Machine learning for modeling and understanding the Earth system

Gustau Camps-Valls
Image Processing Laboratory
Universitat de València

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

Dobro jutro!

Let’s start!

Disclaimer:
many methods, many problems ahead!

Earth science – on the what, when, why and how questions

2 weeks ago …

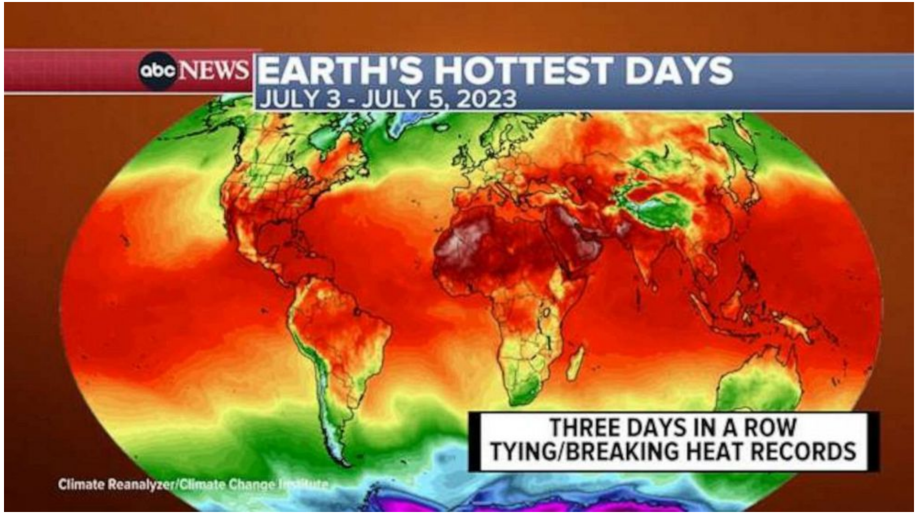
REUTERS® World Business Markets Sustainability Legal Breakingviews Technology Inv

Environment

World breaks hottest-day record for third time this week, U.S. agency says

By **Jake Spring**

July 7, 2023 5:15 PM GMT+2 · Updated 18 days ago



Last week …

The New York Times

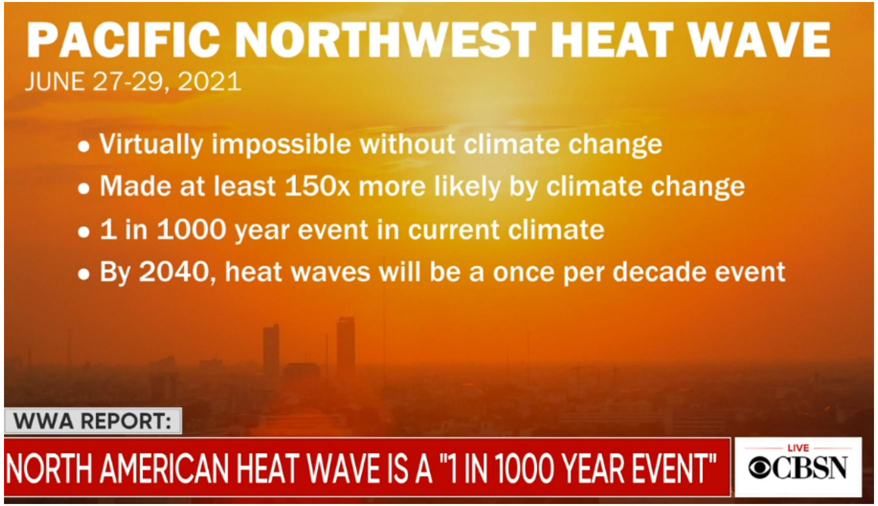
Extreme Heat | U.S. Forecast | Europe Forecast | Global Heat Tracker | U.S. Weather by the Numbers

Extreme Heat

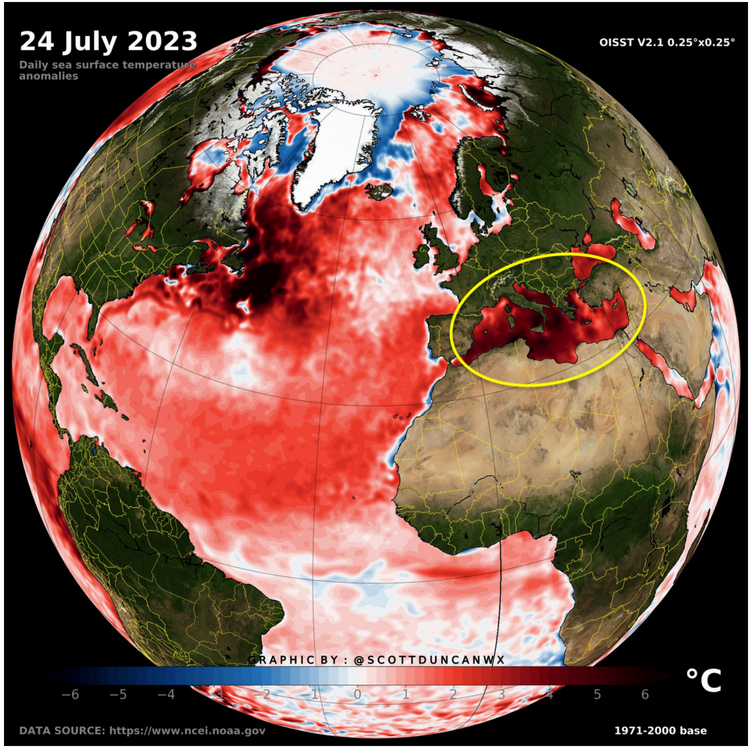
Phoenix Breaks Record With 19 Consecutive Days 110 Degrees or Higher

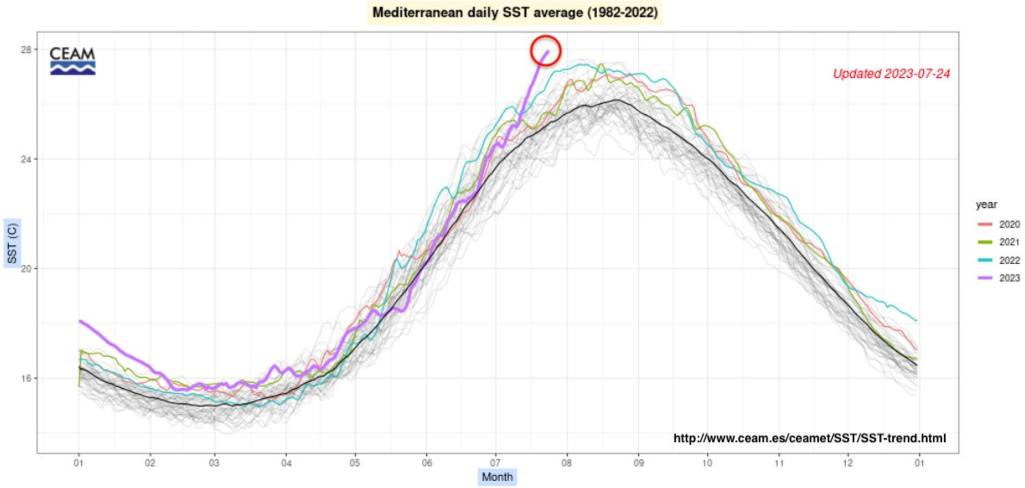
Much of the Northern Hemisphere is experiencing withering high temperatures, which scientists warn are increasingly likely.

Published July 18, 2023 Updated July 20, 2023



Yesterday ...


24 July 2023
Daily sea surface temperature anomalies
OISST V2.1 0.25°x0.25°
GRAPHIC BY : @SCOTTDUNCANWX
°C
−6 −5 −4 −3 −2 −1 0 1 2 3 4 5 6
DATA SOURCE: https://www.ncei.noaa.gov
1971-2000 base


Mediterranean daily SST average (1982-2022)
CEAM
Updated 2023-07-24
SST (C)
28
24
20
16
01 02 03 04 05 06 07 08 09 10 11 12 01
Month
year
2020
2021
2022
2023
http://www.ceam.es/ceamet/SST/SST-trend.html

This appeared today!

nature communications

Explore content ⌄ About the journal ⌄ Publish with us ⌄

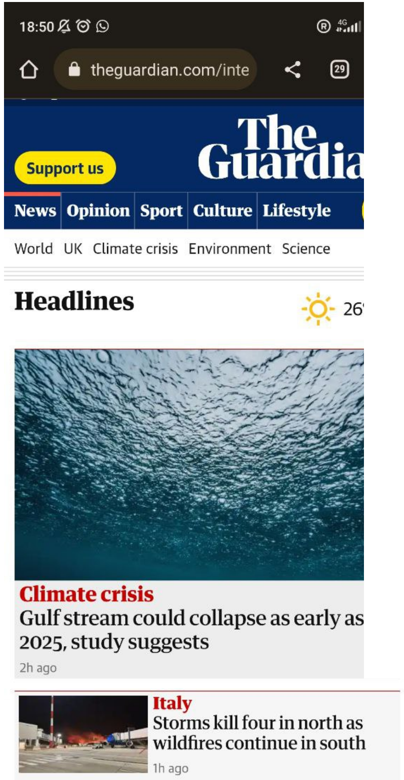
nature > nature communications > articles > article

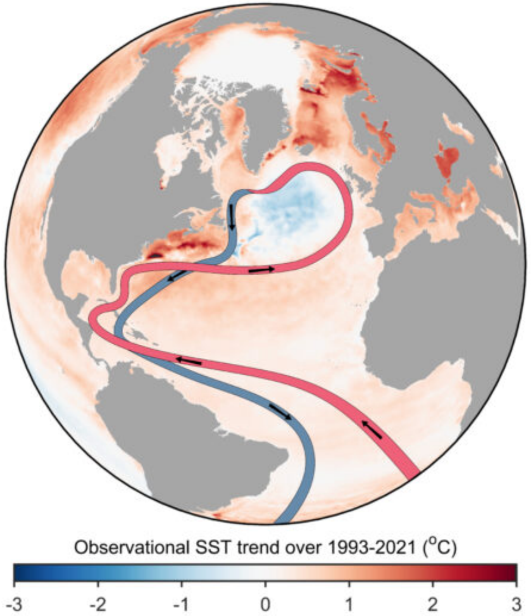
Article | Open Access | Published: 25 July 2023

Warning of a forthcoming collapse of the Atlantic meridional overturning circulation

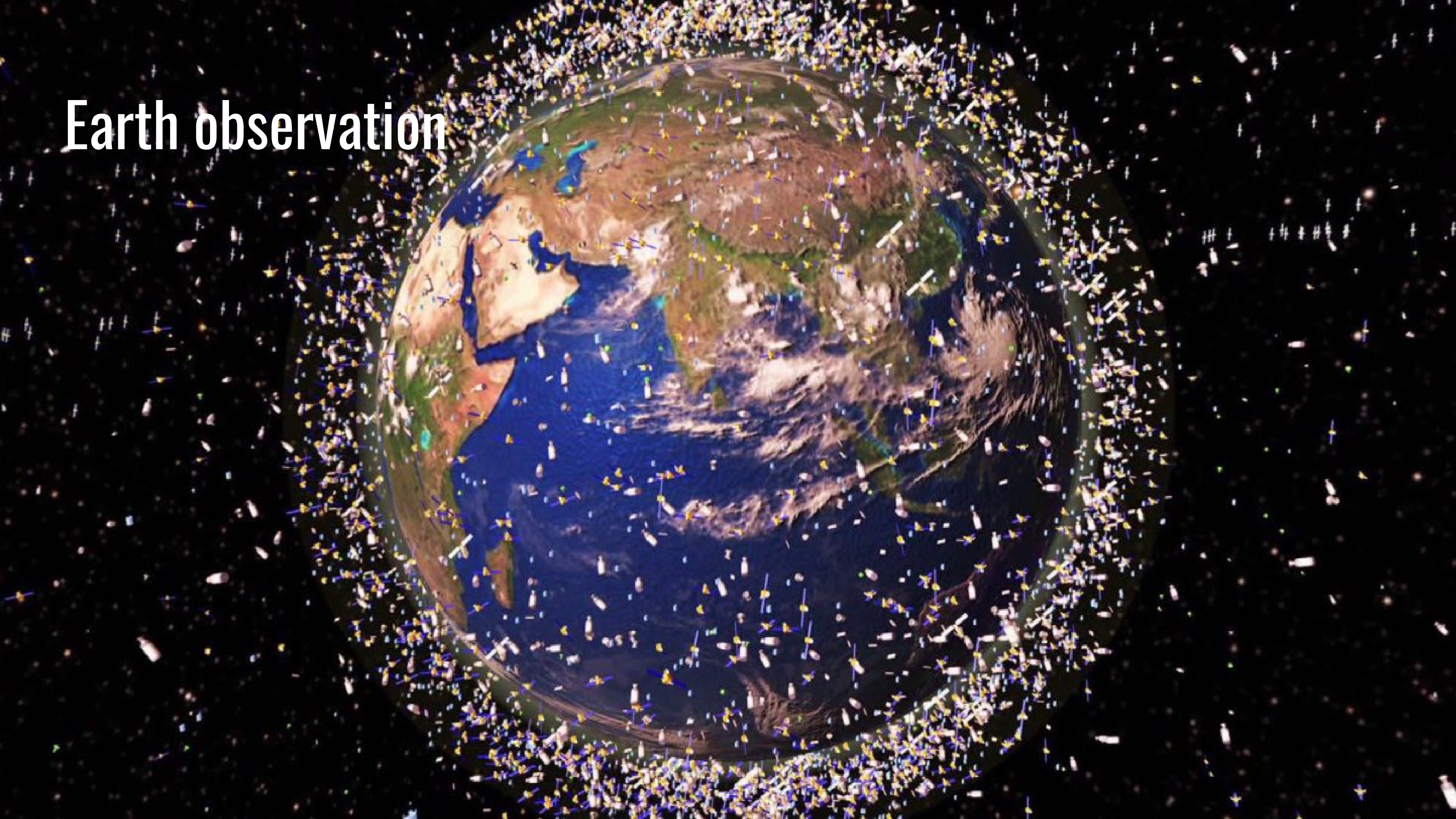
Peter Ditlevsen & Susanne Ditlevsen

Nature Communications **14**, Article number: 4254 (2023) | Cite this article





Earth observation

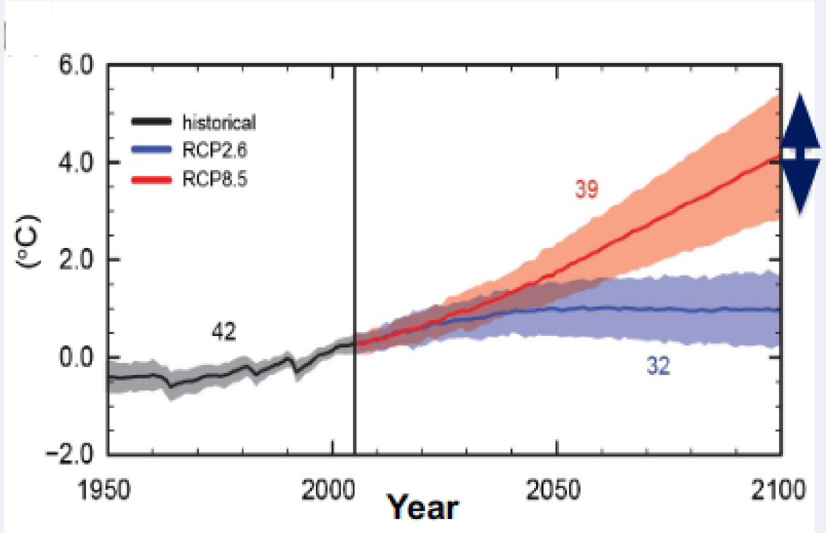


The planet as a hypercube …

«play video 1»

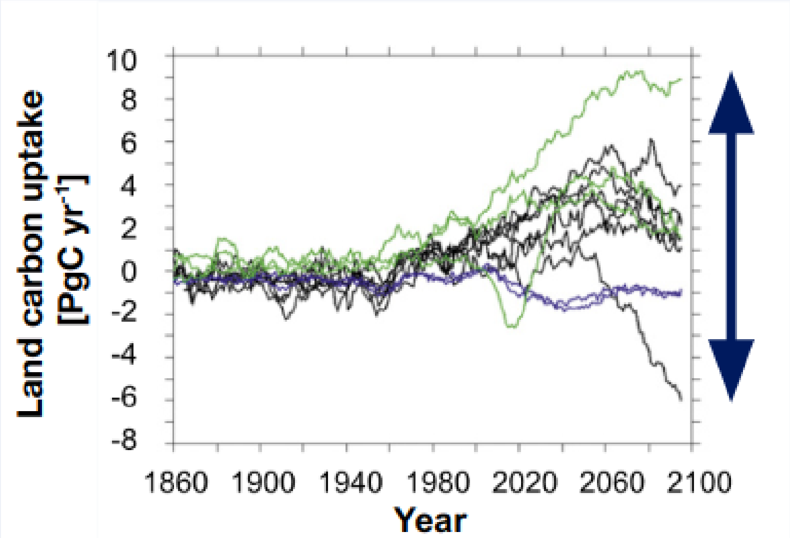
All models are wrong, many are not even useful

Global warming projections



Large Uncertainties
➢ **Clouds**

Land carbon uptake projections

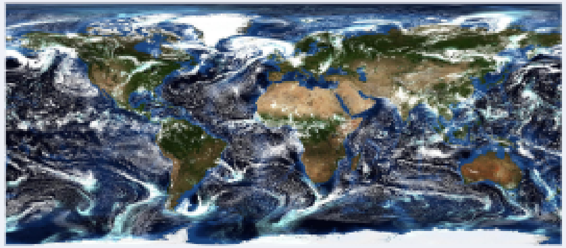


Large Uncertainties
➢ **Land-surface processes**

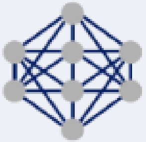
The perfect storm

1. Massive data from Earth observation

2. High-resolution cloud resolving models



3. Progress in machine learning



Agenda for today

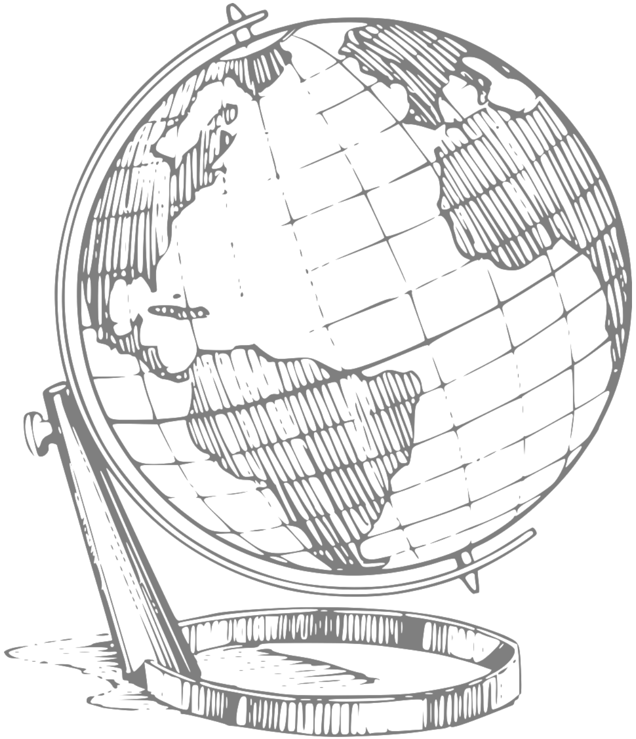
- **Part I: Introduction: why do we need ML?**
- **Part II: ML for Earth sciences**
- **Part III: The challenges**
- **Part IV: Physics-aware Machine Learning**
- **Part V: Explainable AI**
- **Part VI: Pragmatic causality**

Part I
Why ML for the Earth sciences?

AI helps in...

- **AI for <u>prediction</u> of essential climate variables**
- **AI for <u>detection</u> of extreme events**
- **AI for <u>characterization</u> of Earth status and health**
- **AI for <u>attribution of causes</u> of changes and anomalies**
- **AI to <u>optimize</u> resources**
- **AI to <u>understand human role</u> in the planet**

Did you say ‘modeling & understanding’?



Forecasting	Detecting anomalies	What?
Quantifying uncertainty	Learn to represent	When?
Causal relations	Discover laws	How & Why?

Part II
ML & DL for the Earth sciences

Why deep learning works, after all?

- DL are powerful machine learning models
- DL deals well with spatio-temporal-spectral multidimensional data
- DL can incorporate inductive priors by new losses & architectures
- DL is now a democratized, ready-to-use, commodity tool for users
- DL can be used for regression, classification, clustering, visualization

Deep learning for the Earth sciences works ...

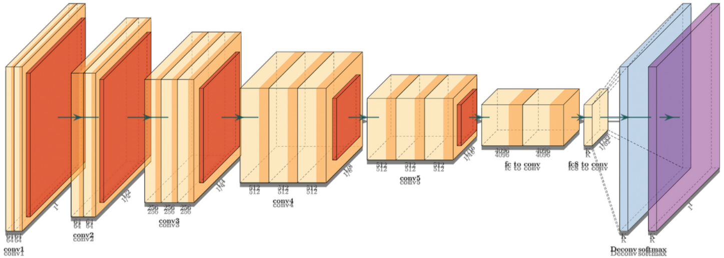
PERSPECTIVE

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]

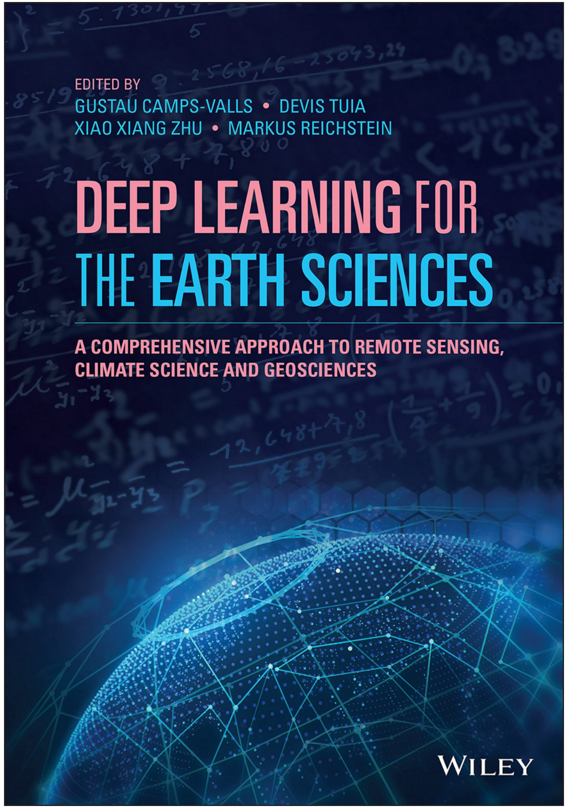
Machine learning approaches are increasingly used to extract patterns and insights from the ever-increasing stream of geospatial data, but current approaches may not be optimal when system behaviour is dominated by spatial or temporal context. Here, rather than amending classical machine learning, we 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.



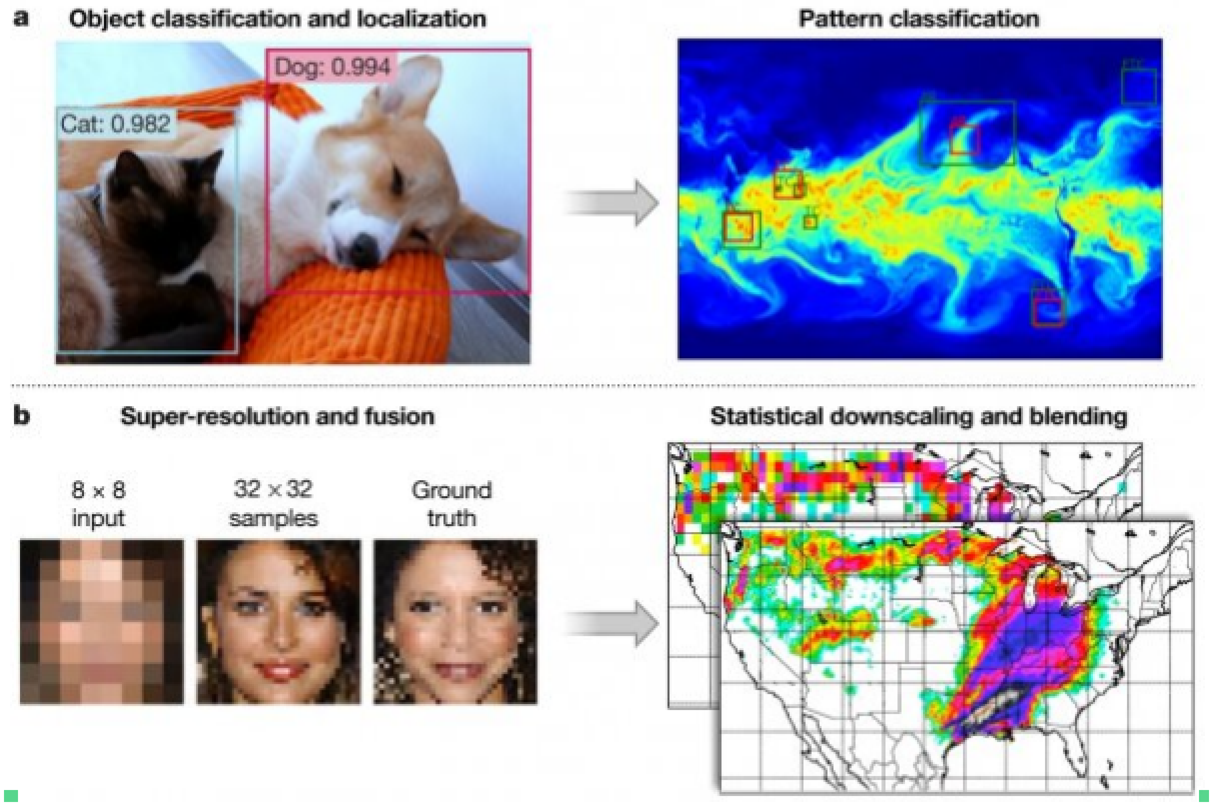
Reichstein, Camps-Valls et al, Nature, 2019

Camps-Valls, Tuia, Xiang, Reichstein. Wiley & Sons book, 2021

GitHub **https://github.com/DL4ES/DL4ES**

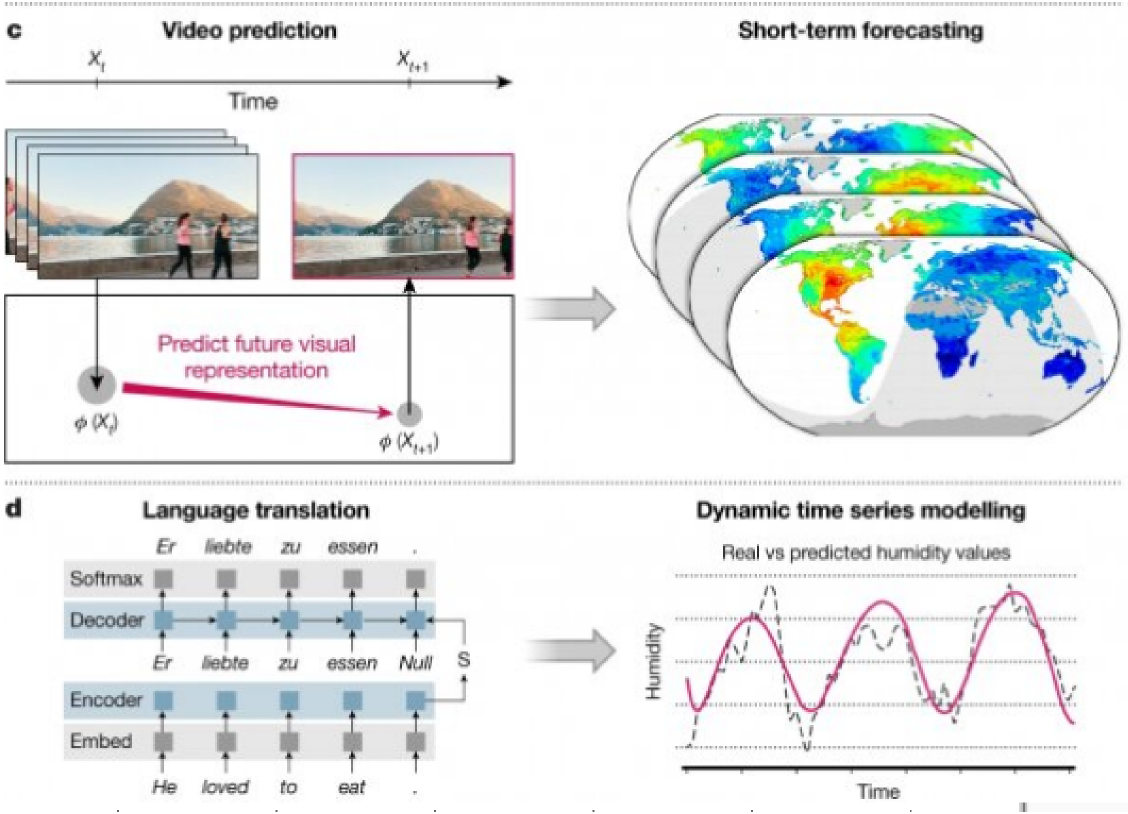


Detect, localize, superresolve ...



Reichstein, Camps-Valls et al, Nature, 2019
Camps-Valls, Tuia, Xiang, Reichstein.
Wiley & Sons book, 2021

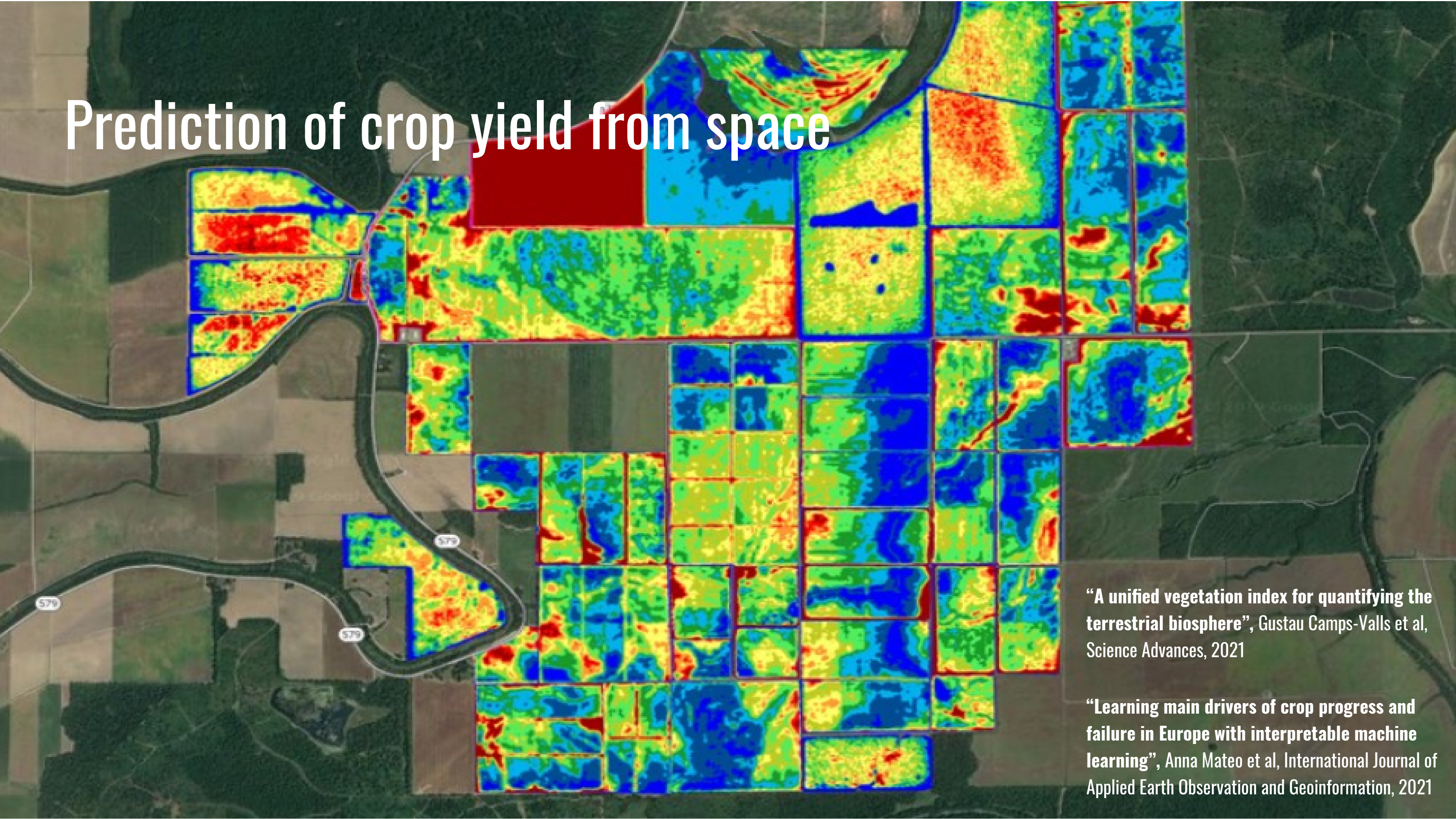
Predict, anticipate, forecast ...



Reichstein, Camps-Valls et al, Nature, 2019
Camps-Valls, Tuia, Xiang, Reichstein.
Wiley & Sons book, 2021

Part II.1
… and we deploy them in the wild!

Prediction of crop yield from space



“A unified vegetation index for quantifying the terrestrial biosphere”, Gustau Camps-Valls et al, Science Advances, 2021

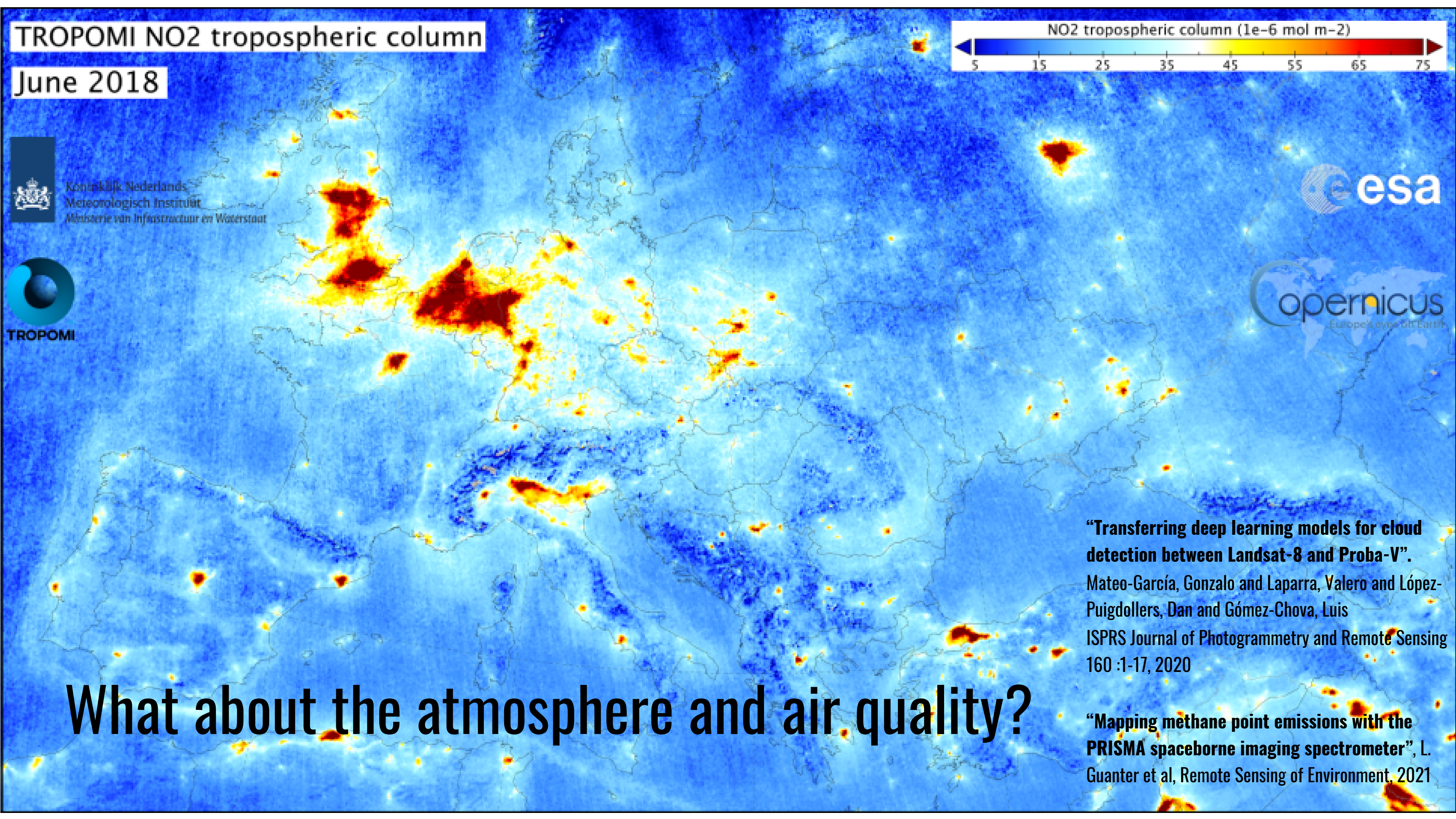
“Learning main drivers of crop progress and failure in Europe with interpretable machine learning”, Anna Mateo et al, International Journal of Applied Earth Observation and Geoinformation, 2021

Coastlines, water bodies and oceans?

“Learning Relevant Features of Optical Water Types” Blix, K. and Ruescas, A. and Johnson, E. and Camps-Valls, G. IEEE Geoscience and Remote Sensing Letters, 2022

“Estimation of Oceanic Particulate Organic Carbon with Machine Learning” Sauzède, R and Johnson, J Emmanuel and Claustre, H and Camps-Valls, G and Ruescas, AB. ISPRS Annals of the Photogrammetry, 2 :949--956, 2020

“Predicting regional coastal sea level changes with machine learning”, V Nieves, C. Radin & G. Camps-Valls, Scientific Reports, 2021



What about the atmosphere and air quality?

"Transferring deep learning models for cloud detection between Landsat-8 and Proba-V". Mateo-García, Gonzalo and Laparra, Valero and López-Puigdollers, Dan and Gómez-Chova, Luis ISPRS Journal of Photogrammetry and Remote Sensing 160 :1-17, 2020

"Mapping methane point emissions with the PRISMA spaceborne imaging spectrometer", L. Guanter et al, Remote Sensing of Environment, 2021

scientific reports

Article | Open Access | Published: 31 March 2021

Towards global flood mapping onboard low cost satellites with machine learning

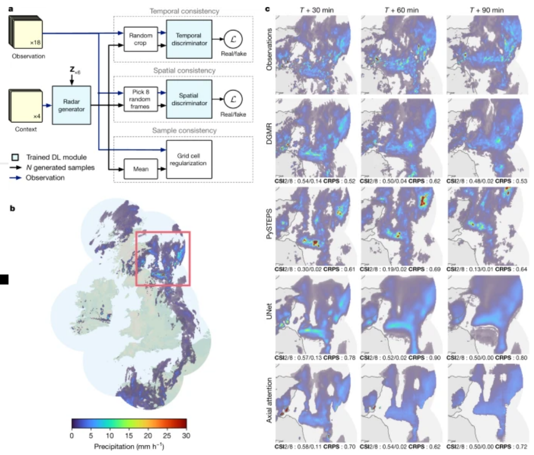
Gonzalo Mateo-Garcia, Joshua Veitch-Michaelis, Lewis Smith, Silviu Vlad Oprea, Guy Schumann, Yarin Gal, Atılım Güneş Baydin & Dietmar Backes

Scientific Reports **11**, Article number: 7249 (2021) | Cite this article nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › articles › article

Article | Open Access | Published: 29 September 2021



Skilful precipitation nowcasting using deep generative models of radar

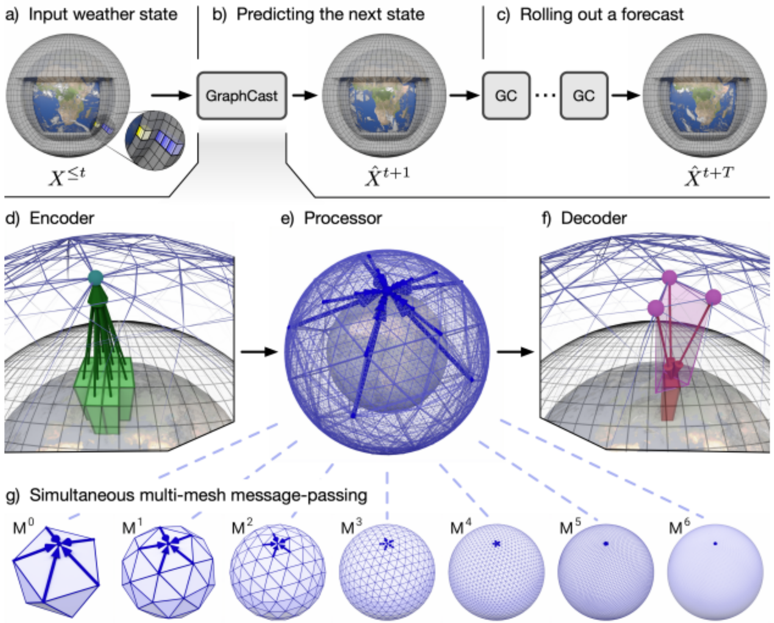
Suman Ravuri, Karel Lenc, Matthew Willson, Dmitry Kangin, Remi Lam, Piotr Mirowski, Megan Fitzsimons, Maria Athanassiadou, Sheleem Kashem, Sam Madge, Rachel Prudden, Amol Mandhane, Aidan Clark, Andrew Brock, Karen Simonyan, Raia Hadsell, Niall Robinson, Ellen Clancy, Alberto Arribas & Shakir Mohamed ✉

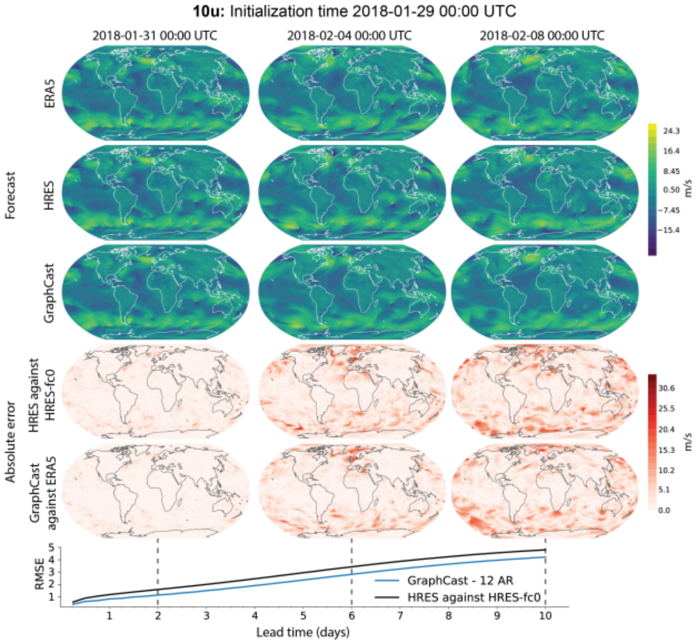
Nature **597**, 672–677 (2021) | Cite this article

GraphCast: Learning skillful medium-range global weather forecasting

Remi Lam[*,1], Alvaro Sanchez-Gonzalez[*,1], Matthew Willson[*,1], Peter Wirnsberger[*,1], Meire Fortunato[*,1], Alexander Pritzel[*,1], Suman Ravuri[1], Timo Ewalds[1], Ferran Alet[1], Zach Eaton-Rosen[1], Weihua Hu[1], Alexander Merose[2], Stephan Hoyer[2], George Holland[1], Jacklynn Stott[1], Oriol Vinyals[1], Shakir Mohamed[1] and Peter Battaglia[1]

[*]equal contribution, [1]DeepMind, [2]Google





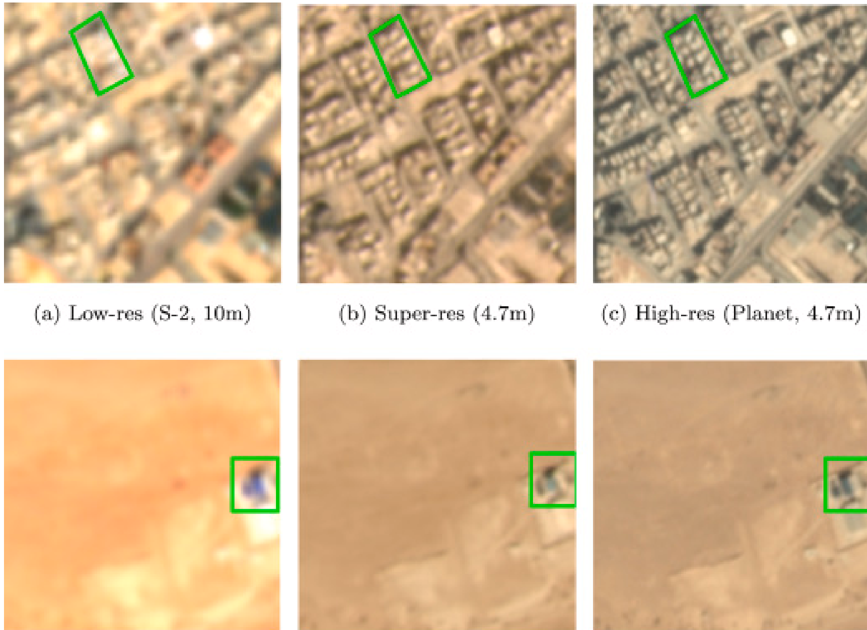
ELSEVIER

ISPRS Journal of Photogrammetry and Remote Sensing

Volume 195, January 2023, Pages 1-13

Multi-spectral multi-image super-resolution of Sentinel-2 with radiometric consistency losses and its effect on building delineation

Muhammed T. Razzak [a], Gonzalo Mateo-García [b], Gurvan Lecuyer [c], Luis Gómez-Chova [b], Yarin Gal [a], Freddie Kalaitzis [a]



(a) Low-res (S-2, 10m) (b) Super-res (4.7m) (c) High-res (Planet, 4.7m)

DL for climate change mitigation

Contents lists available at ScienceDirect

Landscape and Urban Planning

journal homepage: www.elsevier.com/locate/landurbplan

ELSEVIER

Landscape and Urban Planning

Research Paper

Roofpedia: Automatic mapping of green and solar roofs for an open roofscape registry and evaluation of urban sustainability

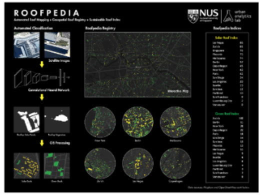
Abraham Noah Wu [a,1], Filip Biljecki [a,b,*,2]

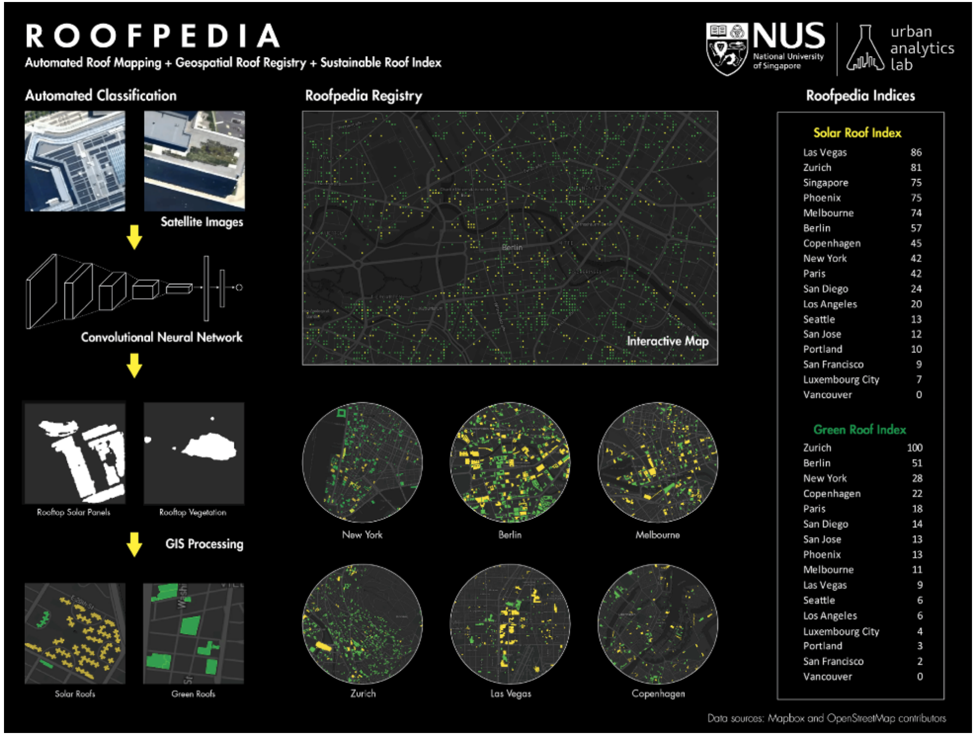
[a] Department of Architecture, National University of Singapore, Singapore
[b] Department of Real Estate, National University of Singapore, Singapore

HIGHLIGHTS

- There is a lack of open data on urban rooftop typology and current use of roofs.
- A deep learning and GIS workflow to map and quantify green and solar roofs.
- A generated dataset that covers 17 cities, scalable to include more locations.
- An index to benchmark the proliferation of green and solar roofs in cities.

GRAPHICAL ABSTRACT





"Roofpedia: Automatic mapping of green and solar roofs for an open roofscape registry and evaluation of urban sustainability."

Wu, Abraham Noah, and Filip Biljecki. Landscape and Urban Planning 214 (2021): 104167.

DL for wealth, energy & activity analysis

Poverty Prediction with Public Landsat 7 Satellite Imagery and Machine Learning

Anthony Perez
Department of Computer Science
Stanford University
Stanford, CA 94305
aperez8@stanford.edu

Christopher Yeh
Department of Computer Science
Stanford University
Stanford, CA 94305
chrisyeh@stanford.edu

George Azzari
Department of Earth System Science
Stanford University
Stanford, CA - 94305
gazzari@stanford.edu

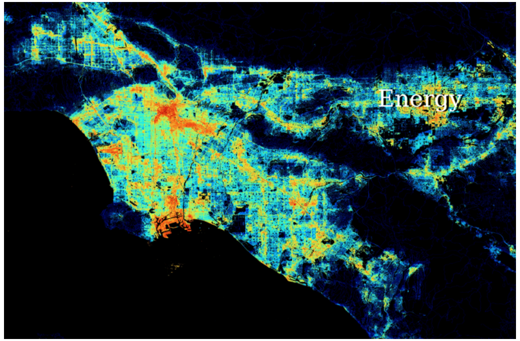
Marshall Burke
Department of Earth System Science
Stanford University
Stanford, CA - 94305
mburke@stanford.edu

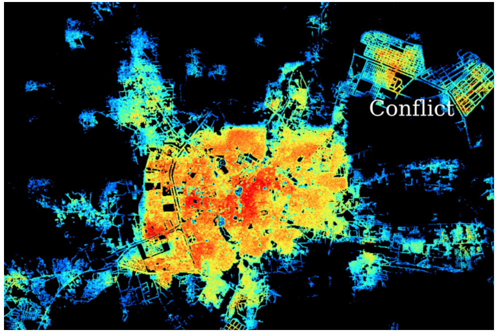
David Lobell
Department of Earth System Science
Stanford University
Stanford, CA - 94305
dlobell@stanford.edu

Stefano Ermon
Department of Computer Science
Stanford University
Stanford, CA - 94305
ermon@cs.stanford.edu

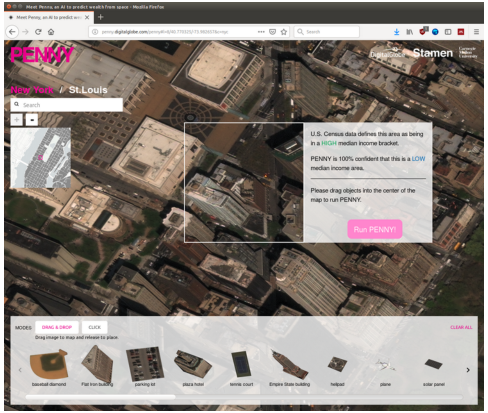
Model	Mean Train r^2	Mean Test r^2	Aggregate Residual r^2
Nightlights / GBT	0.63	0.66	1.0
VGG-F, RGB / ridge	0.71	0.64	0.69
VGG-F, 9 Band / ridge	0.68	0.63	0.70
ResNet-18, 9 Band / ridge	0.69	0.64	0.73
ResNet-34, 9 Band / ridge	0.70	0.65	0.74
Jean et al. [8]	0.53	0.54	0.76

"Poverty prediction with public Landsat 7 satellite imagery and machine learning." Perez, Anthony, et al. arXiv:1711.03654 (2017).



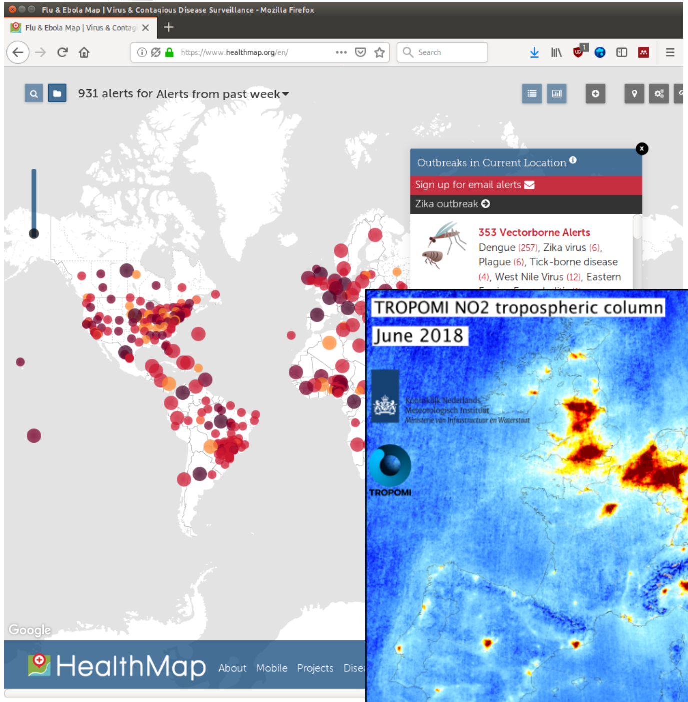


NASA's black marble – https://blackmarble.gsfc.nasa.gov/

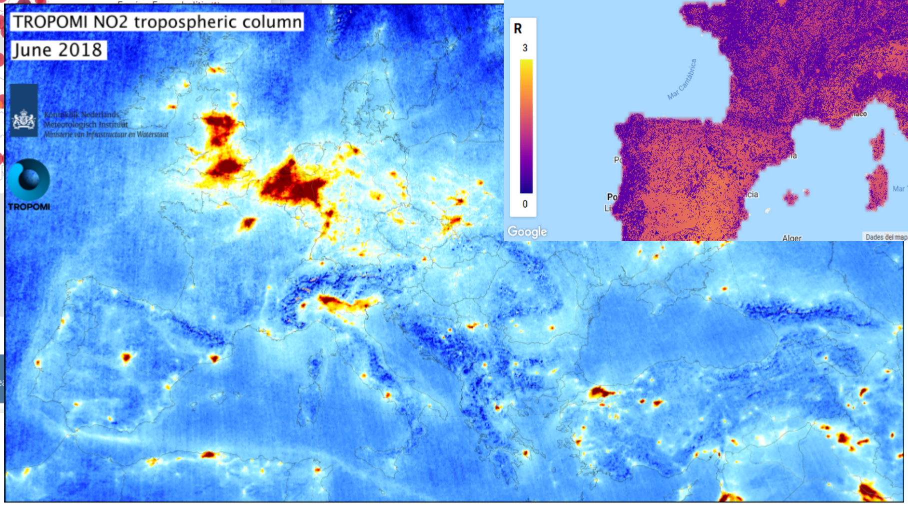


Global wealth map
http://penny.digitalglobe.com

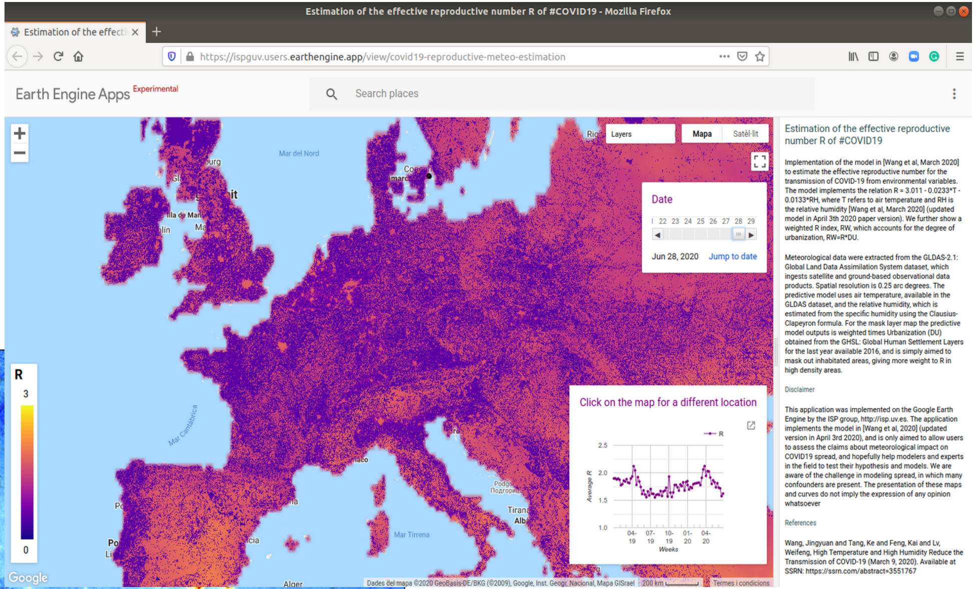
DL for health analysis



https://healthmap.org/en/



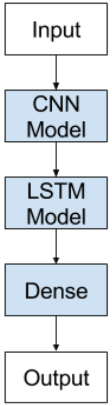
https://ispguv.users.earthengine.app/view/covid19-reproductive-meteo-estimation

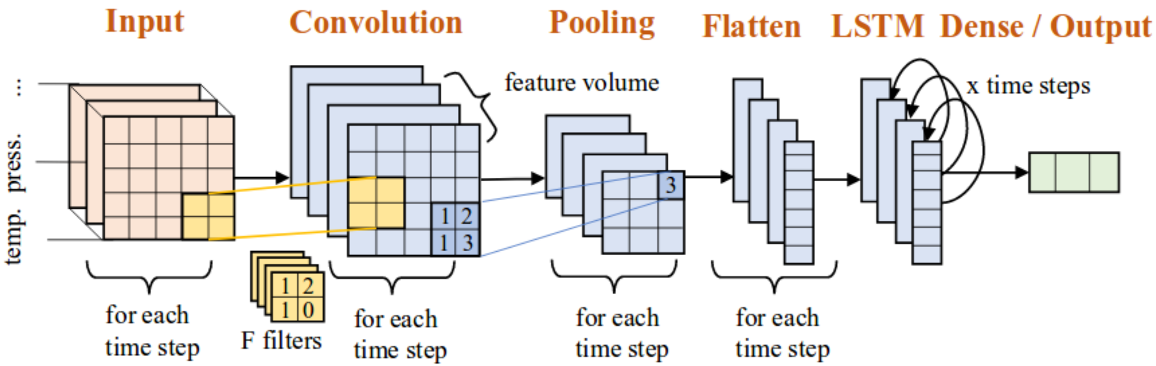


Part II.2
some deep learning
for spatio-temporal data analysis

Spatio-temporal data: convolutional and recurrent networks

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





Reichstein, Camps-Valls et al, Nature, 2019
Camps-Valls, Tuia, Xiang, Reichstein Wiley & Sons book, 2021

Multiscale & multisource wavelet neural networks

- **Wavelet decomposition as a ‘saliency detector’ of interesting regions**
- **Divide-and-conquer strategies for detection**

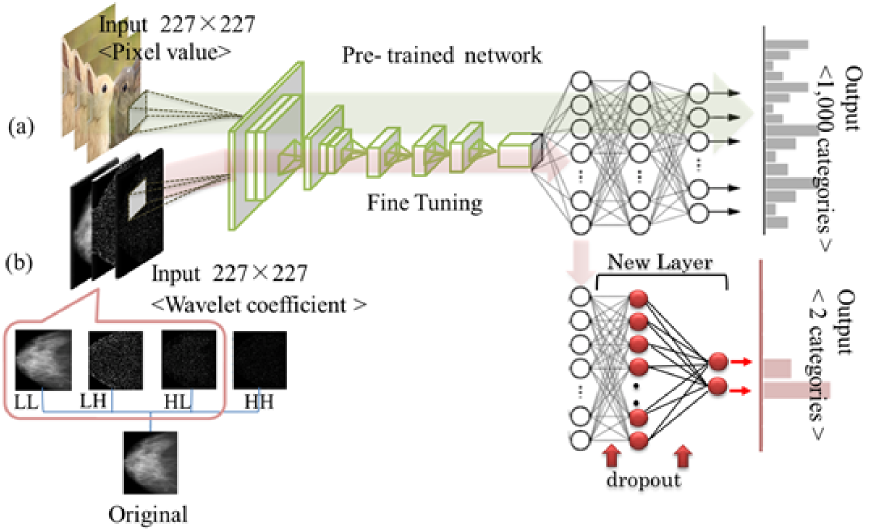
Chaos | ARTICLE | scitation.org/journal/cha

Forecasting of extreme flood events using different satellite precipitation products and wavelet-based machine learning methods

Cite as: Chaos 30, 063115 (2020); doi: 10.1063/5.0008195
Submitted: 19 March 2020 · Accepted: 14 May 2020 ·
Published Online: 2 June 2020

View Online | Export Citation | CrossMark

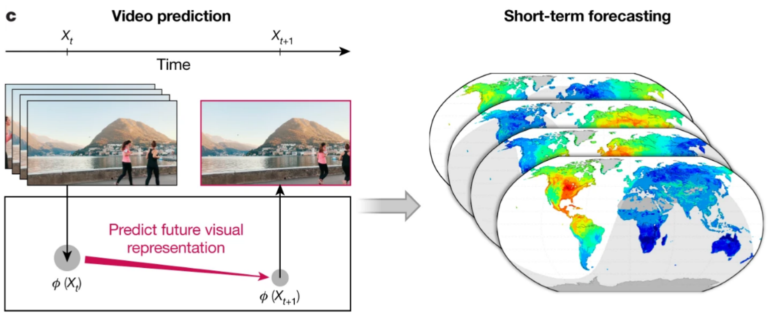
Pavan Kumar Yeditha,[1] Venkatesh Kasi,[1] Maheswaran Rathinasamy,[1] and Ankit Agarwal[2,a)]

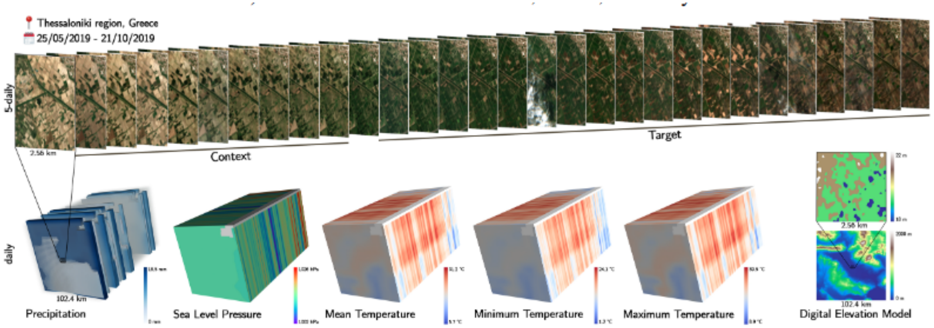


Reichstein, Camps-Valls et al, Nature, 2019
Camps-Valls, Tuia, Xiang, Reichstein Wiley & Sons book, 2021

Forecasting and tracking as a video prediction task

- **Many video prediction techniques from computer vision are widely applicable in extreme event tracking and forecasting**



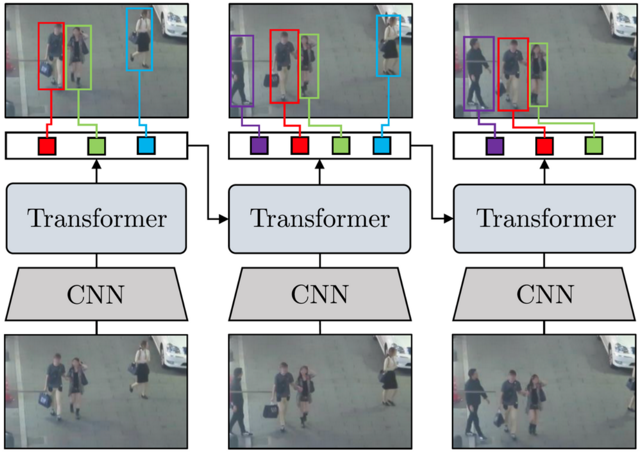


EarthNet2021 dataset and challenge

- **Methods:** U-Net, ARCON, OLS, …

Transformers and attention mechanisms

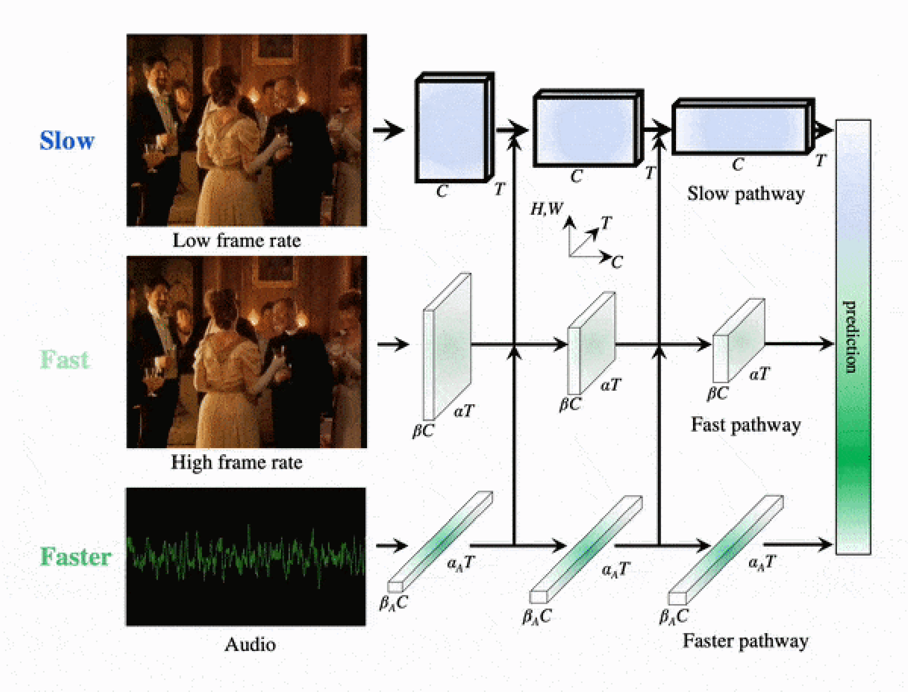
- **Transformers capture multiscale and multisource data relations**
- **Attention mechanisms rooted on visual neuroscience and natural language processing**



TrackFormer: Multi-Object Tracking with Transformers, 2021

Fast and slow multisource neural networks

- Networks to process speed-varying processes
- Different focused branches to fuse multisource information
- Opportunities to model legacy effects & persistence



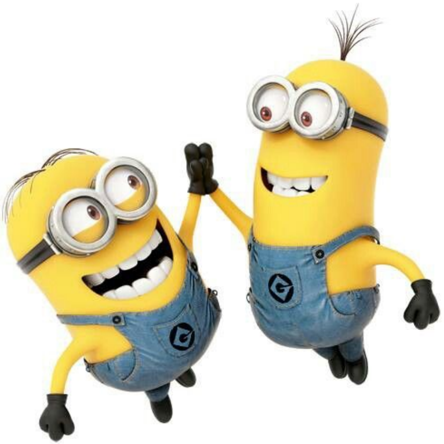
Audiovisual SlowFast Networks for Video Recognition, 2020

Part II.3
advances in kernel methods

- kernel indices

- feature extractors

1- Kernel (vegetation) indices



SCIENCE ADVANCES | RESEARCH ARTICLE

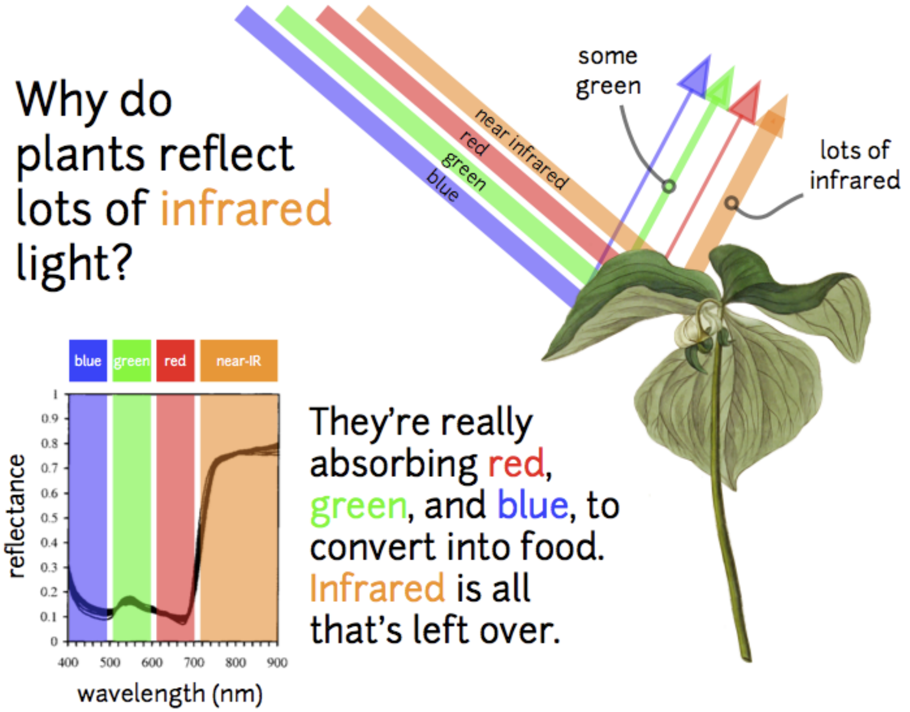
ENVIRONMENTAL STUDIES

A unified vegetation index for quantifying the terrestrial biosphere

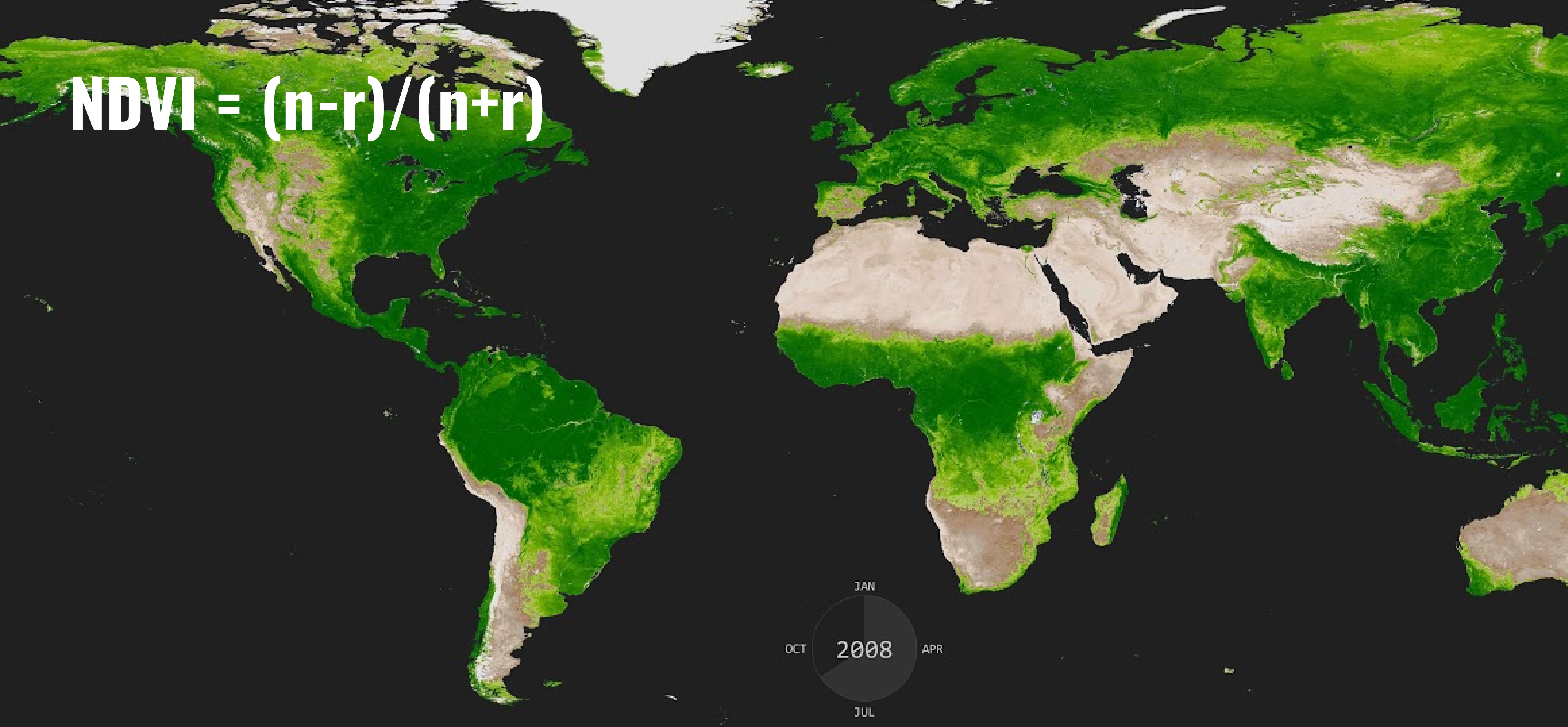
Gustau Camps-Valls[1]*, Manuel Campos-Taberner[2], Álvaro Moreno-Martínez[1,3], Sophia Walther[4], Grégory Duveiller[5], Alessandro Cescatti[5], Miguel D. Mahecha[6,7,8], Jordi Muñoz-Marí[1], Francisco Javier García-Haro[2], Luis Guanter[9], Martin Jung[4], John A. Gamon[10,11], Markus Reichstein[4], Steven W. Running[3]

Empirical vegetation indices derived from spectral reflectance data are widely used in remote sensing of the biosphere, as they represent robust proxies for canopy structure, leaf pigment content, and, subsequently, plant photosynthetic potential. Here, we generalize the broad family of commonly used vegetation indices by exploiting all higher-order relations between the spectral channels involved. This results in a higher sensitivity to vegetation biophysical and physiological parameters. The presented nonlinear generalization of the celebrated normalized difference vegetation index (NDVI) consistently improves accuracy in monitoring key parameters, such as leaf area index, gross primary productivity, and sun-induced chlorophyll fluorescence. Results suggest that the statistical approach maximally exploits the spectral information and addresses long-standing problems in satellite Earth Observation of the terrestrial biosphere. The nonlinear NDVI will allow more accurate measures of terrestrial carbon source/sink dynamics and potentials for stabilizing atmospheric CO_2 and mitigating global climate change.

Vegetation indices



Method	Formulation	ρ
GI	R_{672}/R_{550}	0.52 (0.09)
GVI	$(R_{682}-R_{553})/(R_{682}+R_{553})$	0.66 (0.07)
Macc	$(R_{780}-R_{710})/(R_{780}+R_{680})$	0.20 (0.29)
MCARI	$[(R_{700}-R_{670})-0.2(R_{700}-R_{550})]/(R_{700}/R_{670})$	0.35 (0.14)
MCARI2	$1.2[2.5(R_{800}-R_{670})-1.3(R_{800}-R_{550})]$	0.71 (0.12)
mNDVI	$(R_{800}-R_{680})/(R_{800}+R_{680}-2R_{445})$	0.77 (0.12)
mNDVI_{705}	$(R_{750}-R_{705})/(R_{750}+R_{705}-2R_{445})$	0.80 (0.07)
mSR_{705}	$(R_{750}-R_{445})/(R_{705}+R_{445})$	0.72 (0.07)
MTCI	$(R_{754}-R_{709})/(R_{709}+R_{681})$	0.19 (0.26)
mTVI	$1.2[1.2(R_{800}-R_{550})-2.5(R_{670}-R_{550})])$	0.73 (0.07)
NDVI	$(R_{800}-R_{670})/(R_{800}+R_{670})$	0.77 (0.08)
NDVI2	$(R_{750}-R_{705})/(R_{750}+R_{705})$	0.81 (0.06)
NPCI	$(R_{680}-R_{430})/(R_{680}+R_{430})$	0.72 (0.08)
NPQI	$(R_{415}-R_{435})/(R_{415}+R_{435})$	0.61 (0.15)
OSAVI	$1.16(R_{800}-R_{670})/(R_{800}+R_{670}+0.16)$	0.79 (0.09)
PRI	$(R_{531}-R_{570})/(R_{531}+R_{570})$	0.77 (0.07)
PRI2	$(R_{570}-R_{539})/(R_{570}+R_{539})$	0.76 (0.07)
PSRI	$(R_{680}-R_{500})/R_{750}$	0.79 (0.08)
RDVI	$(R_{800}-R_{670})/\sqrt{(R_{800}+R_{670})}$	0.76 (0.08)
SIPI	$(R_{800}-R_{445})/(R_{800}-R_{680})$	0.78 (0.08)
SPVI	$0.4[3.7(R_{800}-R_{670})-1.2(R_{530}-R_{670})]$	0.70 (0.08)
SR	R_{800}/R_{680}	0.63 (0.12)
SR1	R_{750}/R_{700}	0.74 (0.07)
SR2	R_{752}/R_{690}	0.68 (0.09)
SR3	R_{750}/R_{550}	0.75 (0.07)
SR4	R_{672}/R_{550}	0.76 (0.10)
SRPI	R_{430}/R_{680}	0.76 (0.09)
TCARI	$3[R_{700}-R_{670})-0.2(R_{700}-R_{550})(R_{700}/R_{670})]$	0.53 (0.13)
TVI	$0.5[120R_{750}-R_{550})-200(R_{670}-R_{550})]$	0.70 (0.10)
VOG	$R_{740}/(R_{720}$	0.76 (0.06)
VOG2	$(R_{734}-R_{747})/(R_{715}+R_{726})$	0.72 (0.09)
NAOC	Area in $[643, 795]$	0.79 (0.09)

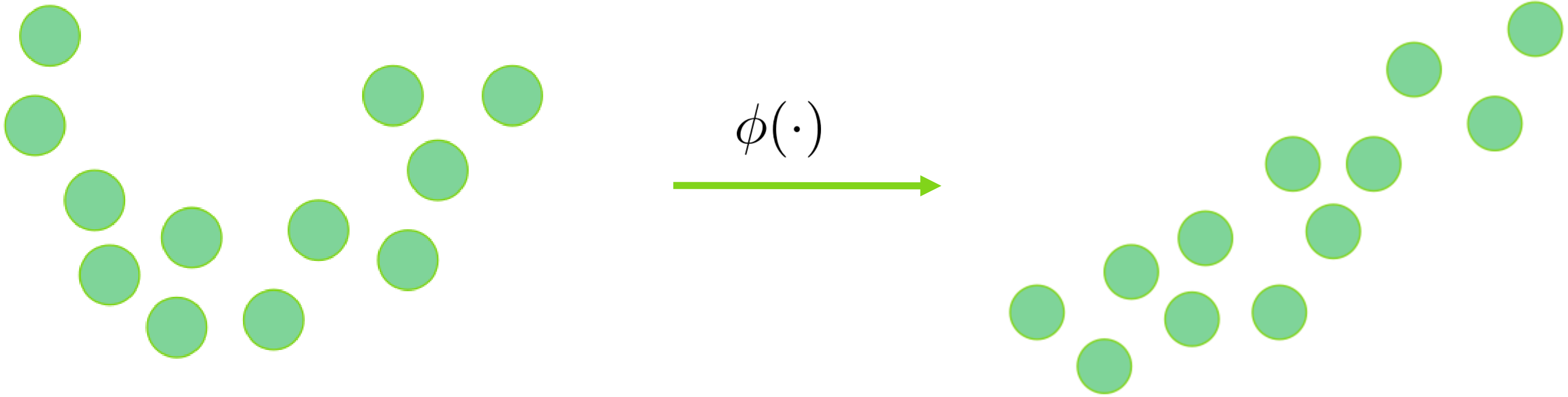


Normalized Difference Vegetation Index (NDVI)
Collected by MODIS on NASA's Terra Satellite

A simple observation ...

- All indices try to linearize the problem with arbitrary, yet sensible, transformations...
- Why not accounting for all possible transformations jointly?

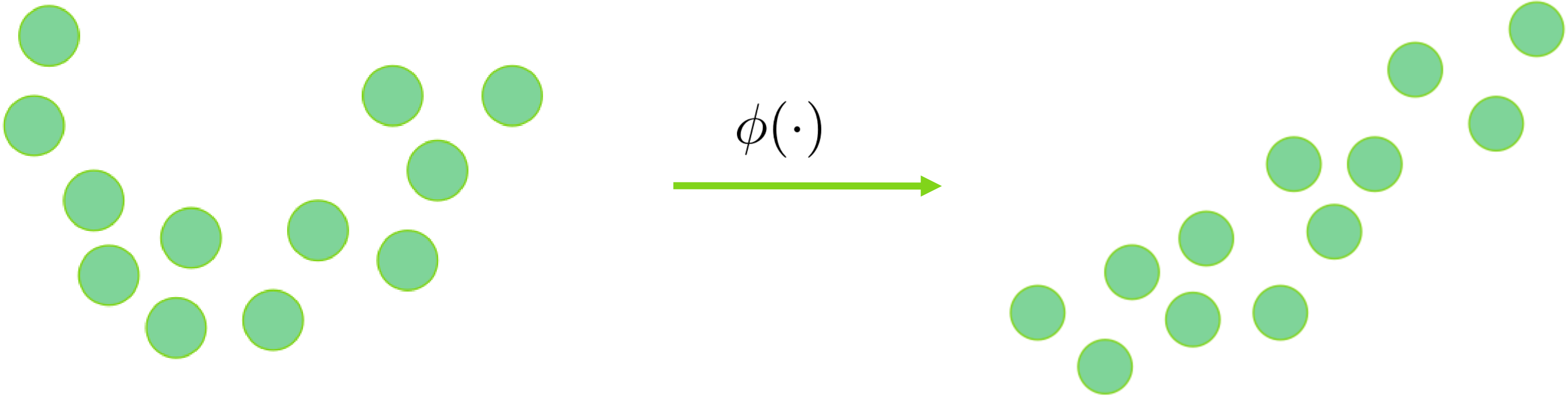
Kernel methods to the rescue …



$$k(a, b) = a^\top b$$

$$k(a, b) = \boldsymbol{\phi}(a)^\top \boldsymbol{\phi}(b)$$

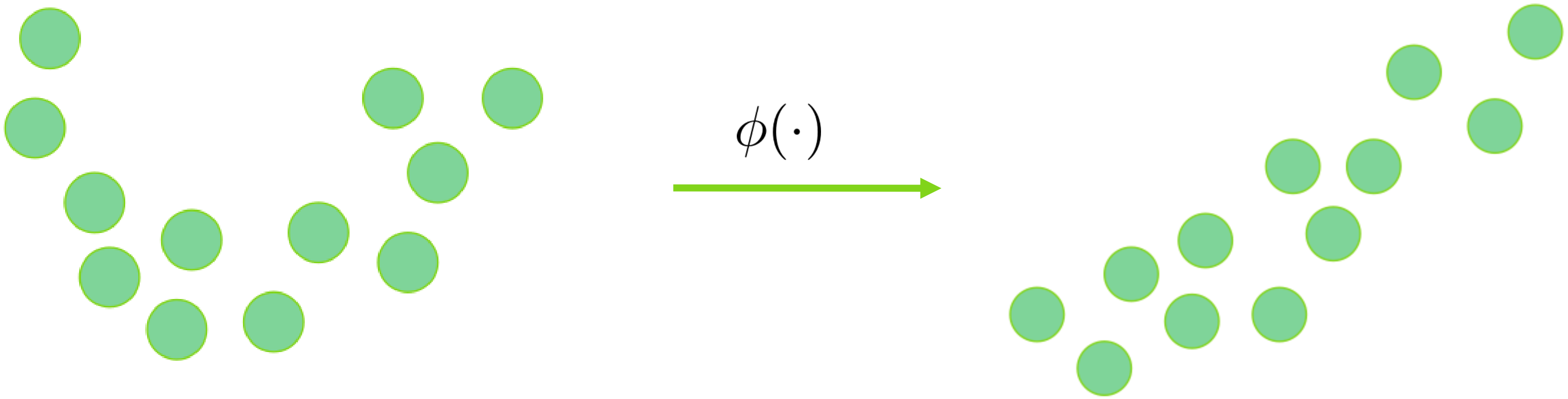
Kernel methods to the rescue …



$k(a, b) = a^\top b$

$k(a, b) = \exp(-\|a - b\|^2/(2\sigma^2))$

Kernel methods to the rescue ...



$k(a, b) = a^\top b$

$\boldsymbol{\phi}(a) \approx \{a, a^2, a^3, \ldots\}$

Kernel NDVI

$$\mathrm{NDVI} = \frac{n - r}{n + r}$$

Kernel NDVI

$$\mathrm{kNDVI} = \frac{\|\phi(n) - \phi(r)\|^2}{\|\phi(n) + \phi(r)\|^2}$$

Kernel NDVI

$$\mathrm{kNDVI} = \frac{k(n,n) - k(n,r)}{k(n,n) + k(n,r)}$$

$$k(n,r) = \exp\left(-\frac{\|n-r\|^2}{2\sigma^2}\right)$$

Kernel NDVI

$$\text{kNDVI} = \tanh\left(\left(\frac{n-r}{2\sigma^2}\right)^2\right)$$

$$k(n,r) = \exp\left(-\frac{\|n-r\|^2}{2\sigma^2}\right) \quad \sigma = \frac{1}{2}(n+r)$$

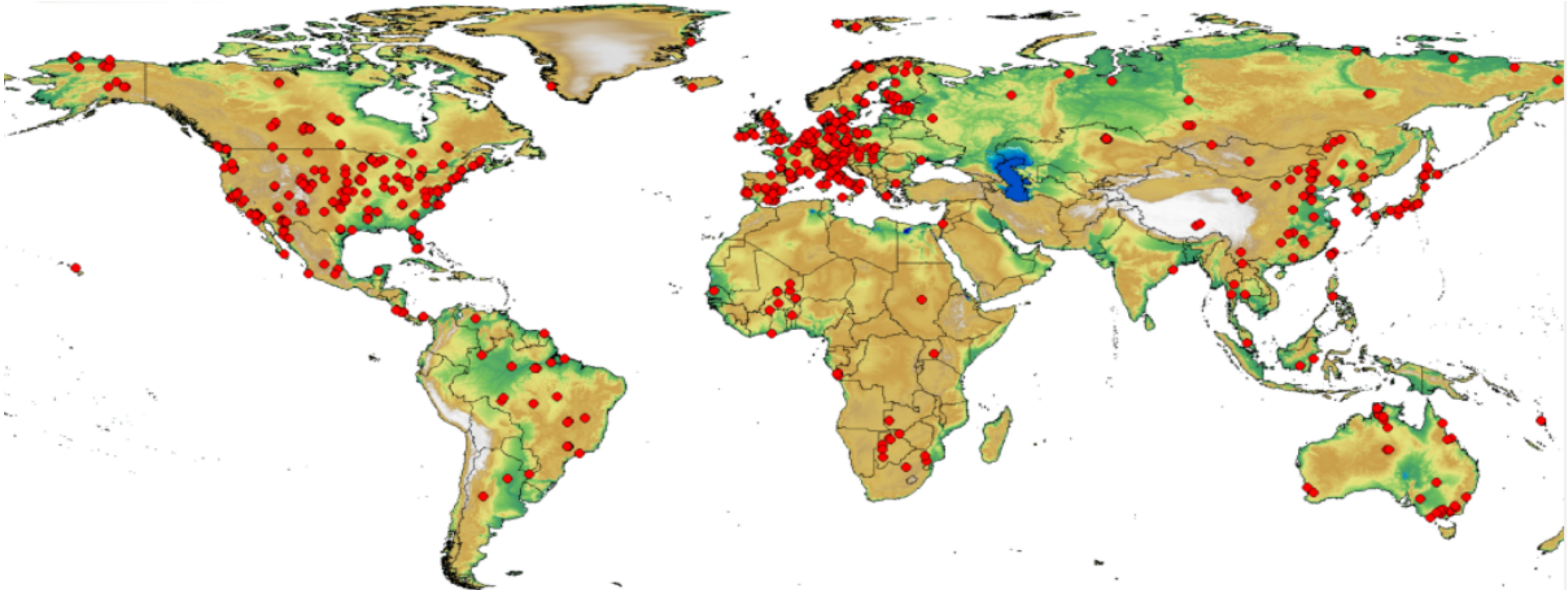
Kernel NDVI

$$\mathrm{kNDVI} = \tanh(\mathrm{NDVI}^2)$$

$$k(n, r) = \exp\left(-\frac{\|n - r\|^2}{2\sigma^2}\right) \quad \sigma = \frac{1}{2}(n + r)$$

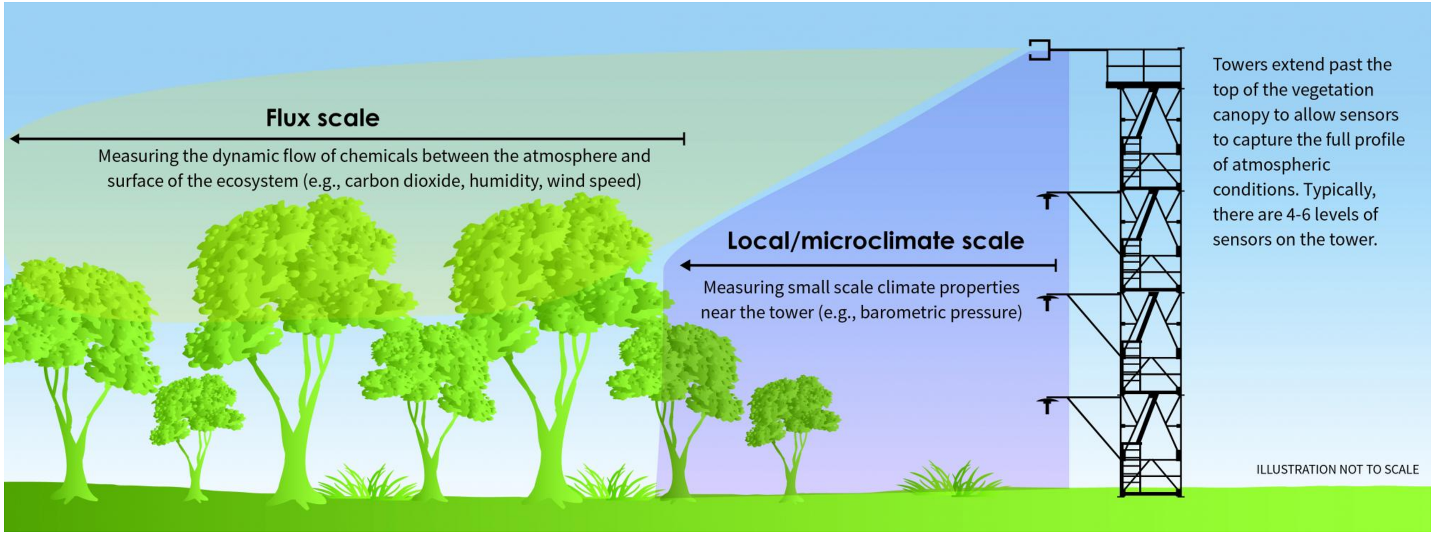
Accurate proxy to gross primary production

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



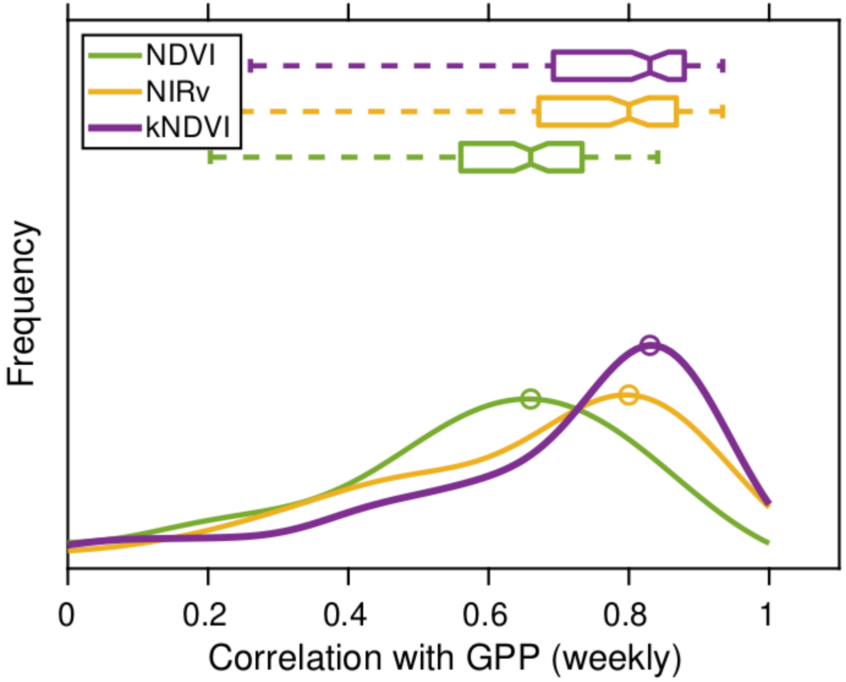
Accurate proxy to gross primary production

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



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

Accurate proxy to gross primary production

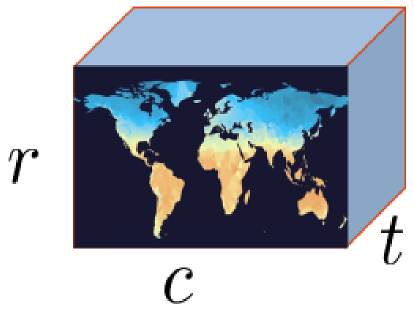


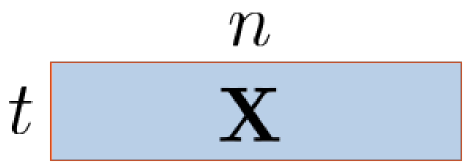
		GPP		
	Biome	**NDVI**	**NIRv**	**kNDVI**
C1	**NF**	0.59	**0.65**	0.62
C2	**EBF**	0.37	**0.45**	**0.45**
C3	**DBF**	0.61	0.79	**0.82**
C4	**MF**	0.69	0.84	**0.86**
C5	**SH**	0.57	0.68	**0.72**
C6	**SAV**	0.63	**0.74**	**0.74**
C7	**GRA**	0.61	0.71	**0.72**
C8	**CRO**	0.51	**0.58**	**0.58**
	ALL	0.59	**0.68**	**0.68**

2- ROCK-PCA: Rotated complex kernel PCA

PCA

- **Find projections of maximum variance**

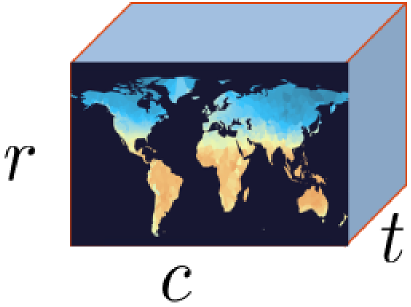


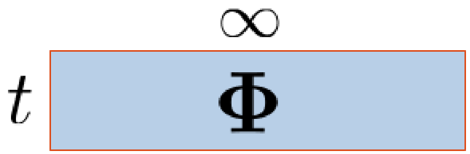


- Covariance $\mathbf{C} = \mathbf{X}^\top \mathbf{X}$
- Gram $\mathbf{G} = \mathbf{X}\mathbf{X}^\top$
- Diagonalization $\mathbf{G}\mathbf{V} = \mathbf{\Lambda}\mathbf{V}$

KPCA

- **Find projections of maximum variance in a higher dimensional Hilbert space**





- Covariance $\mathbf{C}_{\mathcal{H}} = \mathbf{\Phi}^\top \mathbf{\Phi}$
- Gram $\mathbf{K} = \mathbf{\Phi}\mathbf{\Phi}^\top$
- Diagonalization $\mathbf{K}\boldsymbol{A} = \mathbf{\Lambda}\mathbf{A}$

ROCK-PCA

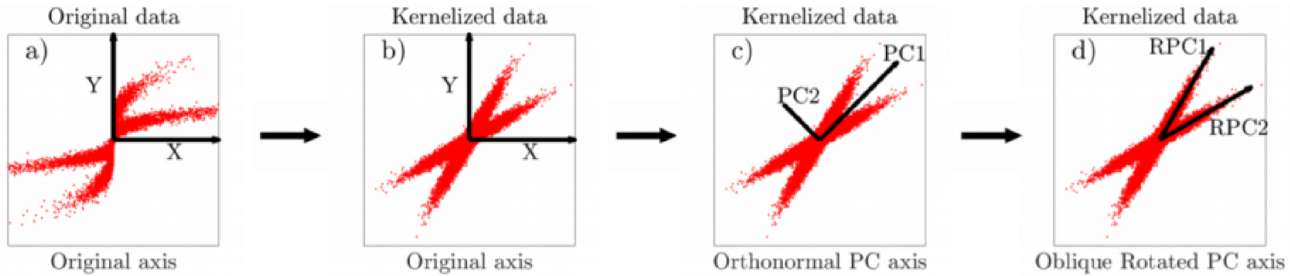
- **Complex:** Map data to the complex domain with the Hilbert transform

$$\mathbf{x}_{\mathsf{H}}(t) = \mathbf{x}(t) + j\mathsf{H}(\mathbf{x}(t)) \qquad x_h(t) := \mathsf{H}(x(t)) = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{x(\tau)}{t-\tau}\, d\tau$$

- **Nonlinear:** Map the complex data into a higher dimensional kernel Hilbert space

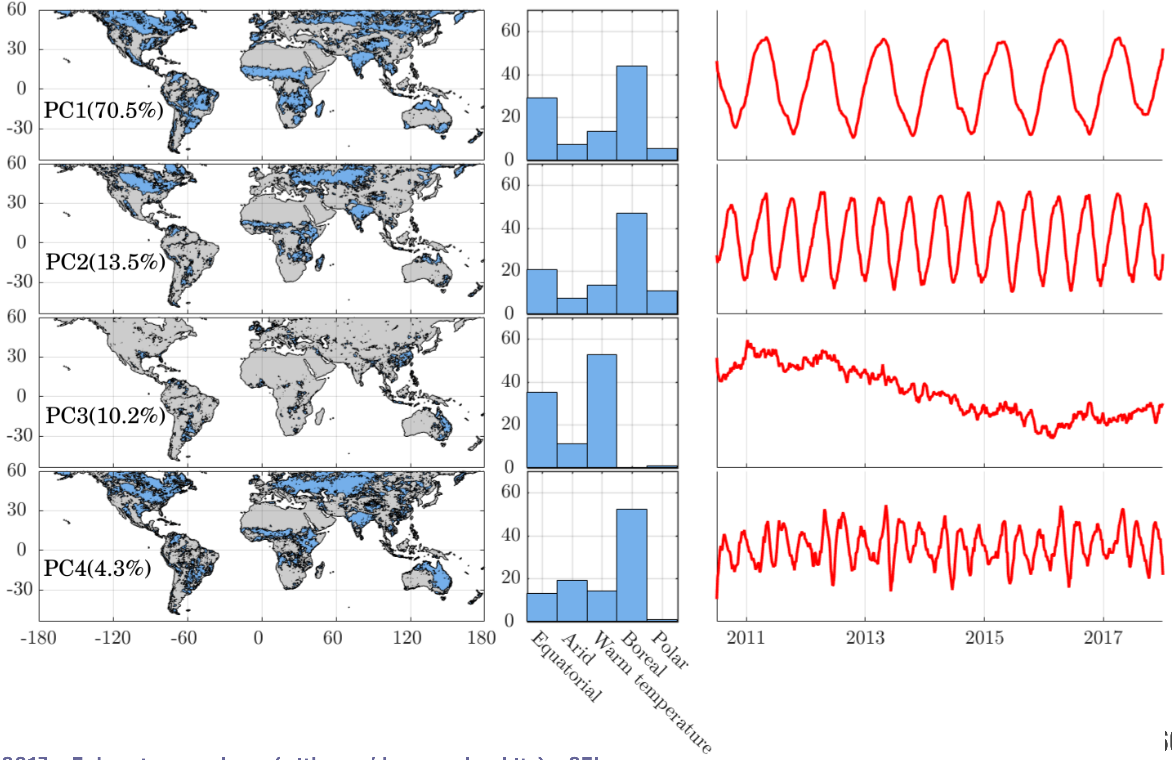
$$\mathbf{G}_{\mathsf{H}} = \mathbf{X}_{\mathsf{H}}\mathbf{X}_{\mathsf{H}}^{H} = \mathbf{G} + j\tilde{\mathbf{G}}_h \in C^{t\times t} \qquad \mathbf{K}_{\mathsf{H}} = \mathbf{\Phi}_{\mathsf{H}}\mathbf{\Phi}_{\mathsf{H}}^{H} = \mathbf{K} + j\mathbf{K}_h$$

- **Eigendecompose $\mathbf{K}_{\mathsf{H}}$ to obtain $\mathbf{V}_{\mathsf{H}}$**
- **Extra Promax (oblique) rotation:** $\mathbf{B}_p = \mathbf{R}\mathbf{V}_{\mathsf{H}} \qquad \mathbf{b}_p = \mathbf{b}^p / \|\mathbf{b}^p\|$



SM decomposition

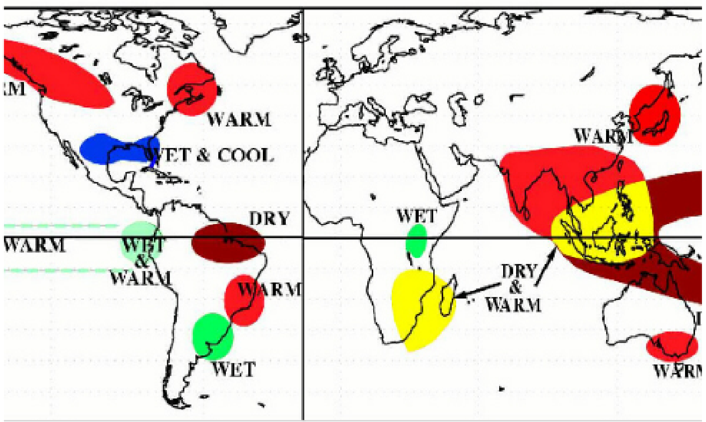
- Strong compression
- 1st: annual oscillation
  - Boreal
  - Equatorial
- 2nd: seasonal oscillation
  - Boreal
  - Croplands
- 3rd: intrannual trend, ENSO

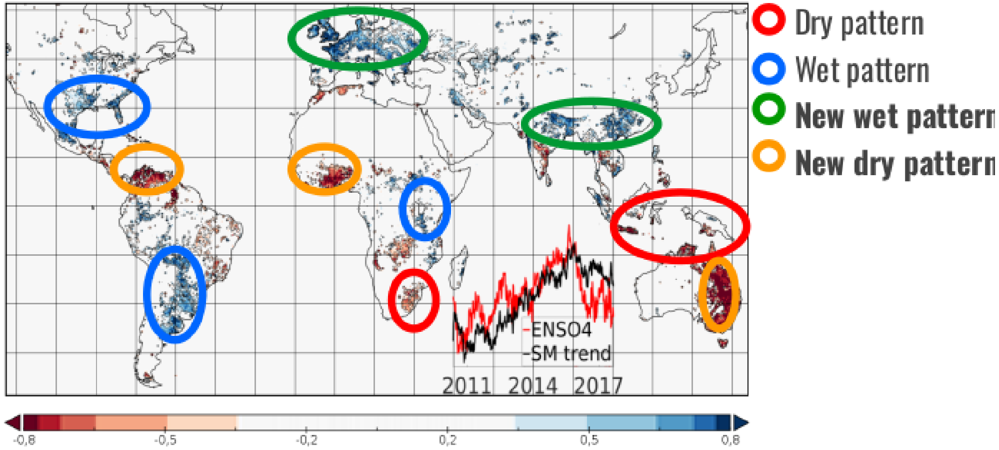


SMOS-BEC (Barcelona Expert Center) + June-2010 to June-2017 + 5 days temporal gap (with asc./desc avgd. orbits) + 25km res.

SM decomposition

- 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

Pause café

Part II.5
Gaussianizing the Earth

PDF estimation is the core of statistics, machine learning and info theory

$$\sum_x p(x) \log p(x) \; \sum_{xy} p(x,y) \log p(y|x) \; \sum_x p(x) \log q(x) \; \sum_x p(x) \left(\frac{\partial p_\theta(x)}{\partial \theta} \right)^2$$

$$\sum_x p(x) \log p(x) \; \sum_{xy} p(x,y) \log p(y|x) \; \sum_x p(x) \log q(x) \; \int_{\mathcal{X}} p(x) \log(p(x)/q(x)) \mathrm{d}x$$

$$\sum_x p(x) \left(\frac{\partial p_\theta(x)}{\partial \theta} \right)^2 \; \sum_x p(x) \log p(x) \; \sum_{xy} p(x,y) \log p(y|x) \; \sum_x p(x) \log q(x)$$

$$\sum_x p(x) \left(\frac{\partial p_\theta(x)}{\partial \theta} \right)^2 \; \int_{\mathcal{X}} p(x) \log(p(x)/q(x)) \mathrm{d}x \; \sum_x p(x) \log p(x)$$

$$\sum_{xy} p(x,y) \log p(y|x) \; \sum_x p(x) \log q(x) \; \sum_x p(x) \left(\frac{\partial p_\theta(x)}{\partial \theta} \right)^2 \; \sum_x p(x) \log p(x)$$

$$\sum_{xy} p(x,y) \log p(y|x) \; \sum_x p(x) \log q(x) \; \sum_x p(x) \left(\frac{\partial p_\theta(x)}{\partial \theta} \right)^2 \; \sum_x p(x) \log p(x)$$

$$\sum_{xy} p(x,y) \log p(y|x) \; \sum_x p(x) \log q(x) \; \int_{\mathcal{X}} p(x) \log(p(x)/q(x)) \mathrm{d}x$$

$$\sum_x p(x) \left(\frac{\partial p_\theta(x)}{\partial \theta} \right)^2$$

Gaussianization for PDF estimation

Why? Statistical independence of features is useful to ...

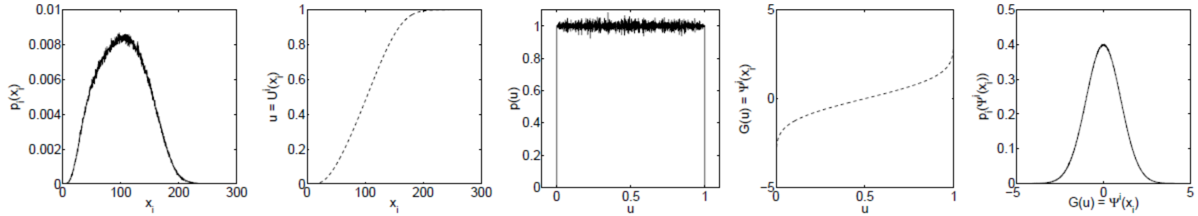
- ... process dimensions independently, no dim curse
- ... tackle the PDF estimation problem directly
- ... and estimate multivariate IT measures

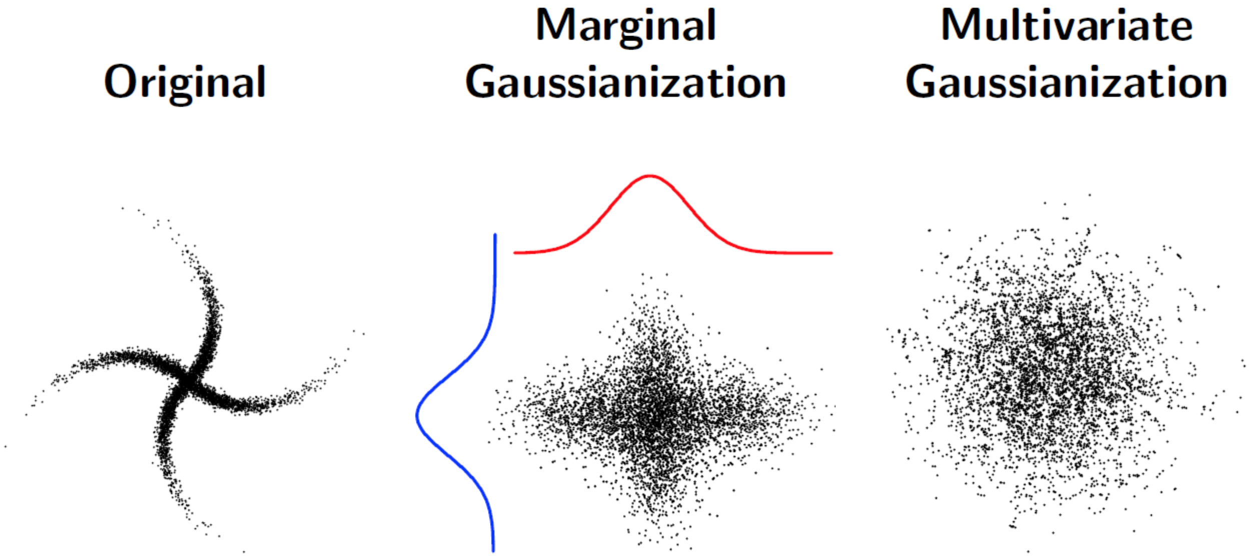
Marginal (univariate) Gaussianization is easy!

Gaussianization in each dimension, $\mathbf{\Psi}^i_{(k)}$, can be decomposed into two consecutive equalization transforms:

1. Marginal uniformization, $U^i_{(k)}$, based on the cdf of the marginal PDF
2. Gaussianization of a uniform variable, $G(u)$, based on the inverse of the cdf of a univariate Gaussian: $\mathbf{\Psi}^i_{(k)} = G \odot U^i_{(k)}$

$$
\begin{aligned}
u = U^i_{(k)}(x_i^{(k)}) &= \int_{-\infty}^{x_i^{(k)}} p_i(x_i'^{(k)})\, dx_i'^{(k)} \\
G^{-1}(x_i) &= \int_{-\infty}^{x_i} g(x_i')\, dx_i'
\end{aligned}
$$




Original
Marginal Gaussianization
Multivariate Gaussianization

RBIG = Rotate and marginally Gaussianize

- √ An orthogonal transform does not affect Gaussianity
- √ Univariate Gaussianization is unique

Rotation-based Iterative Gaussianization (RBIG)

Definition

Given a D-dimensional random variable $\mathbf{x}^{(0)} = [x_1, \ldots, x_D]^\top$ do

$$\mathcal{G} : \mathbf{x}^{(k+1)} = \mathbf{R}_{(k)} \mathbf{\Psi}_{(k)}(\mathbf{x}^{(k)}), \quad k = 1, \ldots, K$$

Rotation-based Iterative Gaussianization (RBIG)

Definition

Given a D-dimensional random variable $\mathbf{x}^{(0)} = [x_1, \ldots, x_D]^\top$ do

$$\mathcal{G} : \mathbf{x}^{(k+1)} = \mathbf{R}_{(k)} \boldsymbol{\Psi}_{(k)}(\mathbf{x}^{(k)}), \quad k = 1, \ldots, K$$

Properties

- $\surd$ is invertible and differentiable
- $\surd$ is valid under any rotation transform (PCA, ICA, random!)
- $\surd$ converges! (negentropy and MI reduce in each iteration)
- $\surd$ is fast (only marginal Gaussianization and rotations needed)
- $\surd$ is a deep neural net! (normalizing flow)
- $\times$ is relatively robust to high-dim spaces
- $\times$ is a meaningless transform

The change-of-variables formula

Let $\mathbf{x} \in \mathbb{R}^D$ be a r.v. with PDF $p_\mathbf{x}(\mathbf{x})$. Given some bijective, differentiable transform of $\mathbf{x}$ into $\mathbf{y}$ using $\mathcal{G} : \mathbb{R}^D \to \mathbb{R}^D$, $\mathbf{y} = \mathcal{G}(\mathbf{x})$, the PDFs are related:

$$p_\mathbf{x}(\mathbf{x}) = p_\mathbf{y}(\mathcal{G}(\mathbf{x})) \left| \frac{d\mathcal{G}(\mathbf{x})}{d\mathbf{x}} \right| = p_\mathbf{y}(\mathcal{G}(\mathbf{x})) \left| \nabla_\mathbf{x} \mathcal{G}(\mathbf{x}) \right|$$

where $|\nabla_\mathbf{x}\mathcal{G}|$ is the determinant of the transform's Jacobian matrix

RBIG for density estimation, $p_{\mathbf{x}}(\mathbf{x}) = p_{\mathbf{y}}(\mathcal{G}(\mathbf{x}))\ |\nabla_{\mathbf{x}}\mathcal{G}(\mathbf{x})|$

- PDF of a multivariate Gaussian:

$$p_{\mathbf{y}}(\mathbf{y}) = p_{\mathbf{y}}(\mathcal{G}(\mathbf{x})) \propto \exp(-\frac{1}{2}(\mathcal{G}(\mathbf{x}) - \boldsymbol{\mu}_{\mathbf{y}})^\top \boldsymbol{\Sigma}^{-1}(\mathcal{G}(\mathbf{x}) - \boldsymbol{\mu}_{\mathbf{y}}))$$

- Jacobian is the product of Jacobians:

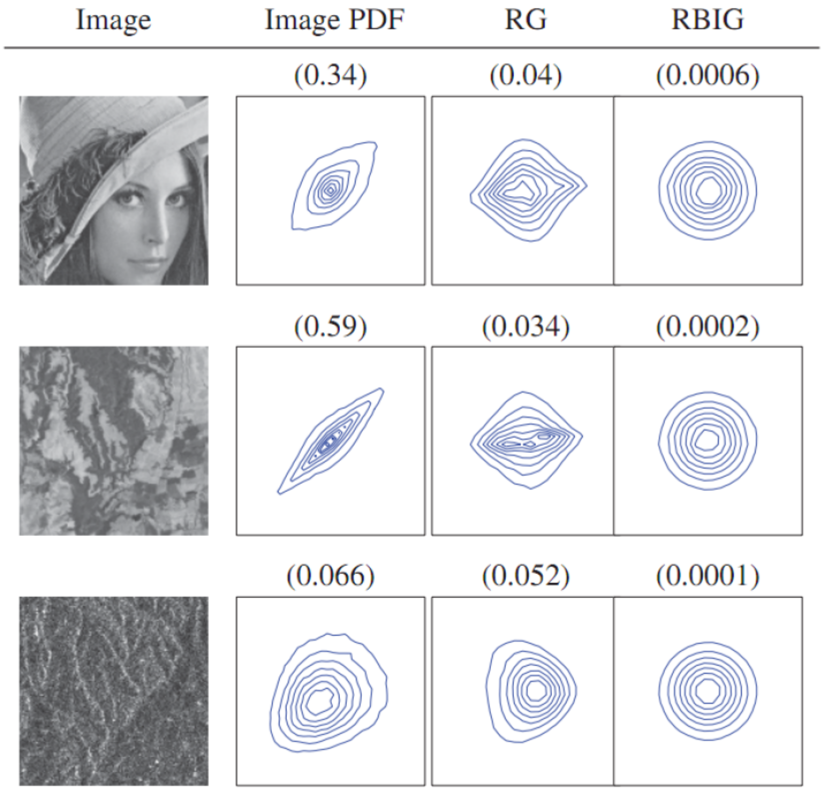
$$\nabla_{\mathbf{x}}\mathcal{G} = \textstyle\prod_{k=1}^{K} \mathbf{R}_{(k)} \nabla_{\mathbf{x}^{(k)}} \boldsymbol{\Psi}_{(k)}$$

$$\nabla_{\mathbf{x}^{(k)}} \boldsymbol{\Psi}_{(k)} = \begin{pmatrix} \frac{\partial \boldsymbol{\Psi}^1_{(k)}}{\partial x_1^{(k)}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\partial \boldsymbol{\Psi}^d_{(k)}}{\partial x_d^{(k)}} \end{pmatrix}, \quad \frac{\partial \boldsymbol{\Psi}^i_{(k)}}{\partial x_i^{(k)}} = \frac{\partial \mathcal{G}}{\partial u} \frac{\partial u}{\partial x_i^{(k)}}$$

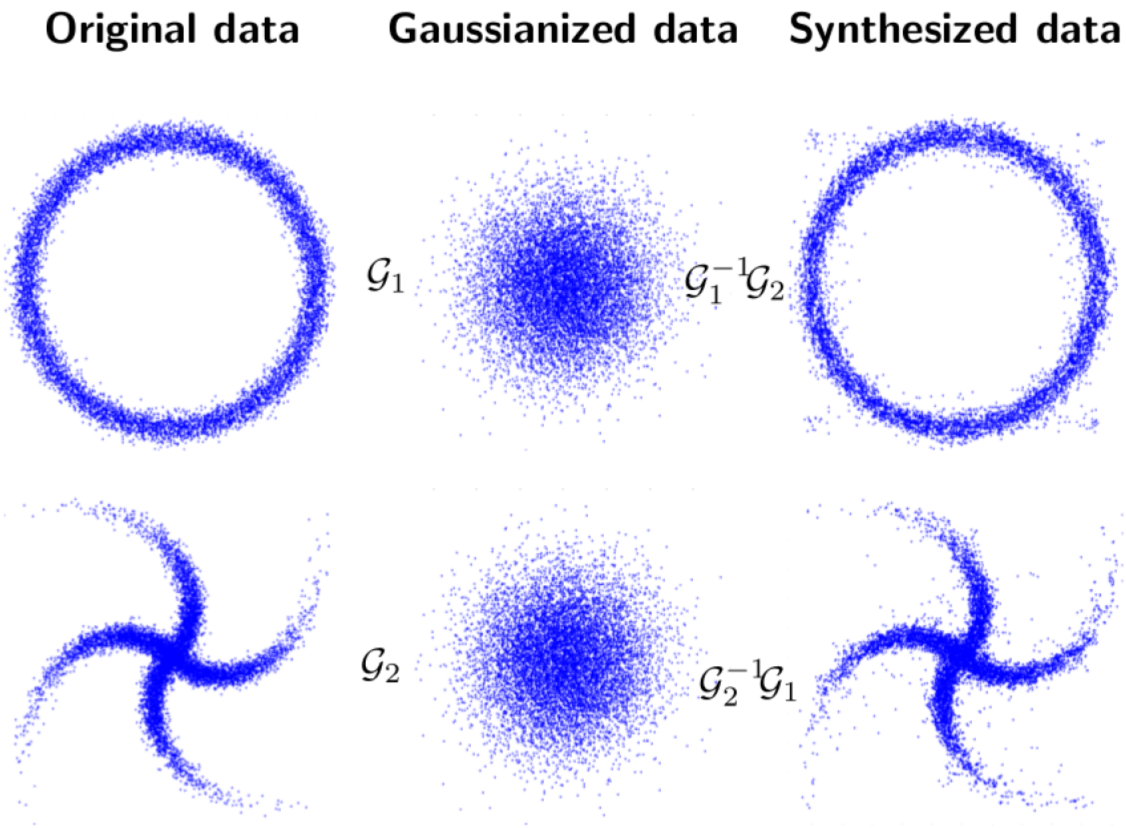
- Invertible:

$$\mathcal{G} : \mathbf{x}^{(k+1)} = \mathbf{R}_{(k)} \boldsymbol{\Psi}_{(k)}(\mathbf{x}^{(k)}) \rightarrow \mathcal{G}^{-1} : \mathbf{x}^{(k)} = \boldsymbol{\Psi}^{-1}_{(k)}(\mathbf{R}^\top_{(k)} \mathbf{x}^{(k+1)})$$

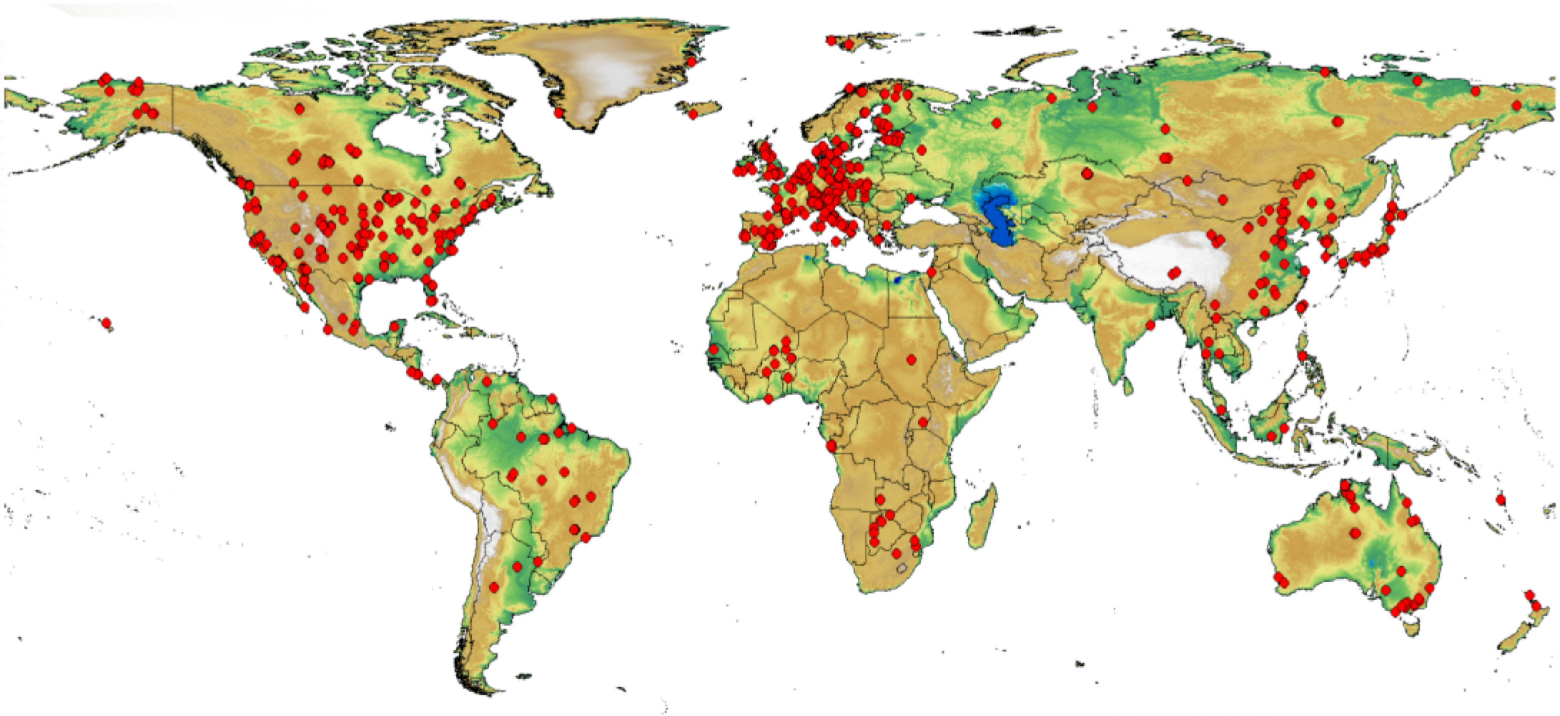
It works in arbitrary natural and remote sensing images



Synthesis in low dimensions

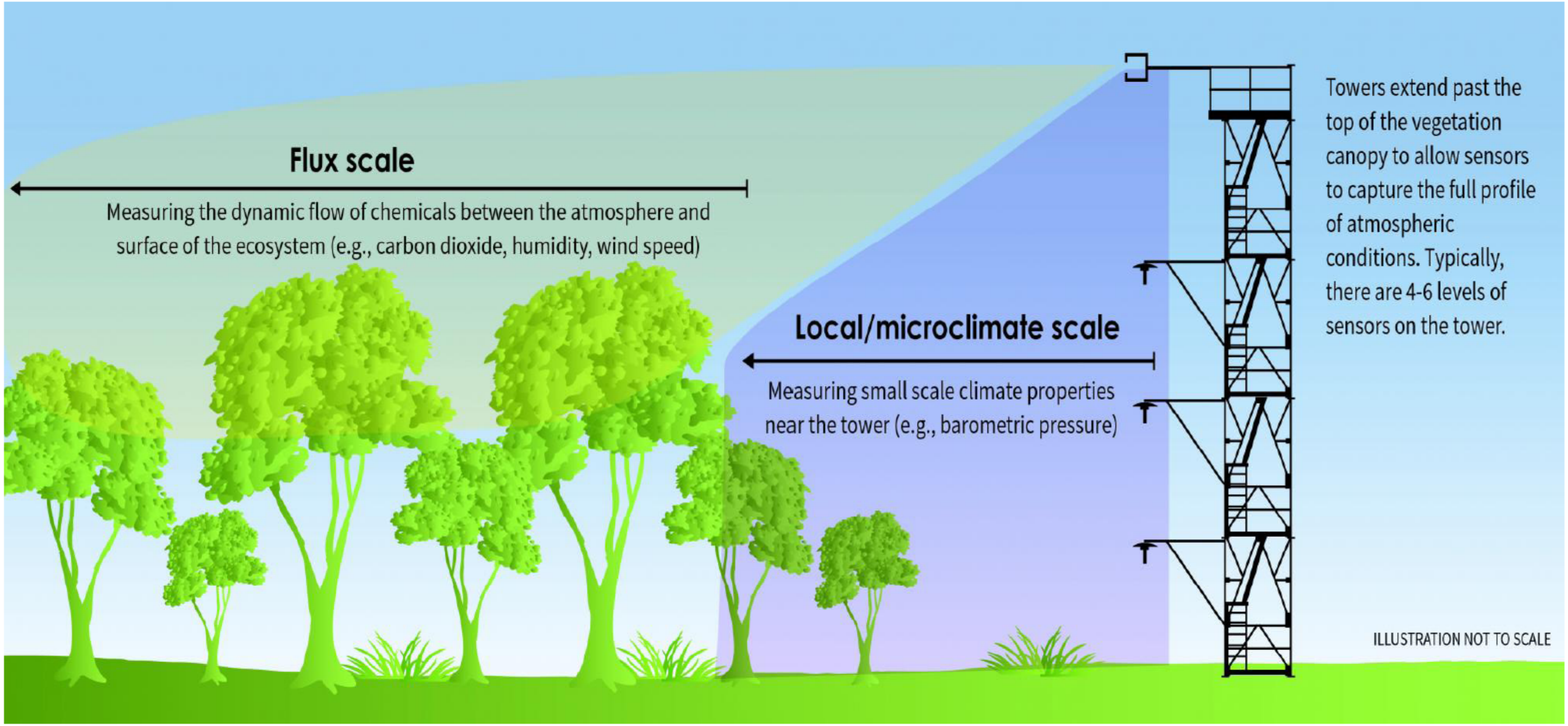


Synthesis in moderate dimensions



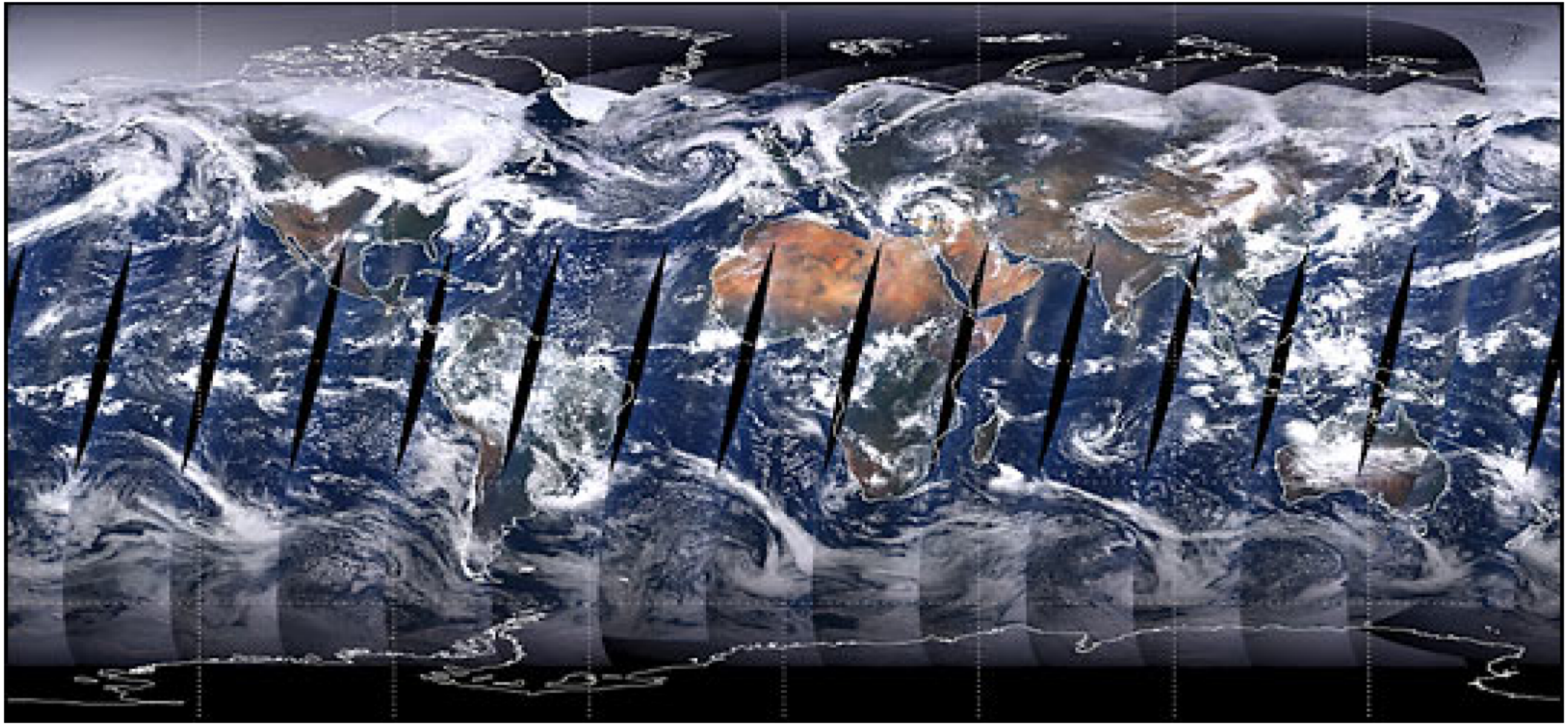
FLUXNET: a network for micro-meteorological tower sites

Synthesis in moderate dimensions



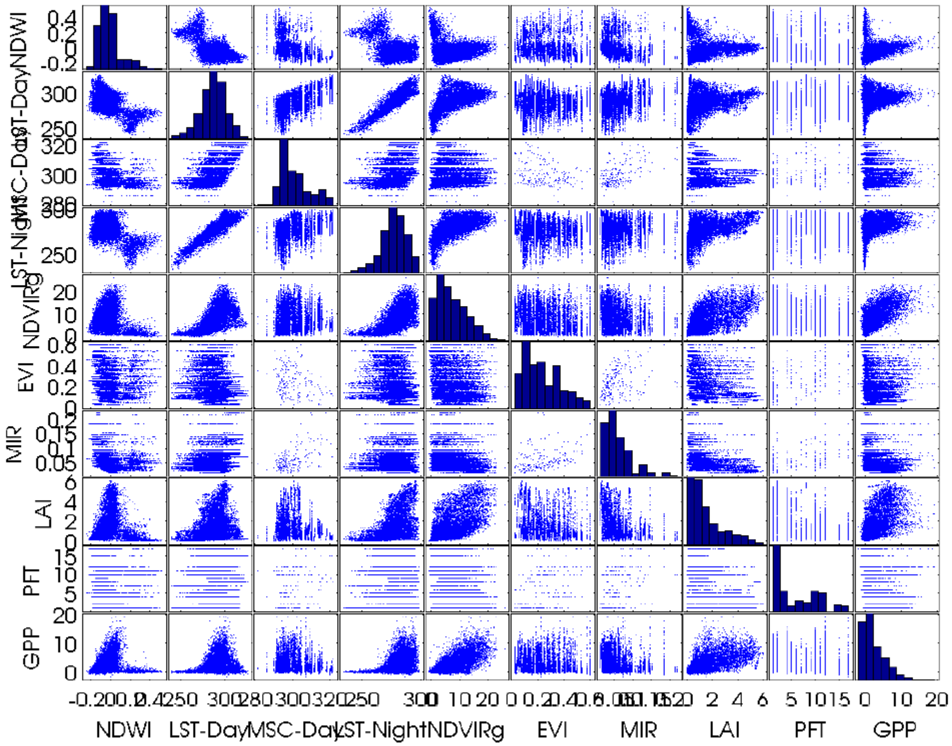
Sensors allow estimating turbulent exchange of carbon dioxide (CO_2), latent and sensible heat, CO_2 storage, net ecosystem exchange, energy balance, ...

Synthesis in moderate dimensions

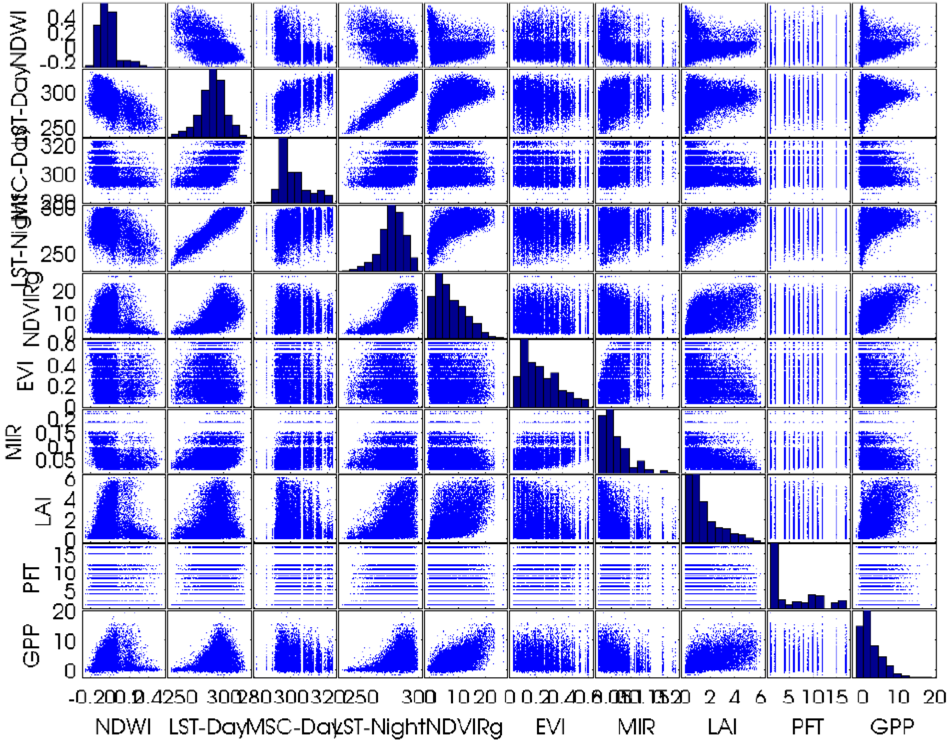


MODIS sensor: 36 channels, 8-daily, 500 m

Synthesis in moderate dimensions



Synthesis in moderate dimensions

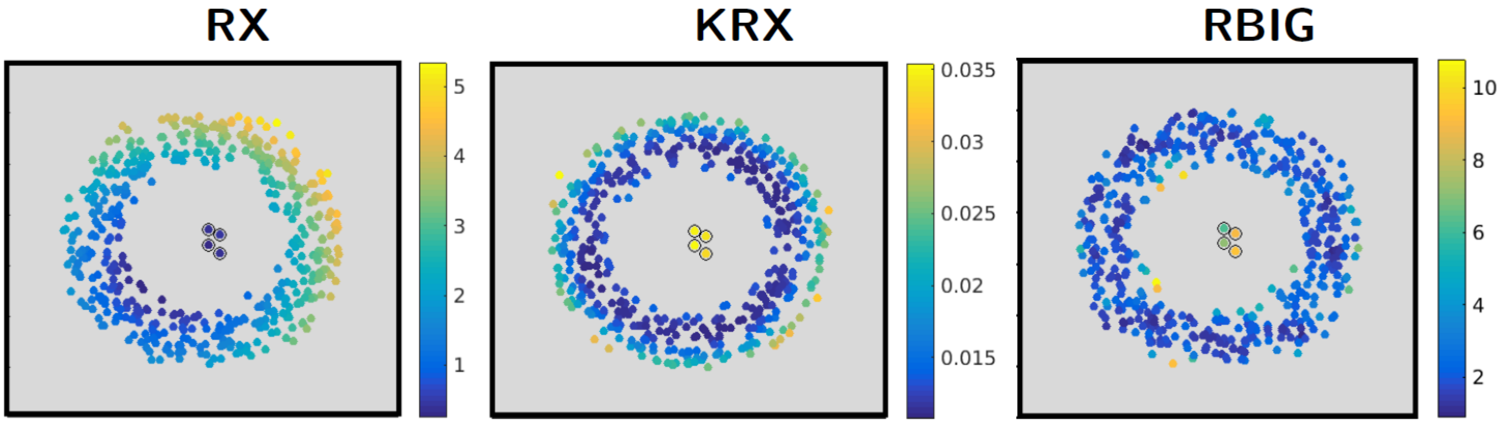


Synthesis in moderate dimensions

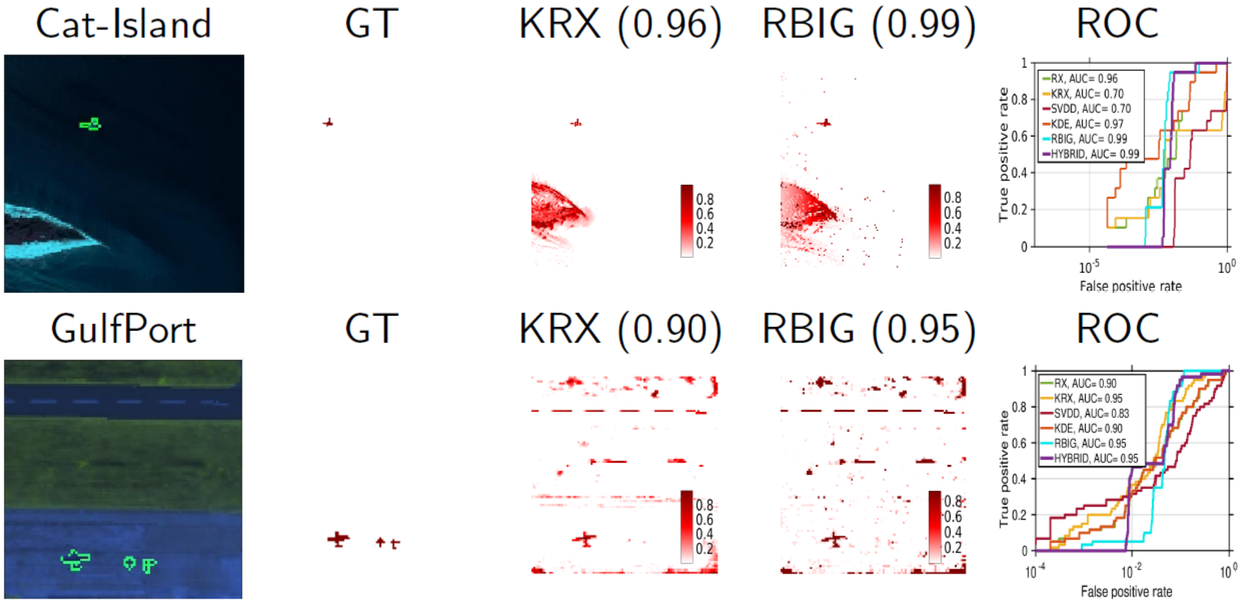
real, $n = 10^4$	ME	RMSE	MAE	R
LR	-0.01	1.82	1.28	0.78
GPR	+0.03	1.72	1.14	**0.81**
real+syn, $n = 10^6$	ME	RMSE	MAE	R
LR	-0.01	1.80	1.27	0.79
GPR	**-0.00**	**1.63**	**1.03**	**0.83**

Anomaly and extreme detection

$$A_{\mathrm{RX}}(\mathbf{x}) = (\mathbf{x}-\mu)^\top \mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)$$
$$A_{\mathrm{KRX}}(\mathbf{x}) = \tilde{k}(\mathbf{x},\cdot)^\top(\mathbf{K}\mathbf{K})^{-1}\tilde{k}(\mathbf{x},\cdot)$$
$$A_{\mathrm{RBIG}}(\mathbf{x}) \propto \frac{1}{p_{\mathrm{RBIG}}(\mathbf{x})}$$



Anomaly and extreme detection



RBIG framework allows to compute all IT measures

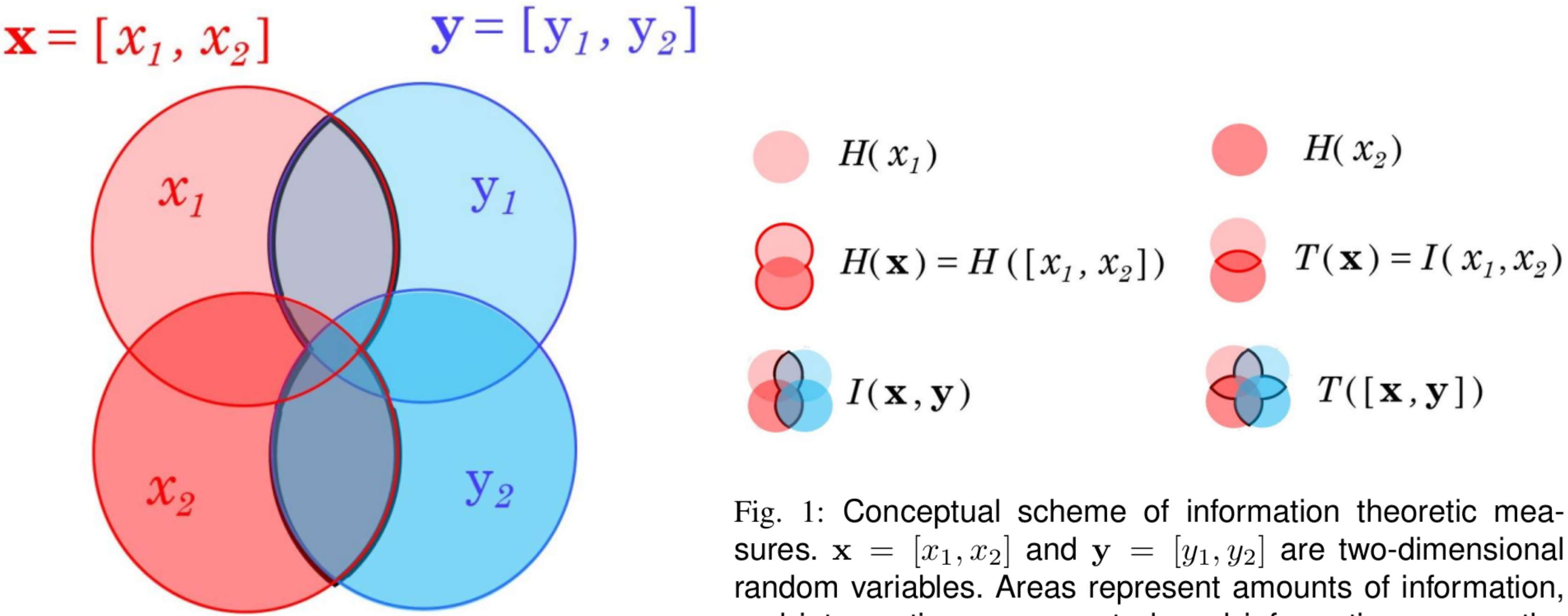


Fig. 1: Conceptual scheme of information theoretic measures. $\mathbf{x} = [x_1, x_2]$ and $\mathbf{y} = [y_1, y_2]$ are two-dimensional random variables. Areas represent amounts of information, and intersections represent shared information among the corresponding variables and dimensions. Examples of entropy, total correlation and mutual information are given.

RBIG framework allows to compute all IT measures

1. Total Correlation (aka multi-information)

$$TC = \sum_{k=1}^{K} \left(D \; h(\mathcal{N}(0,1)) + \sum_{d=1}^{D} h(\mathbf{x}_d^{(k)}) \right)$$

RBIG framework allows to compute all IT measures

1. Total Correlation (aka multi-information)

$$TC = \sum_{k=1}^{K} \left(D\ h(\mathcal{N}(0,1)) + \sum_{d=1}^{D} h(\mathbf{x}_d^{(k)}) \right)$$

2. Multidimensional entropy (and negentropy):

$$H(\mathbf{x}) = \sum_{d=1}^{D} h(\mathbf{x}_d) - TC(\mathbf{x})$$

RBIG framework allows to compute all IT measures

1. Total Correlation (aka multi-information)

$$TC = \sum_{k=1}^{K} \left(D \; h(\mathcal{N}(0,1)) + \sum_{d=1}^{D} h(\mathbf{x}_d^{(k)}) \right)$$

2. Multidimensional entropy (and negentropy):

$$H(\mathbf{x}) = \sum_{d=1}^{D} h(\mathbf{x}_d) - TC(\mathbf{x})$$

3. Kullback-Leibler divergence: $D_{KL}(\mathbf{x}\|\mathbf{y}) = TC(\mathcal{G}_x(\mathbf{y}))$

RBIG framework allows to compute all IT measures

1. Total Correlation (aka multi-information)

$$TC = \sum_{k=1}^{K} \left(D\ h(\mathcal{N}(0,1)) + \sum_{d=1}^{D} h(\mathbf{x}_d^{(k)}) \right)$$

2. Multidimensional entropy (and negentropy):

$$H(\mathbf{x}) = \sum_{d=1}^{D} h(\mathbf{x}_d) - TC(\mathbf{x})$$

3. Kullback-Leibler divergence: $D_{KL}(\mathbf{x} \| \mathbf{y}) = TC(\mathcal{G}_x(\mathbf{y}))$
4. Conditional independence

$$\begin{aligned} I(\mathbf{x}, \mathbf{y}|\mathbf{z}) &= H(\mathbf{x}, \mathbf{z}) + H(\mathbf{y}, \mathbf{z}) - H(\mathbf{x}, \mathbf{y}, \mathbf{z}) - H(\mathbf{z}) \\ &= TC(\mathbf{x}, \mathbf{y}, \mathbf{z}) - TC(\mathbf{x}, \mathbf{z}) - TC(\mathbf{y}, \mathbf{z}) \end{aligned}$$

with the null hypothesis distribution $p(I(\mathbf{x}, r(\mathbf{y})|\mathbf{z}))$

But ... how to estimate total correlation?

1: Given data $\mathbf{x}^{(0)} = [x_1, \ldots, x_D]^\top \in \mathbb{R}^D$
2: Learn the sequence of Gaussianization transforms $\mathbf{y} = \mathcal{G}(\mathbf{x})$
3: Compute the cumulative reduction in mutual information

$$TC = \sum_{k=1}^{K} \left(D\ h(\mathcal{N}(0,1)) + \sum_{d=1}^{D} h(\mathbf{x}_d^{(k)}) \right)$$

Total Correlation

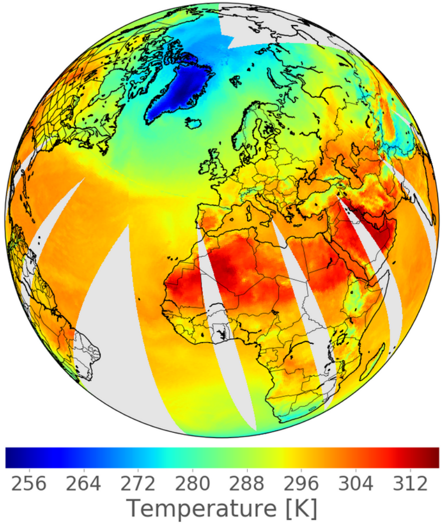
TABLE 1: Relative mean absolute errors in percentage for total correlation estimation on known PDFs. Best value in dark gray, second best value in bright gray.

		D_x	**RBIG**	**kNN**	**KDP**	**expF**	**vME**	**Ens**
Gaussian	-	3	1.5	2.5	159.2	1.2	8.5	9.8
		10	3.1	31.2	>100	0.2	33.9	44.9
		50	1.3	32.7	>100	0.1	>100	38.7
		100	0.8	31.0	89.9	0.1	94.2	34.9
Rotated	-	3	1.70	1.80	82.90	16.80	1.90	9.40
		10	8.30	27.20	>100	11.00	24.20	38.70
		50	7.70	51.10	>100	15.10	>100	59.40
		100	7.50	57.80	>100	15.50	>100	64.50
Student	$\nu = 3$	3	7.01	13.55	>100	94.03	>100	66.59
		10	32.93	16.73	>100	67.32	>100	15.27
		50	18.18	12.02	>100	29.44	>100	24.65
		100	12.71	17.41	>100	21.12	>100	28.63
	$\nu = 5$	3	26.61	52.76	>100	89.74	81.85	133.12
		10	23.94	19.74	>100	49.60	>100	12.31
		50	10.10	16.87	>100	20.29	>100	32.14
		100	7.10	22.53	>100	15.39	>100	34.96
	$\nu = 20$	3	88.27	>100	>100	48.56	>100	>100
		10	3.05	11.86	>100	10.51	>100	19.93
		50	3.07	33.17	>100	4.54	>100	52.62
		100	1.31	35.56	>100	3.43	>100	49.46

TABLE 2: Relative mean absolute errors in percentage for entropy estimation on known PDFs. Best value in dark gray, second best value in bright gray.

		D_x	**RBIG**	**kNN**	**KDP**	**expF**	**vME**	**Ens**
Gaussian		3	1.5	2.5	159.2	1.2	8.5	9.8
		10	3.1	31.2	>100	0.2	33.9	44.9
		50	1.3	32.7	>100	0.1	>100	38.7
		100	0.8	31.0	89.9	0.1	94.2	34.9
Rotated		3	2.8	4.7	127.2	37.2	3.6	22.7
		10	17.4	45.2	263.8	23.9	4.5	62.0
		50	7.6	46.0	140.2	14.2	87.6	53.1
		100	5.2	43.50	113.9	12.1	94.3	48.3
Student	$\nu = 3$	3	0.56	0.62	35.7	11.5	3.25	2.11
		10	2.81	1.45	138.2	15.9	52.9	1.80
		50	6.12	3.37	198.7	22.43	175.4	6.96
		100	6.88	8.45	237.3	25.34	164.9	13.59
	$\nu = 5$	3	0.27	0.66	24.9	3.50	1.24	2.00
		10	1.16	1.26	96.2	5.63	59.23	1.23
		50	2.80	4.77	147.5	9.61	202.3	8.81
		100	3.17	10.6	187.7	11.4	194.9	16.2
	$\nu = 20$	3	0.27	0.49	19.2	0.70	1.41	1.76
		10	0.54	0.82	70.6	1.6	46.6	0.30
		50	0.93	6.62	107.3	3.37	219.7	11.06
		100	0.69	13.4	139.6	4.23	214.2	19.24

Total Correlation



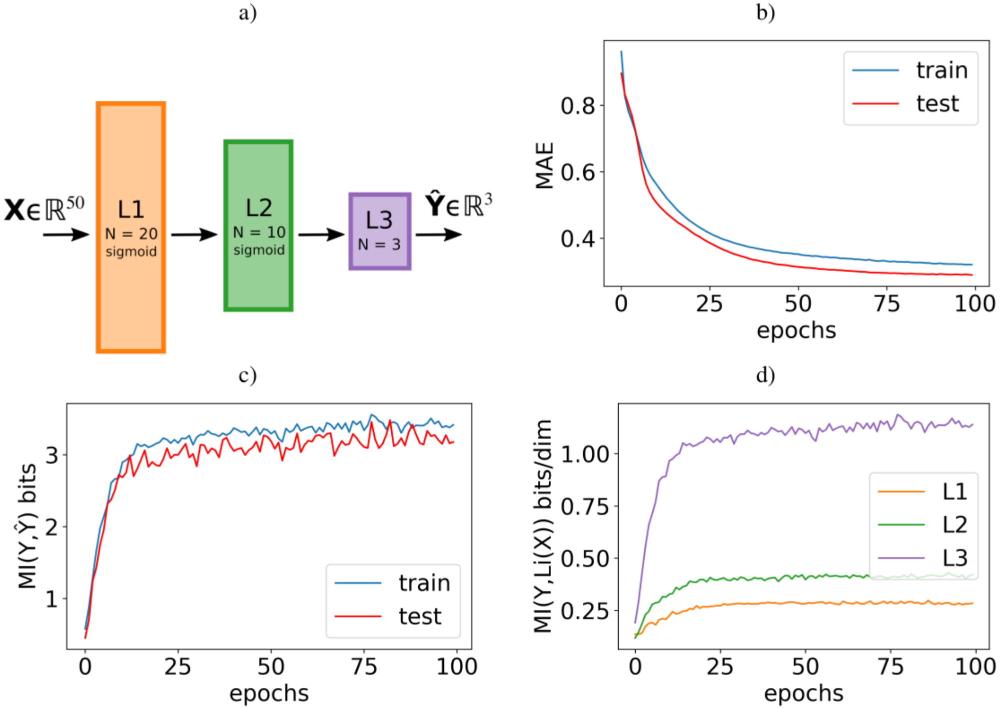
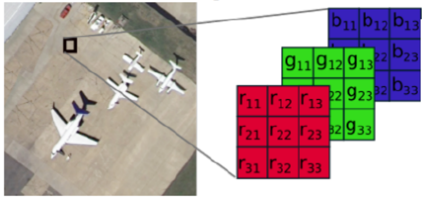


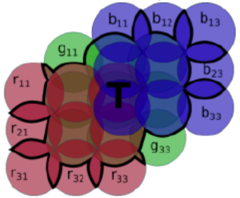
Fig. 8: Learning in artificial neural networks from RBIG estimations of mutual information: evolution of I during the training of an ANN. a) Configuration of the considered neural network. b) Error evolution. c) Evolution of I between the predicted output and the actual data. d) Evolution of I per dimension between the output of each layer and the actual data.

Information in high spatial resolution images

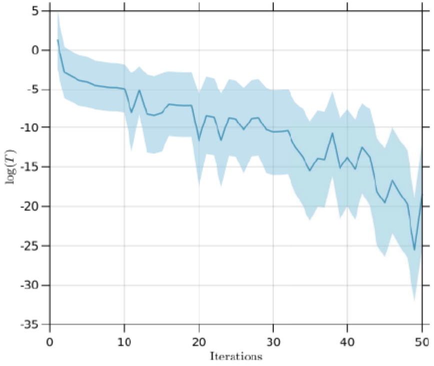
(a) Data partition



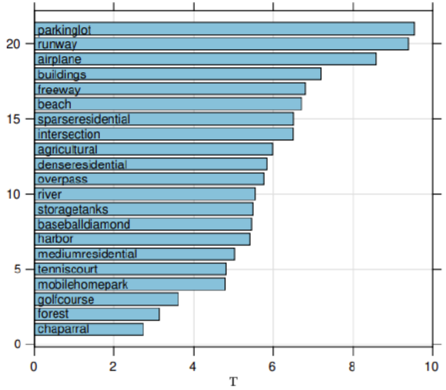
(b) Measured magnitude



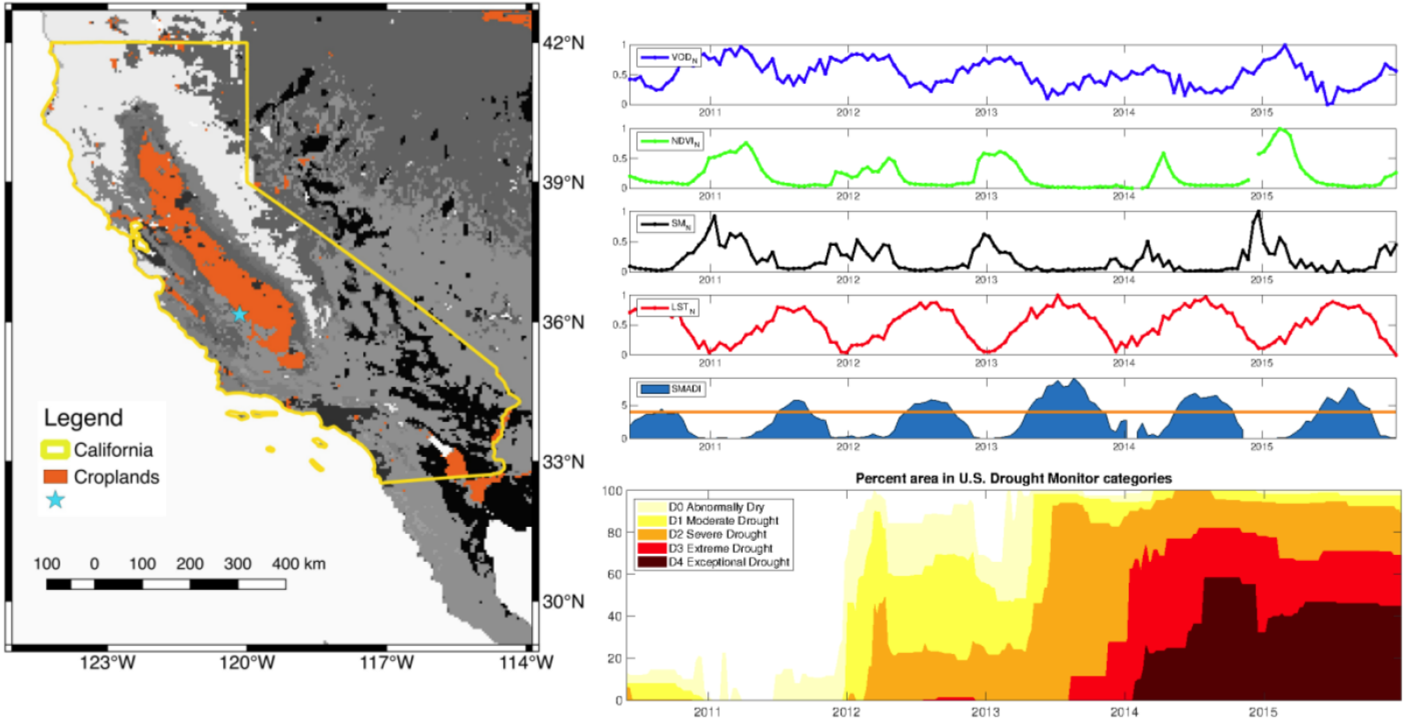
(c) Averaged convergence



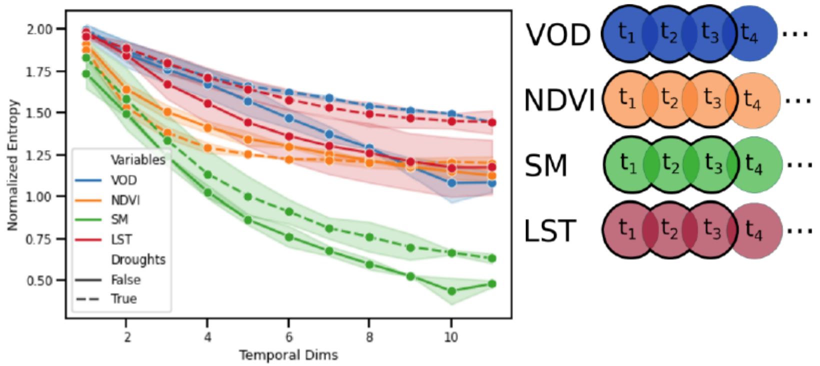
(d) Class-specific T

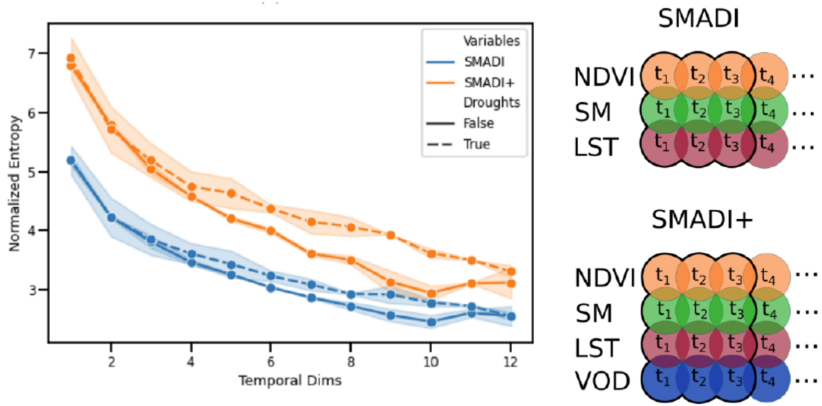


Information in terrestrial biosphere over time



Information in terrestrial biosphere over time





Conclusions

Take-home messages

- $\surd$ Simple, Fast, Versatile, Hyperparameter free
- $\surd$ Info bottleneck with multivariate measures
- $\surd$ Many applications possible, use it!
  - https://isp.uv.es/rbig.html
  - https://github.com/IPL-UV/rbig_jax

Future steps

- Train all layers at the same time
- Conditional Independence Test
- Conditional Density Estimation

References

- *"Iterative Gaussianization: From ICA to random rotations"* V. Laparra, G. Camps-Valls, J. Malo, IEEE Transactions on Neural Networks, 22(4):537549, Apr 2011
- *"Gaussianizing the Earth,"* J. Johnson, V. Laparra, M. Piles, and G. Camps-Valls, in IEEE Geoscience & Remote Sensing Magazine, 2020.
- *"Information Theory in Density Destructors,"* Johnson, J.E. Laparra, V. Santos-Rodriguez, R., Camps-Valls, G., Malo, J., International Conference on Machine Learning (ICML), 2019
- *"Information Theory Measures using Gaussianization,"* V. Laparra, E. Jonhson, G. Camps-Valls, R. Santos-Rodrguez, Jess Malo, IEEE Transactions on Information Theory, submitted, 2020

Pause café

Agenda for today

- Part I: Introduction: why do we need ML?
- Part II: ML for Earth sciences
- **Part III: The challenges**
- **Part IV: Physics-aware Machine Learning**
- **Part V: Explainable AI**
- **Part VI: Pragmatic causality**

Part III
The challenges

AI promises to transform scientific discovery ...

Deep learning challenges

- Do Models respect Physics Laws?
- What did the ML model learn?
- Do they get cause-effect relations?

The New York Times

Opinion

OP-ED CONTRIBUTORS

Eight (No, Nine!) Problems With Big Data

By Gary Marcus and Ernest Davis

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio &

Archive > Volume 538 > Issue 7623 > News Feature > Article

NATURE | NEWS FEATURE

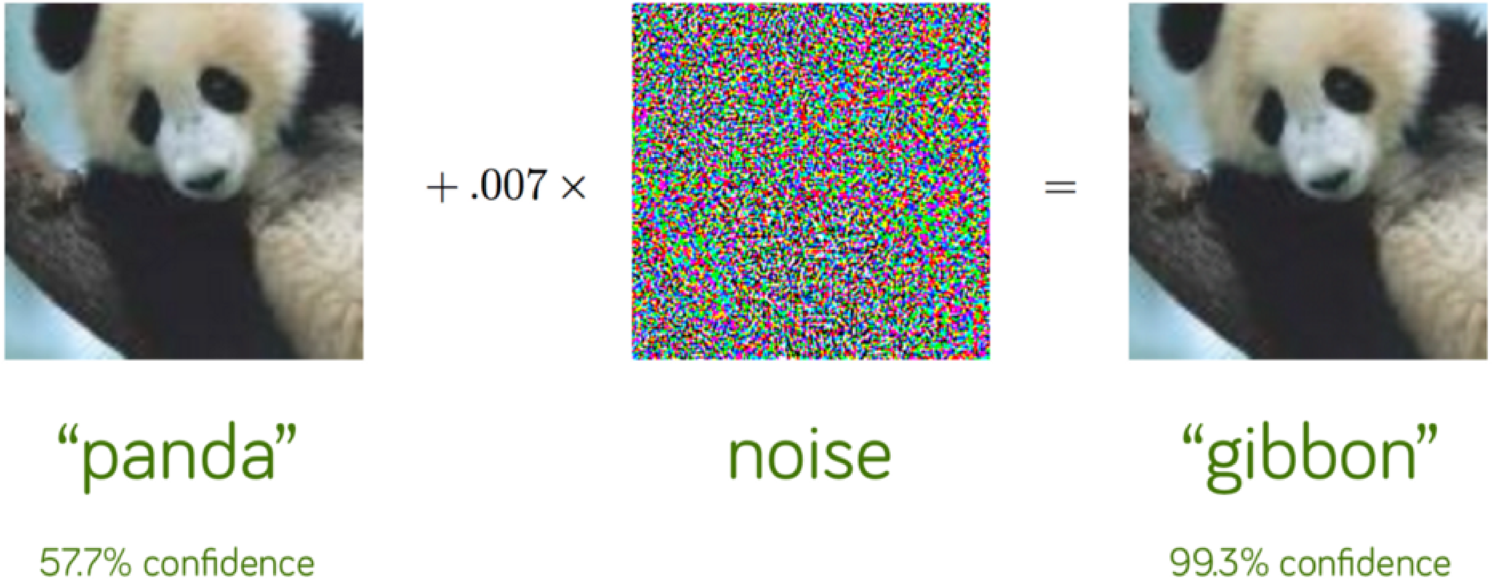
Can we open the black box of AI?

Artificial intelligence is everywhere. But before scientists trust it, they first need t understand how machines learn.

Davide Castelvecchi

Microsoft Tay (and chatGPT too) learns biases

Automatic systems can be fooled with adversarial examples



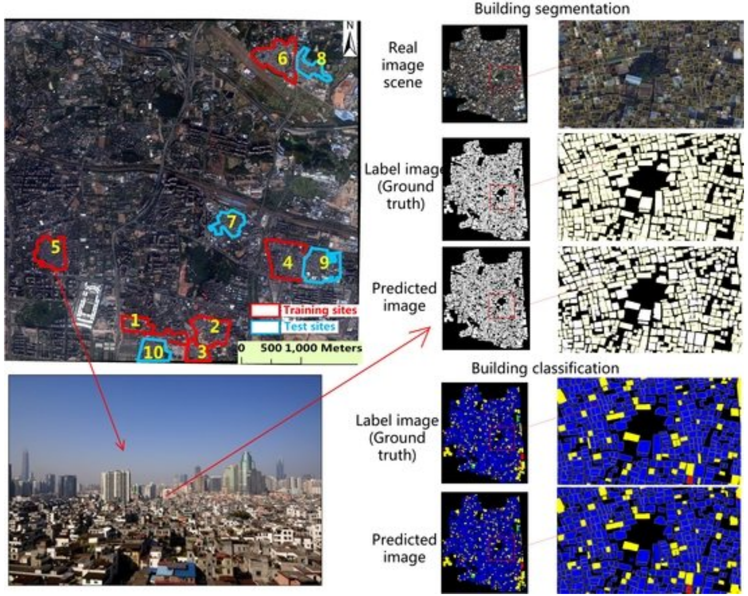
DL suffers hallucinations (aka confabulations, delusions)

- Confident response not justified by the training data

PA

summarise this article https://www.nytimes.com/2023/03/11/technology/chatgpt-prompts-to-avoid-content-filters.html

The article discusses how ChatGPT, an AI language model, is being used to circumvent content filters that are designed to detect and block certain types of content. By generating prompts that are less likely to trigger these filters, ChatGPT can be used to create content that might otherwise be censored. While this technology could be useful for those seeking to express themselves freely, it also raises concerns about the potential misuse of AI-generated content, as well as the difficulty of regulating such content in the future. The article concludes by noting that there is still much to learn about the impact of AI language models like ChatGPT on the way we communicate online.



Current AI doesn't get context, composition and meaning

Current AI doesn't get semantics ...

AI doesn't understand physics and causality!


THERE IS
light
AT THE END
of the
TUNNEL

Deep learning challenges

- Do Models respect Physics Laws? **Physics-aware machine learning**
- What did the ML model learn? **Explainable AI**
- Do they get cause-effect relations? **Causal inference & discovery**


The New York Times

Opinion

OP-ED CONTRIBUTORS

Eight (No, Nine!) Problems With Big Data

By Gary Marcus and Ernest Davis


nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio &

Archive > Volume 538 > Issue 7623 > News Feature > Article

NATURE | NEWS FEATURE

Can we open the black box of AI?

Artificial intelligence is everywhere. But before scientists trust it, they first need t understand how machines learn.

Davide Castelvecchi

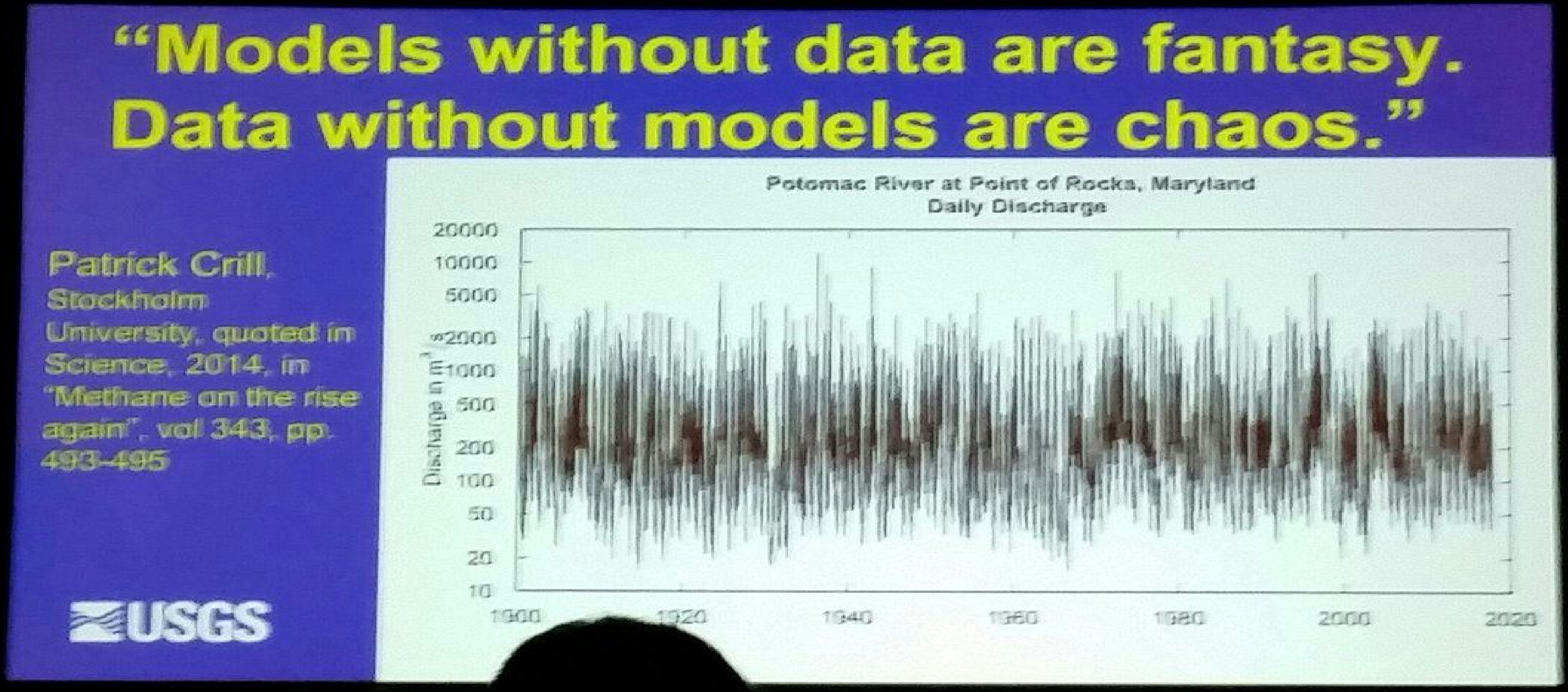
Part IV
Physics-aware machine learning

Physics-*aware** machine learning

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

* aka physics-guided, physics-informed, physics-constrained, science-guided, …

The truth is that…



At AGU 2017, New Orleans, USA

A simple taxonomy

A — Data-model blending
- Joint Gaussian processes
- Distribution regression

B — Surrogate modeling
- Gaussian processes
- Bayesian optimization

C — Learning to parameterize
- Variational inference
- Monte Carlo, Gibbs

D — Learning physics
- Sparse regression
- Latent force models

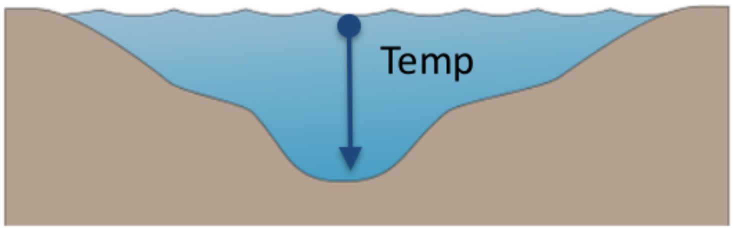
"Living in the Physics - Machine Learning Interplay for Earth Observation"
Camps-Valls et al. AAAI Fall Series 2020 Symposium on Physics-guided AI for Accelerating Scientific Discovery, 2020. arxiv.org/abs/2010.09031

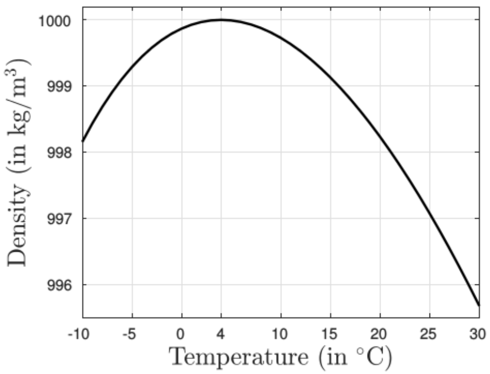
A- Constrained optimization

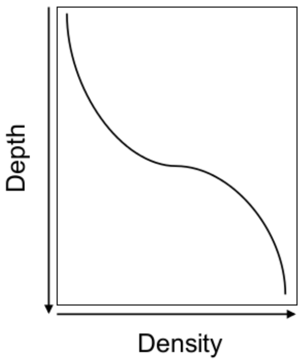
- ML minimizing model errors & violations of the physical laws

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

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







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

A- Fair optimization

- ML minimizing errors & predictions independent of sensitive factors

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

- Independence measured with HSIC

$$I := \mathrm{HSIC}(\mathcal{Y}, \mathcal{H}, \mathbb{P}_{\mathbf{ys}}) = \|\mathbf{C}_{ys}\|_{\mathrm{HS}}^2$$

- Closed form solution with kernels

$$\boldsymbol{\Lambda} = (\tilde{\mathbf{K}} + \lambda \mathbf{I} + \frac{\mu}{n^2} \tilde{\mathbf{K}} \tilde{\mathbf{K}}_S)^{-1} \mathbf{Y}$$

- Probabilistic interpretation with GPs:

$$f \sim \mathcal{GP}\left(0, k(\cdot, \cdot) - k_{\mathbf{X}\cdot}^\top (\mathbf{KHLH} + \delta^{-1}\mathbf{I})^{-1} \mathbf{HLH} k_{\mathbf{X}\cdot}\right)$$

"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

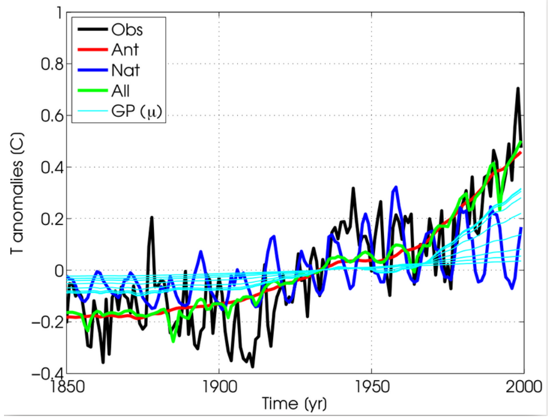
"Physics-aware Nonparametric Regression Models for Earth Data Analysis". Cortés & Camps-Valls. Environmental Research Letters,, 2022

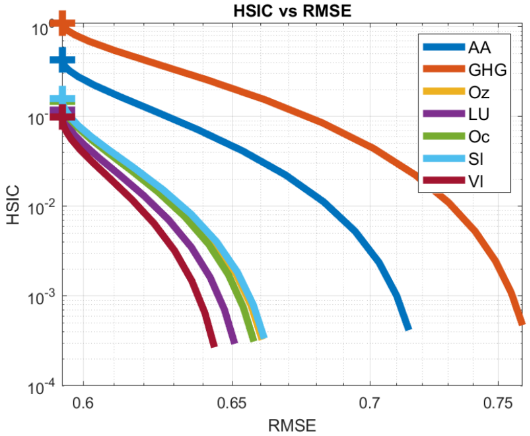
"Kernel Dependence Regularizers and Gaussian Processes with application to Algorithmic Fairness" Zhu Li, Perez-Suay, Camps-Valls and Sejdinovic, Pattern Rec. 2022

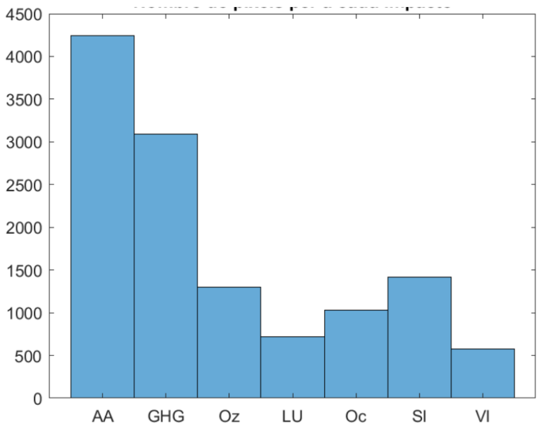
A- Fair optimization

- ML minimizing errors & predictions independent of human 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

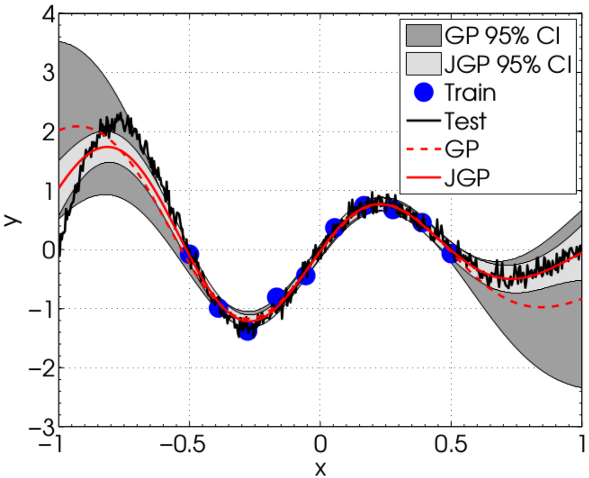
“**Kernel Dependence Regularizers and Gaussian Processes with application to Algorithmic Fairness**” Zhu Li, Perez-Suay, Camps-Valls and Sejdinovic, , Pattern Rec. 2022

A- Blending observations and simulations for extrapolation

- Let ML talk to physical models for extrapolation

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

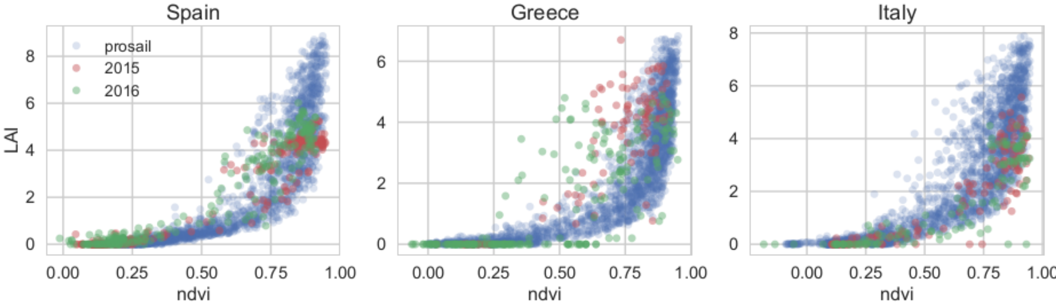
$$\Omega(\hat{y}, \Phi) = \mathrm{Cost}_s(y_s, \hat{y}_s)$$



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 ...

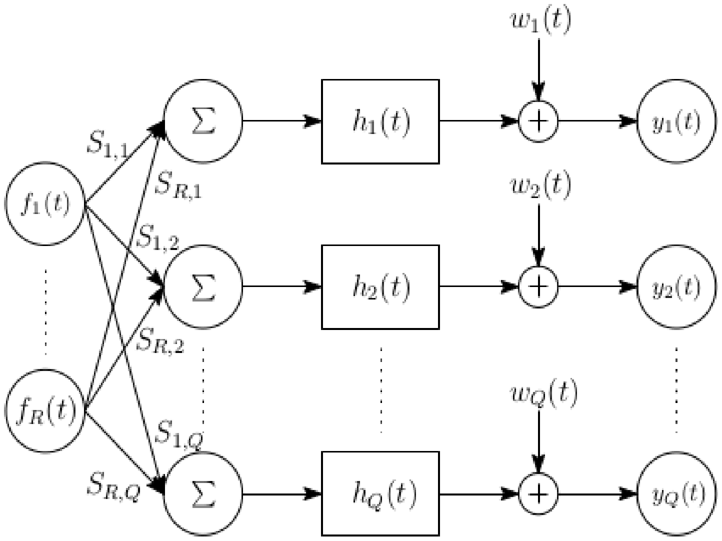


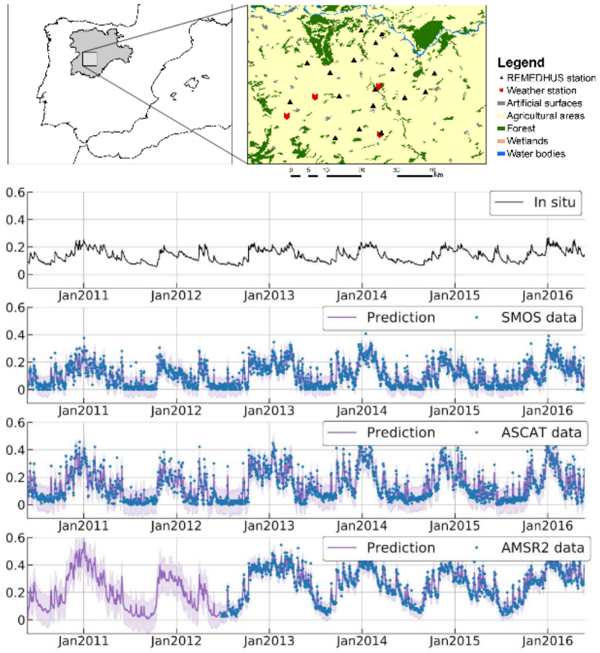
"**Joint Gaussian Processes for Biophysical Parameter Retrieval**" Svendsen, Martino, Camps-Valls, IEEE TGARS 2018

"**Physics-aware Gaussian processes in remote sensing**" Camps-Valls, G. et al. Applied Soft Computing, 2018.

A- Convolution processes & encoding ODEs

- Encode ODEs governing the system + Learn latent forces driving it

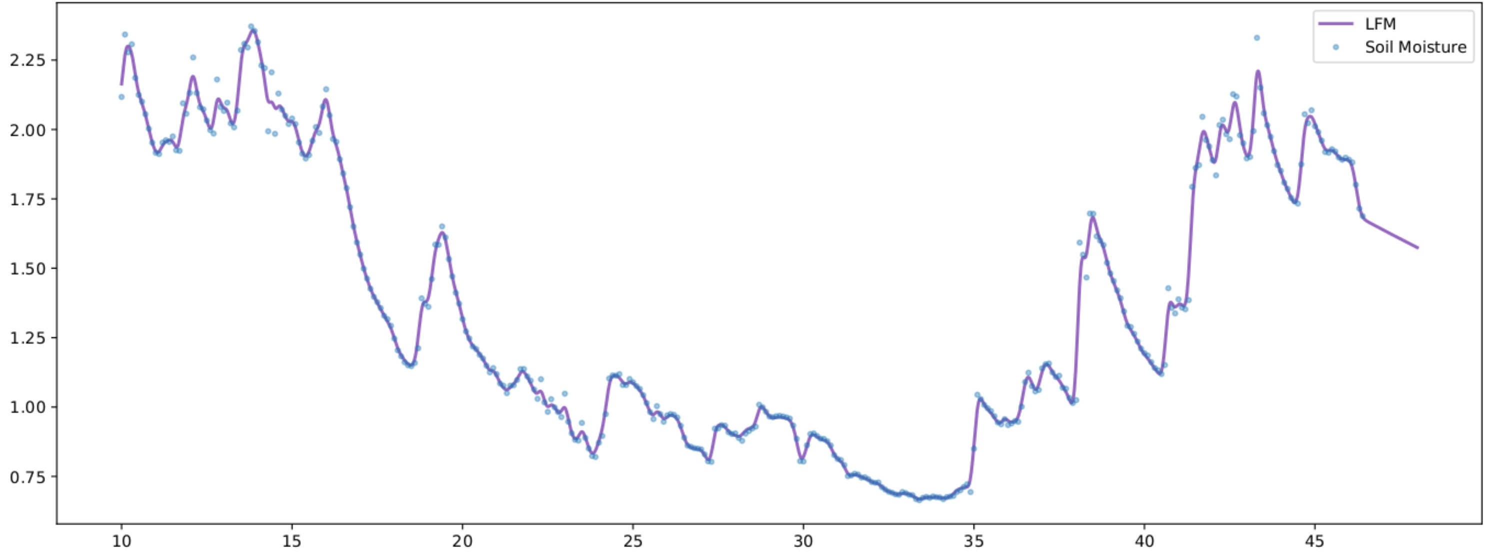




"Integrating Domain Knowledge in Data-driven Earth Observation with Process Convolutions" Svendsen, Muñoz, Piles, Camps-Valls, IEEE TGARS. 2021

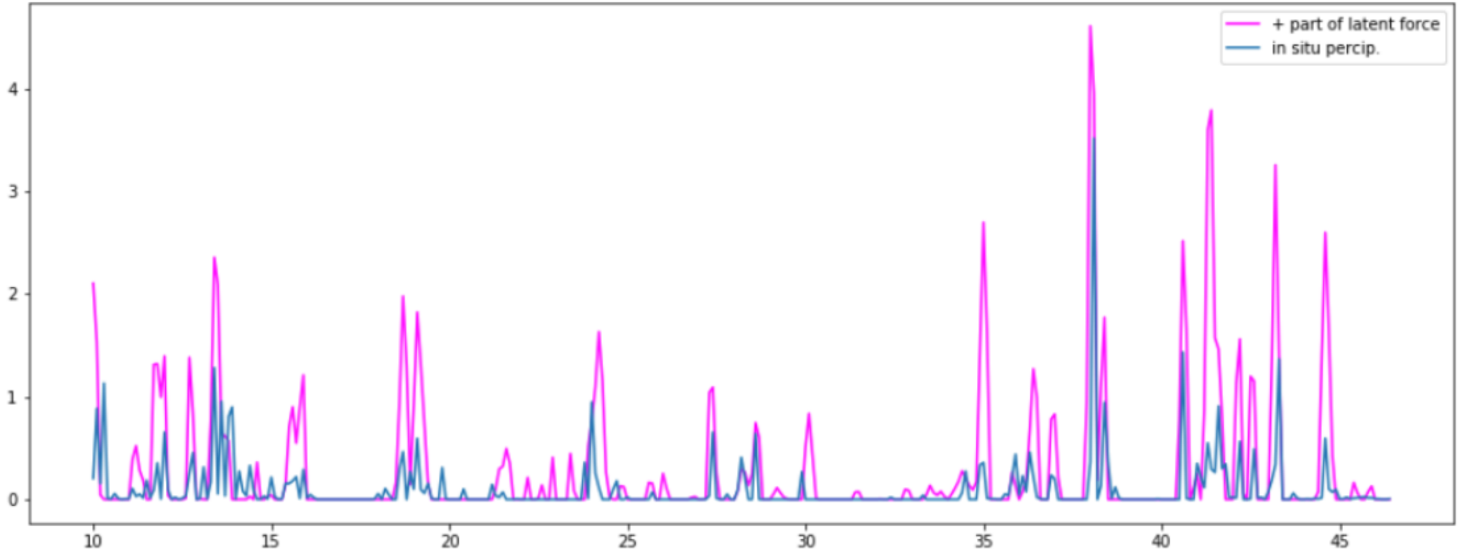
A- Convolution processes & encoding ODEs

- LFM-GP learns to estimate SM from uneven sampled time series ...



A- Convolution processes & encoding ODEs

- ... and also learns driving forces, and one resembles precipitation
- ... plus the time-decay constant of the ODE!



A- Hybrid neural networks

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Computers and Chemical Engineering 8 (2004) 7 5–766

Computers & Chemical Engineering

www.elsevier.com/locate/compchemeng

Combining first principles modelling and artificial neural networks: a general framework

R. Oliveira*

REQUIMTE/CQFB, Centro de Química Fina e Biotecnologia, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, P-2829-516 Caparica, Portugal

$$\frac{\mathrm{d}\boldsymbol{c}}{\mathrm{d}t} = \mathbf{KH}(\mathbf{c})\boldsymbol{\rho} - D\boldsymbol{c} + \boldsymbol{u} \quad (3a)$$

$$\boldsymbol{\rho} = \mathbf{N}(\mathbf{c}, \mathbf{W}) \quad (3b)$$

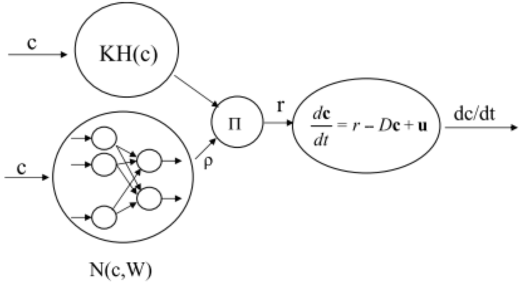
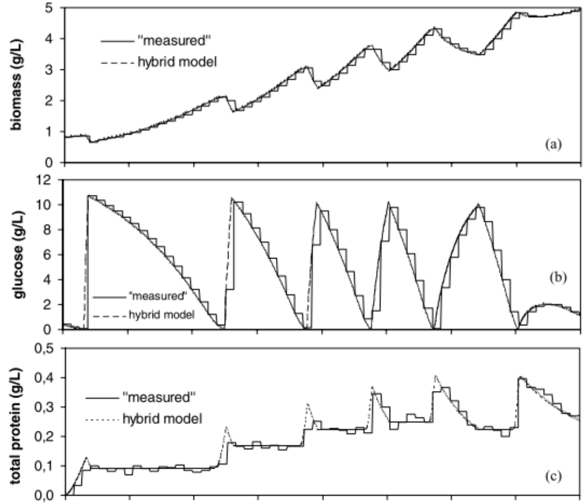


Fig. 1. General hybrid model structure for stirred tank bioreactors.



A- Hybrid neural networks

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]

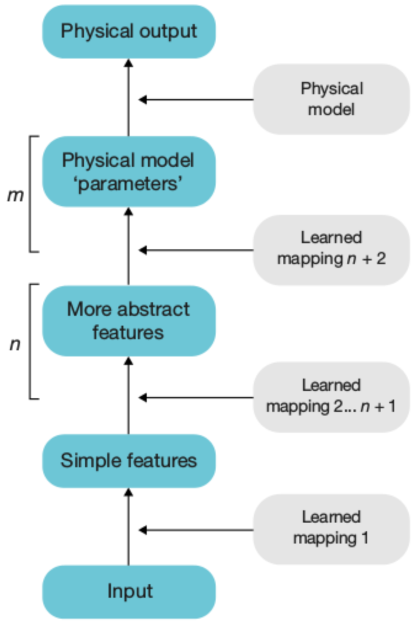
Machine learning approaches are increasingly used to extract patterns and insights from the ever-increasing stream of geospatial data, but current approaches may not be optimal when system behaviour is dominated by spatial or temporal context. Here, rather than amending classical machine learning, we 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.

A- Hybrid neural networks

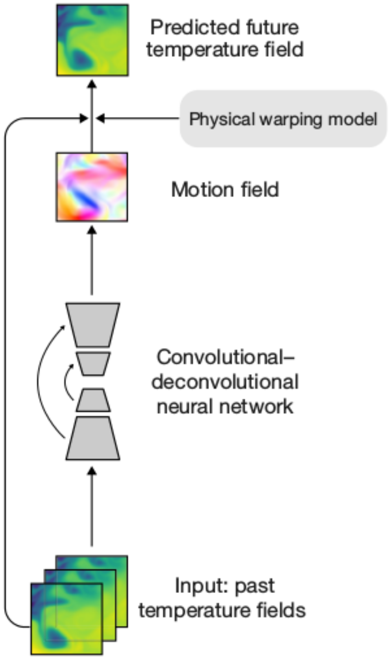
- ML that learns laws of physics (e.g. model-data consistency, mass and energy conservation)

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.

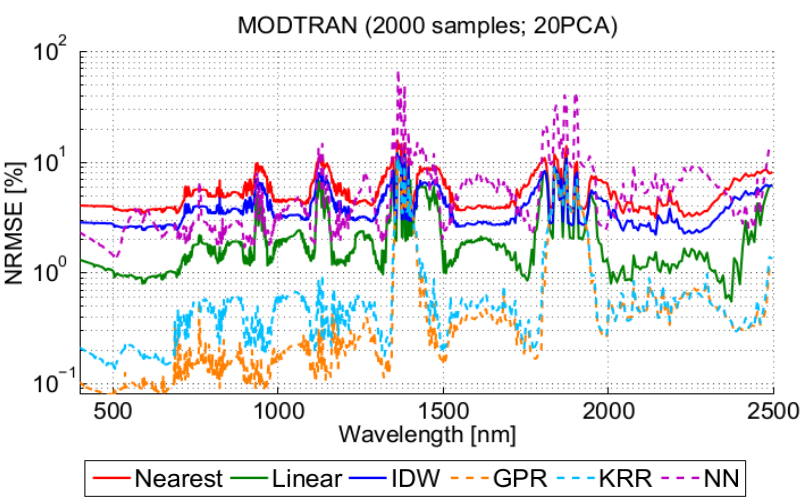
B: A motion field is learned with a convolutional-deconvolutional net, and the motion field is further processed with a physical model

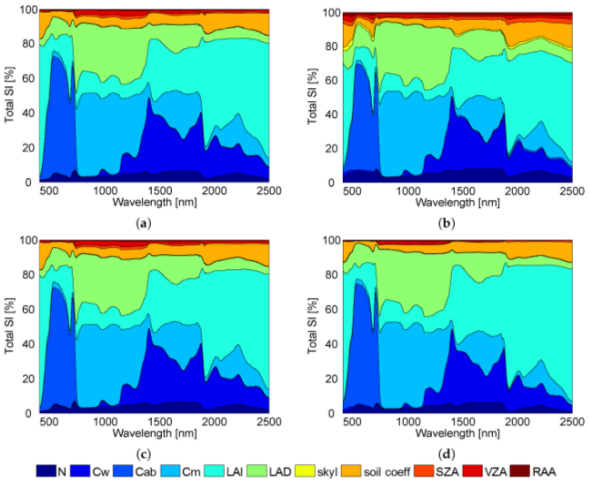


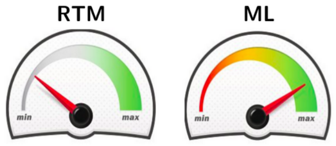
"Deep Learning for Physical Processes: Incorporating Prior Scientific Knowledge".
de Bezenac, Pajot, & Gallinari, arXiv:1711.07970 (2017).

B- Emulating complex codes

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





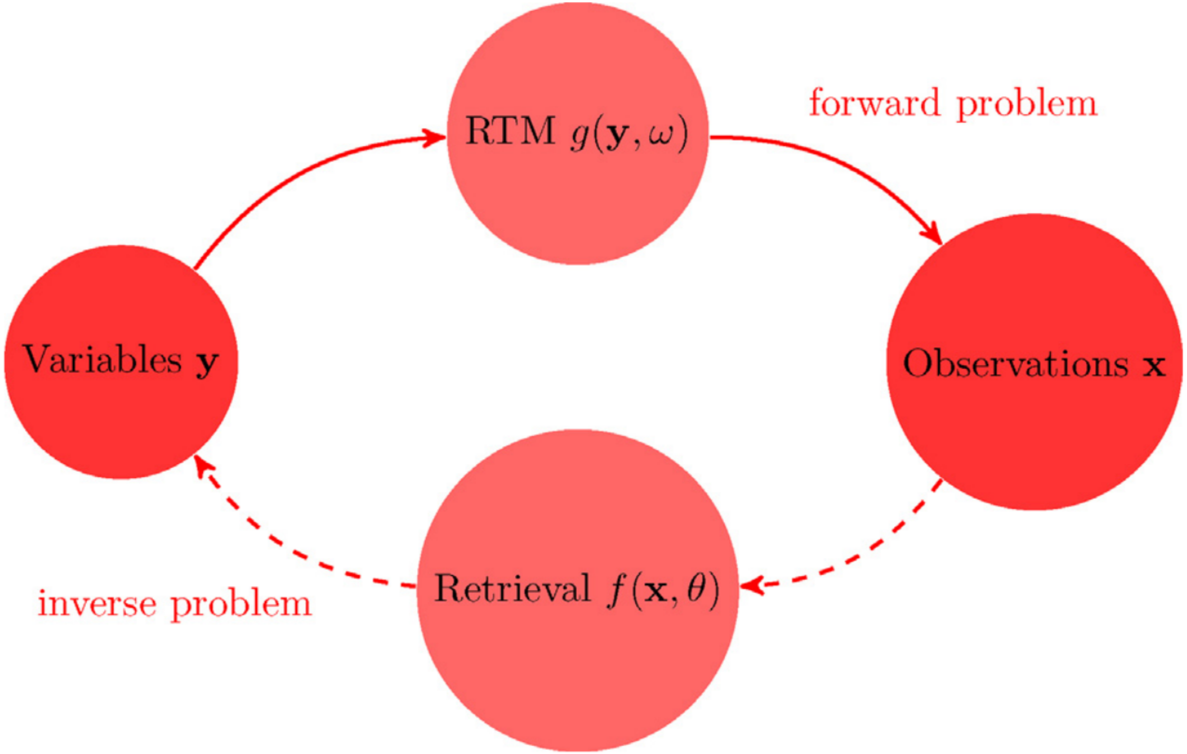


2.3–24.5 s/pix	0.1–1.3 ms/pix
0%	RMSE = 0.1 – 5%

"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

C- Parametrizations with variational inference

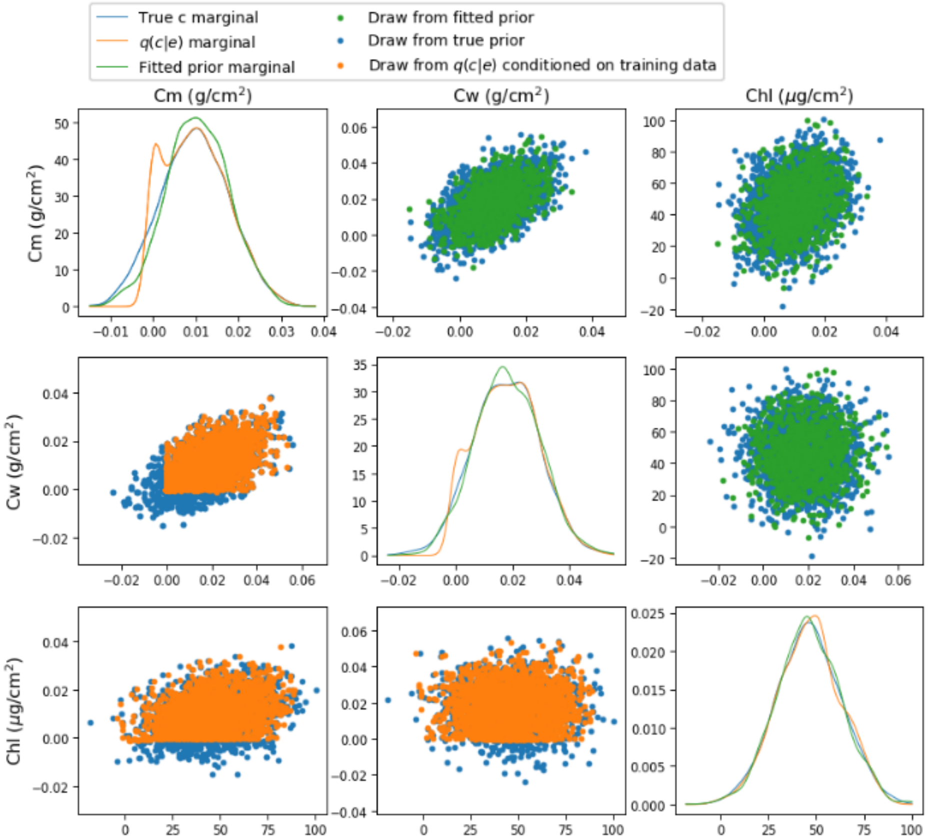


C- Parametrizations with variational autoencoders

- An RTM is a deterministic model mapping parameters ('causes',*c*) to radiances ('effects', *E*)
- Assume a Gaussian prior $P(c) = \mathcal{N}(\mu_\phi, \Sigma_\phi)$
- The evidence/marginal likelihood is hard to integrate w/ RTM inside the Gaussian mean!

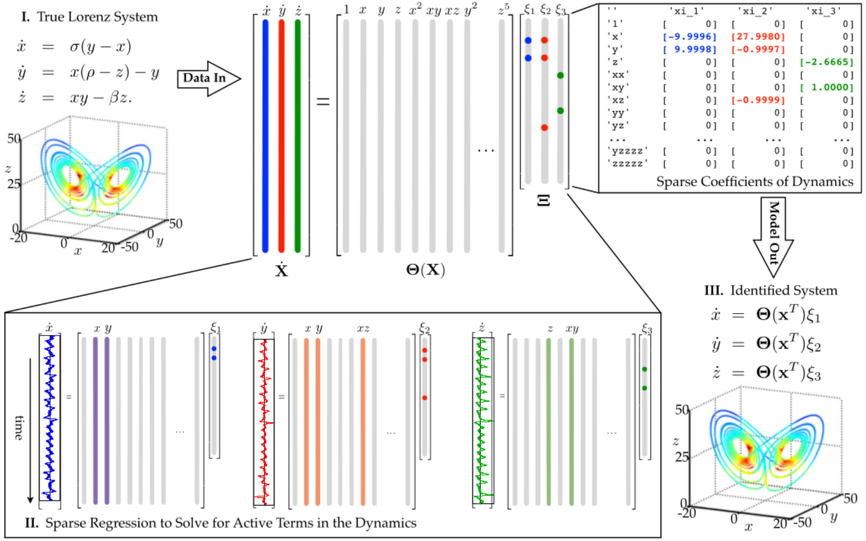
$$P(E|c) = \mathcal{N}(E|RTM(c), \sigma \mathrm{I})$$

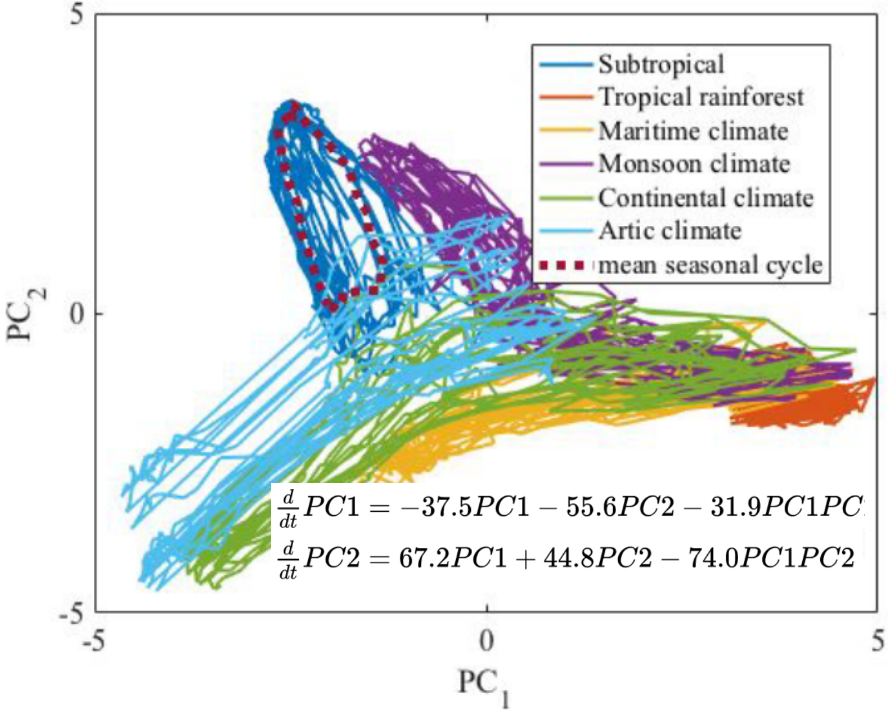
- VAE is orders of magnitude faster than MCMC, but problems with multimodal distributions



"Variational inference over radiative transfer model for biophysical parameter retrieval"
D. Svendsen, L. Martino, V. Laparra, G. Camps-Valls, *Machine Learning*, 2021

D- Discover ODEs from data





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

“**Discovering Differential Equations from Earth Observation Data**” Adsuara, J.E.; Camps-Valls, G.; Reichstein, M. and Mahecha, M. IGARSS 2020

Part V
XAI: Towards transparent models

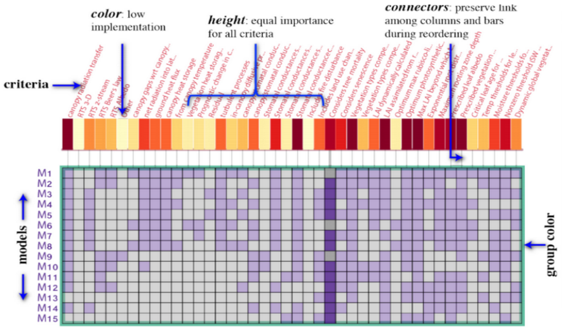
A taxonomy

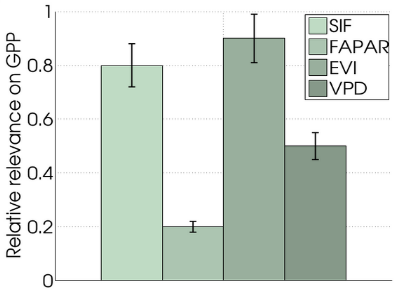
Methods	HSI	ANN	Mechansim		
CAM with global average pooling [41], [90]	✗	✓	Decomposition	Saliency	Perceptive Interpretability
+ Grad-CAM [42] generalizes CAM, utilizing gradient	✓	✓			
+ Guided Grad-CAM and Feature Occlusion [67]	✗	✓			
+ Respond CAM [43]	✗	✓			
+ Multi-layer CAM [91]	✗	✓			
LRP (Layer-wise Relevance Propagation) [13], [52]	✗	N.A.			
+ Image classifications. PASCAL VOC 2009 etc [44]	✗	✓			
+ Audio classification. AudioMNIST [46]	✗	✓			
+ LRP on DeepLight. fMRI data from Human Connectome Project [47]	✗	✓			
+ LRP on CNN and on BoW(bag of words)/SVM [48]	✗	✓			
+ LRP on compressed domain action recognition algorithm [49]	✗	✗			
+ LRP on video deep learning, *selective relevance method* [51]	✗	✓			
+ BiLRP [50]	✗	✓			
DeepLIFT [56]	✗	✓			
Prediction Difference Analysis [57]	✗	✓			
Slot Activation Vectors [40]	✗	✓			
PRM (Peak Response Mapping) [58]	✗	✓			
LIME (Local Interpretable Model-agnostic Explanations) [14]	✓	✓	Sensitivity		
+ MUSE with LIME [84]	✓	✓			
+ Guidelinebased Additive eXplanation optimizes complexity, similar to LIME [92]	✓	✓			
# Also listed elsewhere: [55], [68], [70], [93]	N.A.	N.A.			
Others. Also listed elsewhere: [94]	N.A.	N.A.	Others		
+ Direct output labels. Training NN via multiple instance learning [64]	✗	✓			
+ Image corruption and testing Region of Interest statistically [65]	✗	✓			
+ Attention map with autofocus convolutional layer [66]	✗	✓			
DeconvNet [71]	✗	✓	Inversion	Signal	
Inverting representation with natural image prior [72]	✗	✓			
Inversion using CNN [73]	✗	✓			
Guided backpropagation [74], [90]	✗	✓			
Activation maximization/optimization [37]	✗	✓	Optimization		
+ Activation maximization on DBN (Deep Belief Network) [75]	✗	✓			
+ Activation maximization, multifaceted feature visualization [76]	✗	✓			
Visualization via regularized optimization [77]	✗	✓			
Semantic dictionary [38]	✗	✓			
Decision trees	N.A.	N.A.	Verbal		
Propositional logic, rule-based [81]	✗	✗			
Sparse decision list [82]	✗	✗			
Decision sets, rule sets [83], [84]	✓	✗			
Encoder-generator framework [85]	✗	✓			
Filter Attribute Probability Density Function [86]	✗	✗			
MUSE (Model Understanding through Subspace Explanations) [84]	✓	✓			

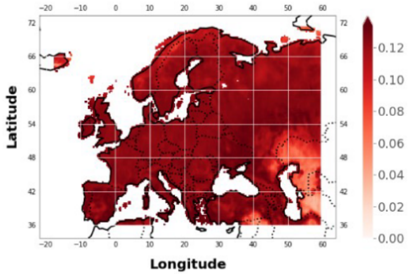
“A Survey on Explainable Artificial Intelligence(XAI): towards Medical XAI”, Tjoa 2019

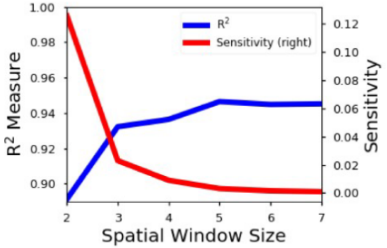
“Advancing Deep Learning For Earth Sciences: From Hybrid Modeling To Interpretability”, Camps-Valls, G. and Reichstein, M. and Zhu, Z. and Tuia, D. IEEE IGARSS (2020)

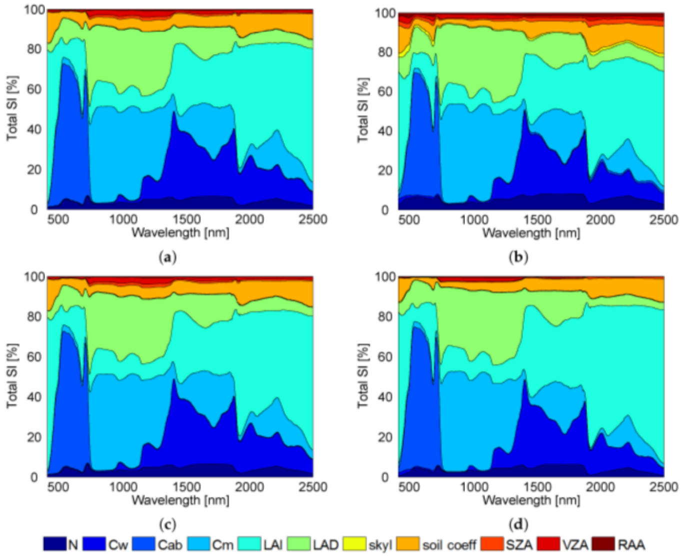
1: Sensitivity analysis











"Remote Sensing Feature Selection by Kernel Dependence Estimation", Camps-Valls, G. Mooij, JM. Schölkopf, IEEE-GRSL, 2010.

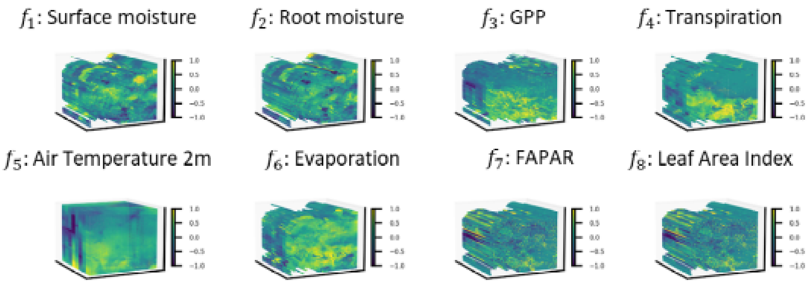
"A guided hybrid genetic algorithm for feature selection with expensive cost functions", M. Jung, J. Zscheischler, Procedia, 2013.

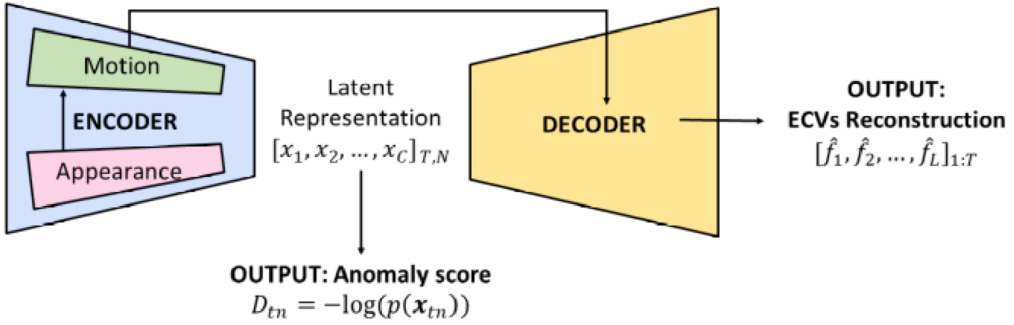
"Kernel methods and their derivatives: Concept and perspectives for the Earth system sciences" Johnson, JE, Laparra, V, Perez, A, Mahecha, M., Camps-Valls, G. PLOS ONE 2020

Detect, anticipate, understand climate extremes

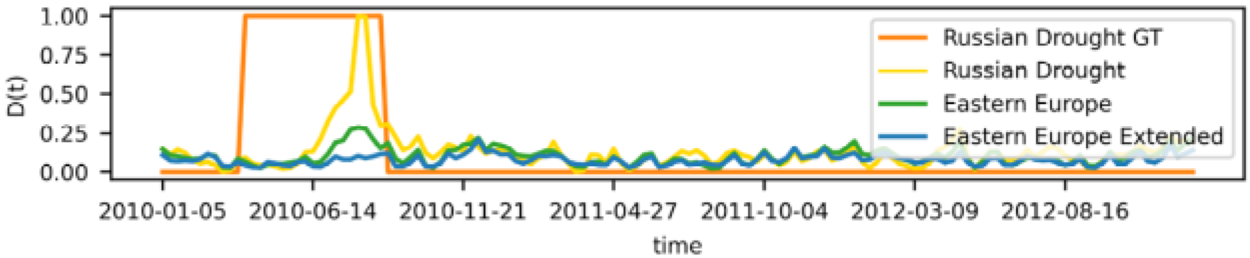
«play video 2»

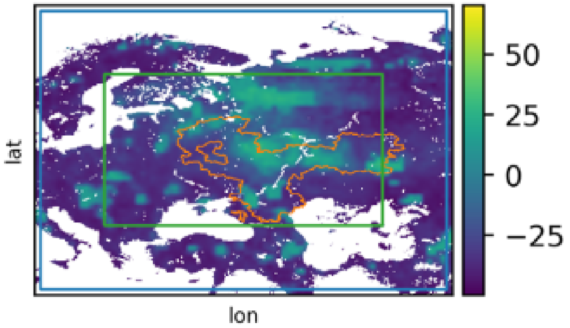
1: Sensitivity analysis





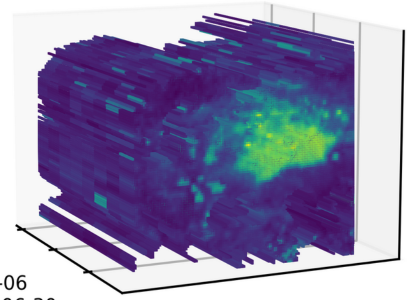
Spatio-temporal drought detection: Russian heat wave in 2010

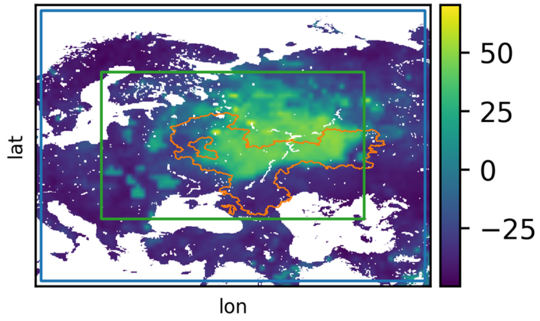


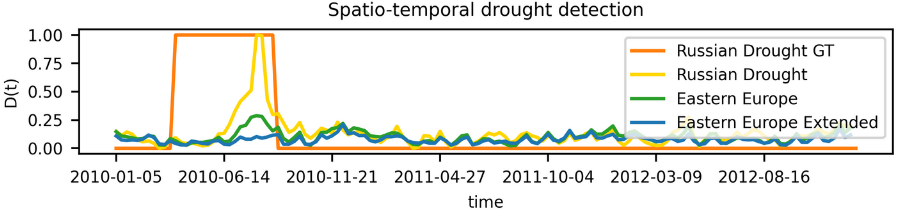


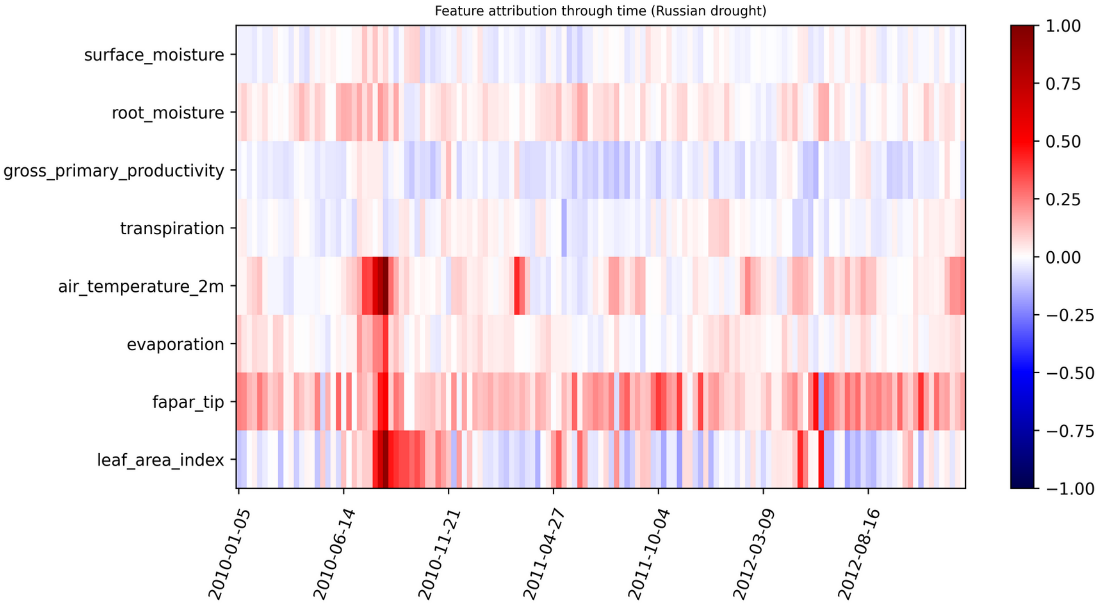
"Spatio-Temporal Gaussianization Flows for Extreme Event Detection". Miguel-Angel Fernández-Torres and J. Emmanuel Johnson and María Piles and Gustau Camps-Valls
EGU General Assembly, Geophysical Research Abstracts, Online, 19-30 April 2021

1: Sensitivity analysis





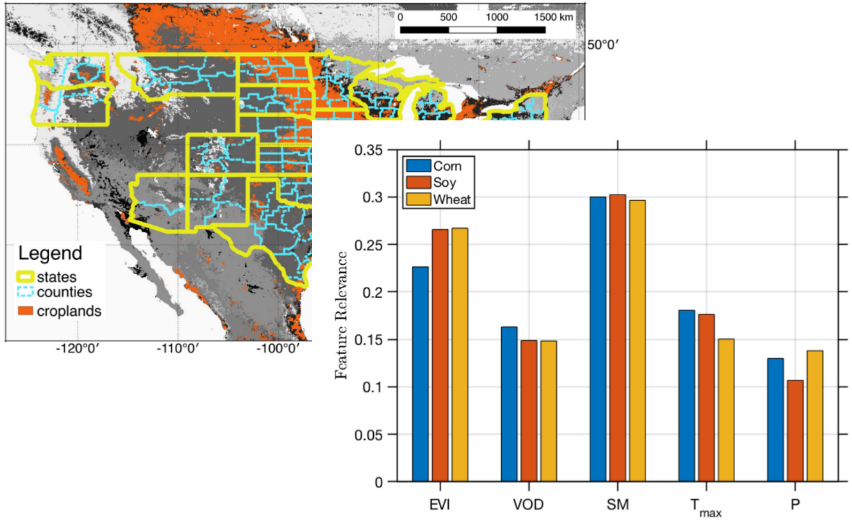


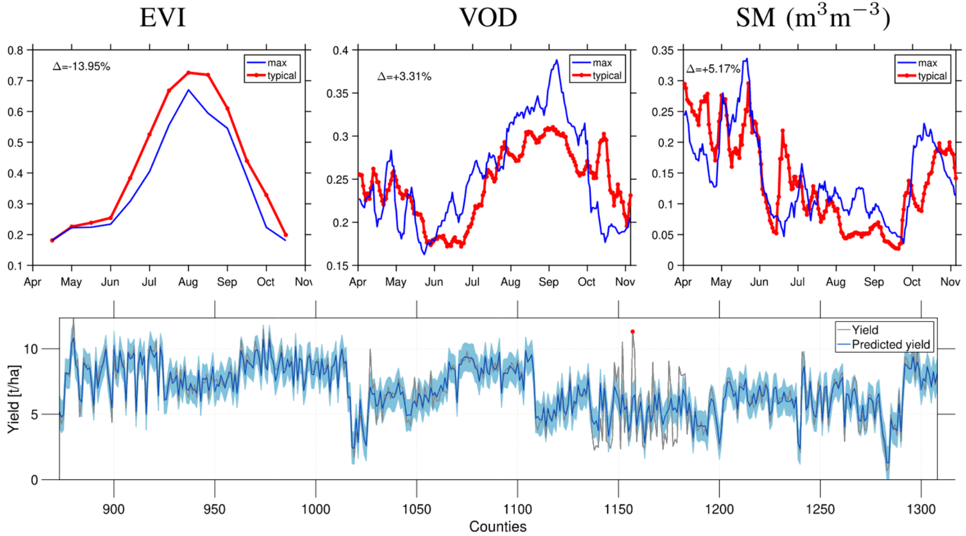


"Spatio-Temporal Gaussianization Flows for Extreme Event Detection". Miguel-Ángel Fernández-Torres and J. Emmanuel Johnson and María Piles and Gustau Camps-Valls
EGU General Assembly, Geophysical Research Abstracts, Online, 19-30 April 2021

2: Model decomposition

- **Gaussian processes** $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^\top \mathbf{x}_j + \nu \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{x}_i - \mathbf{x}_j\|^2\right) + \sigma_n^2 \delta_{ij}$

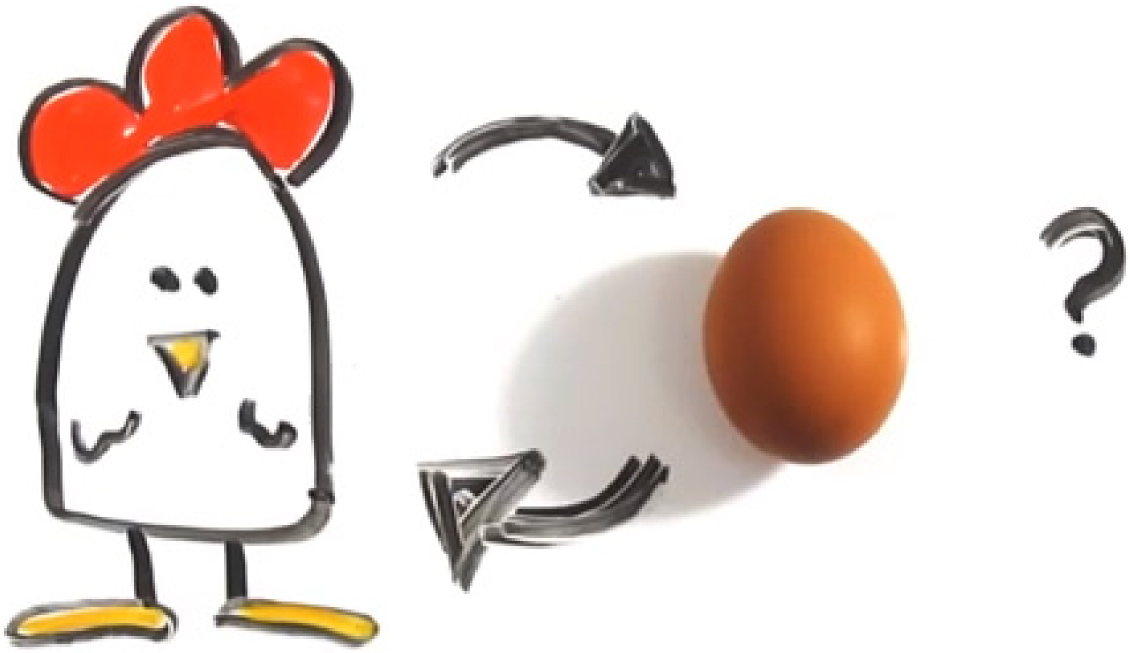




"Estimating and Understanding Crop Yields with Explainable Deep Learning in the Indian Wheat Belt". Wolanin, Mateo-Garcia, Camps-Valls, Gomez-Chova, et al. Environmental Research Letters, 2020
"Crop Yield Estimation and Understanding with Gaussian Processes" Martinez-Ferrer, L, Piles, M. and Camps-Valls, G. IEEE Geoscience and Remote Sensing Letters, 2020

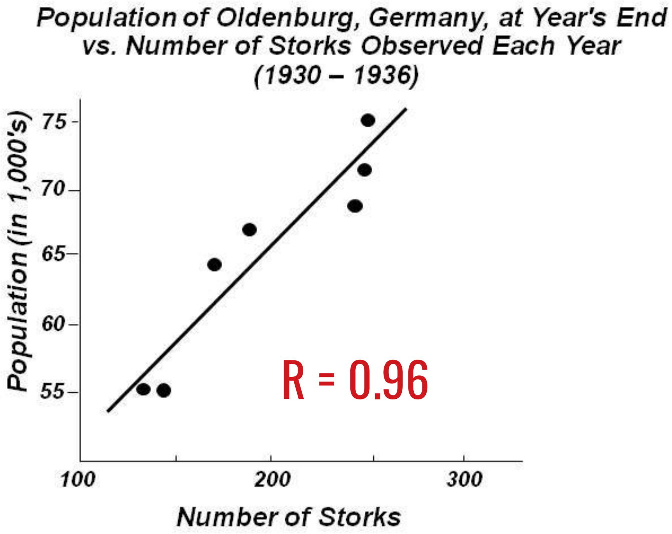
Part VI
All is about (pragmatic) causality

Causal inference



Correlation is not causation, but then what’s causality?

- Correlation is not causation!
- Correlation: you can be right for the wrong reasons!



Source: Statistics for Experimenters, by Box, Hunter & Hunter

Storks Deliver Babies ($p = 0.008$)

KEYWORDS:
Teaching;
Correlation;
Significance;
p-values.

Robert Matthews
Aston University, Birmingham, England.
e-mail: rajm@compuserve.com

Summary
This article shows that a highly statistically significant correlation exists between stork populations and human birth rates across Europe. While storks may not deliver babies, unthinking interpretation of correlation and *p*-values can certainly deliver unreliable conclusions.

◆ INTRODUCTION ◆

Introductory statistics textbooks routinely warn of the dangers of confusing correlation with causation, pointing out that while a high correlation coefficient is indicative of (linear) association, it cannot be taken as a measure of causation. Such warnings are typically accompanied by illustrative examples, such as the correlation between the reading skills of children and their shoe size, or the apparent relationship between educational level and unemployment (see e.g. Freedman *et al.* 1998). However, such examples are often either trivially explained via an obvious confounder (e.g. age, in the case of reading age and shoe size) or are not obviously cases of mere association (e.g. educational level may indeed be at least partly responsible for time spent unemployed). In what follows, I give an example based on genuine data of an association which is clearly ludicrous, but which cannot be so easily dismissed as non-causal via an obvious confounder.

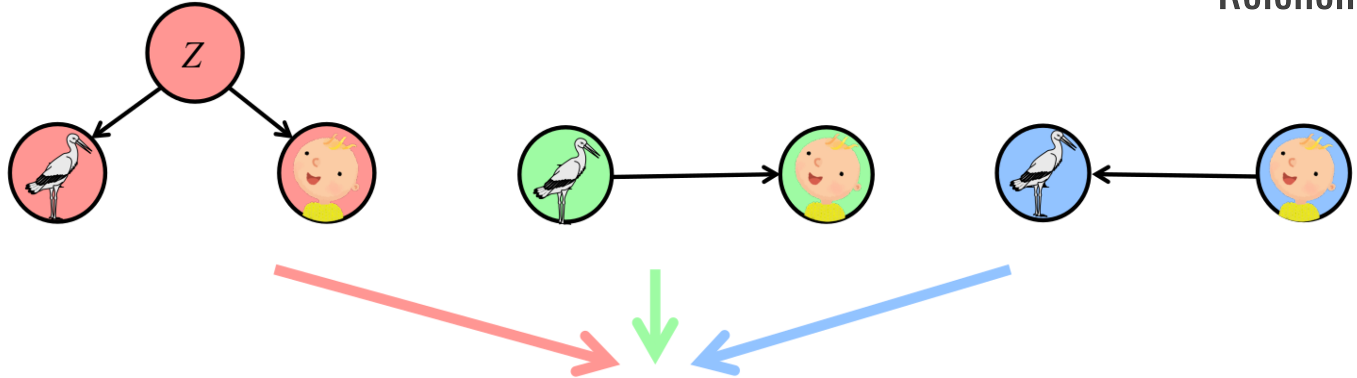
My starting point is the familiar folk tale that babies are delivered by storks. The origins of this connection are believed to lie partly in the association between storks and the concept of women as bringers of life, and also in the bird's feeding habits, which were once regarded as a search for embryonic life in water (Cooper 1992). The legend lives on to this day, with neonate-bearing storks being a regular feature of greetings cards celebrating births.

While it is (I trust) obvious that the legend is complete nonsense, it is legitimate to ask precisely how one might set about refuting it scientifically. If one were approaching the question in the same way that many other links are investigated (e.g. suspected links between diet and cancer risk), one may well decide to carry out a correlational study, to see if the number of storks in a country bears a simple relationship to the number of human births in that country. Although the presence of a statistically significant degree of correlation cannot be taken to imply causation, its absence would certainly constitute evidence against a simple relationship. This possibility can quickly be investigated in the present case using standard hypothesis testing, with the null hypothesis being the absence of any correlation between the number of storks and the number of live births in a particular country. This I now proceed to do.

Common Cause Principle

- If X and Y are statistically dependent, then there exists Z causally influencing both
- Z screens X and Y from each other: X and Y become independent given Z

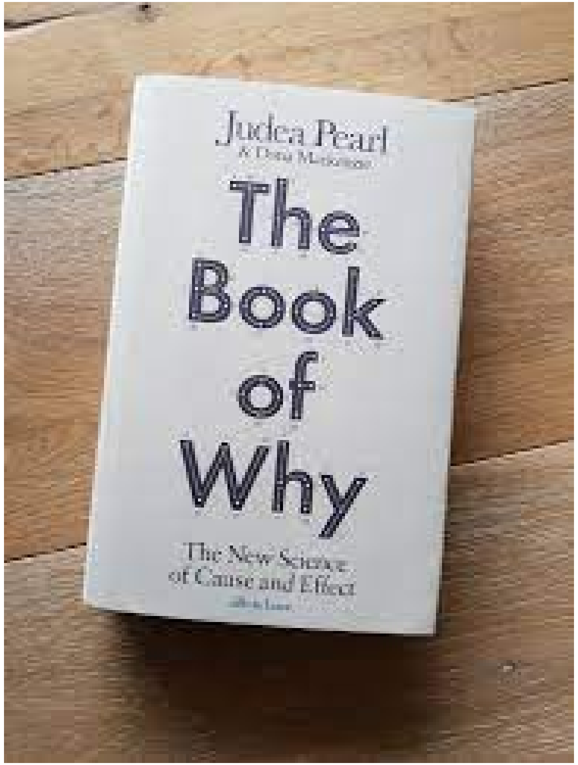
-- Reichenbach, 1956

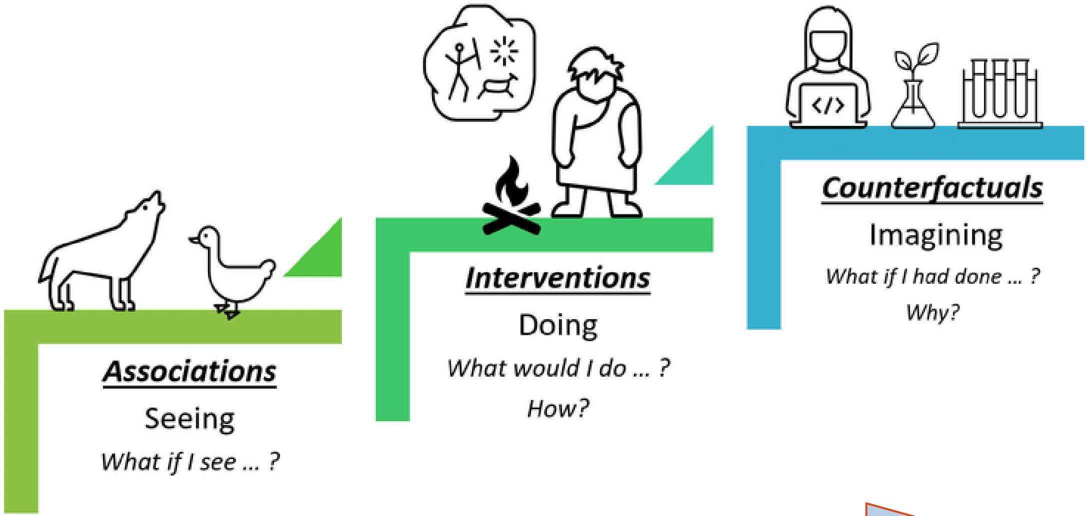


$$p(X,Y)$$

$\sum_z p(x|z)p(y|z)p(z)$ $\quad$ $p(x)p(y|x)$ $\quad$ $p(x|y)p(y)$

The rungs of inference …

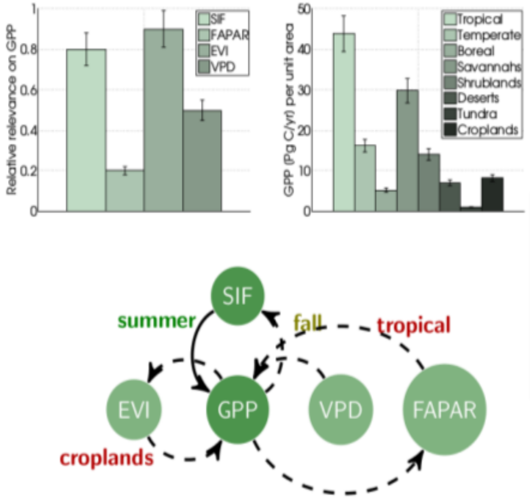


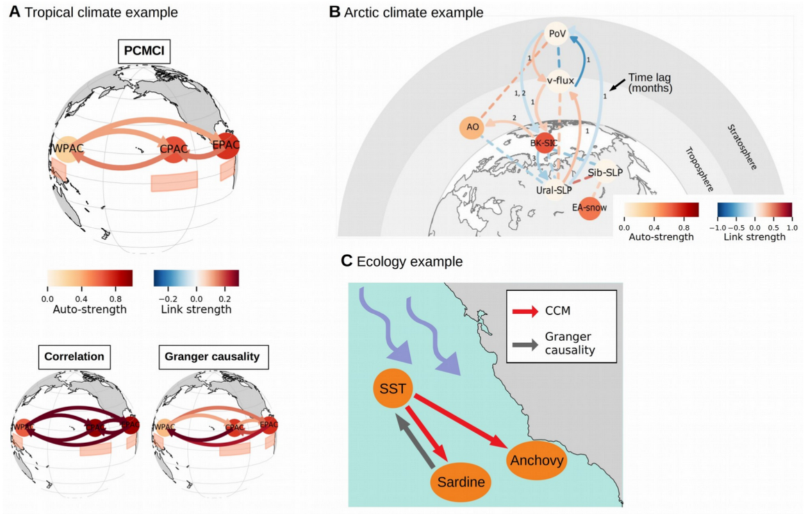


Deep learning world
Interaction world
Scenarios world

Causal inference in Earth and climate sciences

- **Causality 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 (2019). **CauseMe: http://causeme.uv.es**

Methods for causal discovery from data

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]

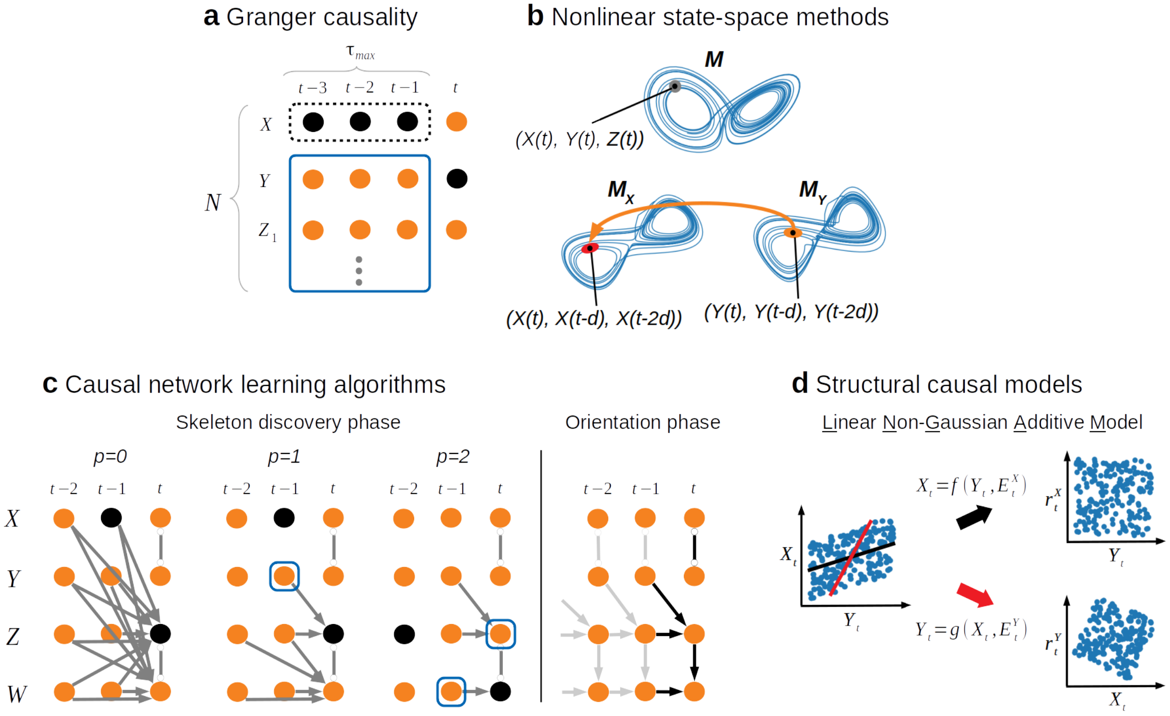
IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING 1

Causal Inference in Geoscience and Remote Sensing From Observational Data

Adrián Pérez-Suay, Member, IEEE, and Gustau Camps-Valls, Fellow, IEEE

Abstract—Establishing causal relations between random variables from observational data is perhaps the most important challenge in today's science. In remote sensing and geosciences, this is of special relevance to better understand the earth's system and the complex interactions between the governing processes.

with societal, economical, and environmental challenges, such as climate change [2], [3]. There is an urgent need for tools that help us to observe and study the earth system. Nowadays, machine learning and signal processing play a crucial role in

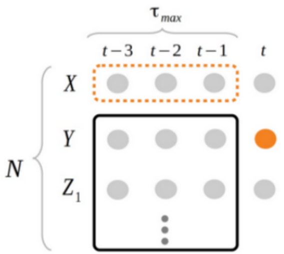


"**Inferring causation from time series with perspectives in Earth system sciences**", Runge, Bathiany, Bollt, Camps-Valls, et al. Nat Comm., 2019

"**Causal Inference in Geoscience and Remote Sensing from Observational Data,**" Pérez-Suay and Camps-Valls, IEEE Trans. Geosc. Rem. Sens, 2018

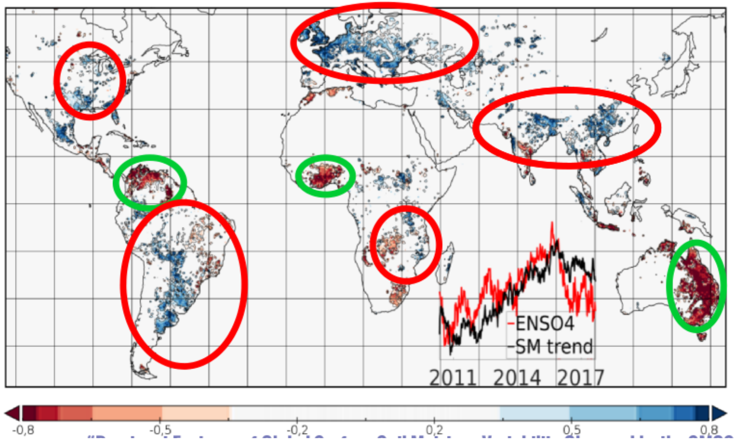
"**CauseMe: An online system for benchmarking causal inference methods,**" Muñoz-Marí, Mateo, Runge, Camps-Valls. In preparation (2019). **CauseMe: http://causeme.uv.es**

1- Nonlinear Nonstationary Granger Causality (XKGC)

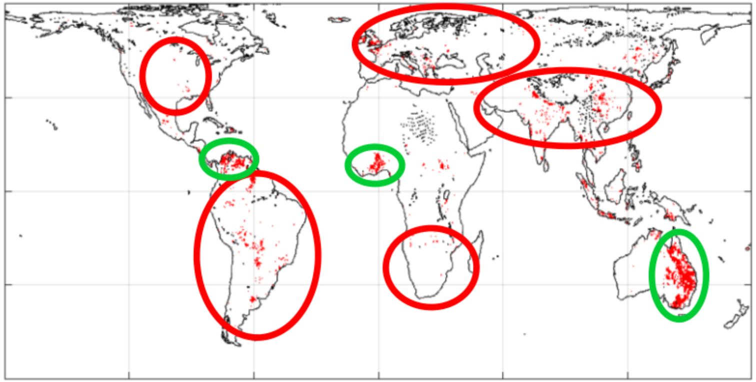


$$\boxed{\begin{aligned} Y_{t+1} &= a^\top X_t + \varepsilon_t^Y \\ Y_{t+1} &= b_1^\top Y_t + b_2^\top X_t + \varepsilon_t^{Y|X} \\ X \rightarrow Y &\leftrightarrow \mathbb{V}[\varepsilon_t^Y] \ll \mathbb{V}[\varepsilon_t^{Y|X}] \end{aligned}}$$

$$\boxed{\begin{aligned} a_H &= (K(X_t, X_t') + \varepsilon_t^Y)^{-1} Y_{t+1} \\ b_H &= (L([Y_t, X_t], [Y_t', X_t']) + \varepsilon_t^{Y|X})^{-1} Y_{t+1} \\ X \rightarrow Y &\leftrightarrow \mathbb{V}_H[\varepsilon_t^Y] \ll \mathbb{V}_H[\varepsilon_t^{Y|X}] \end{aligned}}$$



ENSO4 $\rightarrow$ SM



- Causality is sharper than correlation
- ENSO4 “causes” SM in very dry (Sahel) and very wet (tropical rain forests)

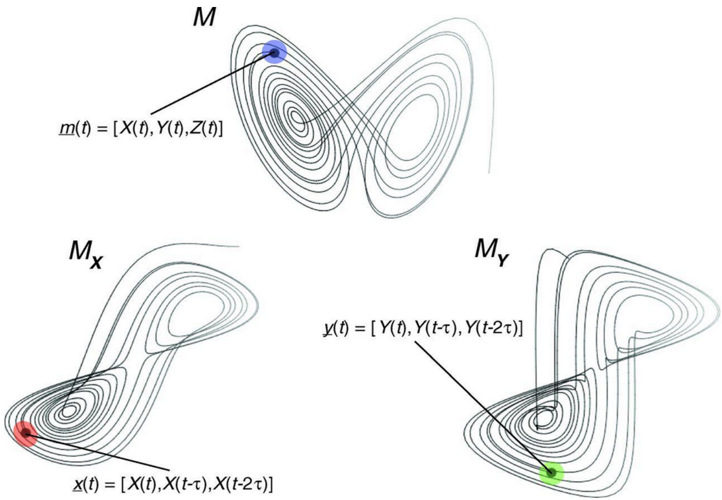
“Inferring causation from time series with perspectives in Earth system sciences”, Runge, Bathiany, Bollt, Camps-Valls, et al. Nat Comm., 2019

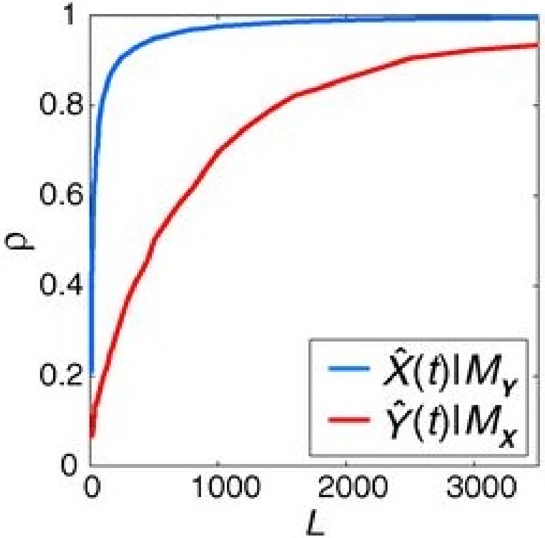
“Causal Inference in Geoscience and Remote Sensing from Observational Data,” Pérez-Suay and Camps-Valls, IEEE Trans. Geosc. Rem. Sens, 2018

“Explicit Granger Causality in Kernel Hilbert Spaces” Diego Bueso, Maria Piles, Gustau Camps-Valls, Physical Review E 102 :062201, 2020

2- Robust Convergent Cross Mapping (RCCM)

- X→Y if you can recover manifold structure Y from manifold structure X
- RCCM: bootstrap resampling for robust estimation + IGCI correction for entropy asymmetry
- Results in carbon and water fluxes

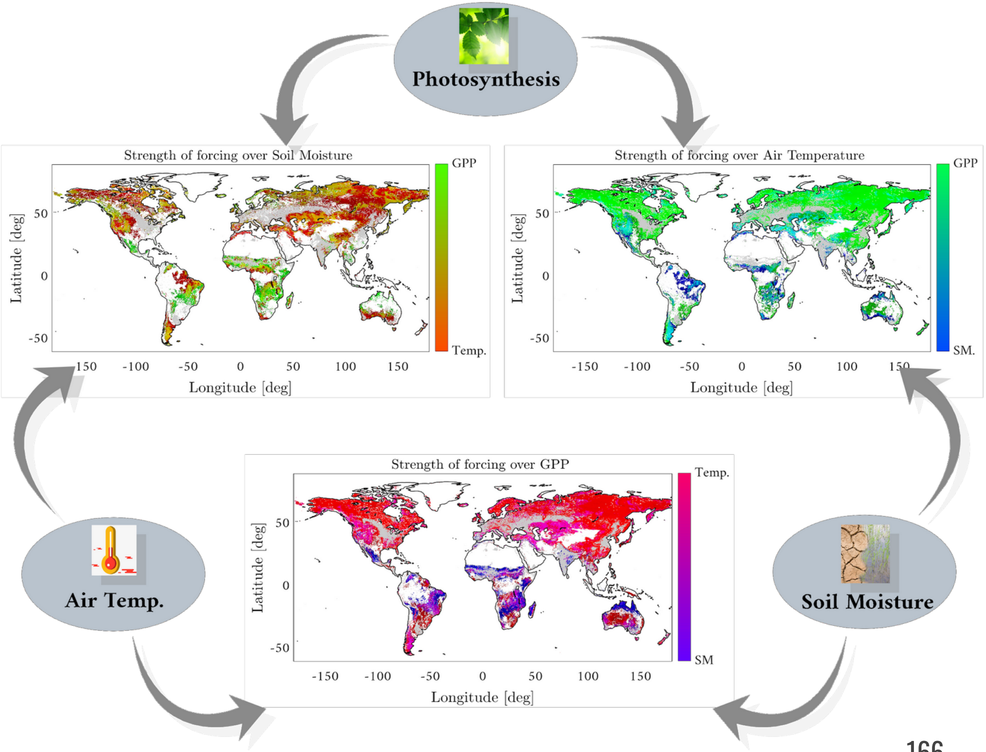




"Inferring causal relations from observational long-term carbon and water fluxes records" E. Diaz, G. Camps-Valls et al. Scientific Reports, 2022

2- Robust Convergent Cross Mapping (RCCM)

- Causality on (GPP, Tair, SM)
- Causal maps capture common 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



"Inferring causal relations from observational long-term carbon and water fluxes records" E. Diaz, G. Camps-Valls et al. Scientific Reports, 2022

3- PC with momentary conditional independence (PCMCI)

- Smart fast algorithm to test conditional independence and decide causal arrows
- Test correlation of residuals-cause in both directions

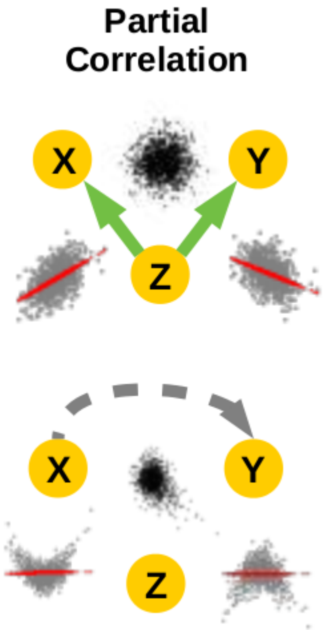
Partial correlation test of $X \perp\!\!\!\perp Y \mid \mathbf{Z}$:

1. Assuming linear model

$$X = \mathbf{Z}\beta_X + \epsilon_X$$

$$Y = \mathbf{Z}\beta_Y + \epsilon_Y$$

2. Test correlation of residuals $\rho(r^X, r^Y)$



“**Causal network reconstruction from time series**” Runge, J. Chaos, 2018

3- PC with momentary conditional independence (PCMCI)

- Smart fast algorithm to test conditional independence and decide causal arrows
- Test correlation of residuals-cause in both directions

Gaussian process + distance correlation test:

1. Assuming nonlinear additive Gaussian

$$X = f_X(\mathbf{Z}) + \epsilon_X$$

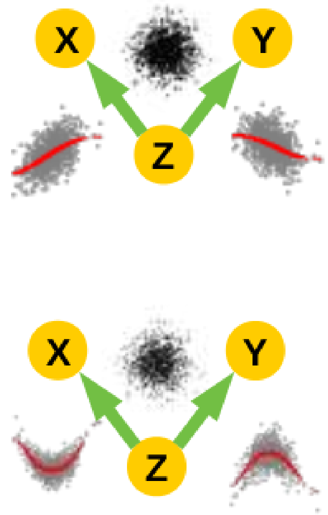
$$Y = f_Y(\mathbf{Z}) + \epsilon_Y$$

$$\epsilon_{\cdot} \sim \mathcal{N}(0, \sigma^2)$$

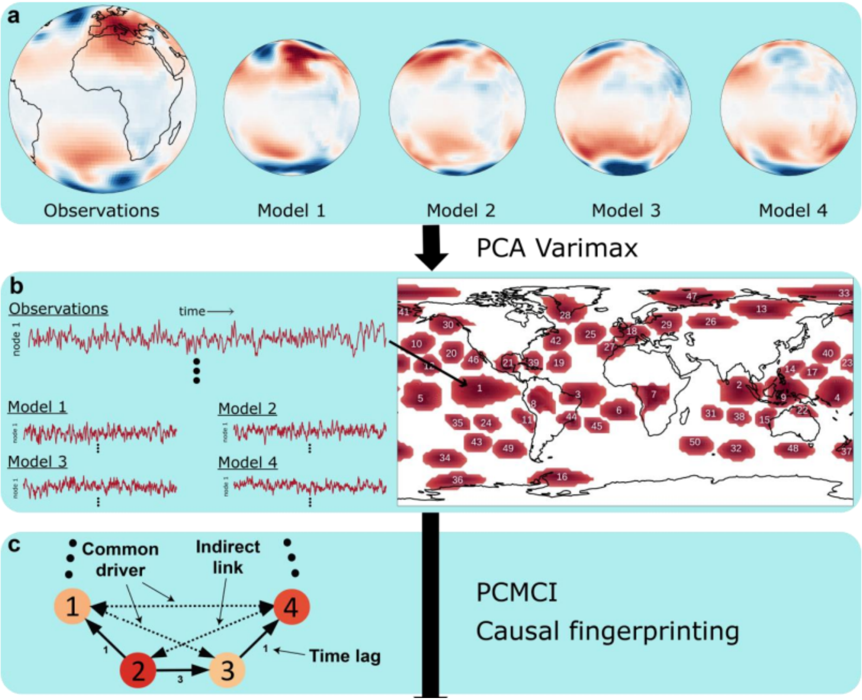
$$f_{\cdot}(\mathbf{Z}) \sim \mathcal{GP}(0, k(\mathbf{Z}, \mathbf{Z}'))$$

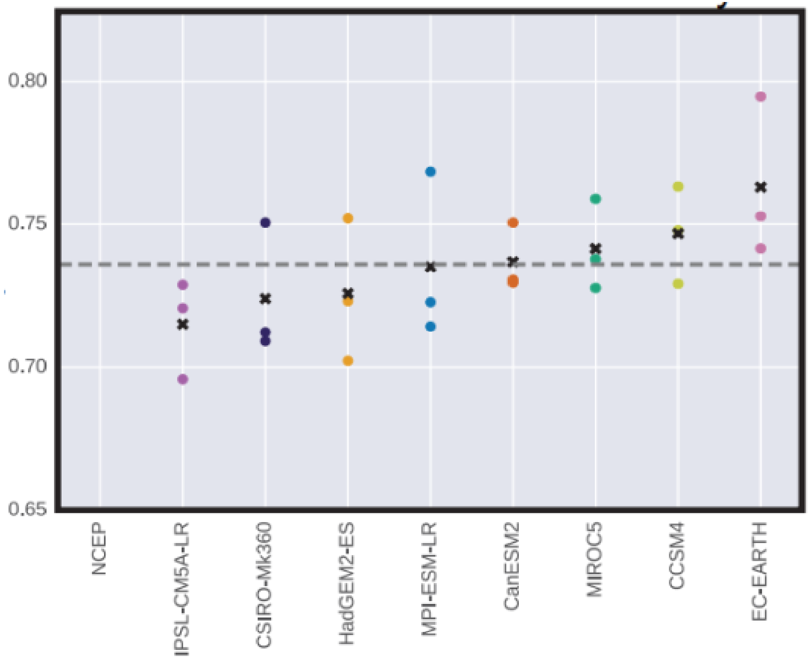
2. Test independence of residuals with *distance correlation coefficient*

Gaussian Process
Distance Correlation



3- PC with momentary conditional independence (PCMCI)





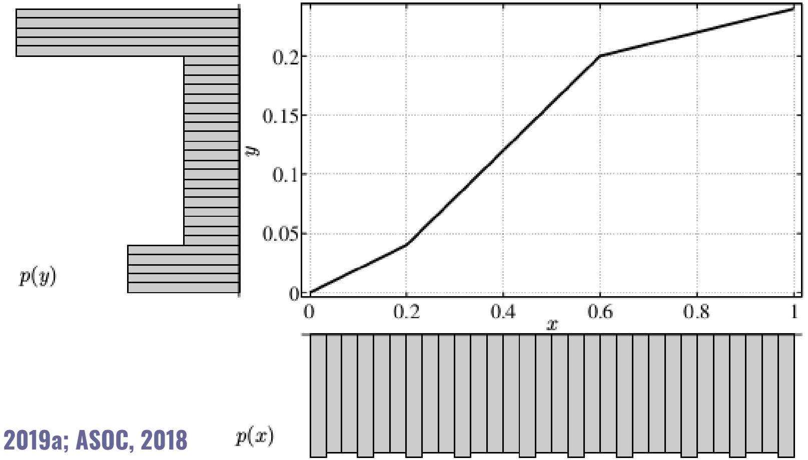
"**Causal networks for climate model evaluation and constrained projections**". Nowack, Runge, Eyring, Haigh, Nature Comm, 2020

4- Structural causal models

- Causality with two variables only!?

$$p(x, y) = p(y|x)p(x) = p(x|y)p(y)$$

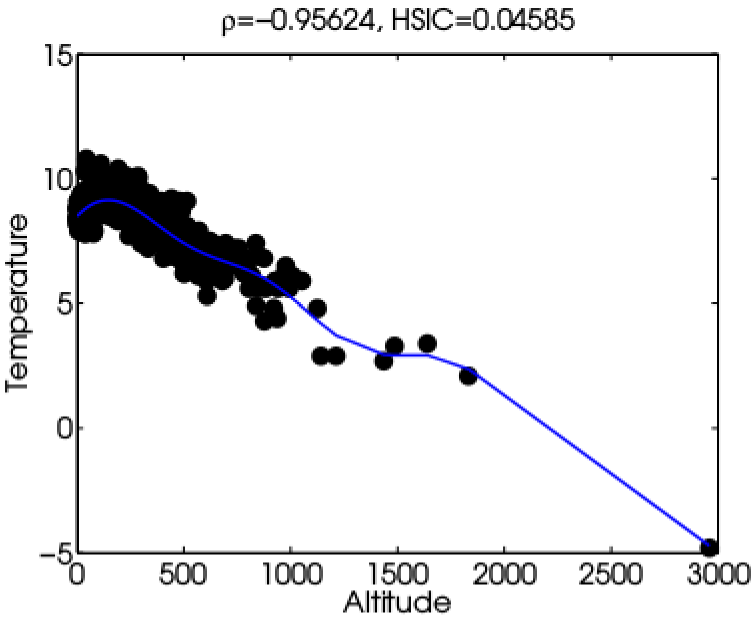
- <u>Idea</u>: Cause should be independent of the generating mechanism

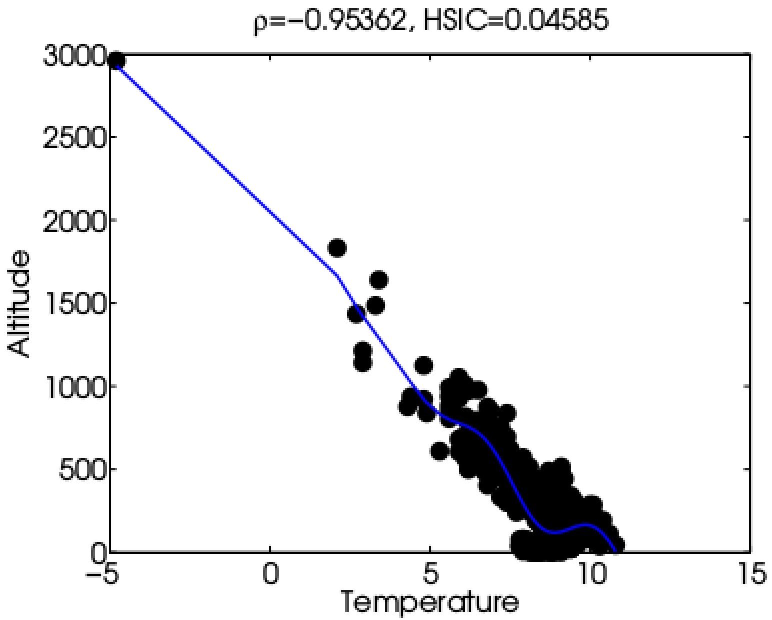


Mitrovic et al, 2018; Perez & Camps-Valls, TGARS, 2019a; ASOC, 2018

4- Structural causal models - example

- Step 1: fit a forward and an inverse nonlinear regression

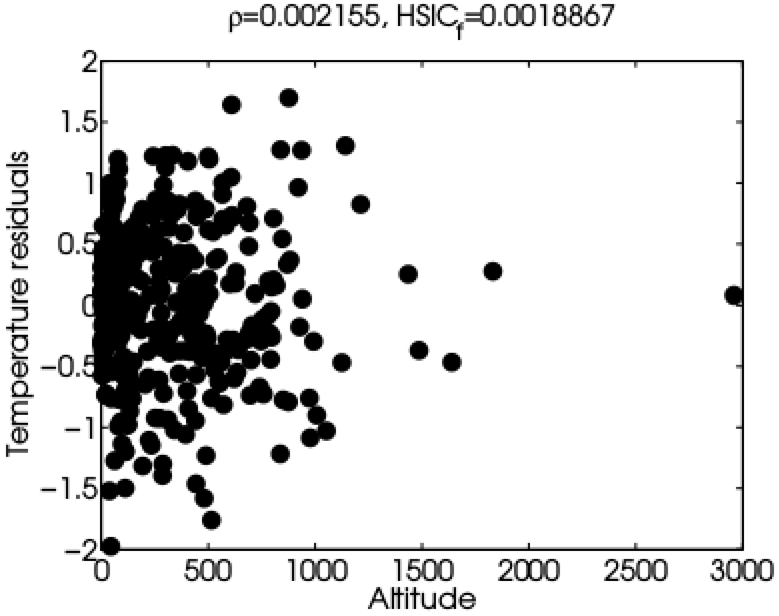


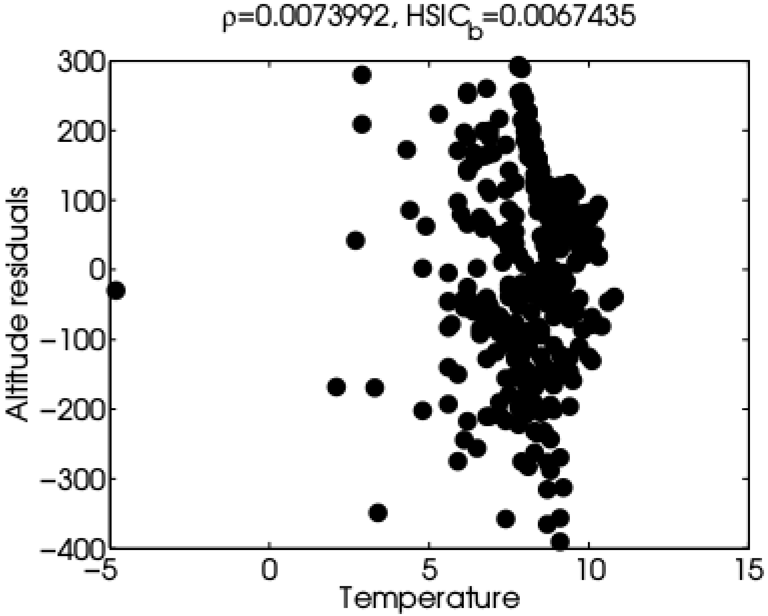


Mitrovic et al, 2018; Perez & Camps-Valls, TGARS, 2019a; ASOC, 2018; Hoyer et al., NIPS 2008.

4- Structural causal models - example

- Step 2: test independence between residuals and the assumed cause

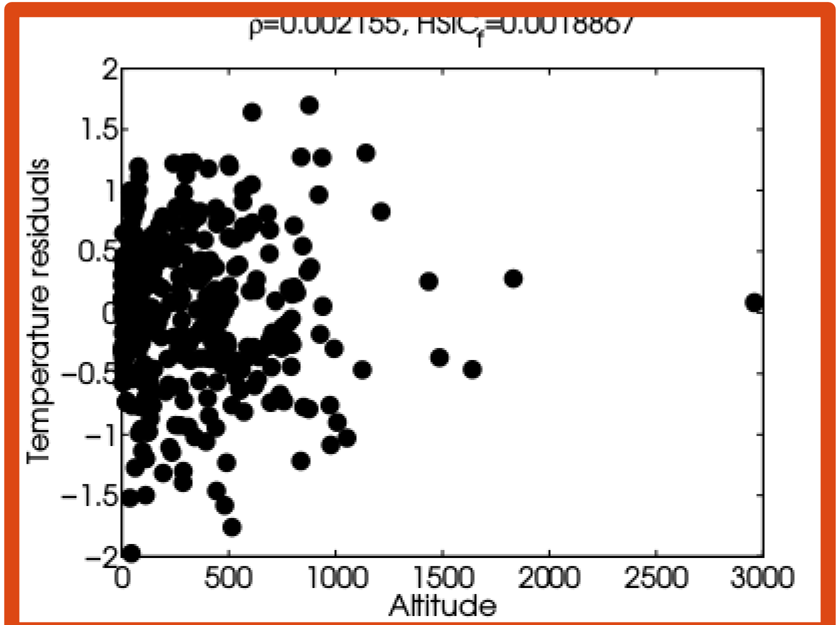


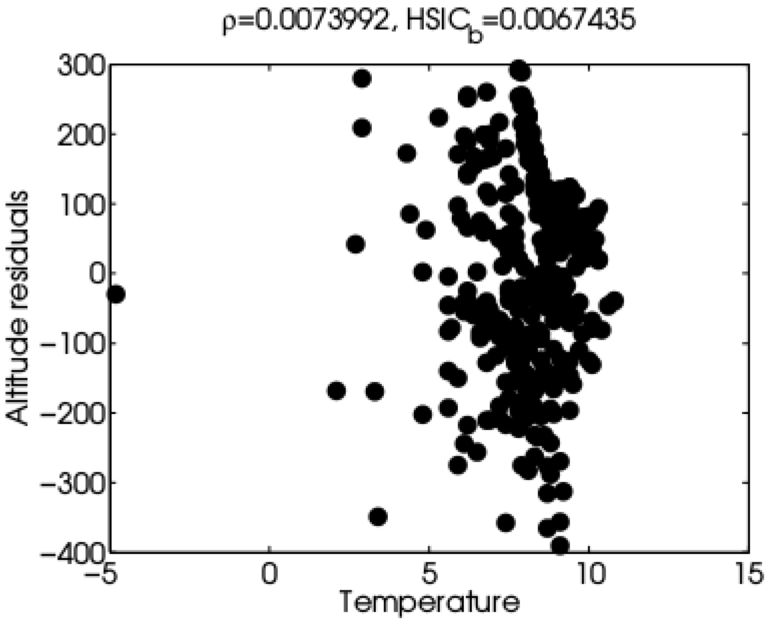


Mitrovic et al, 2018; Perez & Camps-Valls, TGARS, 2019a; ASOC, 2018; Hoyer et al., NIPS 2008.

4- Structural causal models - example

- Step 3: the direction of causation is the most independent

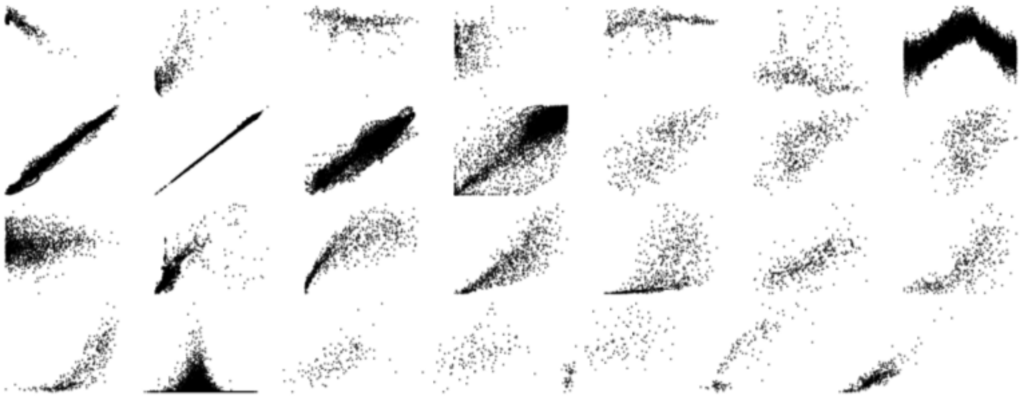




Mitrovic et al, 2018; Perez & Camps-Valls, TGARS, 2019a; ASOC, 2018; Hoyer et al., NIPS 2008.

4- Structural causal models

- Excellent preliminary results in synthetic examples, model emulation, and RTMs

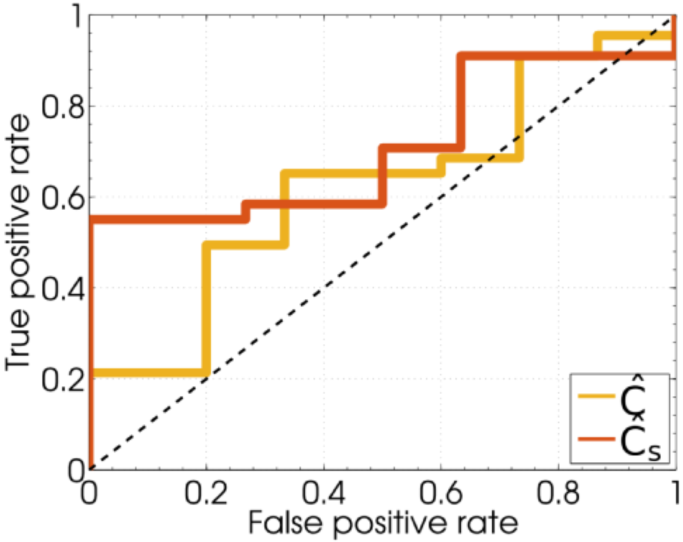


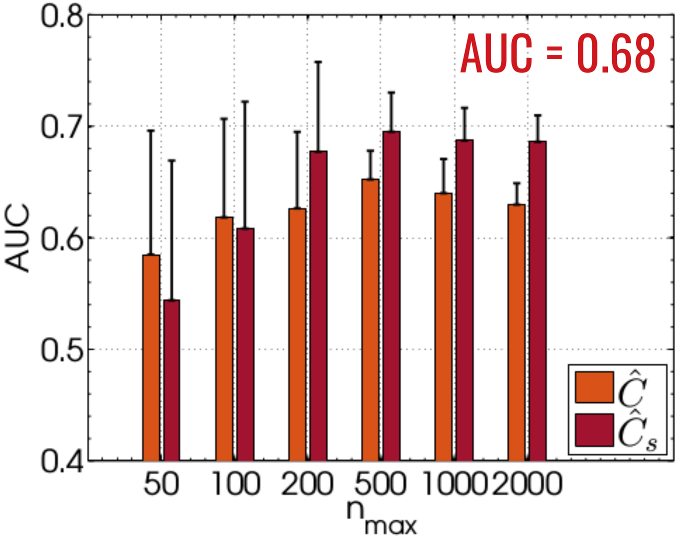
id	x	y	Cause
pair0001	Altitude	Temperature	$\rightarrow$
pair0002	Altitude	Precipitation	$\rightarrow$
pair0003	Longitude	Temperature	$\rightarrow$
pair0004	Altitude	Sunshine hours	$\rightarrow$
pair0020	Latitude	Temperature	$\rightarrow$
pair0021	Longitude	Precipitation	$\rightarrow$
pair0042	Day of the year	Temperature	$\rightarrow$
pair0043	Temperature at t	Temperature at t+1	$\rightarrow$
pair0044	Pressure at t	Pressure at t+1	$\rightarrow$
pair0045	Sea level pressure at t	Sea level pressure at t+1	$\rightarrow$
pair0046	Relative humidity at t	Relative humidity at t+1	$\rightarrow$
pair0049	Ozone concentration	Temperature	$\leftarrow$
pair0050	Ozone concentration	Temperature	$\leftarrow$
pair0051	Ozone concentration	Temperature	$\leftarrow$
pair0072	Sunspots	Global mean temperature	$\rightarrow$

Mitrovic et al, 2018; Perez & Camps-Valls, TGARS, 2019a; ASOC, 2018; Hoyer et al., NIPS 2008.

4- Structural causal models

- Excellent preliminary results in synthetic examples, model emulation, and RTMs





Mitrovic et al, 2018; Perez & Camps-Valls, TGARS, 2019a; ASOC, 2018; Hoyer et al., NIPS 2008.

5- Learning causal feature representations

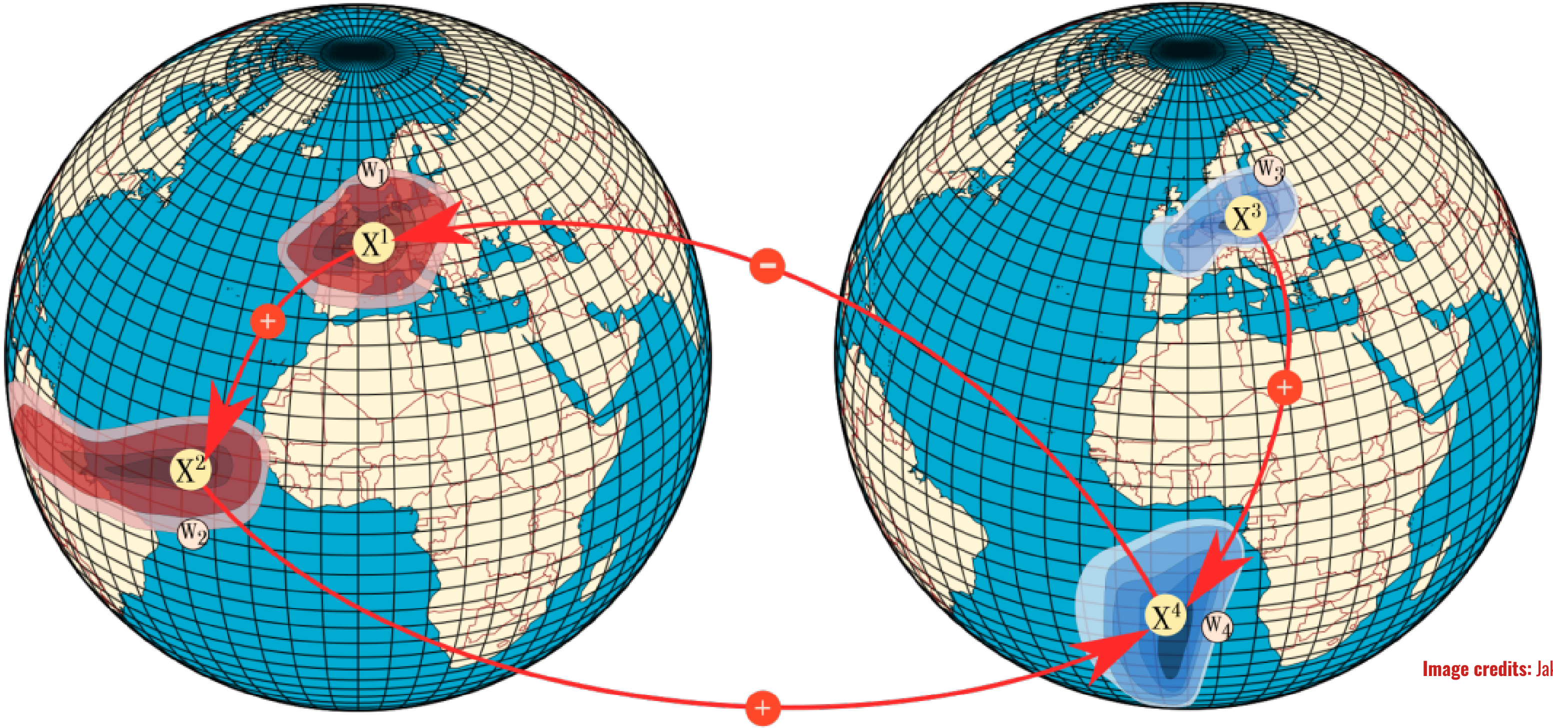
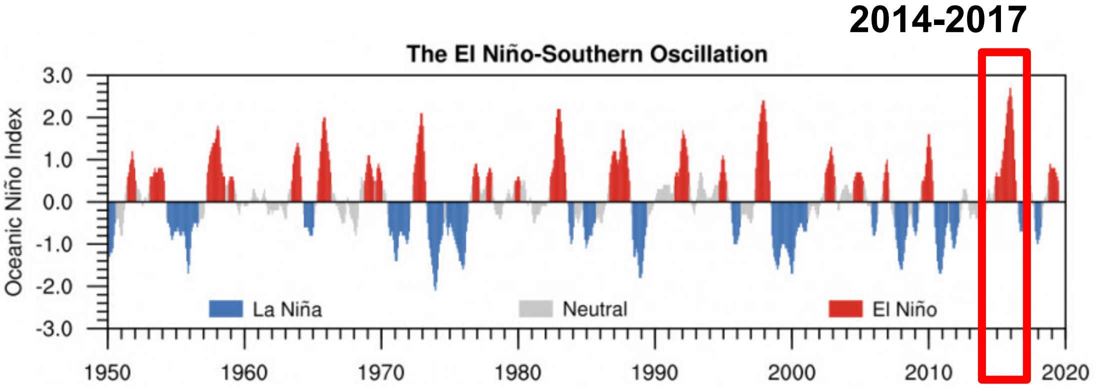


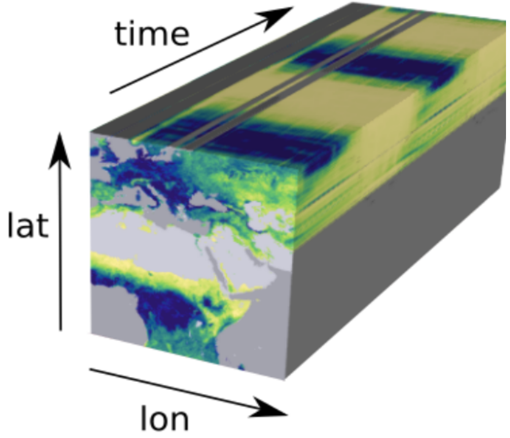
Image credits: Jakob Runge, 2019

Varando, Fernandez, Camps-Valls, Learning Granger Causal Feature Representations, ICML 2021.

5- Learning causal feature representations

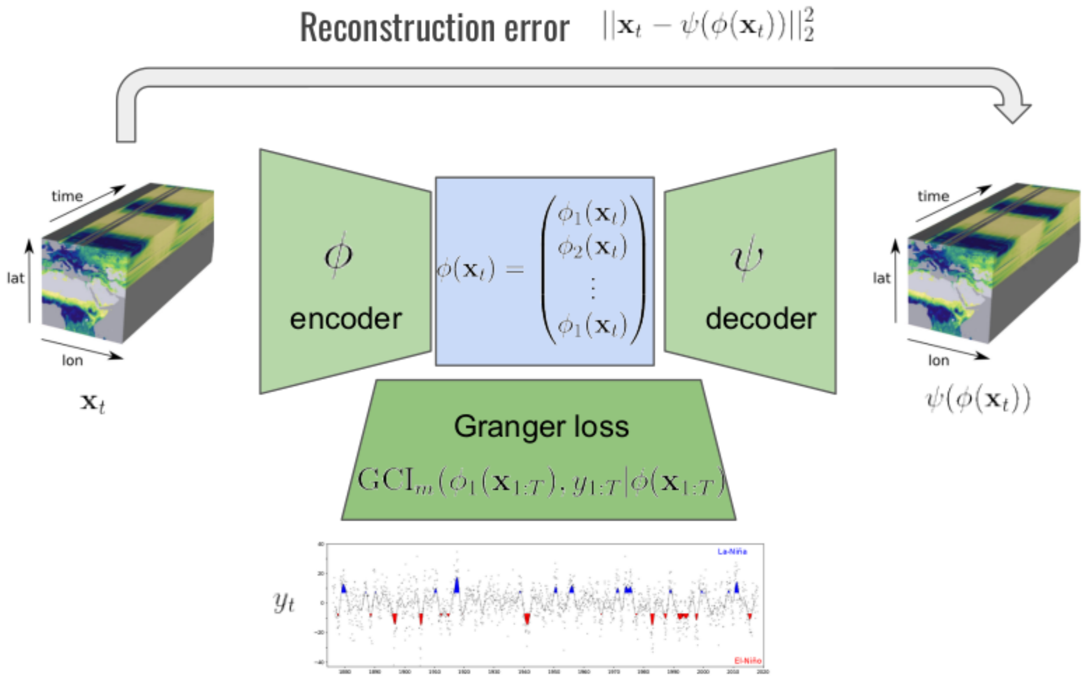
- **ENSO** changes patterns of essential variables like **moisture, greenness & precip.**
- **Goal:** Learn causal impact teleconnections of ENSO on greenness
- NDVI from MODIS in Africa, linear interp, anomalies
- ENSO3.4 index, focus on 2014-2017





Varando, Fernandez, Camps-Valls, Learning Granger Causal Feature Representations, ICML 2021.

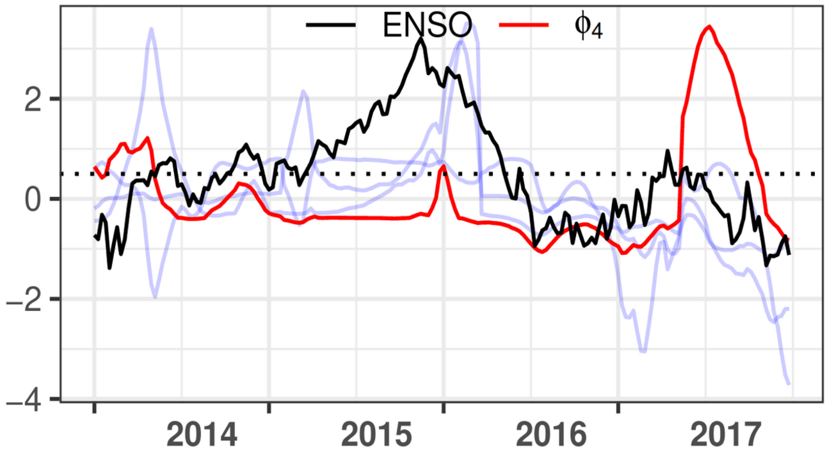
5- Learning causal feature representations



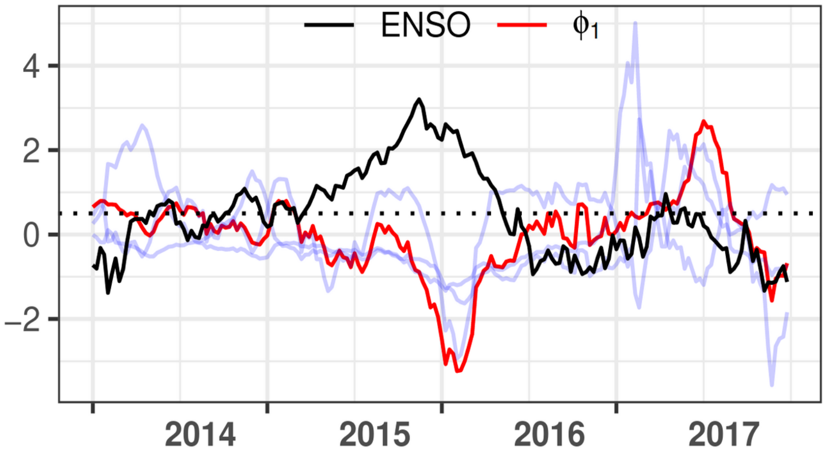
$$\frac{1}{T}\sum_{t=1}^{T} \|\mathbf{x}_t - \psi(\phi(\mathbf{x}_t))\|_2^2 - \beta\, \mathrm{GCI}_m(\phi_1(\mathbf{x}_{1:T}), y_{1:T}|\phi(\mathbf{x}_{1:T}))$$

5- Learning causal feature representations

No Granger penalization $\beta = 0$



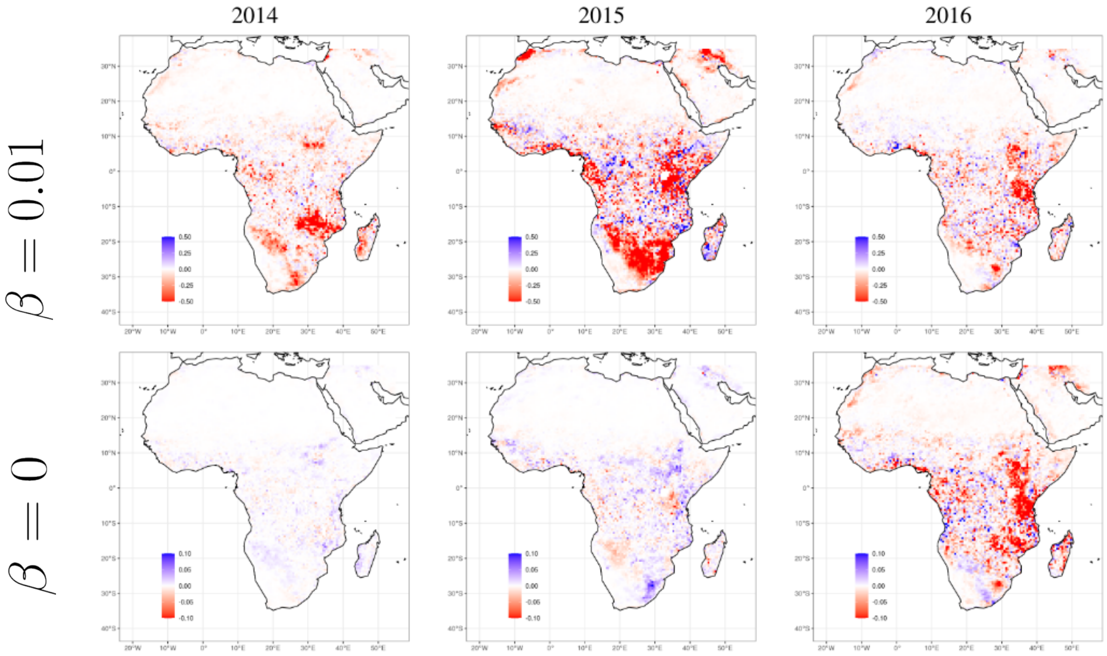
Granger penalization $\beta = 0.01$

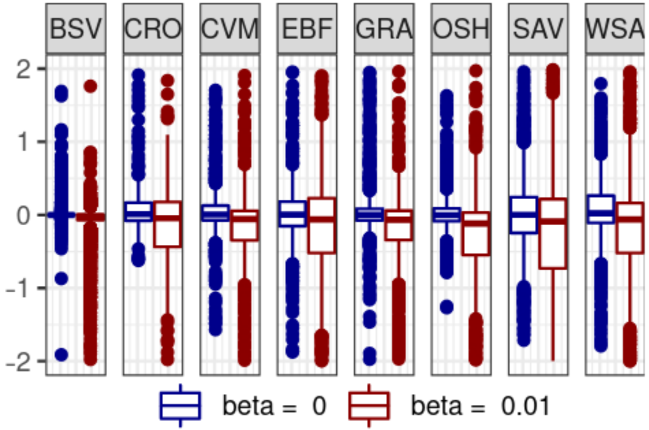


Varando, Fernandez, Camps-Valls, Learning Granger Causal Feature Representations, ICML 2021.

5- Learning causal feature representations

- XAI → Neuron Integrated Gradients (NIG) over the Granger Autoencoder
- Spatially-explicit and temporally resolved activation maps per biome





6- Causality & Disasters

Causal Understanding of Disasters

Causality and “Natural” Disasters

ROBERT A. STALLINGS
Department of Sociology
University of Southern California
rstallin@usc.edu

Worst Cases: Terror and Catastrophe in the Popular Imagination, by **Lee Clarke.** Chicago, IL: University of Chicago Press, 2005. 200 pp. $22.50 cloth. ISBN: 0-226-10859-7.

Heat Wave: A Social Autopsy of Disaster in Chicago, by **Eric Klinenberg.** Chicago, IL: University of Chicago Press, 2002. 320 pp. $15.00 paper. ISBN: 0-226-44322-1.

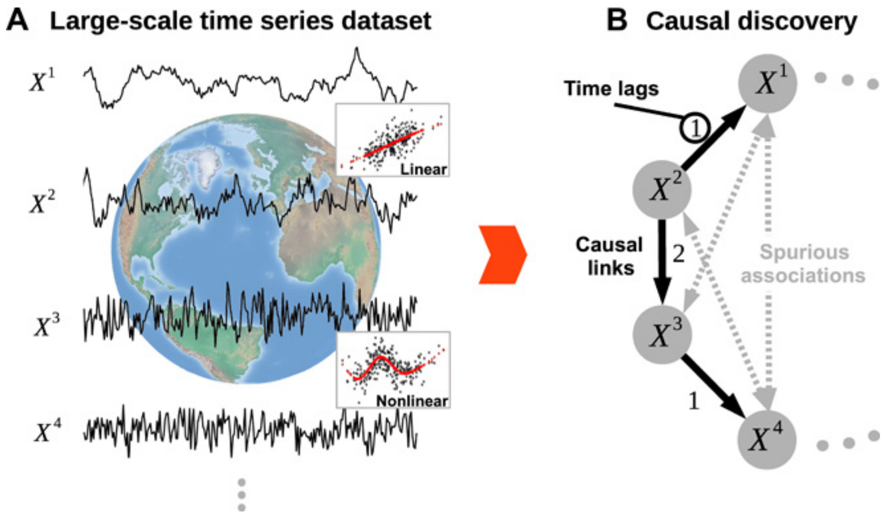
The Vulnerability of Cities: Natural Disasters and Social Resilience, by **Mark Pelling.** Sterling, VA: Earthscan Publications, 2003. 156 pp. $111.57 cloth. ISBN: 1-85383-829-2. $32.60 paper. ISBN: 1-85383-830-6.

Acts of God: The Unnatural History of Natural Disaster in America, by **Ted Steinberg.** Oxford, UK: Oxford University Press, 2003. 320 pp. $19.95 paper. ISBN: 0-19-516545-4.

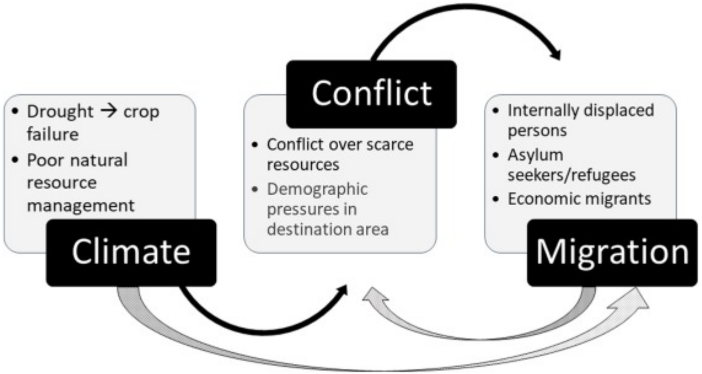
Perils of a Restless Planet: Scientific Perspectives on Natural Disasters, by **Ernest Zebrowski, Jr.** Cambridge, UK: Cambridge University Press, 1997. 320 pp. $24.99 paper. ISBN: 0-521-65488-2.

Understanding Disasters *is about* answering causal queries

- **Causal inference:** draw conclusions about causal relations
- **Causal discovery:** learn relations from data & assumptions
- **Cause-effect estimation:** quantify impacts of interventions



Learning drivers of displacement



nature climate change

Comment | Published: 26 November 2019

Climate migration myths

Ingrid Boas, Carol Farbotko, […] Mike Hulme

Nature Climate Change **9**, 901–903(2019) | Cite this article

476 Accesses | **114** Altmetric | Metrics

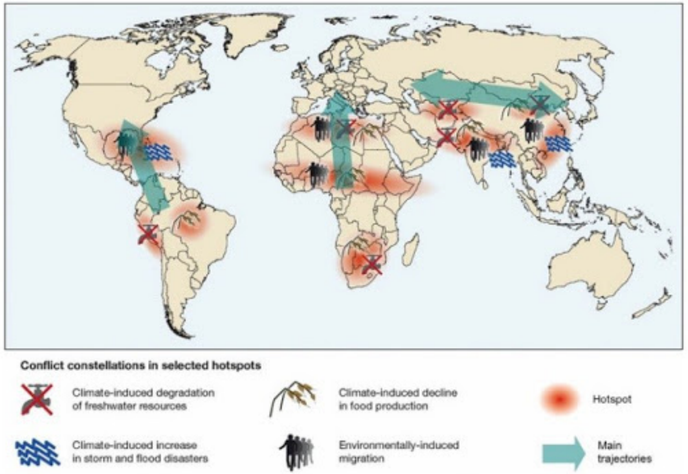
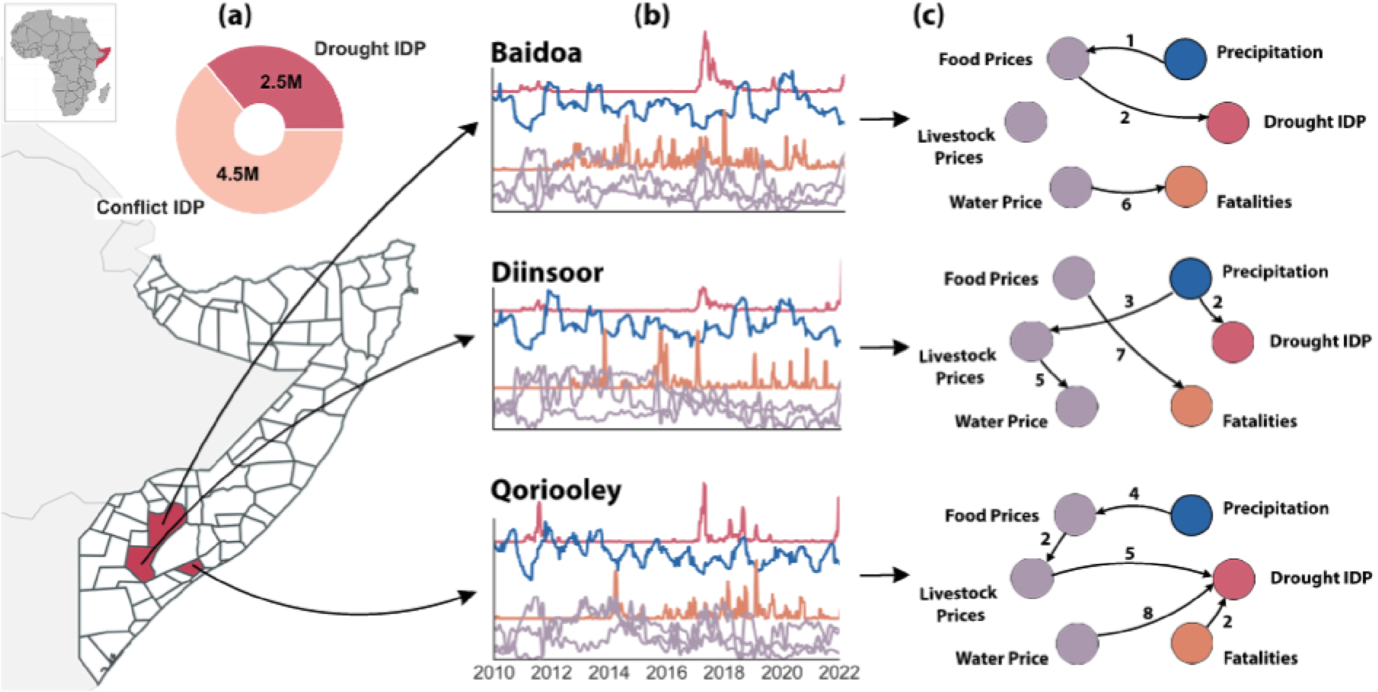


Fig. 11.1 A map of conflict and migration induced by environmental stressors (source: German Advisory Council on Global Change WBGU (2007): Climate Change as a Security Risk arrows added by UNU-EHS)

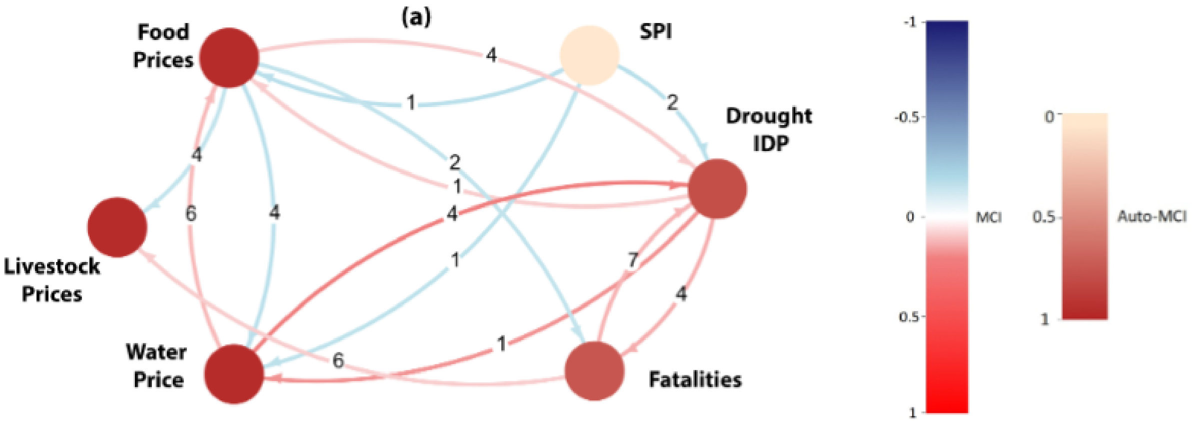
Learning drivers of displacement



"Causal discovery of drought-induced human displacement drivers", Tarraga, Piles, Sevillano, Muñoz, Ronco, Camps-Valls, et al. *Submited*, 2023

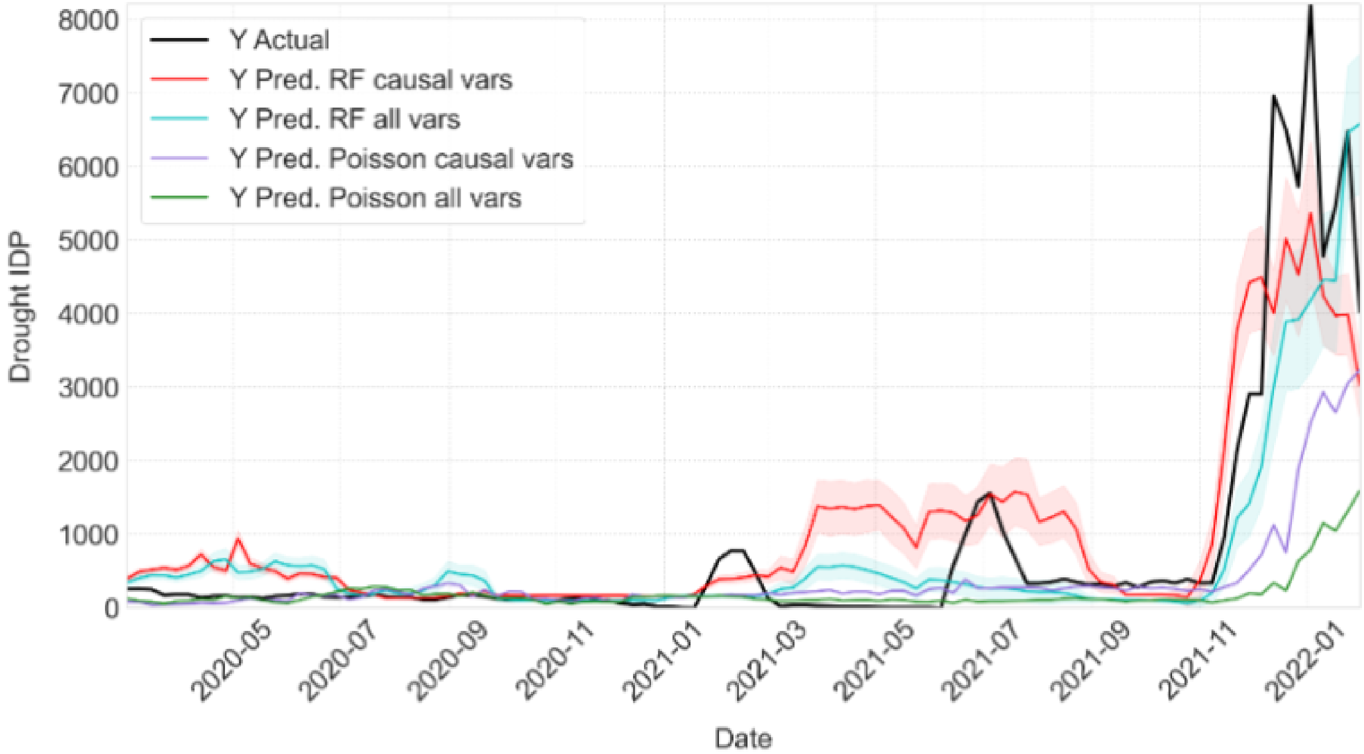
Learning drivers of displacement

Variable	Source	Spatial Resolution	Temporal Resolution
Mean Precipitation	CHIRPS (Chris et al., 2015)	0.05º	Daily
Median NDVI	MODIS TERRA (Didan and Huete, 2015)	1 km	Daily
Mean LST	MODIS TERRA (Wan et al., 2015)	1 km	Daily
Violent Conflict	ACLED (Clionadh et al., 2010)	Geolocated Event	Hourly
Local Market Prices	FSNAU (FSNAU, 2021)	District	Monthly
Drought Displacement	UNHCR PRMN (UNHCR Somalia ID)	District	Weekly
Somalia Districts	UNDP (UNDP)	District	Static
Livelihood Zones	FEWS NET (Fews Net)	Sub-national	Static



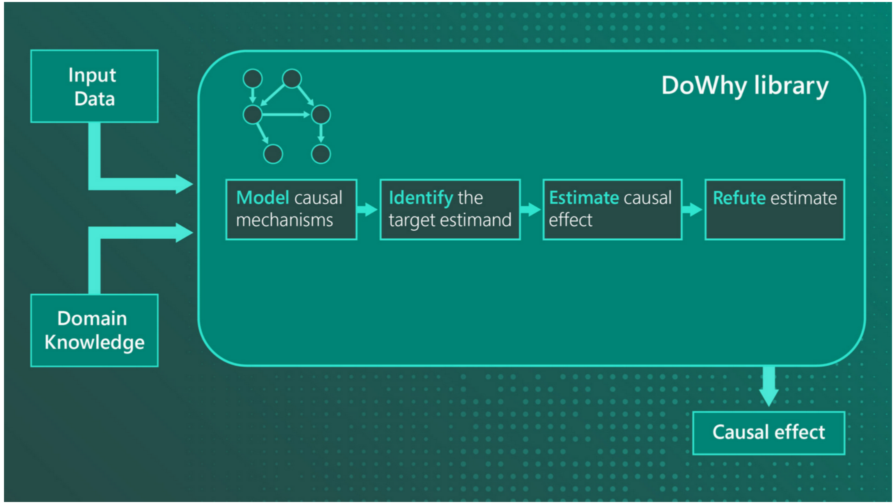
"Causal discovery of drought-induced human displacement drivers", Tarraga, Piles, Sevillano, Muñoz, Ronco, Camps-Valls, et al. *Submited*, 2023

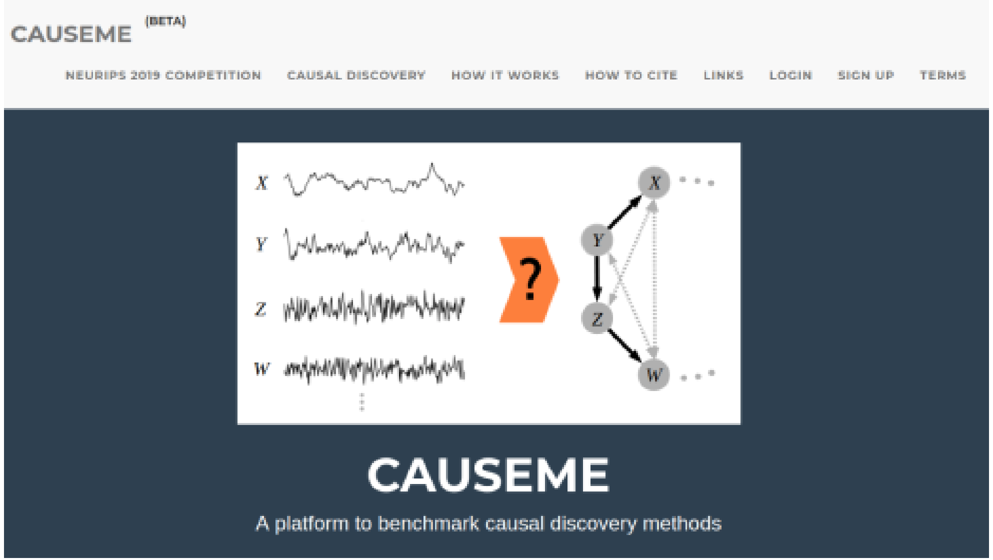
Learning drivers of displacement



"Causal discovery of drought-induced human displacement drivers", Tarraga, Piles, Sevillano, Muñoz, Ronco, Camps-Valls, et al. *Submited*, 2023

Software suites for this …





Conclusions

Take-home message: fitting is not enough!

- You can be right for the wrong reason
- All models are wrong, some are useful
- AI is not deep learning, dude

>> Give Physics, XAI and Causality a Chance

Take-home message: fitting is not enough!

- **AI is a paradigm shift**
  - Excel in classification, (change) detection, parameter retrieval
  - Automate & understand processes
- **Challenges:** interpretability + causal relations + physics consistency
- **Future:**
  - User-centric AI + trustworthiness + accountability
  - Holistic & interdisciplinary education

"Towards a Collective Agenda on AI for Earth Science Data Analysis"
Tuia, Roscher, Wegner, Jacobs, Zhu, and Camps-Valls, G. IEEE Geoscience and Remote Sensing Magazine 2021, arxiv.org/abs/2104.05107

"Living in the Physics - Machine Learning Interplay for Earth Observation"
Camps-Valls et al. AAAI Fall Series 2020 Symposium on Physics-guided AI for Accelerating Scientific Discovery, 2020. arxiv.org/abs/2010.09031

With a lot of help from my friends ...

Research labs

Organizations & Networks

Companies

Funding bodies

ISP at Universitat de València

ISP at Universitat de València - Hiring!