Hybrid & Causal ML in the Earth sciences

Gustau Camps-Valls
Image Processing Laboratory
Universitat de València

isp.uv.es | @isp_uv_es | gustau.camps@uv.es

Outline

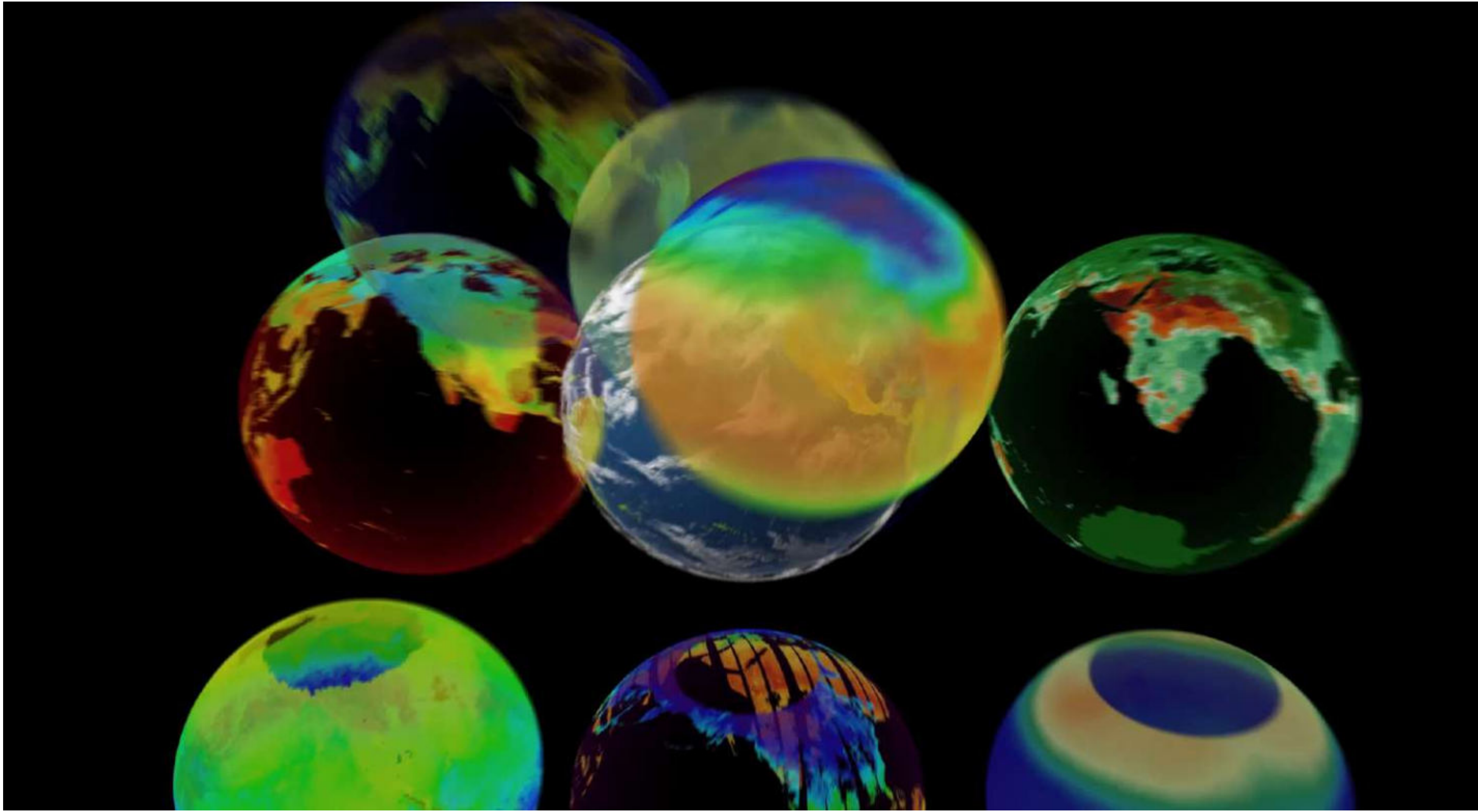
- Part I - ML for Earth sciences
- Part II - Physics-aware Machine Learning
    1. Encode domain knowledge
    2. Emulate complex codes
    3. Learn parametrizations
- Part III – Causal Machine Learning
    1. Learning causal representations
    2. Causal discovery in the wild
    3. Causality with LLMs
- Part IV - Conclusions

Part I
ML for the Earth sciences

Opportunities & challenges

Earth science

Earth observation



DL in Earth and climate sciences – the promise!

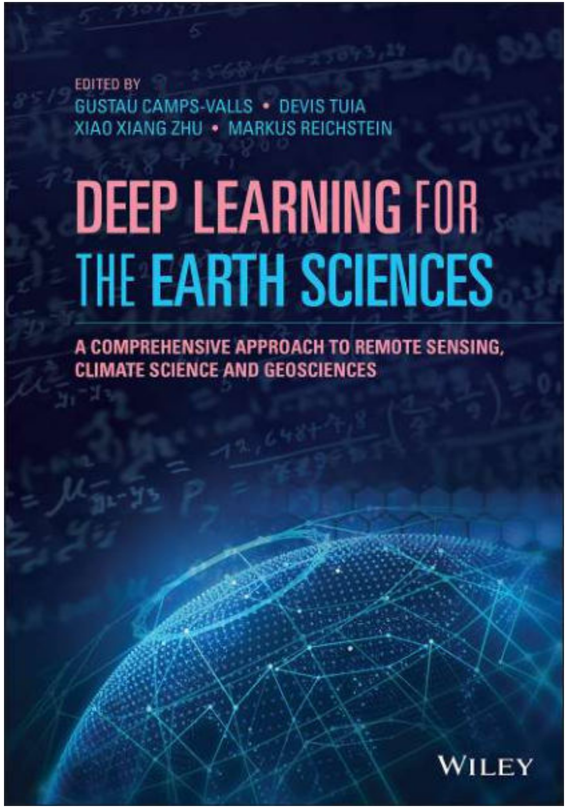
PERSPECTIVE

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

DL in Earth sciences – solved!

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

LETTER

https://doi.org/10.1038/s41586-019-1559-7

Deep learning for multi-year ENSO forecasts

Yoo-Geun Ham[1*], Jeong-Hwan Kim[1] & Jing-Jia Luo[2,3]

PHYSICAL REVIEW RESEARCH **4**, 023028 (2022)

Predicting extreme events from data using deep machine learning: When and where

Junjie Jiang,[1,2] Zi-Gang Huang,[1] Celso Grebogi,[3] and Ying-Cheng Lai[2,4,*]

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

Science RESEARCH ARTI[CLE] Check for updates

Cite as: R. Lam *et al.*, *Science* 10.1126/science.adi2336 (2023).

Learning skillful medium-range global weather forecasting

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

Earth sciences – the *what*, but also the *why & how* questions

- Predict weather / essential climate variables
- Being consistent with domain knowledge
- Understand processes by emulation/parametrization
- Characterize and explain extreme events
- Learn meaningful/causal representations
- Discover causal relations from data
- Attribute causes of changes and anomalies

Earth sciences – not yet!

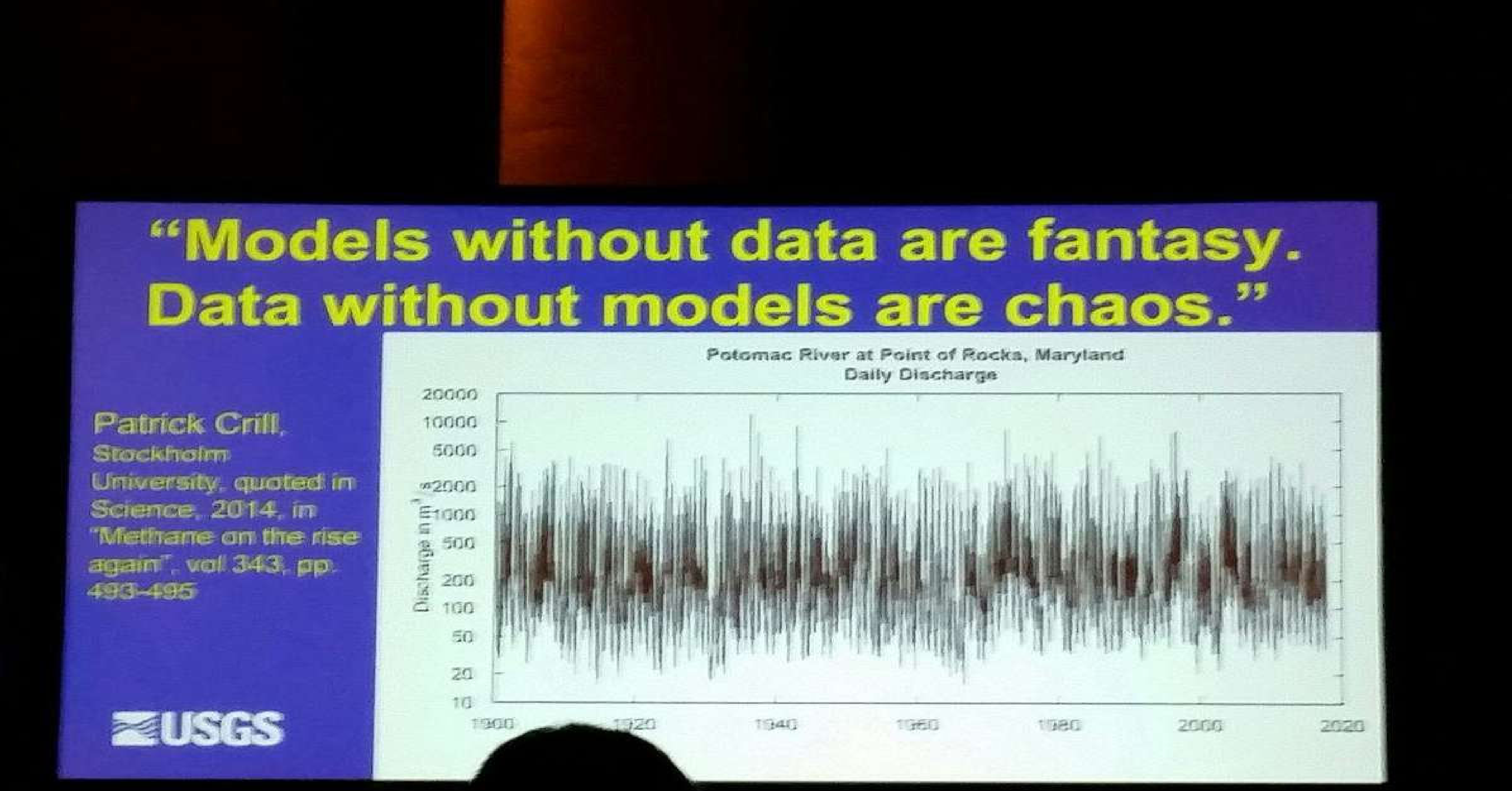
- Predict weather / essential climate variables
- Being consistent with domain knowledge
- Understand processes by emulation/parametrization
- Characterize and explain extreme events
- Learn meaningful/causal representations
- Discover causal relations from data
- Attribute causes of changes and anomalies

Part II

Part III

Part II
Physics-aware ML

* aka physics-guided/informed, domain/science-guided, …



At AGU 2017, New Orleans, USA

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.

1. Encoding domain knowledge

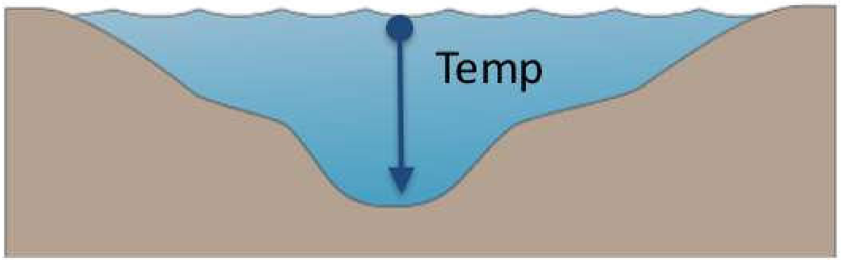
Constrained optimization & hybrid modeling

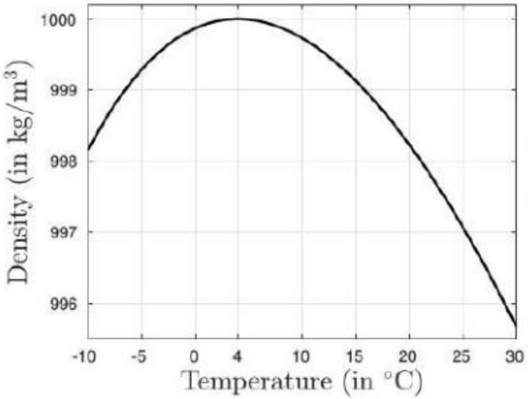
A- Constrained optimization

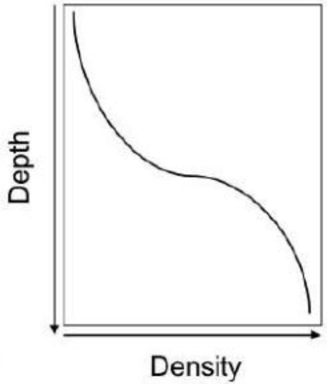
- ML minimizing model errors & violations of the physical laws

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

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







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

B- 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

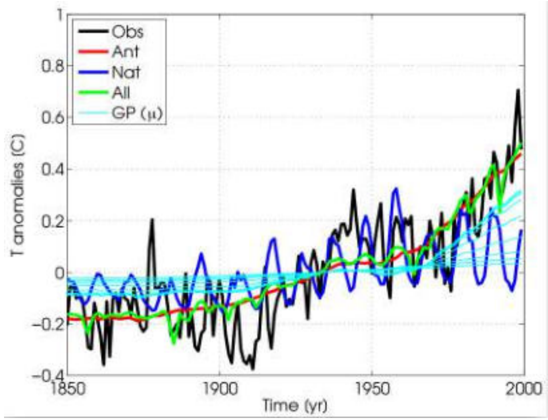
"Physics-aware Nonparametric Regression Models for Earth Data Analysis". Cortés & Camps-Valls. Environmental Research Letters,, 2022

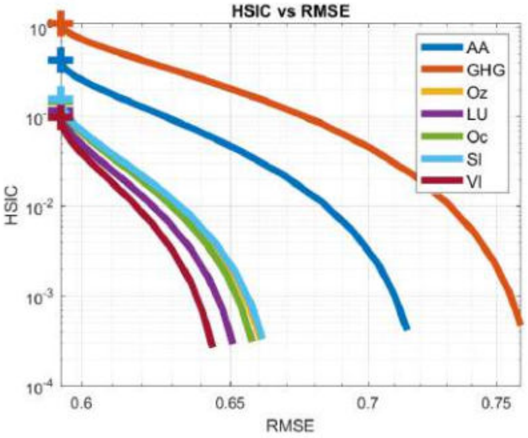
"Kernel Dependence Regularizers and Gaussian Processes with application to Algorithmic Fairness" Zhu Li, Perez-Suay, Camps-Valls and Sejdinovic, Pattern Rec. 2022

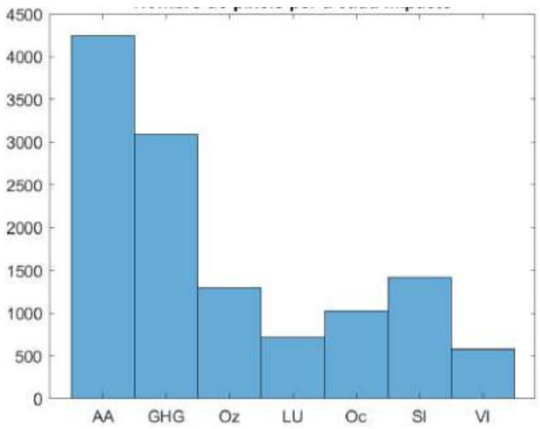
B- Fair optimization

- ML minimizing errors & predictions independent of human factors

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





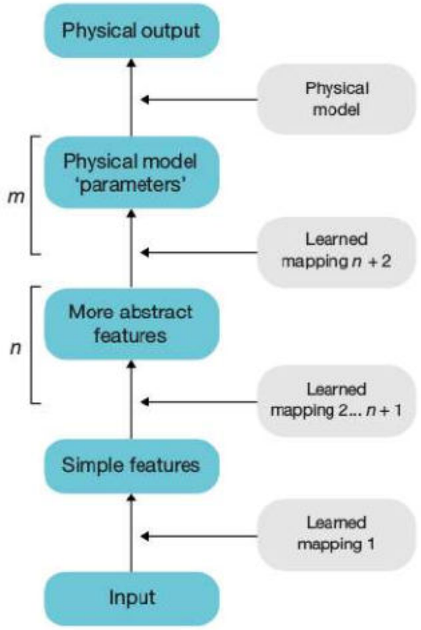


"Fair Kernel Learning" Perez, Laparra, Gomez, Camps-Valls, G. ECML, 2017.

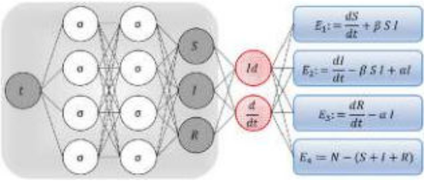
"Consistent Regression of Biophysical Parameters with Kernel Methods" Díaz, Peréz-Suay, Laparra, Camps-Valls, IGARSS 2018

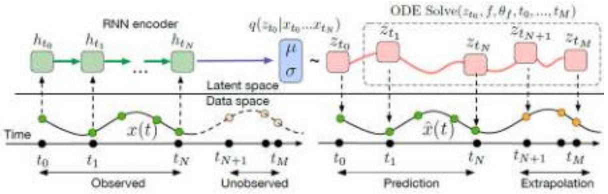
"Kernel Dependence Regularizers and Gaussian Processes with application to Algorithmic Fairness" Zhu Li, Perez-Suay, Camps-Valls and Sejdinovic, , Pattern Rec. 2022

B- Hybrid neural networks



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



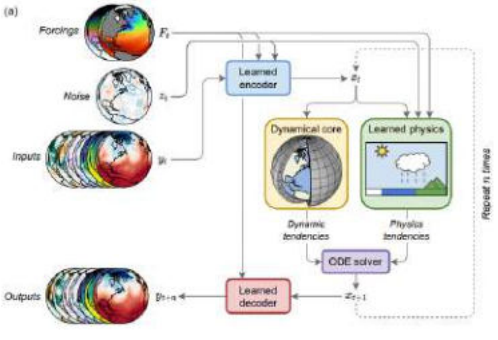


PINNs

...

NeuralODE

NeuralGCM

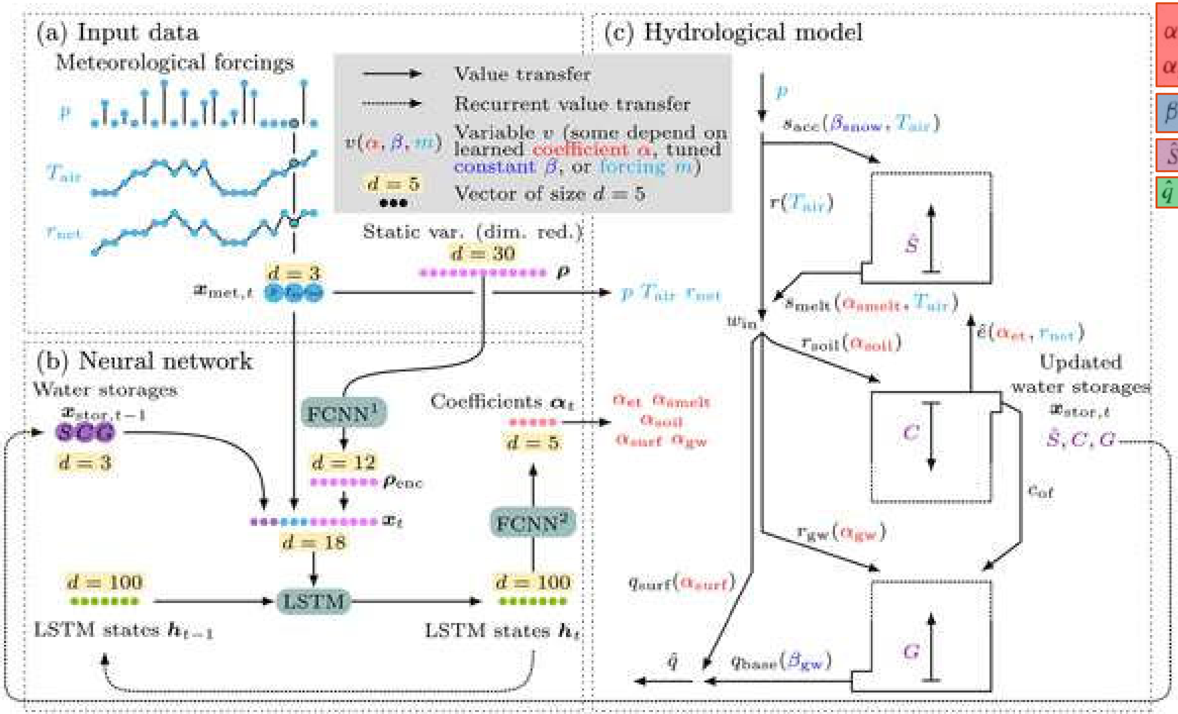


"Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations." Raissi, Maziar, Paris Perdikaris, and George E. Karniadakis. Journal of Computational Physics 378 (2019): 686-707.

"Neural ordinary differential equations." Chen, Ricky TQ, et al. NeurIPS 31 (2018).

"Neural General Circulation Models." Kochkov, Dmitrii, et al. arXiv preprint arXiv:2311.07222 (2023).

B- Hybrid model for the global hydrological cycle



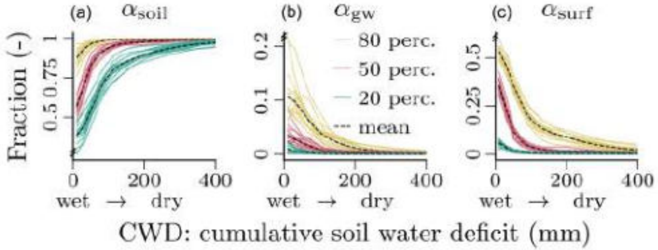
α_{soil} – soil recharge fraction; α_{gw} – groundwater recharge fraction; α_{surf} – surface runoff fraction; α_{smelt} – snowmelt coefficient

β_{snow} – snow undercatch correction constant; β_{gw}: baseflow constant

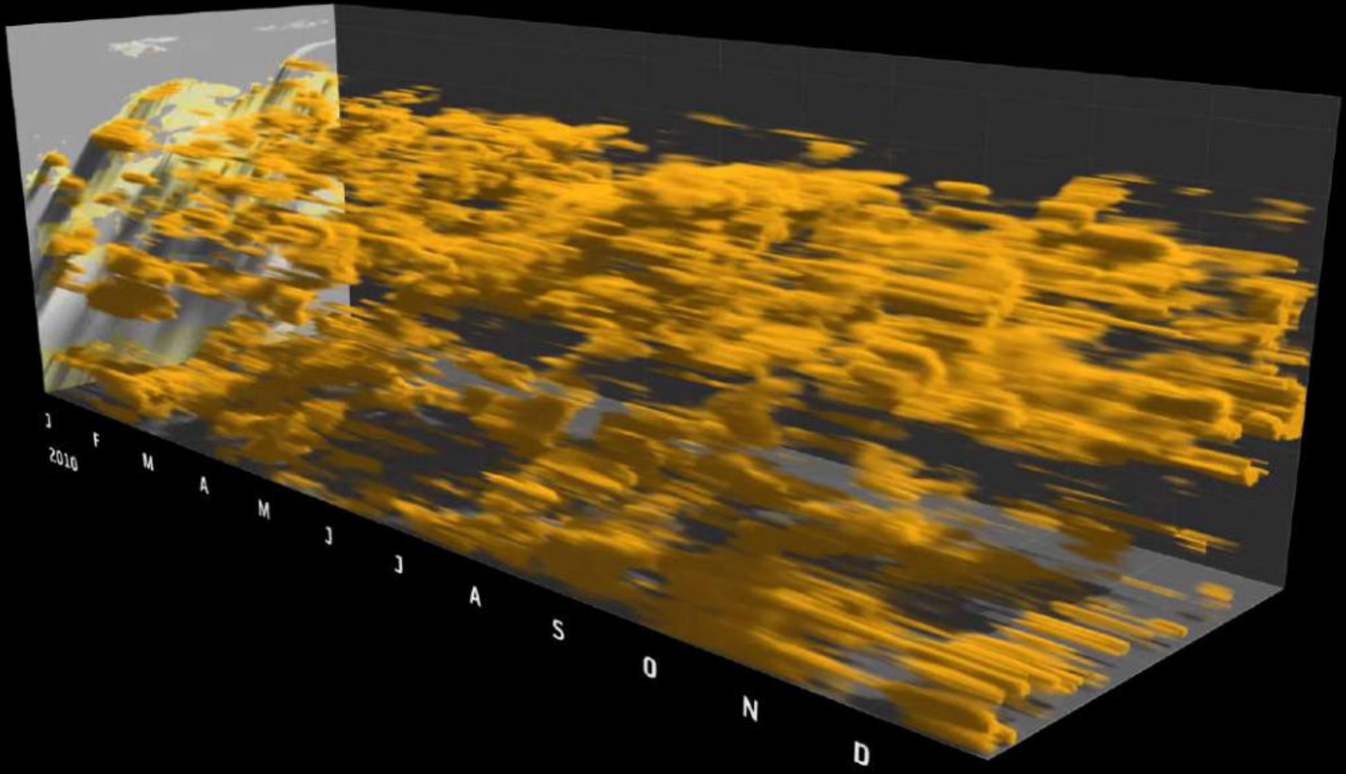
$\hat{S}$ – snow water equivalent; C – cumulative soil water deficit; G – groundwater

$\hat{q}$ – total runoff. $\hat{e}$ – evapotranspiration

	Metric	TWS			SWE			ET			Q		
			MSC	IAV		MSC	IAV		MSC	IAV		MSC	IAV
Global performance	NSE (–)	0.84	0.93	0.54	0.96	0.96	0.22	0.96	0.96	−0.11	0.75	0.78	0.47
	Pearson's r (–)	0.94	0.97	0.80	0.98	0.98	0.87	1.00	1.00	0.67	0.93	0.97	0.81
	SDR (–)	1.15	1.10	1.09	1.02	1.01	1.57	0.99	0.99	1.41	0.93	0.87	1.13
	RMSE (mm)	7.33	4.97	3.27	5.22	5.98	2.16	0.07	0.07	0.02	0.06	0.05	0.03
Local performance	NSE (–)	0.54	0.70	0.26	0.58	0.74	0.15	0.79	0.87	−0.77	0.20	0.17	0.07
	Pearson's r (–)	0.82	0.93	0.67	0.89	0.96	0.64	0.95	0.98	0.60	0.80	0.91	0.62
	SDR (–)	0.98	1.09	0.95	0.91	0.92	0.97	1.03	1.01	1.65	0.98	0.97	1.04
	RMSE (mm)	42.80	22.59	28.72	15.49	13.13	10.60	0.27	0.22	0.14	0.44	0.31	0.27

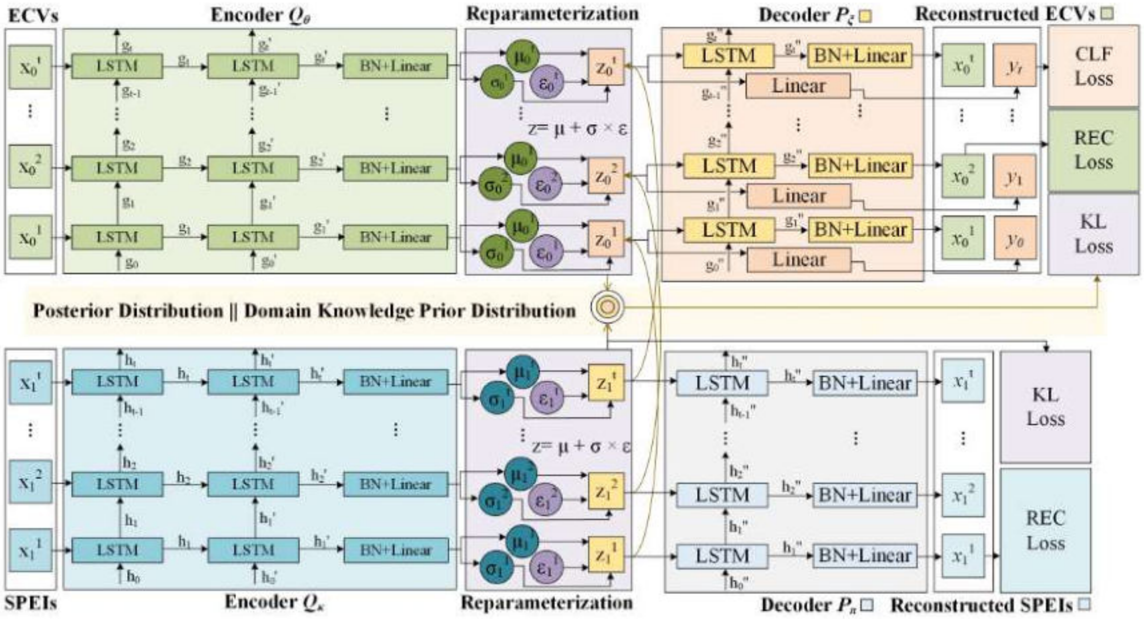


"Towards hybrid modeling of the global hydrological cycle." Kraft, Basil, et al. Hydrology and Earth System Sciences 26.6 (2022): 1579-1614.


J
F
M
A
M
J
J
A
S
O
N
D
2010

C- Extreme event detection, anticipation & explanation

- Multimodal architecture: blend&match latent representations

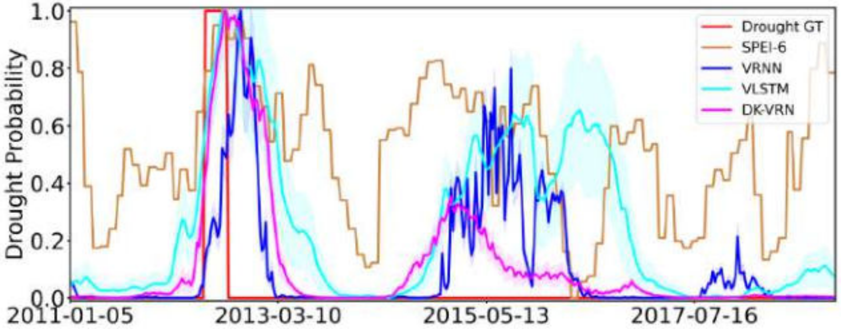


"Domain Knowledge-Driven Variational Recurrent Networks for Drought Monitoring"
Mengxue Zhang, Miguel Ángel Fernández-Torres, Gustau Camps-Valls, Submitted 2023

C- Extreme event detection, anticipation & explanation

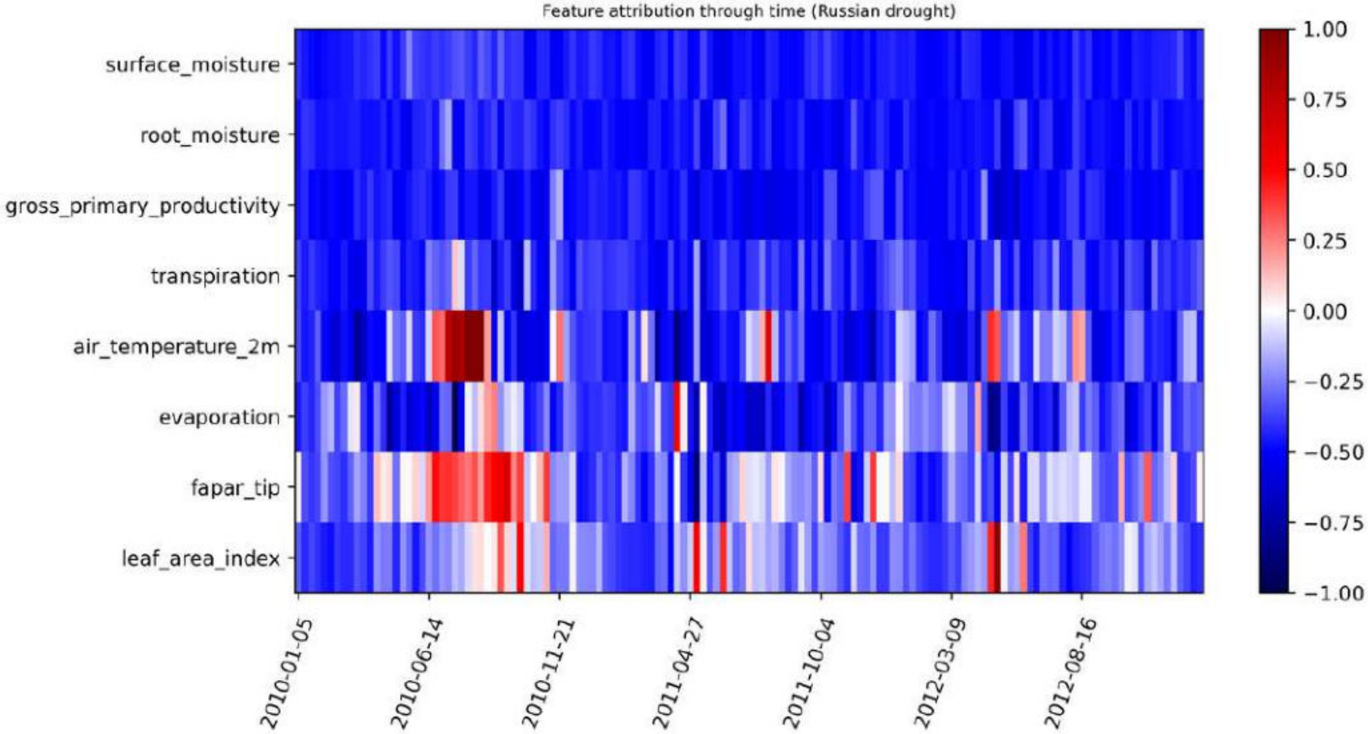
- Transfer across space and time

	CLF-Loss	REC-Loss	KL-Loss	ROC-AUC	PR-AUC	Macro F1
Afghanistan	✓	✗	standard	71.6±5.3	18.7±4.7	51.2±2.1
	✓	✗	domain knowledge	77.0±4.3	26.1±3.9	**53.5±3.0**
	✓	✓	standard	73.4±2.9	19.9±3.5	49.3±3.1
	✓	✓	domain knowledge	**79.7±0.5**	**26.4±1.3**	**53.5±0.7**
Italy	✓	✗	standard	68.5±7.2	15.8±6.5	54.4±4.0
	✓	✗	domain knowledge	74.7±8.3	15.9±9.6	56.7±5.1
	✓	✓	standard	76.0±5.1	12.1±2.6	52.1±1.9
	✓	✓	domain knowledge	**84.3±0.6**	**24.2±2.4**	**55.1±1.1**
Moldova	✓	✗	standard	97.1±3.3	28.6±3.0	59.1±9.1
	✓	✗	domain knowledge	**98.2±3.1**	41.1±7.2	**75.9±9.5**
	✓	✓	standard	92.1±9.4	21.5±8.5	64.6±11
	✓	✓	domain knowledge	94.6±1.1	**43.0±4.0**	71.5±0.4
Russia	✓	✗	standard	80.6±2.8	14.3±2.7	57.6±1.0
	✓	✗	domain knowledge	82.5±2.7	13.7±.0	57.4±2.3
	✓	✓	standard	86.0±3.0	17.7±2.8	60.5±1.6
	✓	✓	domain knowledge	**89.4±0.2**	**18.5±0.4**	**63.2±0.2**
Europe-0	✓	✗	standard	74.8±3.5	13.2±0.9	56.2±1.0
	✓	✗	domain knowledge	80.2±3.2	14.4±3.9	58.0±1.9
	✓	✓	standard	79.6±0.9	16.7±1.3	58.1±1.0
	✓	✓	domain knowledge	**84.3±0.1**	**17.4±0.3**	**61.1±0.3**
Europe-1	✓	✗	standard	70.9±5.5	0.5±0.2	48.8±1.6
	✓	✗	domain knowledge	76.9±4.7	0.5±0.2	48.4±0.7
	✓	✓	standard	76.5±5.0	0.4±0.1	49.2±0.9
	✓	✓	domain knowledge	**82.8±0.4**	**0.7±0.1**	**50.2±0.1**



"Domain Knowledge-Driven Variational Recurrent Networks for Drought Monitoring" Mengxue Zhang, Miguel Ángel Fernández-Torres, Gustau Camps-Valls, Submitted 2023

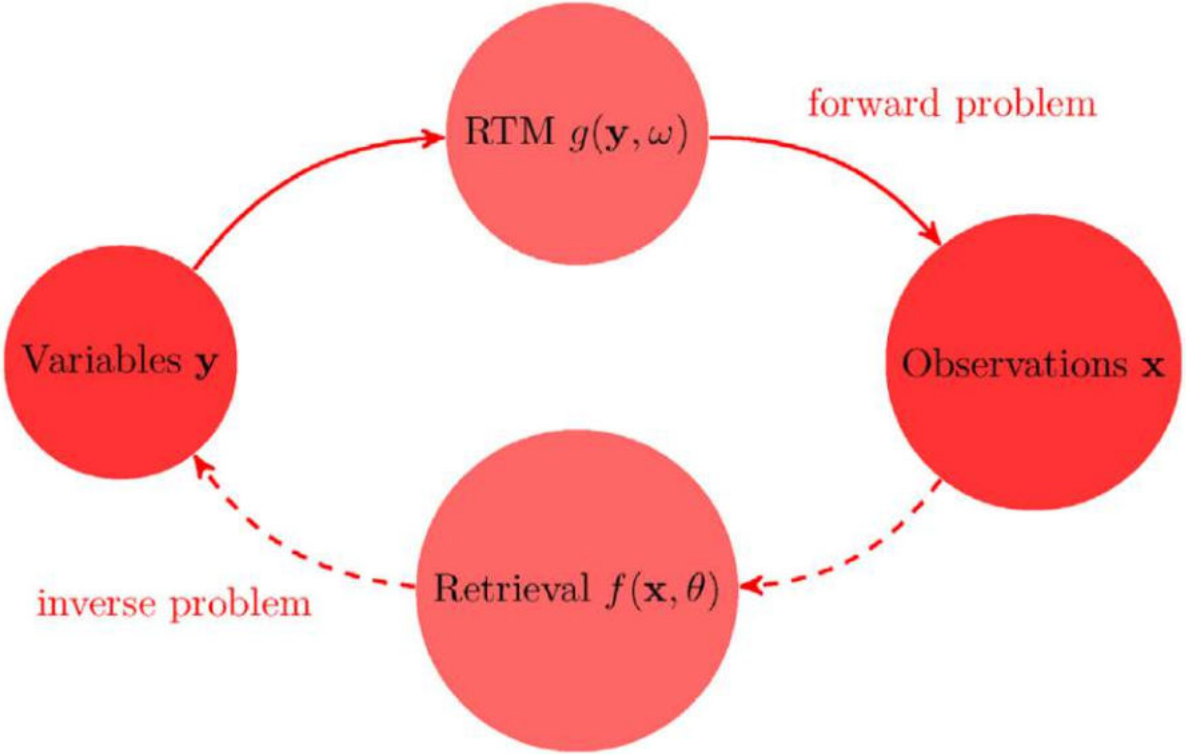
C- Extreme event detection, anticipation & attribution



"Domain Knowledge-Driven Variational Recurrent Networks for Drought Monitoring" Mengxue Zhang, Miguel Ángel Fernández-Torres, Gustau Camps-Valls, Submitted 2023

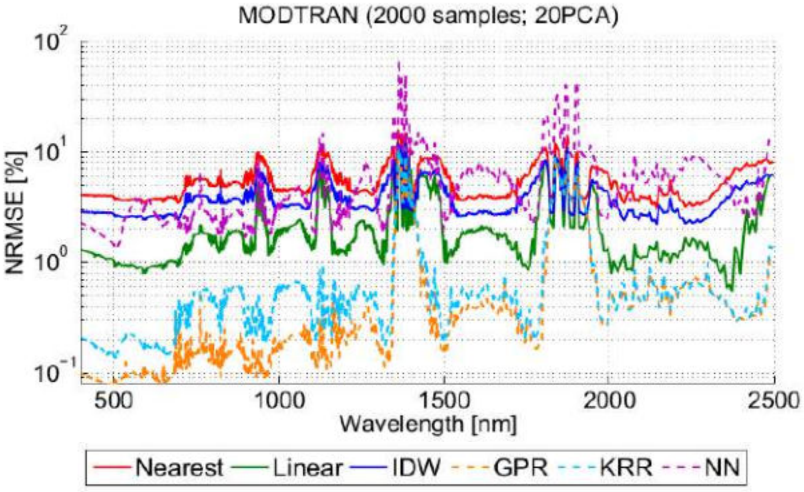
2. Emulation

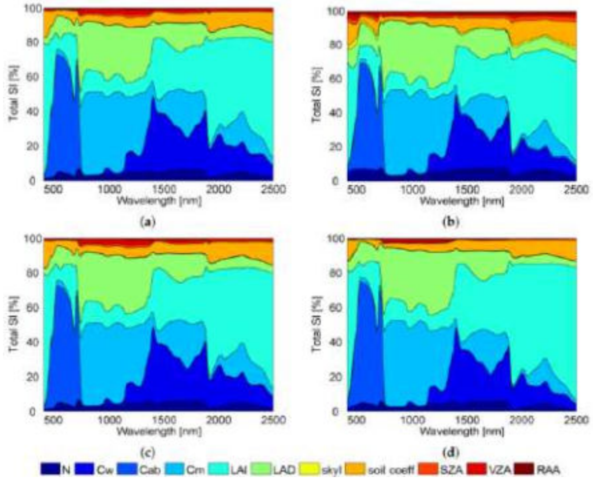
Forward & inverse modeling

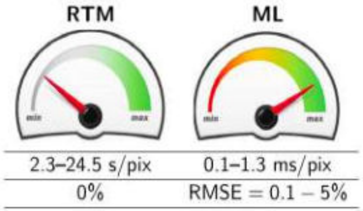


A- Emulating complex codes

- GP Emulation = UQ/UP + Sensitivity analysis + Speed





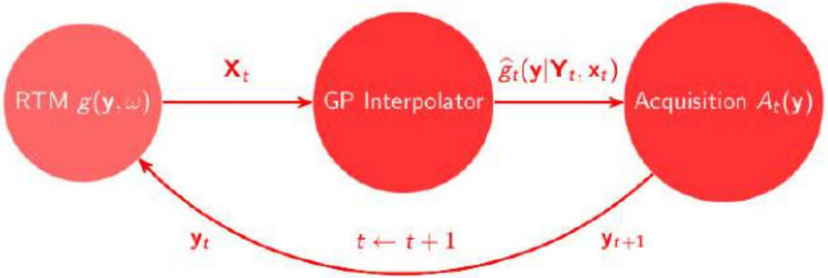


"Emulation of Leaf, Canopy and Atmosphere Radiative Transfer Models for Fast Global Sensitivity Analysis"
Verrelst, Camps-Valls et al Remote Sensing of Environment, 2016

"Emulation as an accurate alternative to interpolation in sampling radiative transfer codes"
Vicent and Camps-Valls, IEEE Journal Sel. Topics Rem. Sens, Apps. 2018

B- Optimizing emulators with GPs

- AGAPE = GP interpolation + Acquisition function

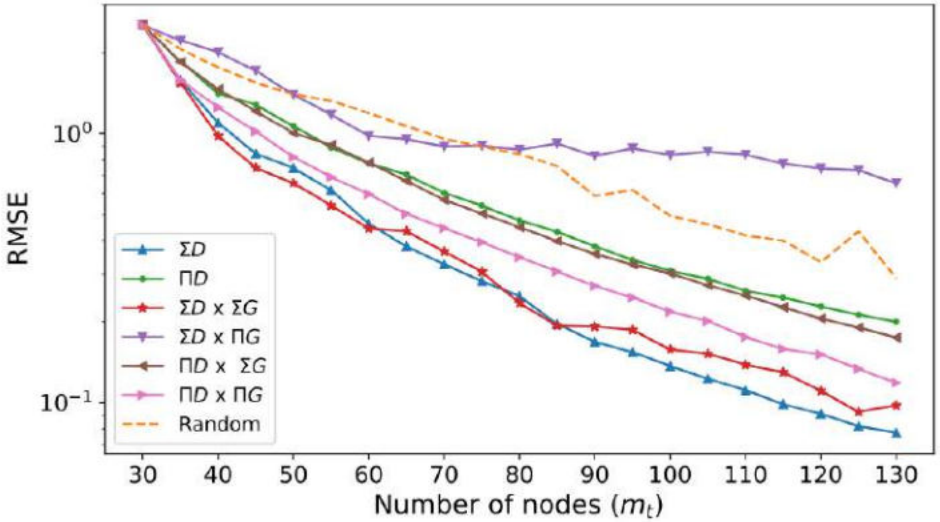


$$\widehat{g}(\mathbf{y}|\mathbf{Y}_t, \mathbf{x}_t) = \mu_{\text{GP}}(\mathbf{y}) = \mathbf{k}^\top \mathbf{K}^{-1} \mathbf{x}_t$$

$$A_t(\mathbf{y}) = [G_t(\mathbf{y})]^{\beta_t} D_t(\mathbf{y}), \quad \beta_t \in [0,1]$$

$$D_t(\mathbf{y}) = \sigma^2_{\text{GP}}(\mathbf{y}) = k(\mathbf{y},\mathbf{y}) - \mathbf{k}^\top \mathbf{K}^{-1} \mathbf{k}$$

$$G_t(\mathbf{y}) = \|\nabla \widehat{g}_t(\mathbf{y}|\mathbf{Y}_t, \mathbf{x}_t)\| = \left\| \sum_{i=1}^{m_t} \alpha_i \nabla k(\mathbf{y}, \mathbf{y}_i) \right\|$$



Active Emulation of Computer Codes with Gaussian Processes. Svendsen, D.H. and Martino, L. and Camps-Valls, G. Pattern Recognition 100 (107103) :1--12, 2020

3. Learning Parametrizations

Model inversion and learning parametrizations

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

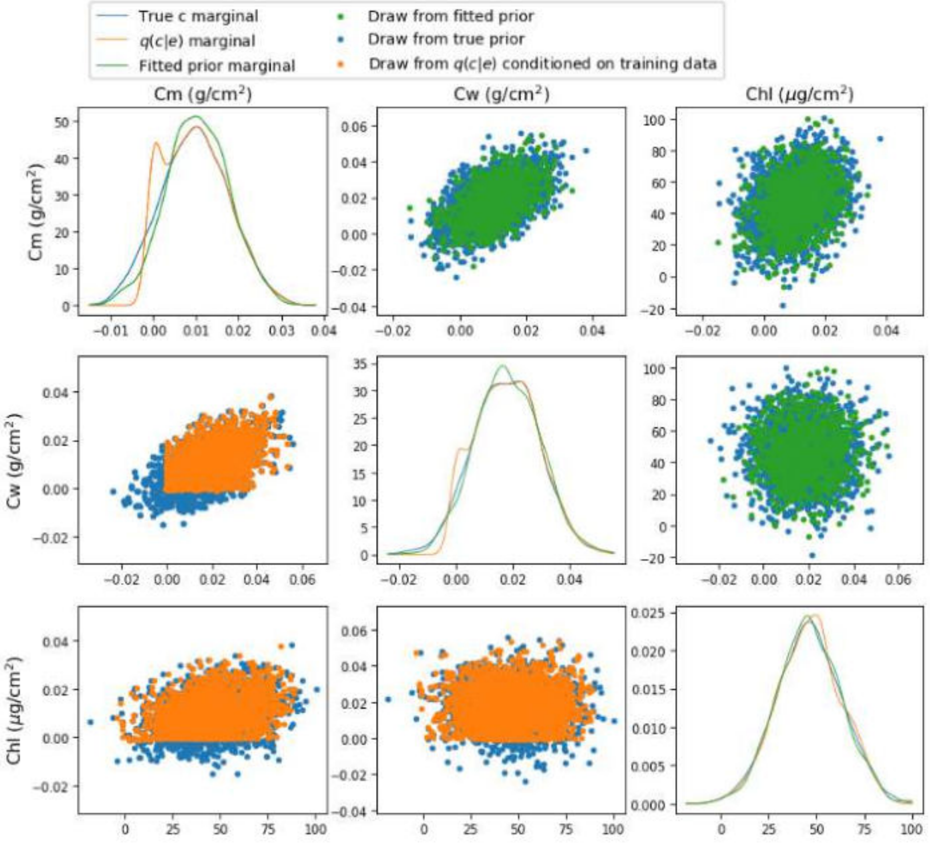
$$P(E|c) = \mathcal{N}(E|RTM(c), \sigma \mathrm{I})$$

- Kingma and Welling (2013)
  - Introduce the variational posterior into the log marginal likelihood
  - Choose the variational posterior to be Gaussian (mean&cov w/ nnet)
  - Compute the expected value of the log-likelihood (KLDs btw. Gaussians easy)
- Unlike in the VAE literature: deterministic decoder + low noise variance in the lik.

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

Model inversion and learning parametrizations

- VAE is orders of magnitude faster than MCMC or ABC, but problems with multimodal distributions
- The VAE scheme provides a posterior approximation
- Readily used for fast inverse modeling & parameterization
- Learn prior distributions of land parameters; canopy water content, chlorophyll and dry matter



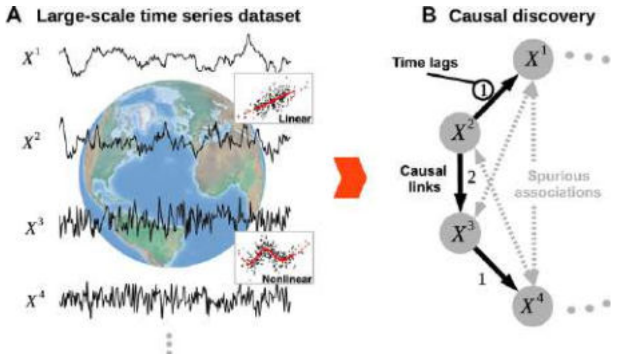
"Variational inference over radiative transfer model for biophysical parameter retrieval"
D. Svendsen, L. Martino, V. Laparra, G. Camps-Valls, *Machine Learning*, 2021

Part III
Causal Machine Learning

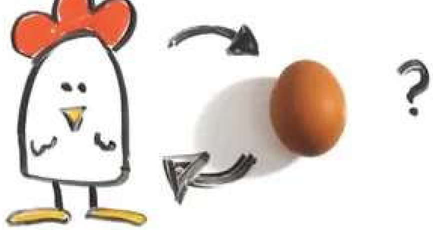
Pragmatic approaches in the wild

Causal understanding *means* better management & robustness

- **Discover causal relations** → being right for the right reasons
- **Identify causal factors** of events → prevent them from occurring
- **Predict occurrence** of disasters → causal forecasting models
- **Evaluate the effectiveness of interventions** → better policies
- **Causal evaluation** → hypothesis testing & model-obs comparison



Challenges in causality



1. Difficulty of identifying causation from data
2. Multidimensionality & collinearity
3. Poor data quality & assumptions (lin/Gauss/iid)
4. Poor data quantity
5. Many factors difficult to measure or quantify
6. Many confounders & sufficiency assumption
7. Hidden/latent factors
8. Nonstationarities in a changing Planet
9. Idenfiability and falsifiability issues
10. Weak or controversial domain knowledge

1. Causal discovery in time series

Causal inference for the Earth system

nature COMMUNICATIONS

PERSPECTIVE

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

Inferring causation from time series in Earth system sciences

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

Contents lists available at ScienceDirect

Physics Reports

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

ELSEVIER

Discovering causal relations and equations from data

Gustau Camps-Valls [a,*], Andreas Gerhardus [b,1], Urmi Ninad [c,b,1], Gherardo Varando [a,1], Georg Martius [d,e], Emili Balaguer-Ballester [f,g], Ricardo Vinuesa [h], Emiliano Diaz [a], Laure Zanna [i], Jakob Runge [b,c]

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

nature reviews earth & environment https://doi.org/10.1038/s43017-023-00431-y

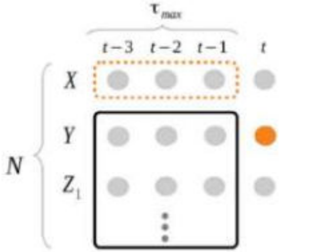
Technical review

Check for updates

Causal inference for time series

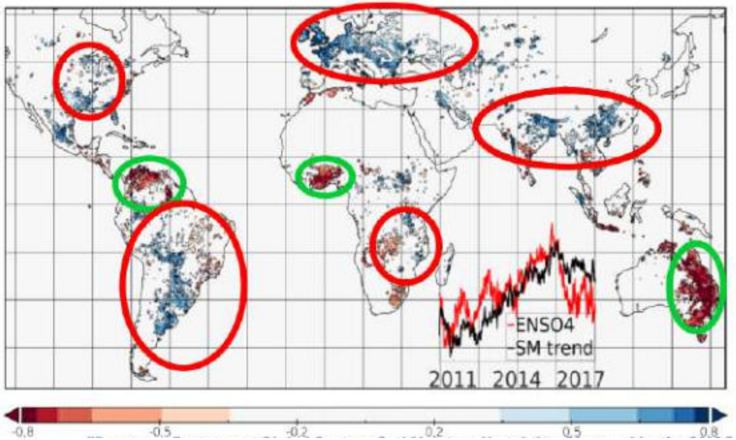
Jakob Runge [1,2], Andreas Gerhardus [1], Gherardo Varando [3], Veronika Eyring [4,5] & Gustau Camps-Valls [3]

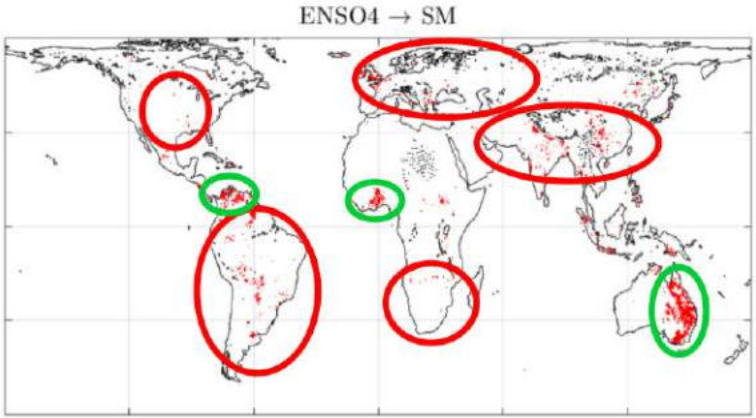
Ex. 1 - Nonlinear Nonstationary Granger Causality (XKGC)



$$Y_{t+1} = a^\top X_t + \varepsilon_t^Y$$
$$Y_{t+1} = b_1^\top Y_t + b_2^\top X_t + \varepsilon_t^{Y|X}$$
$$X \to Y \leftrightarrow \mathbb{V}[\varepsilon_t^Y] \ll \mathbb{V}[\varepsilon_t^{Y|X}]$$

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



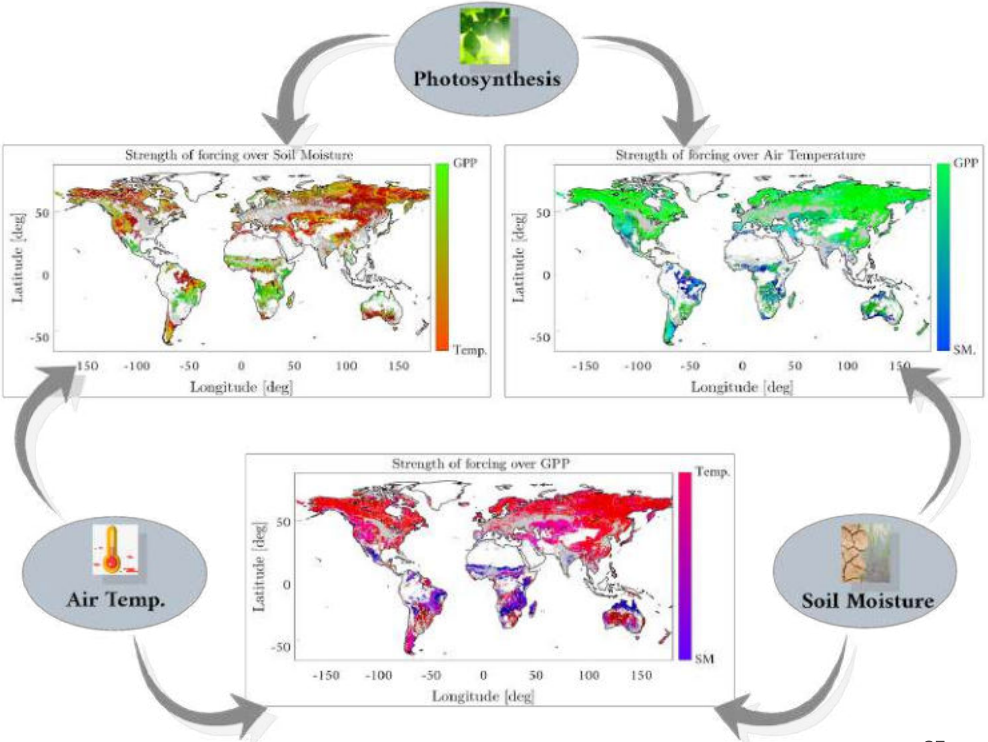


- Causality is sharper than correlation
- ENSO4 “causes” SM in very dry (Sahel) and very wet (tropical rain forests)

“Inferring causation from time series with perspectives in Earth system sciences”, Runge, Bathiany, Bollt, Camps-Valls, et al. Nat Comm (submitted), 2018.
“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, Runge, Camps-Valls. (2019). **CauseMe: http://causeme.uv.es**

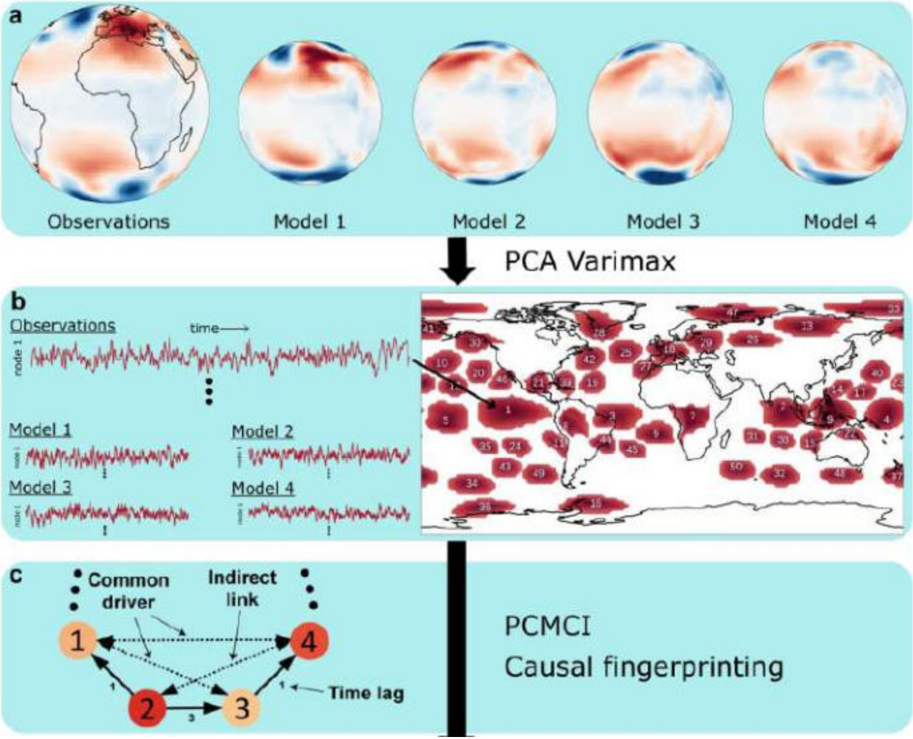
Ex. 2- Robust Convergent Cross Mapping (RCCM)

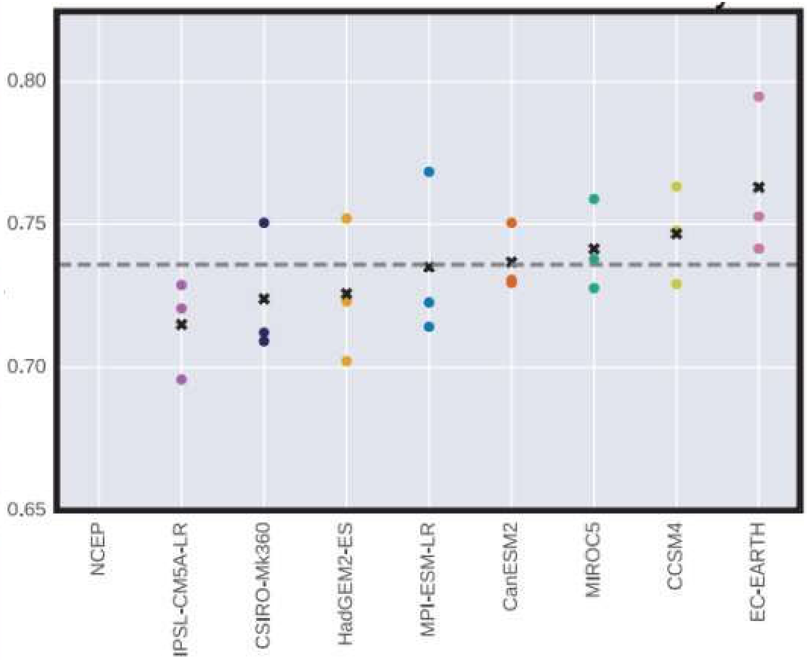
- Causality on (GGP, Tair, SM)
- Causal maps capture general knowledge
- In dry (water-limited) areas, GPP is caused/driven by SM
- Temperature is mainly an effect in boreal regions
- GPP affects SM in dry/savannas/shrubs, possibly related through ET
- SM in boreal regions matches with a reduction in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



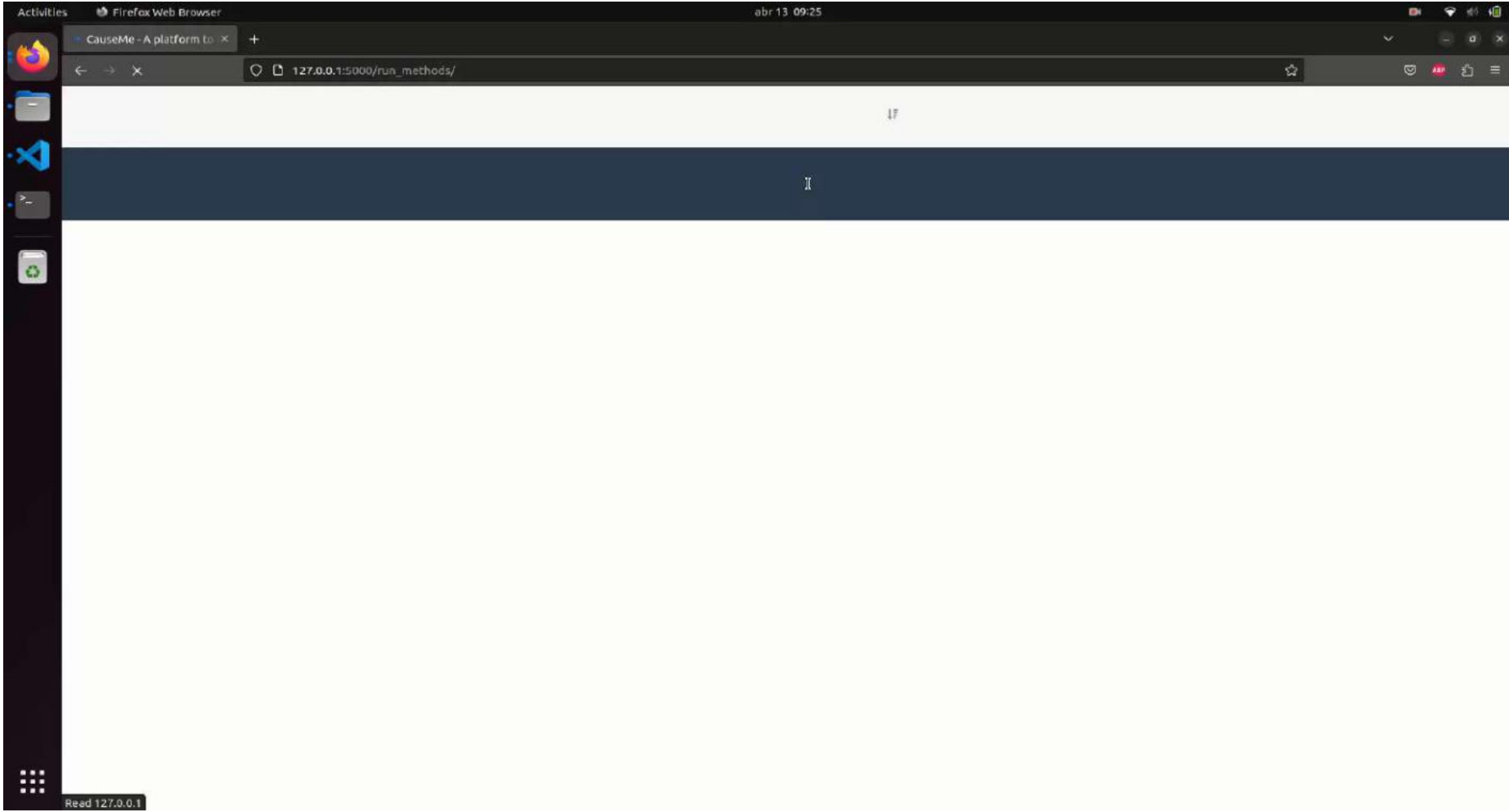
“Inferring causal relations from observational long-term carbon and water fluxes records” E. Diaz, J.E. Adsuara, A. Moreno, M. Piles, G. Camps-Valls,, Sci. Rep 2022

Ex. 3- PC with momentary conditional independence (PCMCI)





«Causal networks for climate model evaluation and constrained projections» Nowack, et al, Nature Comm. (2020)


Activities
Firefox Web Browser
abr 13 09:25
CauseMe - A platform to
127.0.0.1:5000/run_methods/
Read 127.0.0.1

2. Learning causal representations

Learning causal feature representations

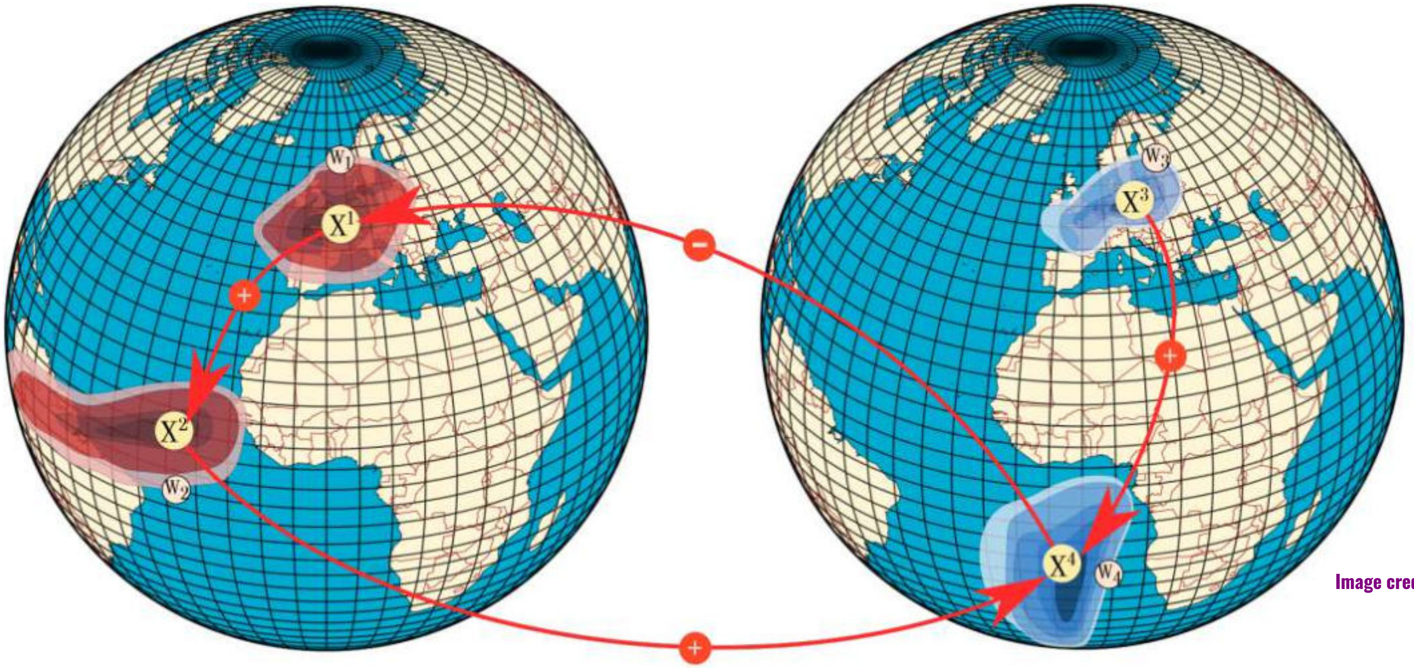
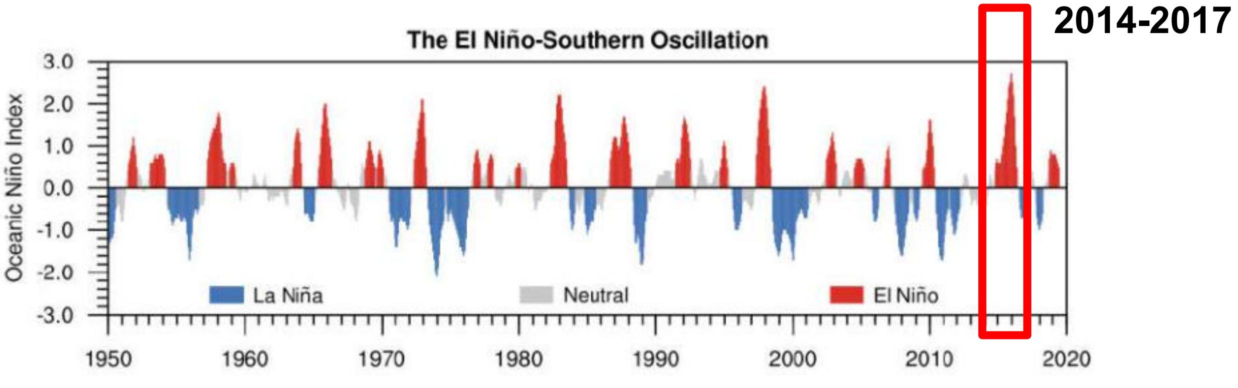


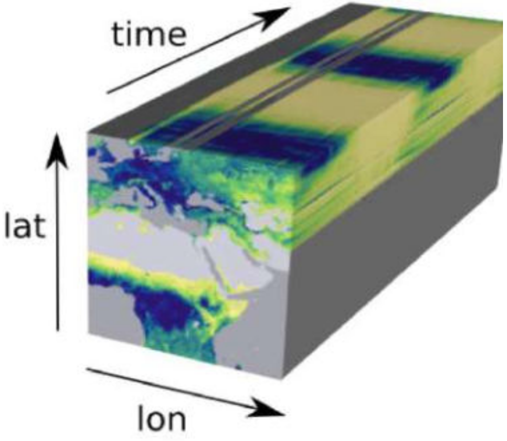
Image credits: Jakob Runge, 2019

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

Learning causal feature representations

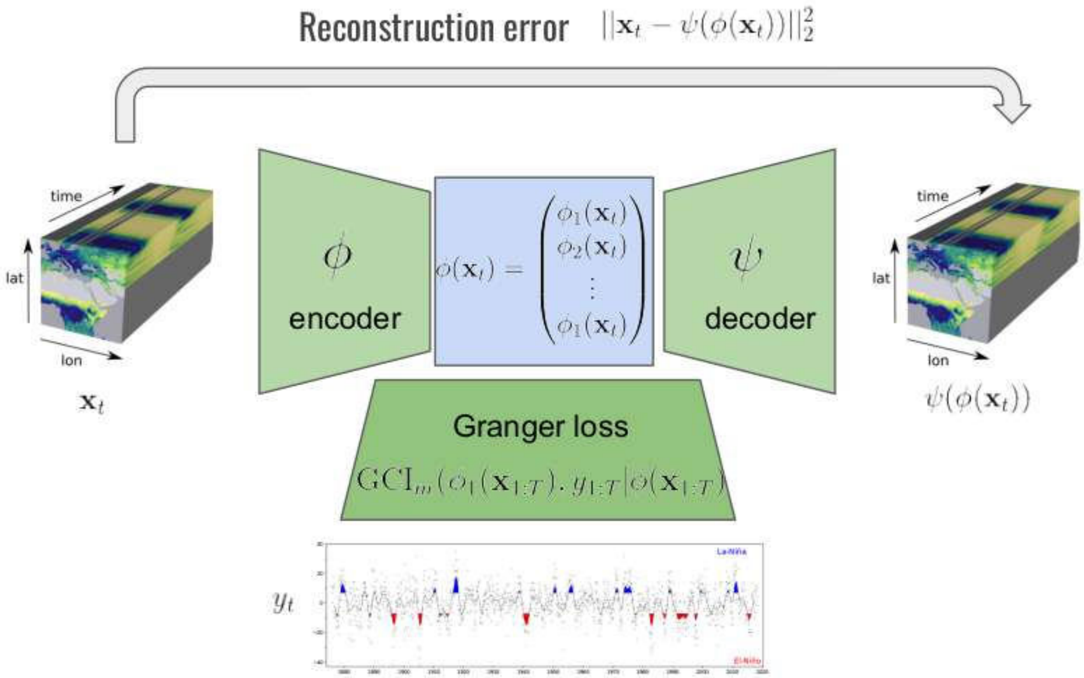
- ENSO influences moisture, greenness & precipitation spatio-temp patterns
- Goal: Learn causal impact teleconnections of ENSO on greenness
- NDVI from MODIS in Africa, linear interp, anomalies
- ENSO3.4 index, focus on 2014-2017





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

Learning causal feature representations

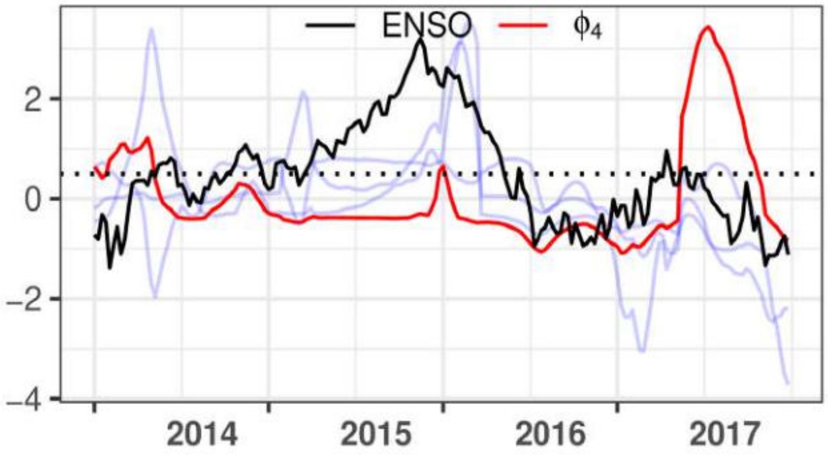


$$\frac{1}{T}\sum_{t=1}^{T} ||\mathbf{x}_t - \psi(\phi(\mathbf{x}_t))||_2^2 - \beta\, \mathrm{GCI}_m(\phi_1(\mathbf{x}_{1:T}), y_{1:T} | \phi(\mathbf{x}_{1:T}))$$

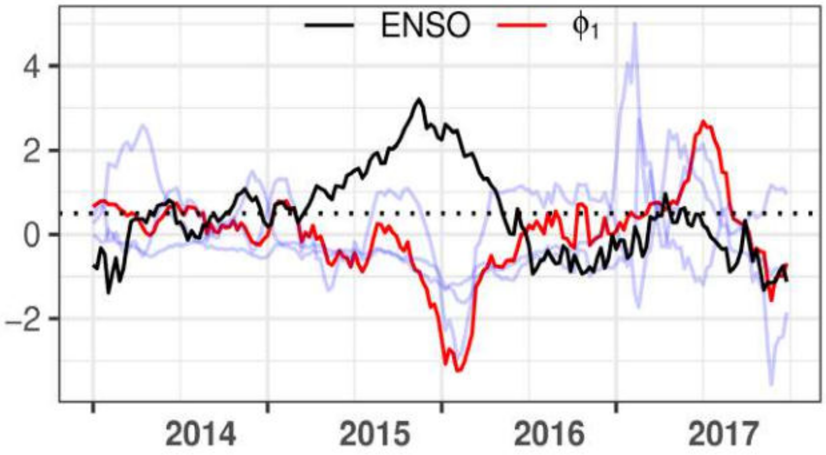
Learning Granger Causal Feature Representations, Varando, Fernandez, Camps-Valls, ICML 2021.

Learning causal feature representations

No Granger penalization $\beta = 0$



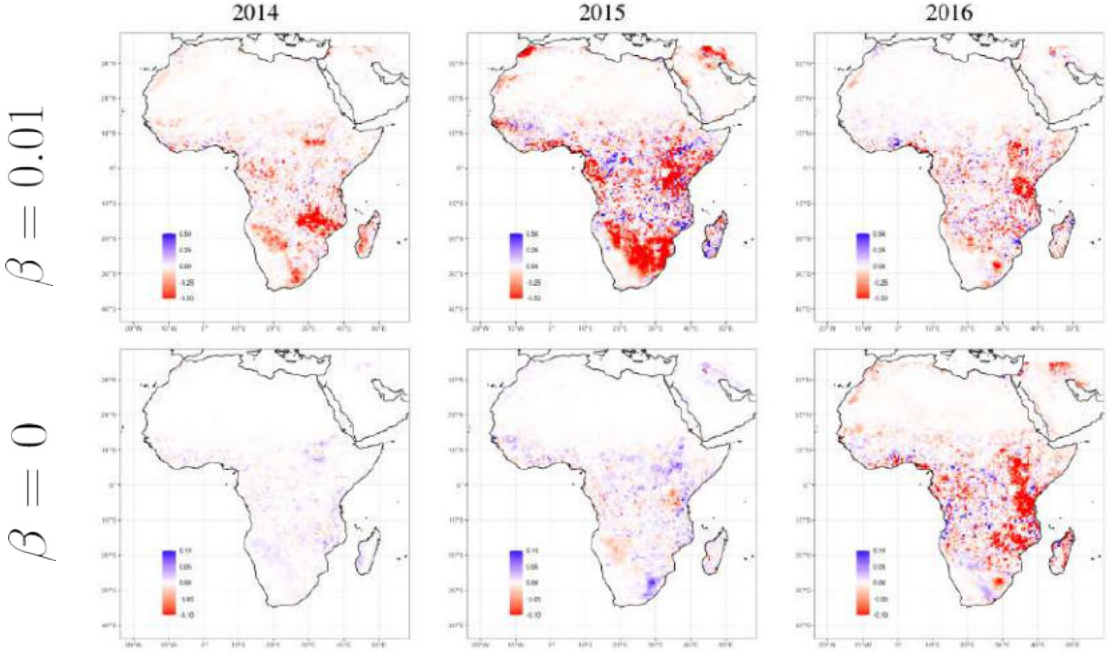
Granger penalization $\beta = 0.01$

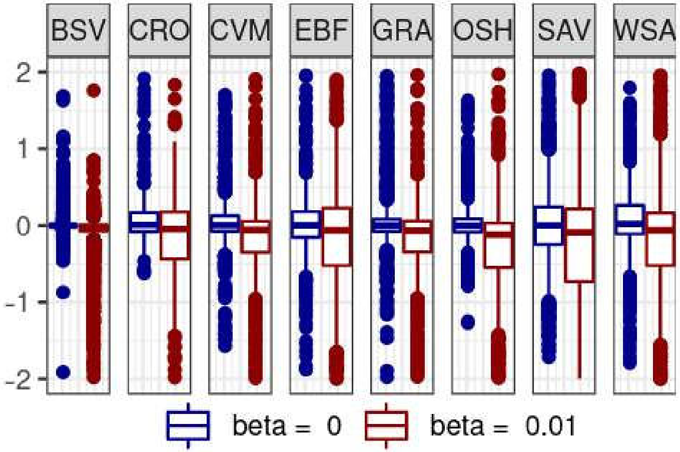


Learning Granger Causal Feature Representations, Varando, Fernandez, Camps-Valls, ICML 2021.

Learning causal feature representations

- XAI (integrated gradients) on the Granger Autoencoder
- Spatio-temporal explicit attributions

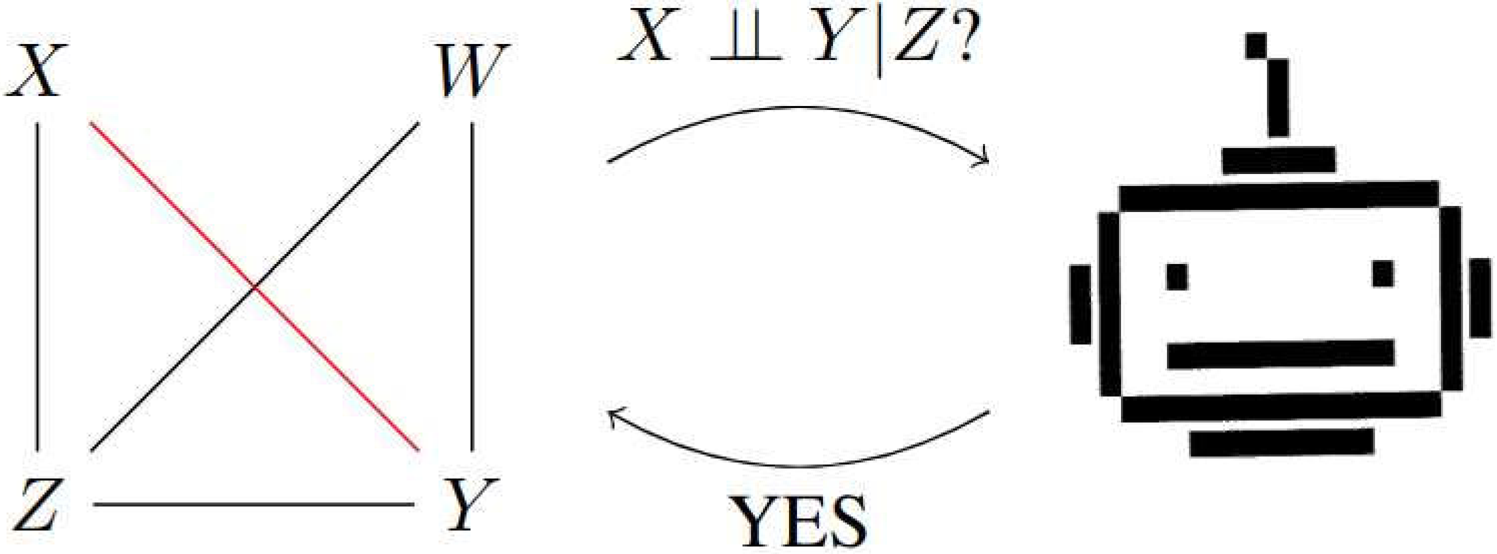




Learning Granger Causal Feature Representations, Varando, Fernandez, Camps-Valls, ICML 2021.

3. Causal discovery with LLMs

Large Language Models for Constrained-Based Causal Discovery



"Large Language Models for Constrained-Based Causal Discovery" K-H Cohrs, G. Varando, G. Camps-Valls, AAAI 2024

Large Language Models for Constrained-Based Causal Discovery

Persona specification
Instructions
Context
Variables description
CI Statement question
Response template

```
system: You are a helpful expert in {field}
and willing to answer questions.

system: You will be asked to provide your
best guess and your uncertainty on the
statistical independence between two
variables potentially conditioned on a set
of variables. Your answer should not be
based on data or observations but on
available knowledge. Even when unsure or
uncertain, provide your best guess (YES or
NO) and the probability that your guess is
correct. Answer only in the required format.

user: {context} Consider the following
variables:
{variables list and description}
is {x} independent of {y} given {z}?

system: Work out the answer in a
step-by-step way to be as sure as possible
that you have the right answer. After
explaining your reasoning, provide the
answer in the following form: [<ANSWER>
(<PROBABILITY>)] where ANSWER is either YES
or NO and PROBABILITY is a percentage
between 0\% and 100\%.
YES stands for "{x} is independent of {y}
given {z}" and NO stands for "{x} is not
independent of {y} given {z}".

For example [NO (50%)] or [YES (50%)].
```

"Large Language Models for Constrained-Based Causal Discovery" K-H Cohrs, G. Varando, G. Camps-Valls, AAAI 2024

Impact on food insecurity

"A total of 6.5 million people face acute food insecurity amid the driest conditions in 40 years(...) A total of 1.84 million children under 5 face acute malnutrition. (...) over 1.5 million drought-driven displacements since the start of the climate crisis."

World Food Programme, Jan 2023

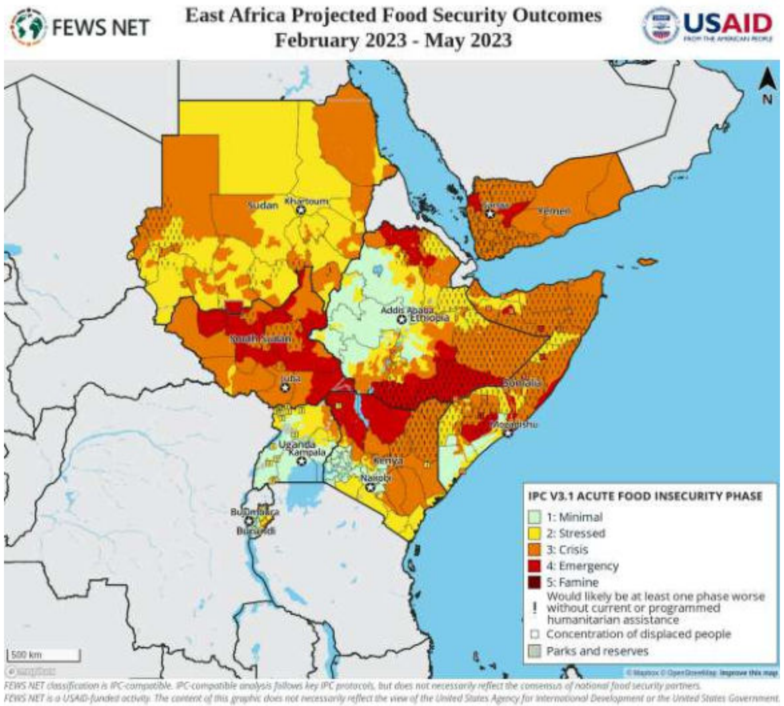
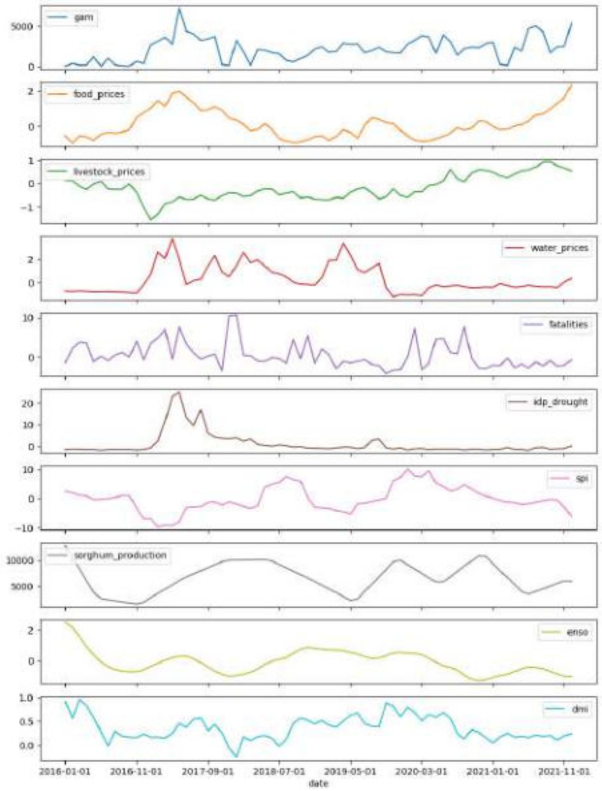


Image credits to: FEWS NET, https://fews.net

Impact on food insecurity

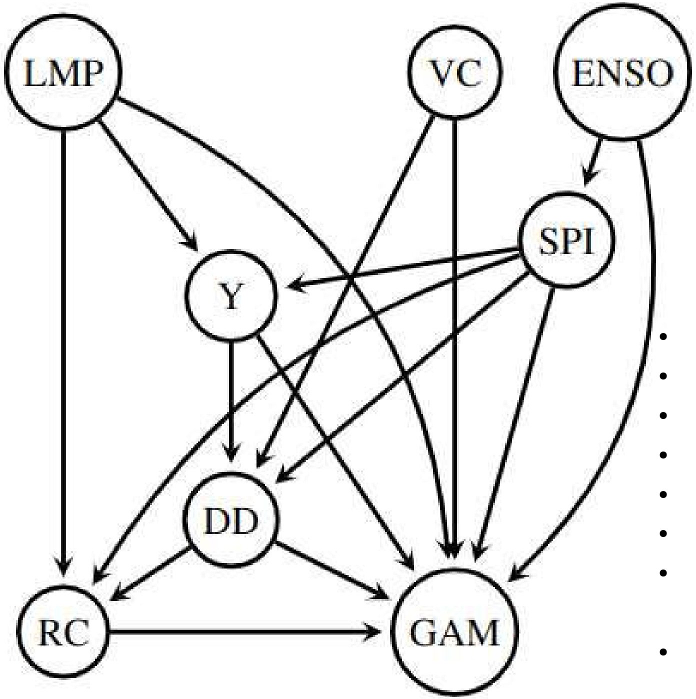
- Monthly data
- 2016 - 2021
- 37 districts
- N~70
- Market/food/livestock/ water prices, displaced people, fatalities, climate variables, humanitarian aid
- Target: malnutrition

Baidoa District



Large Language Models for Constrained-Based Causal Discovery

- Find traces of causal reasoning in model's answers
- Promising, alternative avenue for automated causality
- Useful for fast response, scarce data regimes



- El Niño Southern Oscillation (ENSO)
- Standardized Precipitation Index (SPI)
- Fatalities due to conflicts (VC)
- Local market prices (LMP)
- Sorghum yield production (Y)
- Drought-induced IDP (DD)
- People receiving cash from humanitarian aid (RC)
- Global Acute Malnutrition (GAM).

"Large Language Models for Constrained-Based Causal Discovery" K-H Cohrs, G. Varando, G. Camps-Valls, AAAI 2024

Part IV
Conclusion

Take-home messages

- **Many challenges:** emulate, learn representations, ensure consistency, interpretability, discover causal relations
- **Take 1:** Understanding processes by blending domain knowledge & data
- **Take 2:** Understanding complex systems means answering causal queries