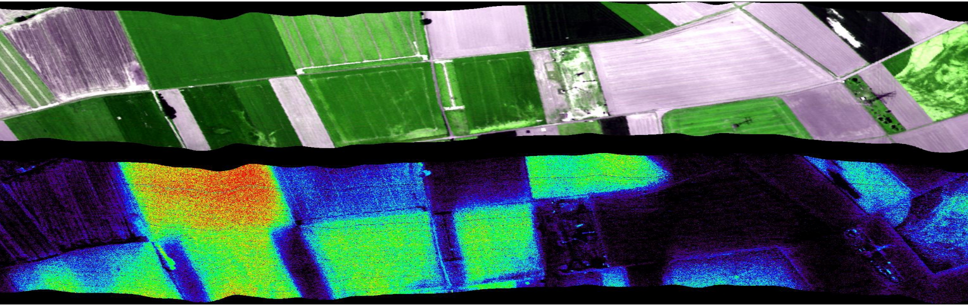
Monitoring Earth Climate Variables with Statistical Inference



Gustau Camps-Valls
Image Processing Laboratory (IPL)
Universitat de València, Spain
✉ gustau.camps@uv.es | http://isp.uv.es

Outline

- **Part 1: Earth observation from space**
- **Part 2: EO needs statistical learning**
- **Part 3: Advances in variable prediction and understanding**
  1. Temporal structures
  2. Learning the covariance
  3. Warping the function
  4. Heteroscedasticity
  5. Multi-output regression
  6. Computational efficiency

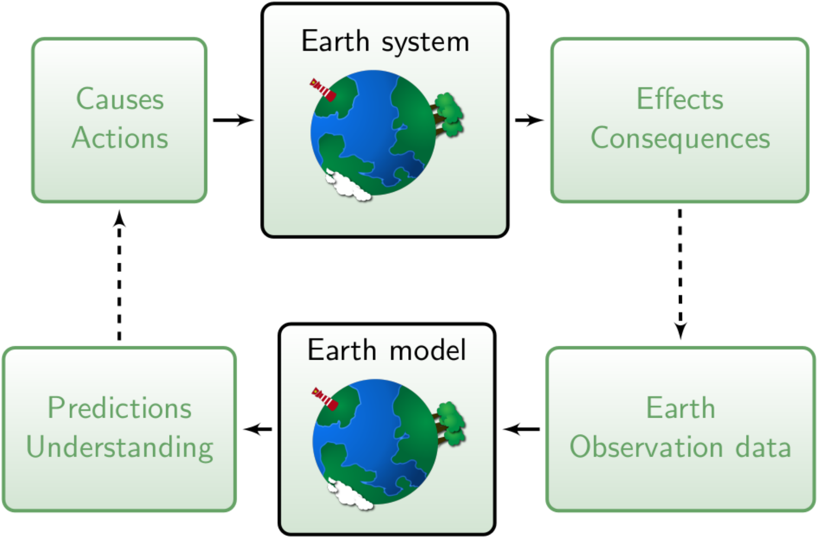
  . . . .

  7. Sensitivity analysis
  8. From regression to causation

  . . . .

  9. GPs in your smartphone
  10. GPs for weather forecasting
- **Part 4: Conclusions: references, code, ads**

Part 1: Earth observation from space

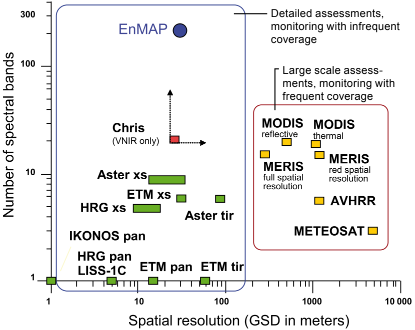

Causes
Actions
Earth system
Effects
Consequences
Earth
Observation data
Earth model
Predictions
Understanding

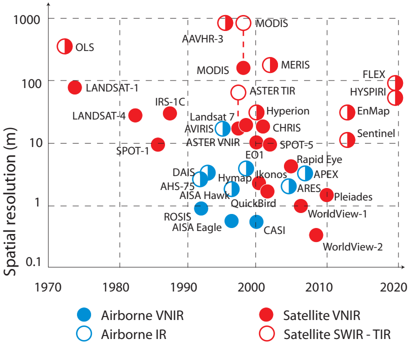
We monitor the Earth constantly ...

NASA Exploration Center, California

EUMETSAT Operations, Darmstadt

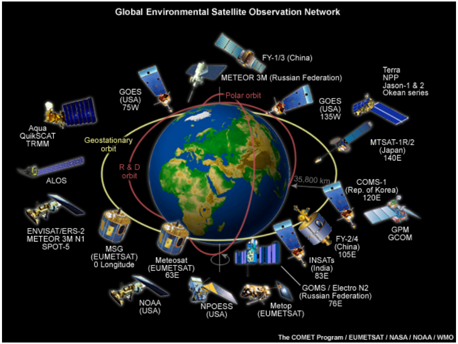
… and with many satellite sensors …



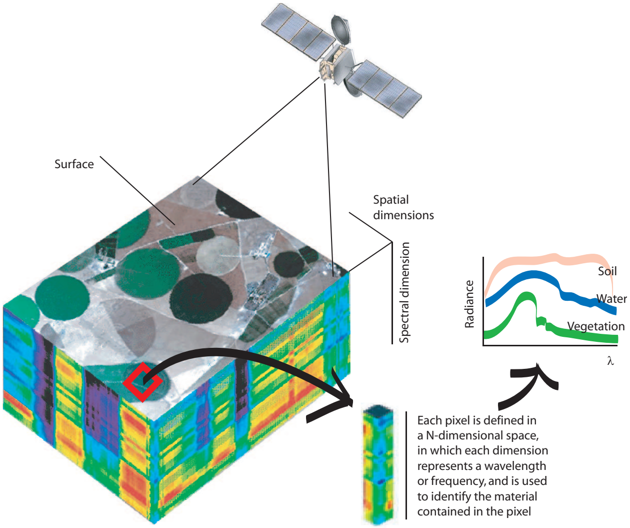


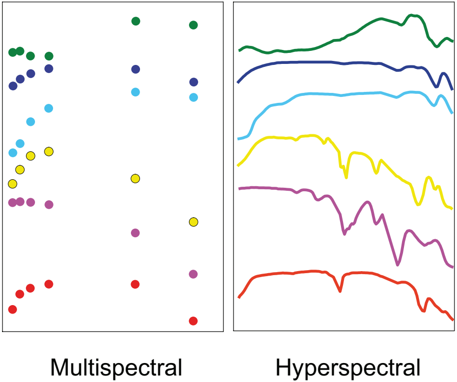
Credits: http://www.enmap.de/

... and many more almost ready to come!



ESA Sentinels, NASA A-train, EnMAP, FLEX, MTG-IRS, HyspIRI, MEOS, ZASat, HIS, HERO, etc.


Surface
Spatial dimensions
Spectral dimension
Radiance
Soil
Water
Vegetation
λ
Each pixel is defined in a N-dimensional space, in which each dimension represents a wavelength or frequency, and is used to identify the material contained in the pixel



- Hyperspectral signals allow finer material characterization
- Absorption, depth, re-emissions and modulated spectral features
- Identification of chemical components and bio-chemical processes
- Estimation of bio-geo-physical parameters

What can we do with these signals?

Geology

- Mineral detection
- Cover homogeneity

Forestry

- Infected trees
- Status monitoring
- Forest clearing

Sea/ice/coastal

- Oil spills monitoring
- Water quality

Precision agriculture

- Crop stress location
- Crop productivity

Atmosphere

- Air quality, pollutants
- Global/local change

Land management

- Crop monitoring/phenology
- Land use/cover change

Defense

- Target detection
- Mine detection

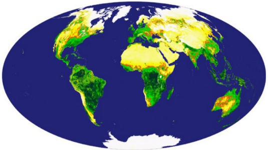
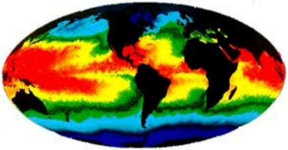
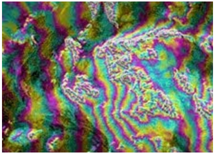
Public safety

- Logistics & operations
- Fire risk, floods

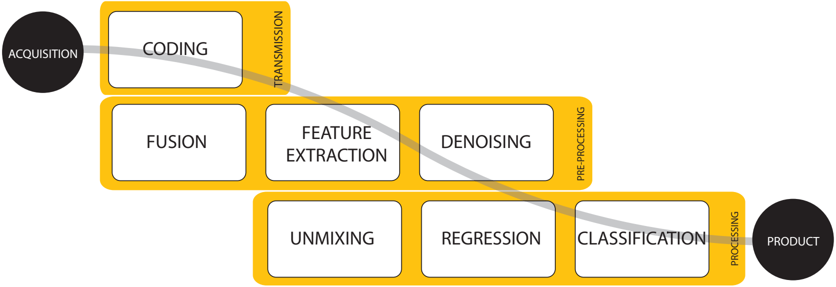
Regulation & Policy making

- Urban growth
- Settlements, population movements

- Visible / NIR / MIR – only day, no clouds
  - Vegetation dynamics and covers
  - Geological maps (structure, mineralogy, oil)
  - Agriculture, urban, forests use
  - Ocean temperature, fitoplancton
  - Meteorology of clouds
  - Layer dynamics
- Thermal Infrared – day and night
  - Clouds temperature and height
  - Forest fires
  - Temperature maps
  - Plumes
  - Heat islands
- Active microwaves – day and night
  - Surface characterization, structure of trees and leaves
  - Moisture, deformation by earthquakes, ...
  - Military applications

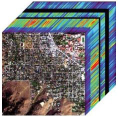




A standard image processing chain:



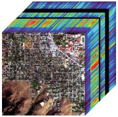
Characteristics of remote sensing data

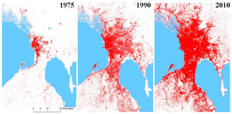
- High spectral resolution $\rightarrow$ moderate spatial resolutions

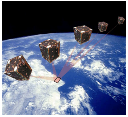


Characteristics of remote sensing data

- High spectral resolution $\rightarrow$ moderate spatial resolutions
- High dimensional: multi-temporal, angular and source

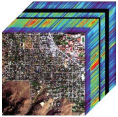


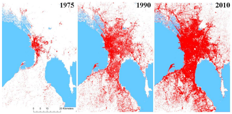


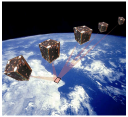


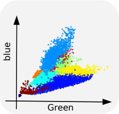
Characteristics of remote sensing data

- High spectral resolution $\rightarrow$ moderate spatial resolutions
- High dimensional: multi-temporal, angular and source
- Non-linear feature relations



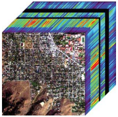


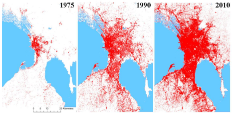


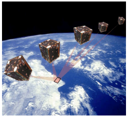


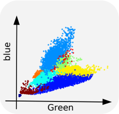
Characteristics of remote sensing data

- High spectral resolution → moderate spatial resolutions
- High dimensional: multi-temporal, angular and source
- Non-linear feature relations
- Non-Gaussian data distributions



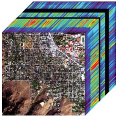


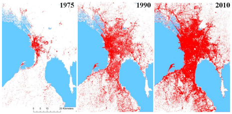


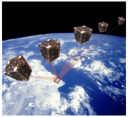


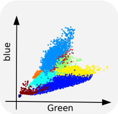
Characteristics of remote sensing data

- High spectral resolution $\rightarrow$ moderate spatial resolutions
- High dimensional: multi-temporal, angular and source
- Non-linear feature relations
- Non-Gaussian data distributions
- Dependent noise, uneven sampling



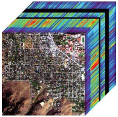
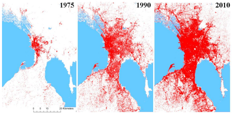


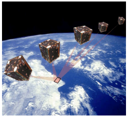
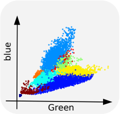




Characteristics of remote sensing data

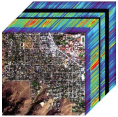
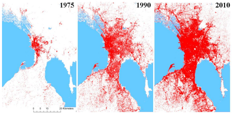
- High spectral resolution $\rightarrow$ moderate spatial resolutions
- High dimensional: multi-temporal, angular and source
- Non-linear feature relations
- Non-Gaussian data distributions
- Dependent noise, uneven sampling
- Few supervised (labeled) information is available (high cost)

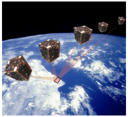
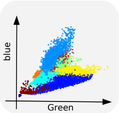



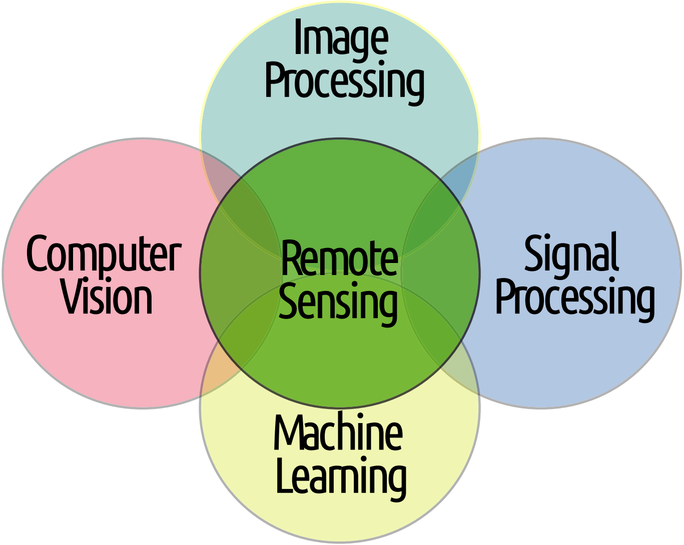
Characteristics of remote sensing data

- High spectral resolution $\rightarrow$ moderate spatial resolutions
- High dimensional: multi-temporal, angular and source
- Non-linear feature relations
- Non-Gaussian data distributions
- Dependent noise, uneven sampling
- Few supervised (labeled) information is available (high cost)
- Tons of data to process in (near) real-time

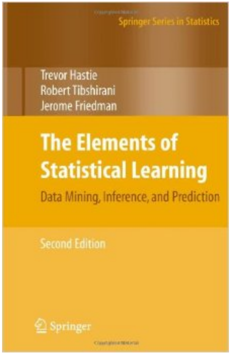




We live at the intersection ...



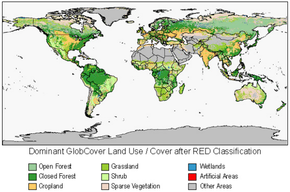
Part 2: Statistical learning



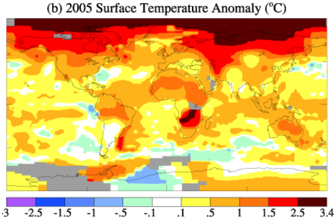
- Statistical learning theory is a framework for machine learning drawing from the fields of statistics and functional analysis
- Statistical learning theory deals with the problem of finding a predictive function based on data
- Given a set of input-output pairs $\mathcal{D} = \{(\mathbf{x}_i, y_i) | i = 1, \ldots, N\}$, learn a function $f(\cdot)$ that predicts outputs for new inputs well, $y^* = f(\mathbf{x}^*)$

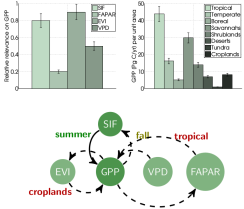
Earth Observation needs Statistical Learning

1. Identify objects and detect changes
2. Estimate the content of bio-geo-physical prameters
3. Assess relative relevance of variables
4. Infer (causal) relations between variables and observations

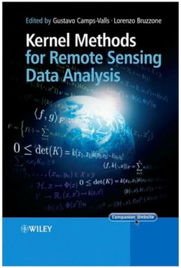


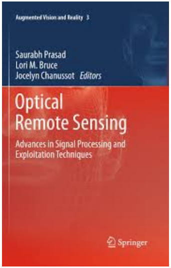
Dominant GlobCover Land Use / Cover after RED Classification

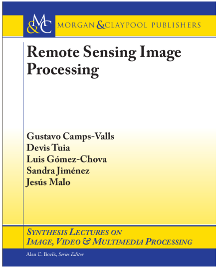




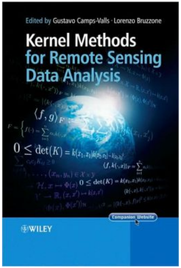
Last decade was dominated by kernel methods ...

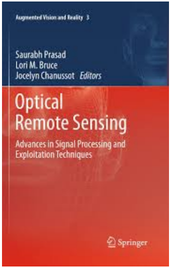


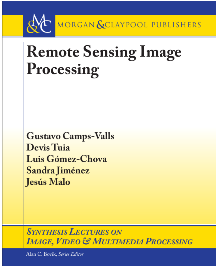




Last decade was dominated by kernel methods ...

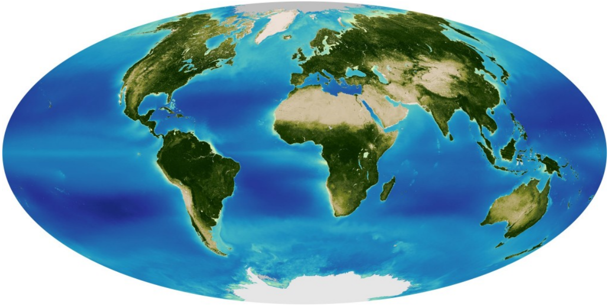




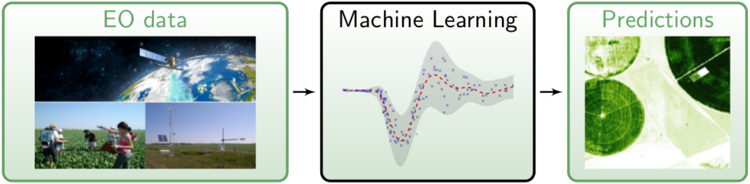


- RS problems are typically nonlinear
- KMs are efficient in high-dimensional low-sampled problems
- KMs allow combination of multi- source/temporal/angular data
- KMs are simple and clean to design, understand and apply
- We have 130 years of solid mathematics, theorems, and bounds

- Estimate biophysical parameters is key to monitor our Planet
- Monitoring land is the most challenging (and interesting) problem:
  - Phenological stage of crops and forests
  - Health status (e.g., development, productivity, stress)
- Implications on agriculture, biofuels and food
- Models typically resort to *in situ* data + remote sensing data



Goal: Transform measurements into biophysical parameter estimates



Data:

- **Input data:**
    - satellite/airborne spectra
    - *in situ* (field) radiometers
    - simulated spectra from RTMs
- **Output results:** estimation of a bio/geo-physical parameter

Taxonomy of retrieval methods

1. The *statistical* inversion models: parametric and non-parametric.
    - Parametric models rely on physical knowledge of the problem and build explicit parametrized expressions that relate a few spectral channels with the bio-geo-physical parameter(s) of interest.
    - Non-parametric models are adjusted to predict a variable of interest using a training dataset of input-output data pairs.
2. *Physical* inversion models: try to reverse RTMs.
    - After generating input-output (parameter-radiance) datasets, the problem reduces to, given new spectra, searching for similar spectra in the dataset and assigning the most plausible ('closest') parameter.
3. *Hybrid* inversion models try to combine the previous approaches.

- **Parametric models based on band ratios are typically used:**
  - Simple
  - Understandable
  - Fast
- **Problems:**
  - Too general and simplistic, not suited to all scenarios
  - Require prior information (and solid physical knowledge)
- **Nonlinear, nonparametric regression typically performs better:**
  - More flexible, adaptive
  - No assumptions about data relations
  - Many methods: neural networks, random forests, SVR ...

What is a Gaussian process?

- It is a probability density over functions
- It is defined by
  - A mean function $m(\mathbf{x})$
  - A covariance function $k(\mathbf{x}, \mathbf{x}')$
- It is expressed as

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}'))$$

- The joint pdf of any subset of points of $f(\mathbf{x})$ is a Gaussian:

$$\begin{bmatrix} f(\mathbf{x}_1) \\ \vdots \\ f(\mathbf{x}_N) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(\mathbf{x}_1) \\ \vdots \\ m(\mathbf{x}_N) \end{bmatrix}, \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \cdots & k(\mathbf{x}_1, \mathbf{x}_N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_N, \mathbf{x}_1) & \cdots & k(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix} \right)$$

Gaussian processes regression in a nutshell

- Input-output data: $\{\mathbf{x}_n \in \mathbb{R}^D, y_n\}_{n=1}^N$
- Observation model:

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

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

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

- Marginal likelihood (aka evidence)

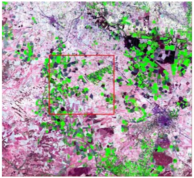
$$\log p(\mathbf{y}) = -\frac{1}{2}\mathbf{y}^\top (\mathbf{K} + \sigma^2 \mathbf{I}_n)^{-1} \mathbf{y} - \frac{1}{2}\log|\mathbf{K} + \sigma^2 \mathbf{I}_n| - \frac{N}{2}\log(2\pi)$$

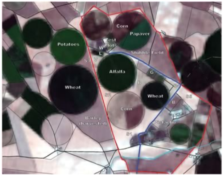
- Slow: $\mathcal{O}(N^2)$ storage and $\mathcal{O}(N^3)$ computing cost

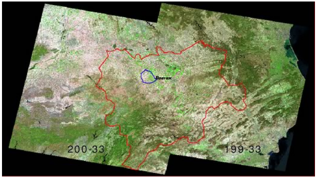
- Nonlinear regression tool
- No assumption about the data relations
- Serious competitor to other nonparametric methods
- Provides confidence intervals
- Learns the relevance of the input bands
- Allows for flexible kernels that encode priors
- Automatic tuning of parameters

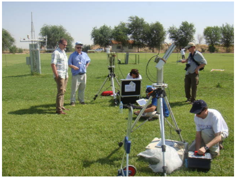
An illustrative real example: Retrieve vegetation chlorophyll content

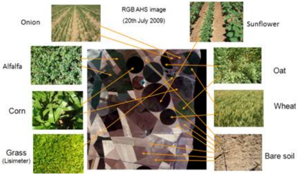
- *In situ* leaf-level *Chl* measured with a calibrated CCM-200 Chlorophyll Content Meter in the field
- CHRIS images (62 bands, 400-1050 nm, 34m nadir res.) –corrected geometrically and atmospherically

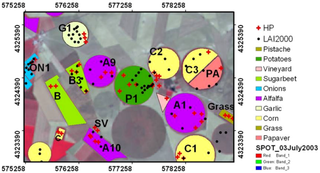






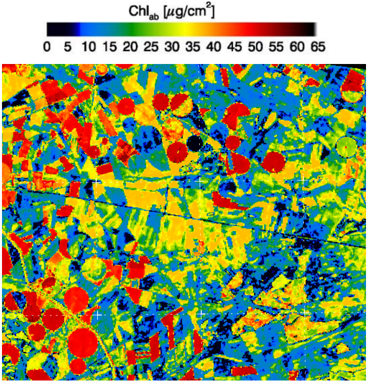


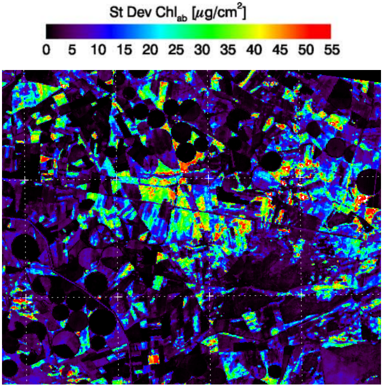




Method	**Formulation**	**|ME|**	**RMSE**	**MAE**	**R**
GI	R_{672}/R_{550}	12.97	28.77	26.58	0.74
mNDVI	$(R_{800}\text{-}R_{680})/(R_{800}+R_{680}\text{-}2R_{445})$	1.20	9.27	7.06	0.79
mNDVI_{705}	$(R_{750}\text{-}R_{705})/(R_{750}+R_{705}\text{-}2R_{445})$	1.22	9.13	6.30	0.80
mSR_{705}	$(R_{750}\text{-}R_{445})/(R_{705}+R_{445})$	2.52	10.94	7.99	0.76
NDVI	$(R_{800}\text{-}R_{670})/(R_{800}+R_{670})$	1.72	9.85	7.34	0.78
NDVI2	$(R_{750}\text{-}R_{705})/(R_{750}+R_{705})$	1.81	9.56	6.79	0.80
OSAVI	$1.16(R_{800}\text{-}R_{670})/(R_{800}+R_{670}+0.16)$	1.72	9.85	7.34	0.78
PRI	$(R_{531}\text{-}R_{570})/(R_{531}+R_{570})$	25.58	35.96	32.14	0.77
PRI2	$(R_{570}\text{-}R_{539})/(R_{570}+R_{539})$	37.84	39.19	37.84	0.76
PSRI	$(R_{680}\text{-}R_{500})/R_{750}$	28.07	37.10	34.18	0.80
RDVI	$(R_{800}-R_{670})/\sqrt{(R_{800}+R_{670})}$	2.12	10.67	8.21	0.76
SIPI	$(R_{800}\text{-}R_{445})/(R_{800}\text{-}R_{680})$	17.18	31.54	28.62	0.76
SR1	R_{750}/R_{700}	3.16	11.76	8.49	0.75
SR3	R_{750}/R_{550}	0.87	9.78	7.51	0.75
SR4	R_{672}/R_{550}	12.97	28.77	26.58	0.74
VOG	R_{740}/R_{720}	0.61	9.68	7.44	0.76
LR	ℓ_2 least squares	4.56	11.52	8.94	0.77
LASSO	ℓ_1 least squares	3.46	12.39	9.56	0.73
TREE	Pruning, min. split = 30	*0.14*	6.98	4.59	0.86
NN	Sigmoid links, one hidden layer	0.93	9.19	6.49	0.77
KRR	RBF kernel	0.73	6.22	5.24	*0.89*
GPR	Anisotropic RBF kernel	1.69	6.57	5.19	*0.95*

simpleR: http://www.uv.es/gcamps/code/simpleR.html





- High confidences (west) were the most sampled fields
- Low confidences (center) in underrepresented areas
  e.g. dry barley, harvested barley, and bright bare soils

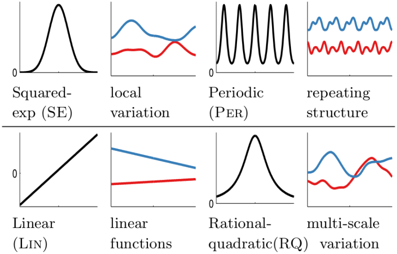
Part 3: Advances in variable prediction and understanding

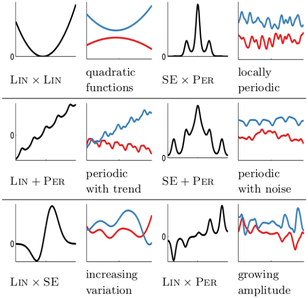
Structured, non-stationary and multiscale

- Typical kernel (covariance) function:

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

- Combine kernels:





[Gahramani, 2013]

Time-based covariance for GP regression

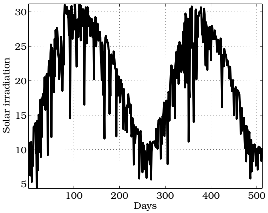
- Standard local relations:

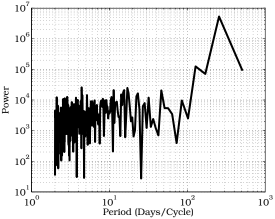
$$k_1(\mathbf{x}_i, \mathbf{x}_j) = \nu \exp\left(-\sum_{f=1}^{F} \frac{(\mathbf{x}_i^f - \mathbf{x}_j^f)^2}{2\sigma_f^2} \right) + \sigma_n^2 \delta_{ij}$$

- Stationary covariance to capture (inexact) periodicity:

$$k_2(t_i, t_j) = \gamma \exp\left(- \frac{2\sin^2[\pi(t_i - t_j)]}{\sigma_t^2} \right) \times \exp\left(- \frac{(t_i - t_j)^2}{2\sigma_d^2} \right)$$

- Combine kernels: $k([\mathbf{x}_i, t_i], [\mathbf{x}_j, t_j]) = k_1(\mathbf{x}_i, \mathbf{x}_j) + k_2(t_i, t_j)$





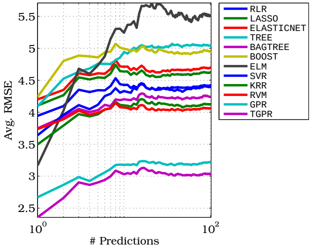
[Salcedo and Camps-Valls, GRSL 2014]

Time-based covariance for GP regression

Prediction of Daily global solar irradiation

Source	Data	Units	min-max
Cimel sunphotometer	Aerosol Optical Depth	-	0.01-1.38
Brewer spectrophotometer	Total Ozone	Dobson	242.50-443.50
Atmospheric sounding	Total Precip. Water	mm	1.33-41.53
GFS	Cloud amount	%	2-79.2
Pyranometer	Measured global solar irradiation	kJ/m^2	4.38-31.15

Method	ME	RMSE	MAE	ρ
RLR	0.27	4.42	3.51	0.76
RLR$_t$	0.25	4.33	3.42	0.78
SVR	0.54	4.40	3.35	0.77
SVR$_t$	0.42	4.23	3.12	0.79
GPR	0.14	3.22	2.47	*0.88*
GPR$_t$	0.13	*3.15*	*2.27*	*0.88*
TGPR	*0.11*	**3.14**	**2.19**	**0.90**



[Salcedo and Camps-Valls, GRSL 2014]

Parameter-free covariance

PCK: Probabilistic cluster kernel:

1: Run EM-GMM clustering T times with K clusters

2: Obtain the posterior probability vector $\boldsymbol{\pi}_i(t,k)$ for $\mathbf{x}_i$:

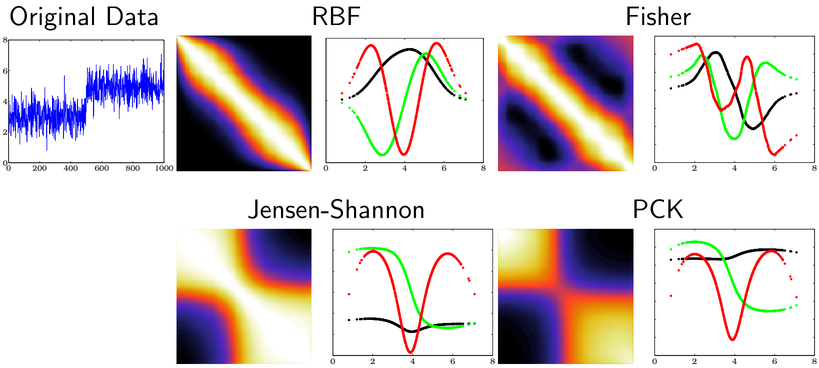
3: Compute:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{Z} \sum_{t=1}^{T} \sum_{k=2}^{K} \boldsymbol{\pi}_i(t,k)^\top \boldsymbol{\pi}_j(t,k) \quad i,j = 1,\ldots,n$$

4: Done!

[Izquierdo and Camps-Valls, NEUCOM, 2013]

Parameter-free covariance



[Izquierdo and Camps-Valls, NEUCOM, 2013]

Parameter-free covariance

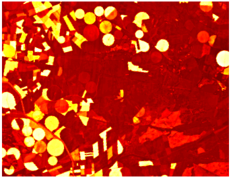
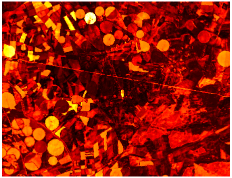
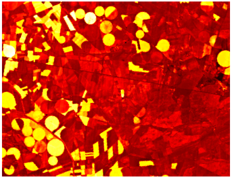
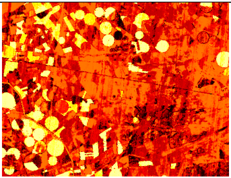
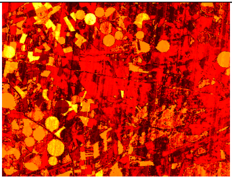
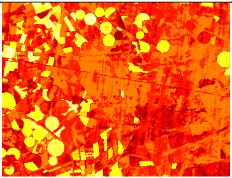
	Chl [$\mu g/cm^2$]	LAI [m^2/m^2]	fCover [%]
	4.552	0.571	0.1154
GPR (ARD)			
	4.277	**0.521**	**0.1113**
GPR (PCK)			

Figure : Estimation maps for Chl, LAI and FCV using CHRIS/PROBA data.

[Izquierdo and Camps-Valls, IEEE TGARS, 2013]

Warped GPR: Learning the output transformation

- No more *ad-hoc* transformations of the observed variable
- Data can be “warped” to look more like a GP:

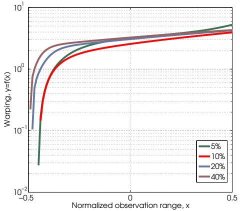
$$y_i = g(f(\mathbf{x}_i) + \varepsilon_i)$$

- Warped GP places another prior for $g(\mathbf{x}) \sim \mathcal{GP}(f, c(f, f'))$

$$g^{-1}(y_i) = y_i + \sum_{\ell=1}^{L} a_\ell \tanh(b_\ell(y_i + c_\ell)), \qquad a_\ell,\ b_\ell \geq 0,$$

- Evidence is analytical, posterior mean is easy to find
- Improved results and some knowledge about the target

METHOD	ME	RMSE	MAE	R
GPR	0.285	2.312	0.451	0.618
WGPR	0.298	2.344	0.445	0.638



[Lazaro, 2013; Camps-Valls, 2014]

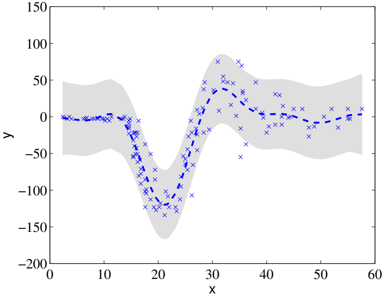
Heteroscedastic GPR

- Heteroscedasticity: Signal and noise relations exist
- Standard GP prior does not capture these relations, $\varepsilon_n \sim i.i.d$
- The GP prior is

$$f(\mathbf{x}) \sim \mathcal{GP}(\mu_0 \mathbf{1}, k_{\boldsymbol{\theta}_f}(\mathbf{x}, \mathbf{x}'))$$

- Good results, but unreasonable uncertainties!

METHOD	ME	RMSE	MAE	R
GPR	0.285	2.312	0.451	0.618
WGPR	0.298	2.344	0.445	0.638



[Lazaro, 2012; Camps-Valls, 2014]

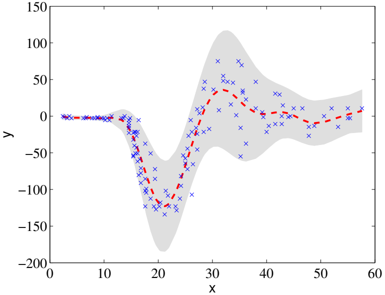
Heteroscedastic GPR

- Heteroscedasticity: Signal and noise relations exist
- Change the GP prior over ε_n to $\varepsilon_n \sim \mathcal{N}(0, e^{g(\mathbf{x}_n)})$
- We place an additional GP prior

$$f(\mathbf{x}) \sim \mathcal{GP}(\mu_0 \mathbf{1}, k_{\boldsymbol{\theta}_f}(\mathbf{x}, \mathbf{x}')) \qquad g(\mathbf{x}) \sim \mathcal{GP}(\mu_0 \mathbf{1}, k_{\boldsymbol{\theta}_g}(\mathbf{x}, \mathbf{x}'))$$

- Improved results and some knowledge about the noise

METHOD	ME	RMSE	MAE	R
GPR	0.285	2.312	0.451	0.618
WGPR	0.298	2.344	0.445	0.638
VHGPR	0.271	2.309	0.442	0.684



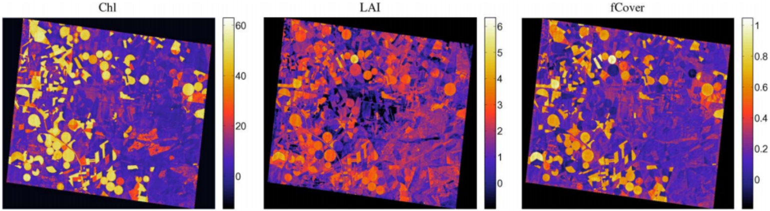
[Lazaro, 2012; Camps-Valls, 2014]

Multioutput GPR

- A model to predict several variables simultaneously
- Constrain the outputs to be physically-meaningful
- Posterior over the unknown output:

$$p(y_*|\mathbf{x}_*,\mathcal{D}) = \mathcal{N}(y_*|\mu_{\mathrm{GP}*},\sigma^2_{\mathrm{GP}*})$$
$$\mu_{\mathrm{GP}*} = \mathbf{k}_{\mathbf{f}*}^\top(\mathbf{K}+\sigma^2\mathbf{I}_n)^{-1}\mathbf{Y} = \mathbf{k}_{\mathbf{f}*}^\top\mathbf{A}$$

- Just efficiency, no actual cross-relations



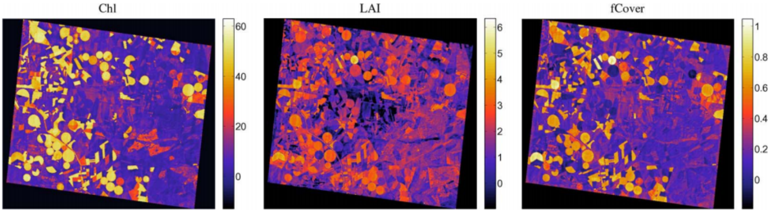
[Camps-Valls, 2012]

Multioutput GPR

- Constrain the outputs to be physically-meaningful
- Response vector as a linear combination of a set of M latent GPs
- Gives a block covariance matrix $[\tilde{\mathbf{K}}_{ij}^m] = k_m(\mathbf{x}_i, \mathbf{x}_j)$, $m = 1, \dots, M$

$$p(y_*|\mathbf{x}_*, \mathcal{D}) = \mathcal{N}(y_*|\mu_{\mathrm{GP}*}, \sigma^2_{\mathrm{GP}*})$$
$$\boldsymbol{\mu}^m_{\mathrm{GP}*} = \tilde{\mathbf{k}}_{\mathbf{f}*}^\top(\tilde{\mathbf{K}} + \sigma^2\mathbf{I}_n)^{-1}\mathbf{Y} = \tilde{\mathbf{k}}_{\mathbf{f}*}^\top\mathbf{A}$$

- Promising results in biophysical parameter retrieval



[Camps-Valls, 2015]

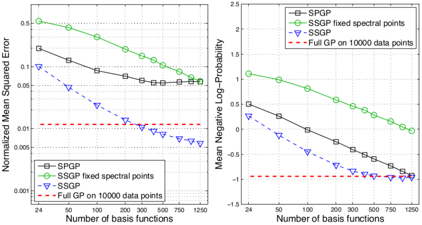
Sparse spectrum Gaussian Processes Regression (SSGPR)

- Speeding up GP is mandatory
- Sparse spectrum GP = standard GP with covariance:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sum_{m=1}^{M} \cos(\mathbf{w}_m^\top(\mathbf{x}_i - \mathbf{x}_j)), \qquad M \text{ basis functions}$$

- Computation time goes from $\mathcal{O}(N^3)$ to $\mathcal{O}(M \cdot N^2)$
- Storage goes from $\mathcal{O}(N^2)$ to $\mathcal{O}(M \cdot N)$

METHOD	ME	RMSE	R	CPU [s]
GPR	0.285	2.312	0.618	1.150
SSGPR	0.296	2.194	0.684	0.530



[Lazaro, 2008; Camps-Valls, 2014]

Sparse spectrum Gaussian Processes Regression (SSGPR)

- Inversion of PROSAIL radiative transfer model
- PROSAIL models leaf properties and canopy bidirectional reflectance
- 10^6 points; train with 400,000 samples and cosine basis
- Prediction of 7 vegetation variables

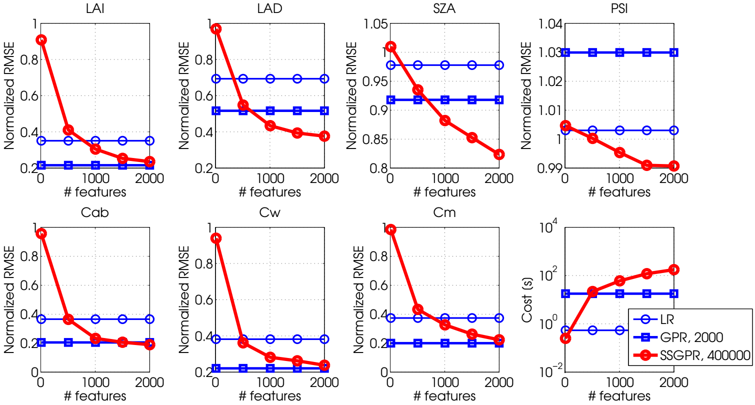
Table : Configuration parameters of the simulated data.

Parameter	Sampling	Min	Max
RTM model: Prospect 4			
Leaf Structural Parameter	Fixed	1.50	1.50
C_{ab}, chlorophyll a+b [μg/cm^2]	$\mathcal{U}(14, 49)$	0.067	79.97
C_w, equivalent water thickness [mg/cm^2]	$\mathcal{U}(10, 31)$	2	50
C_m, dry matter [mg/cm^2]	$\mathcal{U}(5.9, 19)$	1.0	3.0
RTM model: 4SAIL			
Diffuse/direct light	Fixed	10	10
Soil Coefficient	Fixed	0	0
Hot spot	Fixed	0.01	0.01
Observer zenit angle	Fixed	0	0
LAI, Leaf Area Index	$\mathcal{U}(1.2, 4.3)$	0.01	6.99
LAD, Leaf Angle Distribution	$\mathcal{U}(28, 51)$	20.04	69.93
SZA, Solar Zenit Angle	$\mathcal{U}(8.5, 31)$	0.082	49.96
PSI, Azimut Angle	$\mathcal{U}(30, 100)$	0.099	179.83

[Laparra and Camps-Valls, 2015]

Sparse spectrum Gaussian Processes Regression (SSGPR)

- Inversion of PROSAIL radiative transfer model
- PROSAIL models leaf optical properties model and canopy bidirectional reflectance
- 1 million points; train with 400,000 samples and cosine basis



[Laparra and Camps-Valls, 2015]

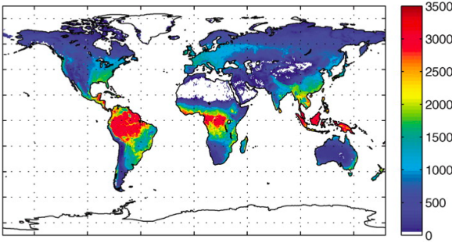
Sensitivity analysis

- Derivatives of the predictive mean $\phi(\mathbf{x})$ wrt features
- Sensitivity of feature j is defined as

$$s_j = \int \left(\frac{\partial \phi(\mathbf{x})}{\partial \mathbf{x}_j} \right)^2 p(\mathbf{x}) \mathrm{d}\mathbf{x} = \ldots = \frac{1}{N} \sum_{q=1}^{N} \left(\sum_{p=1}^{N} \frac{\alpha_p (\mathbf{x}_{p,j} - \mathbf{x}_{q,j})}{\sigma_j^2} k(\mathbf{x}_p, \mathbf{x}_q) \right)^2$$

- Gross Primary Production (GPP) is the largest global CO_2 flux driving several ecosystem functions

GPP	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

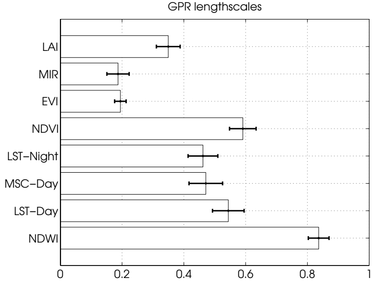


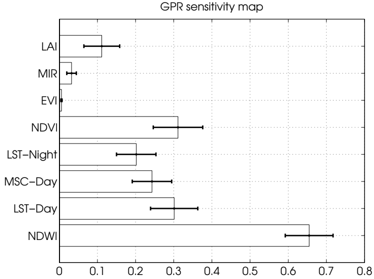
[Jung & Camps-Valls, 2015]

Sensitivity analysis

- Derivatives of the predictive mean $\phi(\mathbf{x})$ wrt features
- Sensitivity of feature j is defined as

$$s_j = \int \left(\frac{\partial \phi(\mathbf{x})}{\partial \mathbf{x}_j}\right)^2 p(\mathbf{x})\mathrm{d}\mathbf{x} = \ldots = \frac{1}{N}\sum_{q=1}^{N}\left(\sum_{p=1}^{N}\frac{\alpha_p(\mathbf{x}_{p,j}-\mathbf{x}_{q,j})}{\sigma_j^2}k(\mathbf{x}_p,\mathbf{x}_q)\right)^2$$



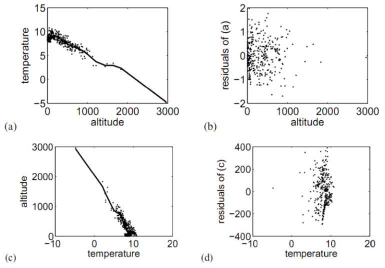


[Jung & Camps-Valls, 2015]

From regression to causation

Hoyer (2008) regression-based framework:

- Perform nonlinear regression from $x \to y$ (and vice-versa, $y \to x$)
- Independent forward residuals? $(y - f(x)) \perp\!\!\!\perp y$
- Independent backward residuals? $(x - g(y)) \perp\!\!\!\perp x$
- p-value of independence tells the right direction of causation



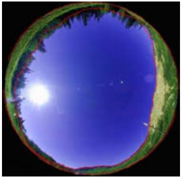
[Camps-Valls, 2012]

Method	p_f	p_b	Conclusion
GP	2.88×10^{-2}	3.54×10^{-12}	alt $\to$ temp
WGPR	7.47×10^{-4}	9.28×10^{-11}	alt $\to$ temp
VHGPR	2.94×10^{-16}	8.83×10^{-23}	alt $\to$ temp
GP	3.86×10^{-61}	1.57×10^{-119}	PPFD(tot)$\to$ NEP
WGPR	2.09×10^{-52}	4.18×10^{-112}	PPFD(tot)$\to$ NEP
VHGPR	5.09×10^{-61}	2.51×10^{-108}	PPFD(tot)$\to$ NEP
GP	1.59×10^{-11}	1.24×10^{-79}	PPFD(diff)$\to$ NEP
WGPR	1.16×10^{-12}	9.00×10^{-79}	PPFD(diff)$\to$ NEP
VHGPR	2.94×10^{-13}	9.90×10^{-78}	PPFD(diff)$\to$ NEP
GP	2.05×10^{-8}	1.56×10^{-112}	PPFD(dir)$\to$ NEP
WGPR	1.30×10^{-15}	3.33×10^{-111}	PPFD(dir)$\to$ NEP
VHGPR	4.53×10^{-16}	1.12×10^{-115}	PPFD(dir)$\to$ NEP

Explicit mapping LAI with GPR and your smartphone

- Leaf area index (LAI) characterizes plant canopies
- LAI: one-sided green leaf area per unit ground surface area
- LAI: Key variable to analyze plants-atmosphere interaction
    - amount of radiation intercepted
    - plant water requirements
    - CO_2 sequestration
    - assimilation of exogenous information in simulation models
    - forecasting purposes
- Hemispherical photography estimates LAI from upward-looking fisheye photographs taken beneath the plant canopy

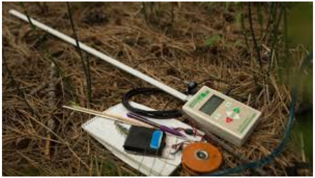


[Campos and Camps-Valls, 2015]

Explicit mapping LAI with GPR and your smartphone

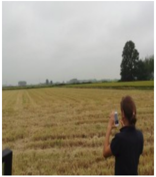
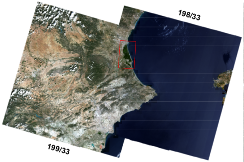
- Leaf area index (LAI) characterizes plant canopies
- LAI: one-sided green leaf area per unit ground surface area
- LAI: Key variable to analyze plants-atmosphere interaction
  - amount of radiation intercepted
  - plant water requirements
  - CO_2 sequestration
  - assimilation of exogenous information in simulation models
  - forecasting purposes
- Ceptometers invert light transmittance models, but expensive, heavy, maintenance

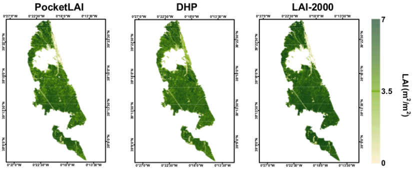




[Campos and Camps-Valls, 2015]

Explicit mapping LAI with GPR and your smartphone

- Leaf area index (LAI) characterizes plant canopies
- LAI: one-sided green leaf area per unit ground surface area
- LAI: Key variable to analyze plants-atmosphere interaction
  - amount of radiation intercepted
  - plant water requirements
  - CO_2 sequestration
  - assimilation of exogenous information in simulation models
  - forecasting purposes
- PocketLAI app: Segmentation (color-based, sky conditions), accelerometer, Poisson model for random leaves spatial distribution

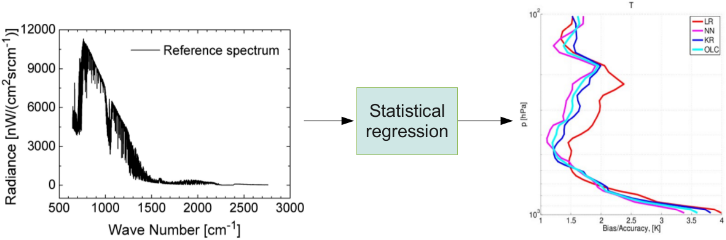



[Campos and Camps-Valls, 2015]

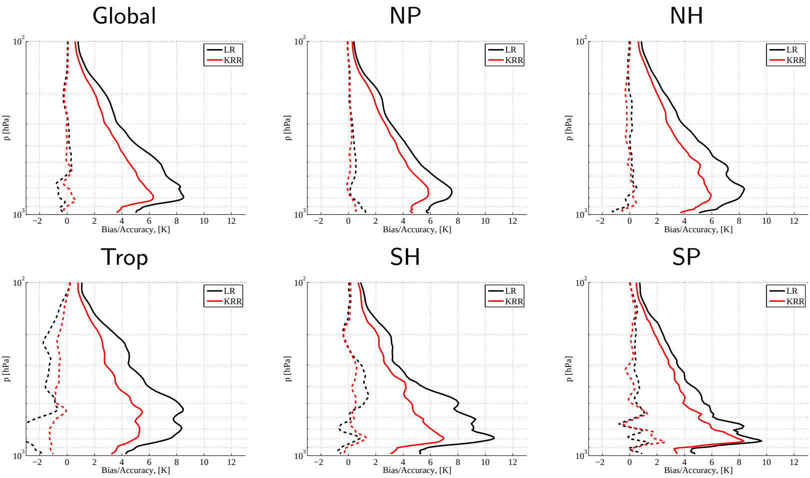
GPR for weather forecasting

- Meteorological satellites carry infrared sounders, e.g. AIRS or IASI
- Big data: terabytes/day, IASI pixels are 8000-dim, 100 atmos. levels
- We run multioutput GPR for predicting temperature, moisture and ozone concentration (among other parameters)
- Gain in RMSE (1.5K), CPU time (seconds vs hours), and cloud detection



[Laparra, Calbet and Camps-Valls, 2015]

GPR for weather forecasting

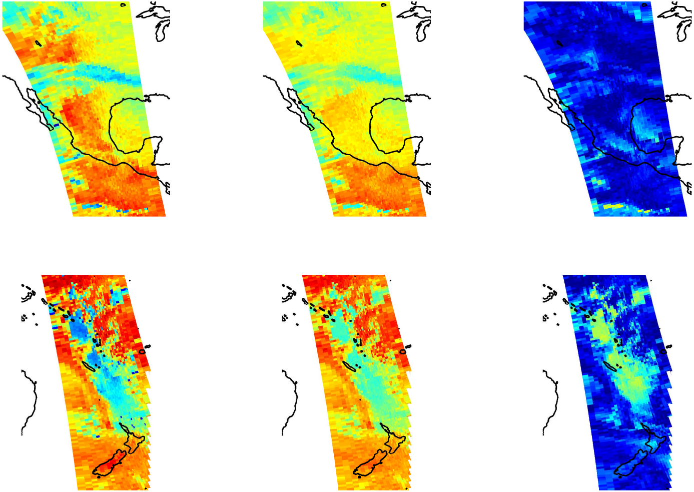


- GPR largely improves PLR results over land (+1.5K)
- Big numerical and statistical differences in all regions

[Laparra, Calbet and Camps-Valls, 2015]

GPR for weather forecasting

PLR | **GPR,** $\mu^2_{\mathbf{GP*}}$ | **GPR,** $\sigma^2_{\mathbf{GP*}}$



[Laparra, Calbet and Camps-Valls, 2015]

Part 4: Conclusions

Conclusions

- **Advances in climate variable estimation from space**
- **Look at the signal structure and act!**
- **Bayesian nonparametrics is a proper framework**
  - Solid Bayesian foundation
  - Excellent prediction capabilities
  - Encoding of prior knowledge and structures
  - Deal efficiently with uncertainties
- **GPR allows feature ranking and causal inference**
- **Code, papers, demos: http://isp.uv.es/**

Conclusions

- **Advances in climate variable estimation from space**
- **Look at the signal structure and act!**
- **Bayesian nonparametrics is a proper framework**
  - Solid Bayesian foundation
  - Excellent prediction capabilities
  - Encoding of prior knowledge and structures
  - Deal efficiently with uncertainties
- **GPR allows feature ranking and causal inference**
- **Code, papers, demos: http://isp.uv.es/**

Thanks! Tuia, Jenssen, Jung, Reichstein, Laparra, Mooij, Schölkopf, Peters, Verrelst, Lázaro-Gredilla, Titsias, Pérez-Cruz, Calbet, Muñoz-Marí, Moreno, Salcedo, Izquierdo, Campos

Visit http://isp.uv.es/

- *Global monitoring of crop photosynthesis with chlorophyll fluorescence*, PNAS 2014.
- *Prediction of Daily Global Solar Irradiation*, IEEE-GRSL 2015.
- *Kernel Framework for Supervised Subspace Learning*, IEEE SPM 2013
- *Advances in Hyperspectral Image Classification*, IEEE SPM 2014
- *Gaussian processess Uncertainty Estimates in Experimental Sentinel-2 LAI*, ISPRS, 2013
- *Gaussian process retrieval of chlorophyll content*, IEEE JSTARS, 2013
- *Retrieval of Canopy Parameters using Gaussian Processes Techniques*, IEEE TGRS, 2012
- *Learning with the Kernel Signal To Noise Ratio*, IEEE PAMI, 2015
- *Sensitivity Analysis of Chlorophyll content prediction*, IEEE TGRS, 2015
- *Retrieval of Biophysical Parameters with Heteroscedastic GP*, IEEE GRSL, 2014
- *Ranking drivers of global carbon and energy fluxes over land*, IEEE IGARSS, 2015
- *Biophysical parameter retrieval with warped Gaussian processes*, IEEE IGARSS, 2015
- *Large-scale random features kernel regression* , IEEE IGARSS, 2015
- *Kernel methods for Remote Sensing Data Analysis*, Wiley and Sons, 2009
- *Kernel methods for Digital Signal Processing*, Wiley and Sons, 2015