A tale (no, 10 tales!) about why embracing Gaussian Processes?

By Gustau Camps-Valls et al. 2020
@isp_uv_es
Advances in Gaussian Processes for Earth Sciences

Physics-aware, interpretability and consistency

Gustau Camps-Valls, Daniel Svendsen, Luca Martino, Adrian Pérez, Maria Piles, and Jordi Muñoz

Image Processing Lab (IPL) - Universitat de València

http://isp.uv.es  @isp_uv_es  gustau.camps@uv.es
Earth observation
Earth observation meets machine learning
Machine learning

\[ F(X) = y \]

- **X**: observations, independent covariates
- **Y**: target, dependent variable
- **F**: machine learning model (nonlinear, nonparametric, flexible, learned from data)
Challenges in ML

- **Consistency issue**: ML models do not respect Physics
- **Learning issue**: ML are excellent approximators, yet no fundamental laws are learned
- **Interpretability issue**: Big data is good to estimate correlations, what about causation?
Gaussian Processes in a nutshell

“A Survey on Gaussian Processes for Earth Observation Data Analysis: A Comprehensive Investigation”

“A Perspective on Gaussian Processes for Earth Observation”
Gustau Camps-Valls, Dino Sejdinovic, Jakob Runge, Markus Reichstein, National Science Review 6 (4) :616-618, 2019
Gaussian Processes in a nutshell

- Input-output data: $D = \{x_n \in \mathbb{R}^D, y_n\}_{n=1}^N$
- Observation model:
  $$y_n = f(x_n) + \varepsilon_n, \quad \varepsilon_n \sim \mathcal{N}(0, \sigma_n^2)$$
- Test point $x_*$ with corresponding output $y_*$
- Posterior over the unknown output:
  $$p(y_* | x_*, D) = \mathcal{N}(y_* | \mu_{GP*}, \sigma_{GP*}^2)$$
  $$\mu_{GP*} = k_{f*}^T(K + \sigma_n^2 I_N)^{-1}y$$
  $$\sigma_{GP*}^2 = \sigma_n^2 + k_{**} - k_{f*}^T(K + \sigma_n^2 I_N)^{-1}k_{f*}$$
- Marginal likelihood optimization:
  $$\log p(y) = -\frac{1}{2} y^T(K + \sigma_n^2 I_N)^{-1}y - \frac{1}{2} \log |K + \sigma_n^2 I_N| - \frac{N}{2} \log(2\pi)$$
Standard GP models for parameter retrieval

Observations, $x$: CHRIS images: 62 bands, 400-1050 nm, 34m
Variables, $y$: *In situ* leaf-level Chl (CCM-200) and LAI (PocketLAI phone app!)
Standard GP models for parameter retrieval

- Vegetation parameters from remote sensing data: chlorophyll content, LAI, vegetation cover

“A Survey on Gaussian Processes for Earth Observation Data Analysis: A Comprehensive Investigation”
The truth is that...
Hybrid modeling, where knowledge lies ...
1: GP for multisource fusion

- Fuse optical and radar information for gap filling, for multiple output predictions
- Several parameters estimated simultaneously improve consistency

\[
f(t) = \begin{bmatrix} f_{PS_1}(t) \\ f_{PS_2}(t) \end{bmatrix} = \begin{bmatrix} a_{1,1}u_1^1(t) + a_{1,2}u_2^1(t) \\ a_{2,1}u_1^2(t) + a_{2,2}u_2^2(t) \end{bmatrix}
\]
2: Distribution regression GP models

- When $X$ is a distribution associated with a scalar $y$
- Distribution regression embeds distributions in Hilbert spaces and runs regression therein
- Many examples: bunch of pixels in a region and associated target variable

\[ P \mapsto \mu_k(P) \rightarrow P \mapsto [\mathbb{E}\phi_1(X), \ldots, \mathbb{E}\phi_s(X)] \in \mathbb{R}^s \]

\[ \langle \mu_k(P), \mu_k(Q) \rangle_{H_k} = \mathbb{E}_{X \sim P, Y \sim Q} k(X, Y) \]
3: Deep Gaussian Processes

- GP also goes deep; improved versatility and expressive power
- Excellent performance in emulation and atmospheric parametrization
4: Physics-aware GPs with constrained optimization

- GP minimizes errors & predictions are indep. of ancillary data

$$\text{FairLoss} = \text{Cost}(y, \hat{y}) + \lambda_1 \| w \|^2_2 + \gamma I(\hat{y}, s)$$
4: Constrained optimization

- GP minimizes errors & predictions are indep. of anthropogenic factors

\[ \text{FairLoss} = \text{Cost}(y, \hat{y}) + \lambda_1 \| w \|^2_2 + \gamma I(\hat{y}, s) \]
5: Joint GP models for forward-inverse modeling

- Let ML talk to physical models

\[
\text{JointLoss} = \text{Cost}(y, \hat{y}) + \lambda_1 ||w||^2 + \gamma \Omega(\hat{y}, \Phi) \\
\Omega(\hat{y}, \Phi) = \text{Cost}_s(y_s, \hat{y}_s)
\]

Setup
- ERMES project: 3 rice sites, 85% European production
- Landsat 8 + in situ measurements + PROSAIL simulations
- In situ LAI measurements: \( r = 70-300 \) (3 countries, 2 years)
- Simulations: \( s = 2000 \) (Landsat 8 spectra and LAI)

Filling the space ...

“Joint Gaussian Processes for Biophysical Parameter Retrieval”
Svendsen, Martino, Camps-Valls, IEEE TGARS 2018

“Physics-aware Gaussian processes in remote sensing”
6: Multioutput GP regression encoding ODEs

- Transfer learning across time, sensors and space: “LFs and noise are GPs + lin.op = a GP!”

- LAI and fAPAR data for Spain and Italy ($Q = 4$ outputs)
- Multioutput improves single output GPs (4.5% gain in MSE)
- Transportability across time/space of estimates

“Gap filling of biophysical parameters with multi-output GPs”

“Learning latent forces from Earth time series”
Svendsen, Muñoz, Piles, Camps-Valls, subm. 2020
7: GP emulation of complex codes

- GP Emulation = Uncertainty quantification/propagation + Sensitivity analysis + Speed


“Emulation as an accurate alternative to interpolation in sampling radiative transfer codes”, Vicent and Camps-Valls, IEEE Journal Sel. Topics Rem. Sens, Apps. 2018
Uncertainty propagation comes with closed-form solutions!

be approximated as a Gaussian. The expectation gives us the same sample GP prior mean, but the resulting equation for the variance of the unknown outputs $y_*$ for a new incoming test input point $x_*$ changes as

$$v^2_{GP} = T_{xx} + k_{xx} - k_1^T (K + \sigma^2 I) T^{-1} k_2$$

(4)

where the effect of the noise in the inputs is represented by $T_{ij} = T(x_i, x_j) = \partial_i \Sigma \partial_j$, and $T_{xx} = T(x_i, x_j)$. We denoted the vector of partial derivatives of $f$ with respect to the sample $x_i$ as:

$$\partial_i := \left[ \frac{\partial f (x_i)}{\partial x_i^1}, \ldots, \frac{\partial f (x_i)}{\partial x_i^D} \right]^T$$

The derivative of the predictive function $f$ (2) in GPs only depends on the derivative of the Kernel function since it is linear with respect to the $\alpha$ parameters

$$\frac{\partial f (x_i)}{\partial x_i^j} = \frac{\partial k_i}{\partial x_i^j} = (\partial k_{ij})^\alpha$$

where $\partial k_{ij} = \left[(\partial k(x_i, x_j) / \partial x_i^j), \ldots, (\partial k(x_i, x_j) / \partial x_i^j) \right]^T$. Please see\(^1\) for a working implementation of the error GP (eGP) model.
9: GP interpretability

- The derivatives of the GP model are analytical!
- Sensitivity analysis for free!

- Look at the feature lengthscales ...

$$k(x_i, x_j) = \nu \exp \left( - \sum_{f=1}^{F} \frac{(x_i^f - x_j^f)^2}{2\sigma_f^2} \right) + \sigma_n^2 \delta_{ij}$$

- Compute feature sensitivity ...

Sensitivities of $\mu(x)$ wrt features

$$s_f = \int \left( \frac{\partial \mu(x)}{\partial x^f} \right)^2 \rho(x) dx \approx \frac{1}{N} \sum_{i=1}^{N} \left( \sum_{j=1}^{N} \rho_{ij} \frac{x_i^f - x_j^f}{\sigma_f^2} k(x_i, x_j) \right)^2$$

- Sensitivity maps yield ...
  - Fast feature ranking and sensitivity analysis
  - Example relevance and informed PDF sampling

"Ranking drivers of global carbon and energy fluxes over land"
9: GP interpretability

- The derivatives of the GP model are analytical!
- Sensitivity analysis for free!
  - Look at the feature lengthscales ...
    \[
    k(x_i, x_j) = \nu \exp \left( - \sum_{f=1}^{F} \frac{(x_{i,f} - x_{j,f})^2}{2\sigma_f^2} \right) + \sigma_n^2 \delta_{ij}
    \]
  - Compute feature sensitivity ... Derivatives of \( \mu(x) \) wrt features
    \[
    s_f = \int \left( \frac{\partial \mu(x)}{\partial x^f} \right)^2 p(x) dx \approx \frac{1}{N} \sum_{q=1}^{N} \left( \sum_{p=1}^{N} \frac{\alpha_p(x_{p,f} - x_{q,f})}{\sigma_f^2} k(x_p, x_q) \right)^2
    \]
  - Sensitivity maps yield ...
    - Fast feature ranking and sensitivity analysis
    - Example relevance and informed PDF sampling

---

"Gaussian Process Sensitivity Analysis for Oceanic Chlorophyll Estimation"
Blix, K. and Jenssen, R. and Camps-Valls, Gustau, IEEE JSTARS, 2017

2: ocean chlorophyll estimation

```
<table>
<thead>
<tr>
<th>Database</th>
<th>ME</th>
<th>RMSE</th>
<th>MAE</th>
<th>( \rho )</th>
</tr>
</thead>
<tbody>
<tr>
<td>SeaBAM (2, 4 and 5)</td>
<td>+0.0037</td>
<td>0.1493</td>
<td>0.1104</td>
<td>0.9679</td>
</tr>
<tr>
<td>SeaWIFS (4, 5 and 6)</td>
<td>-0.0887</td>
<td>0.3149</td>
<td>0.2361</td>
<td>0.9236</td>
</tr>
<tr>
<td>MODIS-Aqua (4, 5 and 6)</td>
<td>+0.0229</td>
<td>0.2461</td>
<td>0.1866</td>
<td>0.9188</td>
</tr>
<tr>
<td>MERIS (synthetic) (5, 6, 7 and 8)</td>
<td>0.004</td>
<td>0.084</td>
<td>0.0232</td>
<td>0.9996</td>
</tr>
<tr>
<td>MERIS (real) (5, 6, 7 and 8)</td>
<td>&lt; 10^{-7}</td>
<td>0.21</td>
<td>0.1464</td>
<td>0.9261</td>
</tr>
</tbody>
</table>
```

- Most sensitive bands between 510 nm and 560 nm
- Preferred bands match across sensors: 555, 547, 510, 443 nm
- Identified sensitive bands match parametric models OC2, CalCOFI, and OC4
- High (low) sensitivity of GP mean (variance) are preferred
10: GP Granger Causality

- Causal inference goes beyond correlation analysis
- Granger causality tests whether the past of X is useful to predict the future of Y
- We introduce a GP-based Granger causality to account for nonlinear GC relations

\[
a_H = (K(X_t, X'_t) + \varepsilon_t^Y)^{-1}Y_{t+1}\\
b_H = (L([Y_t, X_t], [Y'_t, X'_t]) + \varepsilon^Y_t|X)^{-1}Y_{t+1}
\]

\[
X \rightarrow Y \leftrightarrow \nabla_H [\varepsilon^Y_t] \ll \nabla_H [\varepsilon^Y_t|X]
\]

"Revisiting impact of MJO on soil moisture: a causality perspective", AGU 2019
"Explicit Granger causality in Hilbert spaces" Bueso, Piles and Camps-Valls, submitted, 2020
10: GP Granger Causality

- Causality is sharper than mere correlation! Some impacts confirmed, others not!
- ENSO4 “causes” SM in very dry (Sahel) and very wet (tropical rain forests)

"Dominant Features of Global Surface Soil Moisture Variability Observed by the SMOS Satellite" M. Piles et al. Remote Sensing, 2019
Conclusions

- Machine learning in EO and climate
  - Many techniques ready to use
  - Huge community, exciting tools

- Solid mathematical framework to deal with
  - Multivariate data
  - Multisource data
  - Structured spatio-temporal relations
  - Nonlinear feature relations

- Risks & remedies
  - Understanding is more complex
  - Physics consistency a must
  - Physics-driven ML & Explainable AI
“All models are wrong, but some are useful.”

George E. P. Box
Advances in Gaussian Processes for Earth Sciences

Physics-aware, interpretability and consistency

Gustau Camps-Valls, Daniel Svendsen, Luca Martino, Adrian Pérez, Maria Piles, and Jordi Muñoz

Image Processing Lab (IPL) - Universitat de València

http://isp.uv.es  @isp_uv_es  gustau.camps@uv.es