Technical University of Crete
Electronics and Computer Engineering School

Development of hyperspectral microscopy for improving the diagnostic accuracy in leukemia diagnosis

Diploma Thesis by Dimitrios Gkotsoulias

Thesis Committee:

Professor Costas Balas (Supervisor)

Professor Michael Zervakis

Professor Minos Garofalakis

Chania, May 2016

Contents

ACKNOWLEDGEMENTS

I wish to express my appreciation to my professor, Costas Balas, for the opportunity, the guidance and support during the elaboration of this Diploma Thesis.

For their valuable help, suggestions, friendship and kindness I wish to thank the members of the Electronics Lab, Theodoros-Marios Giakoumakis, Athanasios Tsapras and Christos Rossos.

For the procurement of blood and bone marrow stained samples and guidance in medical aspects, I wish to thank the Director of the Institute of Applied Biosciences at CERTH, prof. Kostas Stamatopoulos and Dr. Aliki Xocheli, post-doc researcher at CERTH.

I would also like to thank professors Minos Garofalakis and Michalis Zervakis for their participation as committee for this Thesis.

Last, I would like to express my deep appreciation to my family, for their support, courage and advising through all the years of my studies.

ABSTRACT

Spectral Imaging is a combination of Spectroscopy and Imaging. It is based on the acquisition of a collection of narrow spectral band images (Spectral cube). In Hematology, the May Grunwald-Giemsa (MG-G), a solution mixture of Eosin Y, Azure B and Methylene Blue stains, is commonly used for leukemia diagnosis in blood and bone marrow samples. Diagnosis is based on manual observation of stained tiles. This procedure is time consuming and subjective, as it depends on the experience and judgement of the observer-hematologist. In this diploma thesis, we present a novel Spectral Imaging Microscopy modality, for mapping the concentrations of MG-G substances. Partial Least Squares (PLS) regression algorithm is employed as most efficient in chemo-metrics. Using the spectral cube absorbance data the concentration of each stain is estimated per pixel. Using Dimensionality Reduction Analysis the minimum essential number of spectral bands is estimated, ensuring minimum loss in accuracy and much faster data analysis. Finally, four scaled concentration pseudo-color maps are produced, one for each substance of MG-G and one depicting the proportions of their concentrations, which are used for the identification of the leukemic cells, in an objective and reproducible diagnosis procedure.

1. THEORETICAL BACKGROUND

1.1 Electromagnetic Radiation - Spectrum

Electromagnetic radiation (EMR) is a form of radiant energy released by certain electromagnetic processes. The physical model of EMR is a transverse wave, consisting of an oscillating electrical field E and an oscillating magnetic field M. The two fields are perpendicular to each other as well as to the propagation direction of the wave. The characteristics of the electromagnetic wave wavelength λ (physical length of a full oscillation) and frequency v (number of oscillations per second).

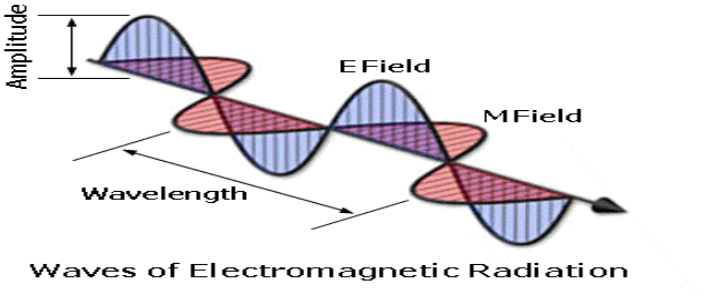


Figure 1. The electromagnetic field

Electromagnetic radiation has characteristics of both wave and particle. This property is called Wave-Particle Duality. Electromagnetic energy is a continuous row of particles or wave energy packages, called photons [1].One photon's energy equals to:

$$\boldsymbol{Ephoton = h * v = h * c/\lambda}$$

Where h is Planck's constant (6.6261_10^34 Js), c is the speed of light (3*10^8 km/h), v is the frequency and λ is the wavelength of the radiation. Depending on wavelength, the electromagnetic spectrum is divided into a number of spectral bands, each one of which interacts with matter in a different way [2]. Those are shown in the figure below along with their wavelengths:

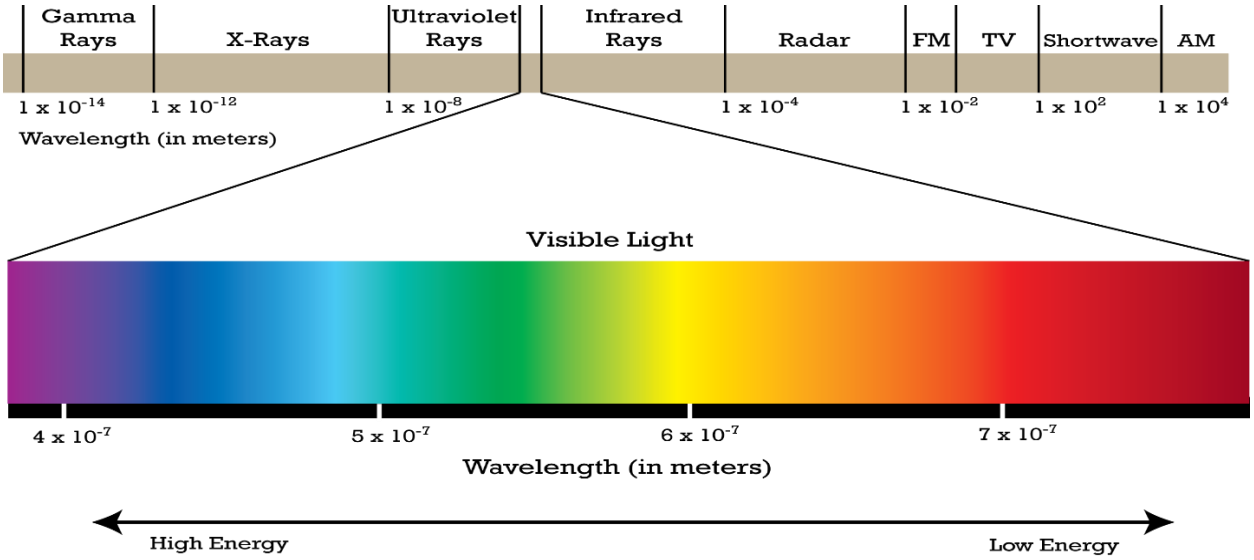


Figure 2. The electromagnetic spectrum

Note that infrared rays are also separated to Near Infrared (NIR), Mid Infrared (MIR), Far Infrared (FIR).

1.2 Spectroscopy – Spectrometry – Imaging spectroscopy

Generally, spectroscopy is the study of interactions between electromagnetic radiation and matter. In expansion to that, we can refer as spectroscopy to any measurement of any quantity as function of wavelength of frequency. Spectroscopic data is often represented by a plot called spectrum which is the response of interest as a function of wavelength or frequency. Spectrometry is the spectroscopic method that is used each time, to estimate the value of a characteristic of interest. Imaging spectroscopy is the application of spectroscopy in every pixel of a special image. Measuring characteristics pixel-wise in different wavelengths can lead to results with great potential in a great range of applications [3].

1.3 Hyperspectral Imaging

Hyperspectral imaging is a modality, hybrid between imaging and spectroscopy, and has been applied to numerous areas of science, now emerging in biomedical engineering. The main idea is acquiring two dimensional images across a wide range of electromagnetic spectrum (UV, visible, infrared) using a two dimensional detector array (CCD or CMOS). Then, a three dimensional dataset of spatial and spectral information is generated. That dataset is called hyper spectral cube. With the spatial information we can locate more accurately the light's interaction with the target, pixel by pixel [4].

We can assume that each pixel, depending on the light it collected from the target on each wavelength, has its own spectral signature (measurement of emitted, reflected or absorbed electromagnetic radiation at specific (varying for each target) wavelengths which can uniquely identify an object), as shown in the picture:

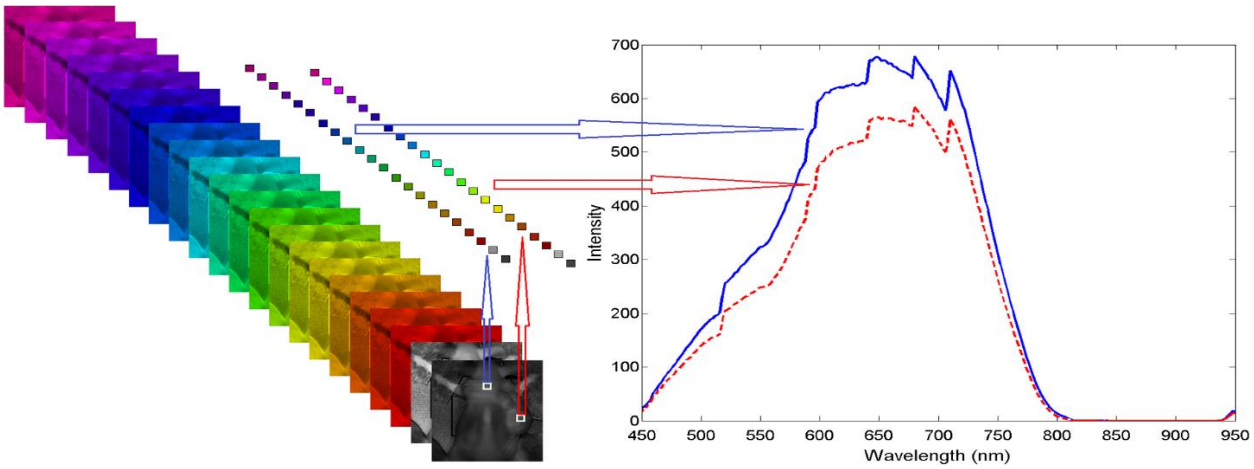


Figure 3. Spectral Characteristics-Signatures

Thus, the images of the hyperspectral cube, can be used to extract unique data concerning the target by classifying the pixels using the spectral signatures pixel by pixel. When light propagates through a biological tissue or sample, undergoes multiple scattering and absorption. Scattering, fluorescence and absorption characteristics are unique for each kind of tissue and when we have a pathological condition, they seem to change depending on the condition's progression. Capturing that reflected, fluorescent or transmitted light with the right tools, in different wavelengths and processing those data, may result to quantitative and significant diagnostic information about the specific biological target's pathology. So in case of biomedical applications, hyperspectral imaging, leads to identification of various pathological conditions and gives us the opportunity for extended noninvasive diagnosis and much more, like surgical guidance etch.

1.4 Spectral Signatures

As we mentioned above, for any given material, when exposed to light, the amount of radiation absorbed, reflected, transmitted varies with the wavelength, as shown in picture above. This property of matters give us the opportunity to uniquely identify different physical or chemical substances and separate them using their spectral signatures [5]. This separation is also known as spectral classification.

1.5 Pixel-wise Spectral Classification

Repeating the classification procedure for every pixel on the spatial coordinates od the spectral cube, is leading to thematic maps, just by giving different values to each class identified. Let us assume as an example that we have a spectral cube of data captured on a tissue with cancerous malignant and benign areas. The comparison of the spectral response of each pixel, establishes a simple rule of classification using the range where the spectral response differs considerably. In our example, we can use the band 297nm or 345nm, so we have for a pixel p: if |p(345)-Normal(345)|< |p(345)-Cancer(345)|→ p is benign, else, p is malignant.

After classification pixel by pixel (or by pixel groups), we produce the thematic map. We can use different colors to separate the pathological areas of the tissue from the healthy ones. For example we can use white for benign areas and black for malignant areas. This is schematically explained on the next image:

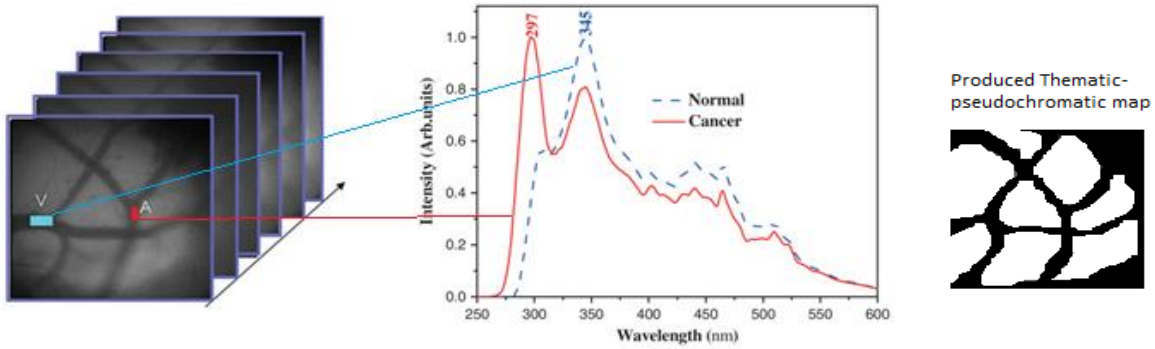


Figure 4. Pixel-wise Spectral classification

When in the produced thematic map by classification or unmixing , different artificial colors or gray shades are used to represent different data (clusters), it is called pseudochromatic map and it's the most common way of representing data in biomedical spectral imaging as it's effectiveness is proven in several uses, from cancern detection to surgical guidance. This procedure is just a quite simple example of how classification is acomplished. Of course there exist many different ways regarding the use ,the features we want to overemphasize and generally the result we want to achieve.

1.6 Quantitative analysis and Chemometrics

1.6.1 Quantitative analysis

Quantitative analysis as used in chemistry, chemical engineering and physics is the determination of the absolute or relative abundance (often expressed as a concentration) of one, several or all particular substance(s) present in a sample.

1.6.2 Chemometrics

Chemometrics is the use of statistical and mathematical techniques in order to extract information from chemical systems using multivariate statistics and applied mathematics to address problems in chemistry, biochemistry, medicine, biology and chemical engineering. The field is generally recognized to have emerged in the 1970s as computers became increasingly exploited for scientific research purposes. Svante Wold and Bruce Kowalski, two pioneers in the field introduced the term chemometrics for first time in science. Wold was a professor of organic chemistry at Umea University, Sweden, and Kowalski was a professor of analytical chemistry at University of Washington, Seattle. Chemometrics usually involves using linear algebra methods to make qualitative or quantitative measurements of chemical data [6] [7] [8] [9].

In this diploma thesis, we have dealt with quantitative analysis and chemometrics in order to deconvolute and estimate the concentrations of the different dies on stained blood and bone marrow samples. These concentrations are measured in Molarity (units: mol/L or M), which

represents the number of moles of a dissolved substance per liter of solution [8]. The algorithm we chose to use is the Partial Least Square (PLS) algorithm. PLS's efficiency on Concentration estimation was validated by Fani Abatzi in her Diploma thesis. Among other deconvolution algorithms this one gave us me most efficient results on microscopy imaging.

1.6.3 Beer-Lampert Law Generalization and use

The assumption that the concentrations of the constituents of interest in the samples are somehow related to the data from a measurement technique (spectral imaging in our case) is the key to quantitative analysis of those data. The ultimate goal is to create a calibration equation (or equations) which, when applied to data of unknown samples measured in the same manner, will accurately predict the quantities of the constituents of interest.

In order to calculate these equations, we have to establish a set of samples with known properties (standards). The standards are then measured by an instrument. Together, this collection of known data (the composition of each standard) and the measured data from the instrument form what is known as a training set. The calibration equations that describe the relationships between these two sets of information are calculated from this data. Beer-Lambert Law defines a simple linear relationship between the spectrum and the composition of a sample. It forms the basis of nearly all other chemometrics methods for spectroscopic data. Simply stated, the law claims that when a sample is placed in the beam of a spectrometer, there is a direct and linear relationship between the amount (concentration) of its constituent(s) and the amount of energy it absorbs.

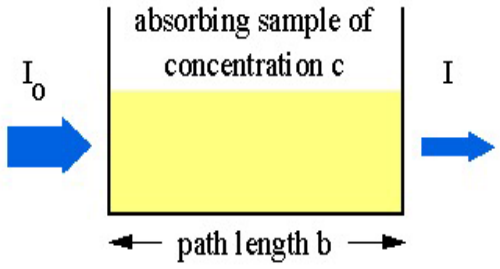


Figure 5. Schematic for Beer-Lambert Law

In mathematical terms (Image 5 for schematic):
Absorbance = log(I0/I) = -log(Transmittance) = log(1/Transmittance) = ε*b*C
Where,
I0= Intensity of the incident light.
I = Intensity of the transmitted light.
T = Transmittance (I/I0), usually expressed as a percentage %T.
C = the concentration of the sample's solution measured in mol/L or M (molarity).
b = the pathlength that the light beam has travelled inside the sample (in cm).
ε = the molar absorptivity of the solution, which is a constant number also proportional to the respective absorbance wavelengths.

The Beer-Lambert law is valid under the following conditions:
1. The solutions are not dense.

2. The only mechanism for the interaction between a dissolved substance and radiation is absorption.
3. The incident radiation to a sample is monochromatic.
4. The sample is in a cuvette (quartz glass in our case) with a uniform intersection.
5. The absorbing molecules act individually (no reaction between substances), means that for n dissolved substances in the mixture, we have: Absorbance = Absorbance(1) + Absorbance(2) +…….+ Absorbance(n) [10].

1.6.4 Univariate Regression (Calibration) Model (UCM)

Univariate linear model, is the one where we have a single response denoted by a single observation. As an example, we have the concentration of one substance as a depended variable and the absorbance on a single wavelength as the independent variable [11]. Symbolizing as y the response, x the absorbance, b the regression coefficients and e the noise, in linear algebra, we have:

$$y = x * b + e$$

Depicted in matrix terms, as:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} b_1$$

Where n, is the number of samples. The above is the simplest kind of linear regression, in which a single variable(x) is associated to a single variable(y) and we have on intercept. Supposing we have more than one wavelength absorbance data, we move to Multiple Linear Regression models (MLR), where the linear relationship between the data is now depicted as:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1d} \\ x_{21} & x_{22} & \dots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nd} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_d \end{bmatrix}$$

Where n is the number of samples and d is the number of wavelengths.

1.6.5 Multiple Linear Regression models (MLR)

As mentioned in 1.6.4 working with non-single variables on regression analysis is introducing MLR. As in univariate regression, the basic idea of a multiple linear model is to specify the relationship between a dependent (response) variable *Y*, and a set of predictor variables X, so that:

$$Y=XB$$

$$Y = b_0 + b_1X_1 + b_2X_2 + ... + b_pX_p$$

In this equation [*b1,b2….,bn*] are the regression coefficients computed from the data (for variables 1 through *p*) and *b0* the regression coefficient for the intercept. Actually, on our work, we have a multiple multivariate regression because X, Y are both non-single variables. We deal with many samples, more than 10 wavelengths, and three mixed substances concentrations.

For example, in chemo-metrics, one could predict substance's concentration in a mixture (Y) as a function of the mixtures spectral absorbance characteristics in certain wavelengths (X). You could use linear regression to estimate the respective regression coefficients from a sample of data, measuring absorbance, and observing the mixture's concentration (or concentrations). Analyzing the estimates of the linear relationships between predictors and observed data, is a key step for making reasonable predictions for new observations [12].

1.6.6 Partial Least Squares algorithm (PLS) and implementations

MLR model has now been extended in a number of ways and is used on more sophisticated data analysis problems. It is the basis for a number of multivariate methods like *discriminant analysis* (i.e., the prediction of group membership from the levels of continuous predictor variables), *principal components regression* (i.e., the prediction of responses on the dependent variables from factors underlying the levels of the predictor variables), and *canonical correlation* (i.e., the prediction of factors underlying responses on the dependent variables from factors underlying the levels of the predictor variables). These methods impose restrictions:

- factors underlying the *Y* and *X* variables are extracted from the $\boldsymbol{Y'Y}$ and $\boldsymbol{X'X}$ matrices, respectively, and never from cross-product matrices involving both the *Y* and *X* variables
- Number of prediction functions can never exceed the minimum of the number of *Y* variables and *X* variables.

Partial least squares regression (PLS) extends multiple linear regression without imposing the restrictions of the aforementioned ways of analysis. In PLS, prediction functions are represented by factors extracted from the $\boldsymbol{Y'XX'Y}$ matrix. The number of such prediction functions that can be extracted typically will exceed the maximum of the number of *Y* and *X* variables making PLS, probably, the least restrictive one of the MLR models. This flexibility allows it to be used in situations where the use of traditional multivariate methods is severely limited, such as when there are fewer observations than predictor variables or very poor a priori data. PLS has become a standard tool in chemometrics, for modeling relations of predictors and responses in multivariate analysis.

In detail, PLS is mostly related to Principal Component Regression (PCR). The two regressions differ in the methods used in extracting factor scores. In short, principal components

regression produces the weight matrix $\boldsymbol{W}$ reflecting **the covariance structure between the predictor variables**, while partial least squares regression produces the weight matrix $\boldsymbol{W}$ reflecting **the covariance structure between the predictor and response variables**. In PLS, instead of first decomposing the spectral matrix into a set of eigenvectors and scores, and regressing them against the concentrations as a separate step, the concentration matrix information are used during the decomposition process. This causes spectra containing higher constituent concentrations to be weighted more heavily than those with low concentrations.
PLS is simply taking advantage of the correlation relationship that already exists between the spectral data and the constituent concentrations. Since the spectral data can be decomposed into its most common variations, so can the concentration data. Generally, PLS generates two sets of vectors and two sets of corresponding scores, one set for the spectral data, and the other for the constituent concentrations. Presumably, the two sets of scores are related to each other through regression, and a calibration model is constructed [10], [12].

As mentioned, PLS performs the decomposition on both the spectral and concentration data simultaneously. As each new factor is calculated for the model, the scores are "swapped" before the contribution of the factor is removed from the raw data. The newly reduced data matrices are then used to calculate the next factor, and the process is repeated until the desired number of factors is calculated. Resulting to a matrix of regression coefficients, the training procedure gives us the ability to predict the concentrations in unknown samples. A detailed description of the standard version of PLS algorithm is [10]:

$Initialize\ j = 1, X1 = X,\ Y1 = Y,$ then proceed to find the g latent variables:

1. $w_j = {X_j}^T * \ y_j / norm({X_j}^T * y_j)$
2. $Let\ t_j = X_j * w_j$
3. $Let\ \widehat{c_j} = {t_j}^T * y_j / {t_j}^T * \ t_j$
4. $Let\ p_j = {X_j}^T * \ t_j / {t_j}^T * \ t_j$
5. $Let\ X_{j+1} = \ X_j - t_j * {p_j}^T$ and $\ y_{j+1} = \ y_j - t_j * \widehat{c_j}$
6. $If\ j = g\ , stop, otherwise\ j = j + 1\ and\ return\ to\ Step$ 1.
7. $Form\ the\ two\ dxg\ matrices\ W, P\ and\ nxg\ matrix\ T\ with\ columns\ w_j, p_j, t_j\ respectively, as\ well\ as\ the\ vector\ \hat{c}\ with\ elements\ \widehat{c_j}.$
8. $\hat{X} = T * P^T$
9. $\hat{y} = T * \hat{c} = X * W * (P^T * W)^{-1} * \hat{c} = X * B$ where B= $W * (P^T * W)^{-1} * \hat{c}$

At Step 9, we have the prediction for the calibration data. Matrix B, is the regression coefficients matrix and can be used for predictions of “uknown” data the same way it is used here. This standard PLS is almost identical to NIPALS algorithm (another PLS implementation), despite PLS is recursive and not iterative as NIPALS, which means in terms of simplicity, standard PLS is simpler, because NIPALS needs determination of iterations number.

A great and simple computational approach on PLS regression is the Statistically Inspired Modification of Partial Least Square algorithm (SIMPLS) [13]. Commonly used in

chemometrics –which is the reason it was developed- SIMPLS is faster and has the same predictive efficiency as classic PLS algorithms. In detail:

$Initialize\ X\ (nxp), Y(nxm)$, A number of factors and a=1

$Y_0 = Y - MEAN(Y)$

$S = X^T * Y_0$ (cross - product)

1. $Let\ q = dominant\ eigenvector\ of\ S' * S\ (from\ Singular\ Value\ Decomposition)$
2. $Let\ r = S * q$ //Y block factor weights
3. $Let\ r = X * r$ //X block factor weights
4. $t = t - MEAN(t)$ //center scores
5. $t = t/norm(t)$ and $r = r/norm(r)$ //normalize scores and weights
6. $p = X^T * t$ //X block factor loadings
7. $q = {Y_0}^T * t$ //Y block factor loadings
8. $u = Y_0 * q$ //Y block factor scores
9. v=p //initialize orthogonal loadings
10. $if\ a > 1\ \ v = v - V * (V^T * p)$ and $u = u - T * (T^T * u)$ else go to Step 11 // make v previous loadings , make u previous t^T values
11. $v = v/norm(v)$ //normalize orthogonal loadings
12. $S = S - v * (v^T * S)$ //deflate S with respect to current loadings
13. $Form\ matrices\ R, T, P, Q, U, V\ from\ r, t, p, q, u, v\ respectively$
14. $If\ a = A\ , stop, otherwise\ a = a + 1\ and\ return\ to\ Step$ 1.
15. $B = R * Q^T$ and $\hat{y} = X * R * Q^T = X * B$ //Extract regression coefficients B and prediction $\hat{y}$.

This implementation of PLS algorithm offers advantages over the standard PLS and NIPALS. The factors are directly computed on the original matrices (mean-centered). The weights have a simpler interpretation than usual and above all, can be calculated without inverse matrix computations. The algorithm does not involve breakdown of X matrix, resulting to speed increase. It also maximizes the covariance criterion.

1.7 Cell and Staining

1.7.1 Cell structure

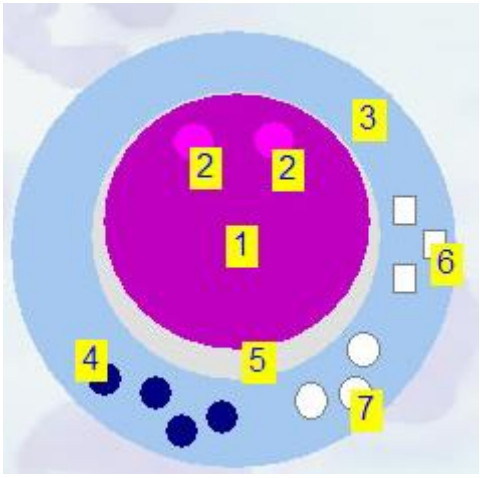


Figure 7. The cell's structure in general

The **cell** (from Latin word *cella*, meaning 'small room') is the basic structural, functional, and biological unit of all known living organisms. A cell is the smallest unit of life that can replicate independently. The things that we need to remember about a human blood cell's structure for better interpretation of this diploma thesis, are on the image, along with the next set of information (marks on the image shows the position of the cellular elements mentioned above) [14]:

1. Nucleus made up of chromatin, DNA, RNA.
2. Nucleoli made up of chromatin RNA.
3. Cytoplasm made up of acids (RNA mostly).
4. Specific granules (Azurophilic, Neutrophilic, Eosinophilic, Basophilic).
5. Perinuclear zone (Hyaloplasm) consisted of mitochondria mostly.
6. Mitochondria and Vacuoles.

1.7.2 What is staining and biochemical stains

Cells in microscope cannot be seen with bare eyes even from a great magnification lens unless they are properly stained. Staining is a technique that enhances contrast in the microscopic vision and makes the features of each kind of cell detectable. Stains and dyes (biomarkers) are used in biology and medicine, biochemistry and mechanics to highlight structures in biological tissues, often with the aid of different microscopes ,cameras and light sources. Stains may be used to define and examine bulk tissues (highlighting, for example, muscle fibers or connective tissue), cell populations (classifying different blood cells, for instance), or organelles within individual cells.

In biochemistry staining involves adding a class-specific (DNA, proteins, lipids, carbohydrates) dye to a substrate to qualify or quantify the presence of a specific compound (like our case). Staining can be used along with optical microscopy, electronical microscopy, hyperspectral microscopy, fluorescent tagging serving similar purposes. Biological staining is also used to mark cells in flow cytometry, and to flag proteins or nucleic acids in gel electrophoresis.

Simple staining is staining with only one stain/dye. There are various kinds of multiple staining, many of which are examples of counterstaining, differential staining, or both, including double staining and triple staining. Staining is not limited to biological materials, it can also be used to study the morphology of other materials like polymers [15] [16].

1.7.3 Staining of blood and bone marrow samples

Staining of blood or bone marrow samples is a routine technique for most of the doctors and researchers, especially on the medical or biological field. Although, in order to be diagnostic weapon and give reliable results, the staining process has to be done very carefully. There's only one exact way for this procedure, depending the case of illness and the sample's characteristics (hematocrit etch). After putting the sample on the tile in the right way, Romanowski technique is used for the staining.In this technique, May Grunwald - Giemsa (MG-G) stain coats the sample.

May Grunwald-Giemsa consists of two pigments. One acid (eosin), and one base (Methylene Blue and Azure B) [17] [18] [19] [20]. This combination gives the final solution the ability to stain the DNA, RNA and other features of the cell, so we are able to observe and qualitative evaluate the morphology of the cells of the sample. Quite significant for the success of this method of staining are [21] [22]:

- The exact PH of the solution needs to be 6.7 exactly. This can be defined with the use of regulatory solution.
- The pigments have to be well stirred.
- Careful exposure of the samples to the solution for the right time.

1.7.4 May Grunwald – Giemsa (MG-G)

As mentioned, May-Grünwald's eosin-methylene blue and Giemsa's azure eosin methylene blue are intended to be used for staining of blood and bone marrow smears in several research and diagnosis procedures. Finally, on the stained smear we have the three substances contained in the stain, coating the cells under different proportions. A typical example of MG-G stained cells is:

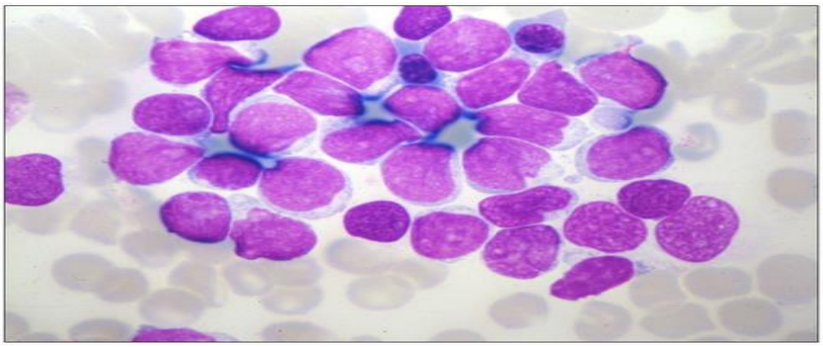


Figure 8. Typical MG-G staining

The purple color of the cell nuclei, is due to molecular interaction between Eosin Y and and Methylene Blue-Azure B-DNA complex. The nuclear acids of the nucleus, the primary grains of protoplasm and the basophil grains are negative charged so they attract the basic stains and constitute the basophil parts of the cell. Their final color after the staining ranges between several shades of blue.

The hemoglobin of the red blood cells, the secondary grains of protoplasm and the eosinophil grains are positive charged therefore they attract the eosin. Those are called eosinophil or acidophilic parts of the cell. Their final color after staining ranges between several shades of red. There also exists secondary grains that attract complex portions of both pigments of the solution, their final color rages between purple and red and are known as neutrophil grains [21] [22]. More information on the substances on the next section. For better visualization I present the next image, where we can see which part of the cell is coated by each substance.

Stain	On Nucleus	On Nucleoli	On Cytoplasm
Methylene Blue	Nucleic acids (DNA-RNA) Nucleic Proteins	RNA	Cytoplasmic Proteins Cytoplasmic RNA Basophil Granules Negative charged Neutrophil granules
Azure B	Nucleic acids (DNA-RNA) Nucleic Proteins	RNA	Cytoplasmic Proteins Cytoplasmic RNA Basophil Granules Negative-charged Neutrophil granules
Eosin Y	Chromatin(DNA) mostly		Reticulocytes Positive-charged hemoglobin parts Neutrophil granules Eosinophil granules

Figure 7. Stain substances-Cell Parts pairing

We need to remember that he final staining depends on the combination of the cellular elements (DNA, RNA, Proteins and others) as well as on the way that the sample is stained (1.7.2).

1.7.4.1 Eosin Y

The name ‘Eosin’ comes from Eos, the Ancient Greek word for 'dawn'. Eosin Y with chemical formula $C_{20}H_6Br_4Na_2O_5$ is a fluorescent acid negative charged compound that binds to and forms salts with basic, or eosinophilic, compounds containing positive charges (such as hemoglobin of the red blood cells , chromatin, proteins that are basic / positive due to the presence of amino acid residues such as Arginine and Lysine) and stains them dark red or pink as a result of the actions of bromine on fluorescein. Structures that attract and get stained with eosin are termed eosinophile [23] [24].

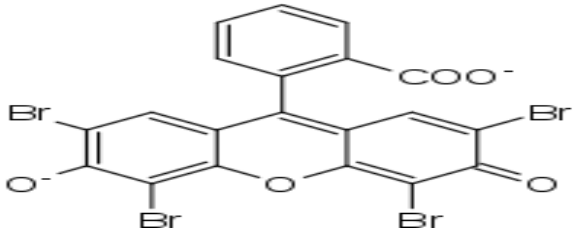


Figure 9. Eosin chemical composition

1.7.4.2 Methylene Blue

Methylene blue is a heterocyclic aromatic chemical compound with the molecular formula$C_{16}H_{18}N_3SCl$. It has many uses in biology and chemistry. At room temperature it appears as a solid, odorless, dark green powder, that yields a blue solution when dissolved in water. The hydrated form has 3 molecules of water per molecule of methylene blue. It is commonly used as a dye or stain (including to hematological samples) and is attracted from the basophil parts of the tissue or cell [23] [25].

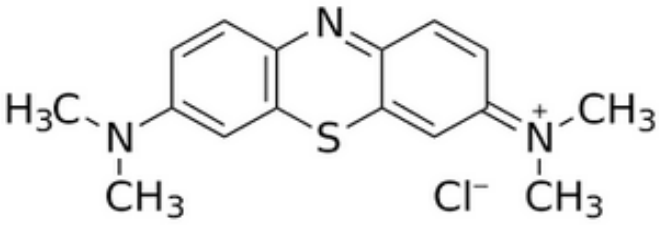


Figure 10. Methylene blue chemical composition

1.7.4.3 Azure B

Azure gets its name from the blue color between blue and cyan, like sky color on a clear day. It is a methylated thiazine dye, formed by the oxidation of Methylene blue, with chemical formula $C_{15}H_{16}ClN_3S$. Azure B is commonly used as a metachromatic basic dye ranging from green (to

chromosomes) and blue (to nucleoli and cytoplasmic ribosomes), to red color (to deposits containing mucopolysaccharides). At room temperature it appears as dark green crystalline powder [23] [26].

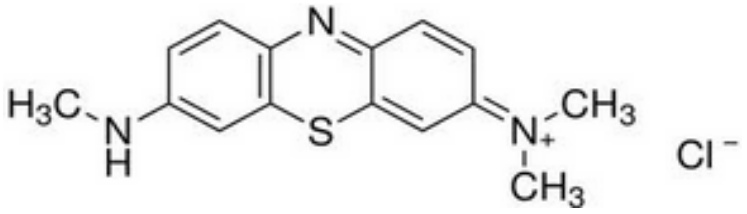


Figure 11. Azure B chemical composition

1.8 Blood – Bone marrow and leukemia disease

1.8.1 Blood

Blood is a constantly circulating fluid providing the body with nutrition ingredients, oxygen, and waste removal. Blood is mostly liquid, with numerous cells and proteins suspended in it. It is thicker than pure water. The average person has about 5 liters (more than a gallon) of blood. A liquid called plasma makes up about half of the content of blood. Plasma contains proteins that help blood to clot, transport substances through the blood, and perform other functions. Blood plasma also contains glucose and other dissolved nutrients.

About half of blood volume is composed of blood cells. Blood cells are produced during the process of hematopoiesis in the bone marrow which is the flexible tissue in the interior of bones. For humans and mammals generally, those cells are separated in three general types:

- Red blood cells, carry oxygen to the tissues
- White blood cells, have significant role to immune system
- Platelets, smaller cells that help blood to clot

Blood is conducted through blood vessels (arteries and veins). Blood is prevented from clotting in the blood vessels by their smoothness, and the finely tuned balance of clotting factors.

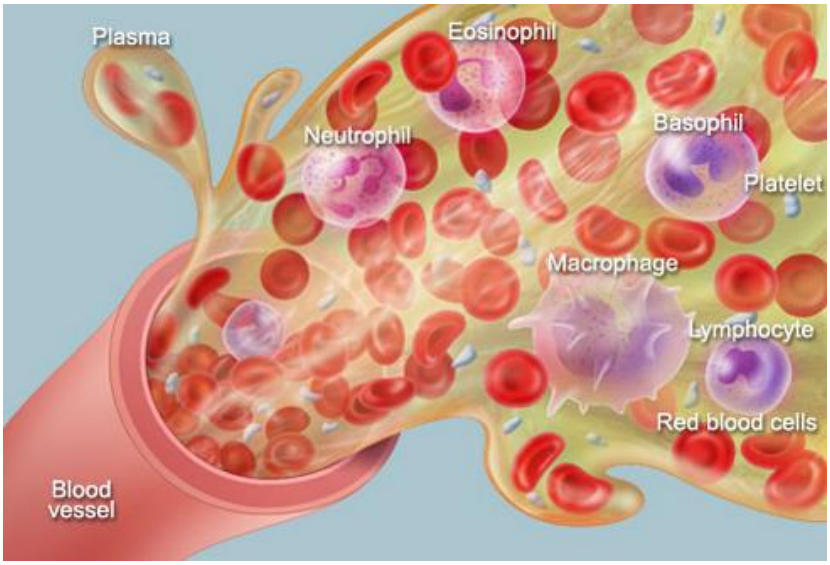


Figure 12. Blood Schematic

1.8.2 Bone Marrow and Hematopoiesis

Bone marrow is the flexible tissue in the interior center of bones. In humans, during the process of **Hematopoiesis**, red blood cells are produced by cores of bone marrow in the heads of long bones. Statistically, bone marrow constitutes 4% of the total body mass of humans. The hematopoietic components of bone marrow produce approximately 500 billion blood cells per day. These cells use the bone marrow vasculature as a gate to the body's systemic circulation. Bone marrow's role is significant also in lymphatic system, producing the lymphocytes that support the body's immune system.

The bone marrow has two types, the Red marrow (*medulla ossium rubra*), which consists mainly of hematopoietic tissue, and the Yellow marrow (*medulla ossium flava*), which is mainly made up of fat cells. Most of White blood cells, Red blood cells, platelets, arise in red marrow.

The stroma, is a tissue not directly involved in the marrow's primary function of hematopoiesis. Yellow bone marrow makes up the majority of bone marrow stroma, as smaller concentrations of stromal cells can be found in red bone marrow. Stroma is indirectly involved in hematopoiesis, since it provides the hematopoietic microenvironment that facilitates hematopoiesis by the parenchymal cells.

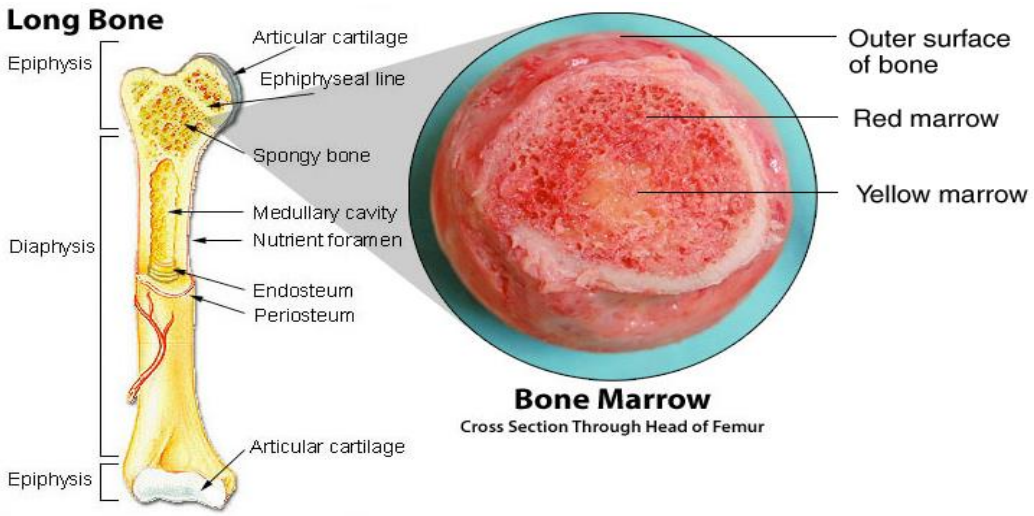


Figure 13. Bone marrow schematic

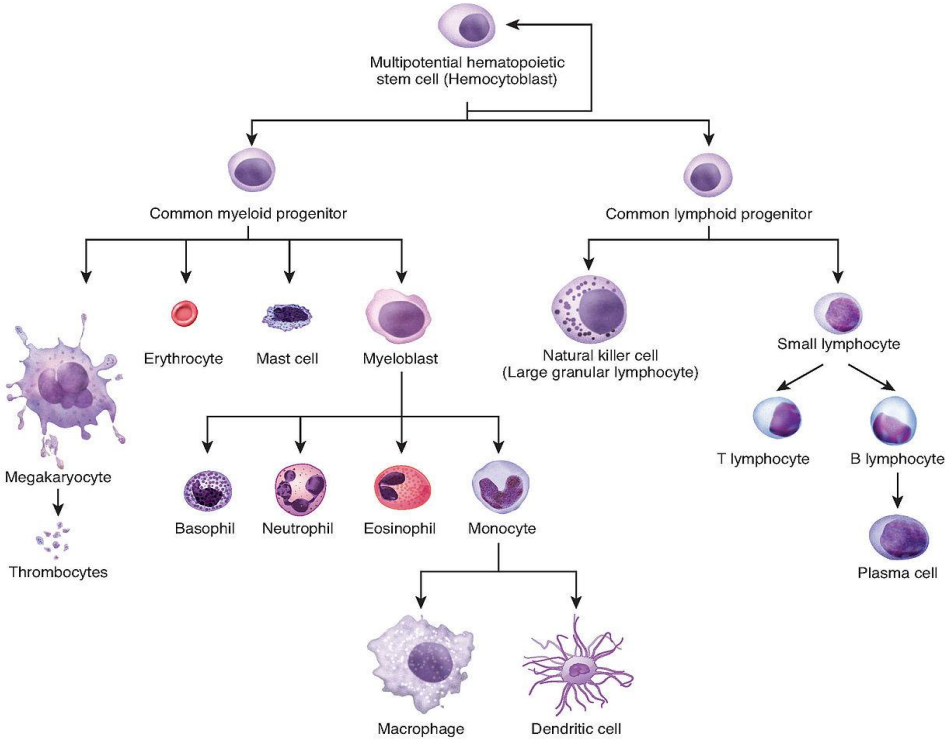


Figure 14. Hematopoiesis procedure

1.8.3 White Blood cells

1.8.3.1 General

White blood cells or leukocytes, are the cells of the immune system that are involved in the operations aimed to protect the body against infections and foreign invaders. They constitute an heterogeneous group of nucleated blood cells. Leukocytes are produced from a multipotent cell in the bone marrow known as a hematopoietic stem cell and are found mainly in the blood and lymphatic system.The total number of leukocytes in humans ranges from 3.5 to 10.5 x 109 / L.

1.8.3.2 Ordinary Types of white blood cells

White blood cells are classified into two major categories: The myeloid **leukocytes** and the **lymphocytes** (20-30%). The second category includes T cells, B cells and natural killer cells.
Five different types of myeloid leukocytes exist. These types are distinguished by their structure and functionality:

- Multinuclear neutrophils (50-70%)
- monocytes (2-6%)
- eosinophils (2-5%)
- basophils (0.2%)

Neutrophils, eosinophil and basophil granulocytes, monocytes, lymphocytes and plasma cells are normally present in the blood. All these kinds of cells are involved in processes of inflammation and fighting infections. **The myeloblasts, the promyelocytes, the myelocytes, the metamyelocytes and leukemic blasts are present in peripheral blood only in case of pathological conditions.**

Reduction on the number of leukocytes below the reference levels (3.5 to 10.5 x 109 / L) is known as leukopenia, while the increase is known as leukocytosis. The measurement of the total number of leukocytes is a parameter without much significance due to the heterogeneity of white cell types. Instead, the measurement of the different leukocyte subpopulations is considered very important and is known as a differential measurement (differential count). The results of differential measurement given in absolute numbers and in relative terms (percentage). Absolute values are diagnostically significant. Below we have an analytical presentation of each kind of leukocyte, their properties and role in many processes:

Lymphocytes

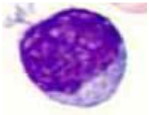


Figure 15. Lymphocyte coated with MG-G stain

Lymphocytes have a round to oval nucleus and size is similar to the size of red blood cells (erythrocytes). The cytoplasm painted gray-blue after staining. The quantity of the cytoplasm to

each lymphocyte can vary greatly. As a result, the cell size varies from 9 to 20 μm. The difference in volume of the cytoplasm is the cause of the process of separation of lymphocytes in small and large ones. This difference reflects the different lymphocyte activation phases. After staining blood Lymphocytes are classified in two main morphological types. The cells of one type are relatively small, usually have no granules and therefore exhibit a large nucleus to cytoplasm ratio (N:C).The cells of the other type, known as large granular lymphocytes (LGL), are larger, have a smaller ratio N:C and contain cytoplasmic, basophil granules. They constitute a percentage of 5-10% of lymphocytes. LGL morphological characteristics are presented in gamma/delta subset of T lymphocytes and natural killer cells.

The normal number of lymphocytes blood is between 1.0 and 3.5 x 109 / L. This number indicates that lymphocytes are the second most common type of cells in blood, after neutrophil granulocytes. The average adult human has about 10^12 lymphocytes. There is no morphological distinction between B and T lymphocytes.

Lymphocytes are mediators of cellular and humoral immunity. Lymphopenia is observed in cases where the number of lymphocytes is lower than usual, whereas lymphocytosis (lymphocytosis) in cases where the number of lymphocytes is increased. Lymphocytosis observed in lymphoproliferative syndromes, and lymphopenia in cases of infections (such as HIV, tuberculosis etch), after radiation treatment and during treatment with immunosuppressive drugs.

Plasma cells

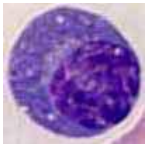


Figure 16. Plasma cell coated with MG-G stain

The plasma cells rarely occur in peripheral blood. A typical plasma cell is larger than a lemphocyte having a diameter from 15 to 20 μm. Eccentric-shaped nucleus with clear perinuclear region and basophil cytoplasm are the main characteristics of plasma sells. Their function is antibodies production. The normal level of plasma cells in peripheral blood is 0 - 0.1 x 109 / L. Their existence in the peripheral blood may be a result of vaccination. The increased number of plasma cells is known as plasmacytosis. Benign plasmatocytosis is observed in cases of viral infections. Release of plasma cells in the peripheral blood sometimes occurs in patients with advanced plasmacytoma (malignant plasma lymphoma).

Monocytes

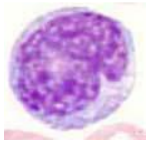


Figure 17. Monocyte coated with MG-G stain

With a diameter of 15-20 μm monocytes is the largest group of peripheral blood cells. Their form is heterogeneous. They form pseudopodia in the outer membrane. The cytoplasm is blue - gray, and very often exhibit basophil granules and organelles. The nucleus can be a horseshoe-formed or present lobes . The normal monocytes levels range from 0.2 - 1.0 x 10^9 / L. Increased number of monocytes is known as monocytosis, and dicreased as monocytopenia.

Monocytes have the very special ability of immigration. When migrating to the tissues they are called macrophages. Monocytes play an important role in acute and chronic infections. They are important components of cell- mediated immunity. The monocytosis associated with various chronic infections (such as tuberculosis , typhoid fever) and malignant diseases (such as Hodgkins). Monocytosis is observed in acute and chronic myelomonocytic leukemia. Monocytopenia occurs in bone marrow aplasia, in hairy cell leukemia after therapy treatment with steroids.

Basophils

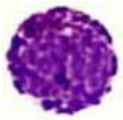


Figure 18. Basophil cell coated with MG-G stain

The basophil granulocyte's diameter ranges from 10 to 14 μm and is smaller than the other granulocytes. The dark purple granules is very densely packed and overlap the nucleus and cytoplasm. The normal level of basophils ranges from 0 to 0.15 x 109 / L. They are rarely encountered in blood smears. Basophils are important in hypersensitivity reactions. They can leave the bloodstream and migrate to surrounding tissues. Increased number of basophils is characterized as basophilia. Basophilia is observed in cases of chronic myelocytic leukemia and other myeloproliferative syndromes.

Eosinophils

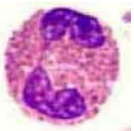
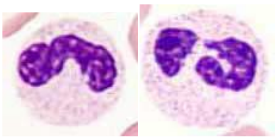


Figure 19. Eosinophil cell coated with MG-G stain

The eosinophil granulocytes have diameter about 16 μm, are circular and very little larger than neutrophilic granulocytes. The granules are light violet to orange-yellow and very densely packed. The nucleus is usually biloba. Eosinophils are rarely found in the blood smear, and their normal levels vary from 0.02 to 0.4 x 10^9 / L. The eosinophil granulocytes have an important role in allergy and parasitic infections. Like neutrophils, they are able to phagocytose and migrate.

The increased number eosinophils is characterized as eosinophilia while reduced as eosinopenia. Eosinophilia mainly caused by allergy and parasitic diseases. Many pharmacological treatments can also cause eosinophilia. Some neoplastic diseases (egHodgkins) occasionally cause eosinophilia. The cases where eosinophilia reason is not detectable, are referred as idiopathic eosinophilia syndromes. Eosinophilia is often detected in cases of stress and acute infections.

Neutrophils



Band neutrophils Segmented Neutrophils

Figure 20. Types of neutrophils coated with MG-G stain

The neutrophilic granulocytes are typically circular and have a light gray cytoplasm. Their diameter is about 14 μm. The granules are stained red-violet to brown. Neutrophils subdivision is based on the structure of their nucleus: band neutrophils and segmented neutrophils. The band neutrophils, are less developed and more immature than the segmented type. Neutrophils with polylobe nucleuses, are separated in more than 4 kinds, and are referred to as hyper-segmented. Their normal level range from 1.6 to 7.4 x 10^9 / L. Lack of neutrophils is called neutropenia (neutropenia), and on the other side, increased number is called neutrophilia (neutrophilia).

One of the main functions of neutrophils is to protect against bacterial infections, phagocytosing and destroying pathogens. Neutrophils can leave the bloodstream and migrate to surrounding tissues, to fight infections. Neutrophilia has several causes. The mobilization of adherent neutrophils is typical stress characteristic (stress leukocytosis). Acute infections and inflammations may result to mobilization of neutrophils from the bone marrow. Neutropenia is associated with a pharmacological treatment of infections (e.g., parvoviruses, malaria) and autoimmune diseases (e.g. systemic lupus erythematosus). The presence of increased hyper-segmented neutrophil count is usually evidence of lack of evidence of vitamin B12 or folic acid (megaloblastic anemia).

1.8.4 Leukemia Disease

1.8.4.1 General

Leukemia is cancer of the blood cells. It’s the situation in which bone marrow fails to produce healthy leukocytes. Cancerous leukocytes produced are characterized by structure abnormalities compared to the

healthy ones as a result of their incomplete development. Those leukocytes grow faster than normal cells, they don't stop growing when they should and are not capable of doing what leukocytes are aimed to do for the human organism crowding out the healthy blood cells in the bone marrow. This may lead to serious health problems like anemia, bleeding, bruising and infections [27].

1.8.4.2 Ordinary types of leukemia

Type of leukemia depends on the type of white blood cell that has become cancerous. There are four main types of leukemia [21] [27]:

- Acute lymphoblastic leukemia, or ALL.
- Acute myelogenous leukemia, or AML.
- Chronic lymphocytic leukemia, or CLL.
- Chronic myelogenous leukemia, or CML.

There are also several less common types of leukemia, such as hairy cell leukemia and others (most of them mentioned in chapter 1.8.3).

1.8.4.3 Typical methods for leukemia diagnosis

Modern diagnosis is based on **repeated complete blood counts and a bone marrow examinations** following **observations of the symptoms**. It is common, that blood tests are unable to show with certainty that a person has leukemia, mostly in cases that the disease is in early stages or remission. The type of leukemia is quite difficult to be determined and needs the doctor's experience and full attention. Also, a lymph biopsy can be performed to diagnose certain types of leukemia in certain pathological situations.

In order to proceed to a diagnosis, the doctor, using his microscope, has to find the field of interest on the dyed tile of blood or bone marrow. After that, he changes the lens in order to have a clearer view and he does the same procedure with an even bigger lens, in order **to separate the kinds of leukocytes on the sample with bare eyes**. This whole procedure needs to be done several times (on many coordinates on the sample) in order to result to a reliable diagnosis. Most of the times this procedure is not enough and the doctors perform **flow cytometry (FACS)** in order to establish a more accurate diagnosis. Even if great doctors and qualified tests exist, many times this determination fails, and wrong treatments are followed resulting to deterioration of the symptoms and deaths [21] [22].

On cases that the disease is in advanced stage**, blood chemistry tests** can be used to determine the degree of liver and kidney damage or the effects of chemotherapy treatment. **X-ray, MRI, or ultrasound are used when we want to check the damages that leukemia has caused on bones, brain, kidneys, spleen and liver** [21] [22].

As we can understand leukemia is a severe disease, with most of the times non-specific symptoms that easily can refer to other diseases and has tricky way of diagnosis. It is in each doctor's hands and experience whether he will achieve to determine the type of leukemia and propose the right treatment. On the next paragraph we have a demonstration of the typical method that a doctor follows to examine a blood tile.

1.8.4.4 Typical examination method of blood-bone marrow stained tile

After sample staining researchers and doctors move to the process of evaluation. During this process:

- The first magnification lens of the microscope (10x) is selected and moved to the proper observation place on the tile. Using this we can calculate qualitatively the number of leukocytes, platelets, and the quality of the cells that possess nucleus. If the number of red blood cells is available (from automated machine analysis), the number of platelets is possible to be calculated.
- Then we change the lens to a 40x magnification (or greater), so we can see more clearly the types of cells that we have. This can lead the doctor to specified diagnosis.
- Using a 100x magnification diving lens, we calculate the platelets number in field of vision, with approximately 100 to 150 red blood cells. We repeat the last step at least ten times and statistically we have the number of platelets to the number of red cells. In the same way we can estimate the number of leukocytes.
- Lastly, we scan the whole sample for any kind of different cell type, as red blood cells with nucleus, grains of many types, bigger than usual platelets, red grains, megakaryocyte nucleuses or rare lymphatic cells with pathological morphology, like hairy cells. This scan is done with both low magnification lens and more than 40x. This way scientists can define the type of white blood cells and the qualitative changes on cell series (red cell series, white cell series, platelets series) [21] [22].

On the next image we can see the pattern that is followed on a tile when the abovementioned process takes place repeatedly:

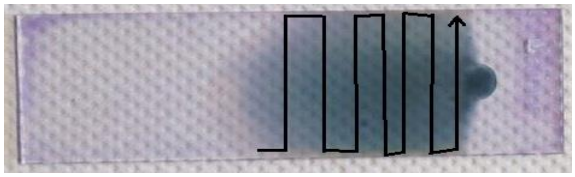


Figure 21. Tile repeated examination pattern

2. SYSTEM DESIGN

During the experimental procedure, we used:

- A Zeiss- Axio Scope.A1microscope
- A XIMEA xiQ USB 3.0 CMOS sensor camera
- A Point Grey- Flea3 USB 3.0 CMOS sensor camera
- A custom tunable light source (LED)
- Blood and bone marrow MG-G stained samples

All these were operated separately and in concurrency using a personal computer, with custom made GUIs (Graphic User Interfaces) in MATLAB. Using those GUIs we are able to handle the TLS and the camera, at the same time, in order to acquire the hyperspectral cube automatically or for other purposes, with settings provided only once to the system. The efficacy of each device was established before starting the experimental procedure.

2.1 Microscope

Image 22. The Zeiss Axio Scope A1

The microscope used for collecting the data, is a Zeiss, model Axio Scope.A1. This microscope due to it's the wide range of interchangeable components made the experimental procedure easier and more efficient. Thus, we replaced the LED light source, with a custom Tunable Light Source (TLS), made in our laboratory (see 2.3, TLS). The microscope is fully

functional on three different lenses, 5x, 30x and 40x [28]. For our experiment, we started the tiles observation, using the 5x lens, then, finding a wide area of interest, we zoomed to 30x, and finally when ended up on a specific area of interest and moved on the 40x lens, in order to result to an image, in which the characteristics of each cell are overemphasized. As we will discuss later, this method is also used by doctors on many areas of the tile, in order to conclude to a diagnosis.

2.2 Hyperspectral Cameras

First of all we have to make clear that we used no more than one camera at a time. It is important because the cost of a two camera system increases dramatically. The two cameras we used during our experiment are:

XIMEA xiQ USB 3.0

XimeaxiQ USB 3.0 is a CMOS, high-speed, low power consumption, color camera, proper for experimental hyperspectral imaging. Using the usb3 bandwidth capability it provides us 1280 x 1024 resolution color images at a maximum of60 FPS [29].

Figure 23 .Ximea xi-Q USB3

Point Grey Flea3 USB 3.0

Point Grey Flea3 USB 3.0 is a high-speed, low power consumption, color camera, proper for hyperspectral imaging. Using the usb3 bandwidth capability it provides us 4096 x 2160 resolution color images at 21 FPS (185 MByte/s).

Figure 24. Pointgrey Flea3

The flea3 features Sony's new IMX121 sensor with "Exmor R" back-illuminated CMOS architecture. This CMOS architecture improves sensitivity and dynamic range for sharp by increasing quantum efficiency and reducing noise, delivering high-quality color images.After testing the system with the XIMEA XiQ usb3 camera, we needed to observe the results more detailed and in higher resolution.This is why we used the Flea 3 on 4096x2180 pixels resolution (Full HD) [30].

2.3 Tunable light source (TLS)

2.3.1 General

The light source used for our experiments is a custom tunable led light source, made in our laboratory by Christos Rossos during the elaboration of his Diploma Thesis [31]. This light source is capable of providing us electromagnetic radiation which extends from UV to near infrared, using a number of narrow band emission LEDs and propagating their light through optical fibers, one for each one. The LEDS emission peaks are: 390nm, 420 nm, 450 nm, 465 nm, 505 nm, 530 nm, 600 nm, 640 nm, 666 nm, 685nm, 720nm, 735nm, 770nm, 810nm, 940nm plus one White LED for RGB view. LEDs are connected to an Arduino for easier handling. Other circuits (microcontrollers) are also used for current stabilization, handling etch.

The tunable LED source is energy efficient and has excellent throughput (85%), great stability, low cost and high tunability, as well as infinite emission choices, no warm-up period and great potential of totally directional light, facts that make it a great scientific instrument for experimental imaging microscopy. As aforementioned, electromagnetic radiation of this source extends from 380 nm (Ultra Violet), to 980 nm (near infrared) with spectral profiles shown in the diagram below (The LED of white light and the 940nm LED are missing from the diagram):

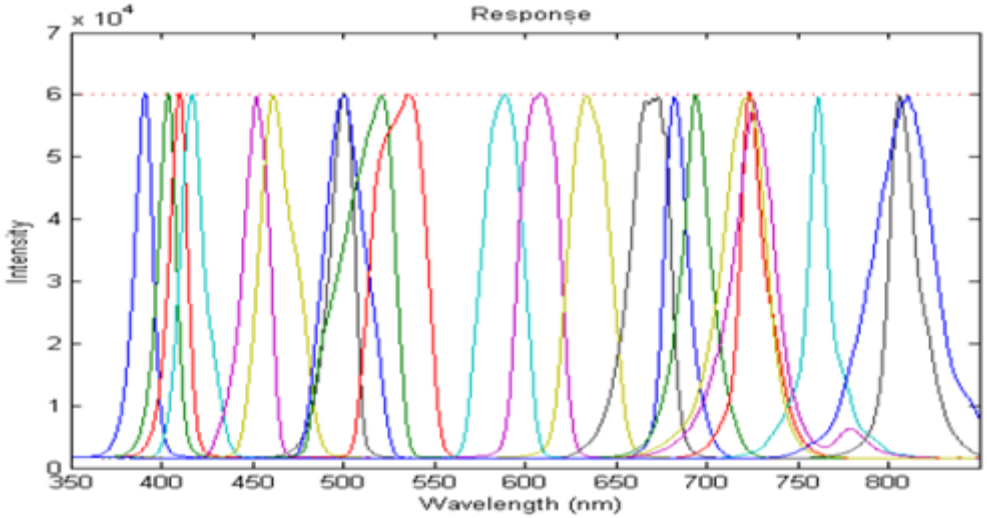


Figure 25. TLS LEDs spectral profiles

2.3.2 TLS efficiency validation

We firstly operated the light source separately, to make sure that it works properly in every single band. A spectrometer is an instrument used to measure properties of electromagnetic radiation over the spectrum. Usually the measured variable is radiation's intensity per wavelength. With the spectrometer, we can produce spectral lines and measure properties for radiation extending to a great range from Gama Rays to Infrared. The spectrometer we used in our experiments is the Ocean Optics USB 2000+ [32]:

Figure 26. Spectrometer Ocean Optics USB 2000+

For each of the 16 LEDs we measured the Bandwidth, the FWHM value (Full Width Half Max), the spectral line, the intensity variation for some time after the opening of the LED. All these measurements were taken in order to make sure that the conditions are proper for the settings that needed to be done for our sensitive experiment.

2.3.3 System calibration for spectral cube acquisition

For the calibration of a system like the one described we have two choices. First one:

- LEDs were calibrated on flat response.
- For each LED, the shutter of the camera was measured in order to achieve flat response on the intensity that the CMOS captures. Setting the microscope on an empty side of the tile, and acquire a spectral cube. On each band we stop and adjust camera's shutter, keeping gain as close to zero as possible.

By the end of this procedure, we have a calibrated spectral imaging system. The second way is:

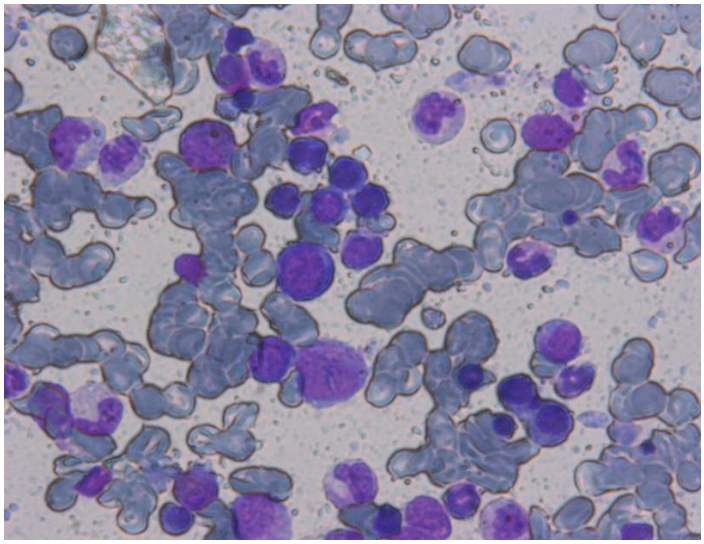
- Setting the microscope on an empty side of the tile.
- Move to the less powerful LED, set it on almost max power and adjust the shutter to a value that results to the wanted intensity.
- After that, keeping the shutter stable, we move to every LED and adjust the power until the system is flattened.

Loading those data on our system from a look-up table, we can achieve flattened spectral cube acquisition, without interruptions and delays.

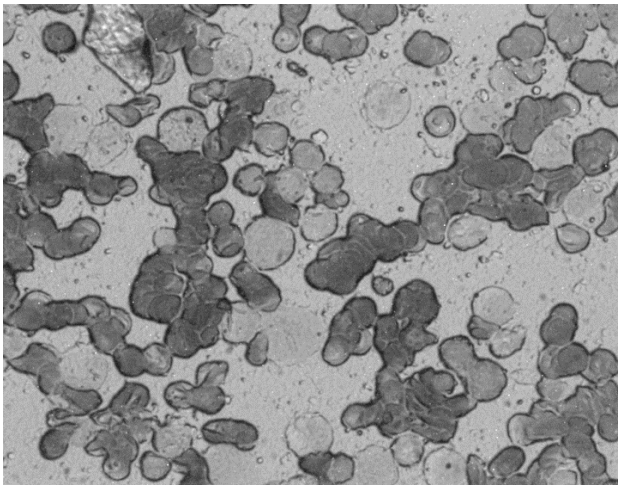
3. METHODS AND SYSTEM VALIDATION

3.1 Spectral Cube acquisition and observation

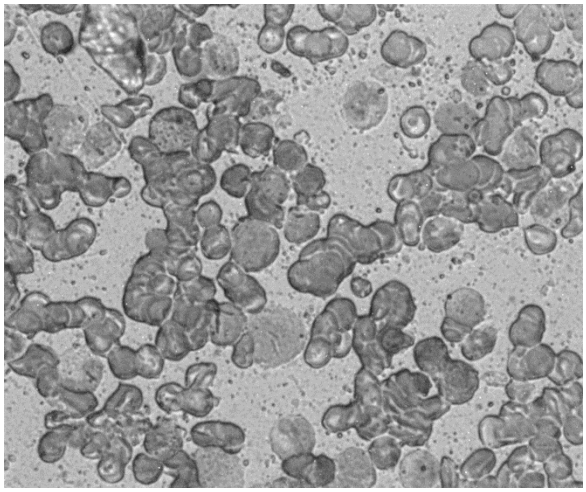
After building and calibrating the system, during and after the process of cubes acquisition we had to observe every detail of the images in every single wavelength in order to understand the differences of the collected spectral images. Before trying to explain the results that we observe on the cube, we have to demonstrate one. On the next pages the acquired spectral cubes are to be analyzed image by image from 390nm to 940 nm, and the spectral properties of the stained bone marrow and blood samples will be revealed. A characteristic spectral cube example is the one above. Notice the changes on the intensity of the single wavelength images as we move from 420nm to near Infrared:



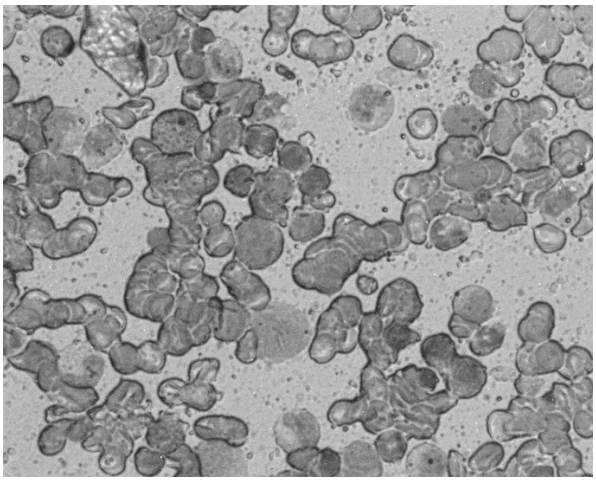
Color Image



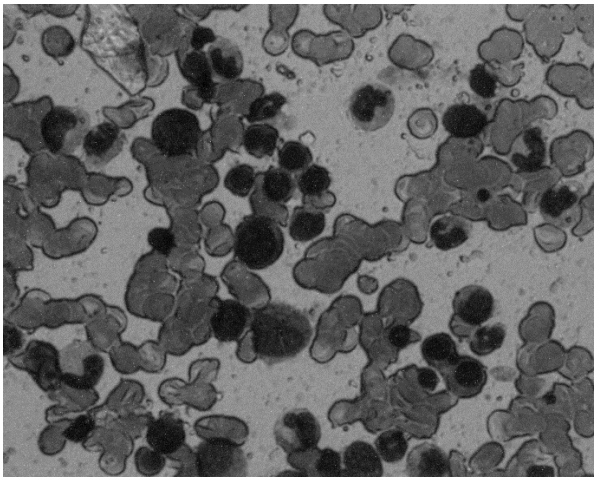
420nm



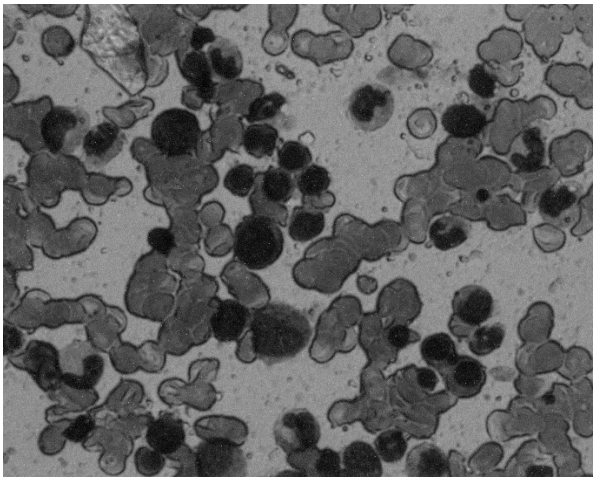
450nm



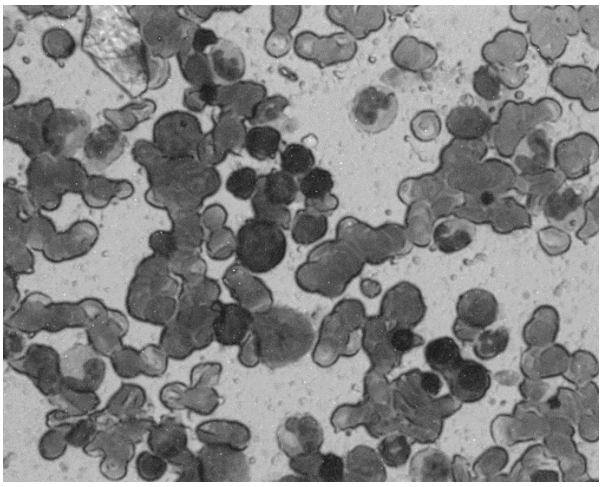
465nm



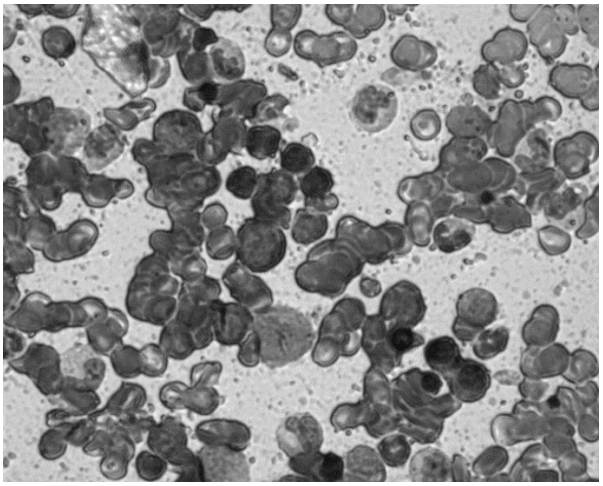
505nm



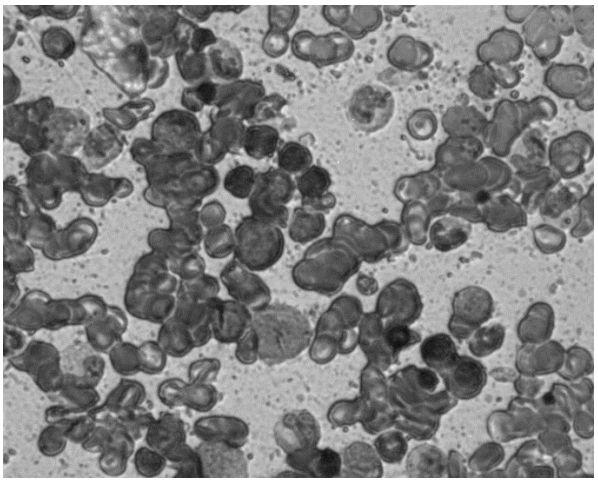
530nm



600nm



640nm



666nm

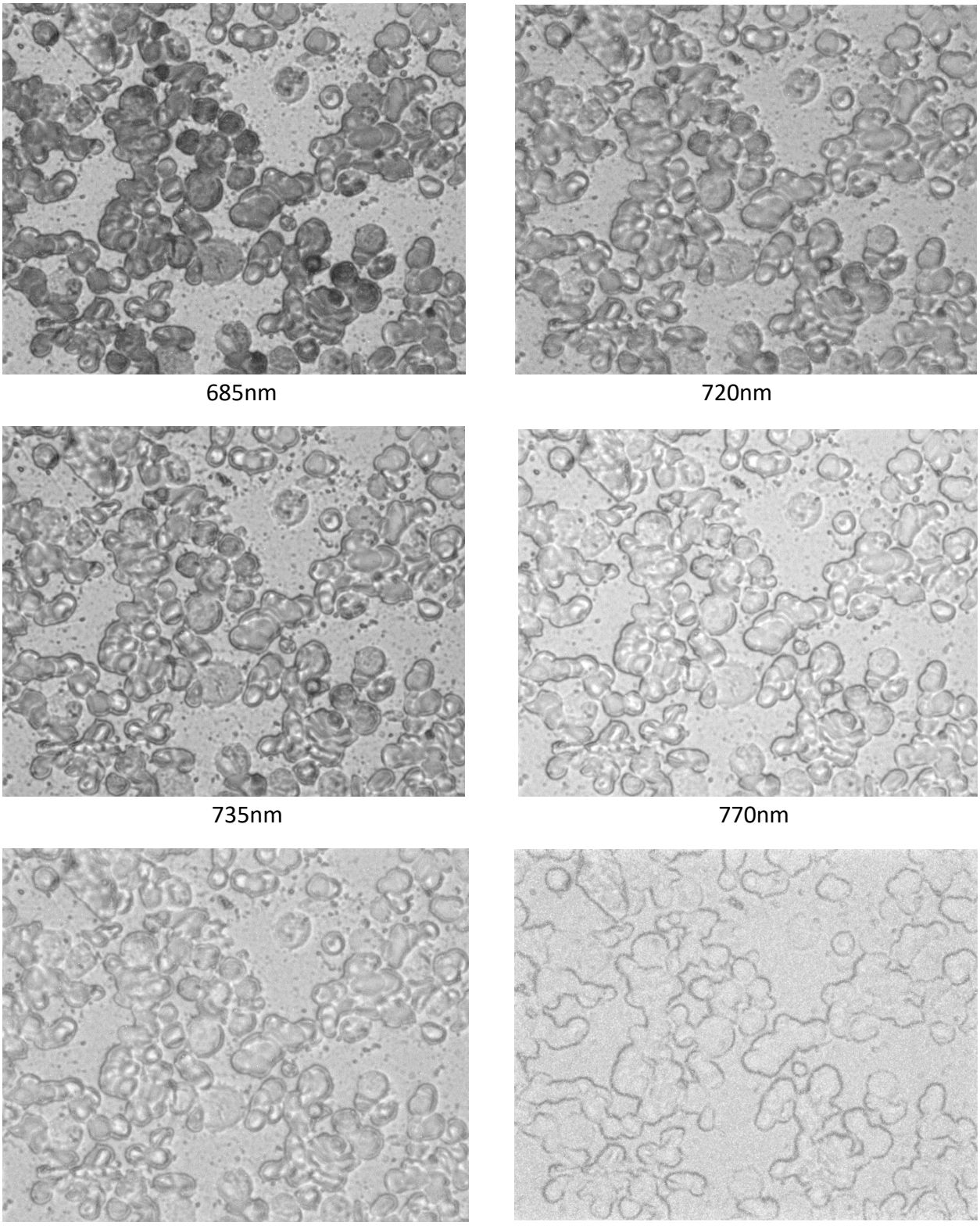


Figure 27. Full Spectral Cube captured by our system

This is a proper example of a full spectral cube because the examined bone marrow area contains many kinds of cells, from Red blood cells, to White blasts. It is easy for us to observe the changes generally, separately on each kind of cell and also some difficulties that will be explained. The separation of these cells is no matter for us now, as the purpose of this thesis is more general.

Observations on the spectral Cube

420 nm

On 420 nm we can clearly observe that the white blood cells (purple-blue on Color Image) are completely invisible, taking the same grayscale level as the background. Same thing with the separated red cells. It worth mentioning that the red cells are sometimes stacked or attached to each other like a stack of coins (Rouleaux). Light transmittance at that point cannot be reliable, as the light-matter interaction changes. This is why we have less transparent points on that band.

450 nm

On 450 nm we observe the white blood cells becoming a little darker than the background. Red blood cells remain dark as on the previous band. We can now clearly observe the first nucleus features appear on the white cells. This is because chromatin, which is stained mostly by eosin Y, starts becoming visible on that wavelength.

465 nm

On 465nm the observed difference between the background and the white cells remains the same, with very little change. Chromatin remains typically same, but with a second observation, we can observe an almost linear increasing relationship of the absorbance with the previous band.

505 nm

The 505 nm is a key band for our experiment. We can clearly observe that the white cells nucleus become dark and formed. The red cells are darkened too, as eosin Y works as a counterstain, but the real change is on the nucleus of the white cells. Eosin Y, which mostly stains the nucleus chromatin has its peak absorbance very close to that spectral point, fact that is now experimentally proven. The chromatin (whole nucleus) appears dark and spotless, and so very easy to observe. The cytoplasm is getting darker but still maintains its great contrast with both nucleus and background.

530 nm

On 530nm we have similar observations to 505nm. Some of the nucleuses appear partially darker, showing particularities on their structure. This is because now we can observe the chromatin stained by eosin Y, but also nucleic acids stained by Azure B and Methylene Blue. The three absorbances are additive, giving a partially very dark result on the nucleus of the cells. Cytoplasm also starts to have certain particularities, concerning the type of the cell. Darkened spots on it reveals basophilic character.

600 nm

Eosin Y has now zero absorbance. As it is easy to understand, now we can observe only the Azure and Methylene Blue absorbance characteristics on the cells. Particularities observed on the previous band are now much clearer as some kinds of cells become lighter than others that maintain their dark features. Nucleic acids concentration is the key to understand those particularities, as both substances stain them. Their existence means greater absorbance, as the staining substance's concentration increases, while their absence means the opposite. We still have a relative similarity on the remarkable behavior of the white cells on the spectrum.

640 nm-666 nm

Here some parts of certain white cells become lighter and others darker, making the differences between the cells more observable. Darker spots on some cell's nucleus or cytoplasm become visible this way. Same here, nucleic and nucleoli acids are the cause of these significant particularities. Still the absorbance of the basic dyes is increased. It is very difficult though, to explain which one of the two basic dyes is the main cause for those great absorbance spots. On these wavelengths their characteristics are similar, making the decomposition process really difficult.

685 nm

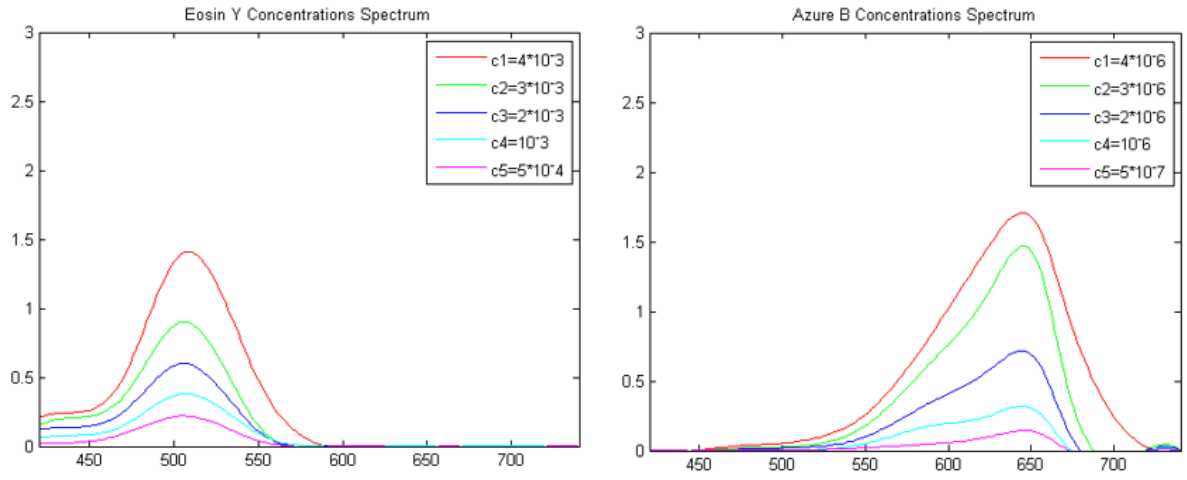
This is a very interesting band, as we can observe that only some parts of the white cells remain dark and overemphasized which means that the absorbance features on those parts are different. As we will analyze on the next chapter, it is due to the excessive concentration-presence of only one of the two basic staining substances. The other substance loses its absorbance characteristics on previous part of the spectrum, making the cells lighter in total, but some parts remain dark as the other substance maintains its absorbance characteristics on this part of the spectrum.

720 nm - 735 nm – 770nm – 810nm – 940 nm

Details fading away as we move to the near Infrared part of the spectrum. The more we increase the wavelength, the more transparent our cells become. As we can understand and will analyze, our staining substances have no absorbance characteristics on near IR and that's the reason for the excessive transparency.

3.2 Staining substances Absorbance Spectra

On this part of the Diploma thesis we will try to give the facts that explain the spectral cube that we demonstrated in section 3.1. As mentioned in section 1.7.3, the stain combination used for our samples is the May Grunwald-Giemsa, which consists of three substances, Eosin Y, Azure B, Methylene Blue. It is easy to understand that each individual components absorbance spectrum, changes (mainly increase or decrease) while the concentration changes. On the next image the three individual components absorbance spectra for Eosin Y [33] [23], Methylene Blue [10], Azure B [34] [23], in different concentrations (Molarity) are presented:



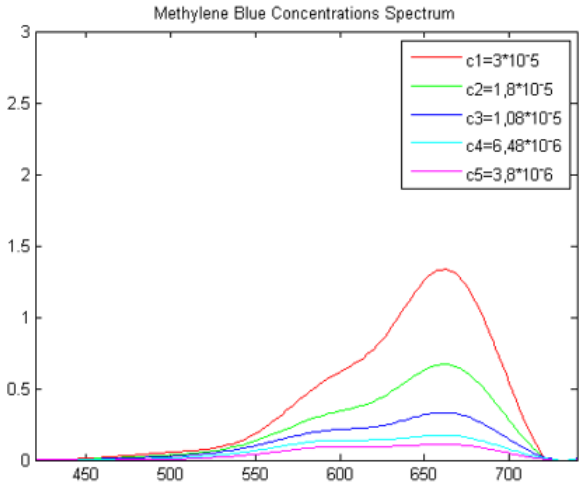
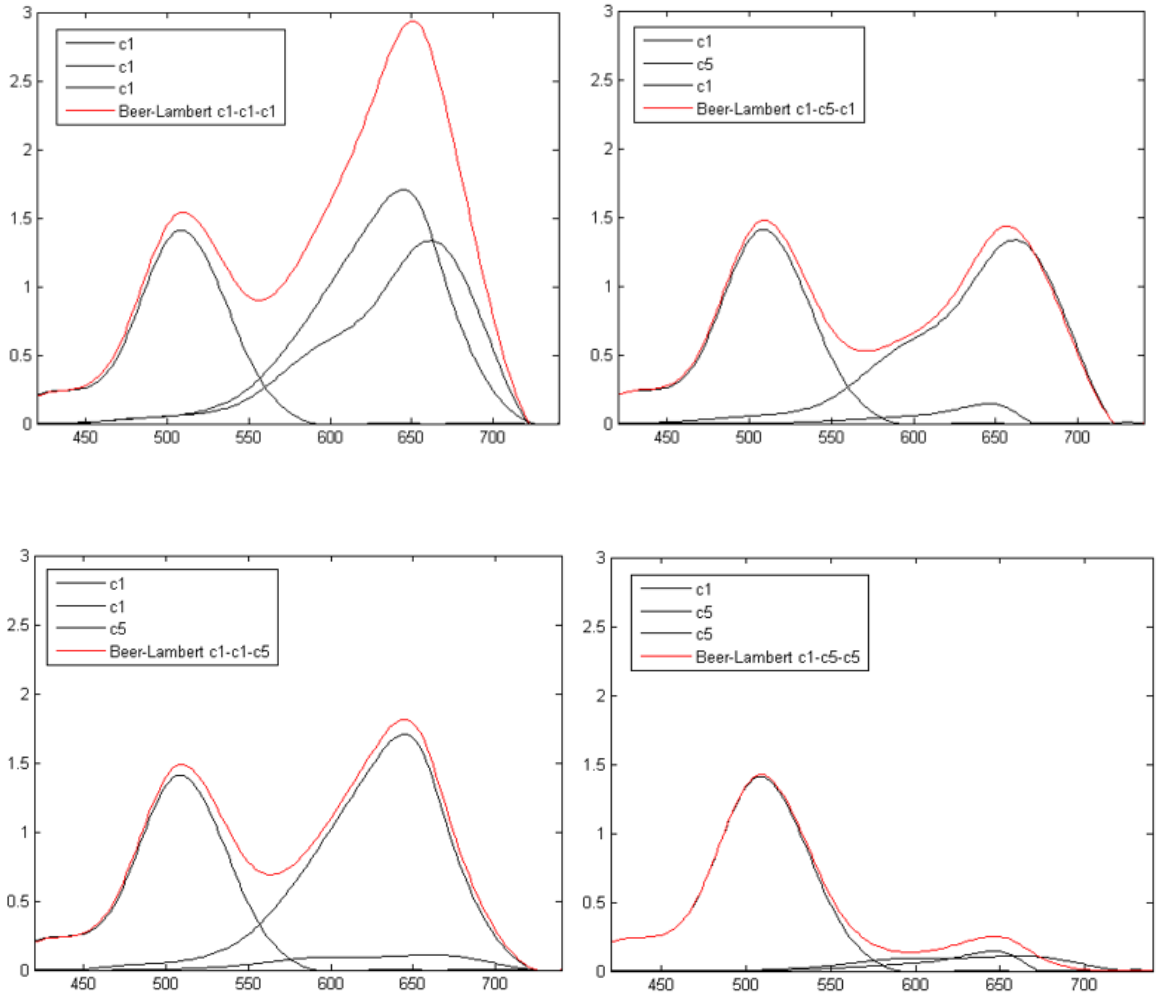


Figure 28. Individual Components Absorbance Spectra



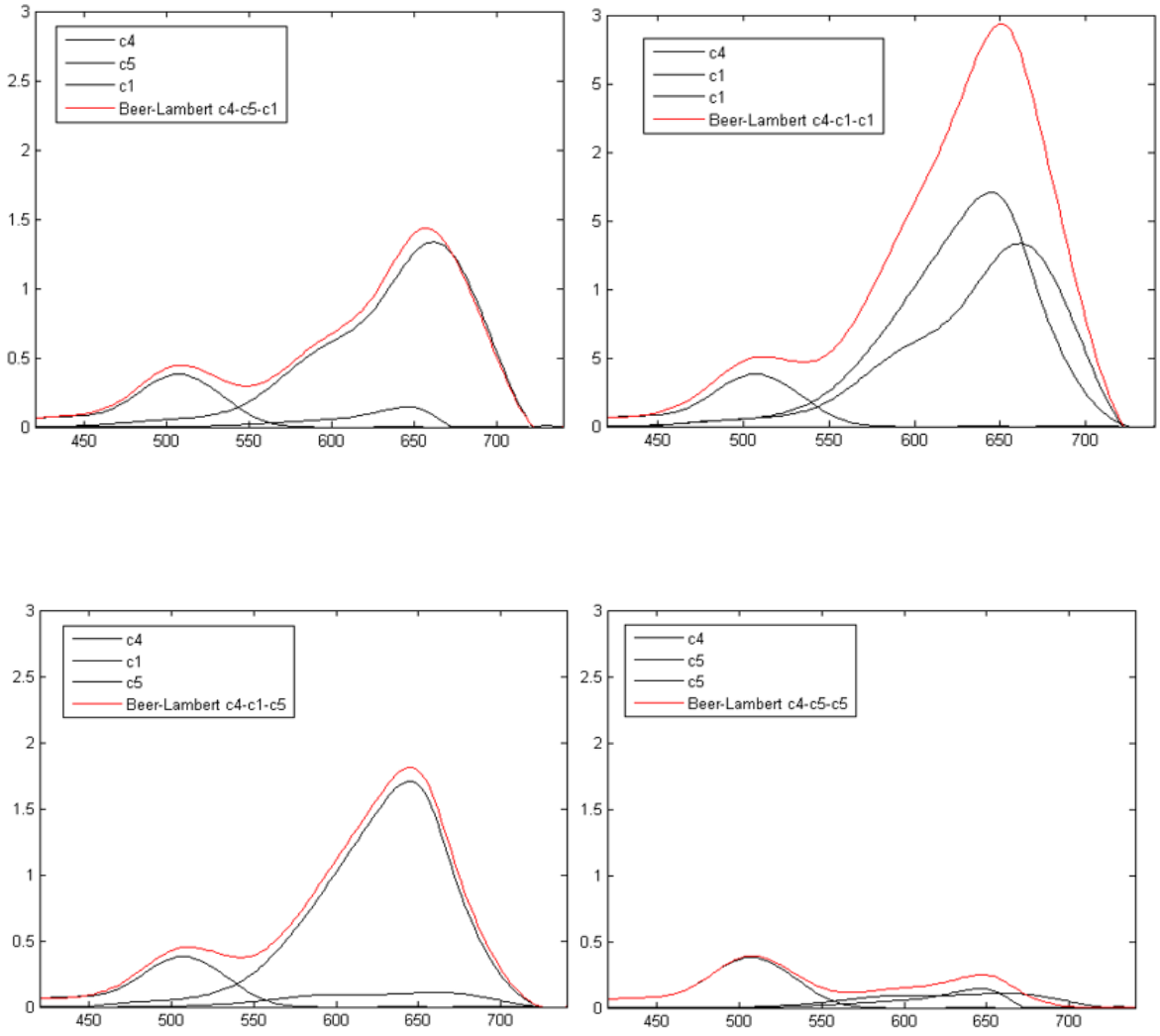


Figure 29. MG-G Mixture Absorbance Spectra

Image 26 shows the individual components spectra. We demonstrate the absorbance spectra of each substance on different concentrations. Using the generalization of Beer-Lambert Law, as explained in section 1.6.3, we can calculate the absorbance spectrum of the mixture of the three substances (for all concentrations combination), using their individual spectra on the wanted concentration (Image 27). For example, on the last part of Image 27, we calculate the mixture spectra (red line) as: Mixture Absorbance(w) = Eosin Y absorbance in C4(w) + Azure B absorbance in C5(w)+Methylene Blue absorbance in C5(w), where w is the wavelength.

It is of great importance to note again that this procedure cannot be applied on mixtures that their substances react in any way, among each other or with the specimen because in this case, the generalization of Beer Lambert Law is not valid. The physical properties of the new substance-s -results of the reaction, may be completely different than the reactants. In our case the three substances of MG-G do not underlie to any kind of chemical bond that could lead to reaction.

Choice of Calibration set for our experiment

In 2014, Fani Abatzi studied on her diploma thesis, concentration prediction and deconvolution efficiency for several algorithms, on many experimental designs and calibration sets [10]. The most efficient one, on Spectral Imaging terms, was the Partial Least Square algorithm. The aforementioned set of calibration is the proposed as the least possible needed, for PLS to give fine results. The experimental design in general terms for a three substances mixture is the following, adding the pure spectra:

Substance 1	Substance 2	Substance 3
C1	C1	C1
C1	C5	C1
C1	C1	C5
C1	C5	C5
C5	C5	C1
C5	C1	C1
C5	C1	C5
C5	C5	C5

Where C1, C2, C3, C4, C5 are the concentrations presented on the dataset above (Image 26-27). The 23 absorbance spectra samples used in this calibration set, were measured from 400 to 700 nm, with 2 nm step. As it is clear, the wavelengths were too many. On the implementation part we will prove that the use of specific selected wavelengths, can lead to the same results, using the same calibration samples.

3.3 Error Estimation

Error estimation and performance evaluation of calibration for regression models that estimate concentrations is associated with two kinds of errors. The first kind and most proper one for our research on this thesis is Root Mean Square Errors (RMSEs). The second is the Standard Error of Regression, which cannot be calculated for PLS or PCR. So, we are going to use RMSEs.

The RMSEs are going to be for us, the metric which will give us a clue about **how close to the initial concentrations are the estimated concentrations** and on the next phase of band reduction, **how close to our initial estimations (full cube) are those estimated by the algorithm calibrated with the reduced number of bands**. Also, we will have a proof that **the concentrations predicted on unknown samples, is valid and close to reality**. RMSEs can be absolute values or be expressed in percentages. The percentages are easier to interpret, because they indicate in % terms, how close we are to the wanted results. **A high percentage of RMSE indicates that we are far from our wanted results, while the lower it becomes, the closer we are to them.**

The first RMSE applied on our system is the Root Mean Square Error of Prediction (RMSEP), that's calculates the RMSE between the Predicted ($Cpred$) and the real value of concentration (C):

$$RMSEP = \sqrt{\frac{\sum_1^n (C - Cpred)^2}{n}} \ ,$$

Where n are the samples in the calibration model. The Relative % Root Mean Square Error of Prediction is (RMSEP%):

$$RMSEP\% = \sqrt{\frac{\sum_1^n \left(\frac{C - Cpred}{C}\right)^2}{n}} * 100\%$$

The second MSE applied on our system is the Root Mean Square Error of Calibration (RMSEC), describes the degree of agreement between the calibration model estimated concentration values for the calibration samples and the accepted true values for the calibration samples.

$$RMSEC = \sqrt{\frac{\sum_1^n (C - Cpred)^2}{DOF}} \ ,$$

Where DOF, is the number of Degrees Of Freedom. The Relative % Root Mean Square Error of Calibration is (RMSEC%):

$$RMSEC\% = \sqrt{\frac{\sum_1^n \left(\frac{C - Cpred}{C}\right)^2}{DOF}} * 100\%$$

In statistic terms, the number of Degrees of Freedom (DOF) is the number of values inthe final calculation of a statistic that are free to vary. More specifically, in chemometricsthe degrees of freedom are the number of data, minus the number of parameters calculated from them. For example, in multivariate regression with p independent variables, the standard error has n-p-1 degrees of freedom. This happens because the degrees of freedom are reduced from n by p+1 numerical constants b_0 ; b_1 ; b2 ; : : : ; b_p, that have been estimated from the sample [39]. This happens when a non-zero intercept exist in the equation or the data are mean-centered, thus b_0 exists. Otherwise the degrees of freedom are n - p.

More specifically, when referring to RMSEC, the right number of dof for PLS (approximately), is the number of samples, n, minus the number of factors (latent variables,3), f, minus 1 if there isn't a non-zero intercept or minus 2 if there is a non-zero intercept or the data are mean-centered.

DOF = n - f or DOF = n - f - 1

4. IMPLEMENTATION AND RESULTS

4.1 Prologue and Implementation details

In Fani Ampatzi's thesis, for a mixture of three components, with concentrations similar to ours (Thymol-Malachite Green- Methylene Blue), the error on unknown data (rRMSEP %) ranges from about 4 to 20 percent, concerning the substance. For example an error metric of RMSEP of 10^-7, is acceptable if our concentrations range from 10^-5 to 10^-6, but it is unacceptable for ranges less than 10^-7. On our metrics, we have very similar values to those, regardless the fact that the deconvolution procedure on our case can easily fail. Observe the spectra presented on the previous section, it is easy to understand that the two basic dyes, have very similar spectral behavior making it difficult to get results of higher accuracy.

The calibration-training set used, is the one mentioned on the previous section 3.3 consisting of the pure spectra and the mixtures spectra. On the following section, the results of prediction of the training data and unknown samples (samples with known concentrations for comparing, but not the training set) are calculated. RMS Errors of four kinds – presented on section 3.4 – are calculated and compared. For convenience, on the next chapters of this diploma thesis, Eosin Y will be referred to as **Comp1** and the two basic substances, as **Comp2** and **Comp3** for Methylene Blue and Azure B, respectively.

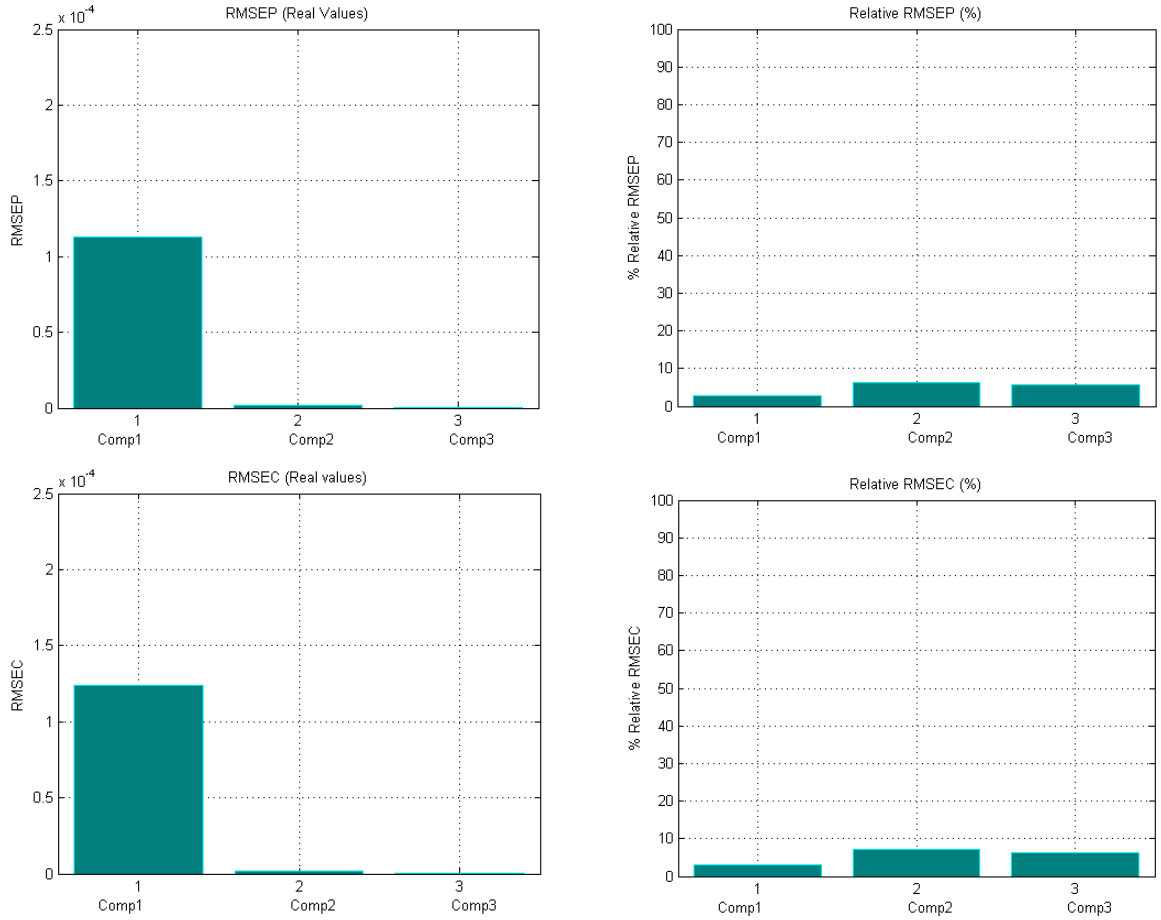
4.2 Results of full cube calibration

4.2.1 Full Cube Calibration set prediction results

First I present the results for the training and the unknown data for full cube calibration. The visible spectrum bands used are 420 nm, 450 nm, 465 nm, 505 nm, 530 nm, 600 nm, 640 nm, 666 nm, 685 nm and the infrared bands 720 nm, 735 nm, 770 nm, 810 nm, 940 nm. After 750 nm, we can clearly observe from the previously presented pure spectra of our substances, that we have zero absorbance. Using this fact, I consider my full cube to be: [420nm, 450nm, 465nm, 505nm, 530nm, 600nm, 640nm, 666nm, 685nm 720nm, 735nm, 770nm]

The next table and bar graphs presents the errors from the full cube calibration on the prediction of the calibration-training concentration data:

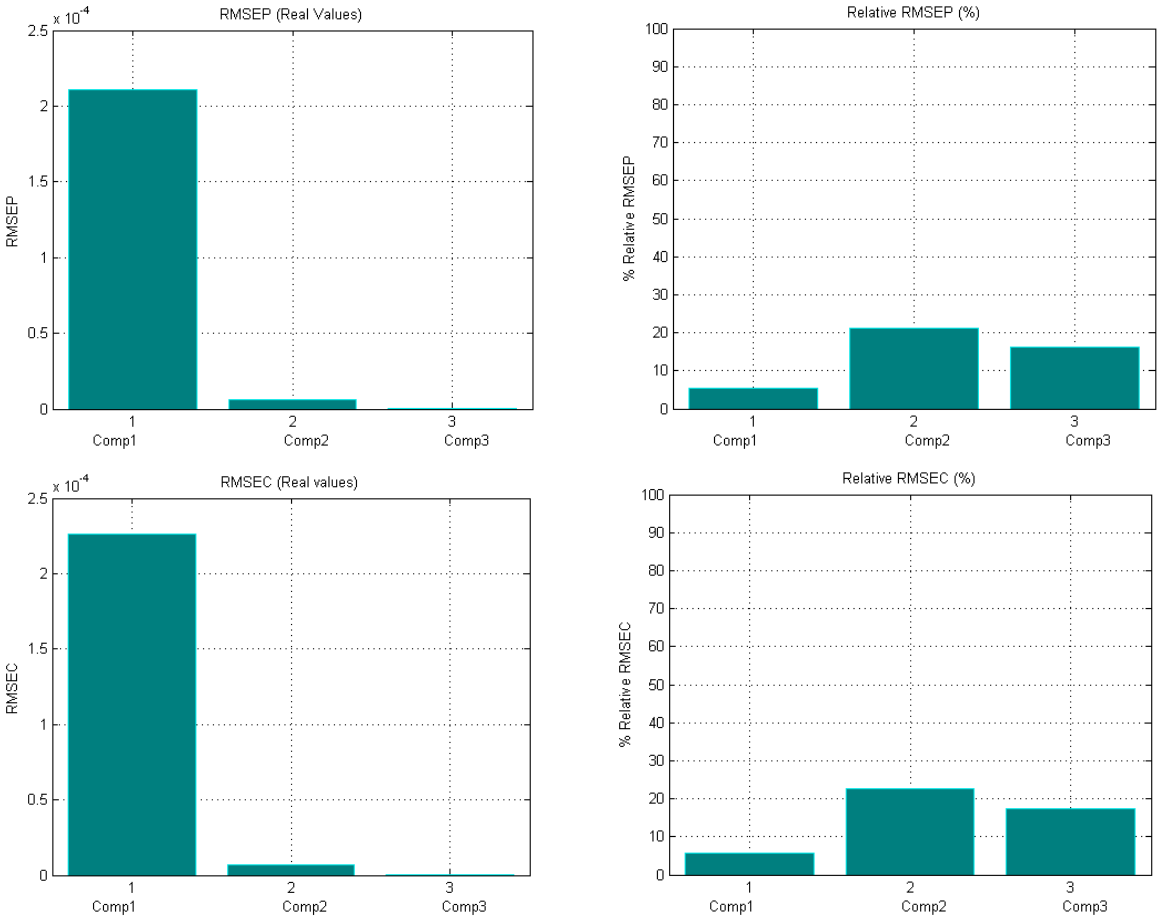
	Comp1	Comp2	Comp3
RMSEP	1.1277e-04	1.9178e-06	2.3232e-07
Relative RMSEP (%)	2.8	6.38	5.81
RMSEC	1.2407e-04	2.1101e-06	2.5561e-07
Relative RMSEC (%)	3.1	6.9	6.39



4.2.2 Full Cube "Unknown" dataset prediction results

The next table and bar graphs presents the errors from the full cube calibration on the prediction of an "unknown" dataset. As expected the errors are increased in comparison with the training data errors, but they are still acceptable for our case.

	Comp1	Comp2	Comp3
RMSEP	2.1054e-04	6.3472e-06	6.5087e-07
Relative RMSEP (%)	5.23	19.16	16.2
RMSEC	2.2616e-04	6.8180e-06	6.8915e-07
Relative RMSEC (%)	5.65	21.16	17.48



It is important here, to mention two facts. First, our full cube calibration on PLS results to similar error results for unknown data as those expected, based on previous study. Secondly, the relative RMSEs in both training and unknown data reveal the efficiency of the algorithm in a more interpretable way. As in mapping, we will be concerned for relative values of concentration, and not the exact ones, the maximum of about 20% error, is more than fine for our experiment. From now on, on this diploma thesis, we consider those full cube results as golden standard for our metrics.

4.3 A-priori data bands reduction (Dimension Reduction)

A common method for faster regression predictive systems, is the Band Reduction or Dimension Reduction method. During this procedure, we try to reduce the needed a-priori data to the least. Knowing we have the least needed dataset in terms of samples, we now have to reduce the number of wavelengths we used, keeping the efficiency of the algorithm close to our golden standard (full cube results).

There are many ways to do the band reduction. First, observing the PLS algorithm weights. Establishing the weights for each band, PLS algorithm, in a way, performs a kind of reduction on

the a-priori data, setting weights close to zero for the data that have decreased correlation. It is possible to keep the bands that's have the biggest weight values, repeat the experiment with the new, reduced data set and compare the results. Another way is by removing one band at a time and comparing, setting some standard error values as threshold. That of course requires lots of combinations among the bands. Also, more sophisticated "pipeline" ways exist, as selecting bands from the dataset and using PLS or other regression algorithms, like PCA, to double check the efficiency or the regression with cross-validation.

For our experiment, our goal is to **result to the use of 6 spectral bands for concentration prediction, without loss of significant information**. It is obvious that the bands on the visible part of the spectrum are the most informative ones. So, we expect from the PLS regression to weight those bands more heavily than the Infrared ones. The next plot is the Weights that our regression calculates for the full cube:

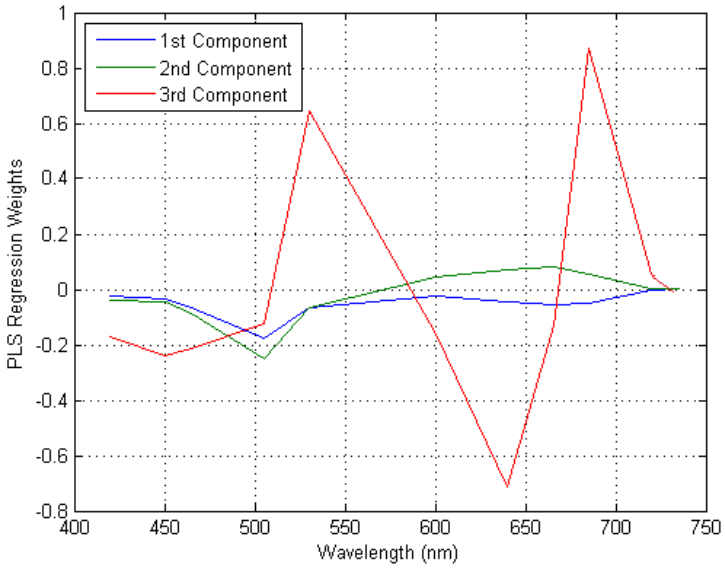


Figure 30. PLS Weights on Wavelengths

From this plot we can extract conclusions about which wavelength is weighted more heavily for each component. The choice of bands has to be done very carefully, so that we will end up with the bands that are weighted more heavily for all components at the same time and not a number of them. For example, if we have to choose among the 666nm and 68nm, we should choose 685nm. That's because the 1st and 2nd component are weighs are almost the same for the two bands, but the 3rd component is weighted much more heavily on 685nm. Same on 450nm with 465nm. The 3rd component is alone more heavily weighted on 450nm, but the weights of 1st and 2nd components are poor and almost the same. This changes slightly on 465nm so, we choose 465nm. The selected bands are shown with the red marks on the next plot:

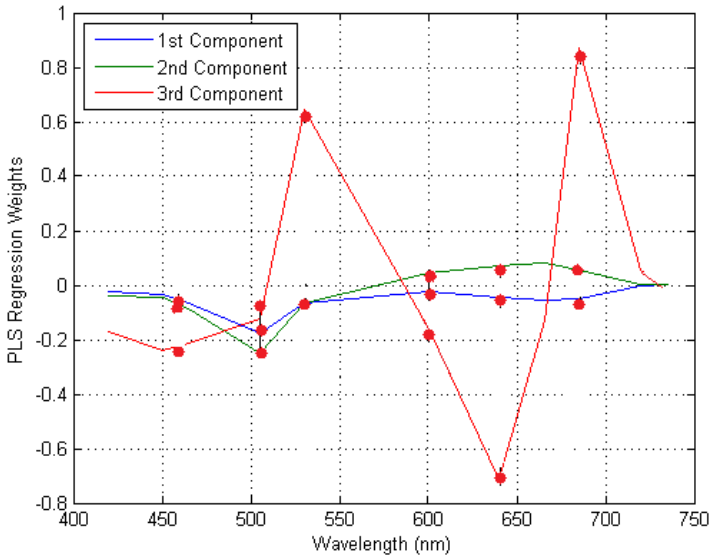


Figure 31. PLS Weights on Wavelengths (Marked selected)

Selected Bands: 465 nm, 505 nm, 530 nm, 600 nm, 640 nm, 685 nm.

Generally, someone could identify the heavy weighted bands just by observing the characteristics of the full cube. The aforementioned procedure is used in order to understand which one of two visually counterpart bands is better for the calibration of the specific algorithm. Also, someone could observe the spectral signatures of the staining substances and several combinations of them. For example, let us assume that we have the MG-G mixture and we measure this spectral data on known concentrations:

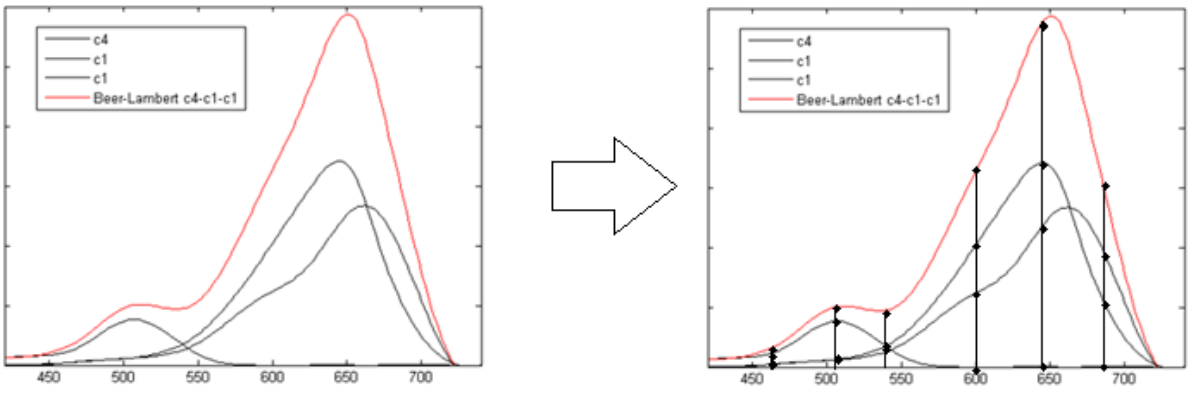
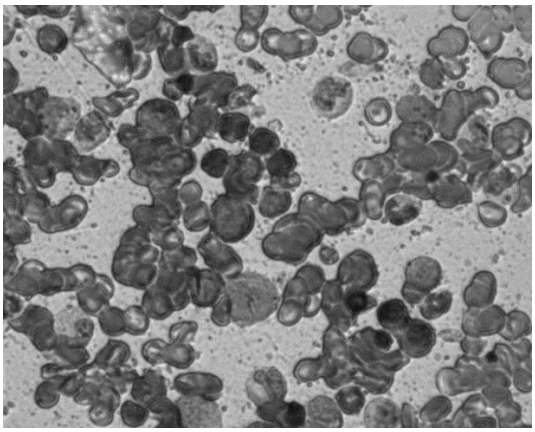
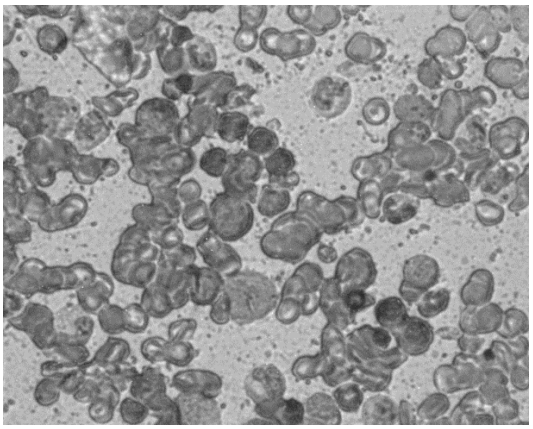


Figure 32. Selected Wavelengths on absorbance spectra

On the second part of the image, the bands of interest, are marked. For example, visually judging, someone could without hesitation use from the given bands, not the 685nm, but the 666nm, which is proven to have the same or more absorbance features:



666nm



685nm

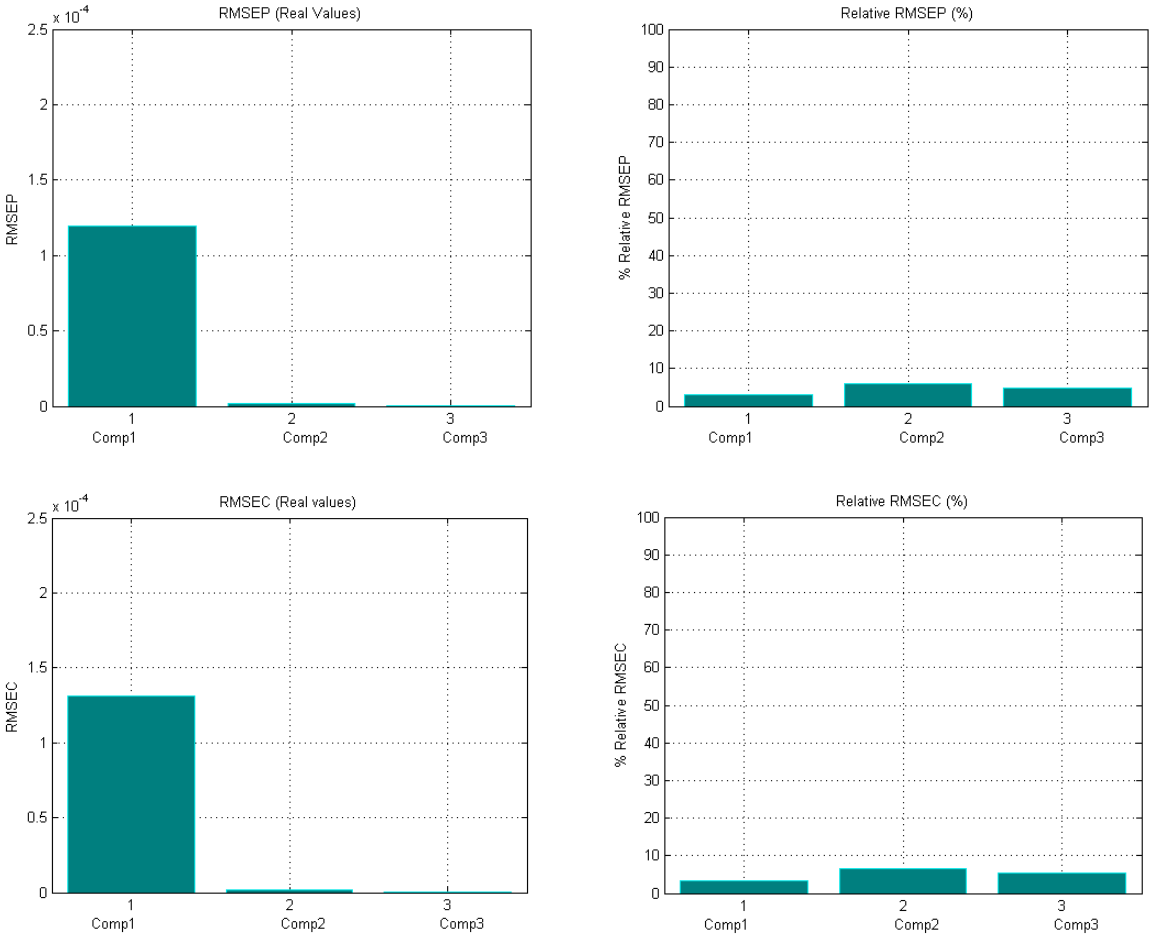
But if we observe, the algorithm prefers the training with the 685nm band, and weights it more heavily because there, we have different spectral characteristics of the individual components. Giving more heavy weights on training, generally leads to better deconvolution results.

4.4 Results of Six selected bands calibration

4.4.1 Six selected bands calibration set prediction results

Now, as aforementioned, we have to check –repeating the experiment- if the errors have been increased, decreased of remain steady. We set a new data set, consisting of only the selected bands and we use it again for calibration of PLS. The next table and bar graphs presents the errors from the 6-selected bands calibration on the prediction of the calibration-training concentration data:

	Comp1	Comp2	Comp3
RMSEP	1.1958e-04	1.8153e-06	1.9211e-07
Relative RMSEP (%)	2.9	6.05	4.8
RMSEC	1.3156e-04	1.9973e-06	2.1137e-07
Relative RMSEC (%)	3.21	6.66	5.28

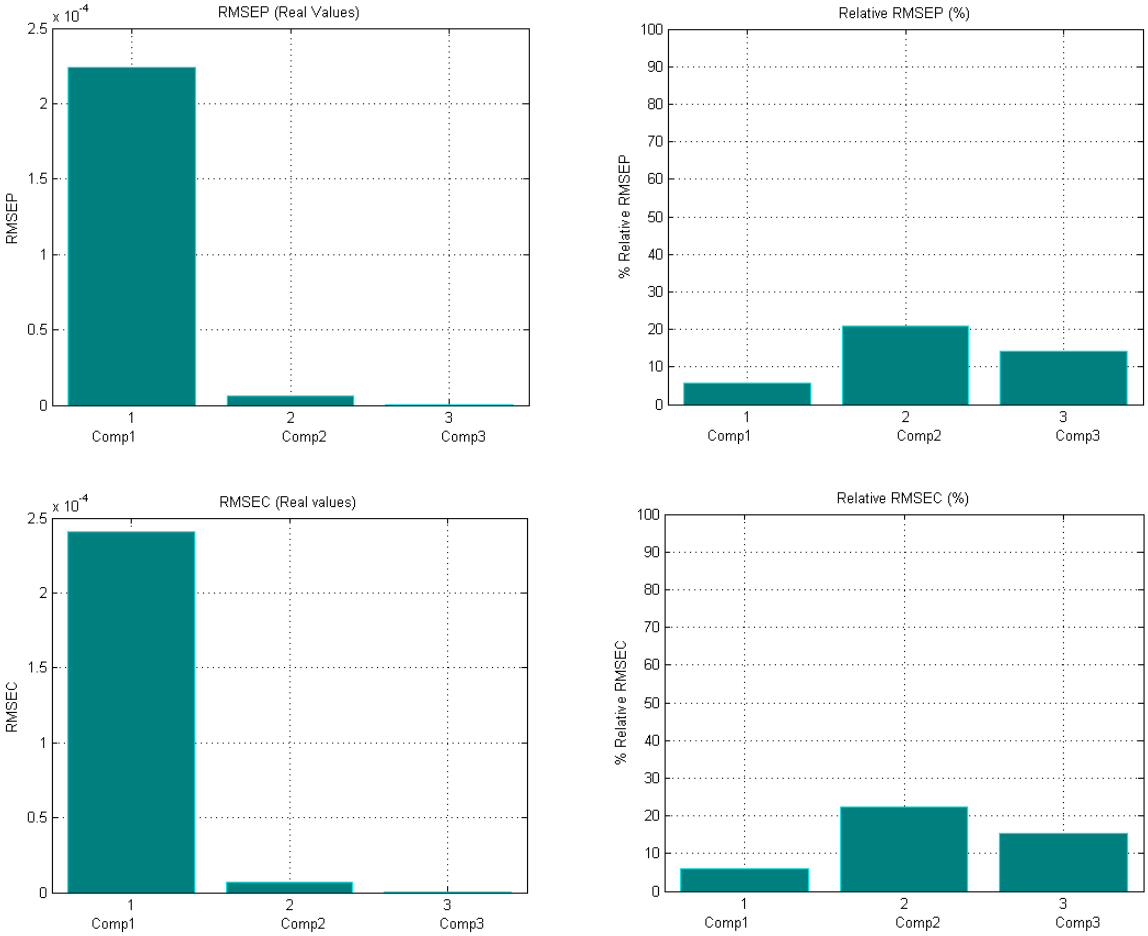


Observe: For Comp1, we have a slight increase of error, but insignificant. On the other two compounds, Comp1 and Comp2, not only we have no error increase, but the errors that shows the differences between the predicted and the real data, are decreased, almost 1% each. Obviously **the regression with selected components had more successful results on this experiment**.

4.4.2 Six selected bands "Unknown dataset" prediction results

The next table and bar graphs presents the errors from the 6 selected bands calibration on the prediction of an "unknown" dataset. As expected the errors are increased in comparison with the training data errors, but they are still acceptable.

	Comp1	Comp2	Comp3
RMSEP	2.2417e-04	6.2877e-06	5.6862e-07
Relative RMSEP (%)	5.6	18.96	14.22
RMSEC	2.4080e-04	6.6541e-06	6.1080e-07
Relative RMSEC (%)	6.02	20.51	15.2

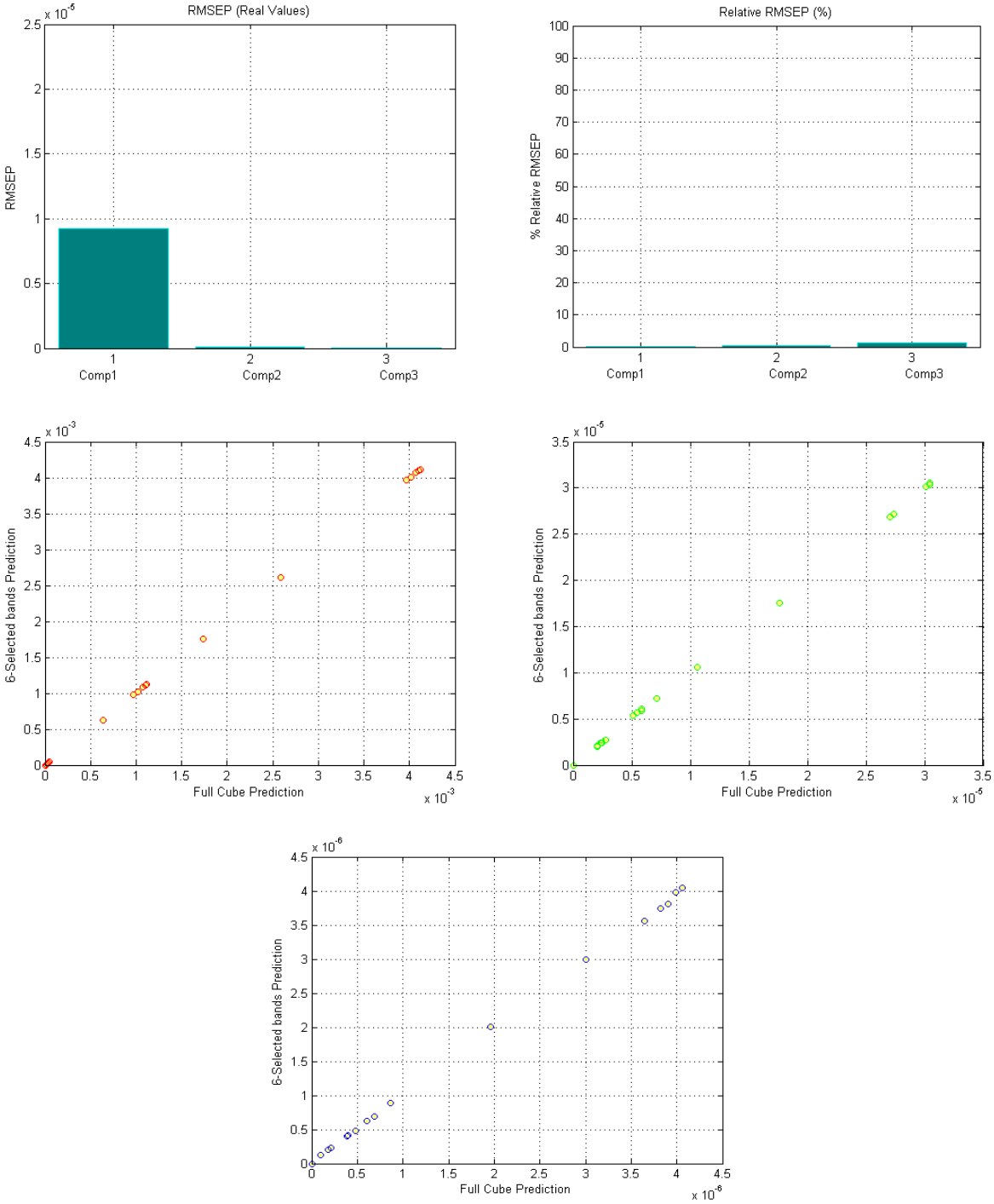


Again, as for the training data prediction, we observe: For Comp1, we have an insignificant increase of error. On the other two compounds, Comp1 and Comp2, not only we have no error increase, but the errors of prediction (predicted and the real data), are decreased, almost 1% each. The efficiency of the algorithm is increased with a six-band training set in comparison with a full-cube (12 bands) training set.

4.5 Comparison between full cube and 6-selected bands results

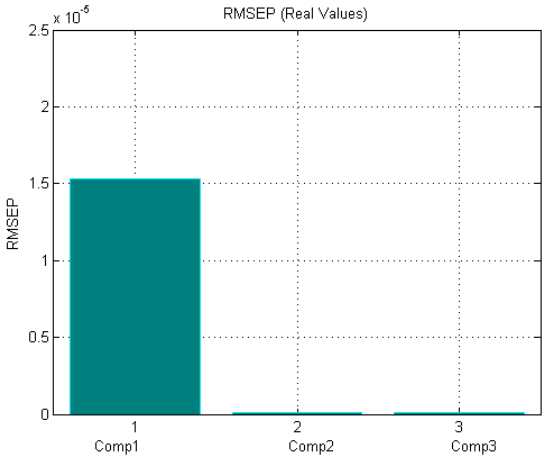
The next table, bar graphs and relation diagrams, presents the errors between the prediction of the training dataset by the full cube calibration and the 6 selected bands calibration.

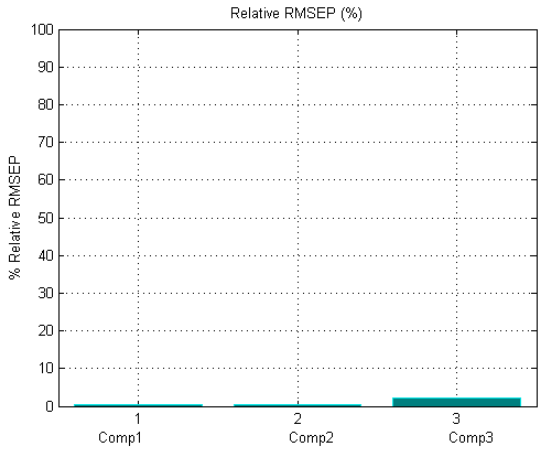
	Comp1	Comp2	Comp3
RMSEP	9.2596e-06	1.3346e-07	5.5058e-08
Relative RMSEP (%)	0.23	0.41	1.06

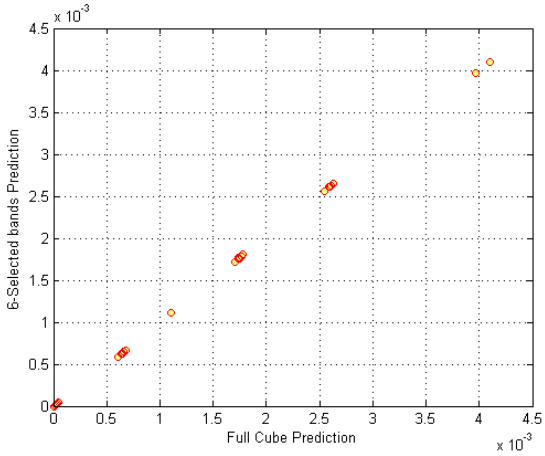


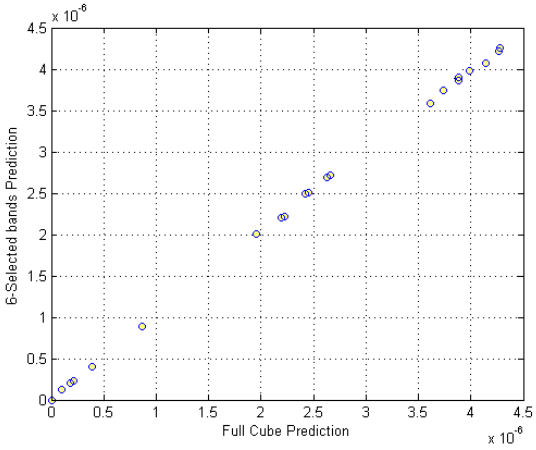
The next table, bar graphs and relation diagrams, presents the errors between the prediction of an "unknown" dataset by the full cube calibration and the 6 selected bands calibration.

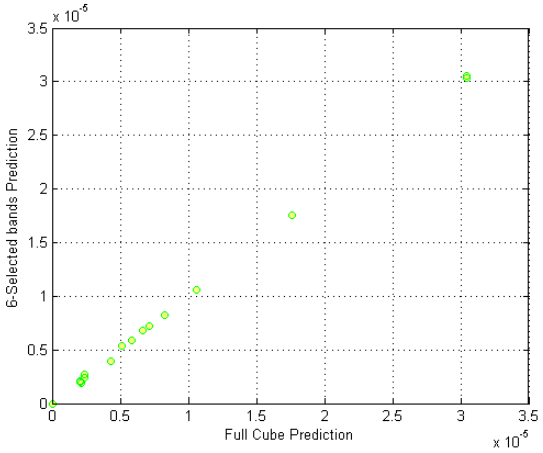
	Comp1	Comp2	Comp3
RMSEP	1.5327e-05	1.2387e-07	8.8244e-08
Relative RMSEP (%)	0.37	0.41	2.06











From all the metrics presented, it is obvious that the band reduction was successful. The results of the 6-selected bands training were not only almost identical to those of the full cube training, but the predicted data appears to be closer to the wanted. On the next chapter of the

diploma thesis, we are moving on spectral imaging data. Using the coefficients matrix from our training set, we predict the concentration data on pixels.

4.6 Spectral Imaging and pixel-wise concentrations prediction

As mentioned, imaging spectroscopy is the application of spectroscopy, spatially in every pixel of a spectral cube. So far, we found that PLS algorithm can deal with samples of 6 wavelengths absorbance spectra and predict the sample's concentration(s) accurately. It is important to mention that for each wavelength of the spectral cube, we chose the dominant channel (R, G or B). Each spectral image captured by our system, is filtered with Wiener filter for noise reduction.

Wiener Filter Reference

Wiener is a low-pass filter that is used on a grayscale image that has been degraded by constant power additive noise. It uses a pixel-wise adaptive method based on statistics estimated from a local neighborhood of each pixel. More specifically, Wiener estimates the local mean variance around each pixel, on [M, N] neighborhood, as:

$\mu = \frac{\sum_{n1,n2\ E\ n} a(n_1,n_2)}{MN}$ and $\sigma = \frac{\sum_{n1,n2\ E\ n} \alpha^2(n_1,n_2)}{MN}$ - μ^2

The wiener estimates are:

$b(n_1,n_2) = \mu + \frac{\sigma^2-\nu^2}{\sigma^2}(a(n_1,n_2)-\mu)$, where ν^2 is the noise variance, calculated as the average of all the local estimated variances.

Let us assume that each pixel of the spatial coordinates of a six-band spectral cube captured by our microscopy system (with its intensities (transmittance) transformed to absorbance using Beer-Lambert, section 1.6.3) is a six-band sample that can be a valid input for our predictive algorithm. Note that for transmittance, on I/Io, as Io we used the value 255, means that the values we get are normalized. Understanding these assumptions, it is easy to interpret how we worked on the spectral images. All we needed was a high quality, calibrated spectral cube consisting of microscopy images of the stained sample (on an area of interest) on the specific 6 wavelengths: 465nm, 505nm, 530nm, 600nm, 640nm, 685nm. The spectral cube gets transformed into matrix, with dimensions X*Y-by-6, where X, Y the spatial coordinates of the spectral cube. Schematic of the procedure on the next image:

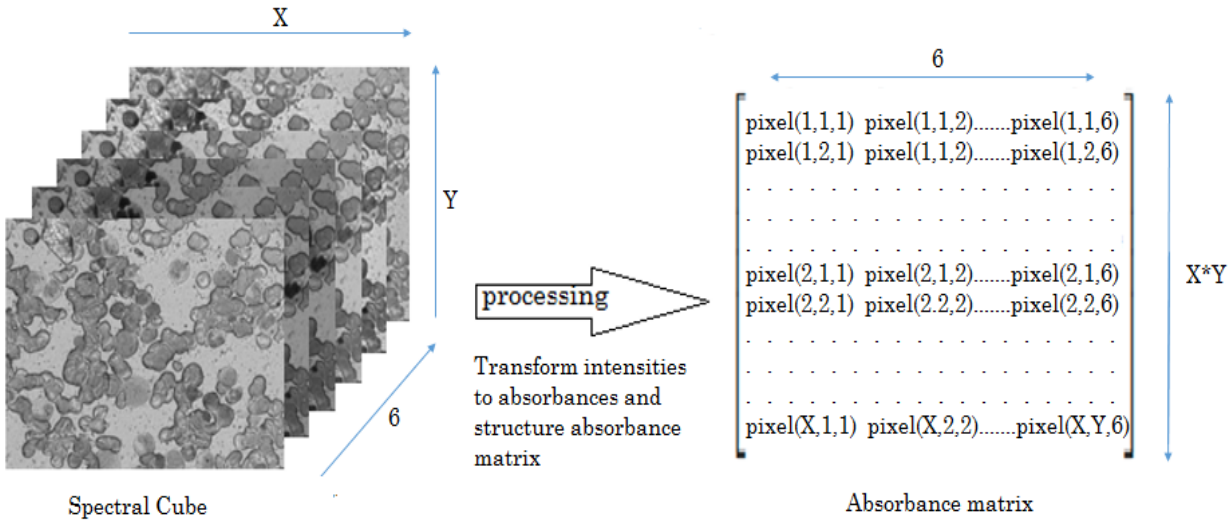


Figure 32. Spectral Cube to Matrix-Input for PLS

Now, using the formed absorbance matrix, the concentrations of the three staining substances can be estimated for each pixel using PLS prediction. The predicted concentrations will be presented as an X*Y-by-3 matrix, that's has the predicted concentrations for substance 1, substance 2, substance 3, on the columns 1, 2, 3 respectively.

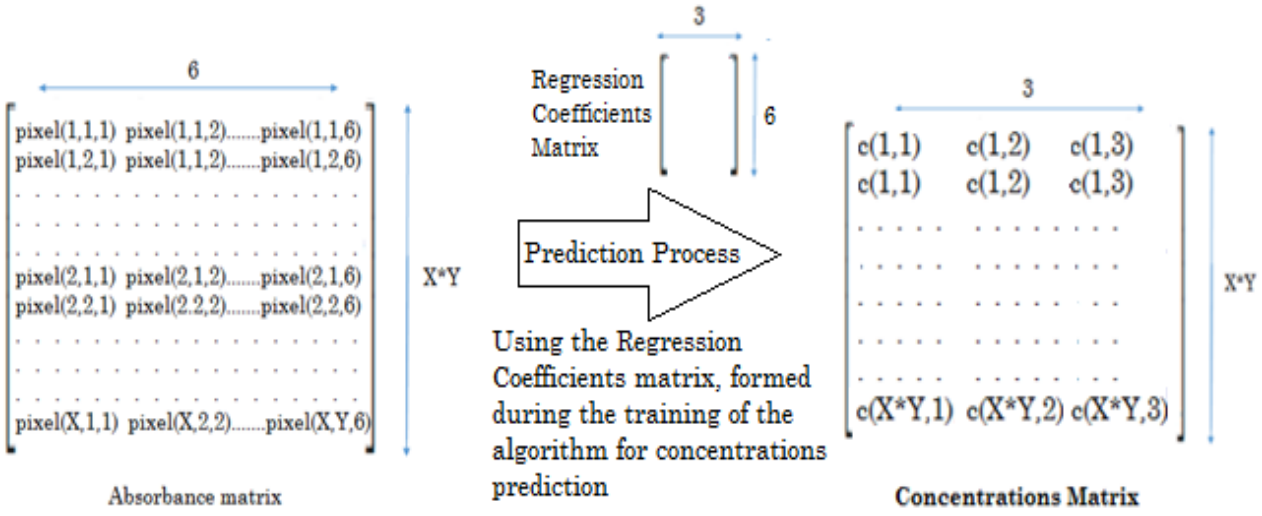


Figure 33. Concentration Prediction using PLS

The columns –each one multiplied with the right constant in order to reach values 0 to 1 (0 to 255)-can be transformed back to image form one by one, knowing the X, Y (spatial coordinates) of the initial spectral cube. After this procedure, we have three grayscale images, that their

common characteristic is that the lighter a pixel is, the bigger concentration it depicts. Schematic of the procedure on the next image:

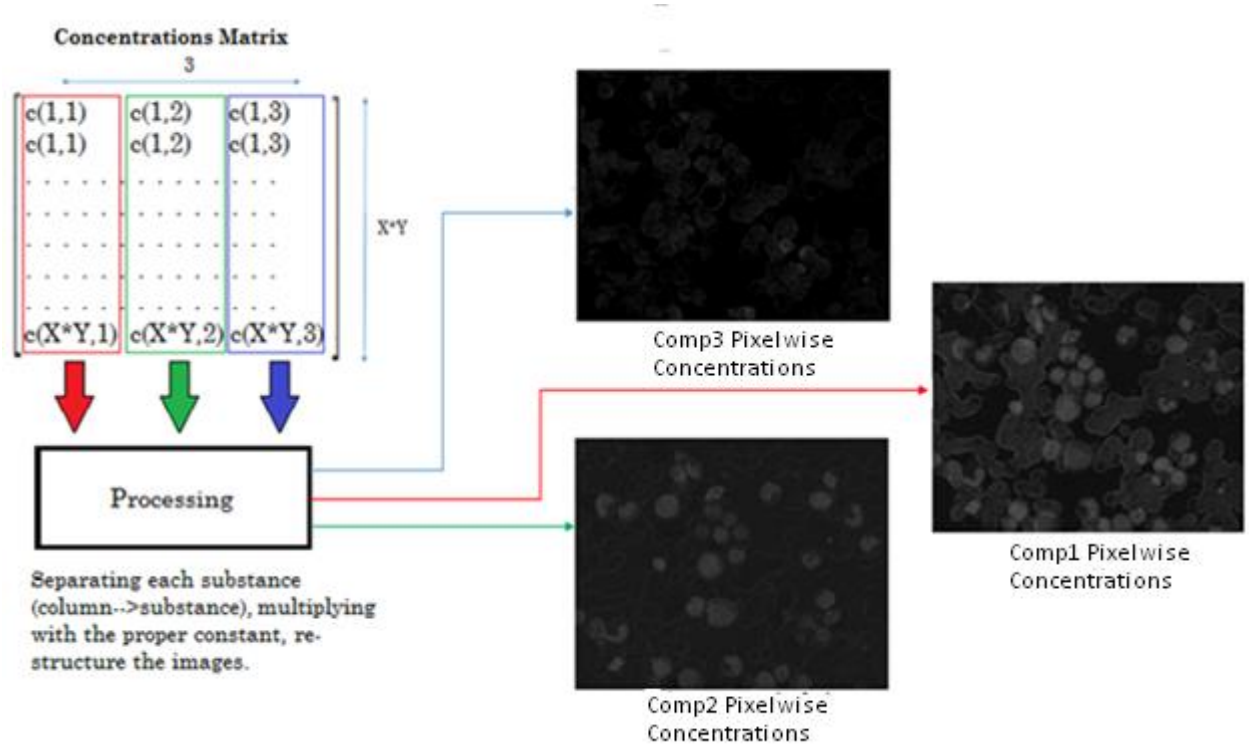


Figure 34. Concentrations to Images

4.7 Mapping Predicted Concentrations on Leukemia tiles

4.7.1 Pseudo-chromatic maps

As described before, on this diploma thesis we follow a procedure in order to result to three images, pixel-wise concentration maps of the area of interest of a Leukemia tile, one for each staining substance. From a six band spectral cube, using the PLS algorithm's prediction (after training the algorithm with our data), we end up after the aforementioned processing, on three grayscale images (0 to 255) shown in figure.

Looking at those images, conclusions can be extracted but not easily. There is a great need for classification of the pixels in a way that the results will be easier to interpret at first by ourselves, and then by the doctors. This way is no other than classification of the concentration values for each substance as we are not exactly interested on the exact value of concentration, but more on a proportional way of depicting the three concentrations.

When the system was ready, we tested it on about 45 spectral cubes, from 15 different blood and bone marrow samples all stained the same way with MG-G. For each substance (Concentration Images), for all examples, we calculated:

- the minimum concentration value
- and maximum concentration value
- the concentration values of the background (places on tile without cells)
- the mean concentration values

It is important for the success of our experiment to mention here, that the minimum and maximum concentrations, as well as the background concentrations were very close or identical in every single example.

Using these data, and of course the medical information that concern the parts of the cell that each substance stains mostly, shown in Figure 7, we are able to perform a classification and from that classification extract **pseudo-chromatic endmember scaled concentration maps (Endmember maps)**. We started with background extraction. The background had to be black, showing no signs of substance existence on the spot. If the background values show concentration of any substance, this will be confusing for the user of the system. For example (Comp 1):

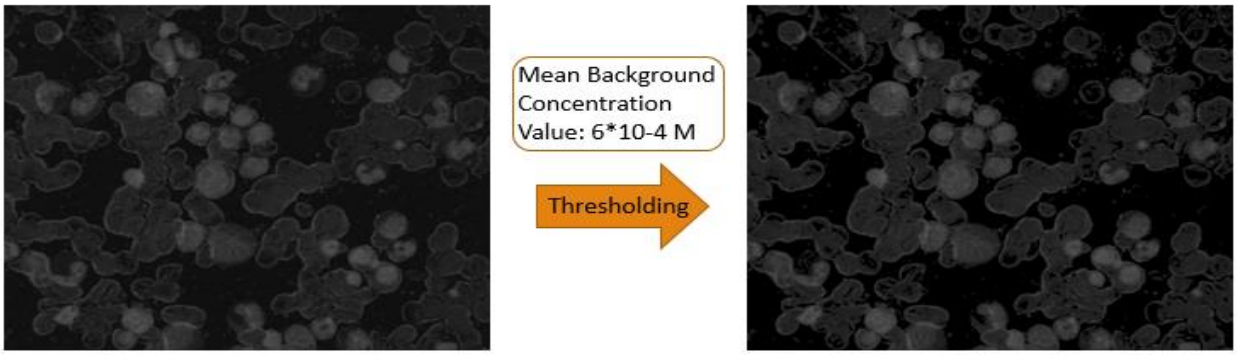


Figure 35. Threshold mapping

Despite that fact, we had to be careful with the background threshold for every substance in order not to lose information concerning the cells. Our purpose is that in the end, the maps will have a zero (black) background, and only cells would appear in certain colors. It is not always true though, because sometimes background reveals stained artificial features, mostly by staining mistakes, dust etch.

The rest of the mapping is done on the same way. The remaining concentration values of each substance were classified, until the max value, with main purpose to overemphasize the features of the cell and at the same time keep an increasing relation. Our first pseudo-maps will be monochromatic, meaning that they will depict the concentration changes in one color's shades. On the next figure we have the first monochromatic scale:

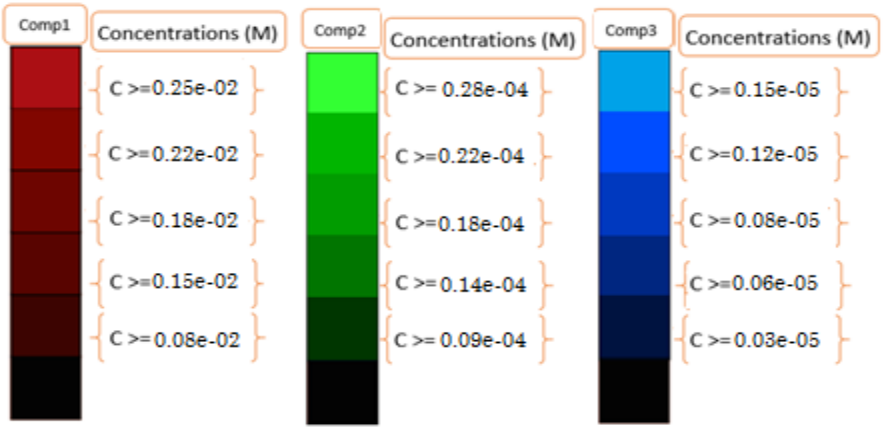


Figure 36. Mono-chromatic Scale

On the same way (and limits), we develop our second pseudo-chromatic scale, multi-colored:

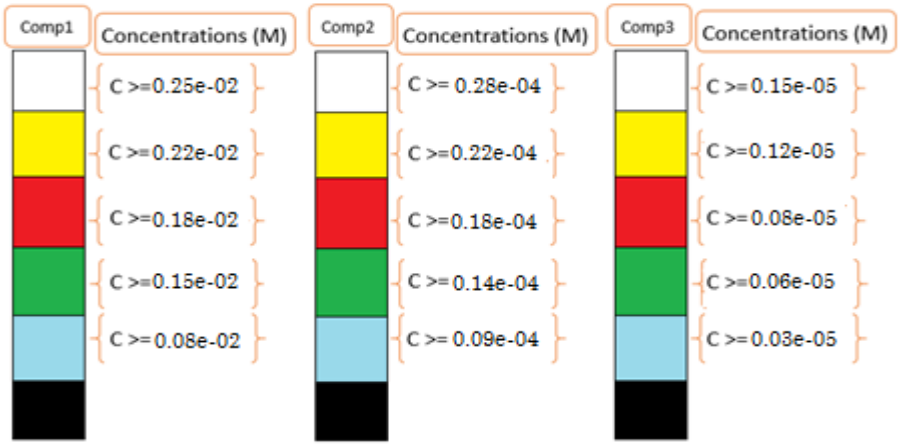
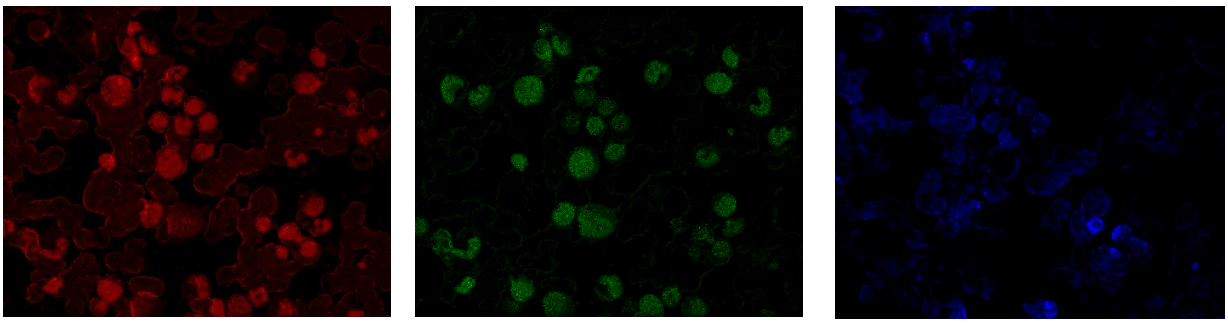


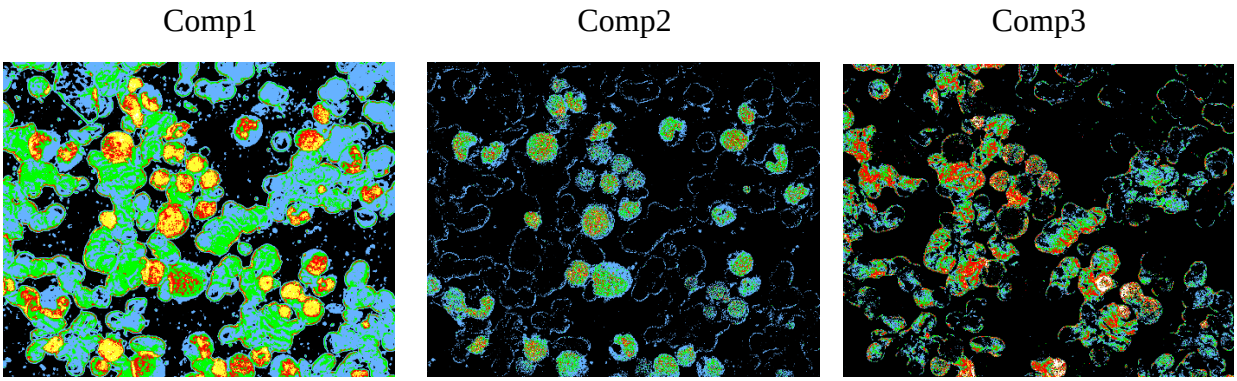
Figure 37. Multi-color chromatic scale

Pixel-wise, the predicted concentrations will get the scaled color values and the maps will be formed. As it is easy to conclude, at this time we have 6 pseudo-maps. As we will present later, the three **monochromatic endmember maps** are made with intention to get used in making another map, the Merged Proportions Map. Example monochromatic pseudo-maps on the next figure:

Comp1 Comp2 Comp3



The three **multi-colored endmember maps** are part of the final system we present. Example on the next figure:



4.7.2 Merged Proportions Map

As understood by its name, the merged proportions map, is a map intended to depict the proportions of each substance chromatically. **The merging of the three abundance maps, result to a single artificial map that clearly depicts the relative concentrations of the three substances**. More specifically, the combination of colors from the three monochromatic pseudo-maps (abundance maps), produce unique color signatures that conclude all the substances concentration information per pixel. For example one pixel that happens be one of those on the nucleus of a cell, has concentration probably by all three substances. It gets Red Channel value by the Comp1 monochromatic endmember pseudo-map, Green Channel value by the Comp2 monochromatic endmember pseudo-map and Blue Channel value by the Comp3 monochromatic endmember pseudo-map. **These results, as we will observe, lead to a unique characterization of each aforementioned part of the cell, as well as overemphasizing structures on the nucleus and cytoplasm, impossible to see with classic microscopy. This can be a useful tool in discriminating different types of white cells and consequently pathologies**. Schematic of the procedure and first example of merged map:

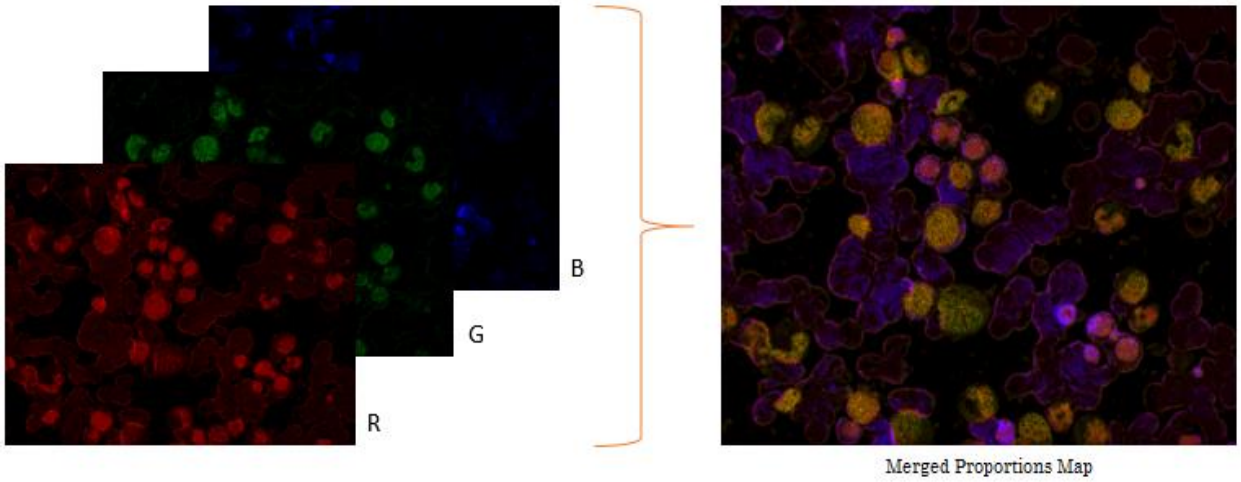
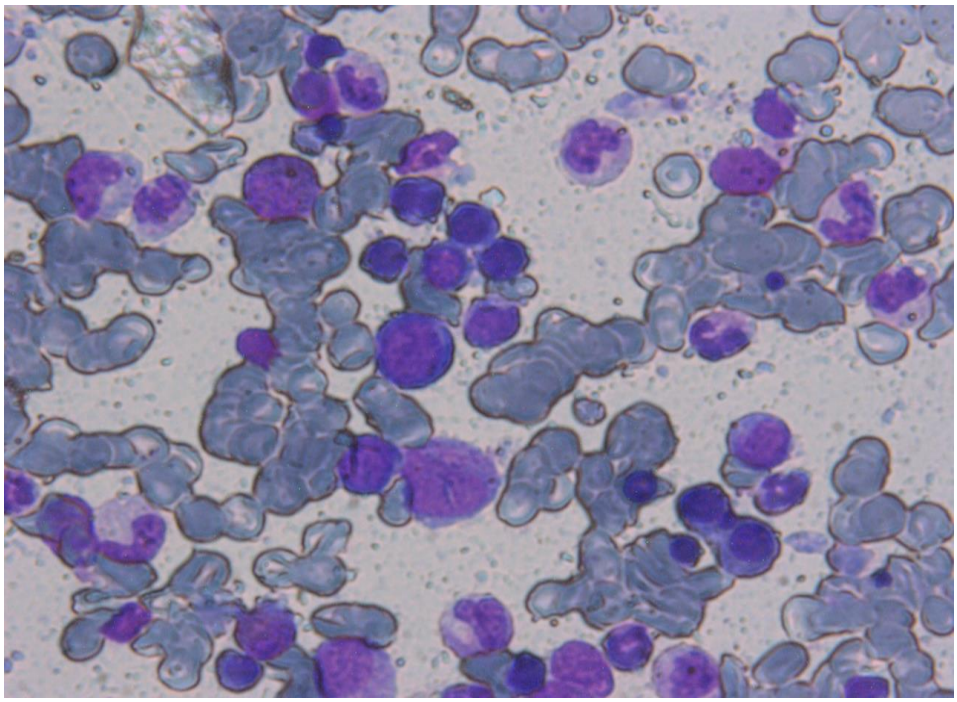


Figure 38. Merged Map Schematic

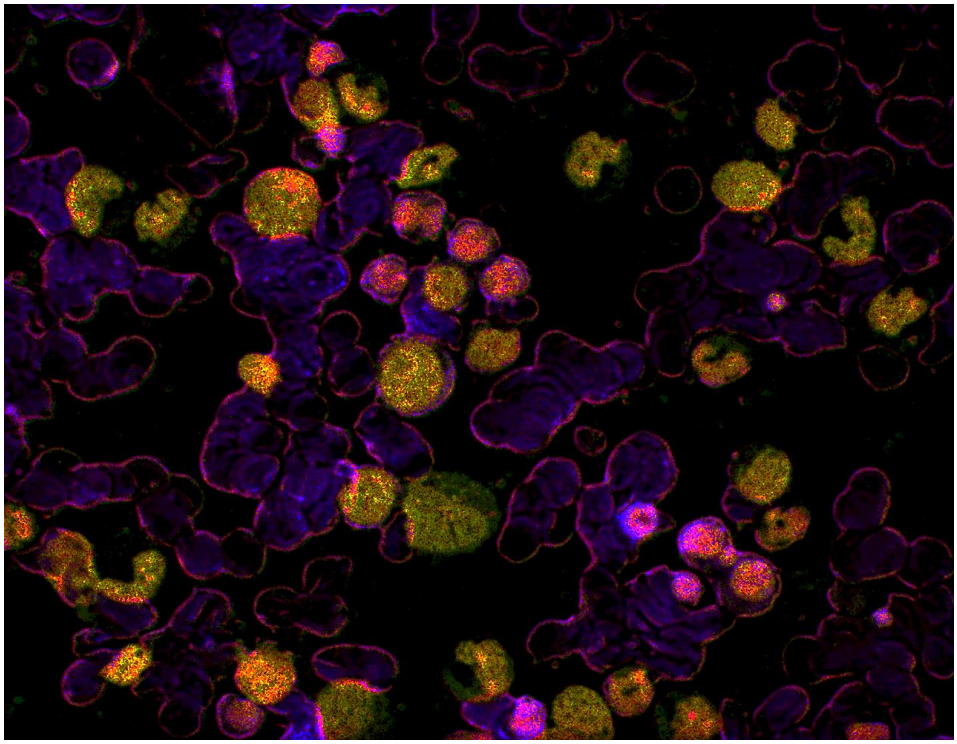
4.8 Examples of mapping and comments

4.8.1 1280x1024 Maps (XIMEA xiQ USB 3.0)

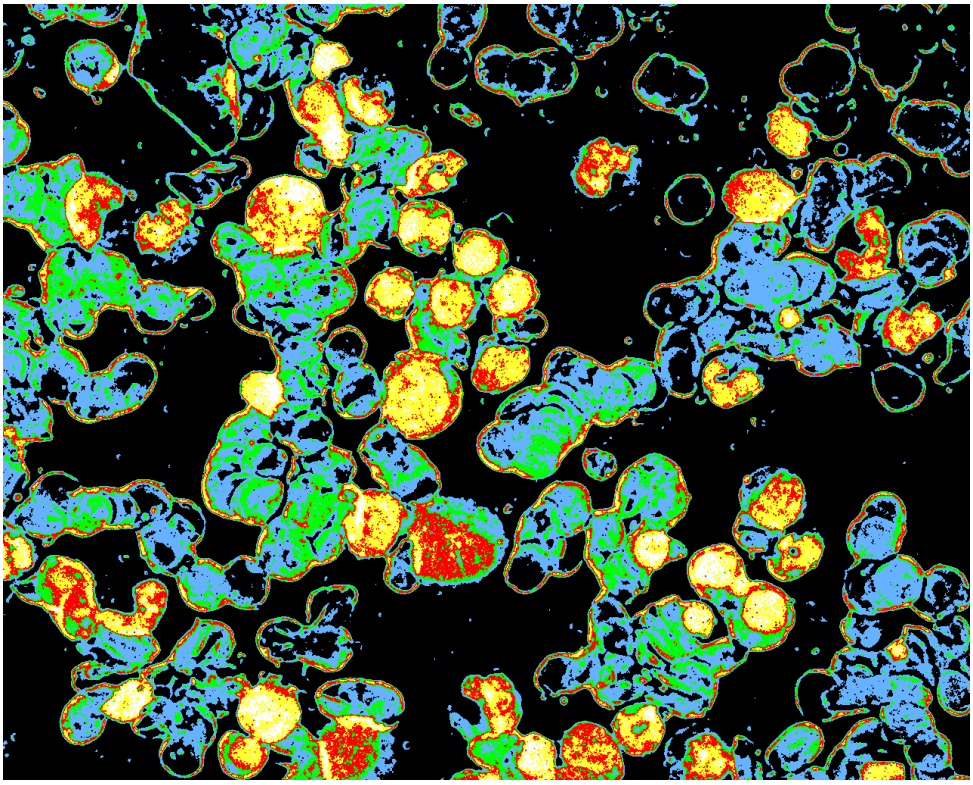
Example 1



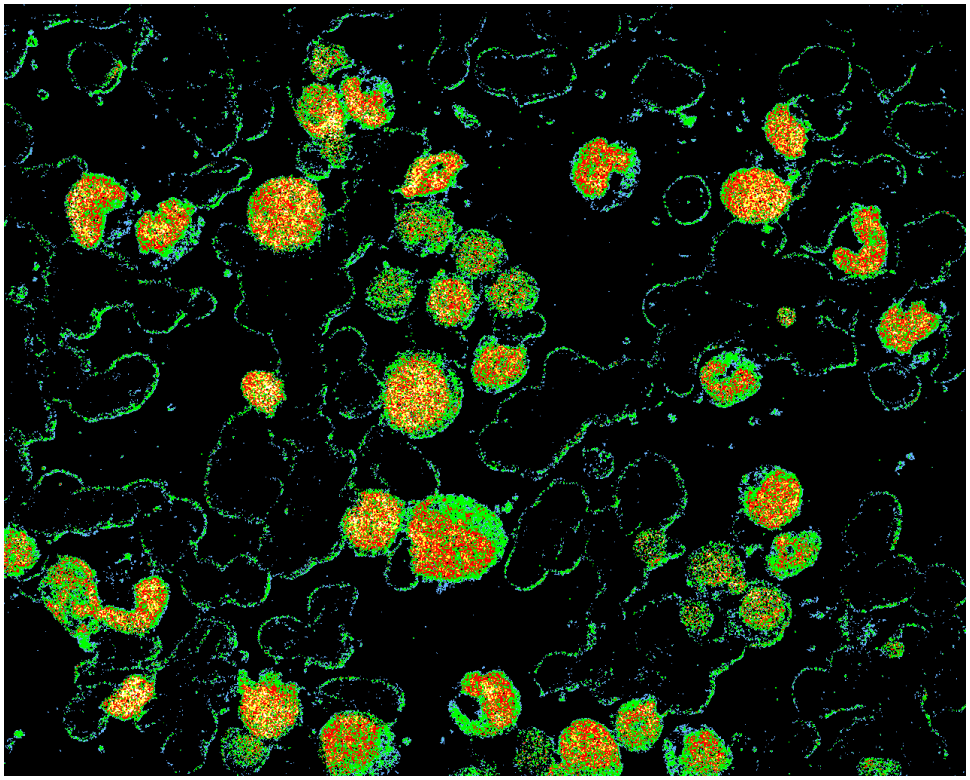
RGB Image



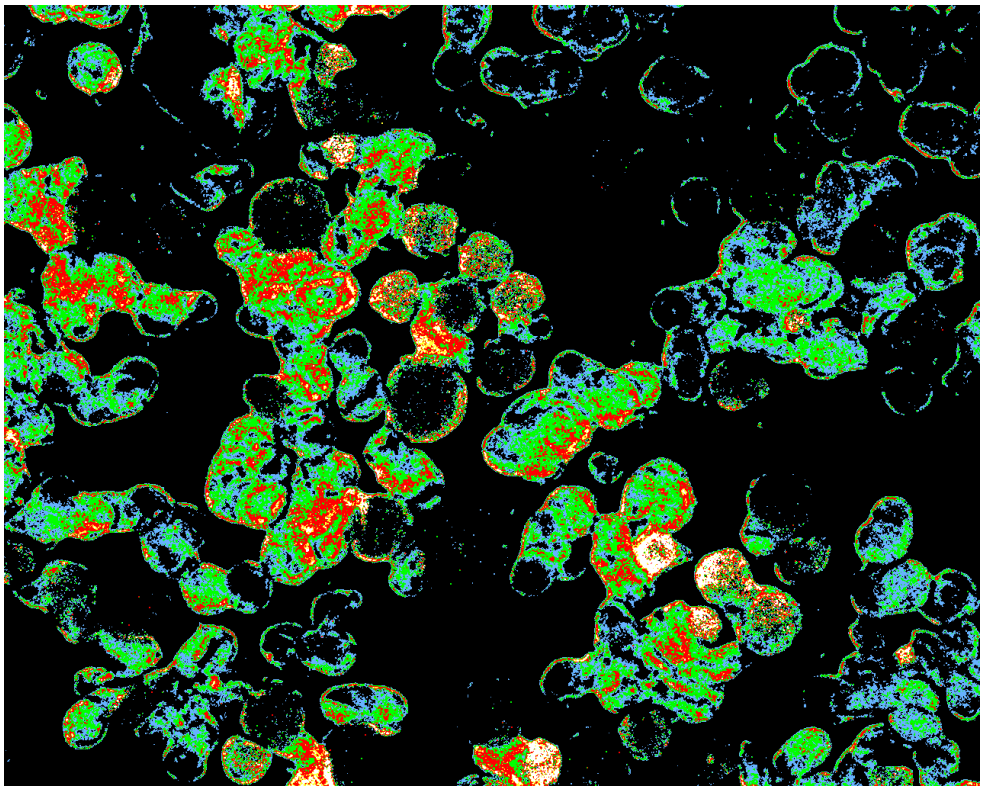
Merged Proportions Map



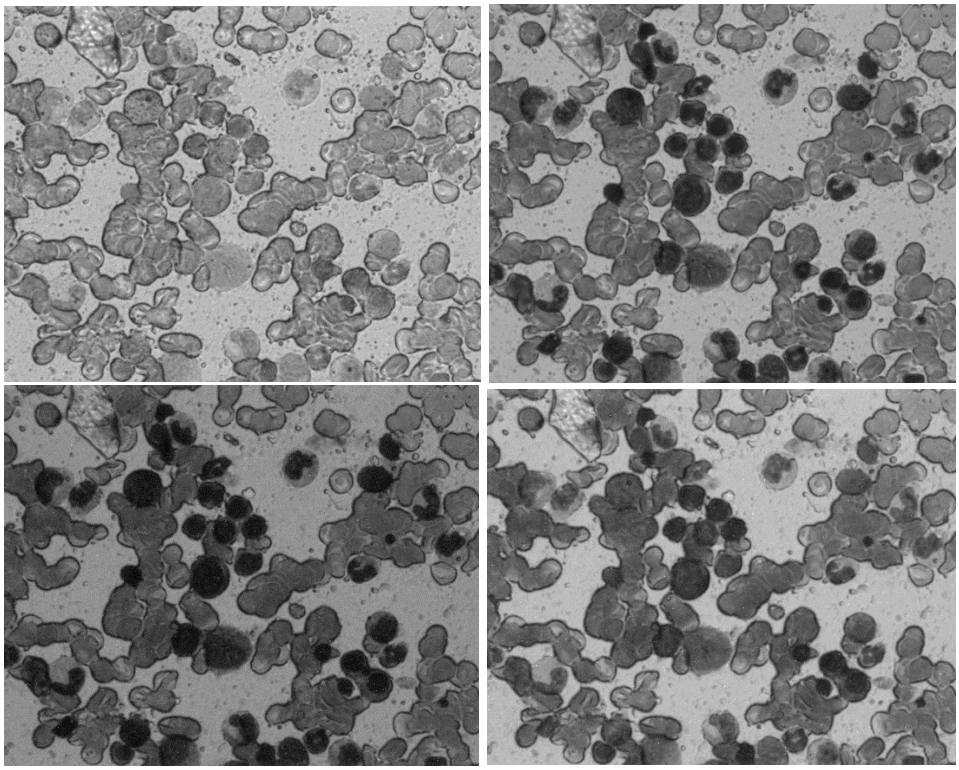
Comp1 Endmember Pseudo-map

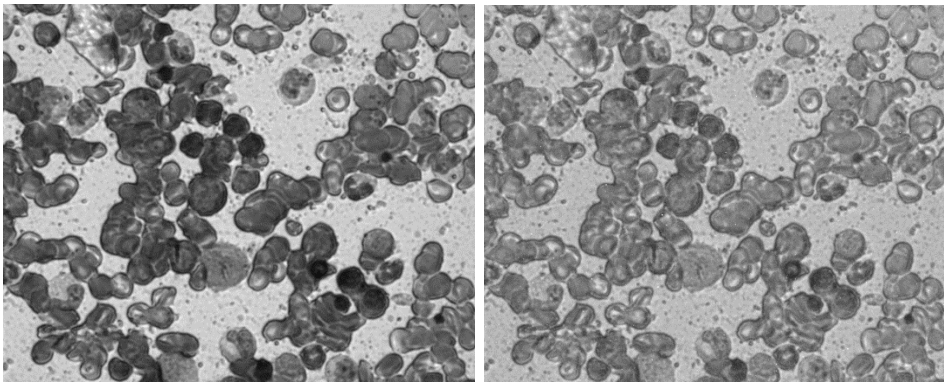


Comp2 Endmember Pseudo-map

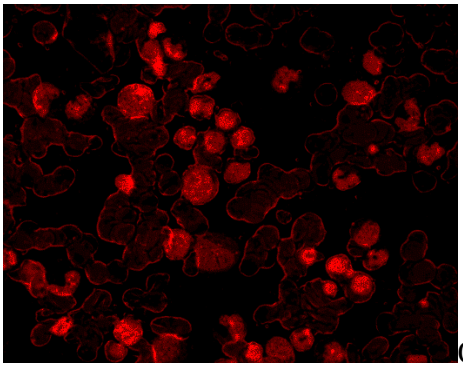


Comp3 Endmember Pseudo-map

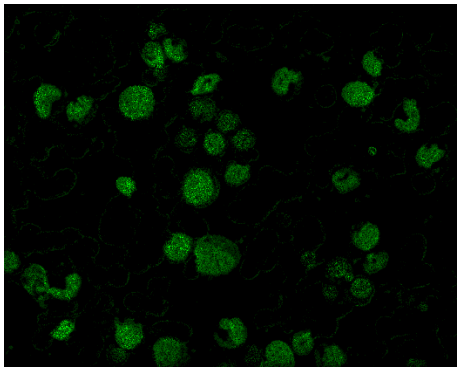




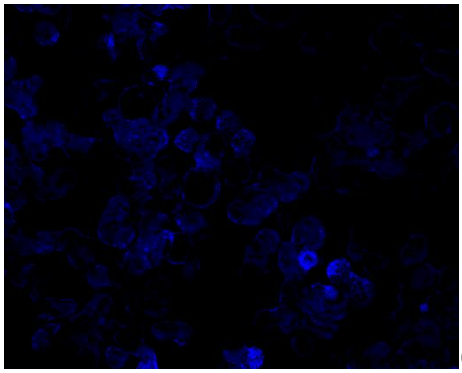
Spectral Cube Images (465nm, 505nm, 530nm, 600nm, 640nm, 685nm respectively)



Comp1 Endmember map



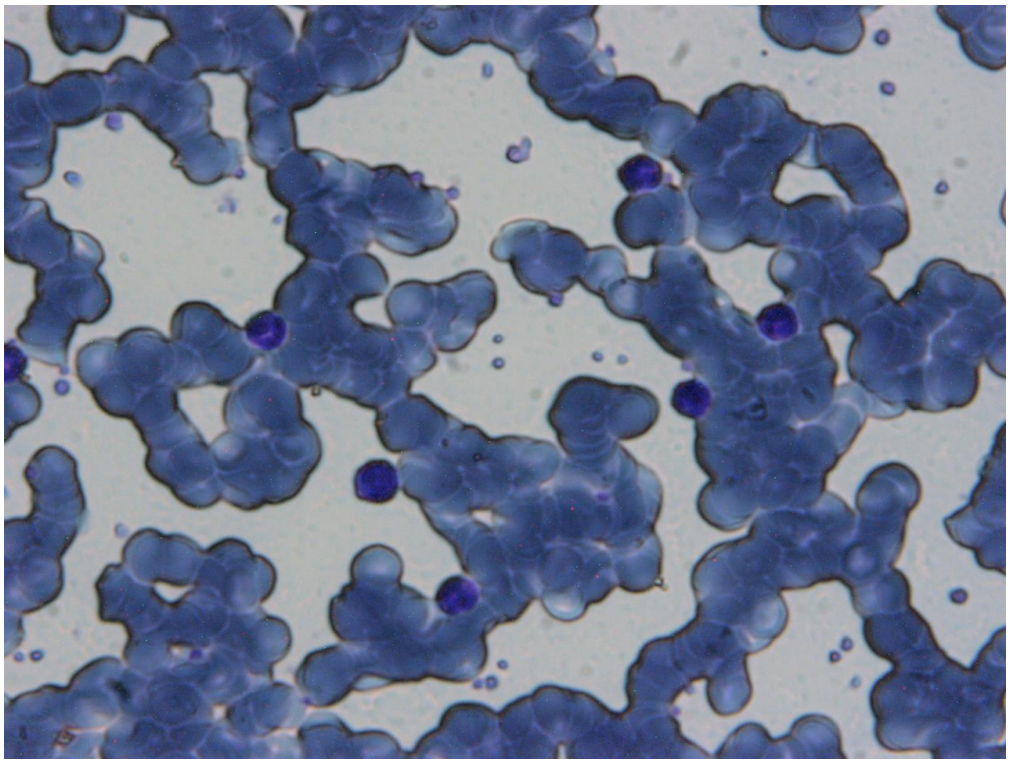
Comp2 Endmember map



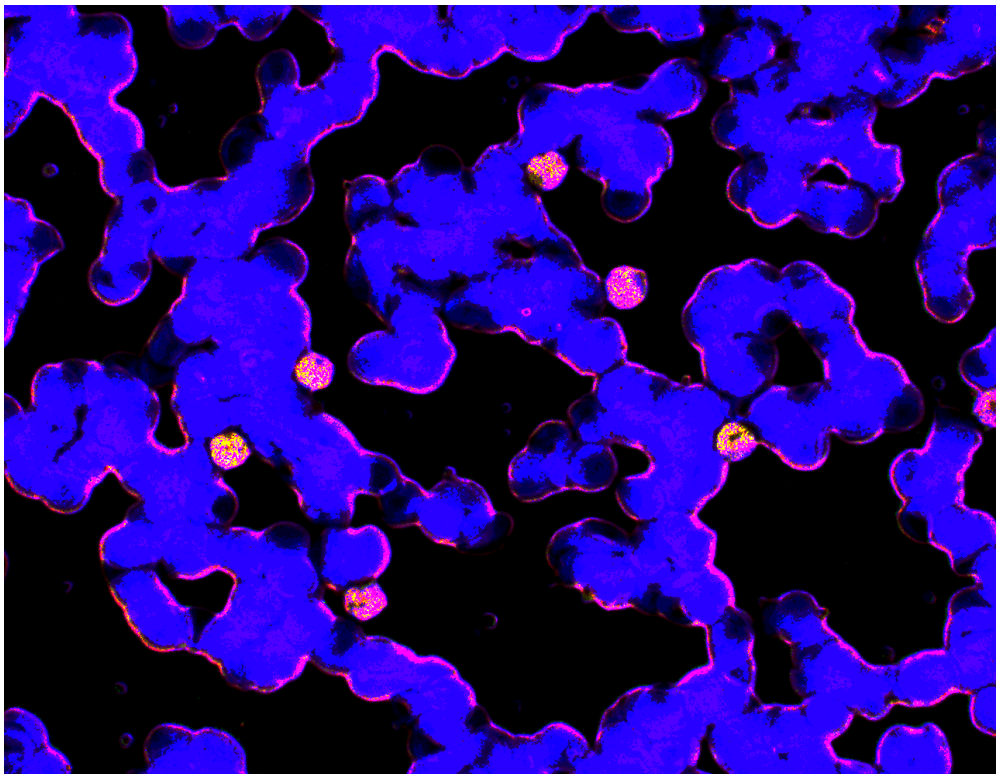
Comp3 Endmember map

Monochromatic Pseudo-maps

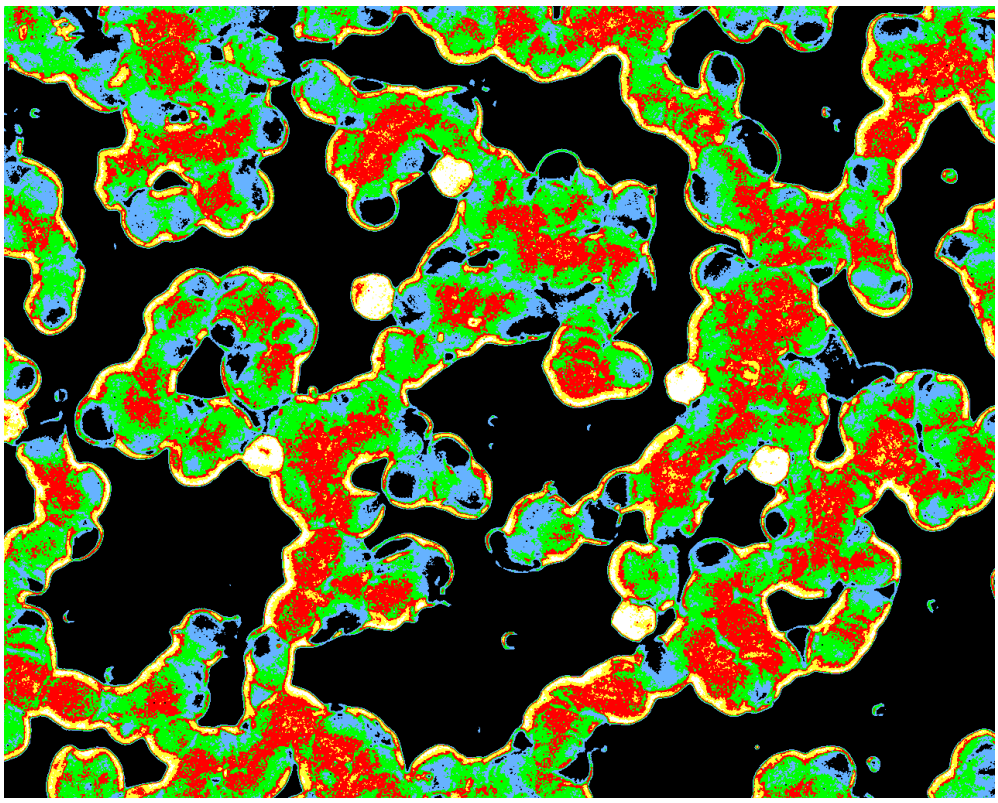
Example 2



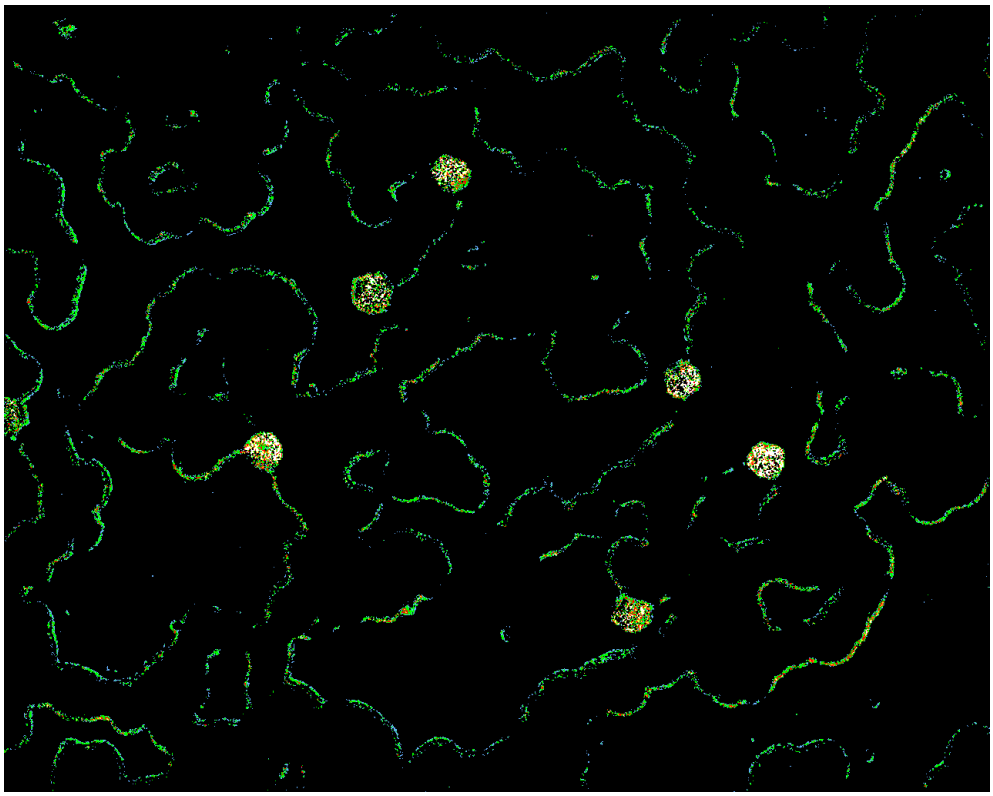
RGB Image



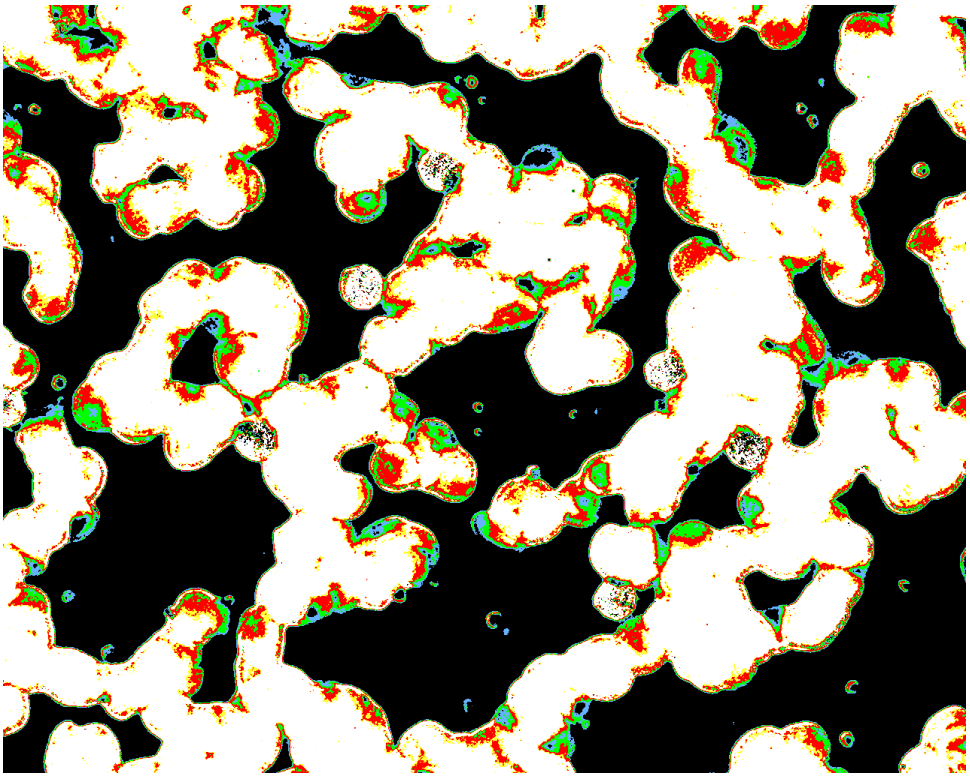
Merged Proportions Map



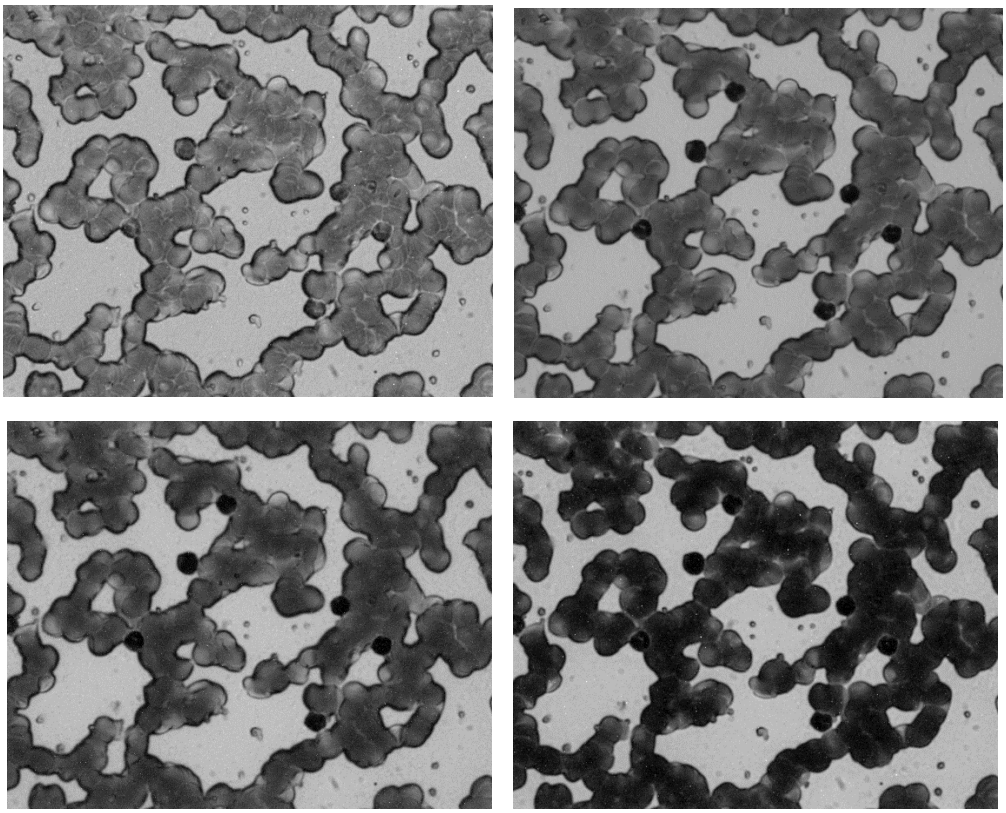
Comp1 Endmember Pseudo-map

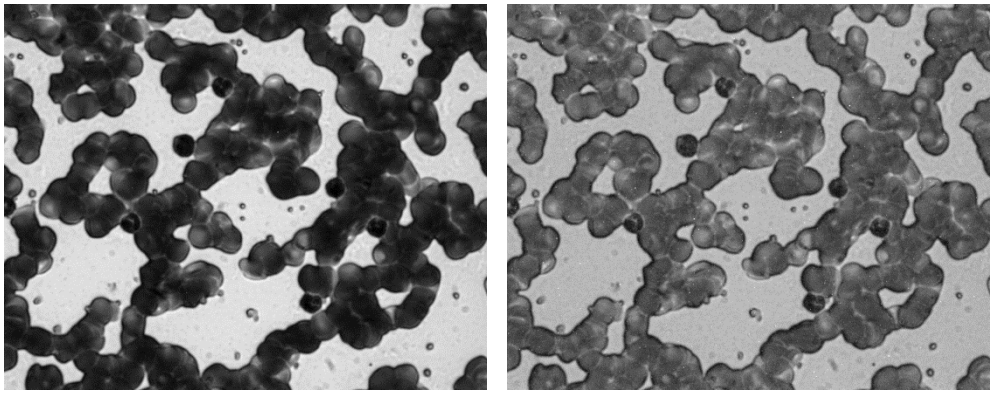


Comp2 Endmember Pseudo-map

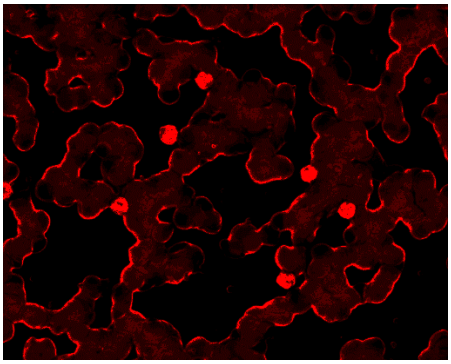


Comp3 Endmember Pseudo-map

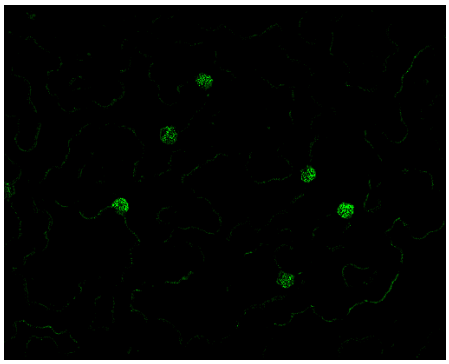




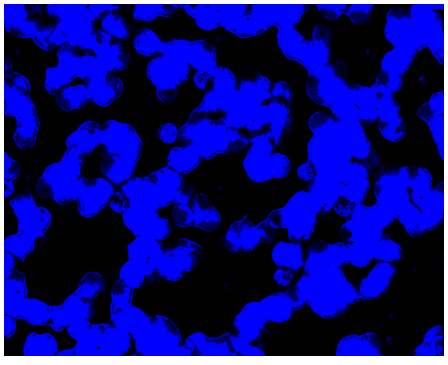
Spectral cube Images (465nm, 505nm, 530nm, 600nm, 640nm, 685nm respectively)



Comp1 Endmember map



Comp2 Endmember map



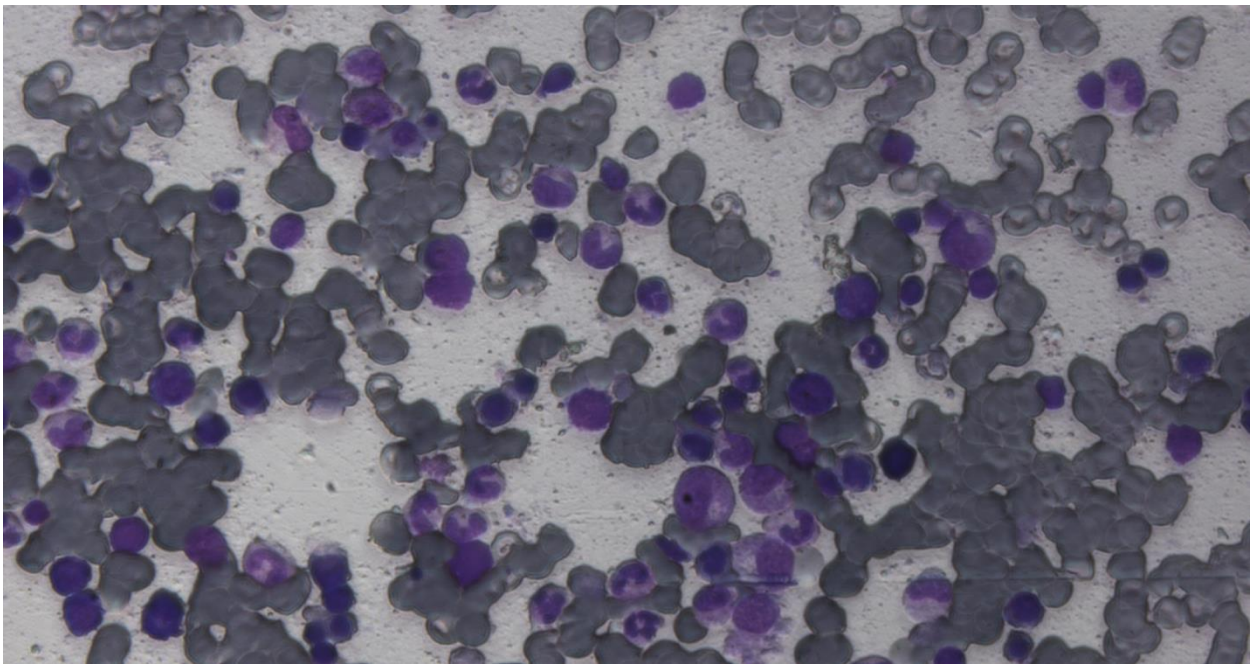
Comp3 Endmember map

Monochromatic Pseudo-maps

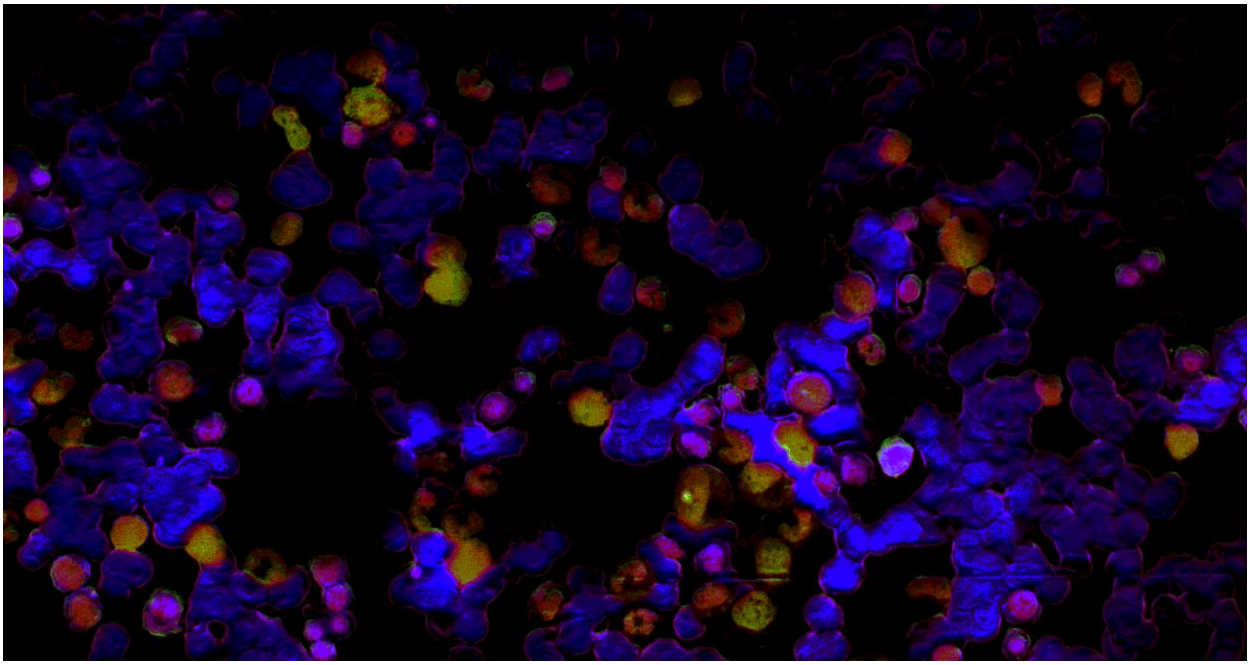
4.8.2 4096x2180 (Full HD) Maps (Pointgrey Flea 3.0)

The difference in quality of imaging and mapping is obvious as we have a great depiction of cell's detailed structure. This kind of resolution gives us the ability of digital zoom without "pixelization", characteristic that is going to be of great value when observing cells.

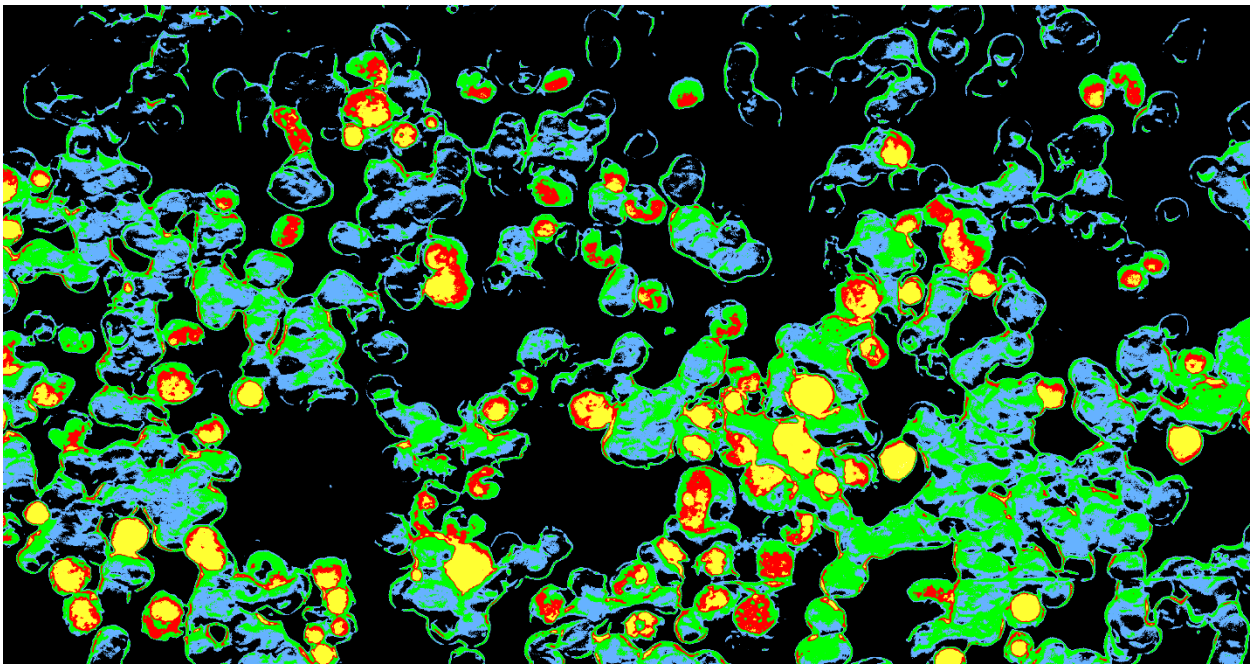
Example 1



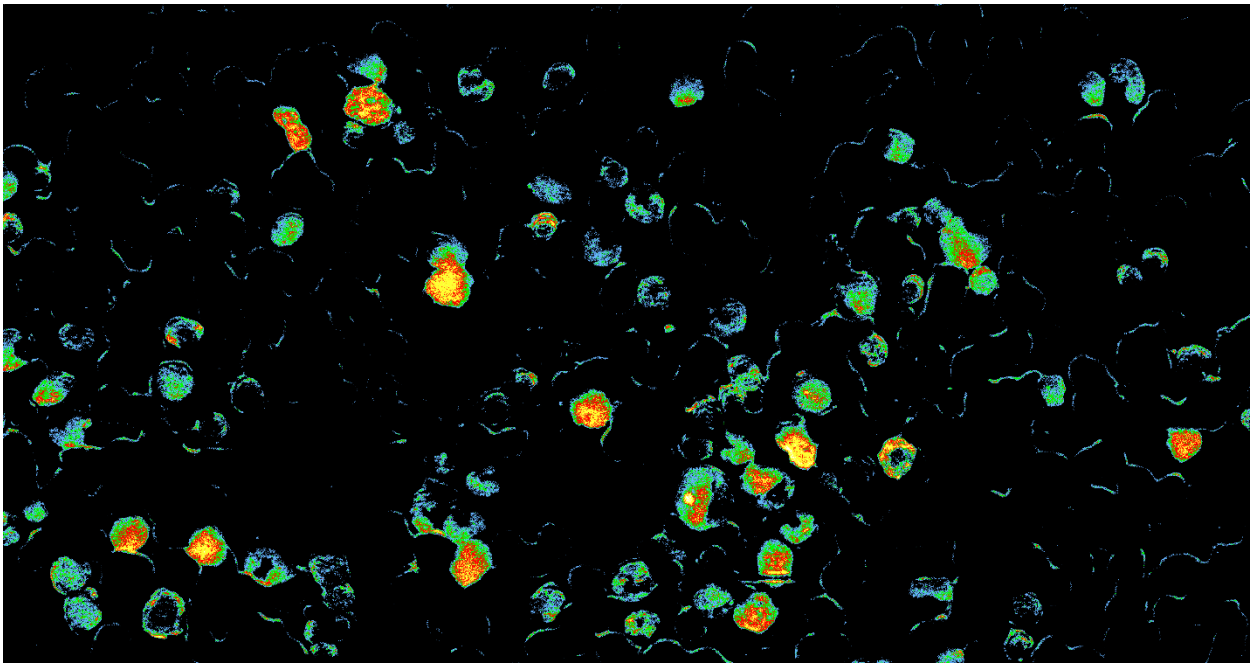
RGB Image



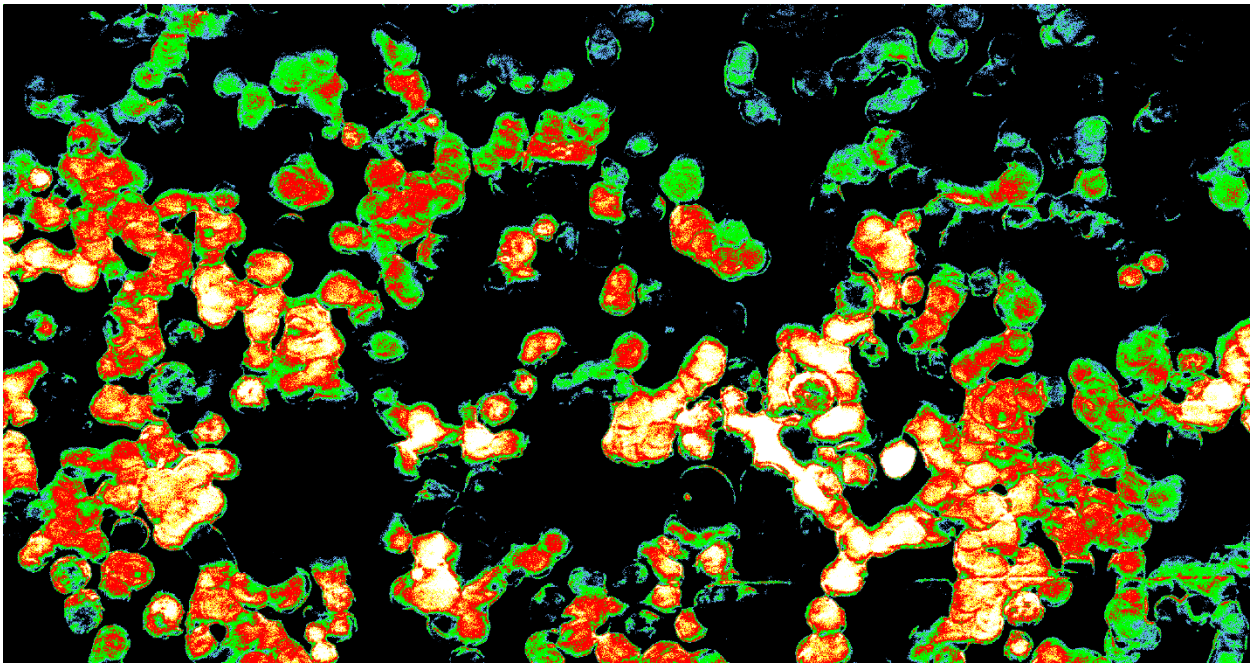
Merged Proportions Map



Comp1 Endmember Pseudo-map

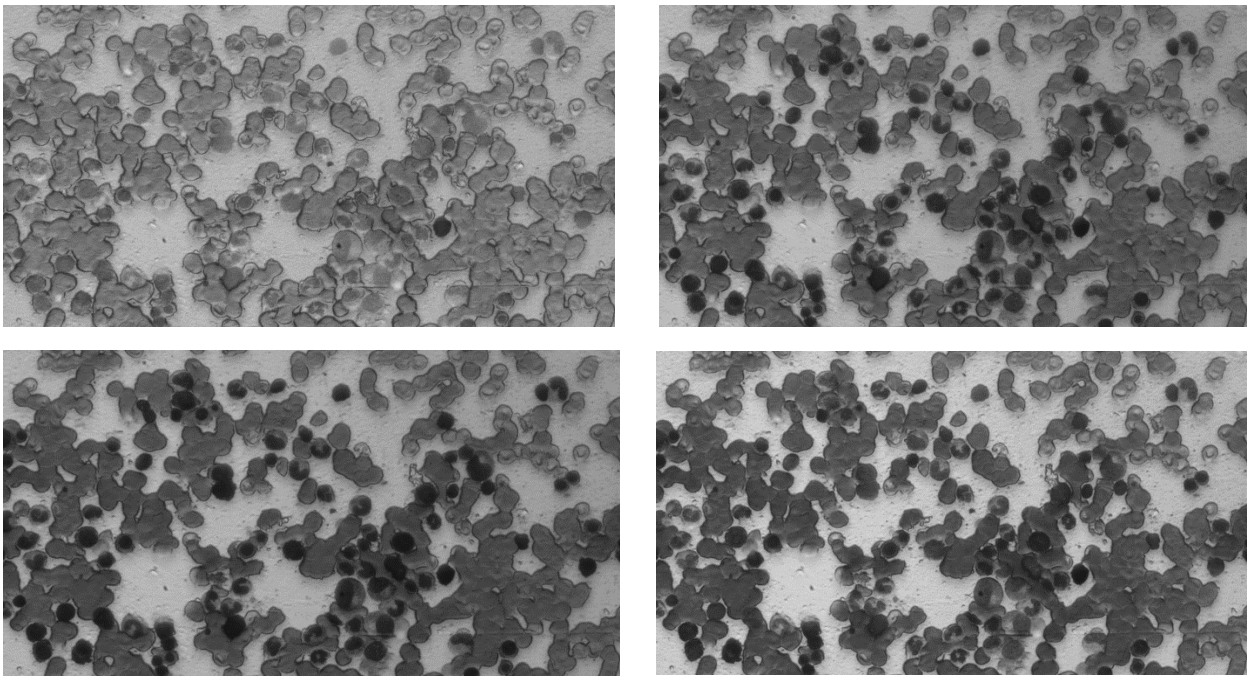


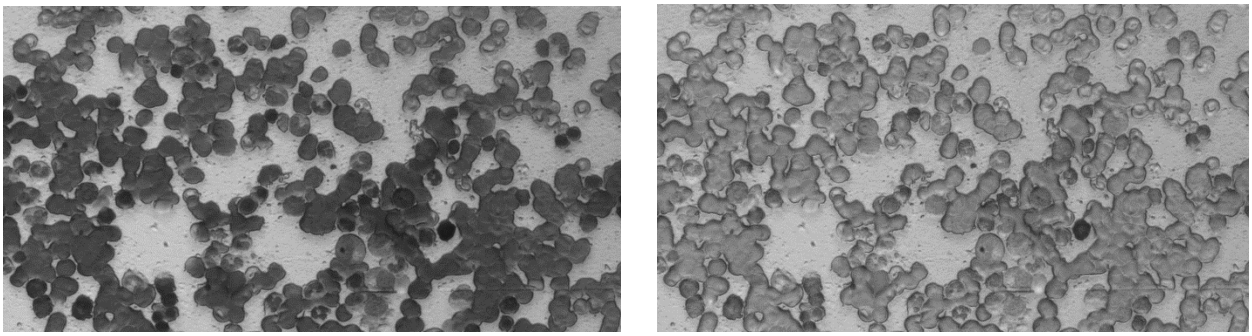
Comp2 Endmember Pseudo-map



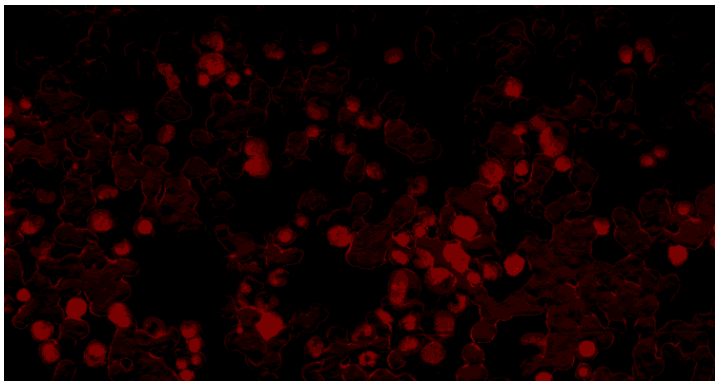
Comp3 Endmember Pseudo-map

Here, the spectral cube as well as the monochromatic endmember maps are presented for observation and comparisons.

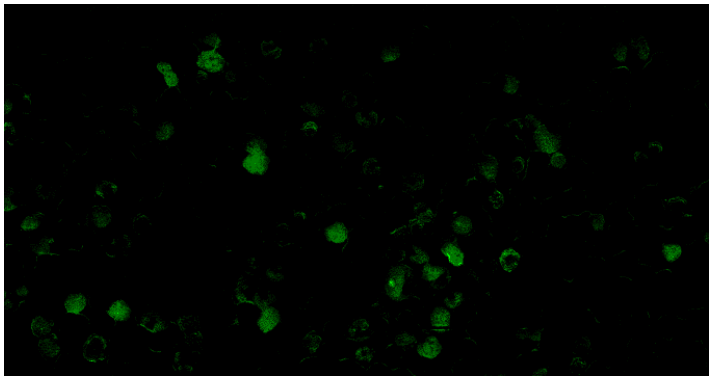




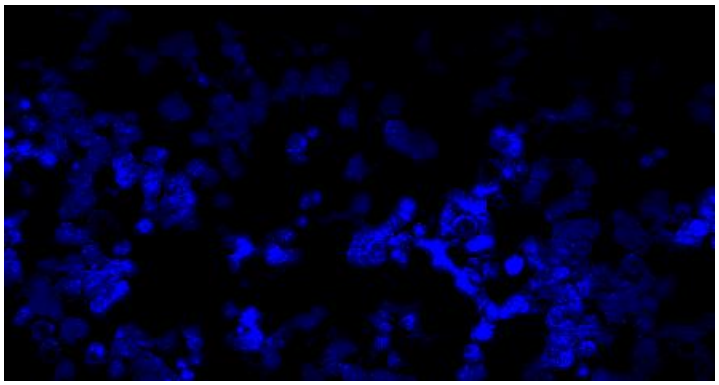
Spectral cube Images N.1 (465nm, 505nm, 530nm, 600nm, 640nm, 685nm respectively)



Comp1 Endmember map



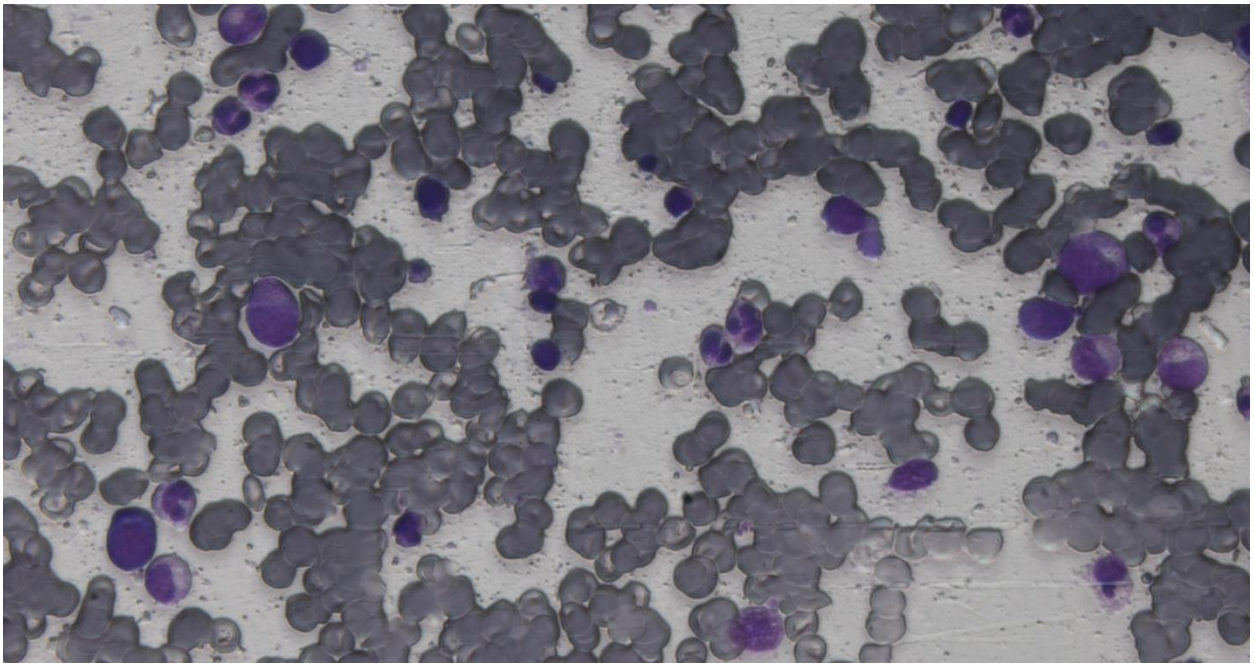
Comp2 Endmember map



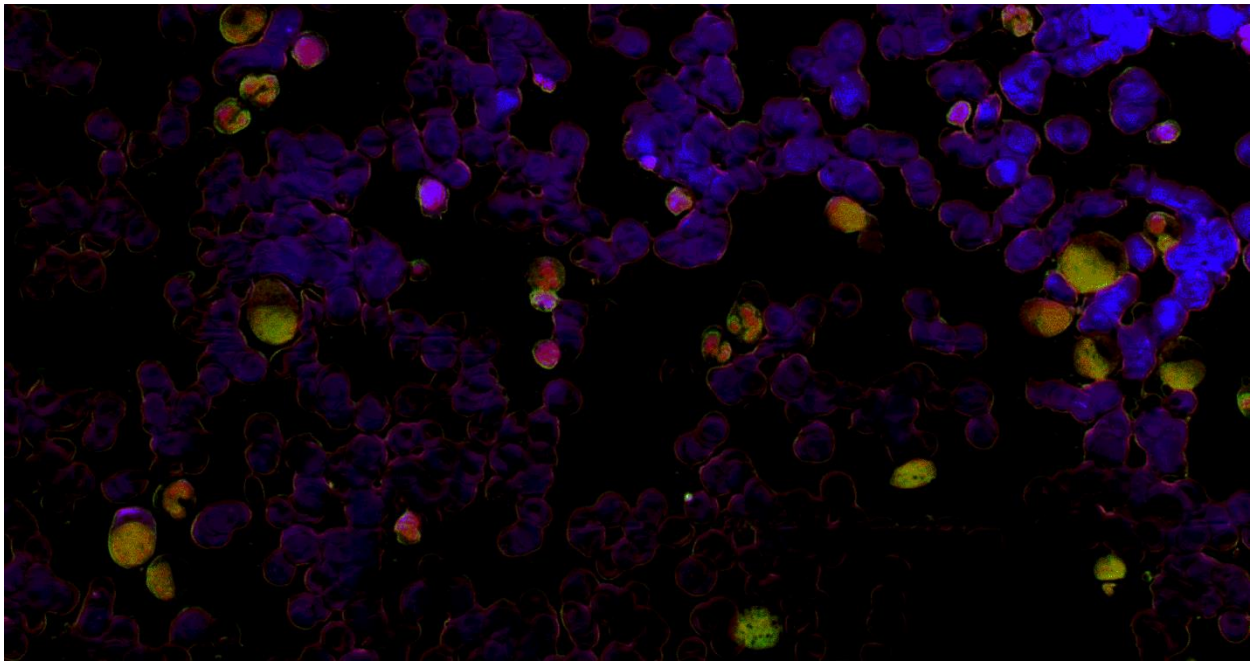
Comp3 Endmember map

Monochromatic Pseudo-maps

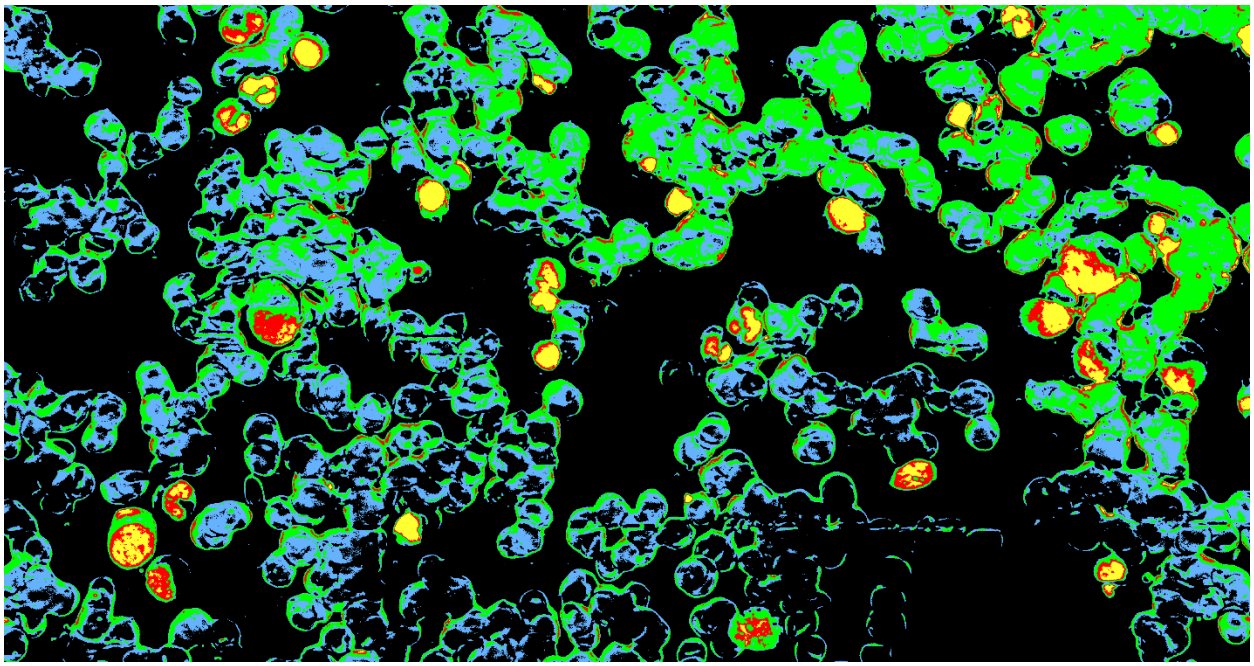
Example 2



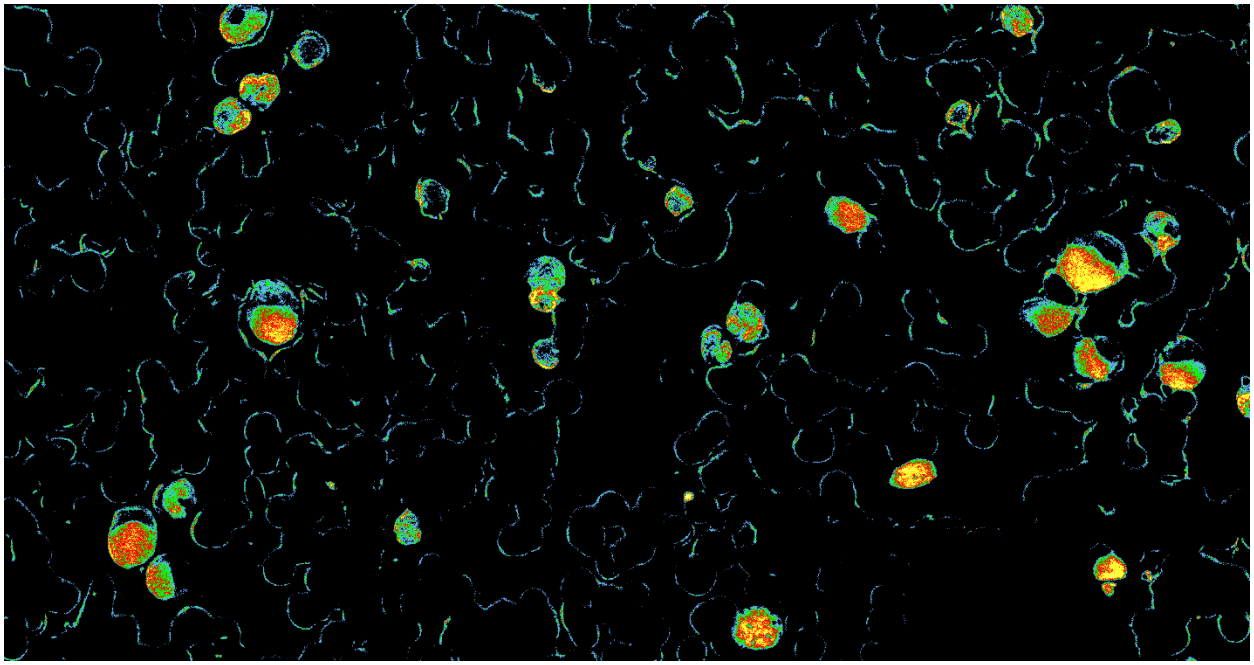
RGB Image



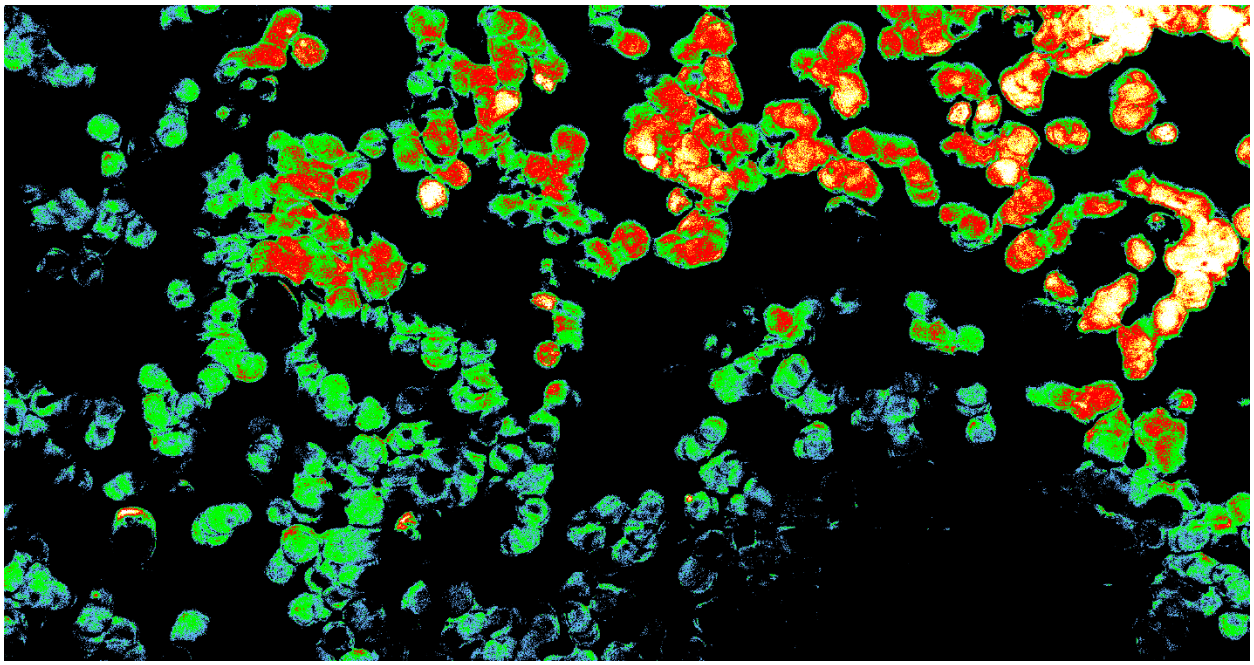
Merged Proportions Map



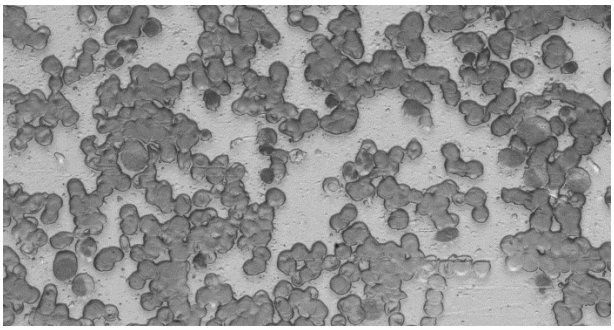
Comp1 Endmember Pseudo-map

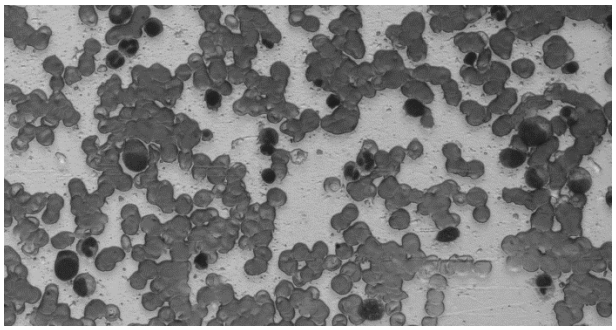


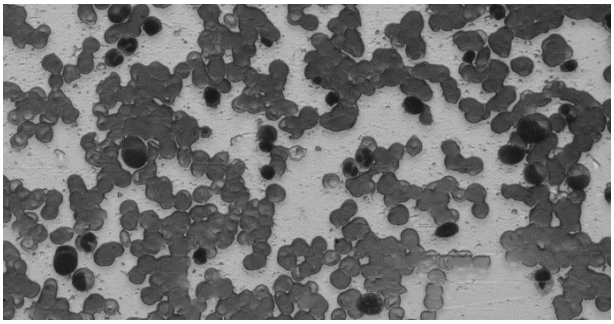
Comp2 Endmember Pseudo-map

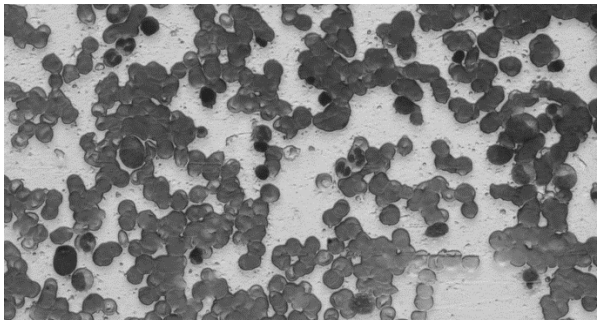


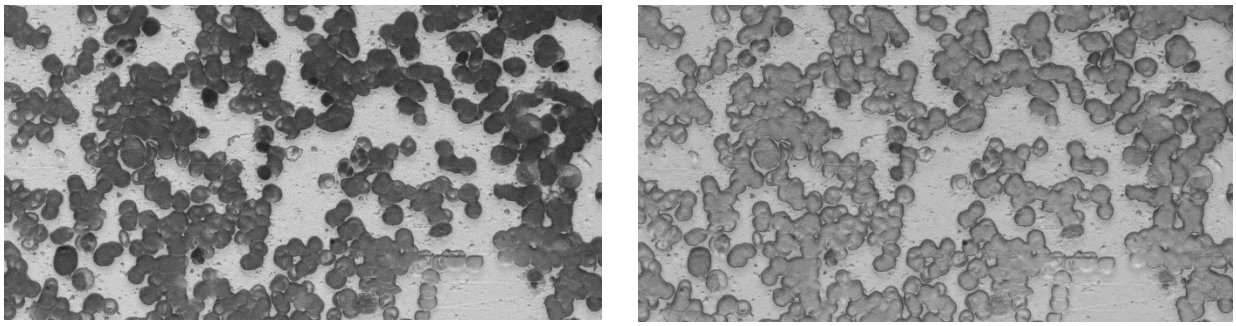
Comp3 Endmember Pseudo-map



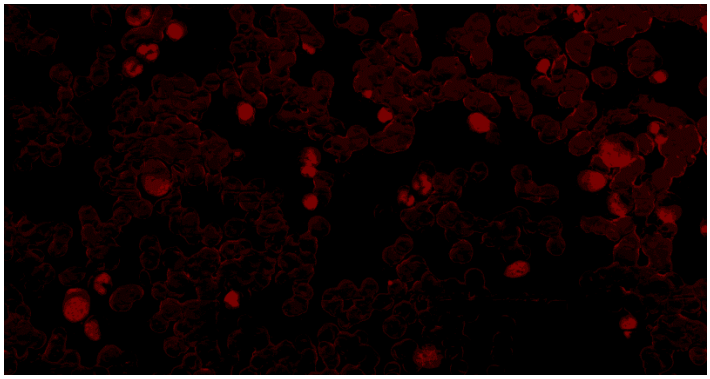




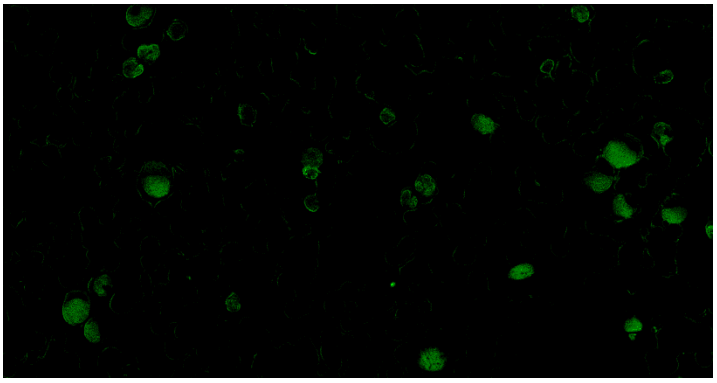




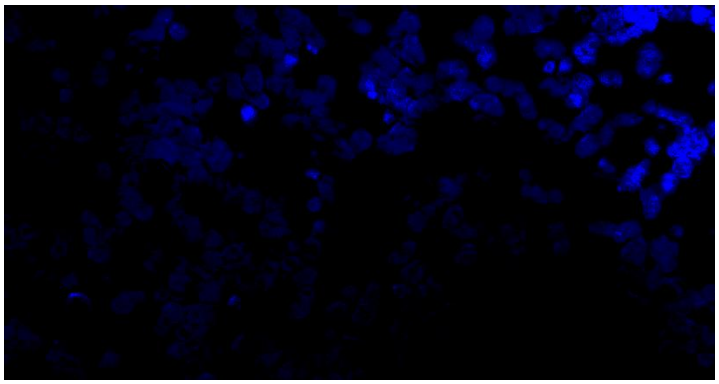
Spectral cube ImagesN.2 (465nm, 505nm, 530nm, 600nm, 640nm, 685nm respectively)



Comp1 Endmember map



Comp2 Endmember map



Comp3 Endmember map

Monochromatic Pseudo-maps

4.8.3 Comments on maps

In chapter 1.7 of this thesis, there is an analysis about the cells structure and staining procedure. Several facts can be extracted by analyzing those information. A general view on the parts that each substance stains on a cell can be acquired, by combining the subchapters 1.7.1 and 1.7.4. Then, these facts can be ascertained on the maps.

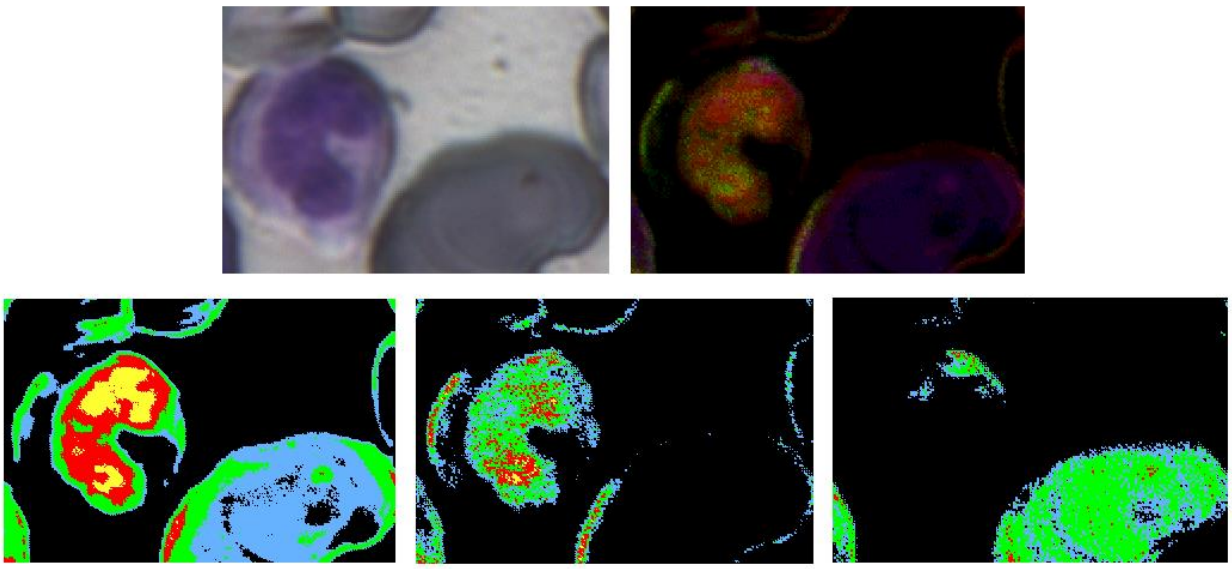
1. The nucleus of each cell, consisting mostly from chromatin, has to be stained by eosin, and partly (or filled) by one or both of the basic substances (Methylene Blue or Azure B) due to the excess presence of nucleic acids (DNA-RNA).
2. As cells become smaller (ex. Lymphocytes) we have to observe the Eosin Y concentration get increased, because we have more chromatin in less space. The same thing applies for the other two substances.
3. The presence of Methylene blue or Azure B concentration on nucleus, depends on the cell and on how carefully the staining procedure was accomplished. Some of the cells have increased concentration of nucleic acids (ex. Lymphocytes or Blasts) and others not (ex. Neutrophil cells).
4. On cytoplasm, we expect combinations the substances, depending on the cell. Cytoplasmic Proteins, Cytoplasmic RNA and Basophil Granules uptake the basic substances (Methylene Blue and Azure B), the same time that Reticulocytes, Positive-charged hemoglobin parts, Eosinophil granules uptake mostly the Eosin Y. Neutrophil granules uptake behavior is depended on their charge. Negative charged mostly uptake the basic substances while the positive charged mostly uptake the Eosin Y (or both with a mechanism that is not yet completely understood).

The aforementioned facts can be observed on the next examples (images), of specific cells. Depending on its type, the cell uptakes are different, resulting to different color and merged map.

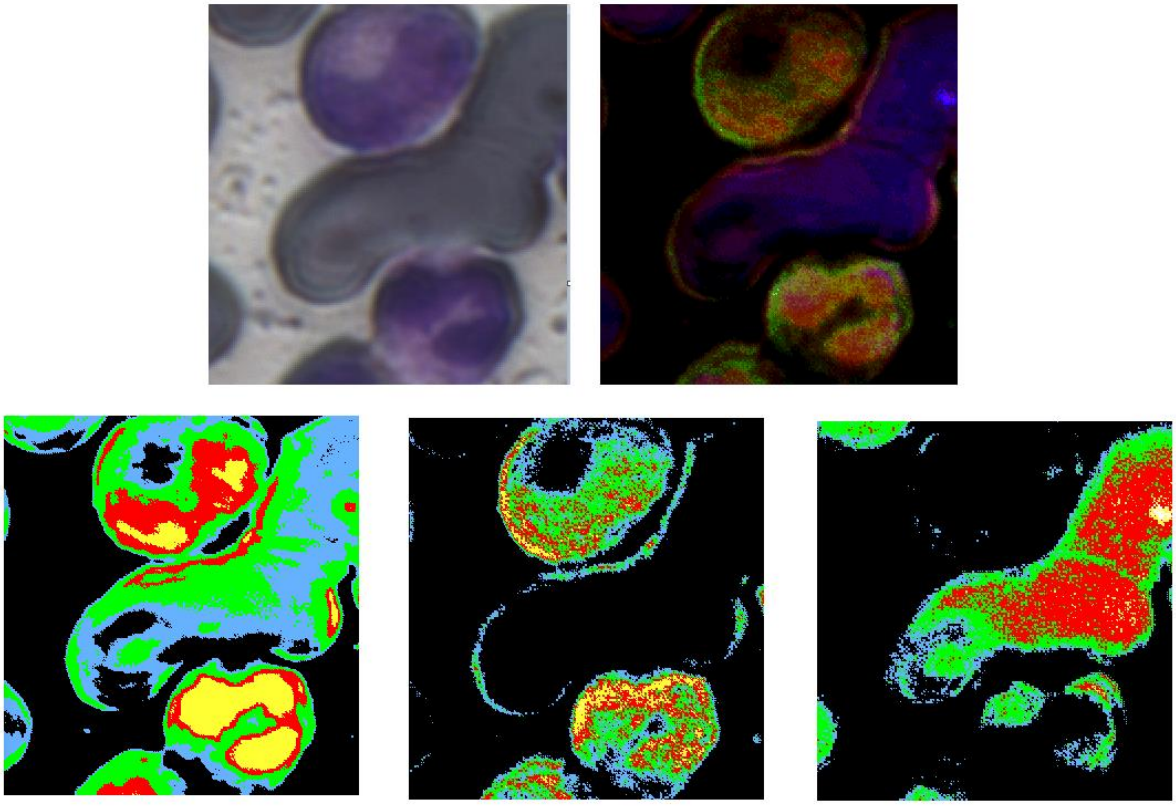
Examples on Regular Polymorphonuclear Neutrophils, Band Cells and eosinophils

We can observe:

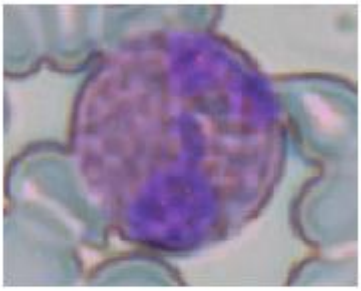
- Eosin Y (Chromatin counterstain) all over the nucleus in different proportions (Red scaled formations). Low concentrations on cytoplasm (in comparison with nucleus).
- Green and blue color artifacts n nucleus (Methylene Blue and Azure B, respectively) depicting the presence of nucleic acids (DNA or RNA). In some parts of the nucleus we have both substances, and the color becomes a mixture between Red, Green and Blue. On cytoplasm we have lower concentrations that on nucleus but partially visible on our maps.
- On the eosinophil cell (Example 3), the aforementioned facts regarding the nucleus are also valid, but on cytoplasm we can clearly observe the eosinophil granules, both in merged and Eosin Y pseudochromatic endmember map (Red shades on cytoplasm).

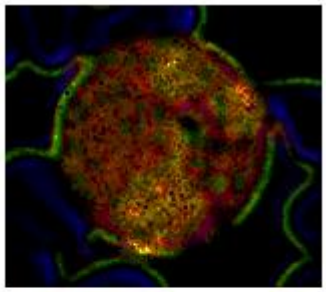


Example 1

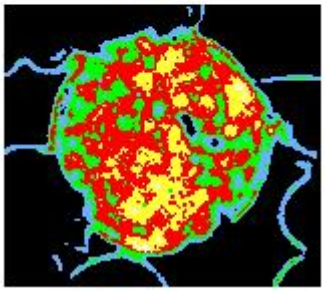


Example 2





Merged Map



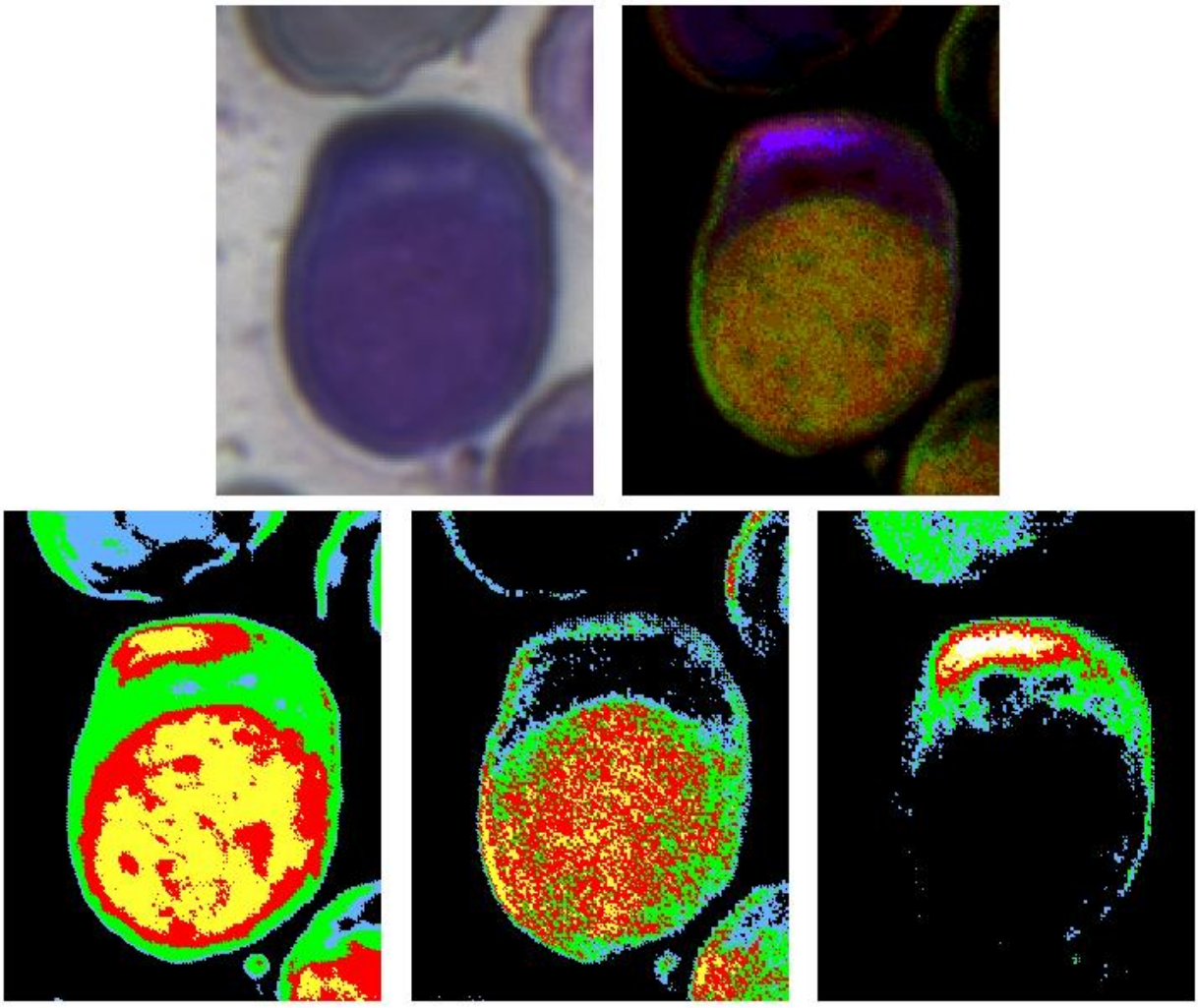
Eosin Y Endmember map

Example 3 (Eosinophil)

Examples on Leukemic Blasts

We can observe:

- Clear nucleus and cytoplasm, fact that helps us extract the Nucleus to Cytoplasm Ratio. This ratio is a clue when doctors try to identify whether a cell is regular or blast (Ratio increase→ Blast). (Example1.Merged)
- Eosin Y (chromatin counterstain, Example1.I) all over the nucleus. Great proportions of nucleic acids (DNA, RNA) on nucleus, overemphasized by green and blue artifacts on merged map (basic dyes concentration increased). (Example1.Merged, Example1.I-II)
- Cytoplasm stained by the basic substances more than a regular cell, indication of nucleic acids and basophil parts presence. (Example1.Merged, Example1.I-II)



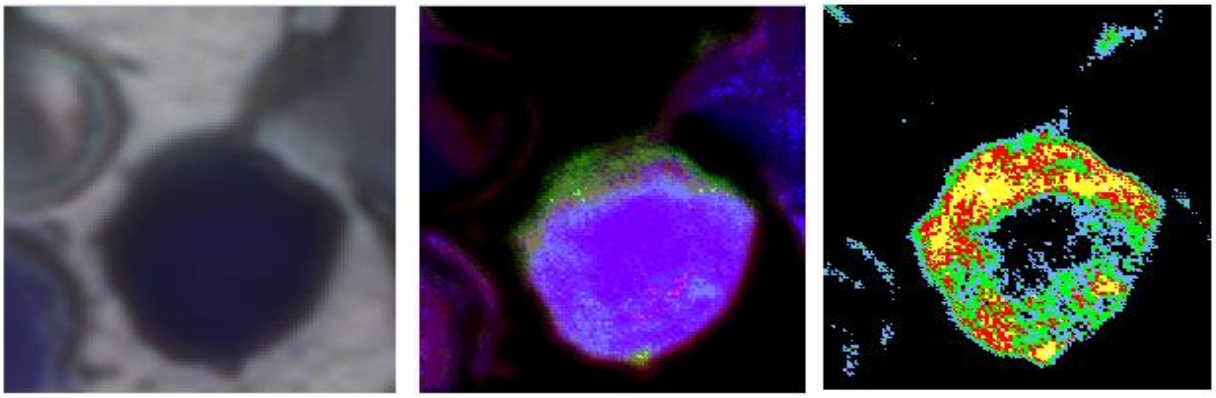
I II III

Example 1

Examples on Basophil cells

We can observe:

- On basophils the exact location of nucleus is undetectable. From bibliography we know that the nucleus of a basophil is a wheel-shaped formation. This can be validated on the image where we have the proportions of the basic substance Methylene Blue (Example1.III) which indicates the presence of nucleic acids (DNA, RNA).
- The enormous presence of three staining basic substances (on different proportions. Especially the basic ones) gives the basophile cell the unique merged color (Example1.II).



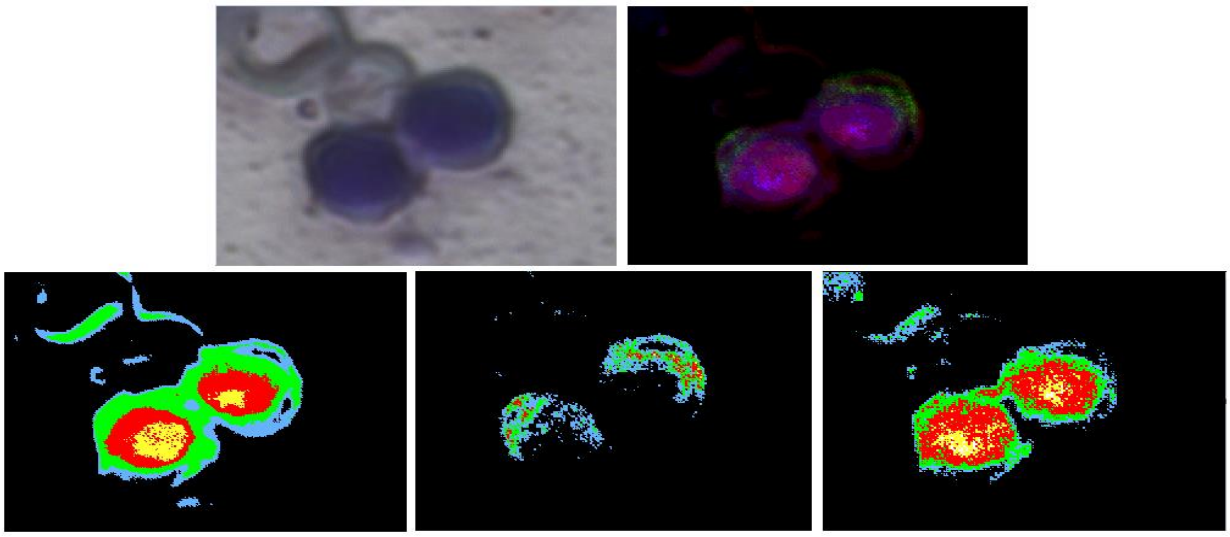
I II III

Example 1

Examples on Lymphocytes

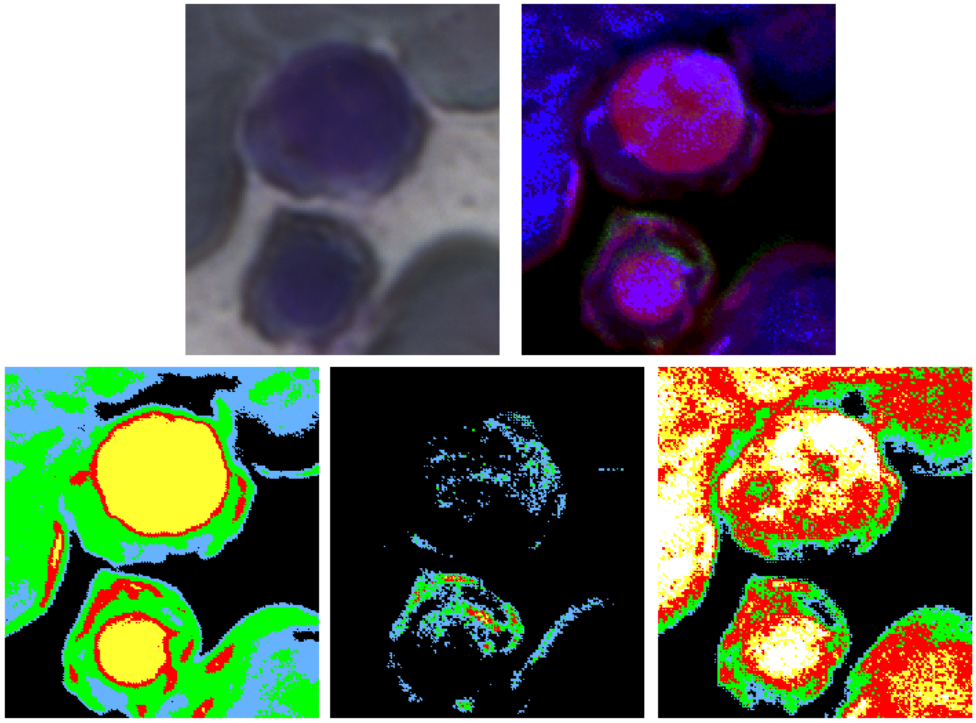
We can observe:

- Much chromatin in little space, gives us visible increased concentrations of Eosin (Clear chromatin space depiction). The same thing applies to nucleic acids concentration.
- The final color is -just like in the case of basophils but in decreased proportions- a combination of all three stains. Despite that fact, artifacts cause by chromatic differentiations, gives us clues about differences between lymphocytes structure and growth rate. As seen in the examples that follow, as well as in the maps on the previous chapter, the artificial color on our merged map is unique for lymphocytes and helps the identification by the examiner.



I II III

Example 1



I II III

Example 2

5. CONCLUSION

5.1 Conclusion, Potential and future work

Our experimental procedure resulted to a six band system, for concentrations prediction and mapping, that works directly on the microscopy tile. The novel system consists of one camera, a six-led tunable light source, and can be attached on every microscope and on almost every personal low power consumption computer. The use of one camera only, the low cost of Tunable light source and the use of personal computer, **widows our system a really cheap one.**

Aiming to improve leukemia diagnosis, the system is capable of helping the examiners identify and classify the exact kinds of cells that they see on the areas of interest on the sample of blood or bone marrow. The cells that appear almost identical using optical microscopy, have now clear differences in terms of staining substances concentration proportions. All cells are depicted in a consisted way, depending on their kind. For example lymphocytes are depicted with a bright purple color. This makes it easier for the doctors to be sure about the identification. **As the identification of cells is the main factor to an accurate diagnosis, using our system, gives the doctor a "second eye", leading to more accurate results and better treatment decisions.**

The whole atlas of Leukemia diagnosis can be changed, in order to include the new data (after calibrating the system with doctors and clinical trials). The proposed procedure is simple, clear and requires nothing more than the same tile that the doctor examines in optical microscopy. **The automation of cell identification procedure through well-trained clustering methods, could lead to a new step of diagnosis, between optical microscopy and flow cytometry**. Flow cytometry, as mentioned, is a time consuming and high-cost procedure that requires several blood and bone marrow samples from the patient. The completion and insertion of our method could save a lot of toil, and the examination costs could decrease dramatically. Also, analyzing our results **could lead to reduction of the chemical staining of samples with various substances, after the regular staining** that is conducted in certain cases during leukemia diagnosis and of course requires more samples from the patient and a lot more effort and working time from the doctors.

It worth mentioning that the May Grunwald – Giemsa substances deconvolution and prediction of concentrations is one of the most difficult ones, due to the similarity in the spectrum of Azure B and Methylene Blue. With simple changes, **the same procedure can be used in almost every set of staining substances with great prediction accuracy and in fluorescence microscopy**. **The system can be used this way on several diseases that require microscopy tile examination.**

Our system has great potential of becoming a Real Time system. The algorithm is quite fast even with High Definition data. It consumes less than one second for cube acquisition and the whole procedure. Using more efficient ways of acquiring the selected bands cube and predict the rest of the bands needed could reduce time consumption even more [35] [36]. Along with PLS algorithm prediction optimization, this may lead to a real time system.

6. REFERENCES

[1] Adamos A.: Αλγοριθμοι δεδομενων υπερφασματικης για την ανιχνευση, τμηματοποιηση και ταυτοποιησηση χαρακτηριστικων διαγνωστικης σημασιας, Diploma Thesis, Technical University of Crete, Greece (2011)

[2] Miramar College: UV-Visible absorption spectroscopy. Online at: http://faculty.sdmiramar.edu/fgarces/LabMatters/Instruments/UV_Vis/Cary50.htm

[3] Epitropou G.: Hyper-spectral imaging and spectral segmentation algorithms for the non-destructive analysis of el-greco's paintings, Diploma Thesis, Technical University of Crete, Greece (2008)

[4] Balas C., Papas C., Epitropou G., Handbook of biomedical optics (Chapter 7- Multi-Hyperspectral Imaging), CRC Press (2011)

[5] Nicholas M. Short, Electromagnetic Spectrum: Spectral Signatures. Online at: https://www.fas.org./irp/imint/docs/rst/Intro/Part2_5.html

[6] Hassan, K.W.: Novel regression methods for spectral data. Master thesis, Lappeen-

Ranta University of Technology, Lappeenranta (2012)

[7] Hibbert, D.B., Gooding, J.J.: Data analysis for chemistry. Oxford University Press

New York (2006)

[8] Krastev, T.: Chemometrics. Available online: http://classifion.sicyon.com/ References/Chemometrics.pdf

[9] James E. Burger: Hyperspectral NIR Image Analysis: Data Exploration, Correction and Regression. PhD thesis, Swedish University of Agricultural Sciences (2006)

[10] Abatzi, F.: Spectral Deconvolution and concentration mapping in complex biochemical stains, Diploma Thesis, Technical University of Crete, Greece (2014)

[11] Gemperlin, P.: Prctical Guide to chemometrics. CRC Press (2010)

[12] Statsoft Inc.: Partial Least Squares (PLS), Available online at: http://www.uta.edu/faculty/sawasthi/Statistics/stpls.html (Basic Ideas and Computational Approach)

[13] Jong, S. "SIMPLS: An Alternative Approach to Partial Least Squares Regression." *Chemometrics and Intelligent Laboratory Systems*. Vol. 18, 1993

[14] Chronolab, Cell Structure, Barcelona, Spain Online at: http://www.chronolab.com/articles/atlas-of-hematology/160-laboratory-hematology/laboratory-blood-analyses/labaratory-diagnostic-acces/morphologycal-examination-of-blood-smears

[15] News Medical: Biomarker - what is a biomarker? Available online: http://www.

news-medical.net/health/Biomarker-What-is-a-Biomarker.aspx

[16] Research Advocary Network: Biomarkers in Cancer Available online: http:

//researchadvocacy.org/images/uploads/downloads/BiomarkerinCancer_

WebDownloadVersion.pdf

[17] Microscopy Giemsa's solution – Product Brochure- Merck KGaA

[18] Giemsa's stain – Product Brochure (Procedure No. GS-10) - SIGMA-ALDRICH, INC

[19] Microscopy May-Grünwald's Solution – Product Brochure- Merck KGaA

[20] J.T. Baker, Product Information May-Grünwald Giemsa Brochure.

[21] Ioannis Georgoulis, Ergasthriakh Aimatologia, Rotonta (2012)

[22] Heckner F., Freund M.: Praktikum der mikroskopischen Hämatologie, Urban & Fischer (2001)

[23] R.W. Sabnis: HANDBOOK OF BIOLOGICAL DYES AND STAINS- SYNTHESIS AND INDUSTRIAL APPLICATIONS, Pfizer Inc. (2010)

[24] StainsFile.: Eosin Y Available online at: http://stainsfile.info/StainsFile/dyes/45380.htm

[25] StainsFile.: Methylene Blue Available online at: http://stainsfile.info/StainsFile/dyes/52015.htm

[26] StainsFile.: Azure B Available online at: http://stainsfile.info/StainsFile/dyes/52010.htm

[27] WebMD, Leukemia. Available online at: http://www.webmd.com/cancer/tc/leukemia-topic-overview#BM_Topic%20Overview

[28] Carl Zeiss Microscopy GmbH, Axio Scope.A1 Brochure, Available online at: http://www.zeiss.com/microscopy/en_de/products/light-microscopes/axio-scope-a1-for-biology.html

[29] XIMEA Inc., xiQ - USB3 Vision Cameras, Available online at: https://www.ximea.com/en/products/usb3-vision-cameras-xiq-line/mq013cg-on

[30] Point Grey Research Inc., Flea3 8.8 MP Color USB3 Vision (Sony IMX121), Available online at: https://www.ptgrey.com/flea3-88-mp-color-usb3-vision-sony-imx121-camera

[31] Rossos C.: Development of a computer controlled tunable wavelength light source from Ultraviolet to Infrared, Diploma Thesis, Technical University of Crete, Greece (2013)

[32] Ocean Optics Inc.: Introduction to spectroscopy in teaching labs using ocean optics spectrometers (2006)

[33] Mahasien H., Nabeel Z.: Effect of concentration Absorption and Fluorescence for Eosin Y in methanol, Al-Mustansiryiah University College of Science (2010)

[34] Nandini R.: A comparative study of metachromasy induced in Azure B by anionic polyelctrolytes, Science Journal of Chemistry (2013)

[35] Iliou D.: Spectral Prediction from Filtered Color CCD Cameras, Diploma Thesis, Technical University of Crete, Greece (2011)

[36] Iliou D.: Hyper spectral data estimation from power dimensionality experimental imaging, Master Thesis, Technical University of Crete, Greece (2014)