclonal antibodies activatably labeled with near infrared fluorophores.," *Breast Cancer Res.*, vol. 14, no. 2, p. R61, Jan. 2012.
- [242] E. L. Bird-Lieberman, A. a Neves, P. Lao-Sirieix, M. O'Donovan, M. Novelli, L. B. Lovat, W. S. Eng, L. K. Mahal, K. M. Brindle, and R. C. Fitzgerald, "Molecular imaging using fluorescent lectins permits rapid endoscopic identification of dysplasia in Barrett's esophagus.," *Nat. Med.*, vol. 18, no. 2, pp. 315–21, Feb. 2012.

- [243] H. Kobayashi, Y. Koyama, T. Barrett, Y. Hama, C. A. S. Regino, I. S. Shin, B.-S. Jang, N. Le, C. H. Paik, P. L. Choyke, and Y. Urano, "Multi-Modal Nano-Probes for Radionuclide and 5-color Near Infrared Optical Lymphatic Imaging," *ACS Nanotechnol.*, vol. 1, no. 4, pp. 258–264, 2007.
- [244] M. V Marshall, J. C. Rasmussen, I.-C. Tan, M. B. Aldrich, K. E. Adams, X. Wang, C. E. Fife, E. a Maus, L. a Smith, and E. M. Sevick-Muraca, "Near-Infrared Fluorescence Imaging in Humans with Indocyanine Green: A Review and Update.," *Open Surg. Oncol. J.*, vol. 2, no. 2, pp. 12–25, Jan. 2010.
- [245] Y. Urano, M. Sakabe, N. Kosaka, M. Ogawa, M. Mitsunaga, and D. Asanuma, "Rapid cancer detection by topically spraying a γ -glutamyltranspeptidase-activated fluorescent probe.," *Sci. Transl. Med.*, 2011.
- [246] D. J. COL. Horwhat and C. FACG, "Chromoendoscopy in the Esophagus," *Visible Hum. J. Endosc.*, vol. 10, no. 3, 2011.