

Article

Spatial Modeling of Precipitation Based on Data-Driven Warping of Gaussian Processes

Vasiliki D. Agou ^{1,†}, Andrew Pavlides ^{1,†} and Dionissios T. Hristopoulos ^{2,*,†} 

¹ School of Mineral Resources Engineering, Technical University of Crete, 73100 Chania, Crete, Greece; vagou@isc.tuc.gr (V.D.A.); apavlidis@isc.tuc.gr (A.P.)

² School of Electrical and Computer Engineering, Technical University of Crete, 73100 Chania, Crete, Greece

* Correspondence: dchristopoulos@ece.tuc.gr; Tel.: +30-2821037688

† These authors contributed equally to this work.

Abstract: Modeling and forecasting spatiotemporal patterns of precipitation is crucial for managing water resources and mitigating water-related hazards. Globally valid spatiotemporal models of precipitation are not available. This is due to the intermittent nature, non-Gaussian distribution, and complex geographical dependence of precipitation processes. Herein we propose a data-driven model of precipitation amount which employs a novel, data-driven (non-parametric) implementation of warped Gaussian processes. We investigate the proposed warped Gaussian process regression (wGPR) using (i) a synthetic test function contaminated with non-Gaussian noise and (ii) a reanalysis dataset of monthly precipitation from the Mediterranean island of Crete. Cross-validation analysis is used to establish the advantages of non-parametric warping for the interpolation of incomplete data. We conclude that wGPR equipped with the proposed data-driven warping provides enhanced flexibility and—at least for the cases studied—improved predictive accuracy for non-Gaussian data.

Keywords: non-Gaussian data; skewed distributions; Gaussian anamorphosis; reanalysis data; kriging; warped Gaussian processes

PACS: 02.50.Fz; 02.60.Ed; 89.60.-k; 92.60.Ry; 05.10.Ln

MSC: 60G15; 60G60; 62F40; 62H11; 62G05; 65C05



Citation: Agou, V.D.; Pavlides, A.; Hristopoulos, D.T. Spatial Modeling of Precipitation based on Data-Driven Warping of Gaussian Processes. *Entropy* **2022**, *24*, 321. <https://doi.org/10.3390/e24030321>

Academic Editor: Antonio M. Scarfone

Received: 10 January 2022
Accepted: 20 February 2022
Published: 23 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Climate change combined with changes in land use is causing increased frequencies of drought and flooding events in many parts of the world. In 2021, extreme rainfall hit the Henan Province of China in July, Western Europe suffered severe flooding in mid-July, while extreme rainfall and flooding also affected the northern Amazon basin in South America and several parts of Africa. At the same time, prolonged droughts plagued several parts of the world. Such adverse impacts have been anticipated by scientists [1]. Climate and land-use changes affect ecosystems and human societies globally. One of the main concerns is their impact on the availability of water resources. It is thus important to better understand and forecast the spatial and temporal patterns of precipitation since these patterns affect the hydrological cycle and are crucial for the sustainability of human life on the planet.

Expert estimates—including in the Sixth Assessment Report of the Intergovernmental Panel on Climate Change—indicate an increase in the global averaged precipitation since 1950 [2]. In certain areas, both the frequency and intensity of heavy precipitation events have increased. In the Mediterranean region, on the other hand, it is expected that summer precipitation will decrease, thus increasing the risk of drought and aridification. In addition, there is evidence that the number of precipitation events has decreased while the intensity per event has increased [3]. Heavy precipitation events are expected to increase in several

Mediterranean countries leading to increased flooding risks [4]. The impact of climate change on societies worldwide has renewed interest in quantitative methodologies that can estimate spatiotemporal climate and weather patterns [5]. Certain areas, including the Mediterranean basin, have been characterized as “climate change hot spots” [6]. Especially in such regions the interplay of climate change and changes in land use is crucial for water resource availability [4].

Accurate spatial models of precipitation are difficult to formulate, due to the variability and intermittent nature of precipitation across different temporal and spatial scales. In the Mediterranean region, the two large water bodies (the Atlantic Ocean and the Mediterranean Sea), as well as the major European mountain ranges are considered the main causes of extreme precipitation [7,8]. The total amount of precipitation received by an area over a specific time window is often modeled by means of parametric, non-Gaussian, probability distributions. If the temporal dimension is taken into account, modeling is further complicated due to the strong seasonal variability and intermittence of precipitation. Furthermore, despite significant progress over the last 30 years, the modeling of interactions between spatial and temporal correlations is an open research topic [9–11].

The spatial patterns of precipitation are calculated by means of stochastic spatial interpolation methods known as Kriging [12–16]. Kriging has been successfully used in environmental, meteorological, and hydrological studies to generate spatial maps based on partial data [17–21]. The predictive equations used by kriging also appear in the framework of Gaussian process regression (GPR) [22,23]. Kriging and GPR methods are based on the assumption of an underlying joint normal (Gaussian) distribution. However, the observed probability distributions of precipitation (as well as other environmental variables) are typically skewed (non-Gaussian) [24,25]. In geostatistical literature, non-Gaussian distributions are treated using nonlinear transformations that restore normality of the marginal distribution in a latent space; for reviews of such transformations see [14,16,23]. The application of nonlinear transformations to achieve normality is known as “Gaussian anamorphosis”. The spatial analysis is carried out in the latent space using the transformed data. Predictions in the observation space are derived by inverse transforming the predictions in the latent space. A similar “warping” approach has been applied to Gaussian process regression [26]. The term “warping” herein refers to the nonlinear transformation of the Gaussian process.

This study has two main objectives. First, we introduce a new, non-parametric (data-driven) warping approach. The warping transformation employs the kernel-based estimate of the cumulative distribution function (CDF) recently presented in [27]. This non-parametric method provides better estimates of the CDF of skewed probability distributions than other commonly used kernel-based methods. Secondly, we show that non-parametric warped Gaussian process regression (wGPR) can be used to model the spatial distribution of non-Gaussian variables such as precipitation. We focus on precipitation amounts because the respective probability distributions vary significantly in time, and the shape of the respective CDFs is not satisfactorily captured by parametric expressions. In order to assess the performance of *non-parametric wGPR*, we compare the results of cross-validation analysis with those obtained by non-warped GPR (i.e., Kriging). We apply cross-validation to a simulated noisy test function used in [26] as well as Leave-One-Out Cross-Validation (LOO-CV) to a reanalysis precipitation dataset from the island of Crete. The spatial dependence in the case of the test function is expressed in terms of a covariance model which corresponds to a linear, damped harmonic oscillator driven by white noise [23]. The enhanced interpolation accuracy provided by this model, which is not well-known in the machine learning and geostatistical literature, motivates its further use in Gaussian process regression.

The remainder of this paper is structured as follows: Section 2 presents the proposed wGPR methodology which involves Gaussian anamorphosis using the kernel-based CDF, spatial interpolation (prediction) of the normalized process employing standard GPR, and generation of the predictive distribution of precipitation by inverting the warping transformation. Section 3 presents an application of wGPR to a one-dimensional (1D)

synthetic dataset which helps to illustrate the method. Section 4 compares GPR and wGPR using an ERA5 precipitation reanalysis dataset. Lastly, Section 5 presents discussion and conclusions and suggests future directions of research.

2. Methodology

Models of space-time processes rely on the spatial and temporal correlations inherent in the data. Most of the commonly used geostatistical approaches as well as GPR are developed to optimally perform Gaussian processes. However, non-Gaussian data are often encountered in nature. For example, precipitation amounts do not follow the Gaussian distribution. Typical models used include the exponential [28], gamma [29–31], lognormal [32–34], Weibull [28], generalized extreme value (GEV) [35–40], and Pareto distributions [41]. The optimal model depends on the geographical location, the climate zone, as well as the analyzed temporal and spatial scale [3,21,27]. It is thus necessary to relax Gaussian assumptions when modeling such data.

The wGPR approach proposed herein tackles non-Gaussian distributions using non-parametric warping of the observation space. The wGPR method involves the following steps: (i) Transformation of the data into standard normal values (normal scores) using a kernel-based warping function to conduct Gaussian anamorphosis; (ii) GP model specification based on variogram estimation; (iii) GPR using the normal scores; (iv) inversion of the warping transform to obtain predictions of precipitation values; and (v) calculation of cross-validation metrics for the assessment of predictive performance. These steps are described in more detail below.

2.1. Introduction to Gaussian Process Regression

A Gaussian process (GP) defines a prior distribution over functions which can then be used for Bayesian regression [22]. Herein we consider Gaussian processes having as input space the geographical coordinates $\mathbf{s} \in D \subset \mathbb{R}^d$ ($d = 1, 2$) of the domain of interest D . More generally, the input space can be extended to include the altitude and time as well as other potentially relevant for precipitation topographic parameters. Since a Gaussian process is fully determined by its mean and covariance kernel, we will denote the GP $z(\mathbf{s})$ by means of $z(\mathbf{s}) = \mathcal{GP}(m(\mathbf{s}); C_0(\mathbf{s}, \mathbf{s}'))$, where $m(\mathbf{s}) : \mathbb{R}^d \rightarrow \mathbb{R}$ is the mean function (expectation) and $C_0(\mathbf{s}, \mathbf{s}') : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ is the covariance (kernel) function; the latter is a non-negative definite function. The mean function and the covariance kernel are determined from a set of *hyperparameters*. In geostatistical parlance, a Gaussian process whose input space is restricted to spatial location is a Gaussian random field [23].

Measurements of the process typically include a noise term $\epsilon(\mathbf{s})$ which represents a collection of independent identically (normally) distributed random variables with zero mean and constant variance σ_ϵ^2 . Then, the observed process is given by

$$x(\mathbf{s}) = z(\mathbf{s}) + \epsilon(\mathbf{s}), \text{ where } \epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2), \quad (1)$$

$\mathcal{N}(\cdot, \cdot)$ denotes the normal probability model, and the symbol \sim implies that the probability distribution of the variable preceding \sim follows the probability law specified after \sim . In the following, $\Phi(\cdot)$ denotes the CDF of the standard normal distribution. The noise variance σ_ϵ^2 is known in geostatistics as the nugget term [16].

Assuming that measurements are available at N points $\{\mathbf{s}_i\}_{i=1}^N$, where $\mathbf{s}_i \in D$, for all $i = 1, \dots, N$, the joint PDF of the data is given by

$$\mathbf{x} := [x(\mathbf{s}_1), \dots, x(\mathbf{s}_N)]^\top \sim \mathcal{N}(\mathbf{m}_s, \mathbf{C}_s), \quad (2)$$

where \top denotes the vector (matrix) transpose, $\mathbf{m}_s = [m(\mathbf{s}_1), \dots, m(\mathbf{s}_N)]^\top$ the vector of expected values, and $[C_s]_{i,j} = [C_0]_{i,j} + \sigma_\epsilon^2 \delta_{i,j}$, for $i, j = 1, \dots, N$ the elements of the covariance matrix.

Let us consider a set of points (in general this set is disjoint from the sampling set), $\{\mathbf{s}_1^*, \dots, \mathbf{s}_p^*\}$ where the unknown values $z(\mathbf{s}_p^*)$, $p = 1, \dots, P$ of the process should be estimated. The joint PDF between the data and the predictions is given by

$$[\mathbf{x}, z(\mathbf{s}_1^*), \dots, z(\mathbf{s}_p^*)]^\top \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{m}_s \\ \mathbf{m}_* \end{bmatrix}, \begin{bmatrix} \mathbf{C}_s & \mathbf{C}_{s,*} \\ \mathbf{C}_{*,s} & \mathbf{C}_{*,*} \end{bmatrix}\right), \quad (3)$$

where $\mathbf{m}_* = [m(\mathbf{s}_1^*), \dots, m(\mathbf{s}_p^*)]^\top$ is the vector of mean values, $[C_{s,*}]_{i,j} = C_0(\mathbf{s}_i, \mathbf{s}_j^*)$ are the elements of the $N \times P$ covariance matrix between sampling and prediction points, $\mathbf{C}_{*,s} = \mathbf{C}_{s,*}^\top$ is the $P \times N$ covariance matrix between the prediction and sampling points, and $[C_{*,*}]_{i,j} = C_0(\mathbf{s}_i^*, \mathbf{s}_j^*)$ are the elements of the $P \times P$ covariance matrix between all pairs of the prediction points.

Since the joint PDF of the Gaussian process is normal, it is straightforward to obtain the conditional (on the data) PDF of the vector $[x(\mathbf{s}_1), \dots, x(\mathbf{s}_N)]^\top$ at the prediction points. The conditional PDF is also jointly normal with mean and covariance given respectively by the posterior mean, \mathbf{m}_{post} and covariance, \mathbf{C}_{post} , i.e.,

$$[z(\mathbf{s}_1^*), \dots, z(\mathbf{s}_p^*)]^\top \mid \text{Data} = \mathcal{N}(\mathbf{m}_{\text{post}}, \mathbf{C}_{\text{post}}), \quad (4)$$

where the posterior mean is given by

$$\mathbf{m}_{\text{post}} = \mathbf{m}_* + \mathbf{C}_{*,s} \mathbf{C}_s^{-1} (\mathbf{x} - \mathbf{m}_s), \quad (5)$$

and the posterior covariance by means of

$$\mathbf{C}_{\text{post}} = \mathbf{C}_{*,*} - \mathbf{C}_{*,s} \mathbf{C}_s^{-1} \mathbf{C}_{s,*}. \quad (6)$$

The GPR Equations (4)–(6) look identical to those of simple kriging [16]. However, in simple kriging, the mean is assumed constant and known, while in GPR the mean can comprise a superposition of basis functions with unknown coefficients (hyperparameters) which are estimated by maximizing the likelihood of the model [22]. Then, GPR is equivalent to universal kriging. More information regarding the relation between kriging and GPR can be found in [22,23]. Model selection in Gaussian processes is based on methods like Bayesian inference and cross-validation. In the former case, a prior function that captures *a priori* beliefs regarding the values of the hyperparameters is used.

Gaussian processes assume that the data follow the multivariate Gaussian distribution and that the observation noise is also Gaussian. The Gaussian assumption simplifies the calculations and leads to explicit predictive expressions. The optimal GPR prediction at a point \mathbf{s}_p^* is given by the conditional mean, $\hat{z}(\mathbf{s}_p^*) = m_{\text{post}}(\mathbf{s}_p^*)$, while the uncertainty is determined by the conditional standard deviation $\hat{\sigma}(\mathbf{s}_p^*) = \sqrt{C_{\text{post}}(\mathbf{s}_p^*)}$. The prediction interval at confidence level $(1 - \alpha) \times 100\%$ for $0 < \alpha < 1$ is given by

$$\left[\hat{z}(\mathbf{s}_p^*) - z_{\alpha/2} \hat{\sigma}(\mathbf{s}_p^*), \quad \hat{z}(\mathbf{s}_p^*) + z_{1-\alpha/2} \hat{\sigma}(\mathbf{s}_p^*) \right], \quad (7)$$

where $z_{\alpha/2} = \Phi^{-1}(\alpha/2)$ and $z_{1-\alpha/2} = \Phi^{-1}(1 - \alpha/2)$ are, respectively, the $(\alpha/2) \times 100\%$ and $(1 - \alpha/2) \times 100\%$ quantiles of the standard normal distribution.

2.2. Warping (Gaussian Anamorphosis) for Non-Gaussian Distributions

If the data follow a skewed probability distribution or exhibit heteroskedasticity (dependence of the variance on the spatial location), the assumption of normality may be inadequate. The standard approach for handling non-Gaussianity applies nonlinear normalizing transformations that restore marginal normality in a latent space. In geostatistical literature, this procedure is known as *Gaussian anamorphosis (GA)*. In the Gaussian process

framework, *warped GPs* incorporate nonlinear “warping” transforms of the observation space [26].

Hence, “warping” applies a nonlinear transformation to a non-Gaussian GP $x(\mathbf{s})$, leading to a *latent Gaussian process* $y(\mathbf{s})$. More precisely, warping is defined as a *monotonic mapping* $g : x \mapsto y$ such that $y(\mathbf{s}) = g[x(\mathbf{s})]$ has Gaussian marginal distribution with zero mean and unit variance [16,23]. Then, $y(\mathbf{s})$ can be modeled as a Gaussian process—under the bold assumption that not only the marginal but also the joint PDF of any vector $\mathbf{y} \in \mathbb{R}^N$ is Gaussian.

If $F_x(x)$ is the marginal CDF of $x(s)$, the *warping transform* $g(\cdot)$ is defined by means of $g : x \mapsto y$ such that

$$y = \Phi^{-1}[F_x(x)] = g(x), \tag{8}$$

The inverse transformation from the latent to the observation space is given by the *inverse warping* function $\tilde{g} = g^{-1}(\cdot)$. The latter is defined by means of $\tilde{g} : y \mapsto x$ so that

$$x = F_x^{-1}[\Phi(y)] = \tilde{g}(y) \tag{9}$$

is given by the monotonic mapping from the latent Gaussian variable y to the observation variable x [14,23]. Often, closed-form expressions can be obtained for the functions $g(\cdot)$ and $\tilde{g}(\cdot)$. In other cases, they are numerically determined from the function composition of the CDF $F_x(\cdot)$ with $\Phi^{-1}(\cdot)$ and of $\Phi(\cdot)$ with the inverse $F_x^{-1}(\cdot)$.

Predictions of $y(\mathbf{s})$ in the warped space can be obtained by applying GPR as shown in Section 2.1. Transferring these predictions to the observation space is straightforward by means of the principle of *quantile invariance*, which states that the quantiles of a probability distribution remain invariant under a monotonic transformation, i.e., if $\Phi(y_\alpha) = \alpha$ and $y_\alpha = g(x_\alpha)$, then it holds that $F_x(x_\alpha) = \alpha$ [23]. Therefore, the predictive distribution in the observation space can be reconstructed from the respective distribution in the warped space by means of the function $\tilde{g}(y)$ defined in Equation (9). More precisely, the optimal prediction is given by

$$\hat{z}(\mathbf{s}_p^*) = \tilde{g}\left(\hat{y}(\mathbf{s}_p^*)\right), \tag{10}$$

while the predictive interval at confidence level $(1 - \alpha) \times 100\%$ is given by

$$\left[\tilde{g}\left(\hat{y}(\mathbf{s}_p^*) - z_{\alpha/2} \hat{\sigma}_y(\mathbf{s}_p^*)\right), \tilde{g}\left(\hat{y}(\mathbf{s}_p^*) + z_{1-\alpha/2} \hat{\sigma}_y(\mathbf{s}_p^*)\right) \right]. \tag{11}$$

Note that Equation (10) returns the median of the marginal predictive distribution in the observation space. This is due to the principle of quantile invariance, taking into account that the conditional mean $\hat{y}(\mathbf{s}_p^*)$ is also the median of the latent variable’s marginal conditional distribution.

2.3. Data-Driven Warping of Gaussian Processes

Non-parametric (data-driven) warping refers to model-free warping functions based on non-parametric estimates of the CDF from the data. Such estimates can be obtained using kernel functions and are more flexible than those provided by parametric models [27,42]. To avoid confusion, we emphasize that GPR is by construction a non-parametric method, in the sense that the underlying process is approximated without invoking a parametric model in the space of functions (albeit the GPs involve a number of hyper-parameters that control the shape of possible functions). So, the term “non-parametric” in relation to the wGPR refers to the data-driven warping function.

Closed-form parametric probability distributions are not sufficiently flexible to provide accurate models for the precipitation amount over different time scales. An example is the amount of precipitation in semi-arid Mediterranean areas, where the optimal model varies significantly across months but also across years for the same month [21]. In addition, parametric models that were accurate for past observations may not adequately capture future precipitation patterns due to climate change and the expected increase of extreme

weather events [43]. Thus, the CDF of the data often cannot be accurately represented by a parametric model.

A non-parametric estimate of the CDF $F_x(x)$ is obtained by integrating kernel density estimators (KDEs), leading to semi-explicit CDF expressions as shown in [27]. The method of kernel cumulative distribution function estimation (KCDE) uses an adaptive plug-in kernel bandwidth based on the theoretical considerations presented in [42]. The KCDE method is shown to provide better estimates of $F_x(x)$ than the empirical (staircase) CDF estimate and kernel density estimation based on the normal-reference rule bandwidth.

A smoothing kernel is a real-valued, non-negative function $K(u; h) = K(\frac{u}{h})$, which respects the properties of normalization, i.e., $\int_{-\infty}^{\infty} du K(u) = 1$, and reflection symmetry, i.e., $K(u) = K(-u)$. The parameter $h > 0$ is the kernel bandwidth.

If $x_{[i]}$ represents the i -th order statistic (i.e., the i -th smallest value) of the sample vector \mathbf{x} , the standard PDF kernel density estimator is given by [44]

$$\hat{f}_K(x) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{x - x_{[i]}}{h}\right). \tag{12}$$

A kernel-based non-parametric estimate of the CDF $F_x(x)$ can be obtained by means of the following weighted sum [27]

$$\hat{F}_K(x) = \sum_{i=1}^N \frac{1}{N} \tilde{K}\left(\frac{x - x_{[i]}}{h}\right). \tag{13}$$

In Equation (13), $\tilde{K}(\cdot)$ is the CDF kernel step defined by means of the following integral

$$\tilde{K}\left(\frac{x - x_{[i]}}{h}\right) = \frac{1}{h} \int_{-\infty}^x dx' K\left(\frac{x' - x_{[i]}}{h}\right). \tag{14}$$

Equation (13) is obtained from Equation (12) using the integral $\hat{F}_K(x) = \int_{-\infty}^x \hat{f}_K(x') dx'$. The CDF kernel steps are smoothed versions of the discontinuous steps used in the staircase CDF estimation. Explicit expressions of CDF kernel steps for various kernel functions are obtained in [27].

Once $\hat{F}_K(x)$ is known, it can be used in Equations (8) and (9) to obtain the warping transform and its inverse. Since it is not in general possible to derive explicit expressions for the warping function and its inverse, the function $g(x)$ is defined in terms of a lookup table which contains CDF values, $\{p_i\}_{i=1}^{N_d}$, at $N_d = 4 \times 10^3$ discretization points; these are uniformly distributed over the interval $[x_{\min} - h, x_{\max} + h]$, where x_{\min} , x_{\max} are respectively the minimum and maximum sample values and h is the kernel bandwidth. The respective values of $g(x)$ are given by means of $\{\Phi^{-1}(p_i)\}_{i=1}^{N_d}$. The inverse transform requires finding x for a given probability level set by $p^* = \Phi(y^*)$. This is accomplished by linear interpolation of $x = F_X^{-1}(p)$ based on the two p_i values in the lookup table nearest to p^* .

2.4. Hyperparameter Estimation

The spatial correlations of the process $x(\mathbf{s})$ —or the transformed GP $y(\mathbf{s})$ —are determined by means of the variogram function; the latter is also known as the second-order structure function in turbulence studies [45]. The variogram is defined as the semi-variance of the process' increments, i.e., $\gamma_Z(\mathbf{r}) = \frac{1}{2} \text{Var}[z(\mathbf{s} + \mathbf{r}) - z(\mathbf{s})]$ [16]. The variogram is purely a function of \mathbf{r} if the increments are stationary (intrinsic hypothesis), a condition less strict than the stationarity of $z(\mathbf{s})$.

In the case of stationary $z(\mathbf{s})$, the covariance kernel and the variogram of the observed process $x(\mathbf{s})$ are connected by means of the equation

$$C(\mathbf{r}) = \sigma^2 - \gamma(\mathbf{r}), \tag{15}$$

where $\sigma^2 = \sigma_\epsilon^2 + \sigma_0^2$ is the total variance and $\gamma(\mathbf{r}) = \gamma_Z(\mathbf{r}) + \sigma_\epsilon^2(1 - \delta_{\|\mathbf{r}\|,0})$. As is known, the observation noise leads to a discontinuity of the variogram at the origin by σ_ϵ^2 (nugget effect).

The variogram estimates are less sensitive to stochastic trends than covariance estimates due to the differencing operation implied in the increments. Thus, stochastic kriging methods are based on the variogram to generate spatial predictions [13,15,16]. An empirical variogram function can be straightforwardly estimated from the data using the *method of moments* [46]. The empirical variogram comprises a set of lag distances and respective estimates of the increment process' semi-variance. This discrete function is then fitted to a theoretical model, i.e., a permissible (conditionally negative definite) function that is well-defined at every possible lag distance \mathbf{r} . This procedure, albeit less efficient than maximum likelihood estimation, provides a computationally fast and visually clear alternative.

A list of commonly used variogram models is given in Appendix A.

2.5. Assessment of Predictive Performance

Cross-validation (CV) is a methodology that employs a set of statistical criteria in order to assess the predictive performance of spatial models. The data are split into two disjoint sets for training and validation. Strategies for selecting training and validation sets are reviewed in [47]. The training set is used to tune the GP hyperparameters. The validation data are used for comparison with the model predictions and assessment of the model's performance [48]. A single training validation (100–400) split is used for the synthetic data example in Section 3.

A common strategy for sparse datasets is leave-one-out cross-validation (LOO-CV)—also known as delete-one CV [49], ordinary CV [50,51] or simply CV [52]. In LOO-CV the training set contains $N - 1$ values and the validation set contains a single value. All N possible partitions of the data into training and validation sets are used. This approach is employed in Section 4 for assessing the GPR and wGPR performance on the precipitation reanalysis data.

The predictive performance of different models is assessed by means of statistical measures which include: the bias or mean error (ME), the mean absolute error (MAE), the root mean square error (RMSE), Pearson's linear correlation coefficient (RP), the Nash-Sutcliffe coefficient (NS), the Empirical interval coverage (CVG), and the Negatively oriented interval score (NINTS) (see Appendix B for the definitions).

3. Application of GPR and Warped GPR to Synthetic Data

We apply the non-parametric wGPR approach to the synthetic 1D example used in [26]. The observation data are given by 100 random samples of

$$x(s) = (\sin(\pi s) + \sigma_\epsilon \epsilon(s))^{1/3}, \quad s \in [-1, 1], \quad (16)$$

where $\sigma_\epsilon = 0.1$ and $\epsilon(s) \sim \mathcal{N}(0, 1)$ provides the noise contamination of the underlying process $z(s) = (\sin(\pi s))^{1/3}$.

The process $z(s)$ is approximated using GPR and wGPR. The performance of the reconstructions is evaluated by comparing the regression estimates with a sample $\{z(s_p^*)\}_{p=1}^{N_v}$ at $N_v = 400$ uniformly distributed validation points in $[-\pi, \pi]$. In the case of GPR, the point predictions are obtained from the marginal posterior mean, Equation (5), the uncertainty is determined from the posterior covariance, Equation (6), and 95.45% prediction intervals are calculated according to Equation (7) with $z_{1-\alpha/2} = -z_{\alpha/2} = 2$. In the case of wGPR, the prediction and the 95.45% predictive intervals are obtained from Equations (10) and (11) respectively.

The empirical variograms for both the sample (in observation space) and its counterpart in the warped space are shown in Figure 1. Note that the spherical and exponential models, which imply continuous but non-differentiable stationary processes, provide poor fits for the empirical variogram (shown by markers), especially in the observation space.

This is due to the fact that the true process is a differentiable function mixed with noise. All the other models tested provide reasonable fits to the empirical variograms (with the inclusion of a small discontinuity term at the origin which reflects the noise). Remarkably, the power-law model ($\gamma(r) = \alpha |r|^{2H}$) which corresponds to a non-stationary, continuous but non-differentiable process also gives a good fit. The best fit for the empirical variograms is provided by the Spartan model given by Equation (A4). In one dimension, the Spartan covariance model represents the covariance function of a damped linear harmonic oscillator driven by white noise [23], (Chapter 9). Hence, it is a suitable candidate for the oscillatory observation process defined by Equation (16).

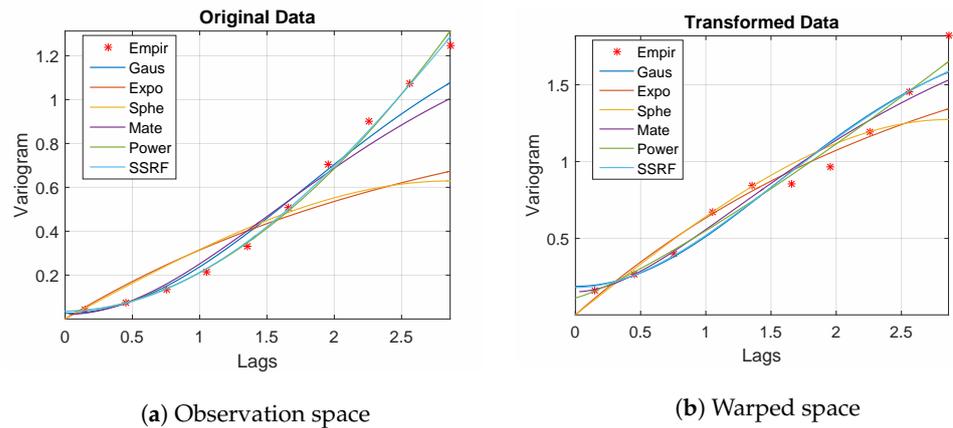


Figure 1. Empirical variograms (markers) and model fits (continuous lines) for the training set data (a) and the warped-space (normalized) data (b).

The bitriangular kernel is used in wGPR for estimating the data-derived CDF based on which the warping is performed. The kernel is defined by means of $K(u) = \frac{3}{2}(1 - |u|)^2$ for $|u| \leq 1$ and $K(u) = 0$ for $|u| > 1$. In both cases (GPR and wGPR) the Spartan variogram (covariance kernel) of Equation (A4) is used. Variograms corresponding to other covariance kernels (e.g., exponential, spherical, and Gaussian) were also studied. In general, non-differentiable kernels (e.g., spherical and exponential) lead to rougher reconstructions (exhibiting slope discontinuities) than the differentiable kernels (e.g., Gaussian and Spartan models).

Table 1 compares cross-validation metrics for the GPR and wGPR methods. Warped GPR achieves better performance in terms of most prediction metrics. High-performing prediction implies ME, MAE, and RMSE close to zero, RP and NS close to one, and small magnitudes of minimum and maximum errors. The bias (ME) is slightly lower for GPR, while the magnitudes of the minimum (ErrMin) and maximum (ErrMax) prediction errors are higher for wGPR than for GPR. The ErrMin and ErrMax values are influenced by the wGPR behavior near the left and right boundaries of the domain, where wGPR is closer than GPR to the observations. However, near the boundary the latter are mostly determined by $\epsilon(s)$ because $z(s) \approx 0$ near $s = \pm\pi$. The non-parametric warping transform used in wGPR is based on the CDF of the observations, thus enhancing adaptability to the data (for better or worse).

Table 1. Cross-validation metrics for GPR and wGPR based on the validation set of 400 points from the function of Equation (16). ME: Mean error. MAE: Mean absolute error. RMSE: Root mean square error. RP: Pearson’s correlation coefficient. NS: Nash-Sutcliffe coefficient. ErrMin: $\min_{s_1^*, \dots, s_p^*} (z(s_p^*) - \hat{z}(s_p^*))$. ErrMax: $\max_{s_1^*, \dots, s_p^*} (z(s_p^*) - \hat{z}(s_p^*))$.

	ME	MAE	RMSE	RP	NS	ErrMin	ErrMax
GPR	−0.012	0.095	0.147	0.985	0.97	−0.50	0.38
wGPR	−0.016	0.050	0.119	0.990	0.98	−0.76	0.64

The approximations of $z(s)$ obtained by means of GPR and wGPR are illustrated in Figure 2: this figure shows the noisy training data (markers), the true process $z(s)$ (black curve), as well as the GPR (magenta line) and wGPR (blue line) approximations with their respective 95.45% prediction intervals. Note that the validation values, $z(s_p^*)$, lie on the black curve (unmarked). As evidenced in the plots of Figure 2, GPR provides a smooth, differentiable approximation of $z(s)$, while wGPR yields a continuous but rough (non-differentiable) approximation that adapts more closely to the training data. In addition, the wGPR approximation is closer to the true $z(s)$ almost everywhere except for the boundaries (as discussed above).

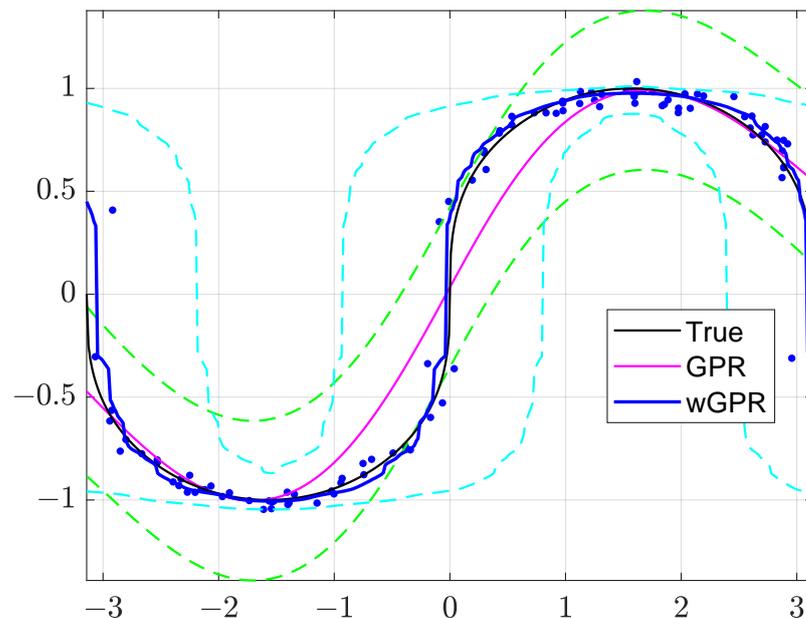


Figure 2. GPR and wGPR approximation of the function in Equation (16). Blue dots: Training set. Black line: The function $z(s)$ plotted versus πs on the horizontal axis. GPR approximations (classical GPR: magenta line, warped GPR, blue line) and 95.45% prediction intervals (GPR: green dash lines, wGPR: cyan dash lines.)

The GPR predictive intervals (green dashed lines) are symmetric around $\hat{z}(s)$, while the wGPR intervals (cyan dashed lines) are asymmetric and their width adjusts to the local slope of $z(s)$. Both prediction intervals contain most of the test values $z(s_p)$, $p = 1, \dots, 400$ (the prediction coverage is 100% for wGPR and $\approx 99\%$ for GPR). On the other hand, GPR achieves this coverage with tighter intervals than wGPR. This is due to the fact that wGPR adapts the warping transform to the CDF of the training data, and is thus influenced by random errors.

4. Application of GPR and Warped GPR to Reanalysis Data

In this section, GPR and wGPR are applied to a sparse, non-Gaussian, precipitation reanalysis dataset.

4.1. Study Area and Data Description

The study area is the island of Crete (Greece) in the southeastern part of the Mediterranean basin. Crete is the largest island in Greece with an area of 8336 km², length of 260 km, width ranging from 12 km to 57 km, and a maximum elevation of 2456 m. The island's climate exhibits a transition from the Mediterranean to semi-arid as is common in Mediterranean regions [21,53]. Temperature and precipitation exhibit significant local variations due to three mountain ranges which are among the highest in Europe. The island is divided into four administrative regional units (from West to East): Chania, Rethymno, Heraklion, and Lasithi.

To investigate the deviations from Gaussian behavior the data are first grouped by location and then by month. The models that were tested include the generalized Pareto, inverse Gaussian, lognormal, t-Scale location, generalized extreme value, Weibull, Gaussian, Birnbaum-Saunders, exponential, extreme value, gamma, Nakagami, logistic, log-logistic, Rayleigh, and Rician probability distributions. According to Akaike’s Information Criterion (AIC), the Nakagami model is optimal at 45 of 65 nodes, the Weibull at 14, the gamma at 4, and the Rayleigh distribution at the remaining two locations. The results based on the Bayesian Information Criterion (BIC) are similar, with the Nakagami model being optimal at 40 of 65 nodes, the Weibull at 12, the gamma at 4, and the Rayleigh distribution at the remaining 9 nodes. For the data grouped by month, the optimal distribution according to AIC is the Nakagami model for 8 out of the 246 wet months, the Weibull for 3, the gamma for 13, the GEV for 25, the Rayleigh for 1, the generalized Pareto for 126, the log-logistic for 2, the lognormal for 3, the Birnbaum-Saunders for 25, and the inverse Gaussian distribution for the remaining 40 wet months.

Table 2. Mean, median, minimum and maximum values (shown across rows) of monthly ERA5 precipitation statistics (shown across the columns) based on 246 monthly values (measured in mm). Each monthly statistic is based on the data at the 65 ERA5 grid nodes. The values for CoV (coefficient of variation), Skew (skewness) and Kurt (kurtosis) are dimensionless. All other values are measured in mm.

	Mean	Median	Min	Max	Std	CoV	Skew	Kurt
Mean	61.25	55.69	26.19	132.70	25.53	0.48	0.82	3.16
Median	59.19	51.78	21.23	123.98	23.67	0.45	0.81	3.04
Minimum	1.75	1.05	0.05	6.10	1.16	0.16	−0.01	1.56
Maximum	198.27	194.15	110.03	375.32	81.54	1.57	2.26	7.75

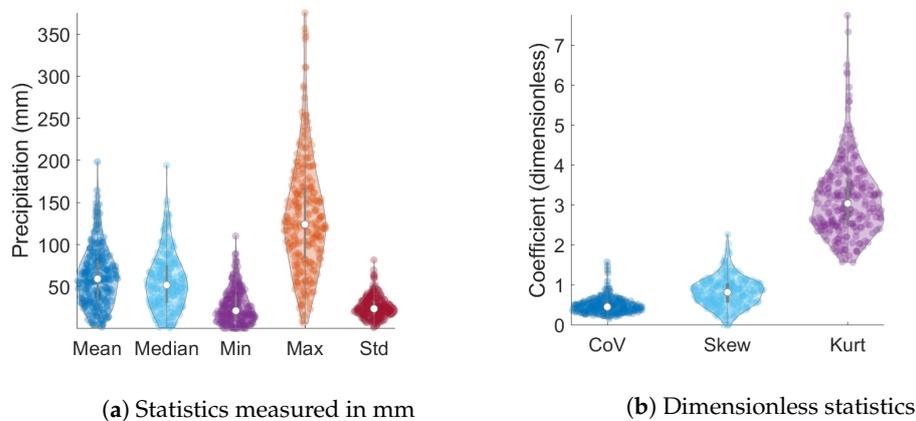


Figure 4. Violin plots for the mean, median, minimum and maximum values of monthly ERA5 precipitation statistics based on 246 monthly values. Each monthly statistic is based on the data at the 65 ERA5 grid nodes. The values for CoV (coefficient of variation), Skew (skewness) and Kurt (kurtosis) are dimensionless. All other values are measured in mm.

In the analysis below, GPR and wGPR are applied to data for each wet-season month at all locations. Hence, a different spatial model is generated for each month. However, monthly data follow a different non-Gaussian probability distribution each month, and hence the warping function in wGPR changes every time. For illustration, the precipitation probability distributions for the year 2008 are investigated. Sixteen parametric probability distribution models (as listed above) were tested. The optimal probability model per each wet season month is presented in Table 3 (based on BIC) (see also Figure 5). The optimal distribution for most months is the generalized Pareto (GP). However, this only means that GP achieves a better BIC value than the other models, but it does not ensure that GP is an accurate representation of the empirical distribution.

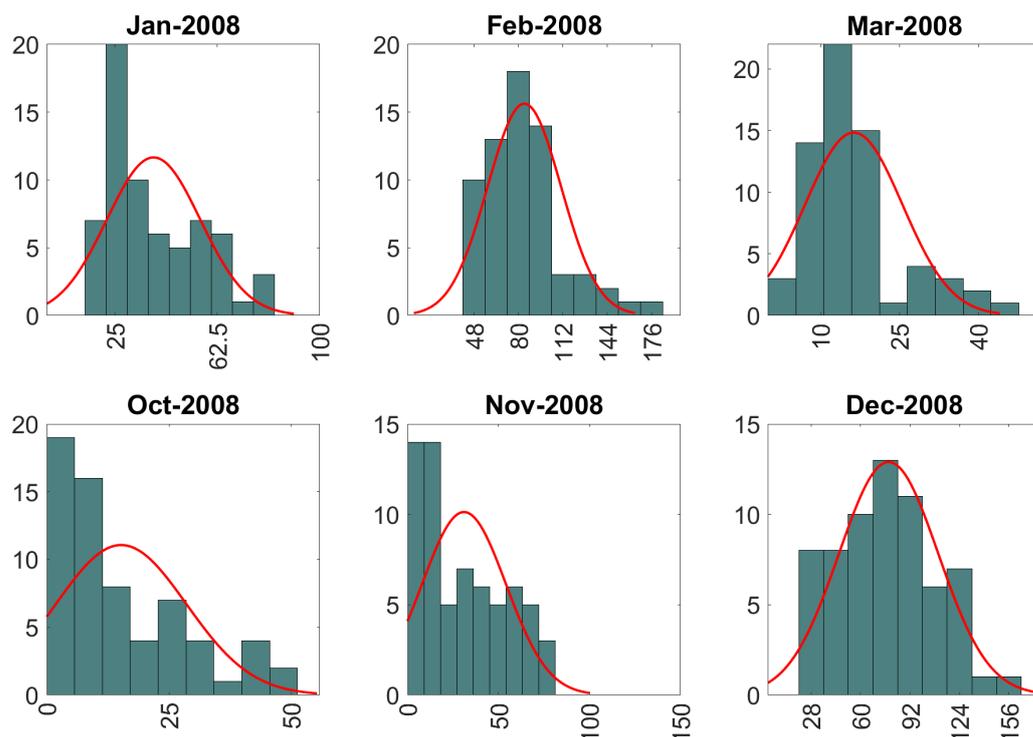


Figure 5. Distribution of monthly precipitation during the wet season of 2008. Histograms are based on ERA5 precipitation data at 65 grid locations over and around the island of Crete. Best fits to the optimal Gaussian PDF models (red line) are also shown. The vertical axis of the histograms represents frequency; the horizontal axis represents precipitation amount measured in mm.

Table 3. Optimal probability distribution fits (based on BIC) for the monthly ERA5 precipitation data in the year 2008. The models studied include the following: “GP”: Generalized Pareto, “InvGauss”: Inverse Gaussian, “Logn”: Lognormal, and “Wei”: Weibull distribution. The optimal probability distributions for each wet-season month are not uniformly the same for different years.

January	February	March	October	November	December
GP	InvGauss	Logn	GP	GP	Wei

4.3. GPR and Warped GPR Comparison Based on the Reanalysis Data

The following section presents LOO-CV analysis using GPR and wGPR for the six wet-season months over the 41-year period. The bitriangular kernel is used to estimate the non-parametric CDF which generates the warping function in wGPR. For both GPR and wGPR the spatial correlations are modeled by means of the exponential, Equation (A1), and Matérn, Equation (A3), variograms. The Spartan variogram, Equation (A5), was also tested, but the results obtained were nearly identical to those obtained with the exponential model. Hence, we kept the latter based on the principle of parsimony (the exponential model involves two hyperparameters compared to three for the Spartan model).

The distribution of LOO-CV metrics are shown in the violin plots of Figures 6–8. Each violin plot is generated by 246 values (6 months × 41 years) of the respective LOO-CV metric which is calculated based on the 65 values at the grid nodes. The LOO-CV metrics used include the mean error (ME), the mean absolute error (MAE), the root mean square error (RMSE), and the Spearman correlation coefficient (RS). In addition to these, we also use two-interval scores (see Appendix B): the interval coverage (CVG) and the negatively oriented interval score (NINTS) [60]. Negatively oriented scores imply that lower scores correspond to better predictions.

The average values of the cross-validation metrics over all 246 time slices are shown in Table 4. The values of the metric have been rounded up to the second decimal place. Based on Table 4 and Figures 6–8 there is practically no difference between the results obtained with the two different covariance kernels, in spite of the fact that the Matérn kernel allows for varying smoothness of the spatial function used to interpolate the data. The main differences are between GPR and wGPR. The mean error (bias) has a smaller magnitude for GPR than for wGPR. This is apparent in both the average values (-0.15 mm versus -0.53 mm) as well as the higher dispersion of the wGPR bias shown in Figure 6a. This behavior is expected since GPR is implemented so as to enforce a zero-bias constraint. Nonetheless, the wGPR bias is still a small fraction of the average minimum of the data (cf. Table 2). The wGPR-derived MAE and RMSE metrics are better than the respective GPR-based values, achieving both lower averages (Table 4) and lower extreme values (see Figures 6b and 7a). The average values of the Spearman correlation are identical for both methods. However, as shown in Figure 7b, the wGPR lower tail of the Spearman correlation distribution is shorter than that of GPR; this implies that wGPR predictions do not lead to poor (rank) correlations with the true data. In terms of the interval scores, wGPR is superior since it leads to lower average NINTS and higher average coverage (CVG) values. In addition, as seen in Figure 8a the lower values of the wGPR interval coverage are higher than those of GPR.

Table 4. Average values of LOO-CV metrics based on the 246 time slices of ERA5 precipitation data for the wet-season months.

	ME	MAE	RMSE	RS	NINTS	CVG
GPR (Expo)	-0.15	7.60	0.25	0.90	67.31	0.97
GPR (Mate)	-0.15	7.60	0.25	0.90	67.31	0.97
wGPR (Expo)	-0.53	7.53	0.21	0.90	65.50	0.98
wGPR (Mate)	-0.53	7.53	0.21	0.90	65.50	0.98

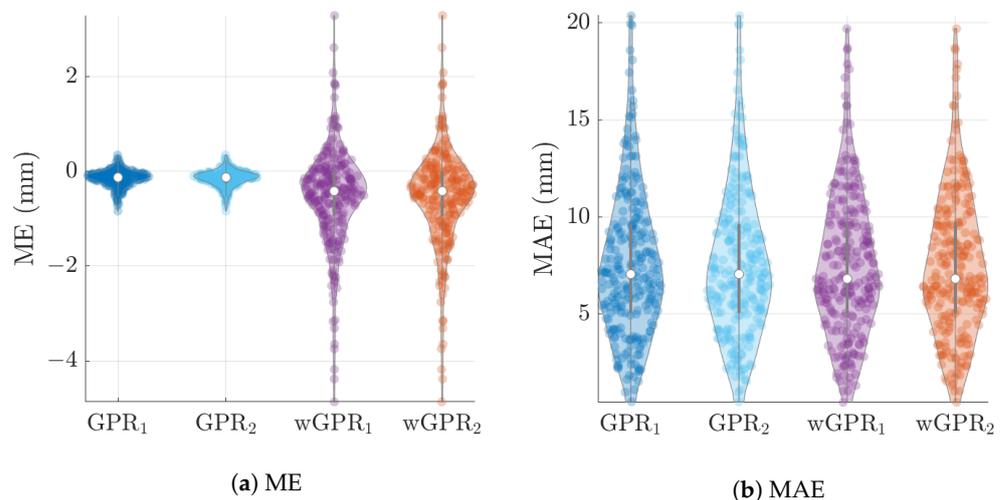


Figure 6. GPR and wGPR LOO-CV mean error (ME) and mean absolute error (MAE) for the wet-season ERA5 precipitation data. The lower indices “1” and “2” refer to the exponential and Matérn models respectively.

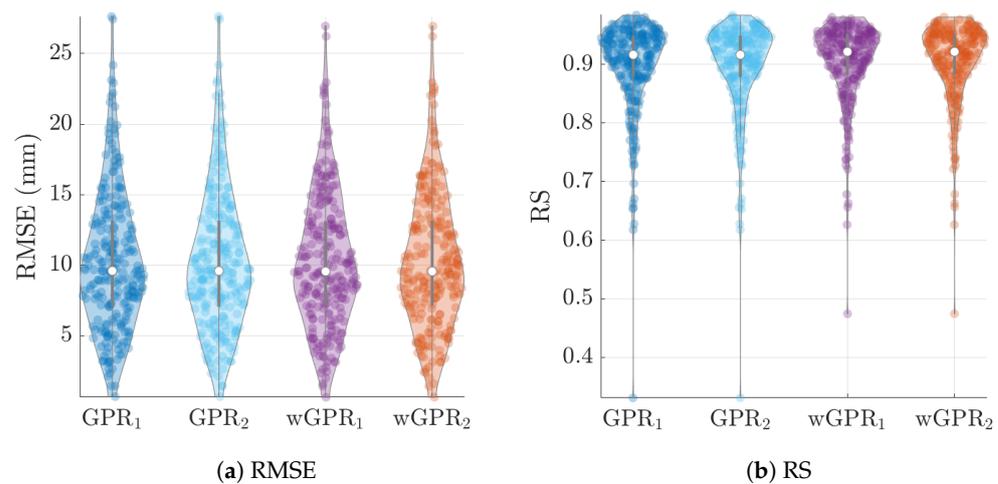


Figure 7. GPR and wGPR LOO-CV root mean error error (RMSE) and the Spearman correlation coefficient (RS) between the true and predicted values for the wet-season ERA5 precipitation data. The lower indices “1” and “2” refer to the exponential and Matérn models respectively.

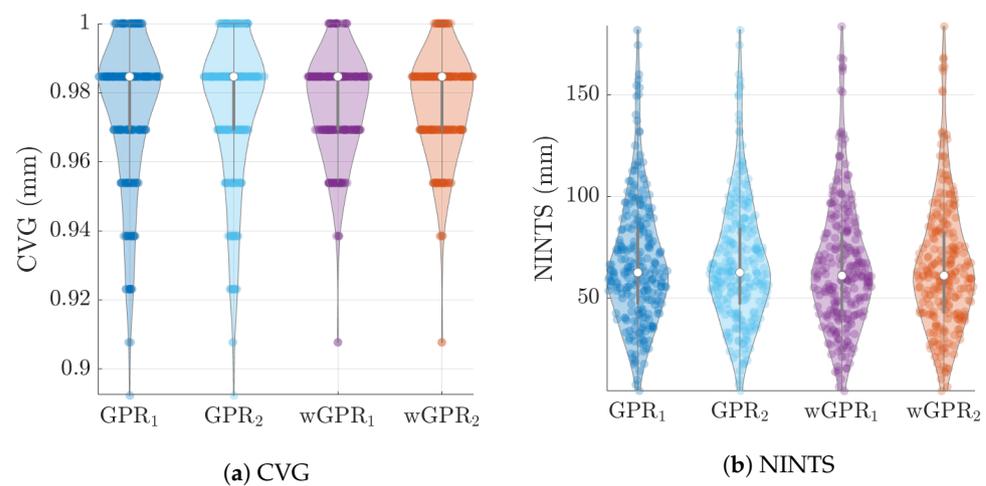


Figure 8. GPR and wGPR LOO-CV for two interval scores: the empirical interval coverage (CVG) and the negatively oriented interval score (NINTS) for the wet-season ERA5 precipitation data. The lower indices “1” and “2” refer to the exponential and Matérn covariance kernels respectively.

5. Discussion and Conclusions

There is a strong interest in the application of Gaussian processes to model spatial and spatiotemporal data [61–63]. In the case of data exhibiting non-Gaussian distribution (e.g., precipitation), nonlinear transforms are applied to the observations in order to allow the application of Gaussian assumptions and methods. In geostatistics, this practice is known as Gaussian anamorphosis [14,16], while in machine learning this approach is known as Gaussian process warping [26,63]. In the study that introduced warped Gaussian processes, the hyperbolic tangent function is used to implement the warping transformation [26]. Other closed-form, nonlinear, monotonic transforms (e.g., square root, Box-Cox, logarithm, Tukey g-and-h) can be used for this purpose [23] (Chapter 14), [16,63,64].

In this study, we introduce a data-driven (non-parametric) warping method for Gaussian processes which employs kernel-based estimates of the data CDF. The latter is derived directly from the observations and conforms to the shape of the probability distribution that underlies the data. This is different from the Snelson et al. [26] approach which used an explicit warping function. To our knowledge the proposed method for warping Gaussian processes is new. The term “non-parametric warping” has also been used to denote

deformations of the input space which can generate non-stationary Gaussian processes [61]. In this context, the warping function deforms the coordinate space so that in the new frame the resulting process can be considered stationary. In addition to applications in Gaussian processes, the non-parametric warping approach presented herein can also be used to perform Gaussian anamorphosis in the geostatistical framework.

Non-parametric warping of Gaussian processes allows greater flexibility than the use of parametric warping models: the shape of the data-driven warping function adapts to the features of the dataset at hand instead of being determined from a closed-form expression. Thus, non-parametric warping can provide improved approximation accuracy compared to GPR as evidenced in the 1D example studied in Section 3 and the precipitation reanalysis data (cf. Section 4). However, as shown in Section 3, non-parametric warping can lead to rough approximating functions if the data are contaminated by noise.

We combine non-parametric warping of Gaussian processes with Gaussian prediction (i.e., GPR/Kriging) and the principle of quantile invariance, in order to derive non-Gaussian marginal predictive distribution functions that capture the characteristics of the data distribution. The geostatistical method of variogram modeling is used to estimate the covariance kernel hyperparameters. Fitting the empirical variogram to theoretical models is a statistically less efficient procedure than likelihood maximization. However, it can lead to significant computational savings for large datasets where the $\mathcal{O}(N^3)$ computational cost of inverting the covariance (Gram) matrix may be prohibitive for likelihood optimization.

Our comparison of GPR and wGPR also employs monthly precipitation reanalysis (ERA5) data for the island of Crete. Reanalysis data provide valuable information regarding meteorological variables and the impact of climate change on their space-time patterns. Reanalysis data are especially useful in areas where environmental monitoring systems are sparse. As mentioned in Section 4, reanalysis datasets typically require bias correction in order to provide locally accurate estimates of precipitation. Bias correction can be conducted in combination with measurements from ground gauges, if such data are available. The GPR and wGPR methods can be used in combination with reanalysis and/or ground datasets to generate enhanced-resolution spatiotemporal maps for water resources management applications. A cross-validation comparison of the two methods (GPR and wGPR) showed that GPR has a lower bias, but wGPR is better with respect to other measures including interval scores. In addition, wGPR exhibits better performance in extreme cases where GPR led to poor cross-validation results. No significant differences are found between covariance kernels capable of exponential-like behavior (e.g., the exponential, Spartan, and Matérn models). This is attributed to the irregular spatial patterns of precipitation over Crete which are marked by exponentially decaying correlations [21].

The warped GPR model can be further investigated along the research directions described below. The present study does not model the dependence of precipitation on altitude. This can be accomplished by means of polynomial trend functions or other function bases. More complex measures of distance can be used in the covariance kernel to account for the anisotropic dependence of spatial precipitation patterns in the West-East and North-South directions of the island [21]. The scaling of the amount of precipitation with respect to the spatial support (i.e., the area over which the amount of precipitation is measured) is an important factor in the assessment of flood risk in ungauged watersheds [65], which needs to be linked to the interpolation procedure. The data from each wet-season month were herein treated as separate time slices. It is in principle possible to view the dataset in a space-time continuum and to adopt a spatiotemporal model for correlations. From the modeling perspective, the main difficulty is the construction of space-time covariance kernels that can adequately capture interactions between spatial and temporal correlations, as mentioned in the Introduction. From the numerical perspective, the covariance kernel inversion for the spatiotemporal problem has a significantly higher computation cost due to the respectively larger size of the dataset; the latter would in this case include $N = N_t \times N_s$ points, where N_s is the number of ERA5 nodes and N_t the total number of time slices. The issue of increased computational complexity can be addressed

by replacing the Gaussian process with a stochastic local interaction model (SLI). The latter expresses space-time correlations in terms of sparse precision matrices. In SLI models the precision matrix is built using a data-adaptive strategy; this can lead to extremely sparse structures, thus reducing the computational cost of interpolation [23,66,67]. In addition to the above, in the space-time framework, a suitable temporal distance should be defined to capture seasonal trends in the precipitation patterns and space-time interaction of the correlations.

In conclusion, we have demonstrated that warped GPR equipped with non-parametric (data-driven) warping functions provides increased flexibility and enhanced accuracy for the spatial prediction of non-Gaussian distributions based on incomplete spatial datasets. The case studies that we examined include a synthetic test function with non-Gaussian noise and precipitation reanalysis data. The probability distributions representative of these datasets involves various degrees of departure from Gaussianity. Based on this evidence, we believe that wGPR will perform well in different applications that involve non-Gaussian data, although this remains to be explicitly demonstrated in future studies. Application of wGPR modeling may lead to poor results for certain types of non-Gaussian probability distributions: these include distributions that allow for isolated spatial extremes, and distributions with discrete spikes (e.g., zero-inflated precipitation during the dry season). The latter could be addressed by means of methods such as the variational, zero-inflated Gaussian process regression [68].

Author Contributions: Conceptualization, D.T.H., V.D.A. and A.P.; methodology, D.T.H., V.D.A. and A.P.; software, D.T.H., V.D.A. and A.P.; validation, V.D.A. and A.P.; formal analysis, D.T.H.; investigation, V.D.A., A.P. and D.T.H.; resources, V.D.A., A.P. and D.T.H.; data curation, V.D.A.; writing—original draft preparation, V.D.A. and D.T.H.; writing—review and editing, D.T.H., V.D.A. and A.P.; visualization, V.D.A.; supervision, D.T.H.; project administration, D.T.H.; funding acquisition, D.T.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research is co-financed by Greece and the European Union (European Social Fund-ESF) through the Operational Programme «Human Resources Development, Education and Lifelong Learning 2014–2020» in the context of the project “Gaussian Anamorphosis with Kernel Estimators for Spatially Distributed Data and Time Series and Applications in the Analysis of Precipitation” (MIS 5052133).

Data Availability Statement: The precipitation reanalysis datasets analyzed during the study are available in the ERA5-ECMWF dataset repository (ERA5|ECMWF). Restrictions apply to the availability of the data. Data was obtained from <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land> (accessed on 9 March 2020).

Acknowledgments: We would like to thank Ioanna Byzyka and Sofia Katakaki, Special Account of Research Funds (ELKE) of the Technical University of Crete for the administrative support of the project.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
GEV	Generalized Extreme Value distribution
GP	Generalized Pareto distribution
GPR	Gaussian process regression
KCDE	Kernel cumulative distribution estimate
KDE	Kernel density estimate
wGPR	Warped Gaussian Process Regression

Appendix A. Variogram Models

Variogram models define via Equation (15) the covariance kernel used in GPR. The models studied herein include the exponential, spherical, Matérn and Spartan models [23]. The relevant expressions are given below. In the following equations, ξ is the characteristic length of the spatial process.

Exponential model:

$$\gamma_0(\mathbf{r}) = \sigma_0^2 [1 - \exp(-\|\mathbf{r}\|/\xi)]. \tag{A1}$$

Spherical model:

$$\gamma_0(\|\mathbf{r}\|) = \begin{cases} \sigma_0^2 \left[1.5 \left(\frac{\|\mathbf{r}\|}{\xi} \right) - 0.5 \left(\frac{\|\mathbf{r}\|}{\xi} \right)^3 \right], & \text{if } \|\mathbf{r}\| \leq \xi, \\ \sigma_0^2, & \text{if } \|\mathbf{r}\| \geq \xi. \end{cases} \tag{A2}$$

Matérn model:

$$\gamma_0(\|\mathbf{r}\|) = \sigma_0^2 \left[1 - \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{\|\mathbf{r}\|}{\xi} \right)^\nu K_\nu \left(\sqrt{2\nu} \frac{\|\mathbf{r}\|}{\xi} \right) \right]. \tag{A3}$$

For the Matérn model, $\nu > 0$ is the smoothness hyperparameter which controls the continuity of the process. $\Gamma(\cdot)$ is the gamma function, and $K_\nu(\cdot)$ is the modified Bessel function of the second kind of order ν .

Spartan model ($d = 1$):

$$\gamma_0(r) = \begin{cases} \sigma_0^2 - \eta_0 e^{-|r|\beta_2/\xi} \left[\frac{\cos(|r|\beta_1/\xi)}{4\beta_2} + \frac{\sin(|r|\beta_1/\xi)}{4\beta_1} \right], & |\eta_1| < 2, \\ \sigma_0^2 - \eta_0 \frac{1+|r|/\xi}{4} e^{-|r|/\xi}, & \eta_1 = 2, \\ \sigma_0^2 - \frac{\eta_0}{\sqrt{\eta_1^2-4}} \left[\frac{e^{-|r|\omega_1/\xi}}{2\omega_1} - \frac{e^{-|r|/\xi\omega_2}}{2\omega_2} \right], & \eta_1 > 2, \end{cases} \tag{A4}$$

Spartan model ($d = 2$):

$$\gamma_0(\|\mathbf{r}\|) = \begin{cases} \sigma_0^2 - \frac{\eta_0}{\pi\sqrt{4-\eta_1^2}} \mathfrak{S} \left[K_0 \left(\frac{\|\mathbf{r}\|}{\xi\omega_2} \right) \right], & |\eta_1| < 2, \\ \sigma_0^2 - \frac{\eta_0\|\mathbf{r}\|}{4\pi\xi} K_{-1} \left(\frac{\|\mathbf{r}\|}{\xi} \right), & \eta_1 = 2, \\ \sigma_0^2 - \frac{\eta_0}{2\pi\sqrt{\eta_1^2-4}} \left[K_0 \left(\frac{\|\mathbf{r}\|}{\xi\omega_2} \right) - K_0 \left(\frac{\|\mathbf{r}\|}{\xi\omega_1} \right) \right], & \eta_1 > 2. \end{cases} \tag{A5}$$

For the Spartan model, the variance σ_0^2 is determined from the hyperparameters η_0, η_1 and ξ as follows [69]:

$$(d = 1) \quad \sigma_0^2 = \begin{cases} \frac{\eta_0}{2\sqrt{2+\eta_1}}, & |\eta_1| < 2, \\ \frac{\eta_0}{4}, & \eta_1 = 2, \\ \frac{\eta_0}{2\sqrt{\eta_1^2-4}} (\omega_1^{-1} - \omega_2^{-1}), & \eta_1 > 2. \end{cases} \tag{A6}$$

$$(d = 2) \quad \sigma_0^2 = \begin{cases} \frac{\eta_0}{2\pi\sqrt{4-\eta_1^2}} \left[\frac{\pi}{2} - \arctan\left(\frac{\eta_1}{\sqrt{4-\eta_1^2}}\right) \right], & |\eta_1| < 2, \\ \frac{\eta_0}{4\pi}, & \eta_1 = 2, \\ \frac{\eta_0}{4\pi\sqrt{\eta_1^2-4}} \ln\left(\frac{\eta_1+\sqrt{\eta_1^2-4}}{\eta_1-\sqrt{\eta_1^2-4}}\right), & \eta_1 > 2. \end{cases} \quad (A7)$$

In the Spartan model $\eta_1 > -2$ is the rigidity hyperparameter (smaller η_1 allow oscillatory behavior of the covariance while $\eta_1 \geq 2$ lead to exponential decay). The coefficients $\beta_{1,2}$ are determined as $\beta_{1,2} = \frac{1}{2}|2 \mp \eta_1|^{1/2}$. The hyperparameters $\omega_{1,2}$ are dimensionless damping coefficients that determine the decay of the slow and fast exponential functions (for $\eta_1 > 2$). They are given by means of $\omega_{1,2}^2 = \left(\eta_1 \mp \frac{\sqrt{\eta_1^2-4}}{2}\right)$.

Appendix B. Cross-Validation Metrics

The following equations define the leave-one-out cross-validation metrics used. The symbol $\hat{z}_{-i}(\mathbf{s}_i)$ denotes the prediction at point $\mathbf{s}_i \in \mathbb{R}^2$ based on the $N - 1$ data points excluding \mathbf{s}_i ; the true value at \mathbf{s}_i is represented by $z(\mathbf{s}_i)$; the spatial average of the data is denoted by $\bar{z}(\mathbf{s}_i)$; and $\bar{\hat{z}}_{-i}(\mathbf{s}_i)$ is the spatial average of the predictions. The above metrics are straightforwardly extended to the case of disjoint training and validation sets.

Mean error (bias) (ME):

$$\epsilon_{\text{bias}} = \frac{1}{N} \sum_{i=1}^N [z(\mathbf{s}_i) - \hat{z}_{-i}(\mathbf{s}_i)]. \quad (A8)$$

Mean absolute error (MAE):

$$\epsilon_{\text{MA}} = \frac{1}{N} \sum_{i=1}^N |\hat{z}_{-i}(\mathbf{s}_i) - z(\mathbf{s}_i)|. \quad (A9)$$

Root mean square error (RMSE):

$$\epsilon_{\text{RMS}} = \sqrt{\frac{1}{N} \sum_{i=1}^N [\hat{z}_{-i}(\mathbf{s}_i) - z(\mathbf{s}_i)]^2}. \quad (A10)$$

Pearson's linear correlation coefficient (RP):

$$\bar{\rho}_P = \frac{\text{COV}(z, \hat{z})}{\sigma_z \sigma_{\hat{z}}} = \frac{\sum_{i=1}^N [z(\mathbf{s}_i) - \bar{z}(\mathbf{s}_i)] [\hat{z}_{-i}(\mathbf{s}_i) - \bar{\hat{z}}_{-i}(\mathbf{s}_i)]}{\sqrt{\sum_{i=1}^N [z(\mathbf{s}_i) - \bar{z}(\mathbf{s}_i)]^2} \sqrt{\sum_{i=1}^N [\hat{z}_{-i}(\mathbf{s}_i) - \bar{\hat{z}}_{-i}(\mathbf{s}_i)]^2}}. \quad (A11)$$

Spearman's rank correlation coefficient is also obtained from Equation (A11) by replacing z, \hat{z} with $R(z), R(\hat{z})$, where $R(\cdot)$ represents the rank function.

Nash-Sutcliffe coefficient (NS):

$$\epsilon_{\text{NS}} = 1 - \frac{\sum_{i=1}^N [\hat{z}_{-i}(\mathbf{s}_i) - z(\mathbf{s}_i)]^2}{\sum_{i=1}^N [z(\mathbf{s}_i) - \bar{z}(\mathbf{s}_i)]^2}. \quad (A12)$$

Empirical interval coverage (CVG):

$$\text{CVG} = \frac{1}{N} \sum_{i=1}^N \text{CVG}_i, \quad (\text{A13})$$

$$\text{CVG}_i = \mathbb{1}(z(\mathbf{s}_i) > l_i \wedge z(\mathbf{s}_i) < u_i) \quad (\text{A14})$$

In the above, $\mathbb{1}(\cdot)$ is the indicator function: $\mathbb{1}(A) = 1$ if A is true and $\mathbb{1}(A) = 0$ if A is false. In addition, l_i and u_i represent, respectively, the $\alpha/2$ and $1 - \alpha/2$ quantiles of the predictive distribution at the point \mathbf{s}_i . The CVG thus represents the fraction of points where the prediction interval (at a specified quantile level $0 < \alpha < 1$) contains the true value of the sample. Herein it is assumed that $\alpha = 4.55\%$.

Negatively oriented interval score (NINTS):

$$\bar{S}_\alpha = \frac{1}{N} \sum_{i=1}^N S_\alpha(l_i, u_i, z(\mathbf{s}_i)), \quad (\text{A15})$$

$$S_\alpha(l, u, x) = (u - l) + \frac{2}{\alpha}(l - x)\mathbb{1}(x < l) + \frac{2}{\alpha}(x - u)\mathbb{1}(x > u). \quad (\text{A16})$$

The l_i and u_i are defined above for the CVG.

References

1. Stocker, T.; Qin, D.; Plattner, G.K.; Tignor, M.; Allen, S.; Boschung, J.; Nauels, A.; Xia, Y.; Bex, V.; Midgley, P.M. (Eds.) *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Cambridge University Press: Cambridge, UK, 2013. [\[CrossRef\]](#)
2. IPCC. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*; Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.L., et al., Eds.; Cambridge University Press: Cambridge, UK, 2021.
3. Varouchakis, E.A.; Corzo, G.A.; Karatzas, G.P.; Kotsopoulou, A. Spatio-temporal analysis of annual rainfall in Crete, Greece. *Acta Geoph.* **2018**, *66*, 319–328. [\[CrossRef\]](#)
4. Zribi, M.; Brocca, L.; Trambly, Y.; Molle, F. *Water Resources in the Mediterranean Region*; Elsevier: Amsterdam, The Netherlands, 2020.
5. Deng, K.A.K.; Lamine, S.; Pavlides, A.; Petropoulos, G.P.; Srivastava, P.K.; Bao, Y.; Hristopoulos, D.; Anagnostopoulos, V. Operational soil moisture from ASCAT in support of water resources management. *Remote Sens.* **2019**, *11*, 579. [\[CrossRef\]](#)
6. Giorgi, F. Climate change hot-spots. *Geoph. Res. Lett.* **2006**, *33*, L08707. [\[CrossRef\]](#)
7. Luterbacher, J.; Xoplaki, E.; Casty, C.; Wanner, H.; Pauling, A.; Küttel, M.; Brönnimann, S.; Fischer, E.; Fleitmann, D.; Gonzalez-Rouco, F.J.; et al. Mediterranean climate variability over the last centuries: A review. In *Mediterranean Climate Variability*; Lionello, P., Malanotte-Rizzoli, P., Boscolo, R., Eds.; Elsevier: Amsterdam, The Netherlands, 2006; Volume 4, pp. 27–148. [\[CrossRef\]](#)
8. Norrant, C.; Douguédroit, A. Monthly and daily precipitation trends in the Mediterranean (1950–2000). *Theor. Appl. Climatol.* **2006**, *83*, 89–106. [\[CrossRef\]](#)
9. Christakos, G. *Spatiotemporal Random Fields: Theory and Applications*; Elsevier: Amsterdam, The Netherlands, 2017.
10. Varouchakis, E.A.; Hristopoulos, D.T. Comparison of spatiotemporal variogram functions based on a sparse dataset of groundwater level variations. *Spat. Stat.* **2019**, *34*, 100245. [\[CrossRef\]](#)
11. Porcu, E.; Furrer, R.; Nychka, D. 30 Years of space–time covariance functions. *Wiley Interdiscip. Rev. Comput. Stat.* **2021**, *13*, e1512. [\[CrossRef\]](#)
12. Christakos, G. *Random Field Models in Earth Sciences*; Academic Press: San Diego, CA, USA, 1992.
13. Cressie, N. *Statistics for Spatial Data*, revised ed.; Series in Probability and Statistics; Wiley: New York, NJ, USA, 1993.
14. Wackernagel, H. *Multivariate Geostatistics*; Springer: Berlin, Germany, 2003.
15. Olea, R.A. *Geostatistics for Engineers and Earth Scientists*; Springer: New York, NY, USA, 1999.
16. Chilès, J.P.; Delfiner, P. *Geostatistics: Modeling Spatial Uncertainty*, 2nd ed.; John Wiley & Sons: New York, NY, USA, 2012.
17. Boer, E.P.; de Beurs, K.M.; Hartkamp, A.D. Kriging and thin plate splines for mapping climate variables. *Int. J. Appl. Earth Obser. Geoinfor.* **2001**, *3*, 146–154. [\[CrossRef\]](#)
18. Guan, H.; Wilson, J.L.; Makhnin, O. Geostatistical mapping of mountain precipitation incorporating autosearched effects of terrain and climatic characteristics. *J. Hydrometeorol.* **2005**, *6*, 1018–1031. [\[CrossRef\]](#)
19. Moral, F.J. Comparison of different geostatistical approaches to map climate variables: Application to precipitation. *Int. J. Climatol.* **2010**, *30*, 620–631. [\[CrossRef\]](#)

20. Verdin, A.; Funk, C.; Rajagopalan, B.; Kleiber, W. Kriging and local polynomial methods for blending satellite-derived and gauge precipitation estimates to support hydrologic early warning systems. *IEEE Tran. Geosci. Remote Sens.* **2016**, *54*, 2552–2562. [[CrossRef](#)]
21. Agou, V.D.; Varouchakis, E.A.; Hristopulos, D.T. Geostatistical analysis of precipitation in the island of Crete (Greece) based on a sparse monitoring network. *Environ. Monit. Assess.* **2019**, *191*, 1573–2959. [[CrossRef](#)]
22. Rasmussen, C.E.; Williams, C.K.I. *Gaussian Processes for Machine Learning*; MIT Press: Cambridge, MA, USA, 2006.
23. Hristopulos, D. *Random Fields for Spatial Data Modeling: A Primer for Scientists and Engineers*; Springer: Dordrecht, The Netherlands, 2020. [[CrossRef](#)]
24. Papalexiou, S.M.; Serinaldi, F.; Porcu, E. Advancing space-time simulation of random fields: From storms to cyclones and beyond. *Water Resour. Res.* **2021**, *57*, e2020WR029466. [[CrossRef](#)]
25. Papalexiou, S.M.; Serinaldi, F. Random fields simplified: Preserving marginal distributions, correlations, and intermittency, with applications from rainfall to humidity. *Water Resour. Res.* **2020**, *56*, e2019WR026331. [[CrossRef](#)]
26. Snelson, E.; Rasmussen, C.E.; Ghahramani, Z. Warped Gaussian processes. *Adv. Neural Inf. Process. Syst.* **2004**, *16*, 337–344.
27. Pavlides, A.; Agou, V.; Hristopulos, D.T. Non-parametric kernel-based estimation of probability distributions for precipitation modeling. *arXiv* **2021**, arXiv:2109.09961.
28. Papalexiou, S.M.; AghaKouchak, A.; Fofoula-Georgiou, E. A diagnostic framework for understanding climatology of tails of hourly precipitation extremes in the United States. *Water Resour. Res.* **2018**, *54*, 6725–6738. [[CrossRef](#)]
29. Ye, L.; Hanson, L.S.; Ding, P.; Wang, D.; Vogel, R.M. The probability distribution of daily precipitation at the point and catchment scales in the United States. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 6519–6531. [[CrossRef](#)]
30. Wilks, D.S. Maximum likelihood estimation for the gamma distribution using data containing zeros. *J. Clim.* **1990**, *3*, 1495–1501. [[CrossRef](#)]
31. Wilks, D.S.; Eggleston, K.L. Estimating monthly and seasonal precipitation distributions using the 30- and 90-day outlooks. *J. Clim.* **1992**, *5*, 252–259. [[CrossRef](#)]
32. Shoji, T.; Kitaura, H. Statistical and geostatistical analysis of rainfall in central Japan. *Comput. Geosci.* **2006**, *32*, 1007–1024. [[CrossRef](#)]
33. Kedem, B.; Chiu, L.S.; North, G.R. Estimation of mean rain rate: Application to satellite observations. *J. Geophys. Res. Atmos.* **1990**, *95*, 1965–1972. [[CrossRef](#)]
34. Cho, H.K.; Bowman, K.P.; North, G.R. A comparison of gamma and lognormal distributions for characterizing satellite rain rates from the tropical rainfall measuring mission. *J. Appl. Meteorol.* **2004**, *43*, 1586–1597. [[CrossRef](#)]
35. Wang, Z.; Zeng, Z.; Lai, C.; Lin, W.; Wu, X.; Chen, X. A regional frequency analysis of precipitation extremes in mainland China with fuzzy c-means and L-moments approaches. *Int. J. Climatol.* **2017**, *37*, 429–444. [[CrossRef](#)]
36. Coles, S. *An Introduction to Statistical Modeling of Extreme Values*; Springer Series in Statistics; Springer: London, UK, 2001.
37. Gellens, D. Combining regional approach and data extension procedure for assessing GEV distribution of extreme precipitation in Belgium. *J. Hydrol.* **2002**, *268*, 113–126. [[CrossRef](#)]
38. Scheuerer, M. Probabilistic quantitative precipitation forecasting using ensemble model output statistics. *Quart. J. R. Meteorol. Soc.* **2014**, *140*, 1086–1096. [[CrossRef](#)]
39. Koutsoyiannis, D. Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records. *Hydrol. Sci. J.* **2004**, *49*, 591–610. [[CrossRef](#)]
40. Moccia, B.; Papalexiou, S.M.; Russo, F.; Napolitano, F. Spatial variability of precipitation extremes over Italy using a fine-resolution gridded product. *J. Hydrol. Reg. Stud.* **2021**, *37*, 100906. [[CrossRef](#)]
41. Baxevani, A.; Lennatsson, J. A spatiotemporal precipitation generator based on a censored latent Gaussian field. *Water Resour. Res.* **2015**, *51*, 4338–4358. [[CrossRef](#)]
42. Botev, Z.I.; Grotowski, J.F.; Kroese, D. Kernel density estimation via diffusion. *Ann. Stat.* **2010**, *38*, 2916–2957. [[CrossRef](#)]
43. Madsen, H.; Lawrence, D.; Lang, M.; Martinkova, M.; Kjeldsen, T. Review of trend analysis and climate change projections of extreme precipitation and floods in Europe. *J. Hydrol.* **2014**, *519*, 3634–3650. [[CrossRef](#)]
44. Parzen, E. On estimation of a probability density function and mode. *Ann. Math. Stat.* **1962**, *33*, 1065–1076. [[CrossRef](#)]
45. Arenas, A.; Chorin, A.J. On the existence and scaling of structure functions in turbulence according to the data. *Proc. Nat. Acad. Sci. USA* **2006**, *103*, 4352–4355. [[CrossRef](#)] [[PubMed](#)]
46. Matheron, G. Principles of Geostatistics. *Econ. Geol.* **1963**, *58*, 1246–1266. [[CrossRef](#)]
47. Sylvain, A.; Alain, C. A survey of cross-validation procedures for model selection. *Stat. Surv.* **2010**, *4*, 40–79. [[CrossRef](#)]
48. Grossman, R.; Seni, G.; Elder, J.; Agarwal, N.; Liu, H. *Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions*; Morgan & Claypool: San Rafael, CA, USA, 2010. [[CrossRef](#)]
49. Li, K.C. Asymptotic optimality for C_p , C_L , cross-validation and generalized cross-validation: Discrete index set. *Ann. Statist.* **1987**, *15*, 958–975. [[CrossRef](#)]
50. Stone, M. Cross-validated choice and assessment of statistical predictions. *J. R. Statist. Soc. Ser. B* **1974**, *36*, 111–147. [[CrossRef](#)]
51. Burman, P. A comparative study of ordinary cross-validation, v -fold cross-validation and the repeated learning-testing methods. *Biometrika* **1989**, *76*, 503–514. [[CrossRef](#)]
52. Efron, B. Estimating the error rate of a prediction rule: Improvement on cross-validation. *J. Am. Statist. Assoc.* **1983**, *78*, 316–331. [[CrossRef](#)]

53. Watrous, L. *Lasithi: A History of Settlement on a Highland Plain in Crete*, xviii ed.; American School of Classical Studies: Princeton, NJ, USA, 1982.
54. Copernicus Climate Change Service C3S. ERA5: Fifth Generation of ECMWF Atmospheric Reanalyses of the Global Climate. 2018. Available online: <https://cds.climate.copernicus.eu/cdsapp#!/home> (accessed on 10 March 2020).
55. Dee, D.; Fasullo, J.; Shea, D. The Climate Data Guide: Atmospheric Reanalysis: Overview & Comparison Tables. Last modified 12 December 2016. Available online: <https://climatedataguide.ucar.edu/climate-data/atmospheric-reanalysis-overview-comparison-tables> (accessed on 7 October 2021).
56. Reichle, R.H.; Liu, Q.; Koster, R.D.; Draper, C.S.; Mahanama, S.P.P.; Partyka, G.S. Land Surface Precipitation in MERRA-2. *J. Clim.* **2017**, *30*, 1643–1664. [[CrossRef](#)]
57. Muñoz Sabater, J.; Dutra, E.; Agustí-Panareda, A.; Albergel, C.; Arduini, G.; Balsamo, G.; Boussetta, S.; Choulga, M.; Harrigan, S.; Hersbach, H.; et al. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data* **2021**, *13*, 4349–4383. [[CrossRef](#)]
58. Xu, X.; Frey, S.K.; Ma, D. Hydrological performance of ERA5 and MERRA-2 precipitation products over the Great Lakes Basin. *J. Hydrol. Reg. Stud.* **2022**, *39*, 100982. [[CrossRef](#)]
59. Google Earth Pro 7.3.4.8248. (14 December 2015). Crete island, Greece, 35°16′12.97″N, 25°1′25.14″E, Eye alt 273.94 km. SIO, NOAA, U.S. Navy, NGA, GEBCO. Image Landsat/Copernicus. Available online: <https://earth.google.com/web/> (accessed on 13 April 2021).
60. Gneiting, T.; Raftery, A.E. Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.* **2007**, *102*, 359–378. [[CrossRef](#)]
61. Marmin, S.; Baccou, J.; Liandrat, J.; Ginsbourger, D. Non-parametric warping via local scale estimation for non-stationary Gaussian process modelling. *Int. Soc. Opt. Photonics* **2017**, *10394*, 1039421.
62. Lu, C.K.; Shafto, P. Conditional deep Gaussian processes: Multi-fidelity kernel learning. *Entropy* **2021**, *23*, 1545. [[CrossRef](#)] [[PubMed](#)]
63. Peters, G.W.; Nevat, I.; Nagarajan, S.G.; Matsui, T. Spatial warped Gaussian processes: Estimation and efficient field reconstruction. *Entropy* **2021**, *23*, 1323. [[CrossRef](#)]
64. Xu, G.; Genton, M.G. Tukey g-and-h random fields. *J. Am. Stat. Assoc.* **2017**, *112*, 1236–1249. [[CrossRef](#)]
65. Barbero, G.; Moisello, U.; Todeschini, S. Evaluation of the Areal Reduction Factor in an Urban Area through Rainfall Records of Limited Length: A Case Study. *J. Hydrol. Engin.* **2014**, *19*, 05014016. [[CrossRef](#)]
66. Hristopulos, D.T.; Agou, V.D. Stochastic local interaction model with sparse precision matrix for space–time interpolation. *Spat. Stat.* **2020**, *40*, 100403. [[CrossRef](#)]
67. Hristopulos, D.T.; Pavlides, A.; Agou, V.D.; Gkafa, P. Stochastic local interaction model: An alternative to kriging for massive datasets. *Math. Geosci.* **2021**, *53*, 1907–1949. [[CrossRef](#)]
68. Hegde, P.; Heinonen, M.; Kaski, S. Variational zero-inflated Gaussian processes with sparse kernels. In Proceedings of the 34th Conference on Uncertainty in Artificial Intelligence (UAI), Monterey, CA, USA, 6–10 August 2018; Globerson, A., Silva, R., Eds.; AUA Press: Corvallis, OR, USA, 2018; Volume 1, pp. 361–371.
69. Hristopulos, D.T. Covariance functions motivated by spatial random field models with local interactions. *Stoch. Environ. Res. Risk Assess.* **2015**, *29*, 739–754. [[CrossRef](#)]