

## COMPUTER SCIENCE

# A perspective on Gaussian processes for Earth observation

Gustau Camps-Valls<sup>1,\*</sup>, Dino Sejdinovic<sup>2</sup>, Jakob Runge<sup>3</sup> and Markus Reichstein<sup>4</sup>

## INTRODUCTION

Earth observation (EO) by airborne and satellite remote sensing and *in situ* observations plays a fundamental role in monitoring our planet. In the last decade, machine learning has attained outstanding results in the estimation of bio-geo-physical variables from the acquired images at local and global scales in a time-resolved manner. Gaussian processes (GPs) [1], as flexible non-parametric models to find functional relationships, have excelled in EO problems in recent years, mainly introduced for model inversion and emulation of complex codes [2]. GPs provide not only accurate estimates but also principled uncertainty estimates for the predictions. Besides, GPs can easily accommodate multimodal data coming from different sensors and from multitemporal acquisitions. Due to their solid Bayesian formalism, GPs can include prior physical knowledge about the problem, and allow for a formal treatment of uncertainty quantification and error propagation.

In remote sensing, we often deal with *radiative transfer models* (RTMs), which implement the equations of energy transfer. These codes are needed for modelling, understanding and predicting

some variables of interest related to the state of the land cover, water bodies and atmosphere. An RTM  $f$  operating in *forward mode* generates a multidimensional *radiance observation*  $y \in \mathbb{R}^p$  seen by the sensor given a multidimensional *parameter state vector*  $x \in \mathbb{R}^d$ ; see Fig. 1. Running forward simulations yields a look-up-table (LUT) of input–output pairs,  $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^n$ . Solving the *inverse problem* implies learning the function  $g$  using  $\mathcal{D}$  to return an estimate  $x_*$  each time a new satellite observation  $y_*$  is acquired. GPs have been used to learn both the often costly forward model  $f$  as well as the inverse model  $g$ . Learning the forward model allows for faster simulations, while learning an inverse model has allowed the provision of physically meaningful,

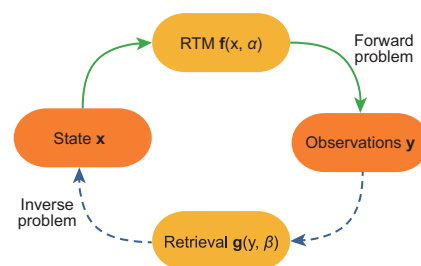
spatially explicit and temporally resolved maps of variables of interest.

Despite great advances in forward and inverse modelling, GP models still have to face important challenges, such as the high computational cost involved or the derivation of faithful confidence intervals. More importantly, we posit that GP models should evolve towards *data-driven physics-aware models* that respect signal characteristics, be consistent with elementary laws of physics, and move from pure regression to observational *causal inference*.

## ADVANCES IN GP INVERSE MODELLING

The most important shortcoming of GPs is their high computational cost and the memory requirements, which grows cubically and quadratically with the number of training points, respectively. Recently, great progress has been made in constructing scalable versions of GPs, demonstrating their utility in big data regimes [3].

An important challenge in Earth observation relates to the fact that data come with complex non-linearities, levels and sources of noise, and non-stationarities. Standard GPs often



**Figure 1.** Forward (solid lines) and inverse (dashed lines) modelling in Earth observation.

assume homoscedastic noise and use stationary kernels though. The current state-of-the-art GP to deal with heteroscedastic noise makes use of a marginalized variational approximation [2]. The method has resulted in excellent performance in estimating biophysical parameters (chlorophyll-a content in plants and water bodies) from acquired reflectances. In many EO applications one transforms the observed variable to linearize or Gaussianize the data via parametric transforms. A *warped* GP model has allowed learning a non-parametric optimal transformation from data, and has shown very good results in predicting vegetation parameters (chlorophyll, leaf area index, and fractional vegetation cover) from hyperspectral images [4]. Another common problem in remote sensing is that of ensuring consistency across products: estimating several related variables simultaneously can incorporate their relations in a single model. A recent latent force model (LFM) GP can encode ordinary/partial differential equations governing the system, and has allowed monitoring of crops, estimation of multiple vegetation covariates simultaneously, and management of missing observations due to the presence of clouds or sensor acquisition problems [5].

Making inferences with GPs is not only about obtaining point-wise estimates but also faithful uncertainty estimates, essential in performing error propagation. Inference should also contemplate *extrapolation* analysis as an ambitious far-end goal. Besides, note that we ultimately aim to characterize model error by comparing simulators to reality, calibrate models by proper estimation of (hyper)parameters, and make uncertainty statements about the world that combine models, data, and their corresponding errors. We think that the Bayesian formalism of GPs is the natural framework to tackle these yet unresolved problems.

## ADVANCES IN GP FORWARD MODELLING

Surrogate modelling, also known as *emulation*, based on GPs is gaining popular-

ity in remote sensing. Emulators are essentially statistical models that learn to mimic the RTM code using a representative dataset  $\mathcal{D}$ . GPs have largely dominated the field for decades and have provided excellent accuracy and physical consistency as studied via sensitivity analysis in the context of vegetation and atmosphere models in [2]. Once the GP model is trained, one can readily perform fast forward simulations, which in turn allows improved inversion. However, replacing an RTM with a GP model requires running expensive evaluations of  $\mathbf{f}$  first. Recent more efficient alternatives construct an approximation to  $\mathbf{f}$  starting with a set of support points selected iteratively [5]. This topic is related to active learning and Bayesian optimization, which might push results further in accuracy and sparsity, especially when modelling complex codes.

RTMs are the result of many decades of scientific research and continuous development, so they often include *ad hoc* rules, heuristics and non-differentiable links that hamper analytic treatment. Emulation allows one to account for input errors, derive predictive variance estimates, infer sensitivity values of parameters, calculate Jacobians and perform uncertainty propagation and quantification analytically. Besides, a lot of physical knowledge used for designing RTMs could be translated in designing priors (e.g. physically plausible parameter values). These excellent capabilities have not been widely exploited in EO applications though.

## TOWARDS PHYSICS-AWARE GP MODELLING

The GP framework allows us to include constraints and priors adapted to *signal features* such as non-stationarity, circularity, spatial-temporal relations, coloured-noise processes and non-i.i.d. relations. Nevertheless, data-driven GP models should be further constrained to provide physically plausible predictions. Recent approaches consider designing joint observation-simulation cross-covariances [5]. Recently we suggested a full framework for hybrid modelling with

machine learning [6], which could be formalized within the GP probabilistic framework too.

Learning dynamical physical systems is very challenging. Recent regression approaches have learned the governing equations of non-linear dynamical systems from data, such as the Lorenz, Navier-Stokes and Schrödinger equations. Models typically impose sparsity and hierarchical modelling, but also a GP probabilistic approach has excelled in discovering ordinary and partial differential, integro-differential and fractional order operators [7].

The integration of physics into GP models not only achieves improved generalization but, more importantly, endorses these grey-box models with *consistency* and *faithfulness*. As a by-product, the hybridization process has an interesting regularization effect, as physics discards implausible models and promotes simpler structures.

## FROM REGRESSION TO CAUSATION

Understanding is more challenging than predicting, especially when no interventional studies can be conducted, as in the Earth sciences. Causal inference from observational data to estimate causal graphical models has become a mature science with effective machine-learning methods to deal with both time series and non-time-ordered data; see [8,9] and references therein. Causal inference methods can be classified roughly into conditional independence or constraint-based approaches and structural causal models. Constraint-based causal discovery algorithms iteratively infer graphical models utilizing conditional independence testing. In [10] a GP-based conditional independence test is combined with a scalable causal discovery algorithm allowing high-dimensional graphical models to be inferred from time-series data. Constraint-based algorithms only allow causal graphical models to be inferred up to a Markov equivalence class. Utilizing additional assumptions, such as on the noise distribution or functional dependence, the class of structural causal

models [8] allows causal directionality to be inferred in such undecidable Markov equivalent cases. Recently, [2,4] introduced the WGP regression in additive noise models to account for post-non-linear effects and heteroscedastic noise respectively, and applied it successfully to a set of geoscience and remote sensing bi-variate problems. Some important challenges in causal inference for the Earth science are still to be solved: how to scale GP models to deal with millions of points, missing data and time aggregation as well as time sub-sampling, and complex spatial-temporal dependence structures. Testing scientific hypotheses, comparing model-vs-data causal graphs, and assessing the impacts of extreme events are just some exciting avenues of further research.

## FUNDING

G.C.V. would like to acknowledge the support from the European Research Council (ERC) under

the ERC Consolidator Grant 2014 project SEDAL (647423).

Gustau Camps-Valls<sup>1,\*</sup>, Dino Sejdinovic<sup>2</sup>, Jakob Runge<sup>3</sup> and Markus Reichstein<sup>4</sup>

<sup>1</sup>Image Processing Laboratory, University of Valencia, Spain

<sup>2</sup>Department of Statistics, University of Oxford, UK

<sup>3</sup>German Aerospace Center, Institute of Data Science, Germany

<sup>4</sup>Max Planck Institute for Biogeochemistry, Germany

\*Corresponding author.

E-mail: [gustau.camps@uv.es](mailto:gustau.camps@uv.es)

## REFERENCES

1. Rasmussen CE and Williams CKI. *Gaussian Processes for Machine Learning*. New York: MIT Press, 2006.
2. Camps-Valls G, Verrelst J and Muñoz-Marí J *et al.* *IEEE Geosci Remote Sens Mag* 2016; **4**: 58–78.
3. Hensman J, Fusi N and Lawrence ND. Gaussian processes for big data. In: Nicholson A and Smyth P (eds). *Uncertainty in Artificial Intelligence*. Oregon: AUAI Press 2013, 282–90.

4. Mateo-Sanchis A, Muñoz-Marí J and Pérez-Suay A *et al.* *IEEE Geosci Remote Sens Lett* 2018; **15**: 1647–51.
5. Camps-Valls G, Svendsen D and Martino L *et al.* *Appl Soft Comput* 2018; **68**: 69–82.
6. Reichstein M, Camps-Valls G and Stevens B *et al.* *Nature* 2019; **566**: 195–204.
7. Raissi M, Perdikaris P and Karniadakis GE. *J Comput Phys* 2017; **348**: 683–93.
8. Peters J, Janzing D and Schölkopf B. *Elements of Causal Inference: Foundations and Learning Algorithms*. Cambridge: MIT Press, 2017.
9. Zhang K, Schölkopf B and Spirtes P *et al.* *Natl Sci Rev* 2018; **5**: 26–9.
10. Runge J. *Chaos* 2018; **28**: 075310.

National Science Review

6: 616–618, 2019

doi: 10.1093/nsr/nwz028

Advance access publication 4 March 2019