Machine learning for modeling and understanding in Earth sciences

Gustau Camps-Valls
Image Processing Lab (IPL)
Universitat de València

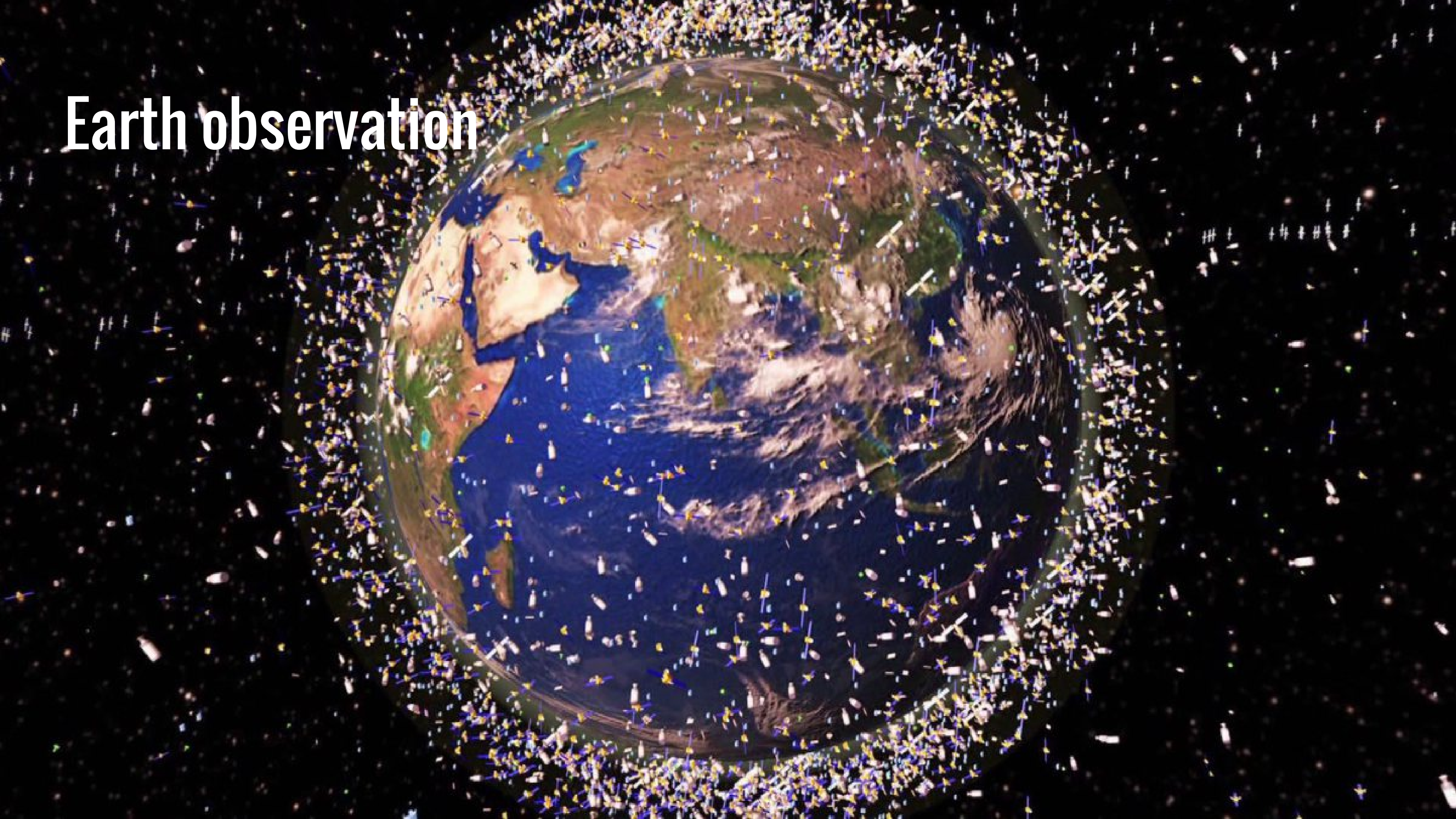
http://isp.uv.es @isp_uv_es gustau.camps@uv.es

Earth observation

Earth observation

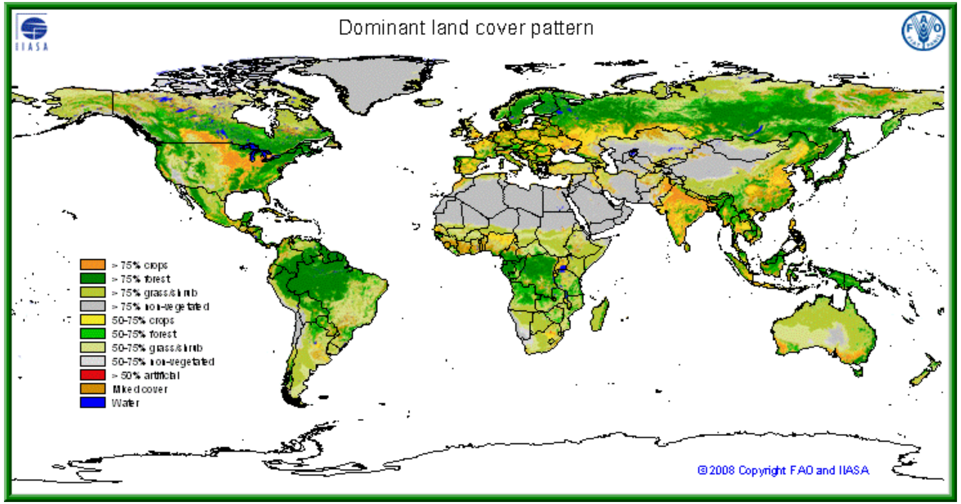
"Earth observation (EO) is the gathering of information about planet Earth's physical, chemical and biological systems via remote sensing technologies supplemented by earth surveying techniques, encompassing the collection, analysis and presentation of data"

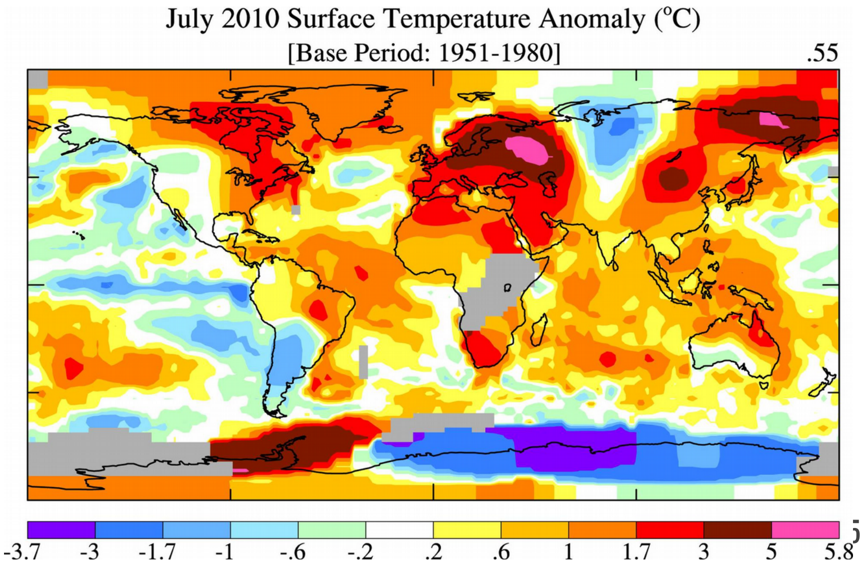
Earth observation



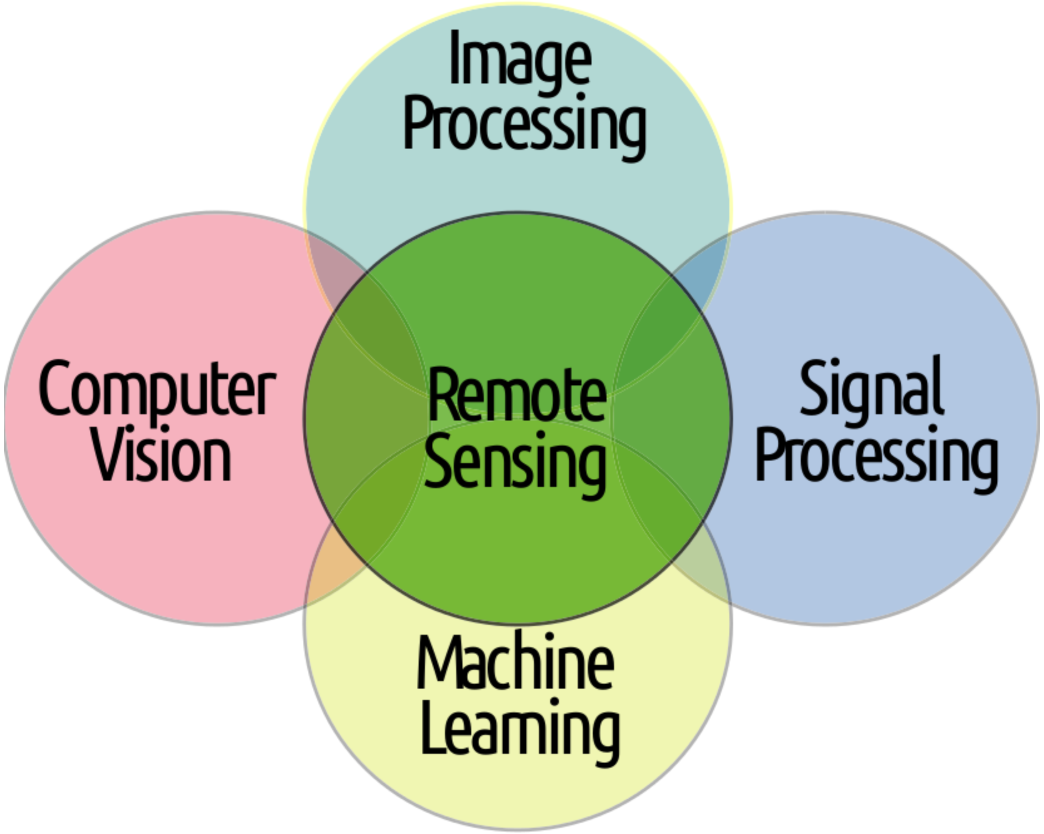
Earth observation applications

- Identify and classify objects, detect clusters & patterns, and detect changes
- Estimate the content of bio-geo-physical and bio-chemical parameters





Earth observation and friends



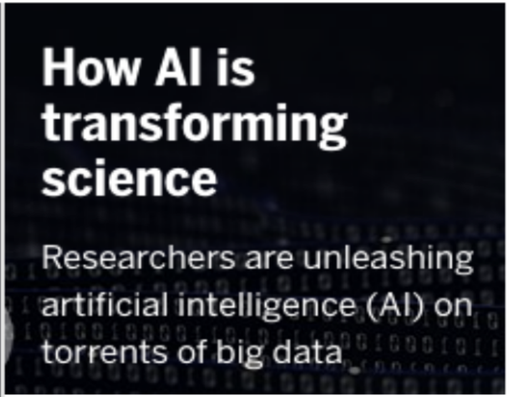
Earth observation meets machine learning

Machine learning

$$F(X) = y$$

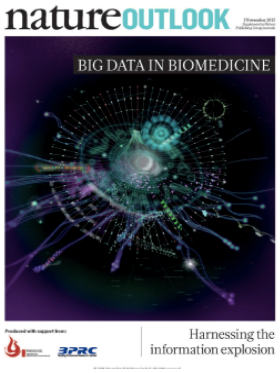
- X: observations, independent covariates
- Y: target, dependent variable
- F: machine learning model (nonlinear, nonparametric, flexible, learned from data)

AI promises to transform scientific discovery ...



"Unlike earlier attempts ... [AI systems] can see patterns and spot anomalies in data sets far larger and messier than human beings can cope with."

July 7 2017 Issue



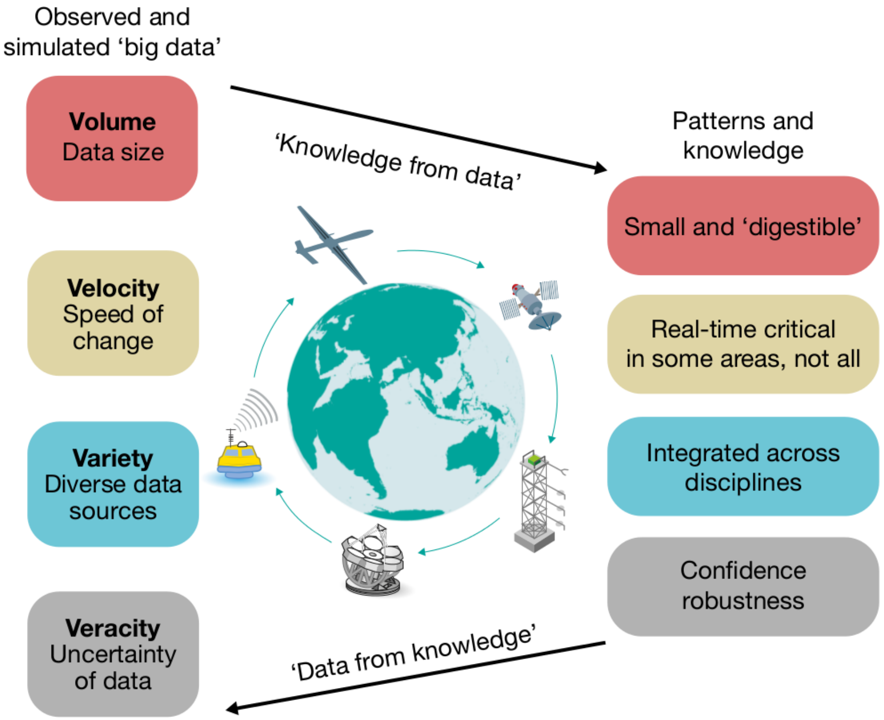
… yet only when some things happen!

- Strong spatial and temporal correlations
- Big data accessible
- Cheap computing resources available
- Fast scalable ML models available
- No expert knowledge needed
- High prediction accuracy is enough

Challenges in Earth system science

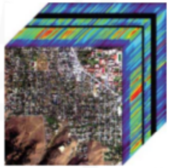
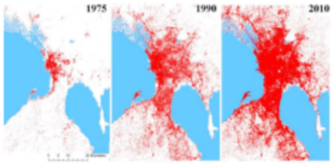
Big data challenges

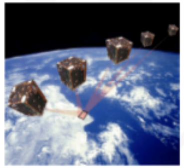
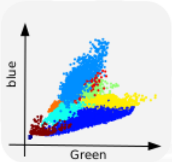
1. Data size now exceeds 100 petabytes, and is growing quasi-exponentially
2. The speed of change exceeds 5 petabytes a year, and acquisition frequencies of 10 Hz or more;
3. Reprocessing and versioning are common challenges
4. Data sources can be multi-dimensional, spatially integrated, from the organ level (such as leaves) to the global level
5. Earth has diverse observational systems, from remote sensing to in situ observations
6. The uncertainty of data can stem from observational errors or conceptual inconsistencies

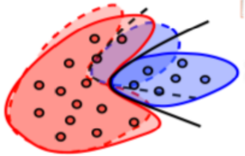
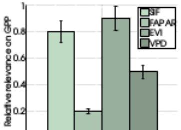


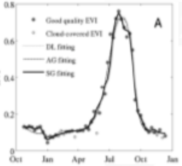
Statistical challenges

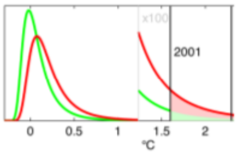
1. High dimensional data: multi-temporal, multi-angular and multi-source
2. Non-linear and non-Gaussian feature relations
3. Data misalignments and distortions
4. Irrelevant features and biased sampling strategies
5. Uneven sampling, skewed distributions and anomalies in the wild
6. Few supervised information is available

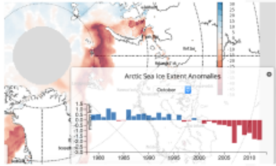









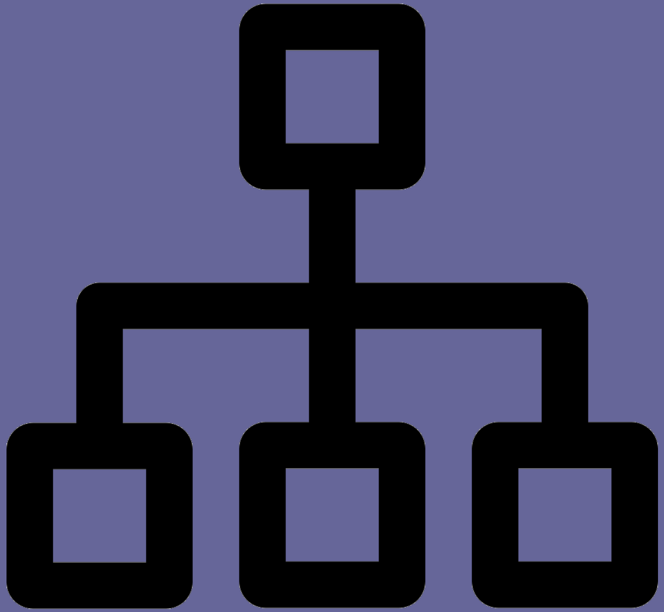

Philosophical challenges

- **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?

Outline

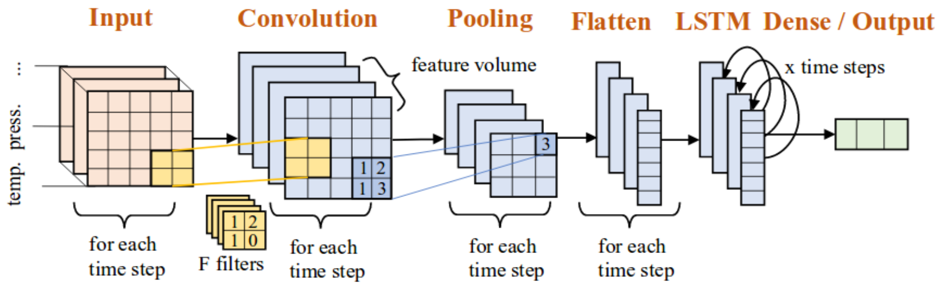
1. Advances in spatio-temporal data processing
   - Classification
   - Regression
2. Big data in the Google cloud
3. Physically-consistent ML
4. Understanding is more important than fitting!

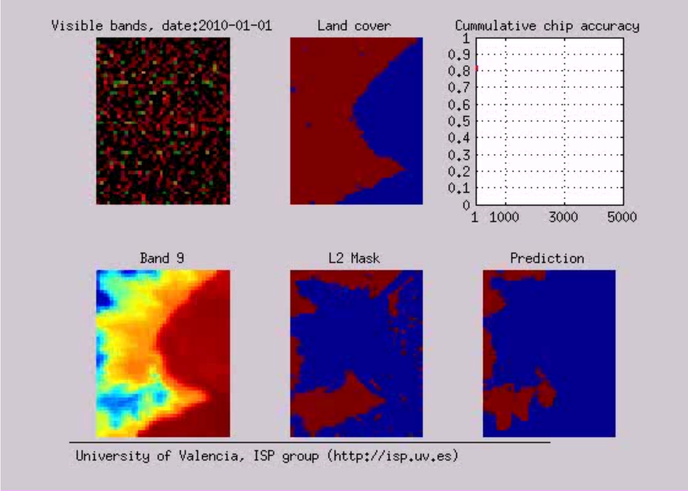
Spatio-temporal data classification



1: Deep neural nets for spatio-temporal classification

- Convolutional neural nets (CNN): hierarchical structure exploits spatial relations
- Long short-term memory (LSTM): recurrent network that accounts for memory/dynamics

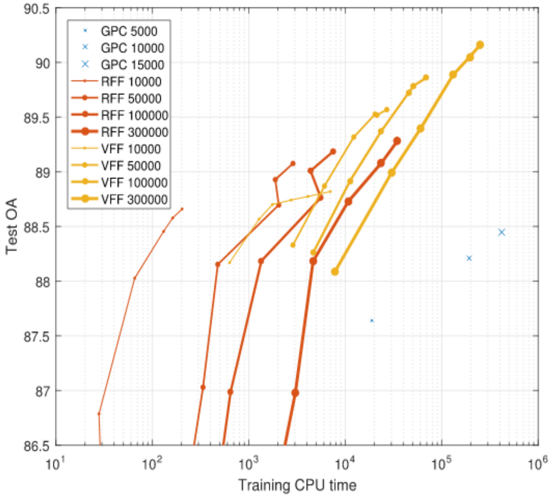


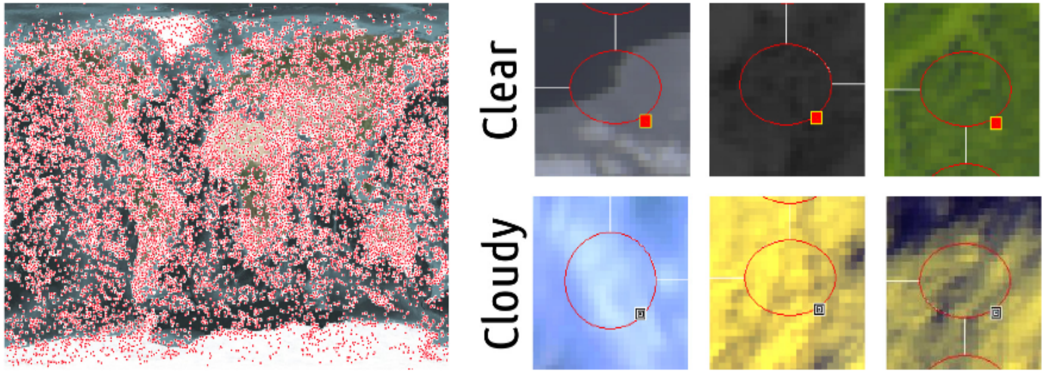


"A Deep Network Approach to Multitemporal Cloud Detection"
Tuia and Camps-Valls, IEEE IGARSS 2018, http://isp.uv.es/code/landmarks.html

2: Probabilistic and scalable classifiers

- Gaussian processes as an alternative to neural nets
- GPs allow a probabilistic treatment, confidence intervals, feature ranking, deep too!
- Gaussian processes start to be scalable ...

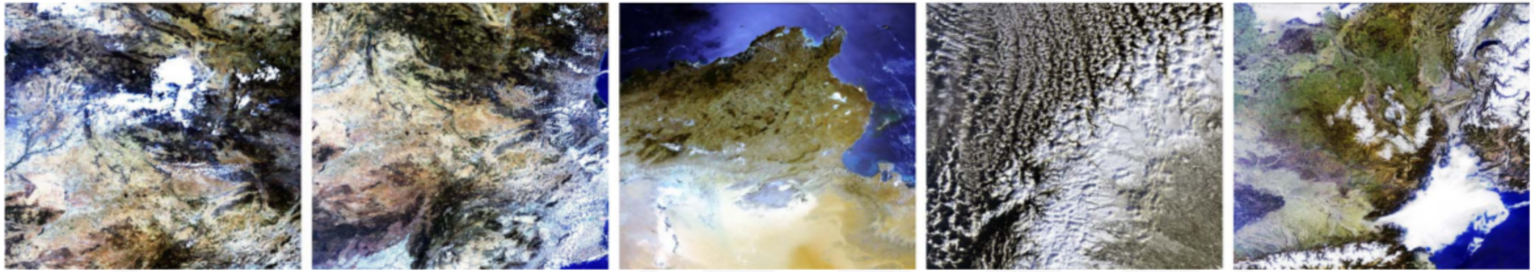


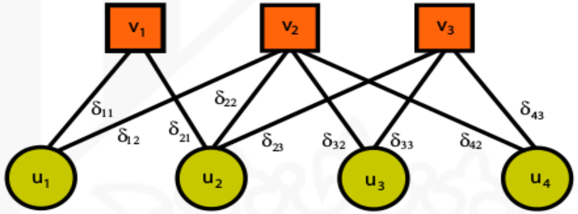


"Remote Sensing Image Classification With Large-Scale Variational Gaussian Processes," Morales, Molina and Camps-Valls, IEEE Trans. Geosc. Rem. Sens, 2018

3: Multitask learning

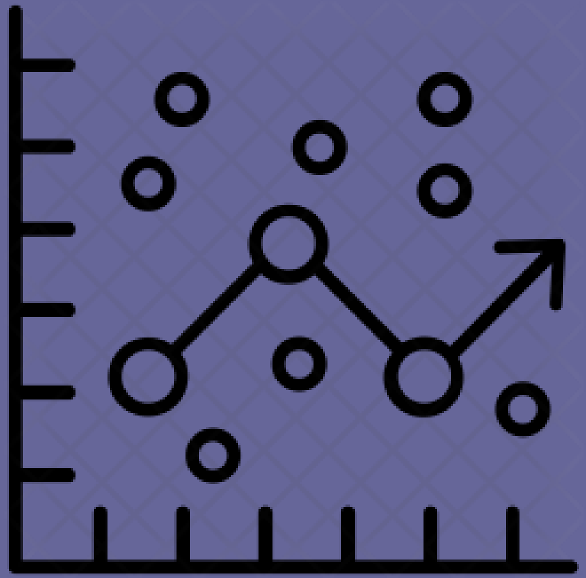
- Multiple inter-related outputs? Data from multiple sources?
- Learn to fuse heterogeneous information





“Multitask Remote Sensing Data Classification”
Leiva and Camps-Valls, IEEE Trans. Geosc, Rem. Sens 2015

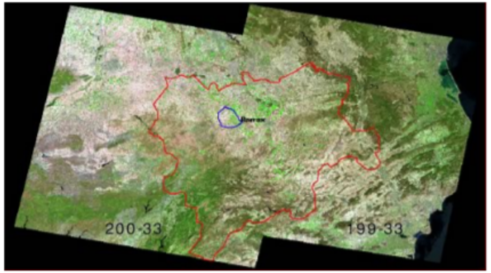
Regression, fitting, parameter retrieval

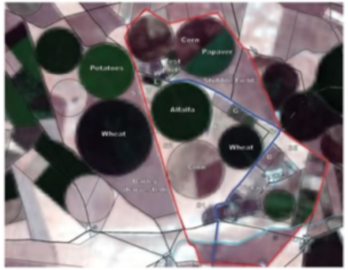


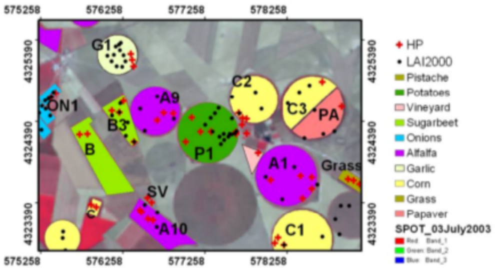
1: Spatializing vegetation parameters from space

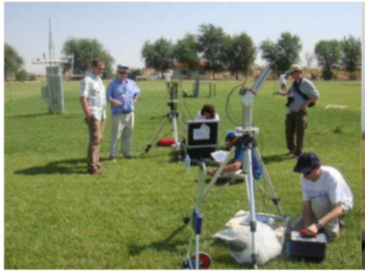
Observations, x: CHRIS images: 62 bands, 400-1050 nm, 34m

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



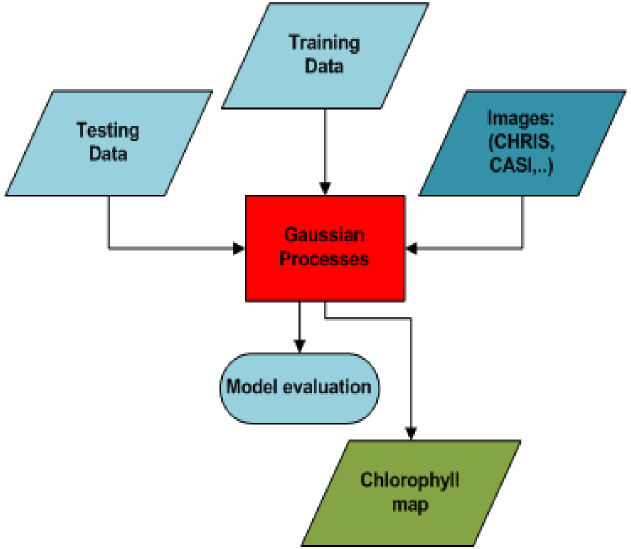


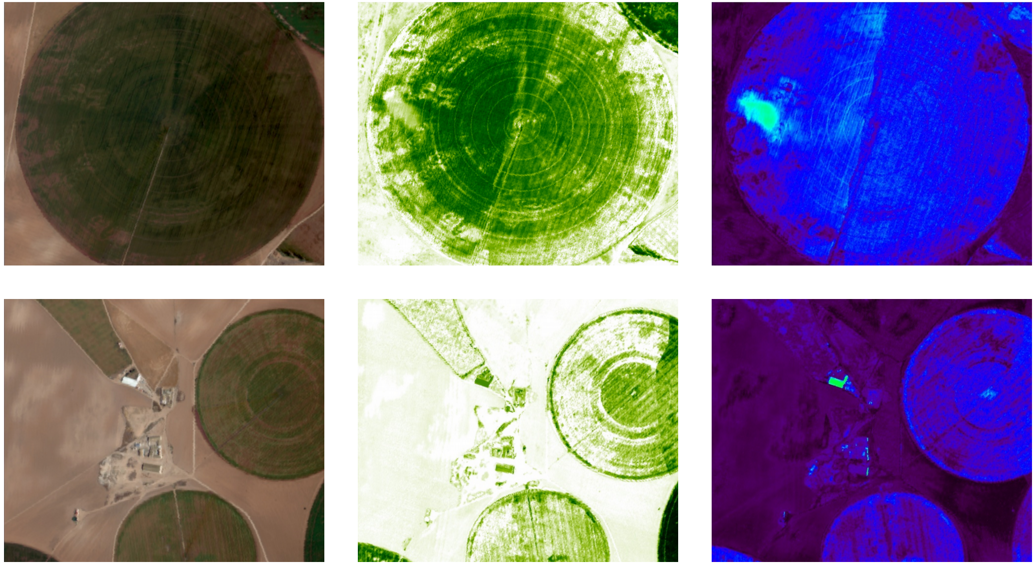




1: Spatializing vegetation parameters from space

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

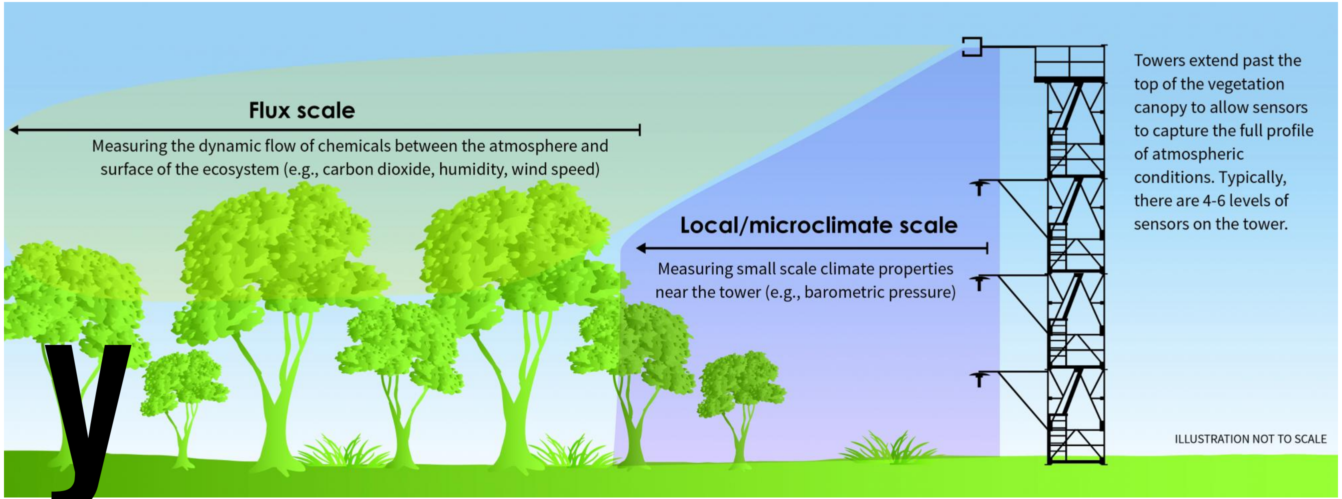




"A Survey on Gaussian Processes for Earth Observation Data Analysis: A Comprehensive Investigation"
Camps-Valls, G. and Verrelst, J. and Muñoz-Marí, et al. IEEE Geoscience and Remote Sensing Magazine 2016

2: Upscaling flux tower data from space

- Sensors allow estimating turbulent exchange of carbon dioxide (CO2), latent and sensible heat, CO2 storage, net ecosystem exchange, energy balance, ...



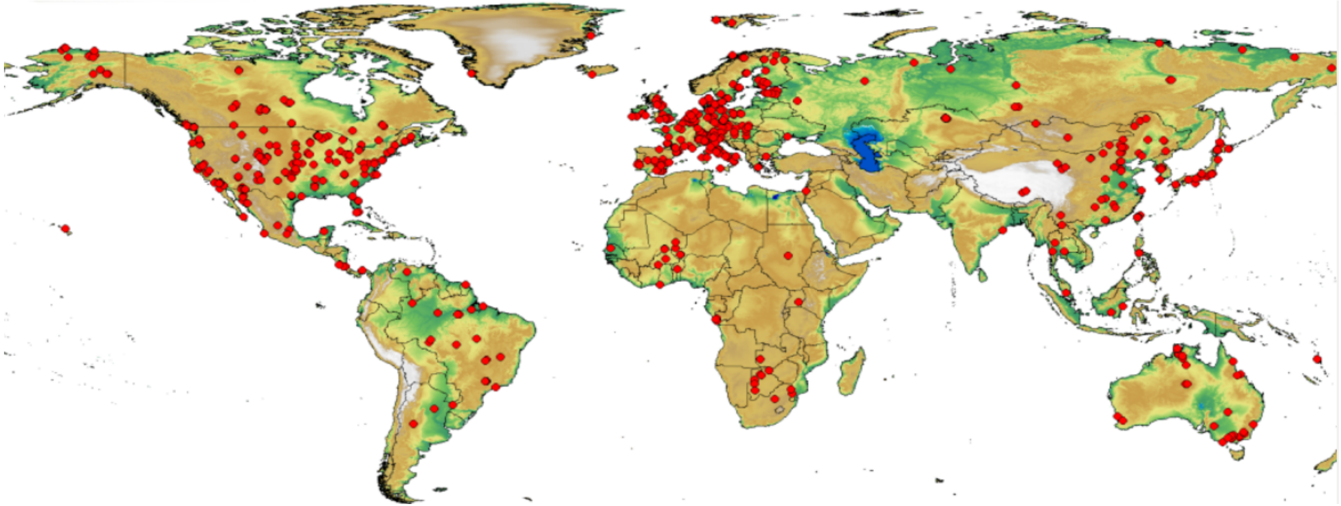
y

- Gross primary productivity
- Terrestrial ecosystem respiration
- Net ecosystem exchange

"Compensatory water effects link yearly global land CO2 sink changes to temperature"
Jung, Reichstein, Schwalm, Camps-Valls, et al. Nature 541 (7638) :516-520, 2017

2: Upscaling flux tower data from space

- FLUXNET: a sensor network of eddy covariances
- Upscaling CO2, energy and heat fluxes



"Compensatory water effects link yearly global land CO2 sink changes to temperature"
Jung, Reichstein, Schwalm, Camps-Valls, et al. Nature 541 (7638) :516-520, 2017

2: Upscaling flux tower data from space

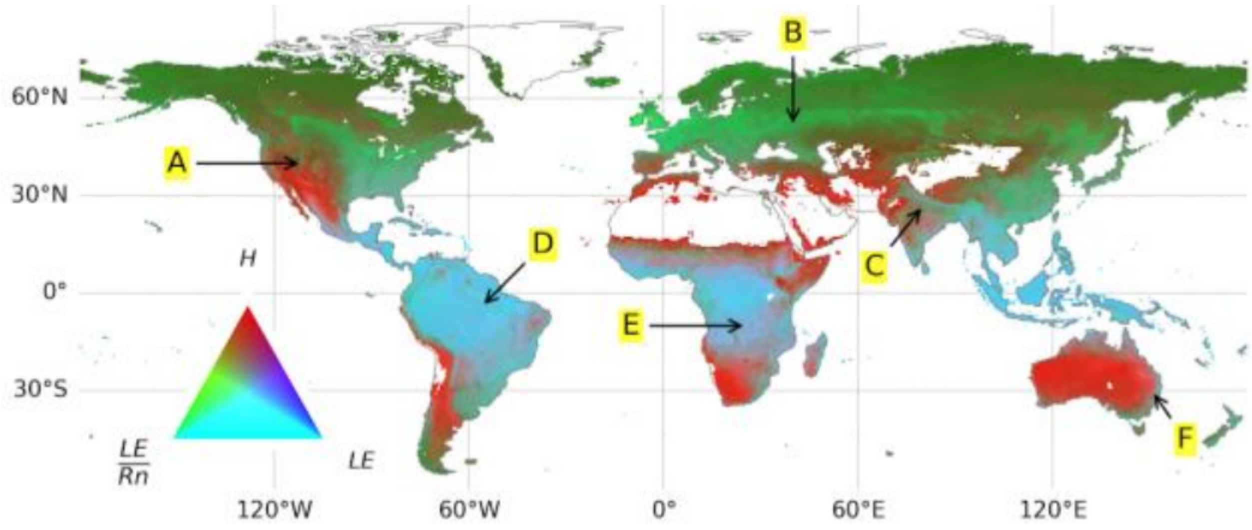
- Upscaling CO2, energy and heat fluxes from eddy covariances

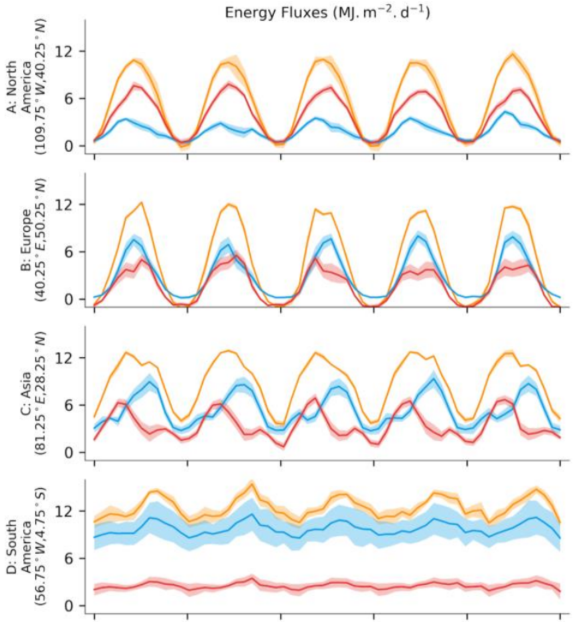
- LAI
- EVI
- NDVI
- LST-Night
- MSC-Day
- LST-Day
- NDWI

"Compensatory water effects link yearly global land CO2 sink changes to temperature"
Jung, Reichstein, Schwalm, Camps-Valls, et al. Nature 541 (7638) :516-520, 2017

2: Upscaling flux tower data from space

- Upscaling CO2, energy and heat fluxes from eddy covariances

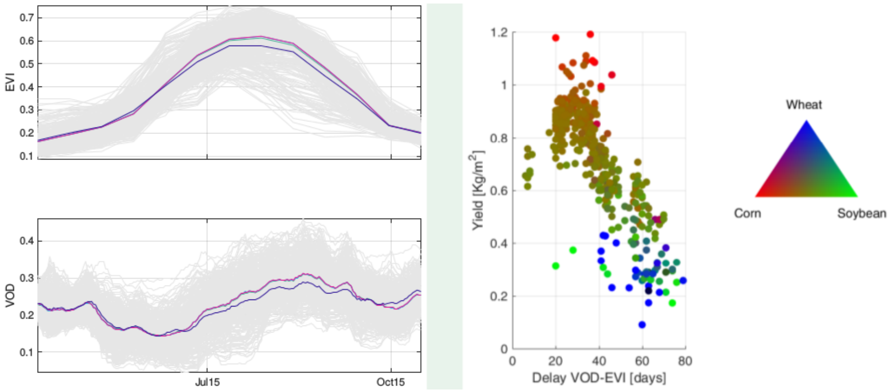


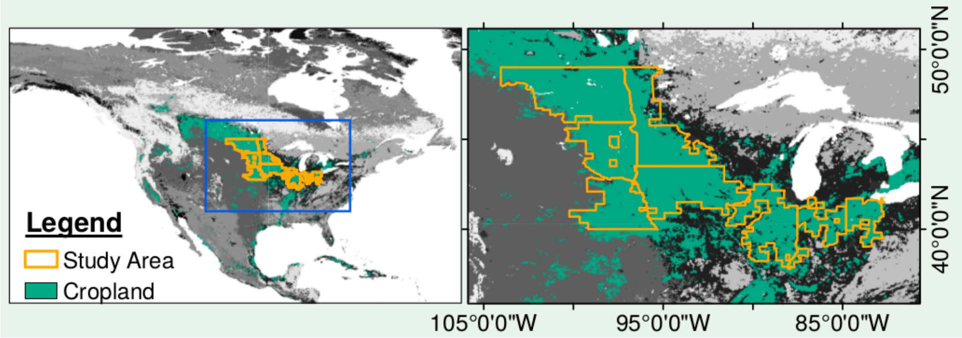


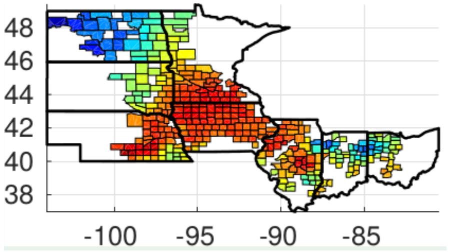
"Compensatory water effects link yearly global land CO2 sink changes to temperature"
Jung, Reichstein, Schwalm, Camps-Valls, et al. Nature 541 (7638) :516-520, 2017

3: Crop yield prediction from multisensory data

- **Crop yield (corn, soybean, wheat) & crop production**





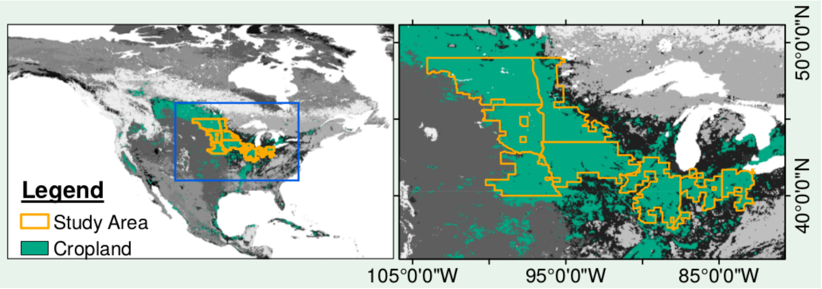


"Nonlinear Distribution Regression for Remote Sensing Applications"
Adsuara, Perez,Muñoz, Mateo, Piles, Camps-Valls, IEEE TGARS 2019

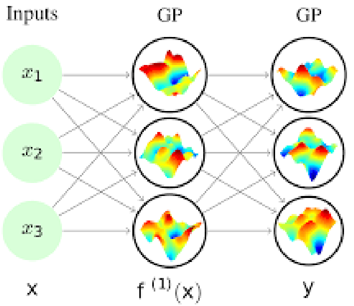
4: Advances in spatio-temporal variable prediction

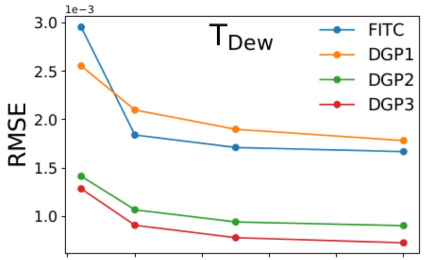
- **STA** is common place in climate informatics, neuroscience, video processing, NLP, ...
- **Current approaches:** CNN + LSTM, space-time Gaussian processes
- **Novel approaches:** distribution regression and variational deep GPs

$$\mathrm{P} \mapsto \mu_k(\mathcal{P}) \to \mathcal{P} \mapsto [\mathbb{E}\phi_1(X), \ldots, \mathbb{E}\phi_s(X)] \in \mathbb{R}^s$$
$$\langle \mu_k(\mathcal{P}), \mu_k(\mathcal{Q}) \rangle_{\mathcal{H}_k} = \mathbb{E}_{X\sim\mathcal{P}, Y\sim\mathcal{Q}} k(X, Y)$$



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



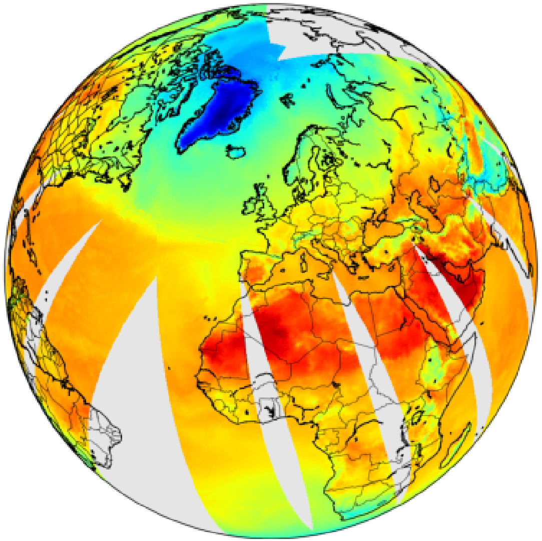


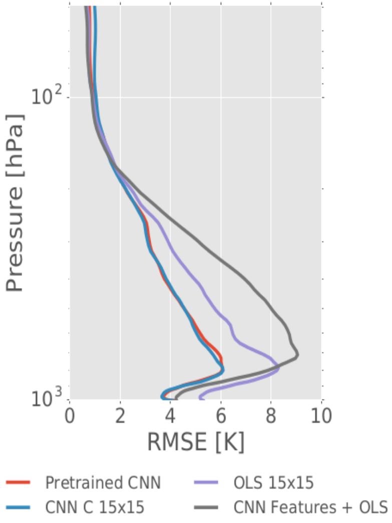
"A Survey on Gaussian Processes for Earth Observation Data Analysis"
Camps-Valls et al. IEEE Geoscience and Remote Sensing Magazine 2016

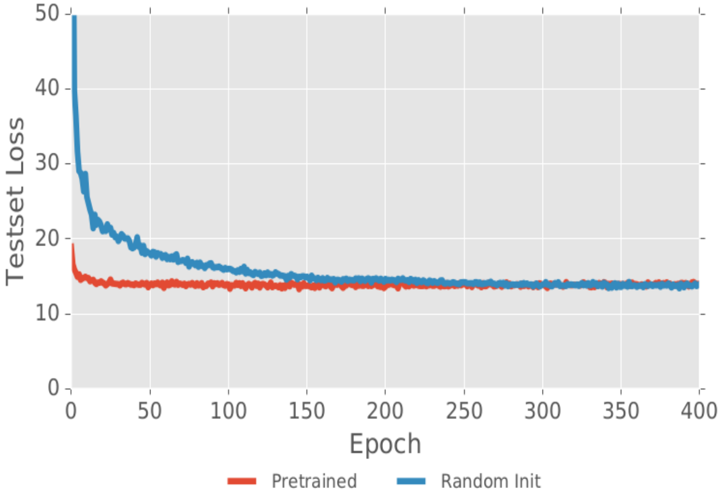
"Deep Gaussian Processes for Retrieval of bio-geo-physical parameters",
Svendsen, Ruescas and Camps-Valls, IEEE Trans. Geosc. Rem. Sens, 2019

5: XYZT Multioutput regression and transfer learning

- Multioutput regression: compactness & speed
- Transfer learning



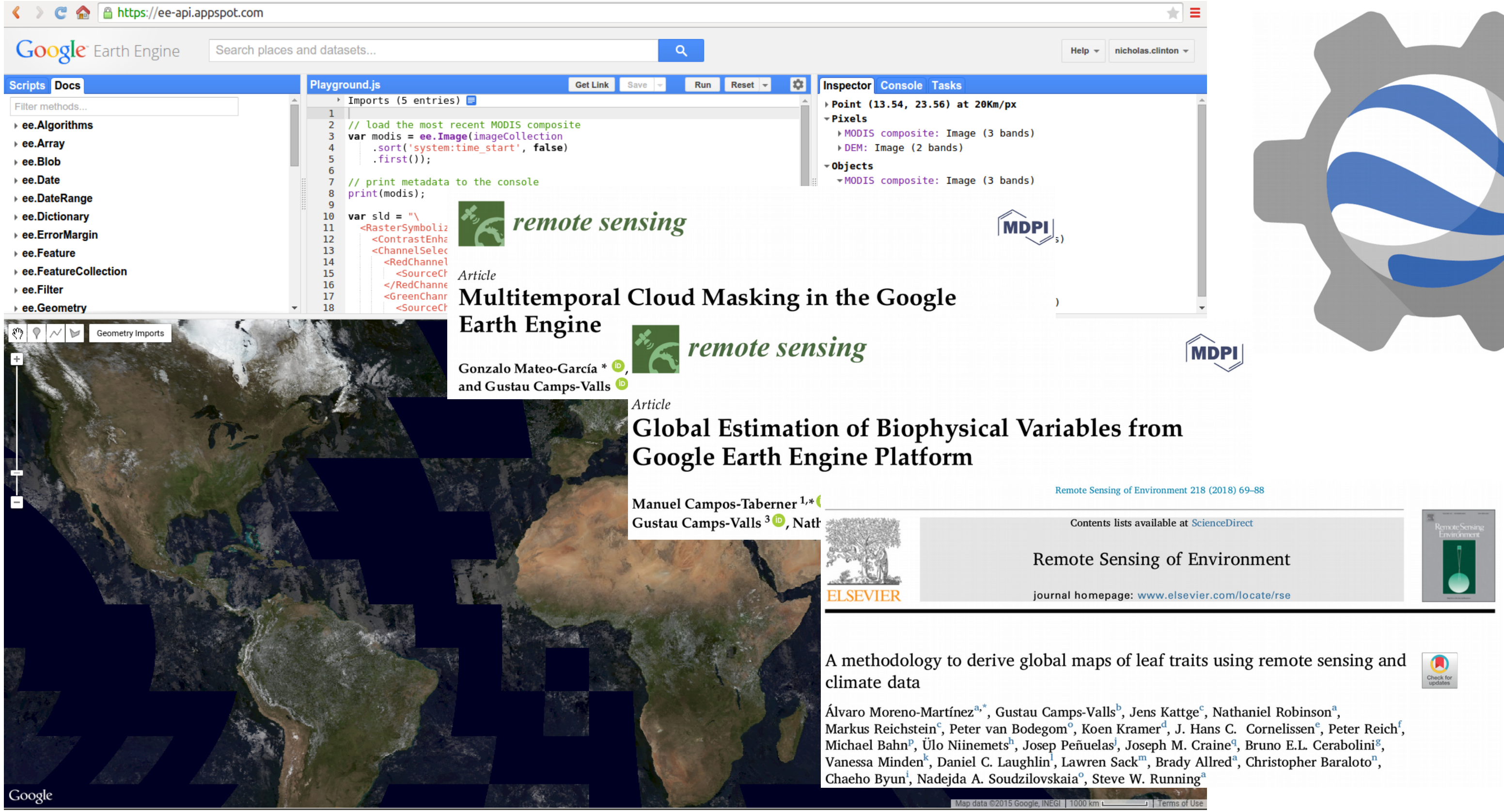




"Statistical Retrieval of Atmospheric Profiles with Deep Convolutional Neural Networks", Malmgren-Hansen, Laparra and Camps-Valls et al, IEEE Trans Geosc. Rem. Sens.. 2019.

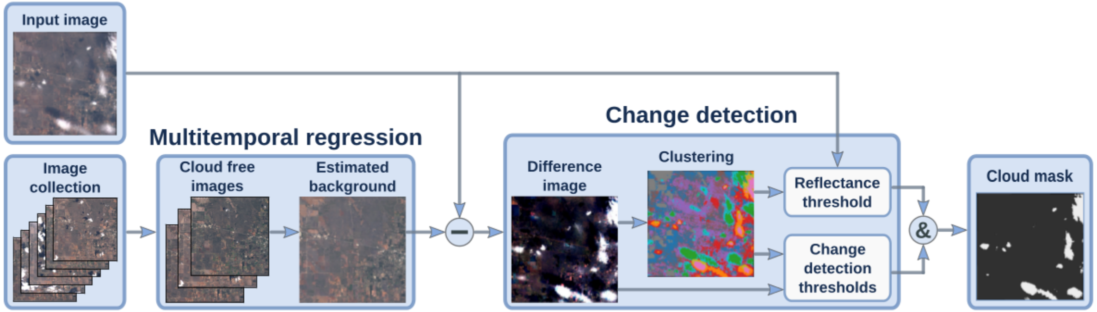
Efficiency

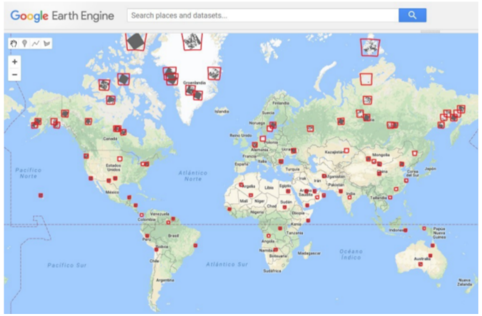
Google Earth Engine (GEE)

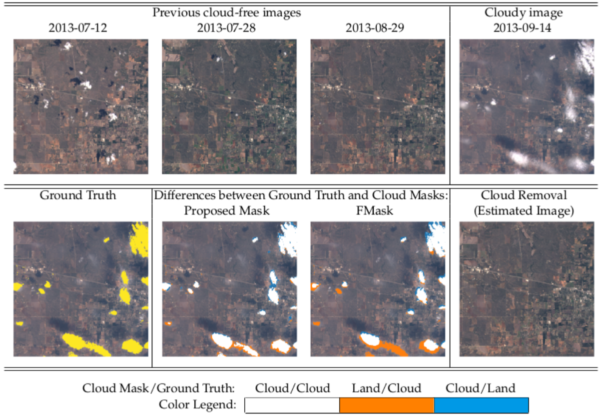


1: Google Earth Engine: cloud detection in the cloud

- Exploit temporal information and change detection



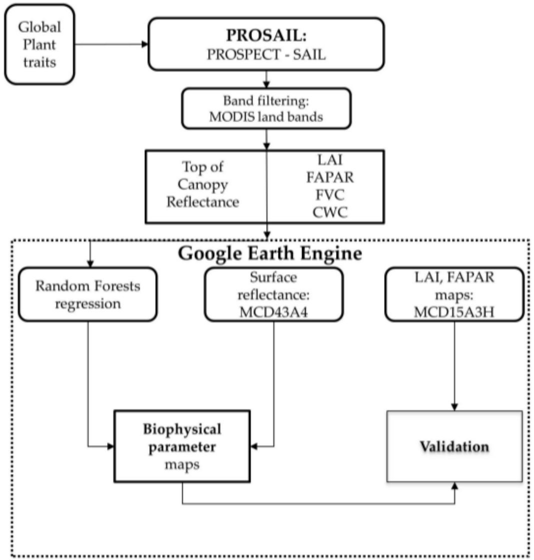




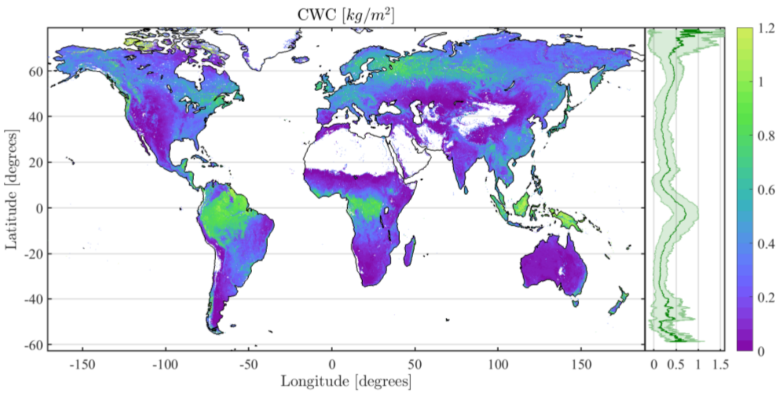
"Multitemporal Cloud Masking in the Google Earth Engine"
Mateo, Gómez, Amorós, Muñoz. and Camps-Valls. Remote Sensing 7 (10) :1079, 2018
"Cloud masking and removal in remote sensing image time series"
Gómez, Amorós, Mateo, Muñoz-Marí and Camps-Valls. Journal of Applied Remote Sensing 11 (1) :015005, 2017

2: Google Earth Engine: biophysical parameter retrieval

- Global maps of LAI, FAPAR, FVC, canopy water content by inverting PROSAIL with ML ...



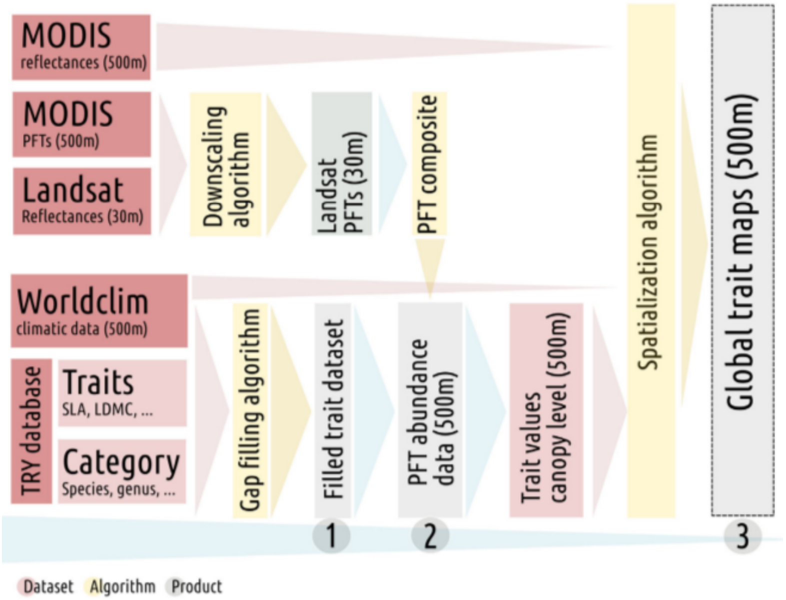
	Parameter	Min	Max	Mode	Std	Type
Leaf	N	1.2	2.2	1.6	0.3	Gaussian
	C_{ab} (μg·cm^{-2})	-	-	-	-	KDE *
	C_{ar} (μg·cm^{-2})	0.6	16	5	7	Gaussian
	C_{dm} (g·cm^{-2})	-	-	-	-	KDE *
	C_w	-	-	-	-	KDE *
	C_{bp}	0	0	0	0	-
Canopy	LAI (m^2/m^2)	0	8	3.5	4	Gaussian
	ALA (°)	35	80	60	12	Gaussian
	Hotspot	0.1	0.5	0.2	0.2	Gaussian
	vCover	0.3	1	0.99	0.2	Truncated Gaussian
Soil	β_s	0.1	1	0.8	0.6	Gaussian

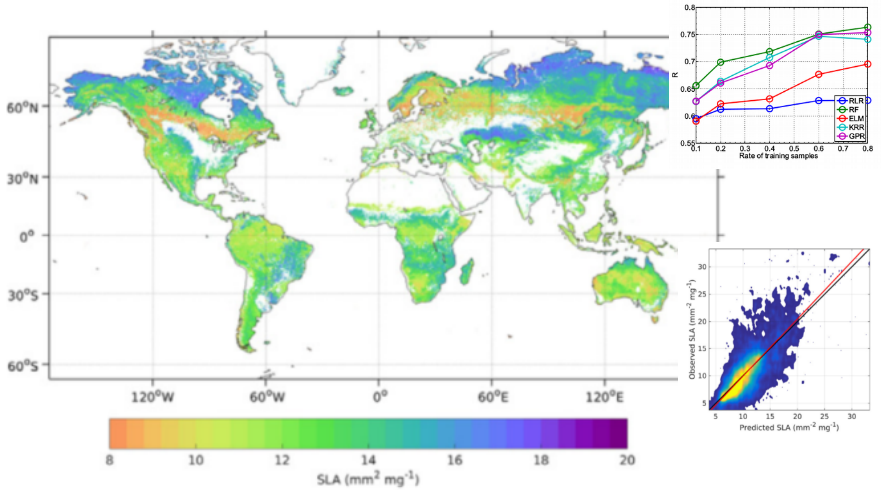


"Global estimation of biophysical variables from Google Earth Engine platform"
Campos, Moreno, Garcia, Camps-Valls, G. et al, Remote Sensing (10):1167, 2018

3: Google Earth Engine: spatialization of plant traits

- Global maps at 500 m resolution of specific leaf area, leaf dry matter content, leaf nitrogen and phosphorus content per dry mass, and leaf nitrogen/phosphorus ratio.



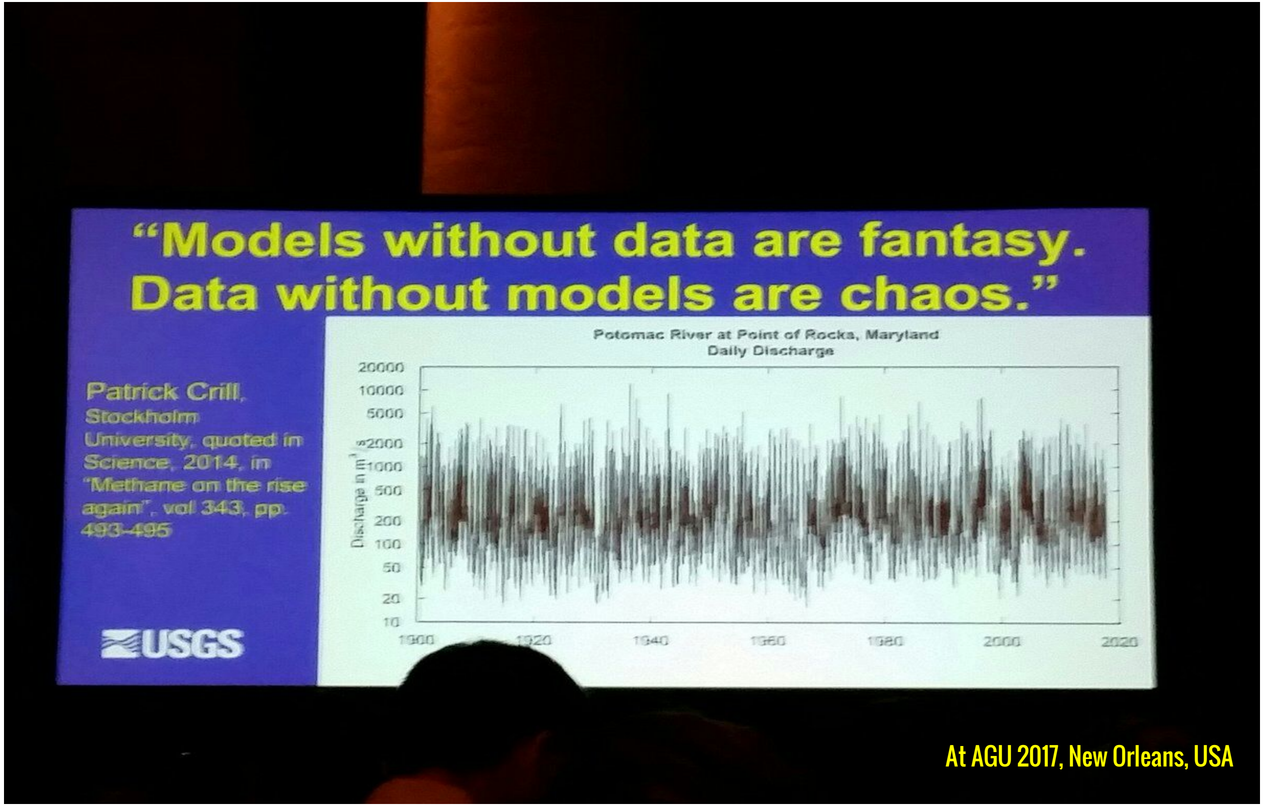


"A methodology to derive global maps of leaf traits using remote sensing and climate data"
Moreno, Camps-Valls, Kattge, Robinson, Reichstein, ... and Running.
Remote Sensing of Environment 218 (12) :69-88, 2018

Physics-aware machine learning

$$F(X, \frac{\partial c}{\partial t} + \mathbf{v}\nabla c = 0) = y$$

The truth is that...



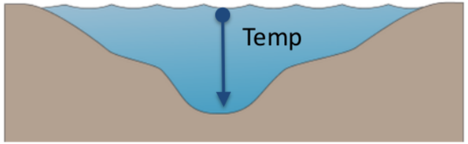
At AGU 2017, New Orleans, USA

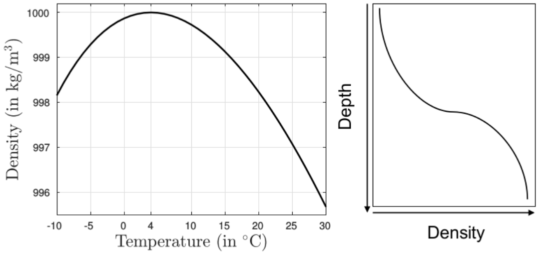
1: Physics-driven ML: constrained optimization

- ML that minimizes model violations and predictions are dependent of physical laws

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

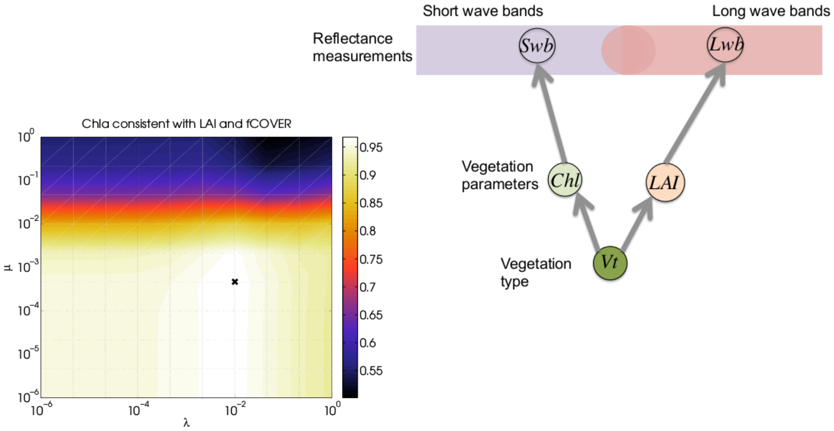
$$\Omega(\hat{y}, \Phi) = \text{sum of physical violations of } \hat{y}$$





"Theory-guided Data Science", Karpatne, A. et al. IEEE Trans. Know. Data Eng., 2017.

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



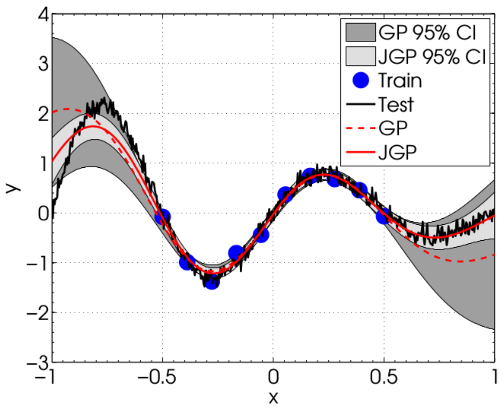
"Fair Kernel Learning" Perez, Laparra, Gomez, Camps-Valls, G. ECML, 2017.
"Consistent Regression of Biophysical Parameters with Kernel Methods" Díaz, Peréz-Suay, Laparra, Camps-Valls, IGARSS 2018

2: Physics-driven ML: joint model-data ML

- Let ML talk to physical models

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

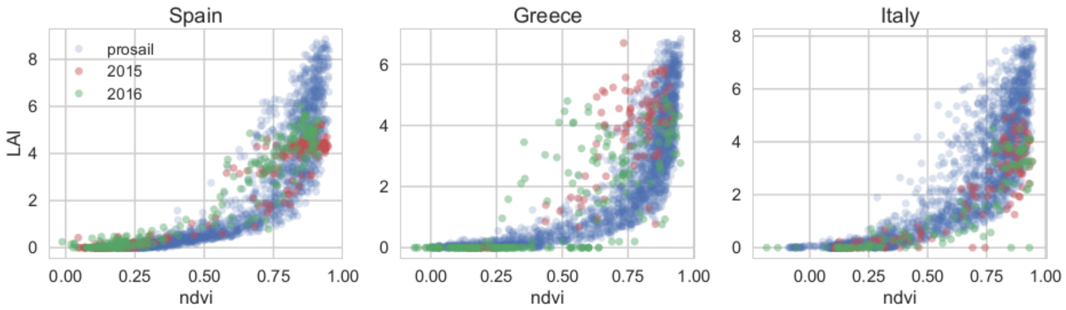
$$\Omega(\hat{y}, \Phi) = \mathrm{Cost}_s(y_s, \hat{y}_s)$$



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)

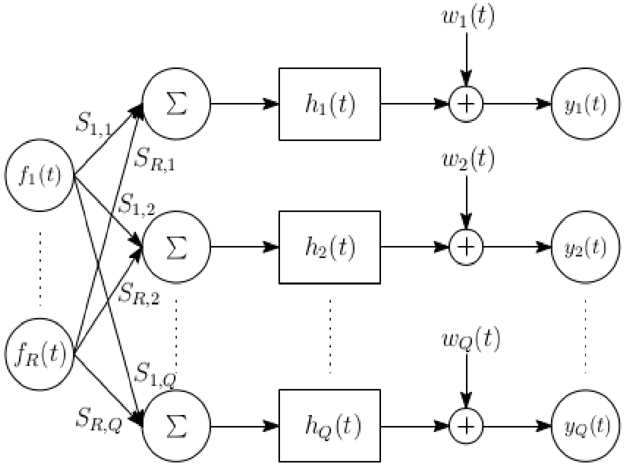
Filling the space ...



"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.

3: Multioutput GP regression encoding ODEs

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



"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, Nat Geosc, 2020

1. **Latent forces** $f_r(t)$: zero-mean GPs with covariance function

$$k_{f_r f_r}(t'-t) \propto \exp\left(-\frac{(t'-t)^2}{2\ell_r^2}\right),$$

as vegetation should be smooth and exhibit local relations

2. **Coupling mechanism** $f_r(t) \leftrightarrow y_q(t)$: linear convolution operator with $h_q(t)$

$$h_q(t) \propto \exp\left(-\frac{t^2}{2\nu_q^2}\right)$$

Green's func. of heat diffusion eq.

as rate of change of $y \propto$ curvature of y

3. **Outputs** as lin. combination of pseudo-outputs plus AWGN:

$$y_q(t) = \sum_{r=1}^{R} S_{rq} y_{rq}(t) + w_q(t), \quad w_q(t) \sim \mathcal{N}(0, \eta_q^2)$$

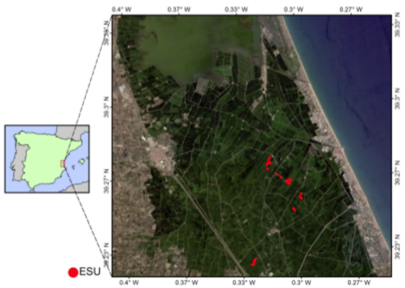
where S_{rq} accounts for the coupling strength

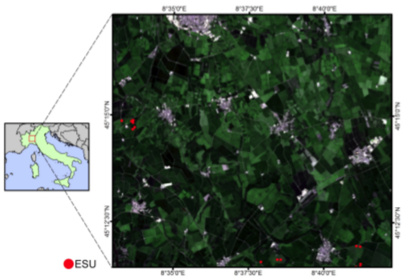
3: Multioutput GP regression encoding ODEs

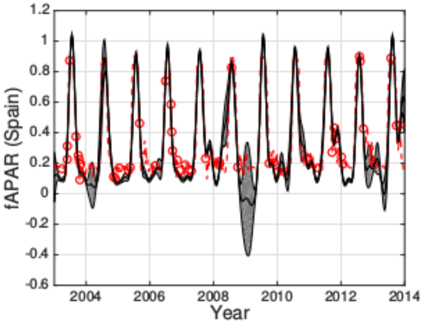
- Example: LAI and FAPAR across time and space

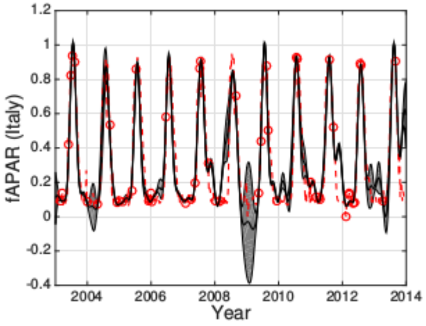
- Time series of LAI and fAPAR variables for rice monitoring
- Spain, Italy, Greece ∼85% of Europe rice production
- H2020 ERMES project: `http://www.ermes-fp7space.eu/`
- Observe inter-annual variability of rice 2003–2014

- 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.
“Latent force GP models for EO time series prediction”
Luengo, Muñoz, Piles, Camps-Valls, IEEE TGARS, 2019

4: Physics-driven ML: hybrid modeling framework

PERSPECTIVE

https://doi.org/10.1038/s41586-019-0912-1

Deep learning and process understanding for data-driven Earth system science

Markus Reichstein[1,2]*, Gustau Camps-Valls[3], Bjorn Stevens[4], Martin Jung[1], Joachim Denzler[2,5], Nuno Carvalhais[1,6] & Prabhat[7]

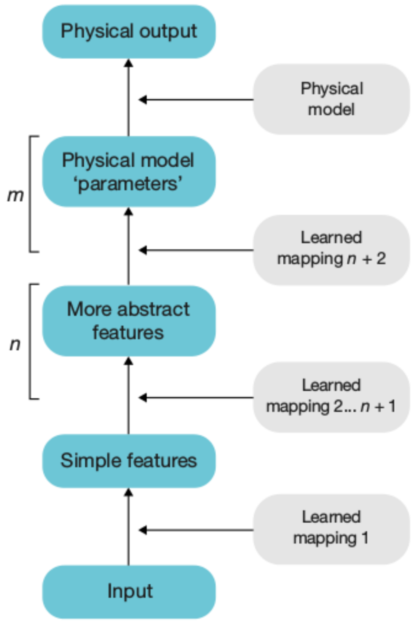
Machine learning approaches are increasingly used to extract patterns and insights from the ever-increasing stream of geospatial data, but current approaches may not be optimal when system behaviour is dominated by spatial or temporal context. Here, rather than amending classical machine learning, we argue that these contextual cues should be used as part of deep learning (an approach that is able to extract spatio-temporal features automatically) to gain further process understanding of Earth system science problems, improving the predictive ability of seasonal forecasting and modelling of long-range spatial connections across multiple timescales, for example. The next step will be a hybrid modelling approach, coupling physical process models with the versatility of data-driven machine learning.

"Deep learning and process understanding for data-driven Earth System Science", Reichstein, Camps-Valls et al. Nature, 2019.

4: Physics-driven ML: hybrid modeling framework

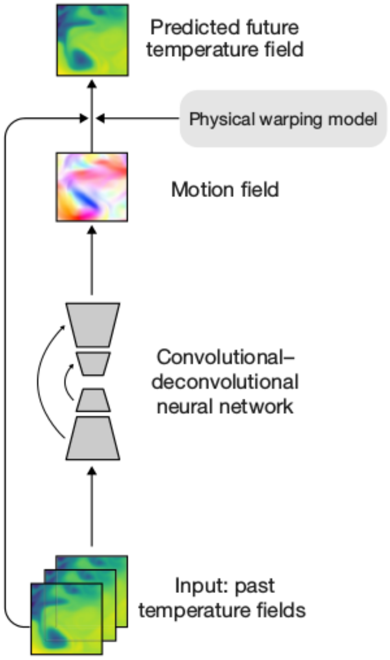
- ML that learns laws of physics (e.g. consistency model-data, convection, advection, mass and energy conservation)

A: "Physisizing" a deep learning architecture by adding one or several physical layers after the multilayer neural network



"Deep learning and process understanding for data-driven Earth System Science"
Reichstein, Camps-Valls et al. Nature, 2019.

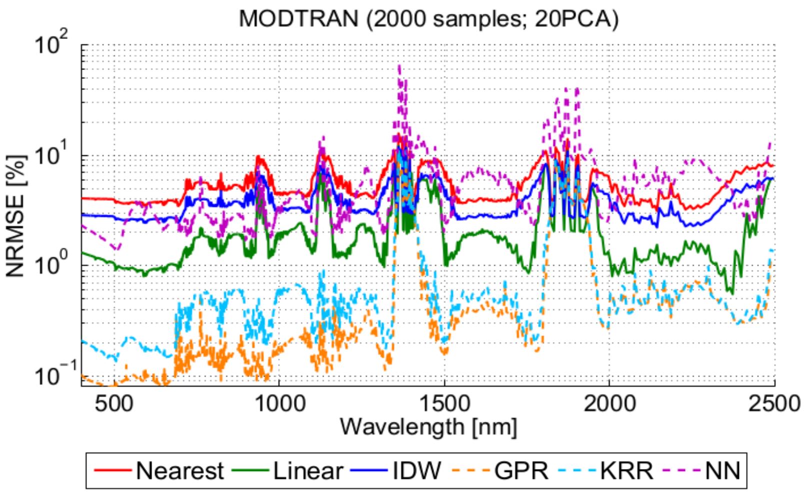
B: A motion field is learned with a convolutional-deconvolutional net, and the motion field is further processed with a physical model

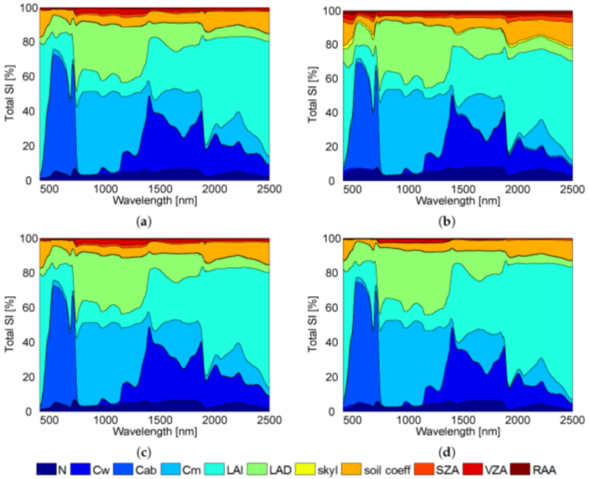


"Deep Learning for Physical Processes: Incorporating Prior Scientific Knowledge".
de Bezenac, Pajot, & Gallinari, arXiv:1711.07970 (2017)

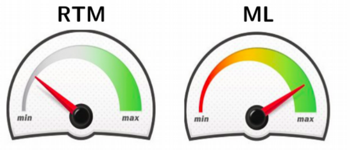
5: Physics-driven ML: emulation of complex codes

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





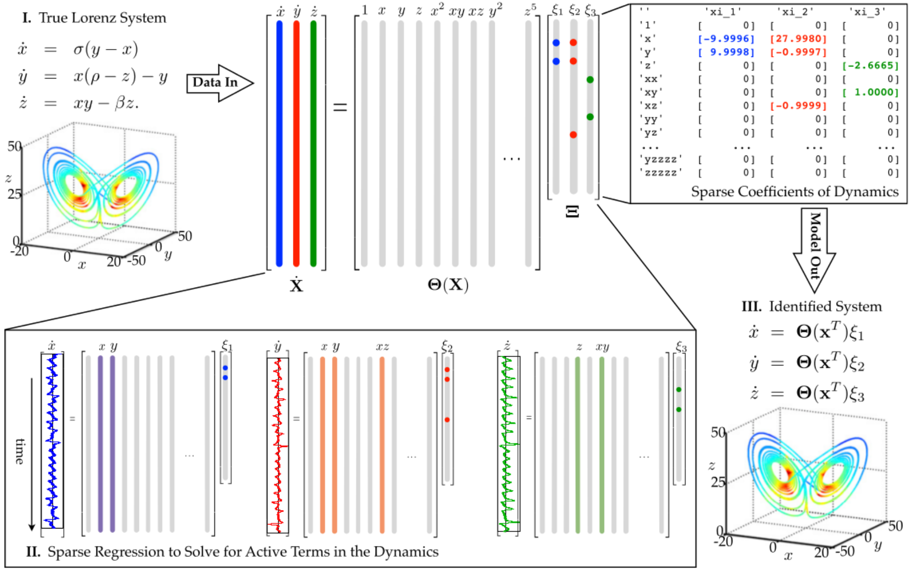
“Emulation of Leaf, Canopy and Atmosphere Radiative Transfer Models for Fast Global Sensitivity Analysis”,
Verrelst, Camps-Valls et al Remote Sensing of Environment, 2016
“Emulation as an accurate alternative to interpolation in sampling radiative transfer codes”,
Vicent and Camps-Valls, IEEE Journal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



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

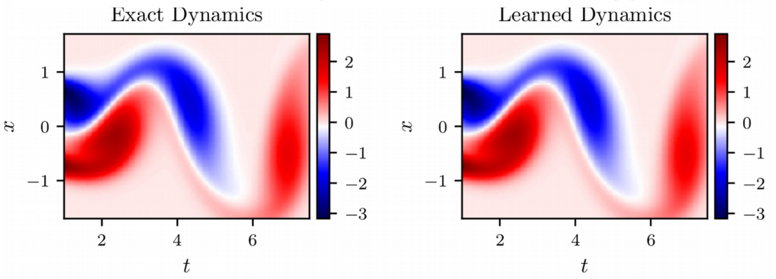
6: Physics-driven ML: learning ODE/PDEs

- Who needs Lorenz?

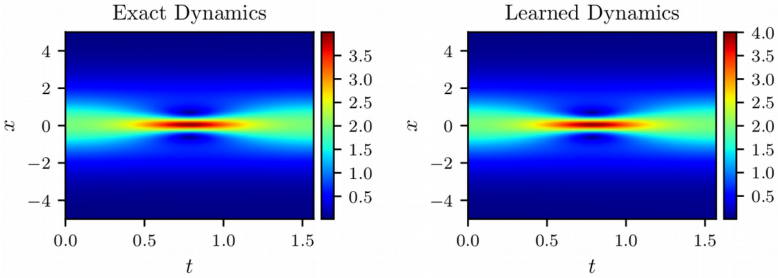


- Who needs Navier Stokes?

$$w_t = -uw_x - vw_y + 0.01(w_{xx} + w_{yy})$$



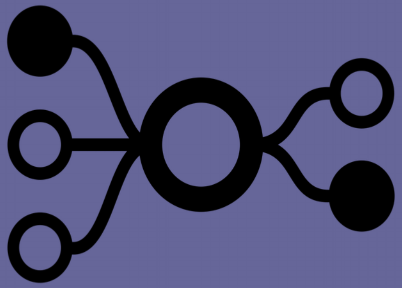
- Who needs Schrödinger? $\psi_t = 0.5i\psi_{xx} + i|\psi|^2\psi$



"Discovering governing equations from data by sparse identification of nonlinear dynamical systems" Brunton, Proctor, Kutz, PNAS 2016

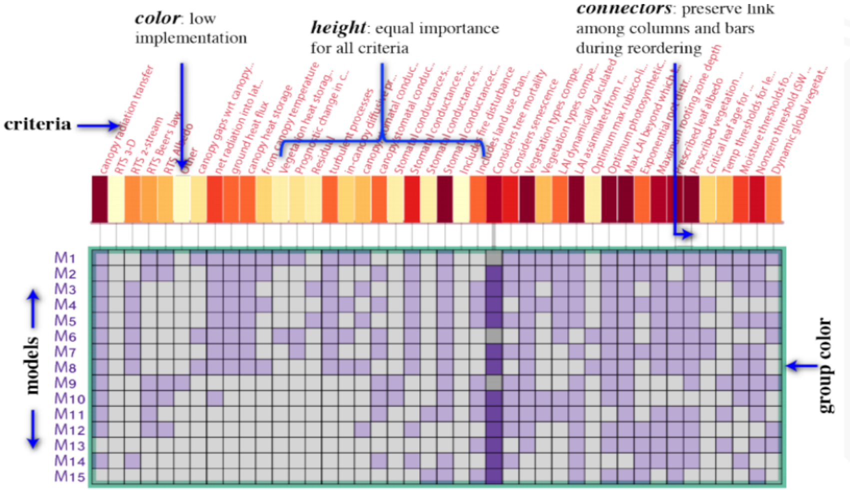
"Deep Hidden Physics Models: Deep Learning of Nonlinear Partial Differential Equations" Raissi, JMLR 2018

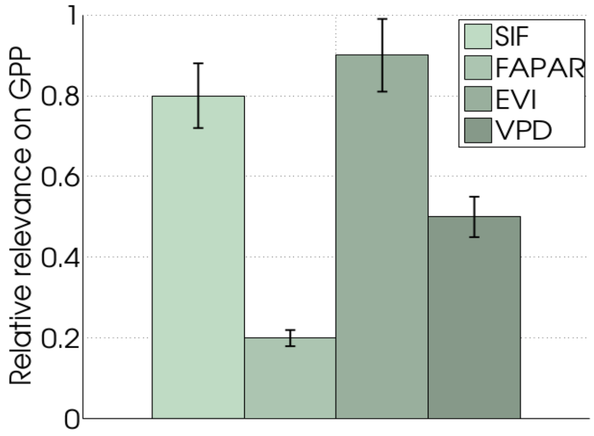
Understanding is more important than fitting!

F(X,Y) = 

1: Feature selection & ranking

- **Filters & wrappers**

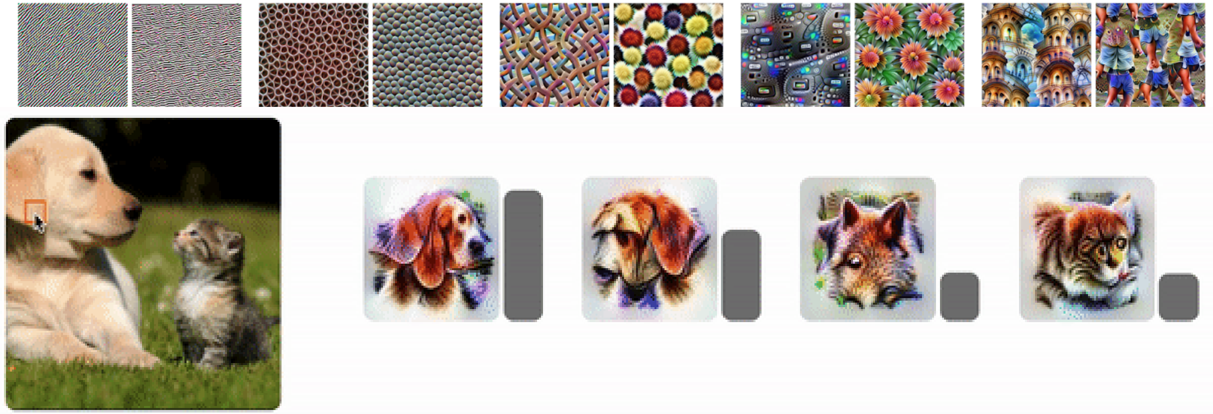


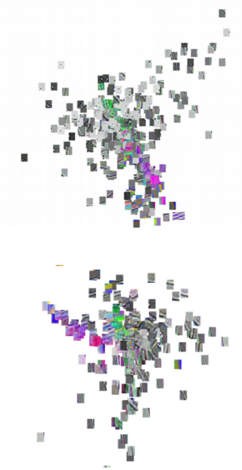


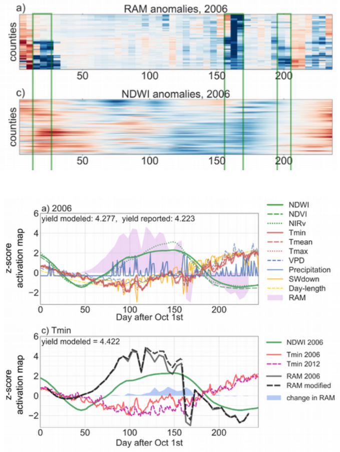
"Remote Sensing Feature Selection by Kernel Dependence Estimation", Camps-Valls, G. Mooij, JM. Schölkopf, IEEE-GRSL, 2010.
"A guided hybrid genetic algorithm for feature selection with expensive cost functions", M. Jung, J. Zscheischler, Procedia, 2013.

2: eXplainable AI (XAI)

- **What did the network learn?** Look at the heatmaps & triggering neurons
- **How do bases change in time and space? And under extremes?**

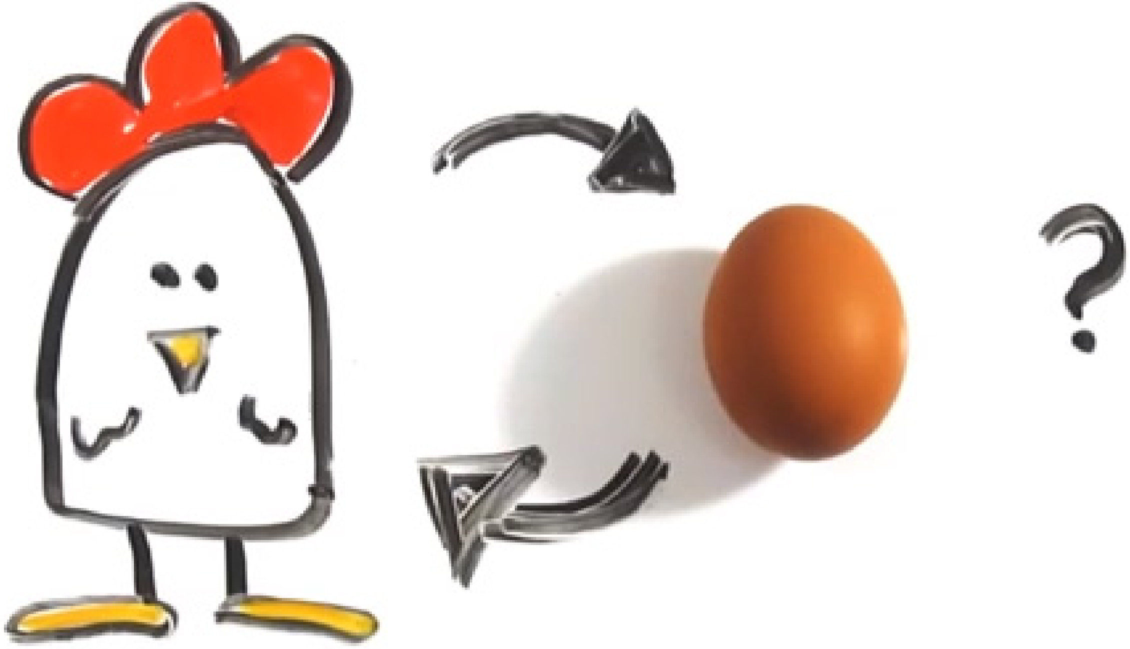






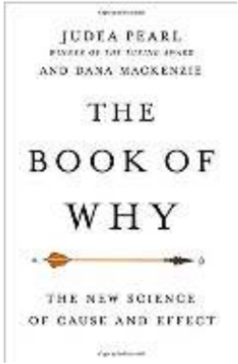
"Visualizing and Understanding Convolutional Networks", Zeriler, et al 2013
"Processing of Extremely high resolution LiDAR and optical data", Campos-Taberner, Camps-Valls et al, 2016
"Understanding convolutional neural nets for crop yield estimation" Wolanin, Guanter, Camps-Valls, ERL 2020

3: Causal inference



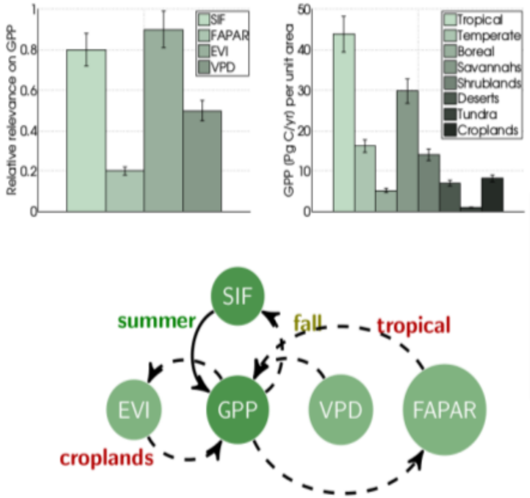
Possible? Yes, under some mild assumptions ...

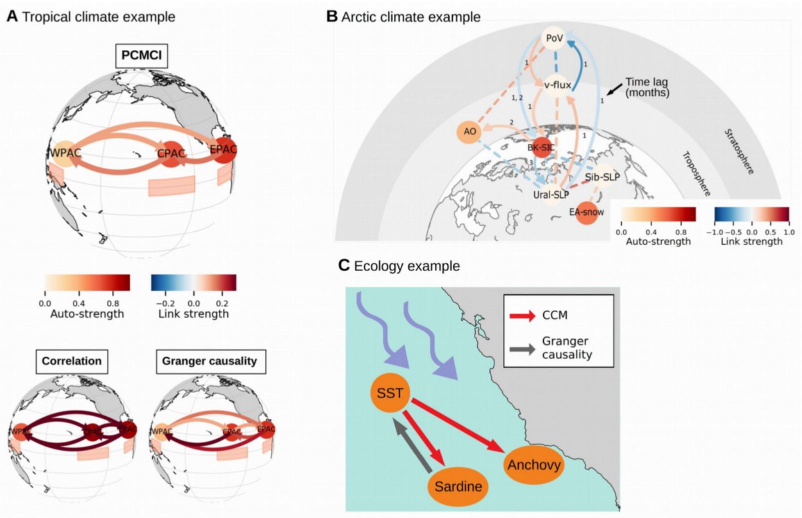
- UCLA: Judea Pearl
- CMU: Peter Spirtes, Clark Glymour, Richard Scheines
- Harvard: Donald Rubin, Jamie Robins
- ETH Zürich: Peter Bühlmann, Nicolai Meinshausen
- MPI Tübingen: Dominik Janzing, Bernhard Schölkopf
- Univ. Amsterdam: Joris Mooij
- Univ. Copenhagen: Jonas Peters
- Aalto Univ.: Patrik Hoyer
- ... and many others



3: Causal inference

- **Causal discovery** learns cause and effects relations from data
- **What for?** Hypothesis testing, model-data comparison, causes of extreme impacts





"Inferring causation from time series with perspectives in Earth system sciences", Runge, Bathiany, Bollt, Camps-Valls, et al. Nat Comm., 2019.
"Causal Inference in Geoscience and Remote Sensing from Observational Data," Pérez-Suay and Camps-Valls, IEEE Trans. Geosc. Rem. Sens, 2018
"CauseMe: An online system for benchmarking causal inference methods," Muñoz-Marí, Mateo, Runge, Camps-Valls. In preparation (2020). **CauseMe: http://causeme.uv.es**

3: Causal inference methods

nature COMMUNICATIONS

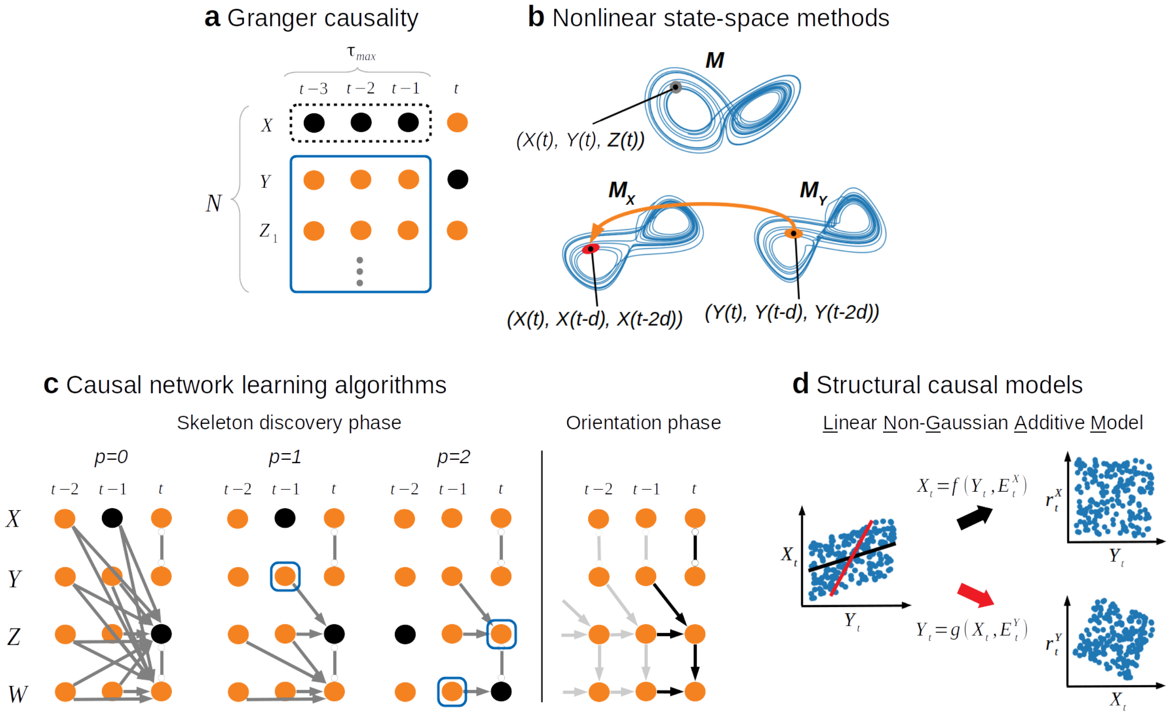
PERSPECTIVE

https://doi.org/10.1038/s41467-019-10105-3 OPEN

Inferring causation from time series in Earth system sciences

Jakob Runge[1,2], Sebastian Bathiany[3,4], Erik Bollt[5], Gustau Camps-Valls[6], Dim Coumou[7,8], Ethan Deyle[9], Clark Glymour[10], Marlene Kretschmer[8], Miguel D. Mahecha[11], Jordi Muñoz-Marí[6], Egbert H. van Nes[4], Jonas Peters[12], Rick Quax[13,14], Markus Reichstein[11], Marten Scheffer[4], Bernhard Schölkopf[15], Peter Spirtes[10], George Sugihara[9], Jie Sun[5,16], Kun Zhang[10] & Jakob Zscheischler[17,18,19]

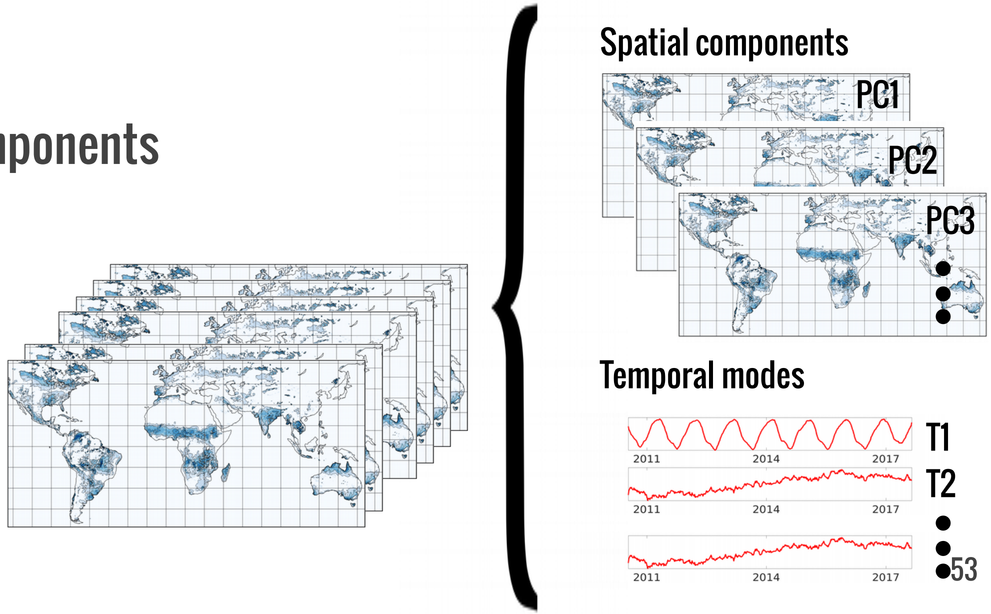
3: Causal inference methods



"**Inferring causation from time series with perspectives in Earth system sciences**", Runge, Bathiany, Bollt, Camps-Valls, et al. Nat Comm., 2019.
"**Causal Inference in Geoscience and Remote Sensing from Observational Data,**" Pérez-Suay and Camps-Valls, IEEE Trans. Geosc. Rem. Sens, 2018
"**CauseMe: An online system for benchmarking causal inference methods,**" Muñoz-Marí, Mateo, Runge, Camps-Valls. In preparation (2020). **CauseMe: http://causeme.uv.es**

Example 1: Spatio-temporal causal analysis of Earth cubes

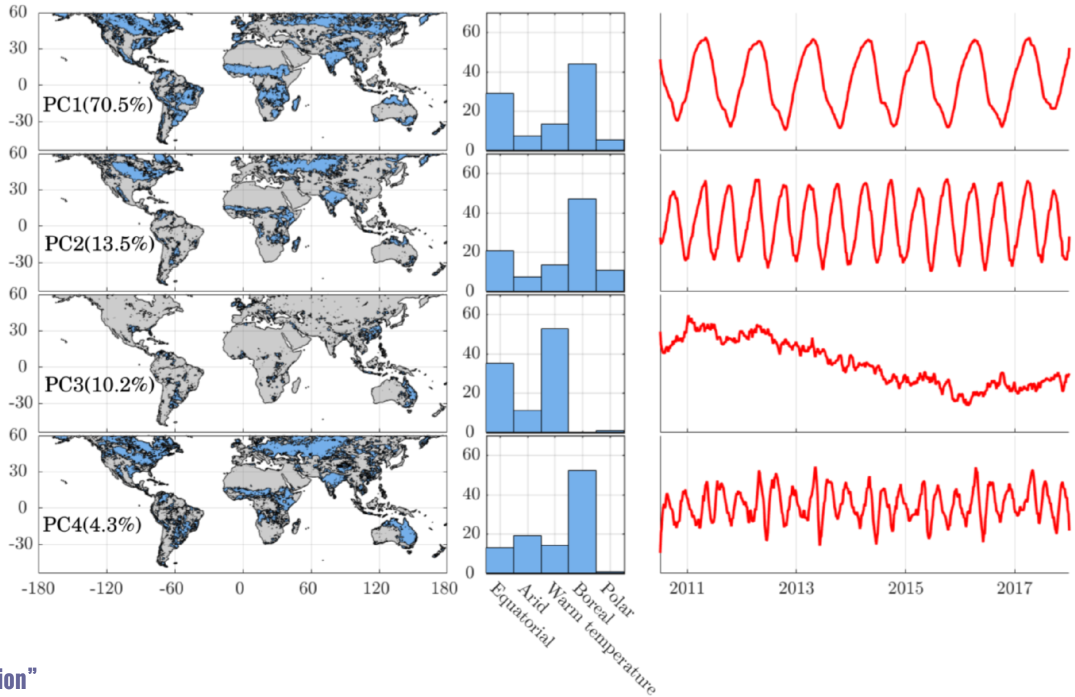
- PCA/EOF is popular, yet cannot cope with nonlinear spatio-temporal relations
- ROCK PCA
  - copes with nonlinearities
  - extracts spatial and temporal components
  - very fast



“Rotated Complex Kernel PCA for spatio-temporal data decomposition”
Bueso, Piles, Camps-Valls, IEEE TGARS, 2020

Example 1: Spatio-temporal causal analysis of Earth cubes

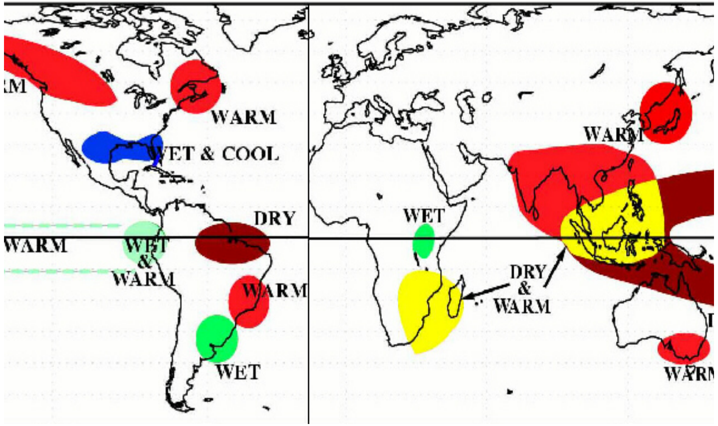
- SM decomposition
  - Meaningful compression
  - Climate-specific modes of variability
  - Boreal and Equatorial modes of SM variability dominate
  - Seasonal and ENSO related temporal modes

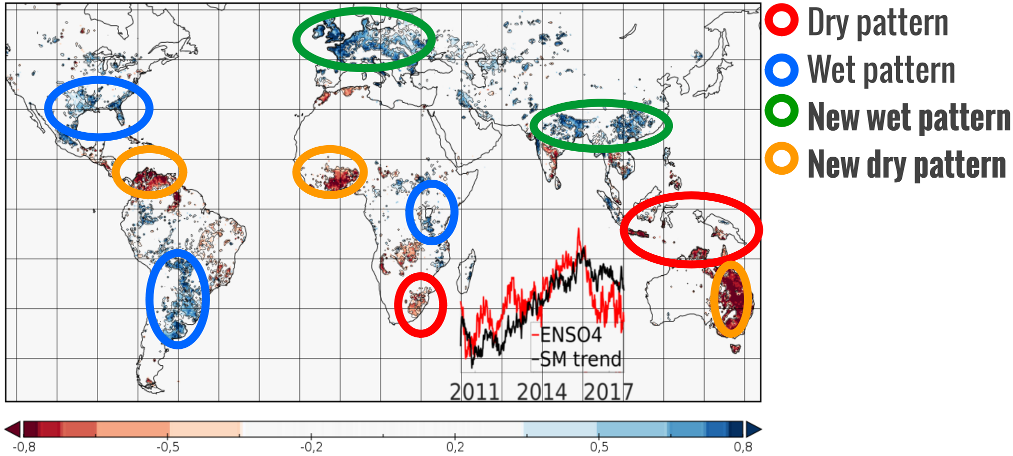


“Rotated Complex Kernel PCA for spatio-temporal data decomposition”
Bueso, Piles, Camps-Valls, IEEE TGARS, 2020

Example 1: Spatio-temporal causal analysis of Earth cubes

- PC3 highly correlates with ENSO + new spatial patterns uncovered





- Nonlinear cross-correlation uncovers unreported SM-ENSO lags

	ENSO 1.2	ENSO 3	ENSO 3.4	ENSO 4
Lag [days]	60	30	25	5
Max Corr	0.56	0.68	0.66	0.8

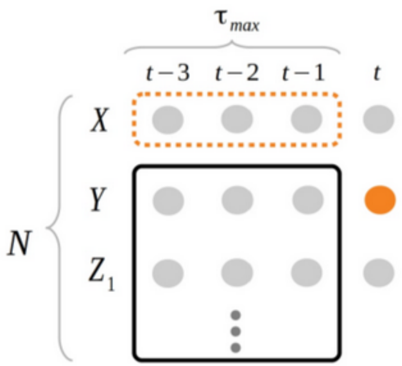
"Rotated Complex Kernel PCA for spatio-temporal data decomposition"
Bueso, Piles, Camps-Valls, IEEE TGARS, 2020

Correlation is not enough: Nonlinear Granger causality

- ENSO4 index and the inter-annual component extracted from VOD and SM
- High correlations, yet … correlation does not imply causation

Correlation is not enough: Nonlinear Granger causality

- Causal inference goes beyond correlation analysis
- Granger causality tests whether the past of X is useful to predict the future of Y



$$Y_{t+1} = a^\top X_t + \varepsilon_t^Y$$

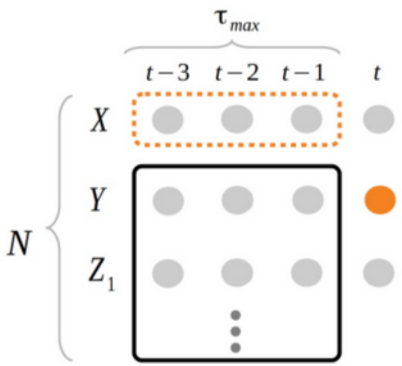
$$Y_{t+1} = b_1^\top Y_t + b_2^\top X_t + \varepsilon_t^{Y|X}$$

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

"Causal inference from Observational Data in Remote Sensing and Geosciences"
Perez-Suay and Camps-Valls, IEEE TGARS 2019
"Inferring causation from time series with perspectives in Earth system sciences"
Runge, J. Bollt, E. Camps-Valls, G. Peters, J. Reichstein, M., Schölkopf, B. et al. Nature Communications, 2019

Correlation is not enough: Nonlinear 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 kernel Granger method to account for nonlinear Granger-causal relations



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

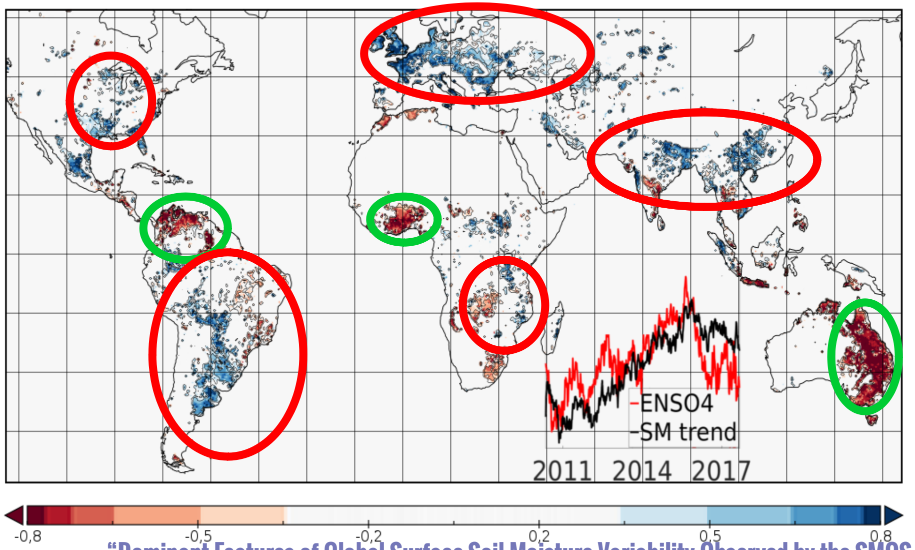
$$b_H = (L([Y_t, X_t], [Y_t', X_t']) + \varepsilon_t^{Y|X})^{-1} Y_{t+1}$$

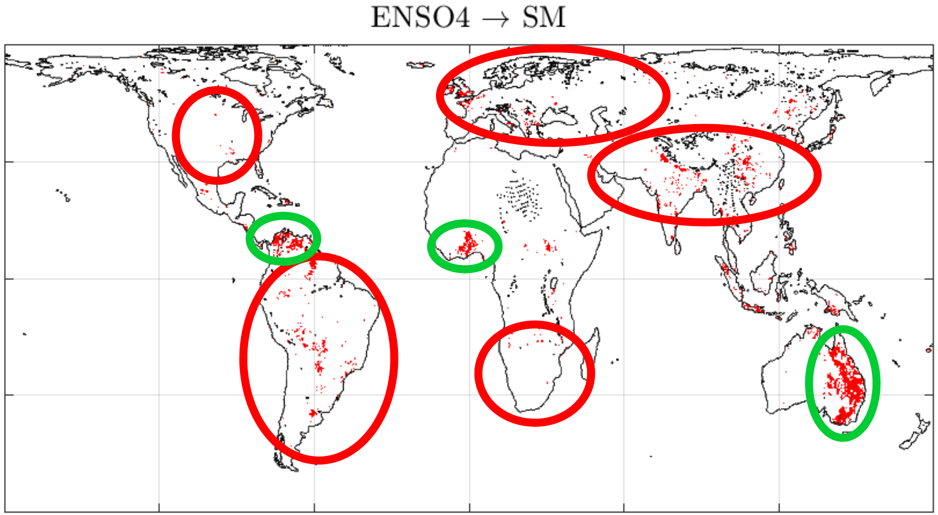
$$X \to Y \leftrightarrow \mathbb{V}_H[\varepsilon_t^Y] \ll \mathbb{V}_H[\varepsilon_t^{Y|X}]$$

"Causal inference from Observational Data in Remote Sensing and Geosciences"
Perez-Suay and Camps-Valls, IEEE TGARS 2019
"Inferring causation from time series with perspectives in Earth system sciences"
Runge, J. Bollt, E. Camps-Valls, G. Peters, J. Reichstein, M., Schölkopf, B. et al. Nature Communications, 2019

Correlation is not enough: Nonlinear 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)

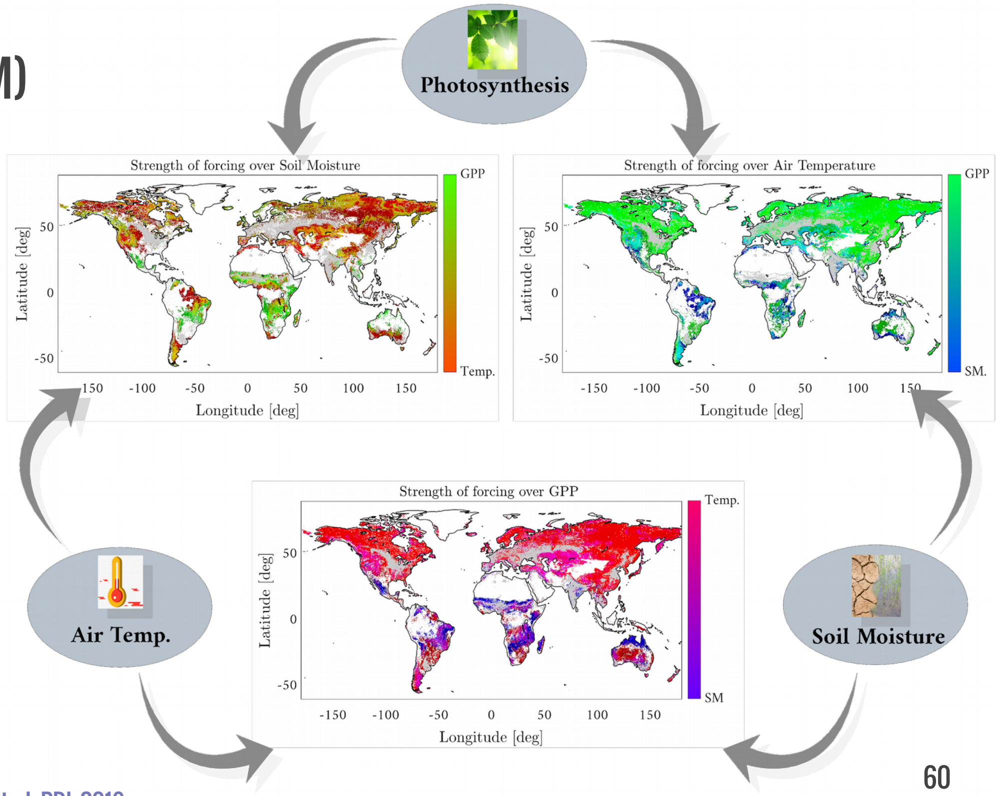




"Dominant Features of Global Surface Soil Moisture Variability Observed by the SMOS Satellite" M. Piles et al. Remote Sensing, 2019

Example 2: Water and energy fluxes causal relations

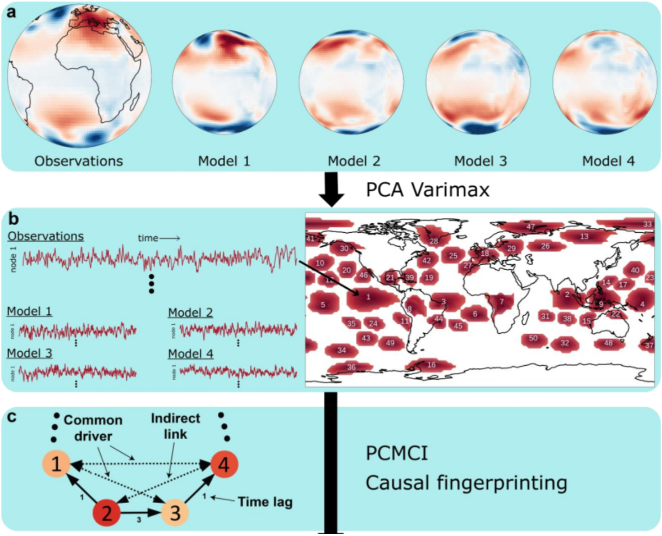
- Unbiased CCM causality on (GPP, Tair, SM)
- Causal maps capture general knowledge
- In dry (water-limited) areas, GPP is caused/driven by SM
- Temperature is mainly an effect in boreal regions
- GPP affects SM in dry/savannas/shrubs, possibly related through ET
- SM in boreal regions matches with a reduction in radiation and temperature

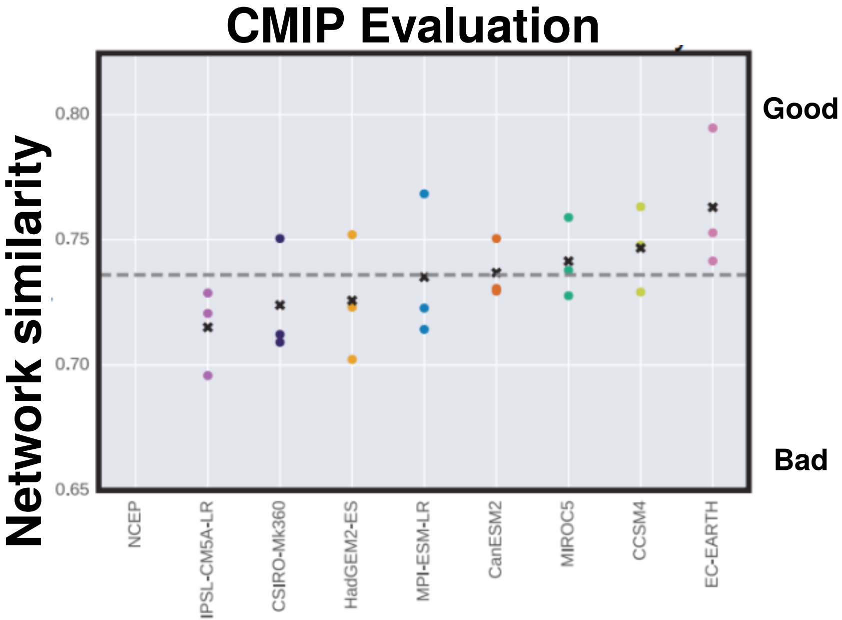


"Unbiased estimation of causal drivers in convergent cross-mapping" G. Camps-Valls et al, PRL 2019

Example 3: climate models (causal) intercomparison

- How similar are the causal mechanisms encoded in the models?
- Do they match observations?

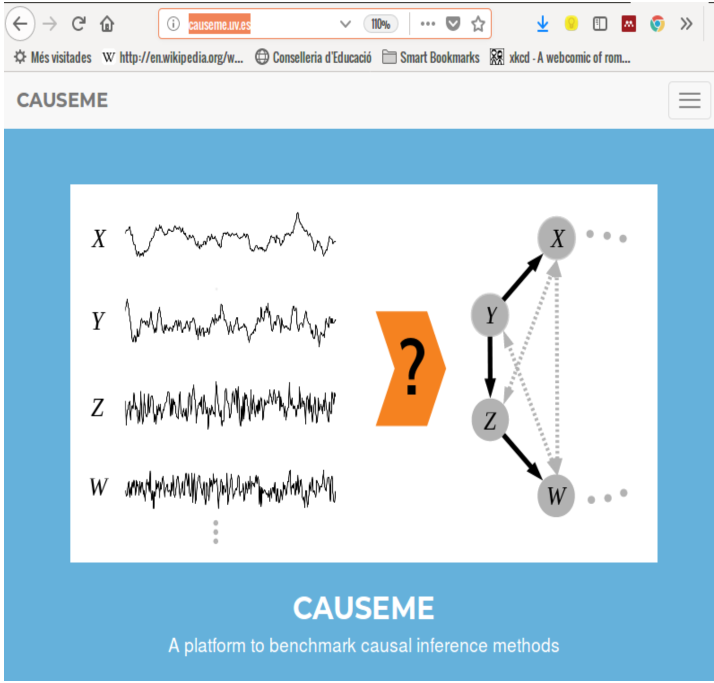




Nowack, Runge, Eyring, Haigh, Nature Comm., subm. (2019)

A platform for causal discovery

- CauseMe: http://causeme.uv.es
  - Download time series with ground truth
  - Run your causal discovery algorithm offline
  - Upload your causal graph
  - Get your results!

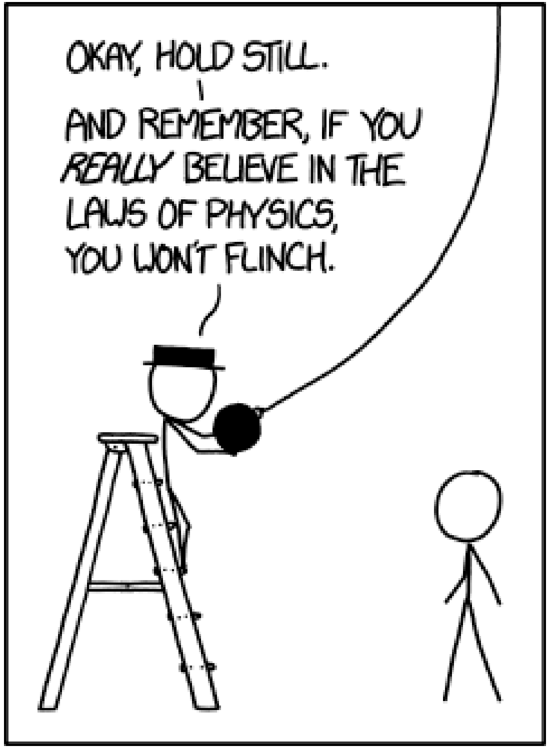


"Inferring causation from time series with perspectives in Earth system sciences"
Runge, Bathiany, Bollt, Camps-Valls, et al. Nat Comm, 2019
"Causal Inference in Geoscience and Remote Sensing from Observational Data,"
Pérez-Suay and Camps-Valls, IEEE Trans. Geosc. Rem. Sens, 2018

Conclusions

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
    - Fitting and classification
- Risks & remedies
    - Understanding is more complex
    - Physics consistency a must
    - Physics-driven ML & Explainable AI



Thanks!

http://isp.uv.es

@isp_uv_es

gustau.camps@uv.es








Propaganda

- ERC CoG: SEDAL: “Statistical Learning for Earth Observation Data Analysis”
- ERC SyG: USMILE: “Understanding and Modeling the Earth System with ML”
  - Need 10 PhD and postdocs!
  - Check out & distribute: isp.uv.es/openings
  - Starting in June/september
- Expertise in:
  - Machine learning
  - (Earth observation) Physics
  - Environmental Sciences
  - Maths
  - Computer science
- València rocks!

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

Propaganda

- ERC SyG: USMILE: “Understanding and Modeling the Earth System with ML”

	Current problems	Standard approximations	Alternative USMILE approximations	WP
Modelling and prediction	Signal relations (spatial, temporal, spectral)	Spatio-temporal relations not included in ML	Convolutional Neural Networks (CNN)[23] and Recurrent Neural Networks (RNN)[28] to account for spatio-temporal relations	1-3
	Strong a-priori assumptions, e.g. linearity, Gaussianity, stationarity	Nonlinear and nonparametric neural networks (NN), kernel and graphical models	Advanced probabilistic Gaussian processes[30] to account for uncertainty quantification, data fusion, and multioutput regression with consistency regularizers	1-3
	Inconsistencies, large uncertainties, overfitting	Standard data assimilation approaches	New paradigm of hybrid modeling[19,30,31], ML respecting and implementing physics laws	1-3
	Anomaly and extreme event detection	Thresholds, univariate or Gaussian assumption	Generative neural networks[32,33] , normalizing flows[34], kernel detectors[35], multivariate anomaly detection[36]	3
Understanding	Black boxes, interpretability	Permutation analysis, global sensitivity analysis[37]	(1) Visualization of convolutional filters and memory units[38,39] in NN; (2) New sparse regularizers[40,41] in regression for (i) selection of covariates, (ii) descriptors and potential causal links [25,42]	3, 4
	Just dependence, no causation, spurious correlations found	Single scale, linearity, Granger causality[25,43]	Conditional independence tests[44,45], multiscale (in space and time) wavelet-based causal schemes	3, 4
Data	High-dimensionality	Linear dimensionality reduction, decoupled space and time, linear principal component analysis (PCA)[25]	Advanced spatio-temporal nonlinear methods based on kernel methods[46] and variational autoencoders[29]	3, 4
	Large volume	High performance computing	Cloud computing	1

http://isp.uv.es

@isp_uv_es

gustau.camps@uv.es

ellis.eu

- Machine Learning for Earth and Climate Sciences (G. Camps-Valls & M. Reichstein)
  - Spatio-temporal anomaly and extreme events detection
  - Dynamic modeling and forecasting
  - Hybrid modeling: linking physics and machine learning
  - Causal inference: learning and explaining representations
  - Earth and climate model emulation and generative modeling