EGU21-14322
https://doi.org/10.5194/egusphere-egu21-14322
EGU General Assembly 2021

Gaussian Process Regression hyperparameter optimization for image time series gap-filling of Earth observation data and crop monitoring

Santiago Belda[1], Matías Salinero[1], Eatidal Amin[1], Luca Pipia[2], Pablo Morcillo-Pallarés[1], and Jochem Verrelst[1]

[1]University of Valencia, Image Processing Laboratory (IPL), Valencia, Spain (santiago.belda@uv.es)
[2]Institut Cartogràfic i Geològic de Catalunya (ICGC), Barcelona, Spain

In general, modeling phenological evolution represents a challenging task mainly because of time series gaps and noisy data, coming from different viewing and illumination geometries, cloud cover, seasonal snow and the interval needed to revisit and acquire data for the exact same location. For that reason, the use of reliable gap-filling fitting functions and smoothing filters is frequently required for retrievals at the highest feasible accuracy. Of specific interest to filling gaps in time series is the emergence of machine learning regression algorithms (MLRAs) which can serve as fitting functions. Among the multiple MLRA approaches currently available, the kernel-based methods developed in a Bayesian framework deserve special attention because of both being adaptive and providing associated uncertainty estimates, such as Gaussian Process Regression (GPR).

Recent studies demonstrated the effectiveness of GPR for gap-filling of biophysical parameter time series because the hyperparameters can be optimally set for each time series (one for each pixel in the area) with a single optimization procedure. The entire procedure of learning a GPR model only relies on appropriate selection of the type of kernel and the hyperparameters involved in the estimation of input data covariance. Despite its clear strategic advantage, the most important shortcomings of this technique are the (1) high computational cost and (2) memory requirements of their training, which grows cubically and quadratically with the number of model's samples, respectively. This can become problematic in view of processing a large amount of data, such as in Sentinel-2 (S2) time series tiles. Hence, optimization strategies need to be developed on how to speed up the GPR processing while maintaining the superior performance in terms of accuracy.

To mitigate its computational burden and to address such shortcoming and repetitive procedure, we evaluated whether the GPR hyperparameters can be preoptimized over a reduced set of representative pixels and kept fixed over a more extended crop area. We used S2 LAI time series over an agricultural region in Castile and Leon (North-West Spain) and testing different functions for Covariance estimation such as exponential Kernel, Squared exponential kernel and matern kernel with parameter 3/2 or 5/2. The performance of image reconstructions was compared against the standard per-pixel GPR time series training process. Results showed that accuracies

were on the same order (12% RMSE degradation) whereas processing time accelerated up to 90 times. Crop phenology indicators were also calculated and compared, revealing similar temporal patterns with differences in start and end of growing season of no more than five days. To the benefit of crop monitoring applications, all the gap-filling and phenology indicators retrieval techniques have been implemented into the **freely downloadable GUI toolbox DATimeS** (Decomposition and Analysis of Time Series Software - https://artmotoolbox.com/).