EGU22-6288
https://doi.org/10.5194/egusphere-egu22-6288
EGU General Assembly 2022


# Data-driven Warping of Gaussian Processes for Spatial Interpolation of Skewed Data

**Dionissios Hristopulos**[1], Vasiliki Agou[2], and Andrew Pavlides[2]
[1]Technical University of Crete, Electrical and Computer Engineering, Chania, Greece (dchristopoulos@isc.tuc.gr)
[2]Technical University of Crete, Mineral Resources Engineering, Chania, Greece

Gaussian processes are a flexible machine learning framework that can be used for spatial interpolation and space-time prediction as well. Gaussian process regression (GPR) is quite similar to the geostatistical kriging method. It encompasses various types of kriging (e.g., simple, ordinary, universal and regression kriging). In addition, it is formulated in an inherently Bayesian framework which allows taking into account a priori beliefs regarding the distribution of the model's hyper-parameters. Thus, it also incorporates Bayesian versions of kriging [1]. GPR is based on the assumption that the stochastic component of the observations follows a Gaussian distribution. However, this is not the case for various environmental variables (e.g., amount of precipitation, hydraulic conductivity, wind speed), which follow skewed probability distributions. The skewness is handled within the geostatistical framework using nonlinear transforms such that the marginal distribution of the data in the latent space becomes normal. This procedure is known as Gaussian anamorphosis in geostatistics. In the context of GPR, the term warped Gaussian process is used to denote the nonlinear transformation of the observations [2]. Gaussian anamorphosis (warping) is usually implemented using explicit, monotonically increasing nonlinear functions. A different approach involves generating the warping function with the help of the empirically estimated cumulative probability distribution of the data. This approach provides flexibility because the transformation is data-driven (non-parametric) and is thus not constrained by specific functional forms. Furthermore, the cumulative distribution function of the data can be accurately estimated using smoothing kernels [3]. We investigate warped Gaussian process regression using synthetic datasets and precipitation reanalysis data 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 demonstrate that warped GPR equipped with data-driven warping provides enhanced flexibility compared to "bare" GPR and can lead to improved predictive accuracy for non-Gaussian data.

Keywords: Gaussian processes, Mediterranean island, non-Gaussian, warping, precipitation

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).

**References**

[1] T. Hristopulos, 2020. Random Fields for Spatial Data Modeling. Springer Netherlands, http://dx.doi.org/10.1007/978-94-024-1918-4.

[2] Snelson, E., Rasmussen, C.E. and Ghahramani, Z., 2004. Warped Gaussian processes. *Advances in neural information processing systems*, *16*, pp.337-344.

[3] Pavlides, A., Agou, V., and Hristopulos, D. T., 2021. Non-parametric Kernel-Based Estimation of Probability Distributions for Precipitation Modeling. *arXiv preprint arXiv:2109.09961*.