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

By Gustau Camps-Valls et al. 2020

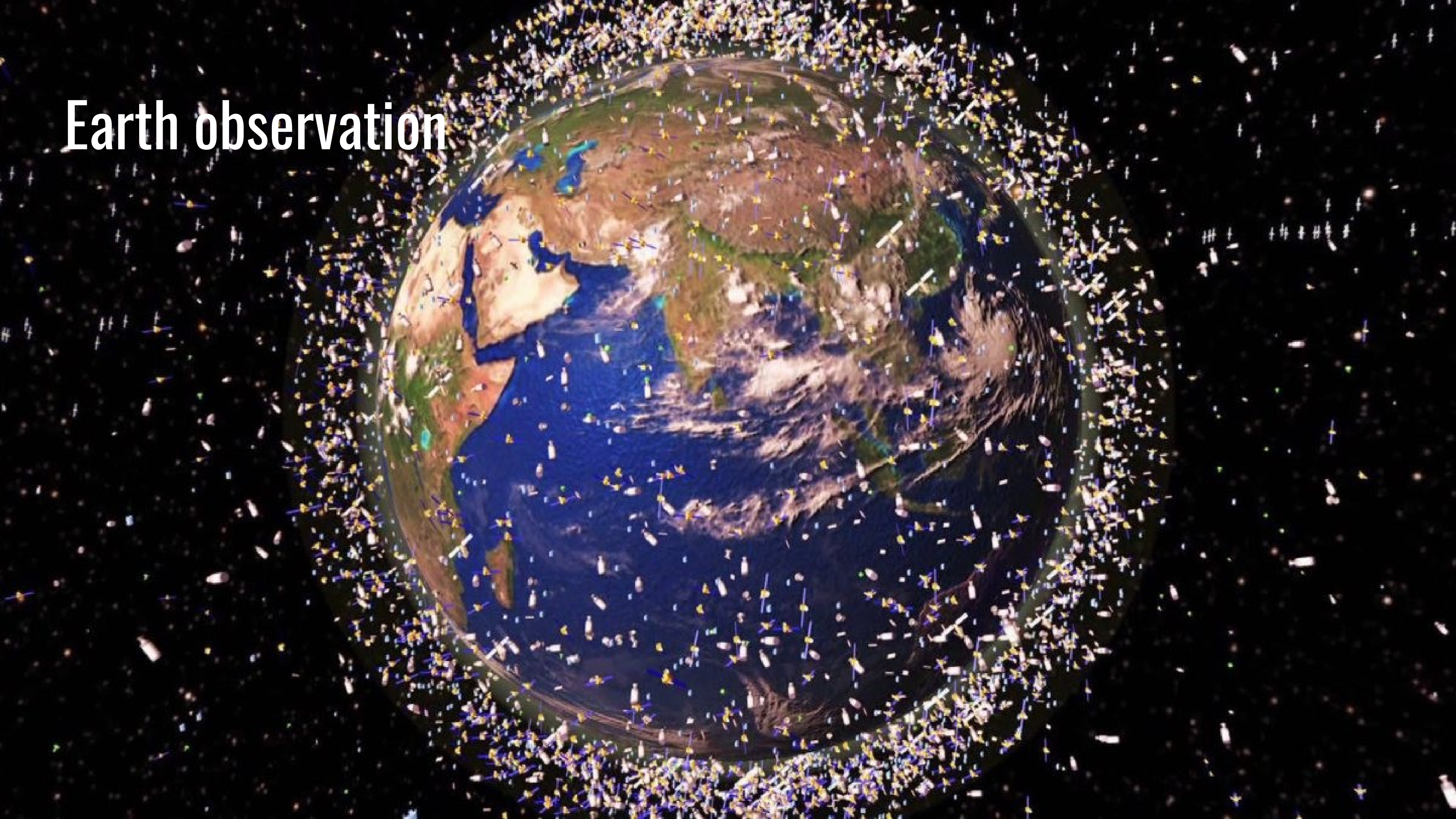
@isp_uv_es

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

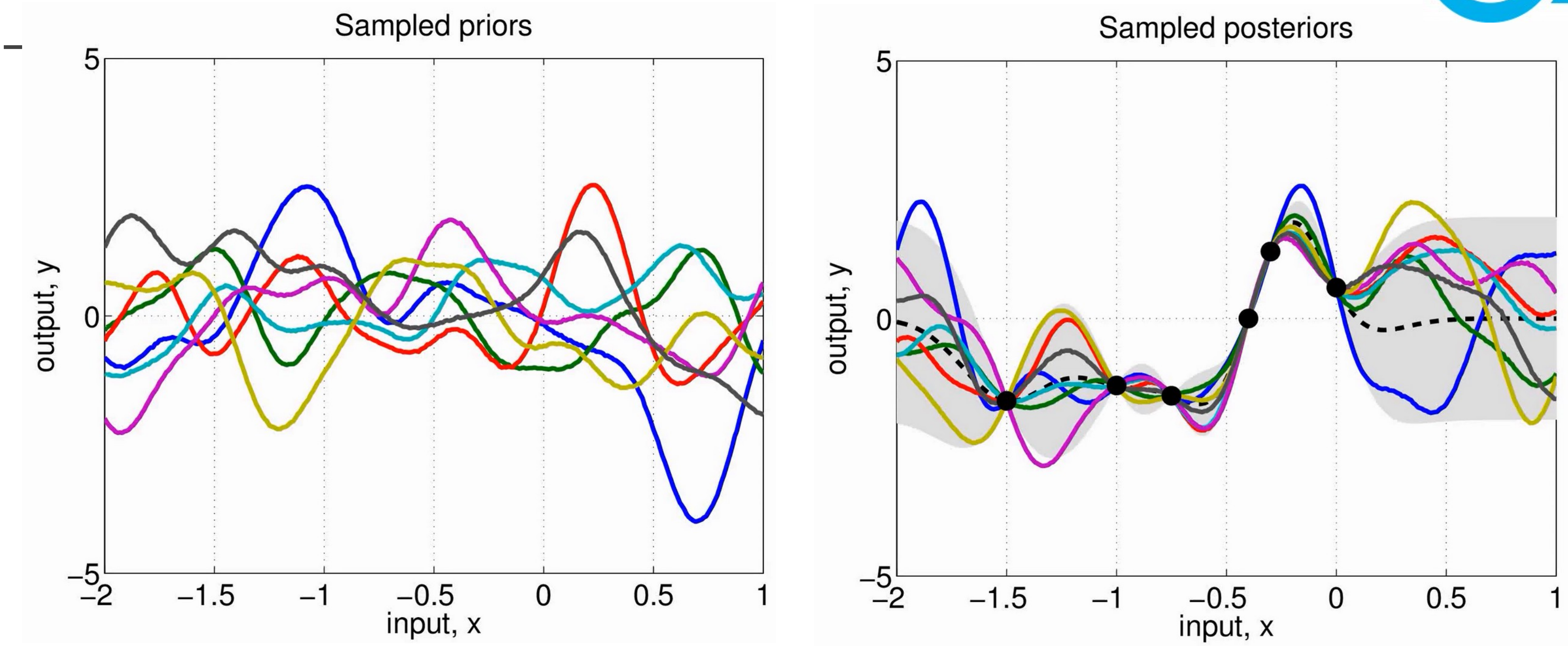
$$F(X) = y$$

- 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"
Camps-Valls, G. and Verrelst, J. and Muñoz-Marí, et al. IEEE Geoscience and Remote Sensing Magazine 2016
" 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: $\mathcal{D} = \{\mathbf{x}_n \in \mathbb{R}^D, y_n\}_{n=1}^N$
- Observation model:

$$y_n = f(\mathbf{x}_n) + \varepsilon_n, \;\; \varepsilon_n \sim \mathcal{N}(0, \sigma_n^2)$$

- Test point $\mathbf{x}_*$ with corresponding output y_*
- Posterior over the unknown output:

$$\begin{aligned} p(y_*|\mathbf{x}_*, \mathcal{D}) &= \mathcal{N}(y_*|\mu_{\mathrm{GP}*}, \sigma^2_{\mathrm{GP}*}) \\ \mu_{\mathrm{GP}*} &= \mathbf{k}_{\mathbf{f}*}^\top(\mathbf{K} + \sigma_n^2 \mathbf{I}_N)^{-1}\mathbf{y} \\ \sigma^2_{\mathrm{GP}*} &= \sigma_n^2 + k_{**} - \mathbf{k}_{\mathbf{f}*}^\top(\mathbf{K} + \sigma_n^2 \mathbf{I}_N)^{-1}\mathbf{k}_{\mathbf{f}*} \end{aligned}$$

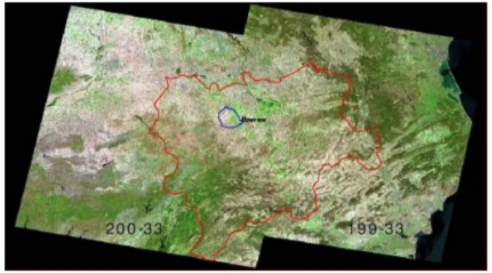
- Marginal likelihood optimization:

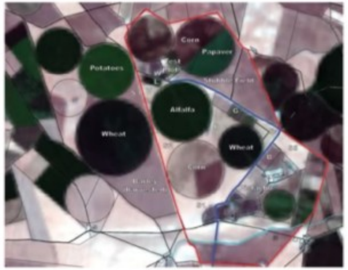
$$\log p(\mathbf{y}) = -\frac{1}{2}\mathbf{y}^\top(\mathbf{K} + \sigma_n^2 \mathbf{I}_N)^{-1}\mathbf{y} - \frac{1}{2}\log|\mathbf{K} + \sigma_n^2 \mathbf{I}_N| - \frac{N}{2}\log(2\pi)$$

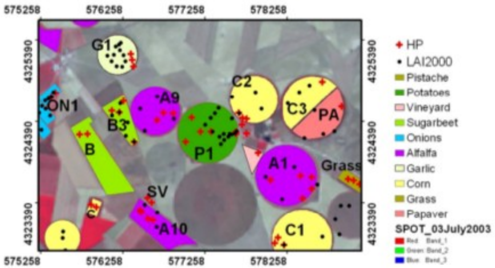
Standard GP models for parameter retrieval

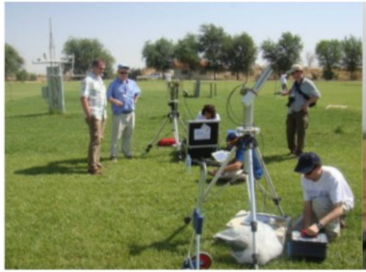
Observations, x: CHRIS images: 62 bands, 400-1050 nm, 34m

Variables, y: *In situ* leaf-level *Chl* (CCM-200) and LAI (PocketLAI phone app!)



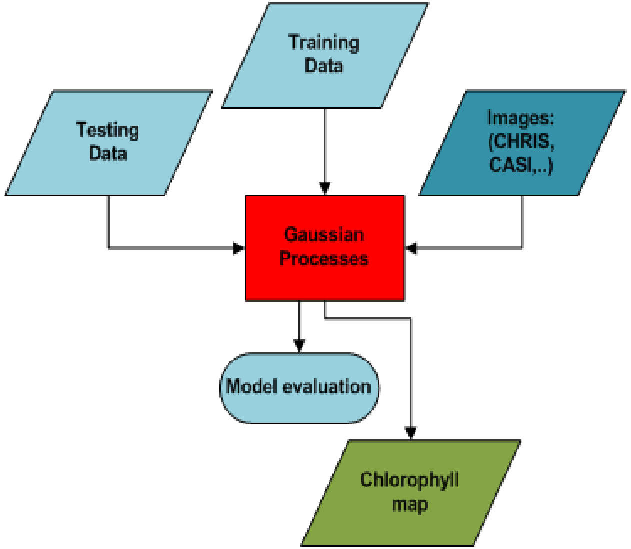


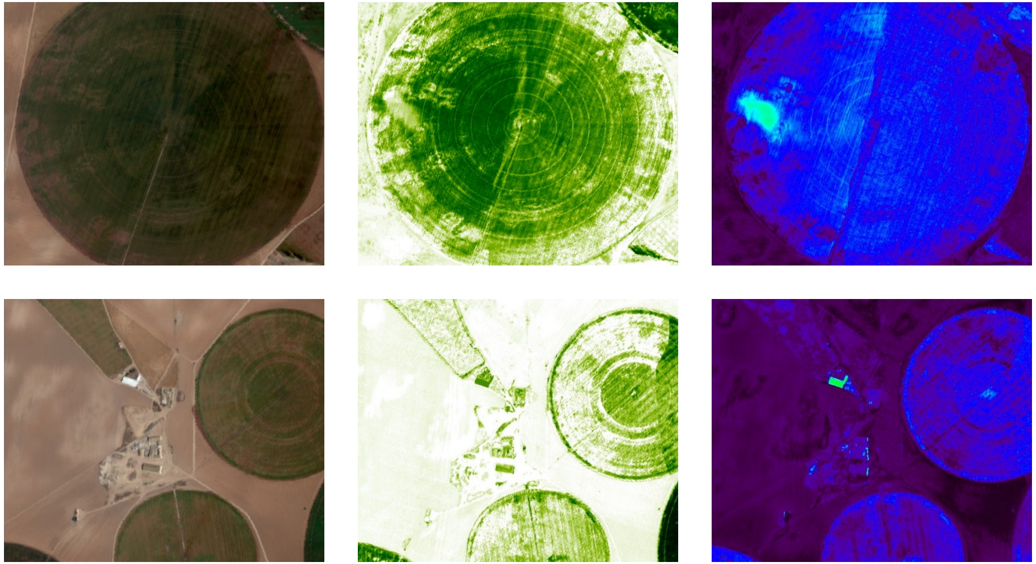




Standard GP models for parameter retrieval

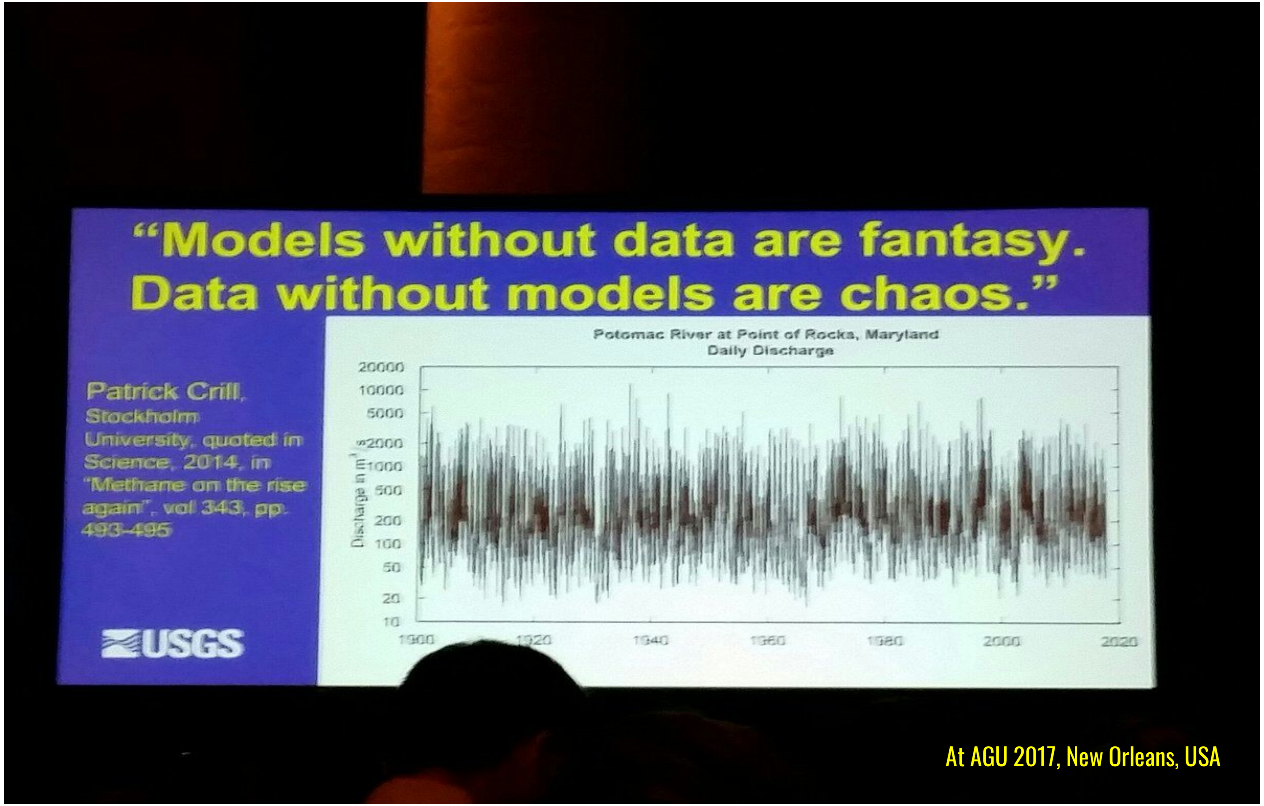
- Vegetation parameters from remote sensing data: chlorophyll content, LAI, vegetation cover



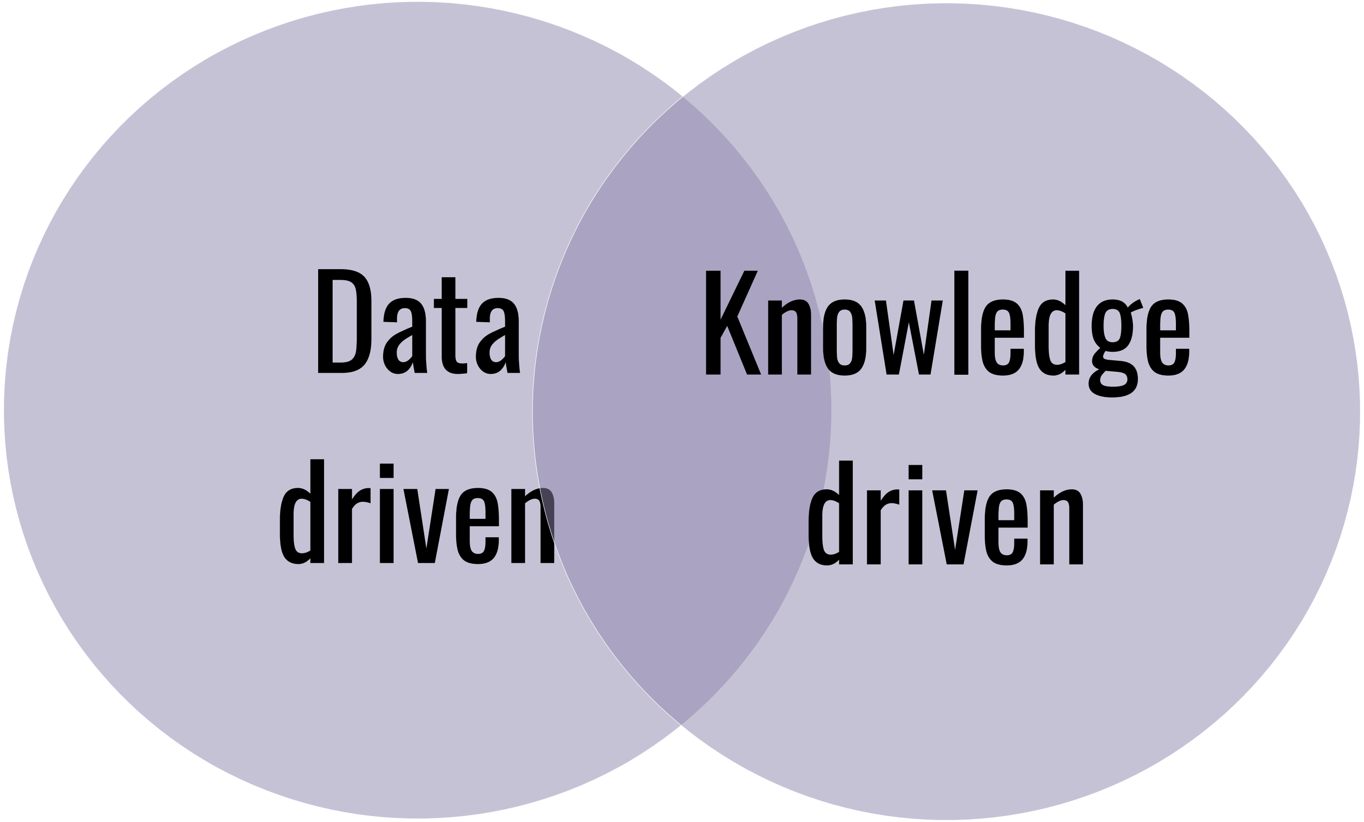


"A Survey on Gaussian Processes for Earth Observation Data Analysis: A Comprehensive Investigation"
Camps-Valls, G. and Verrelst, J. and Muñoz-Marí, et al. IEEE Geoscience and Remote Sensing Magazine 2016

The truth is that...



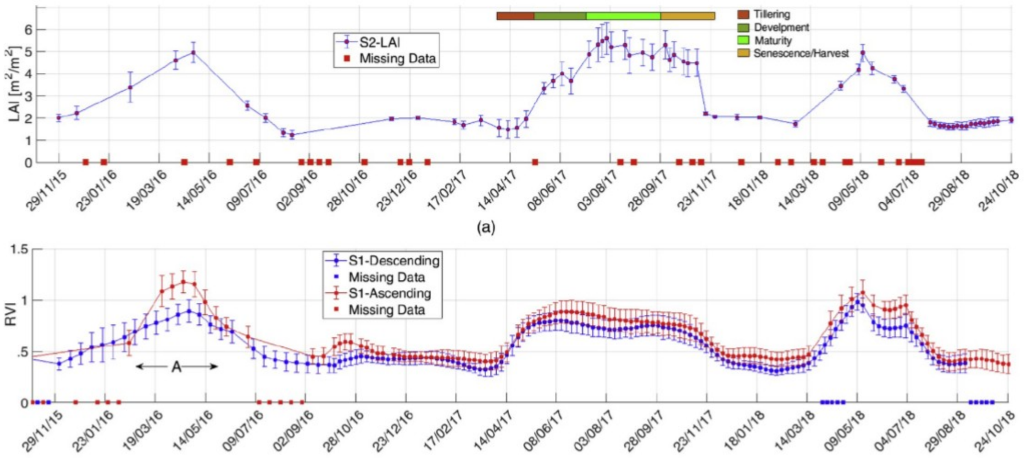
Hybrid modeling, where knowledge lies …



1: GP for multisource fusion

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

$$\mathbf{f}(t) = \begin{bmatrix} f_{PS1}(t) \\ f_{PS2}(t) \end{bmatrix} = \begin{bmatrix} a^1_{1,1}u^1_1(t) + a^1_{1,2}u^1_2(t) \\ a^1_{2,1}u^1_1(t) + a^1_{2,2}u^1_2(t) \end{bmatrix}$$



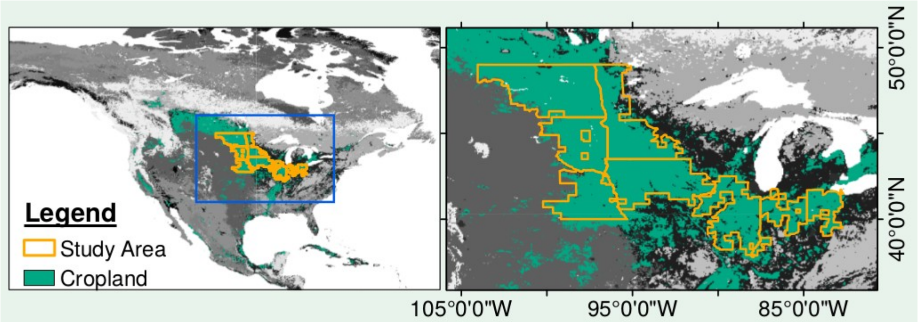
“**Fusing Optical and SAR time series for LAI gap filling with multioutput Gaussian processes**”
Luca Pipia and Jordi Muñoz-Marí and Eatidal Amin and Santiago Belda and Gustau Camps-Valls and Jochem Verrelst
Remote Sensing of Environment 235 :111452, 2019

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

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

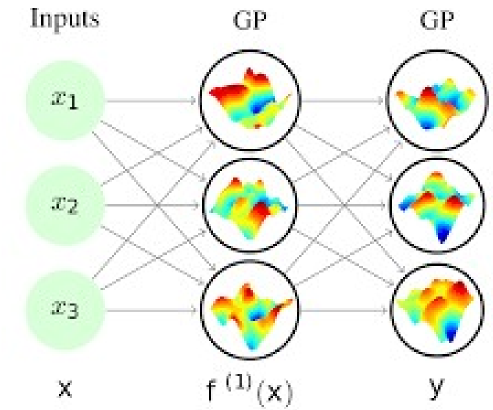
$$\langle \mu_k(\mathcal{P}), \mu_k(\mathcal{Q}) \rangle_{\mathcal{H}_k} = \mathbb{E}_{X \sim \mathcal{P}, Y \sim \mathcal{Q}} k(X, Y)$$

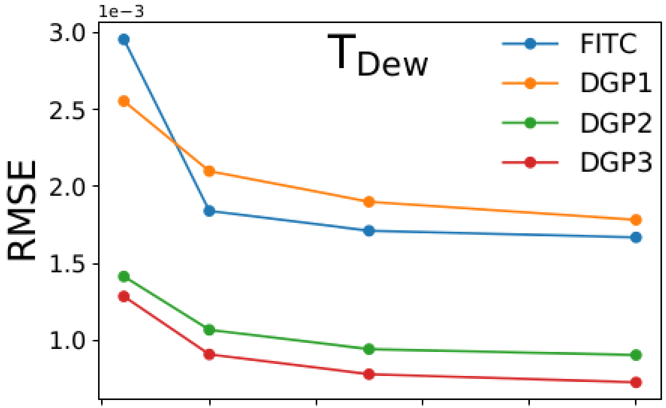


“Nonlinear Distribution Regression for Remote Sensing Applications”
Adsuara, Perez,Muñoz, Mateo, Piles, Camps-Valls, IEEE TGARS 2019

3: Deep Gaussian Processes

- GP also goes deep; improved versatility and expressive power
- Excellent performance in emulation and atmospheric parametrization



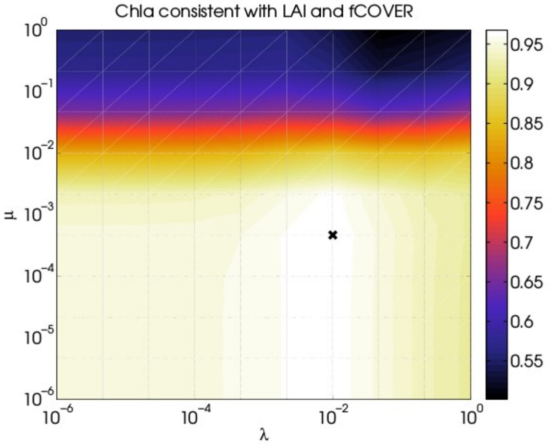


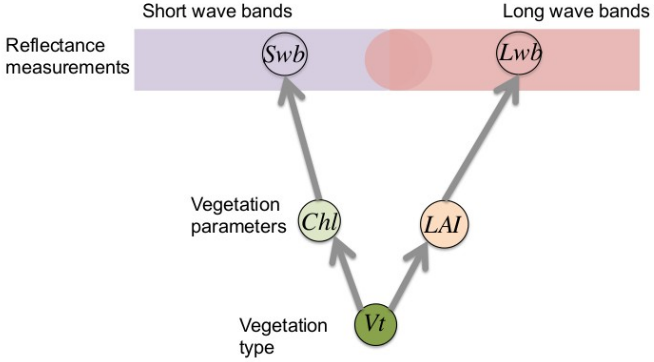
"Deep Gaussian Processes for Retrieval of bio-geo-physical parameters",
Svendsen, Molina, Ruescas and Camps-Valls, ISPRS, 2020

4: Physics-aware GPs with constrained optimization

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

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



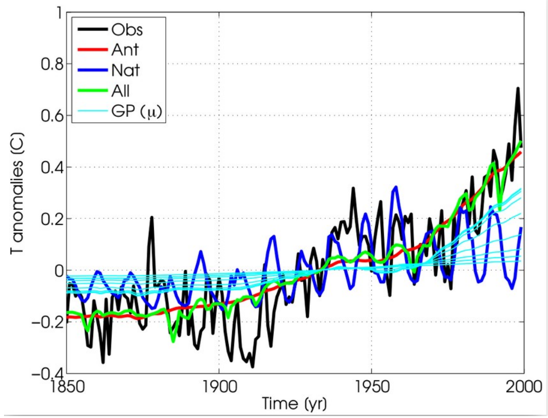


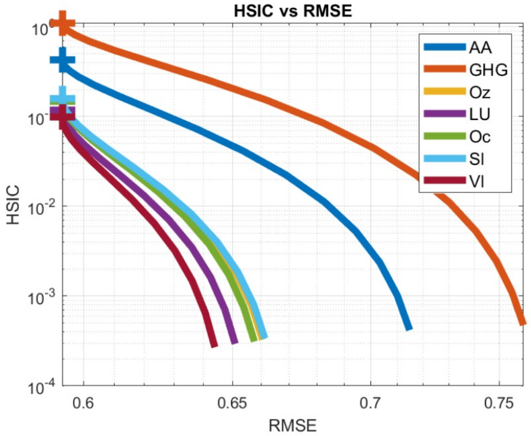
"Fair Kernel Learning" Perez, Laparra, Gomez, Camps-Valls, G. ECML, 2017.
"Consistent Regression of Biophysical Parameters with Kernel Methods" Díaz, Peréz-Suay, Laparra, Camps-Valls, IGARSS 2018
"Physics-aware Gaussian processes in remote sensing" Camps-Valls, G, Svendsen, D, Martino, L. et al, Applied Soft Computing 68 :69-82, 2018

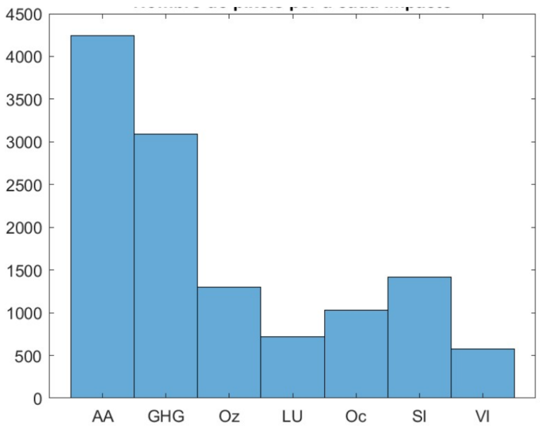
4: Constrained optimization

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

$$\mathrm{FairLoss} = \mathrm{Cost}(y, \hat{y}) + \lambda_1 \|w\|_2^2 + \gamma I(\hat{y}, s)$$







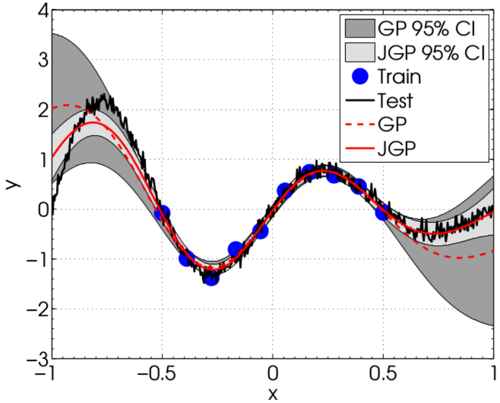
"Fair Kernel Learning" Perez, Laparra, Gomez, Camps-Valls, G. ECML, 2017.
"Consistent Regression of Biophysical Parameters with Kernel Methods" Díaz, Peréz-Suay, Laparra, Camps-Valls, IGARSS 2018
"Physics-aware Gaussian processes in remote sensing" Camps-Valls, G, Svendsen, D, Martino, L. et al, Applied Soft Computing 68 :69-82, 2018

5: Joint GP models for forward-inverse modeling

- Let ML talk to physical models

$$\text{JointLoss} = \text{Cost}(y, \hat{y}) + \lambda_1 \|w\|_2^2 + \gamma \Omega(\hat{y}, \Phi)$$

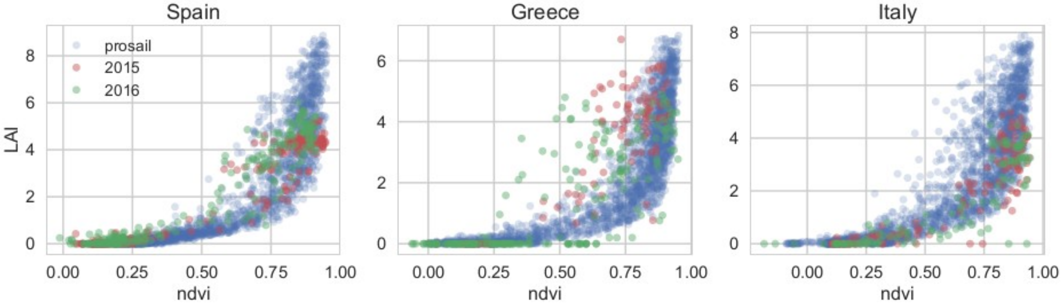
$$\Omega(\hat{y}, \Phi) = \text{Cost}_s(y_s, \hat{y}_s)$$



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)

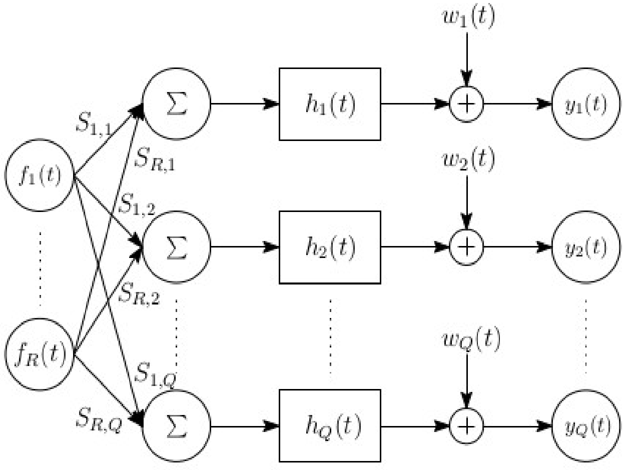
Filling the space ...



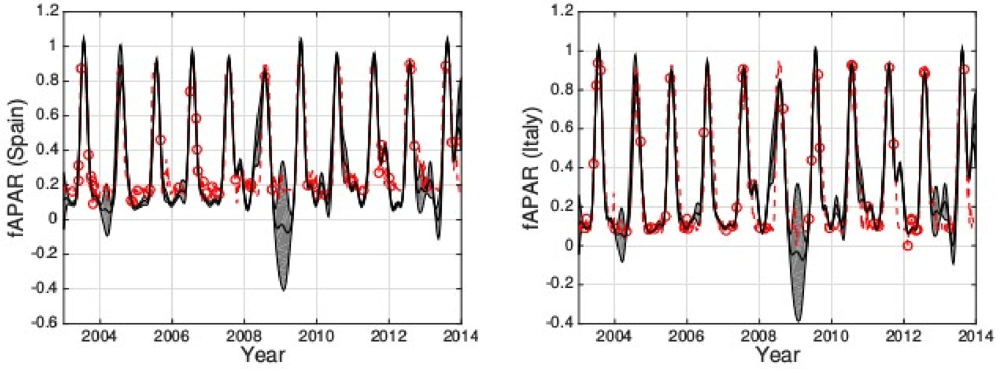
"Joint Gaussian Processes for Biophysical Parameter Retrieval"
Svendsen, Martino, Camps-Valls, IEEE TGARS 2018
"Physics-aware Gaussian processes in remote sensing"
Camps-Valls, G. et al. Applied Soft Computing, 2018.

6: Multioutput GP regression encoding ODEs

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



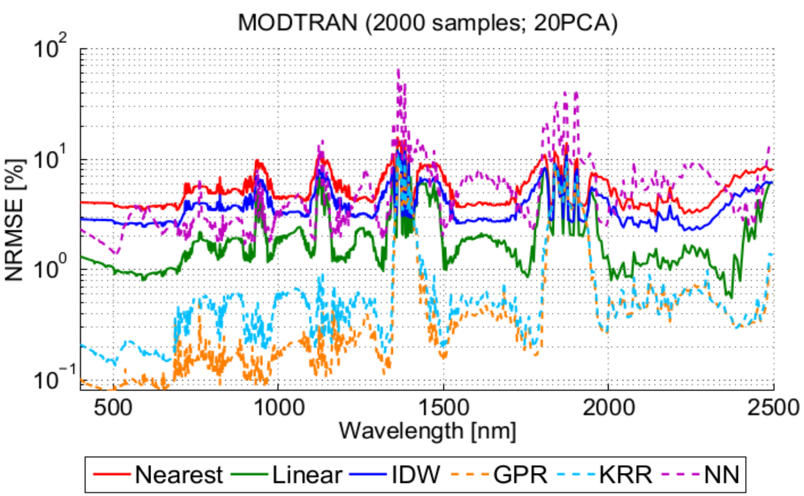
- LAI and fAPAR data for Spain and Italy ($Q = 4$ outputs)
- Multioutput improves single output GPs (4.5% gain in MSE)
- Transportability across time/space of estimates

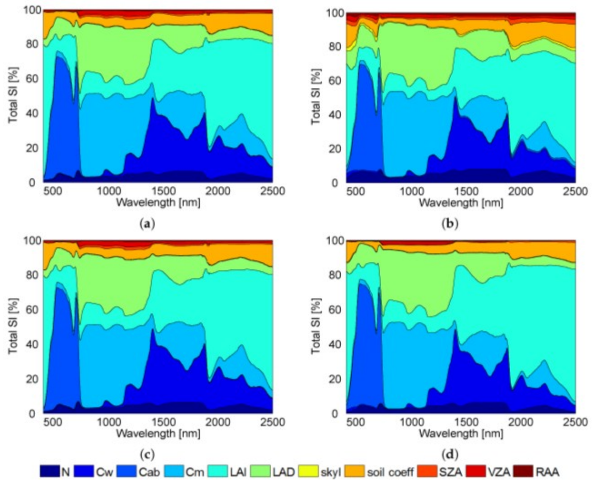


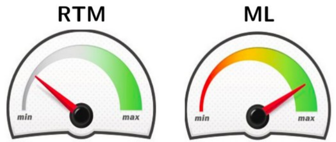
“Gap filling of biophysical parameters with multi-output GPs”
Mateo, Camps-Valls et al, IEEE IGARSS. 2018.
“Learning latent forces from Earth time series”
Svendsen, Muñoz, Piles, Camps-Valls,, subm. 2020

7: GP emulation of complex codes

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







RTM	ML
2.3–24.5 s/pix	0.1–1.3 ms/pix
0%	RMSE = 0.1 – 5%

"Emulation of Leaf, Canopy and Atmosphere Radiative Transfer Models for Fast Global Sensitivity Analysis",
Verrelst, Camps-Valls et al Remote Sensing of Environment, 2016
"Emulation as an accurate alternative to interpolation in sampling radiative transfer codes",
Vicent and Camps-Valls, IEEE Journal Sel. Topics Rem. Sens, Apps. 2018

8: GP for uncertainty propagation

- 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 $\mathbf{x}_*$ changes as

$$v_{eGP}^2 = T_{**} + k_{**} - \mathbf{k}_*^\top \left(\mathbf{K} + \sigma_y^2 \mathbf{I}_N + \mathbf{T}\right)^{-1} \mathbf{k}_* \quad (4)$$

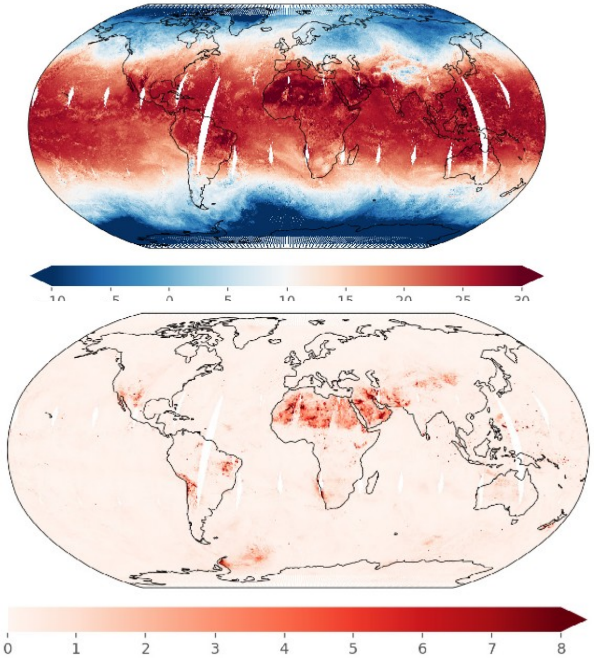
where the effect of the noise in the inputs is represented by $\mathbf{T}_{ij} = T(\mathbf{x}_i, \mathbf{x}_j) = \partial_i^\top \boldsymbol{\Sigma}_x \partial_j$, and $T_{**} = T(\mathbf{x}_*, \mathbf{x}_*)$. We denoted the vector of partial derivatives of f with respect to the sample x_i as:

$$\partial_i := \left[\frac{\partial f(\mathbf{x}_i)}{\partial x_i^1} \cdots \frac{\partial f(\mathbf{x}_i)}{\partial x_i^D}\right]^\top.$$

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 $\boldsymbol{\alpha}$ parameters

$$\frac{\partial f(\mathbf{x}_i)}{\partial x_i^j} = \frac{\partial \mathbf{k}_i \boldsymbol{\alpha}}{\partial x_i^j} = (\partial \mathbf{k}_{ij})^\top \boldsymbol{\alpha}$$

where $\partial \mathbf{k}_{ij} = [(\partial k(\mathbf{x}_i, \mathbf{x}_1)/\partial x_i^j), \ldots, (\partial k(\mathbf{x}_i, \mathbf{x}_N)/\partial x_i^j)]^\top$. Please see[1] for a working implementation of the error GP (eGP) model.



"Accounting for Input Noise in Gaussian Process Parameter Retrieval"
Johnson, J. E. and Laparra, V. and Camps-Valls, G. IEEE GRSL 17 (3) :391-395, 2020

9: GP interpretability

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

- **Look at the feature lengthscales ...**

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

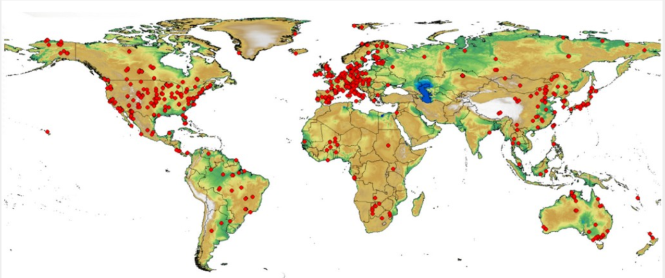
- **Compute feature sensitivity ...** Derivatives of $\mu(\mathbf{x})$ wrt features

$$s_f = \int \left(\frac{\partial \mu(\mathbf{x})}{\partial x^f} \right)^2 p(\mathbf{x}) \mathrm{d}\mathbf{x} \approx \frac{1}{N} \sum_{q=1}^{N} \left(\sum_{p=1}^{N} \frac{\alpha_p (x_p^f - x_q^f)}{\sigma_f^2} k(\mathbf{x}_p, \mathbf{x}_q) \right)^2$$

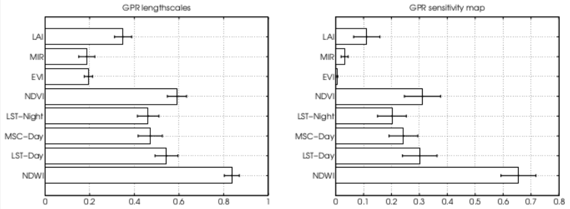
- **Sensitivity maps yield ...**
  - Fast feature ranking and sensitivity analysis
  - Example relevance and informed PDF sampling

"Ranking drivers of global carbon and energy fluxes over land"
Camps-Valls, G. Jung, M. Ichii, K. Papale, D. Tramontana, G. Bodesheim, P. Schwalm, C. Zscheischler, J. Mahecha, M. Reichstein, M. IEEE IGARSS 2015

1: GPP estimation from FLUXNET



	ME	RMSE	MAE	ρ
LR	-0.01	1.83	1.30	0.78
MLP	+0.04	1.92	1.39	0.73
SVR	**+0.01**	1.80	1.23	0.78
GPR	+0.03	**1.76**	**1.16**	**0.80**



9: GP interpretability

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

- **Look at the feature lengthscales ...**

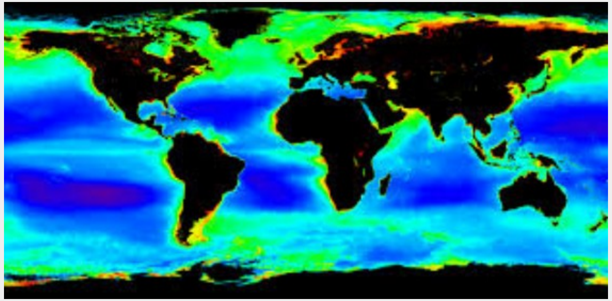
$$k(\mathbf{x}_i, \mathbf{x}_j) = \nu \exp\left(- \sum_{f=1}^{F} \frac{(x_i^f - x_j^f)^2}{2{\sigma_f}^2} \right) + \sigma_n^2 \delta_{ij}$$

- **Compute feature sensitivity ...** Derivatives of $\mu(\mathbf{x})$ wrt features

$$s_f = \int \left(\frac{\partial \mu(\mathbf{x})}{\partial x^f} \right)^2 p(\mathbf{x}) d\mathbf{x} \approx \frac{1}{N} \sum_{q=1}^{N} \left(\sum_{p=1}^{N} \frac{\alpha_p (x_p^f - x_q^f)}{\sigma_f^2} k(\mathbf{x}_p, \mathbf{x}_q) \right)^2$$

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

2: ocean chlorophyll estimation



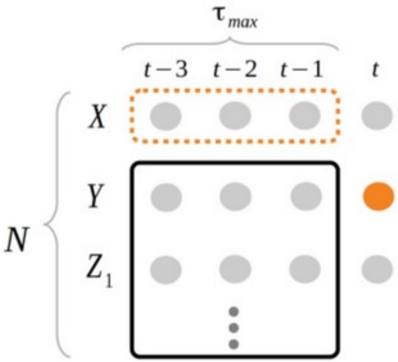
Database	ME	RMSE	MAE	ρ
SeaBAM (2, 4 and 5)	+0.0037	0.1493	0.1104	0.9679
SeaWIFS (4, 5 and 6)	−0.0887	0.3149	0.2361	0.9236
MODIS-Aqua (4, 5 and 6)	+0.0229	0.2461	0.1866	0.9188
MERIS (synthetic) (5, 6, 7 and 8)	0.004	0.084	0.0232	0.9996
MERIS (real) (5, 6, 7 and 8)	$< 10^{-7}$	0.21	0.1464	0.9261

- 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

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

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



$$a_H = (K(X_t, X_t') + \varepsilon_t^{Y})^{-1} Y_{t+1}$$

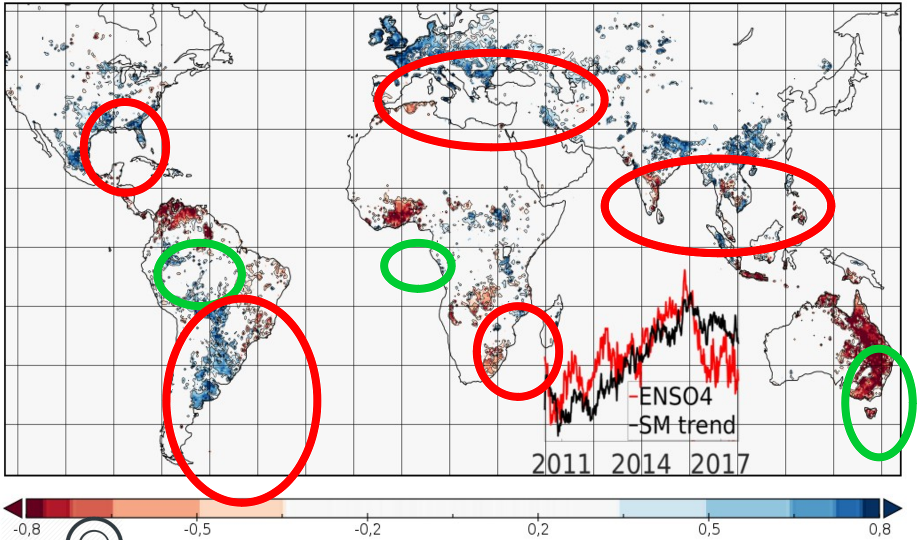
$$b_H = (L([Y_t, X_t], [Y_t', X_t']) + \varepsilon_t^{Y|X})^{-1} Y_{t+1}$$

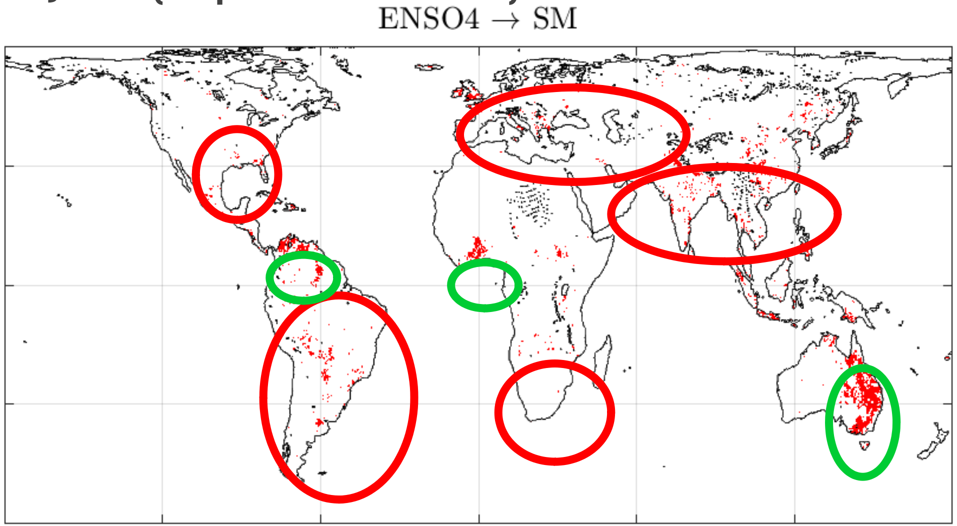
$$X \rightarrow Y \leftrightarrow \mathbb{V}_H[\varepsilon_t^{Y}] \ll \mathbb{V}_H[\varepsilon_t^{Y|X}]$$

“Revisiting impact of MJO on soil moisture:a causality perspective”, AGU 2019
“Cross-Information Kernel Causality: Revisiting global teleconnections of ENSO over soil moisture and vegetation” (2018). Climate Informatics: CI 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)

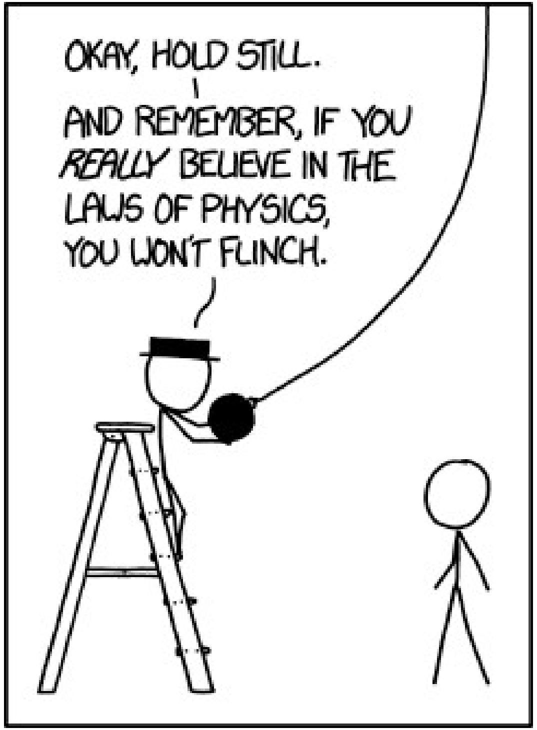


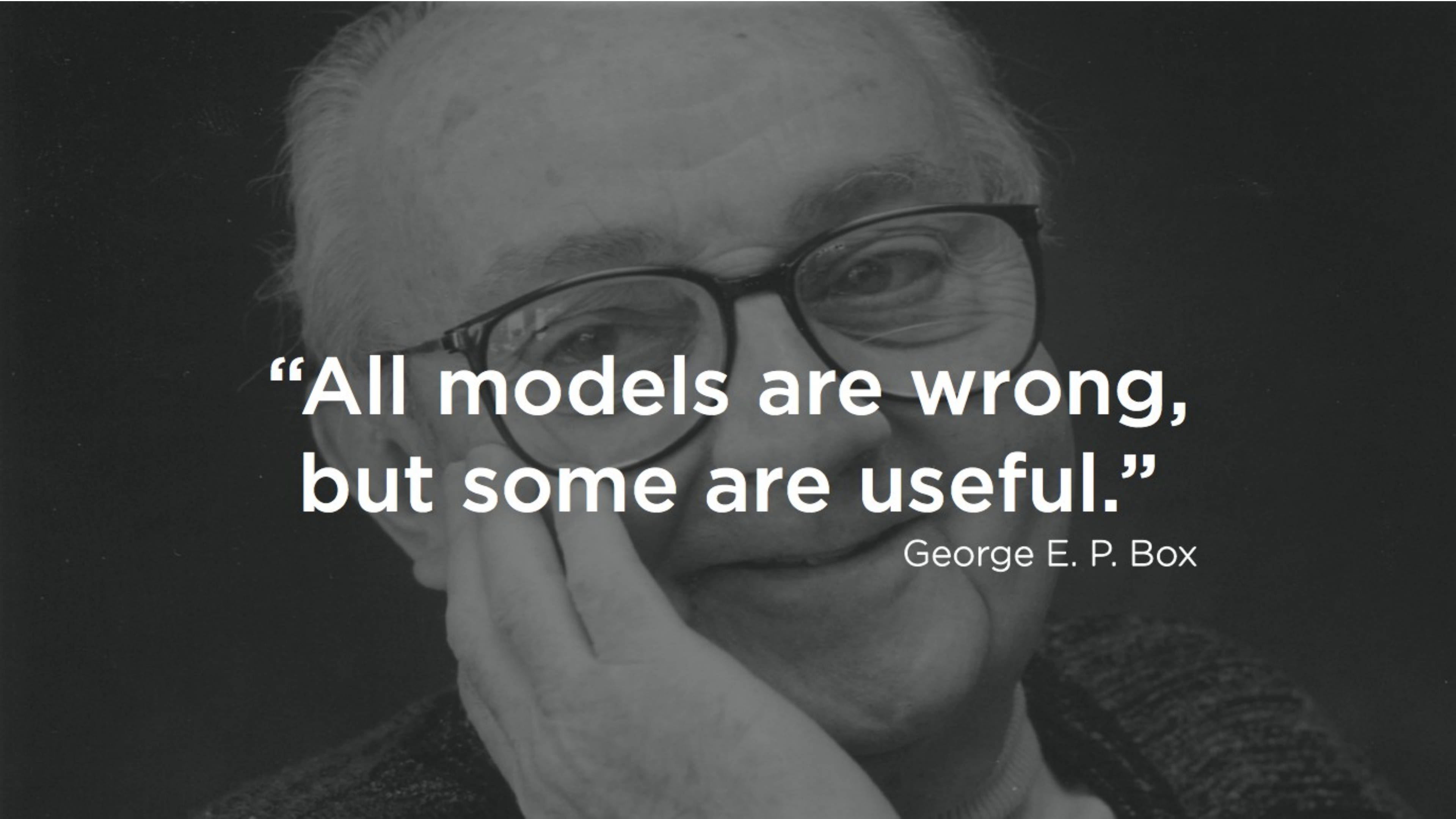


“Dominant Features of Global Surface Soil Moisture Variability Observed by the SMOS Satellite” M. Piles et al. Remote Sensing, 2019

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