Applied Statistics in the Climate Sciences
A mild overview

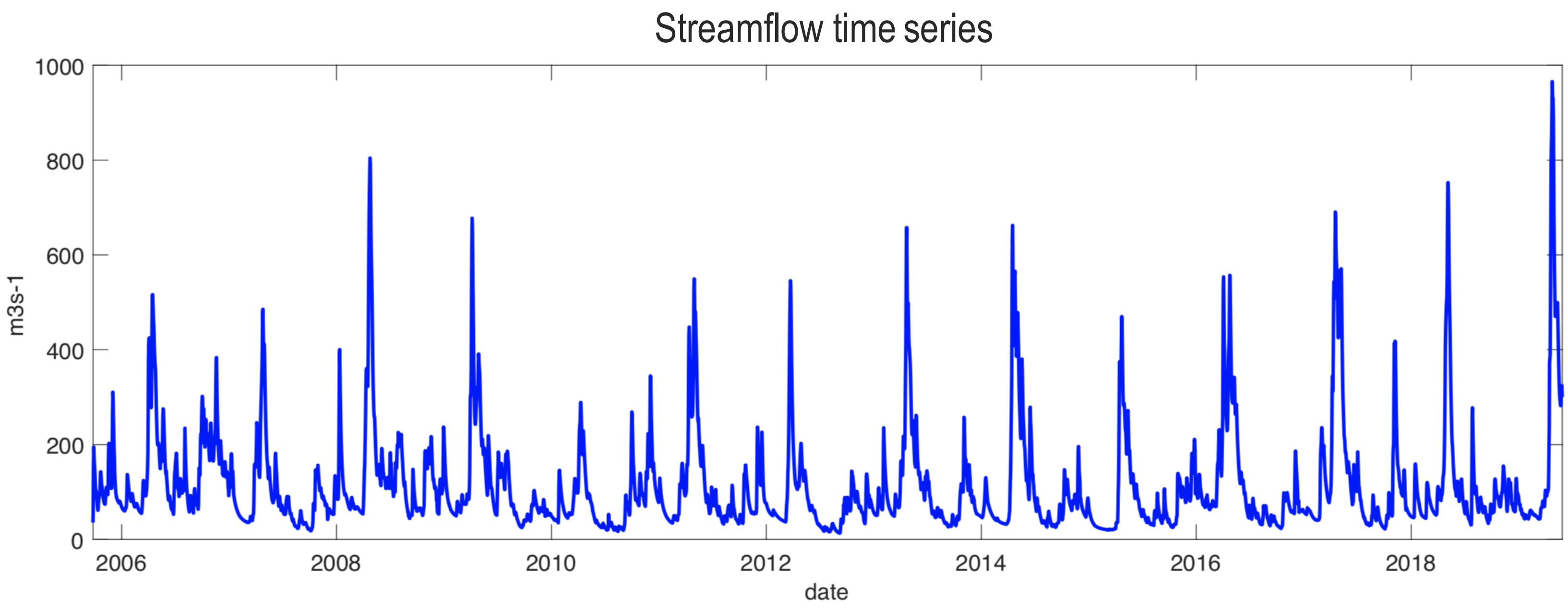
Alexis Hannart

31 October 2019

Outline

- General considerations
- Statistical Methods & Illustrations
- Conclusion

Randomness and determinism



- Is the Rouge river streamflow deterministic or probabilistic ?

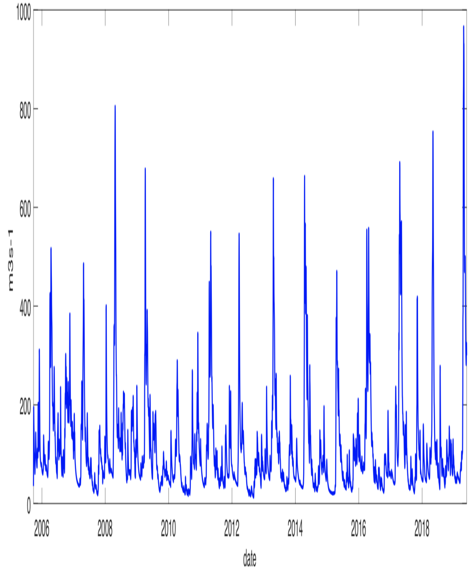
Randomness and determinism

- Is the climate system deterministic or probabilistic ?

Randomness and determinism

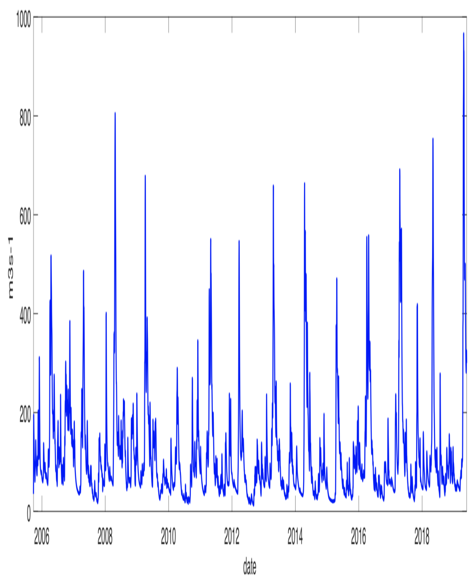
- Is coin tossing deterministic or probabilistic ?

Randomness and determinism



- What is deterministic and what is random ?

Randomness and determinism



- What is deterministic and what is random ?
- Is this question nonsense ?

Flipping a coin

$$\frac{d\vec{X}}{dt} = \vec{\Omega}(t) \times \vec{X}.$$

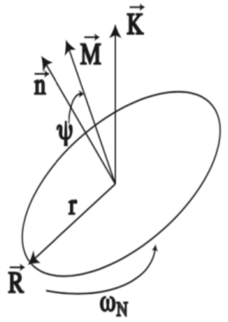


Figure 2: Coordinates of Precessing Coin.

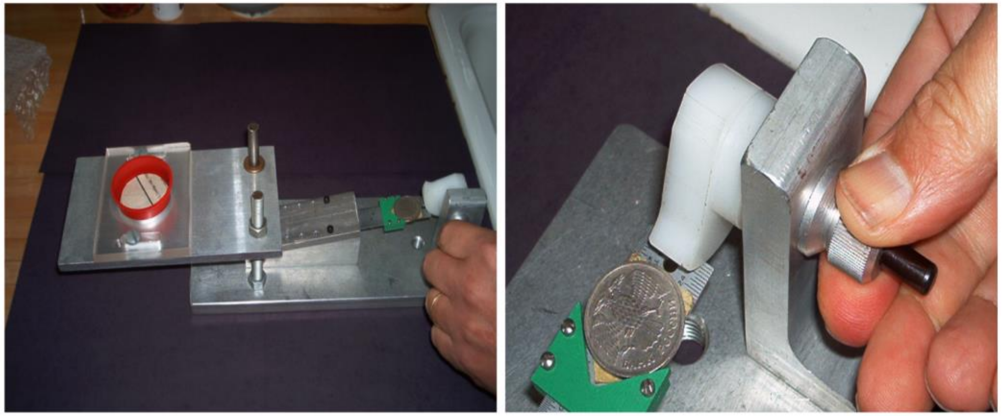


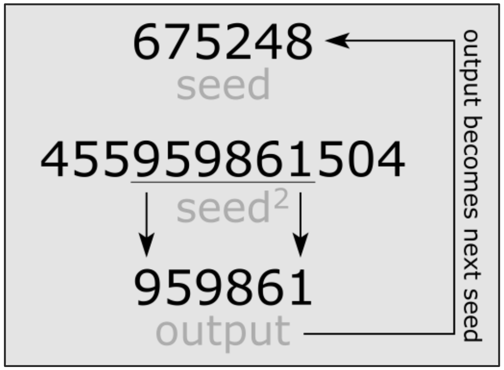
Figure 1.a Figure 1.b

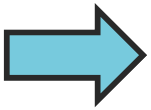
« We conclude that coin-tossing is ‘physics’, not ‘random’. »

Diaconis et al. 2007

Generating a « random » number

‘Middle-square’ algorithm





Deterministic, nonlinear dynamic system.
‘Pseudo-random’.

Deterministic versus Probabilistic

- Everything is deterministic, randomness does not exist in the real world.

- Chaos is not randomness, it is insufficient knowledge about the initial condition (and/or the boundary condition, and/or the dynamic).

⇒ Probabilities are a convenient mathematical tools to describe deterministic systems that are insufficiently known.

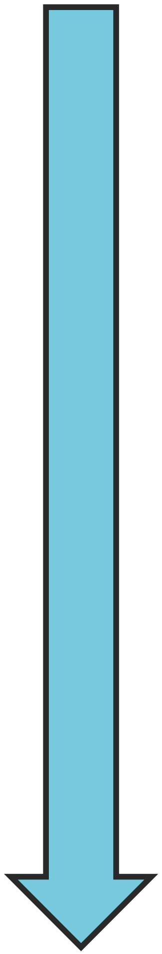
⇒ 'Probabilistic' is not a property of a system, but a modeling choice of the system's observer.

Outline

- General considerations
- Statistical Methods & Illustrations
- Conclusion

Theory and Application

Theory

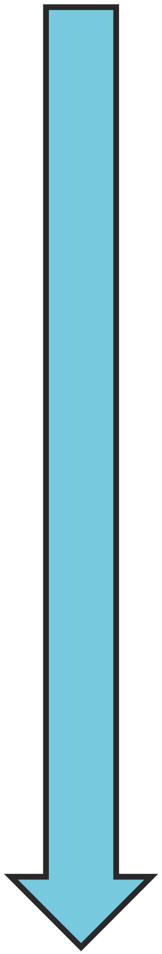


Application

- Mathematics

Theory and Application

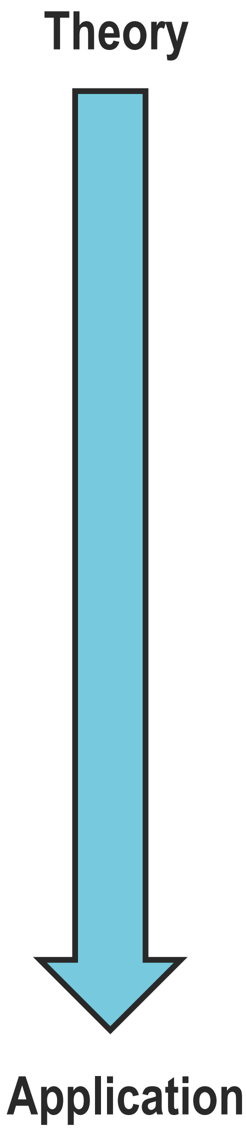
Theory



Application

- Mathematics
- Applied Mathematics

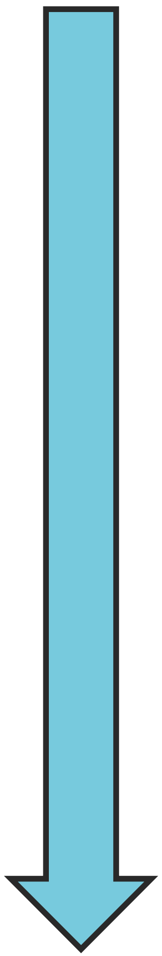
Theory and Application



- Mathematics
- Applied Mathematics
- Statistics

Theory and Application

Theory

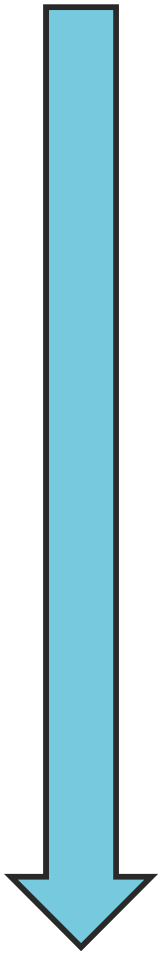


Application

- Mathematics
- Applied Mathematics
- Statistics
- Applied Statistics

Theory and Application

Theory



Application

- Mathematics
- Applied Mathematics
- Statistics
- Applied Statistics
- Application of Applied Statistics to theoretical problems
- Application of Applied Statistics to applied problems

Theory and Application

Theory

- Mathematics
- Applied Mathematics
- Statistics

- Applied Statistics
- Application of Applied Statistics to theoretical problems
- Application of Applied Statistics to applied problems

Application

Theory and Application

Theory → **Application**

- Mathematics
- Applied Mathematics
- Statistics

Theorems

- Applied Statistics

Data

- Application of Applied Statistics to theoretical problems
- Application of Applied Statistics to applied problems

Theory and Application

Theory

Application

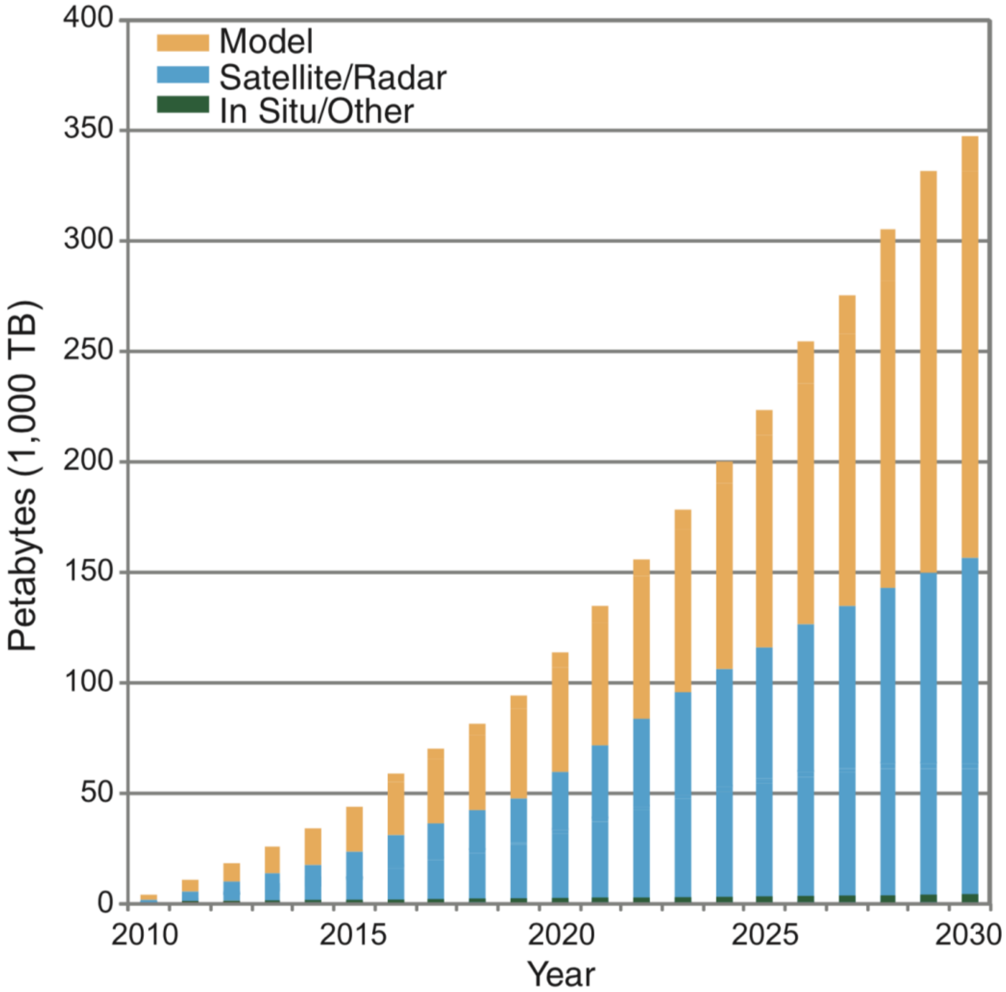
- Mathematics
- Applied Mathematics
- Statistics

- Applied Statistics

Data Science

- Application of Applied Statistics to theoretical problems
- Application of Applied Statistics to applied problems

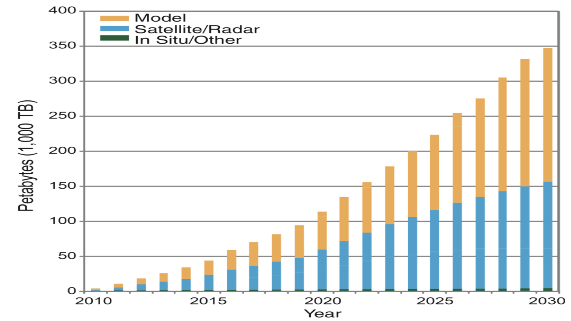
Data volume trend in climate science



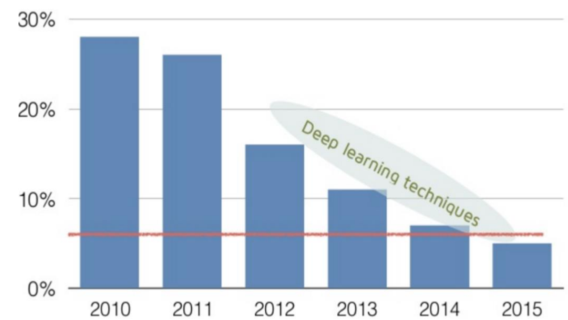
Overpeck et al. 2011

A simple story

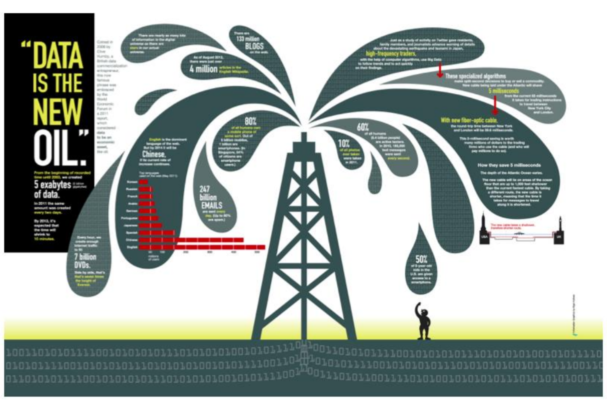
- Exponential trend on data generation and storage,



- Matched by smart algorithms and large computional power,

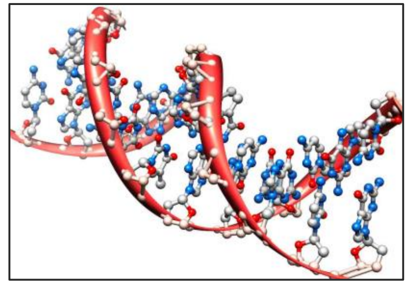


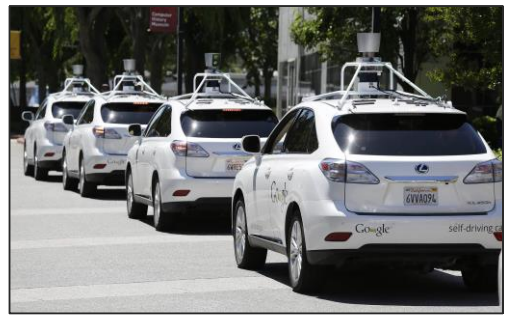
- New applications, products, services, and tools for science.

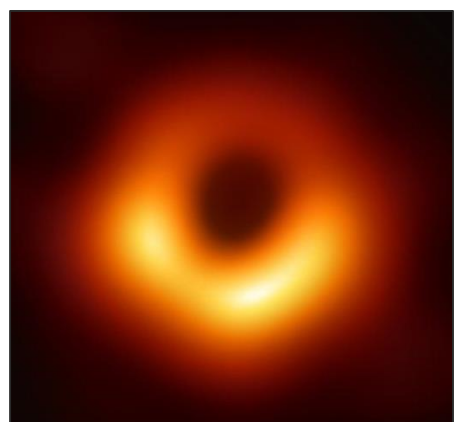


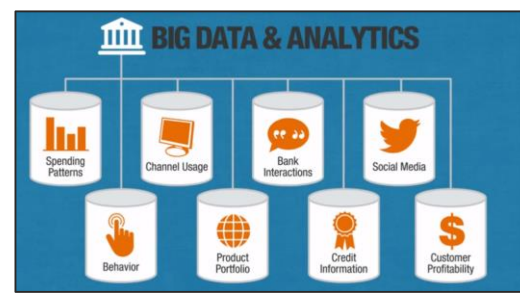
The AI 'fourth revolution'

- Search Engines & Internet
- Health & Genomics
- Astrophysics
- Banking & Finance
- Transport & Logistics
- Marketing & Media
- Energy & Distribution
- Agriculture & Forestry
- Urbanism



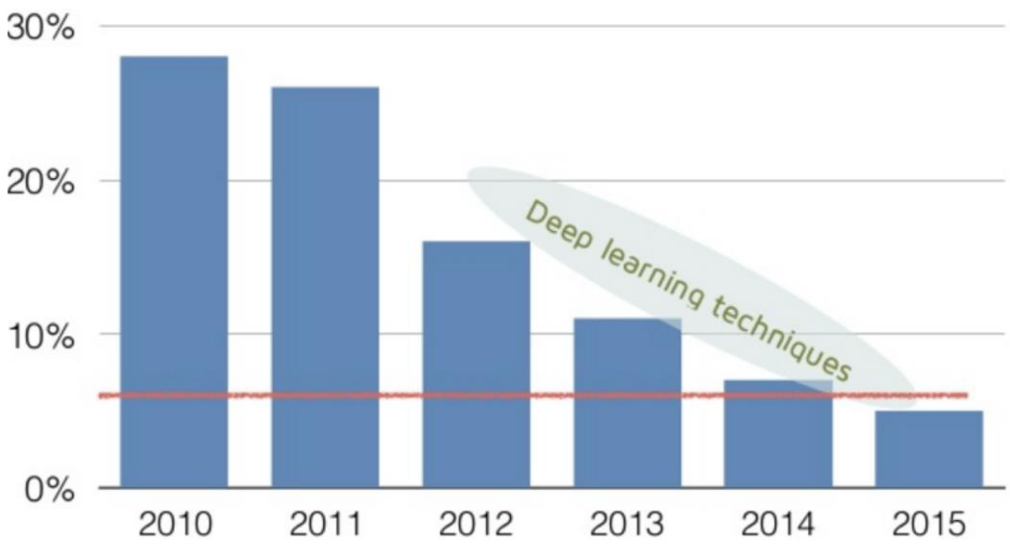


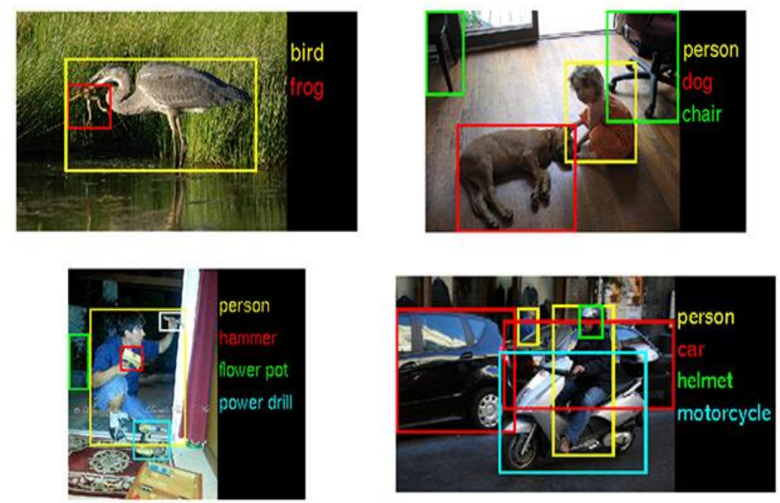




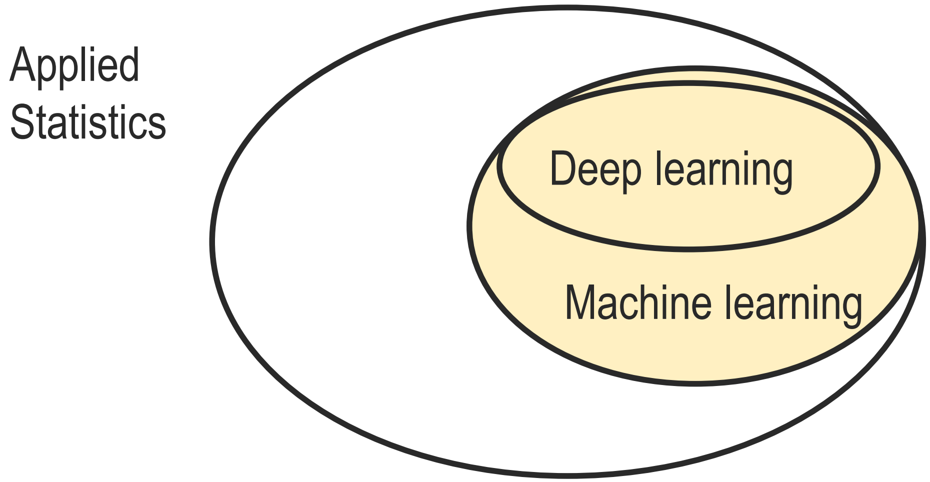
Skill trend in image recognition

ImageNet challenge





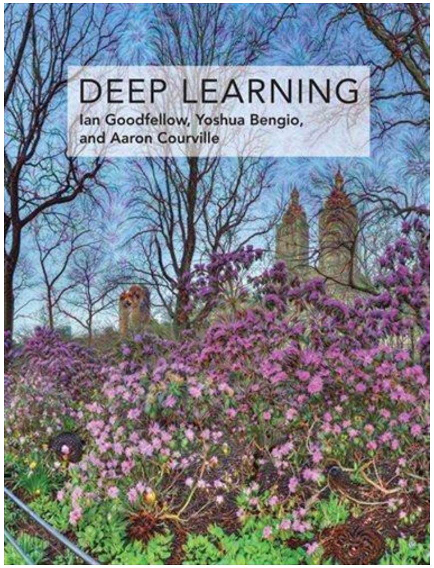
Applied Statistics and Machine learning



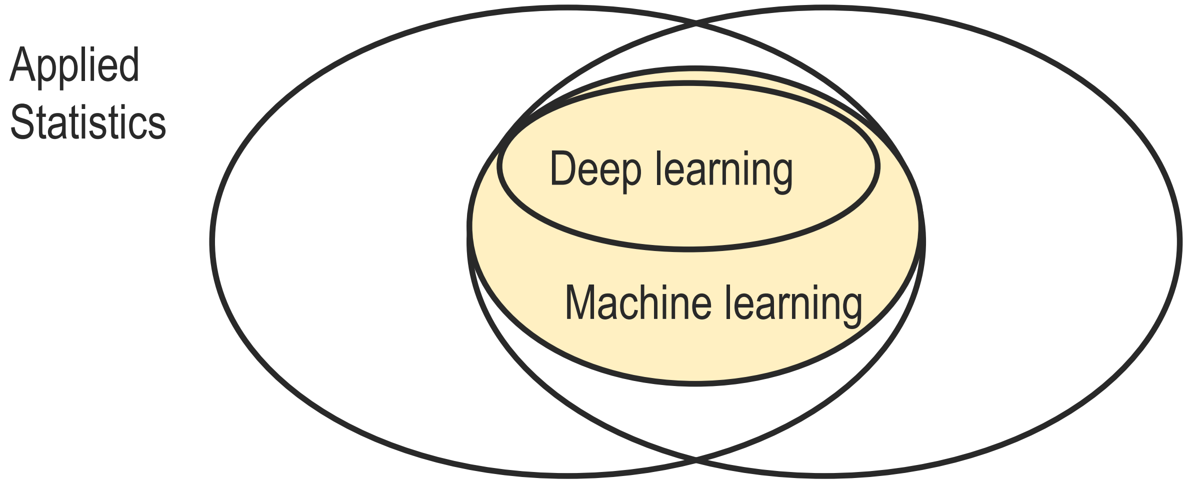
Machine learning is essentially a form of applied statistics:

- increased emphasis on the use of computers to statistically estimate complicated functions,
- decreased emphasis on proving confidence intervals around these functions.

Goodfellow I., Y. Bengio and A. Courville (2016) Deep Learning, MIT Press



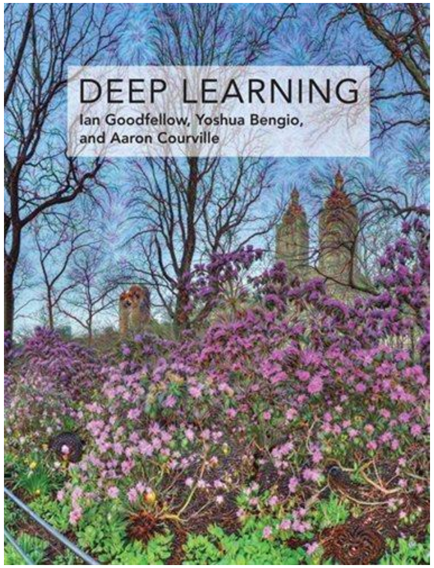
Applied Statistics, Machine learning and AI



Machine learning is essentially a form of applied statistics:

- increased emphasis on the use of computers to statistically estimate complicated functions,
- decreased emphasis on proving confidence intervals around these functions.

Goodfellow I., Y. Bengio and A. Courville (2016) Deep Learning, MIT Press



Theory and Application

Theory

- Mathematics
- Applied Mathematics
- Statistics
- Applied Statistics
- Application of Applied Statistics to theoretical problems
- Application of Applied Statistics to applied problems

Application

Theory and Application

Theory

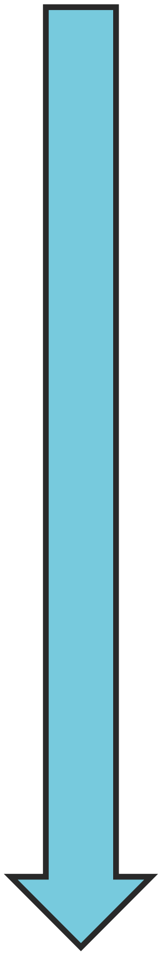
- Mathematics
- Applied Mathematics
- Statistics
- Applied Statistics
- Application of Applied Statistics to theoretical problems

- Application of Applied Statistics to applied problems

Application

Theory and Application

Theory



Application

- Mathematics
- Applied Mathematics
- Statistics
- Applied Statistics
- Application of Applied Statistics to theoretical problems

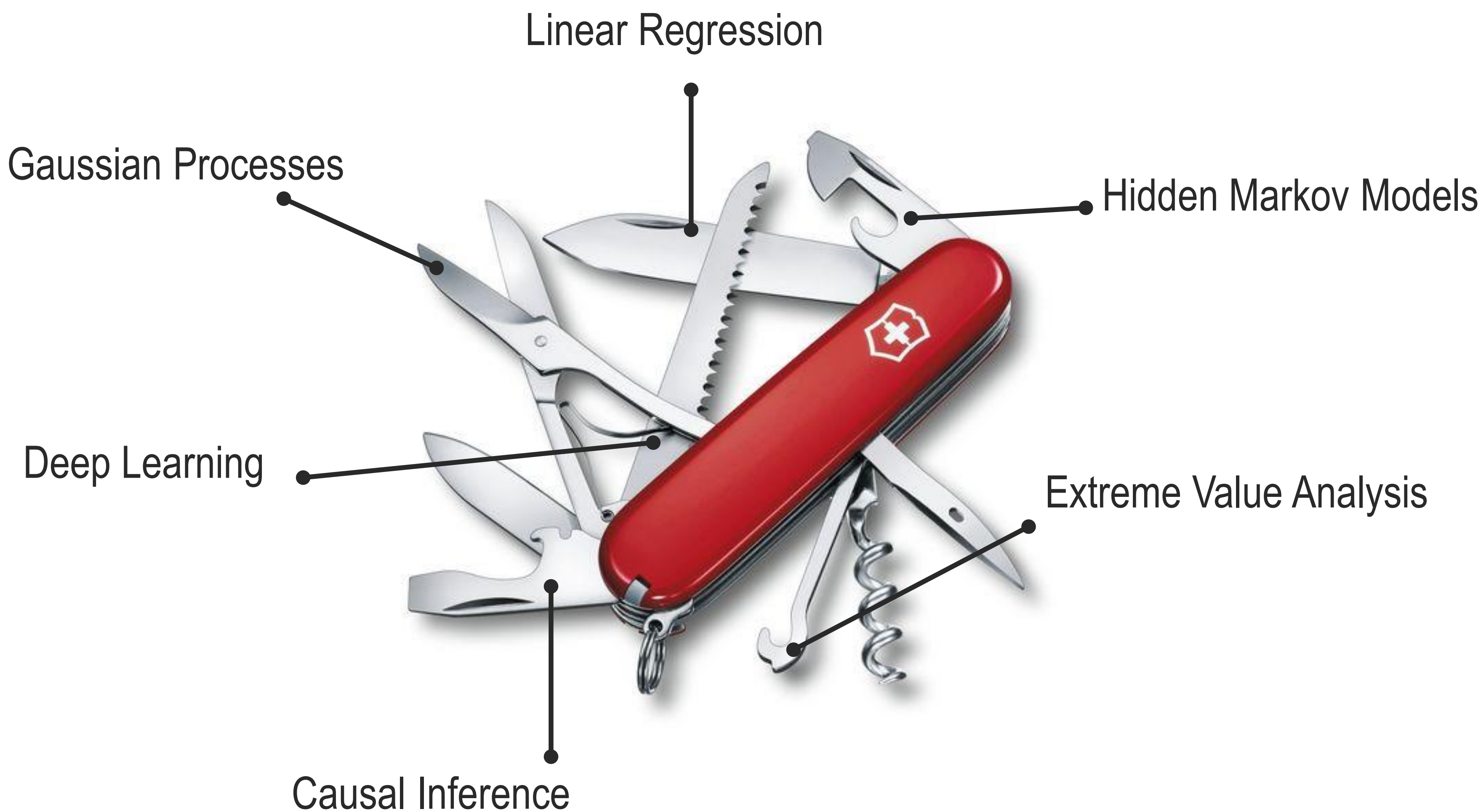
- Application of Applied Statistics to applied problems

Applied Statistics: a swiss knife

Applied Statistics: using existing tools

Applied Statistics: designing more tools

Applied Statistics: a few useful tools



Outline

- General considerations
- Statistical Methods & Illustrations
- Conclusion

Basic principles

$$p(x \mid \theta)$$

Basic principles

$$p(x \mid \theta)$$

$$\widehat{\theta} = \text{argmax}_{\theta} \; \log p(x \mid \theta)$$

Basic principles

$$p(x \mid \theta)$$

$$p(\theta)$$

Basic principles

$$p(x \mid \theta)$$

$$p(\theta)$$

$$p(\theta \mid x) \propto p(x \mid \theta) \,.\, p(\theta)$$

Basic principles

$$p(x \mid \theta)$$

$$p(\theta)$$

$$p(\theta \mid x) \propto p(x \mid \theta) \,.\, p(\theta)$$

$$\widehat{\theta} = \mathrm{argmax}_{\theta}\, p(\theta \mid x)$$

Basic principles

$$p(x \mid \theta)$$

$$p(\theta)$$

$$p(\theta \mid x) \propto p(x \mid \theta) \, . \, p(\theta)$$

$$\widehat{\theta} = \text{argmax}_{\theta} \, p(\theta \mid x)$$

$$\widehat{\theta} = \mathbb{E}(\theta \mid x)$$

Basic principles

$$p(x \mid \theta)$$

$$p(\theta)$$

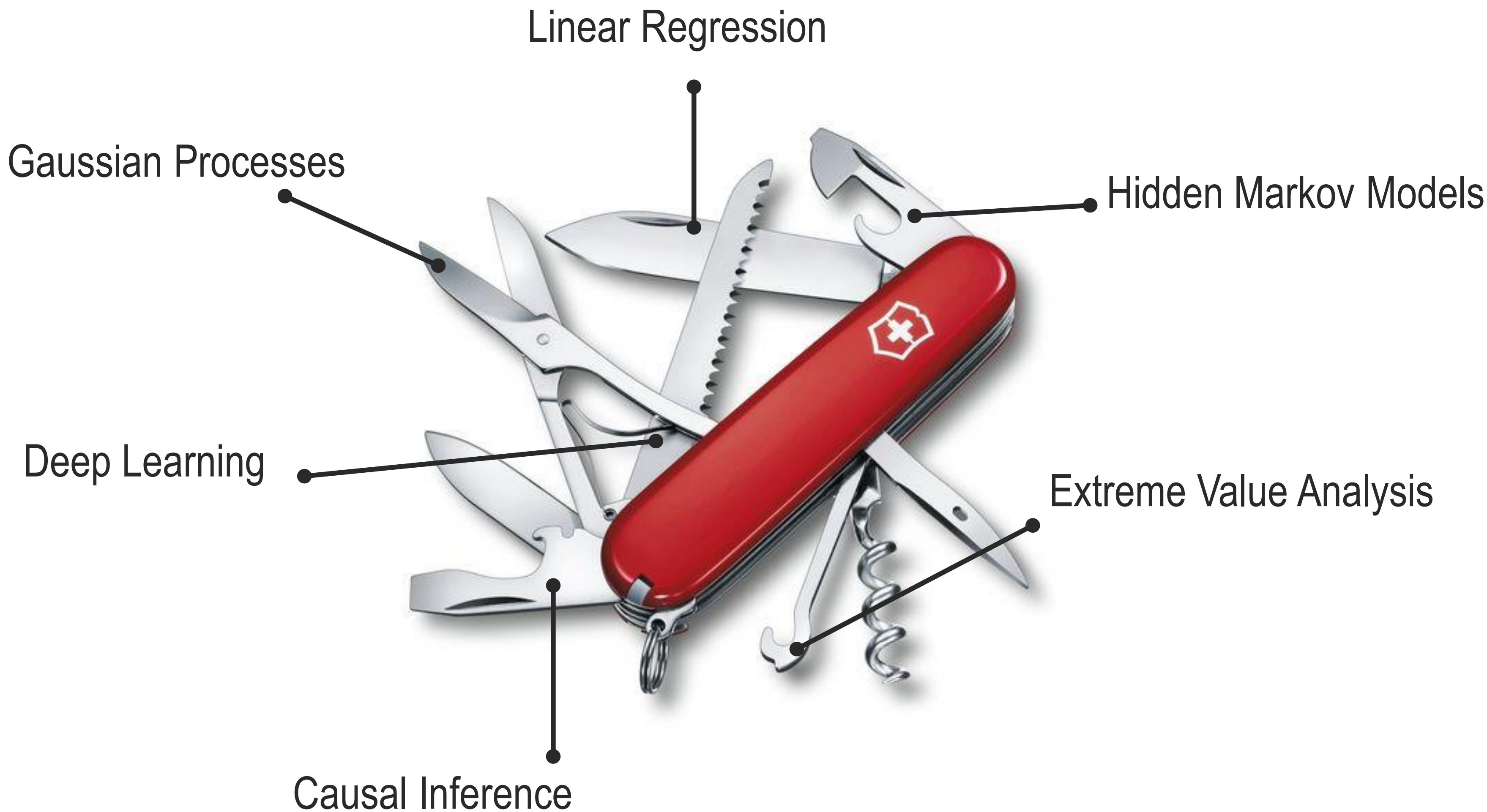
$$p(\theta \mid x) \propto p(x \mid \theta) \,.\, p(\theta)$$

$$\widehat{\theta} = \text{argmax}_{\theta}\; p(\theta \mid x)$$

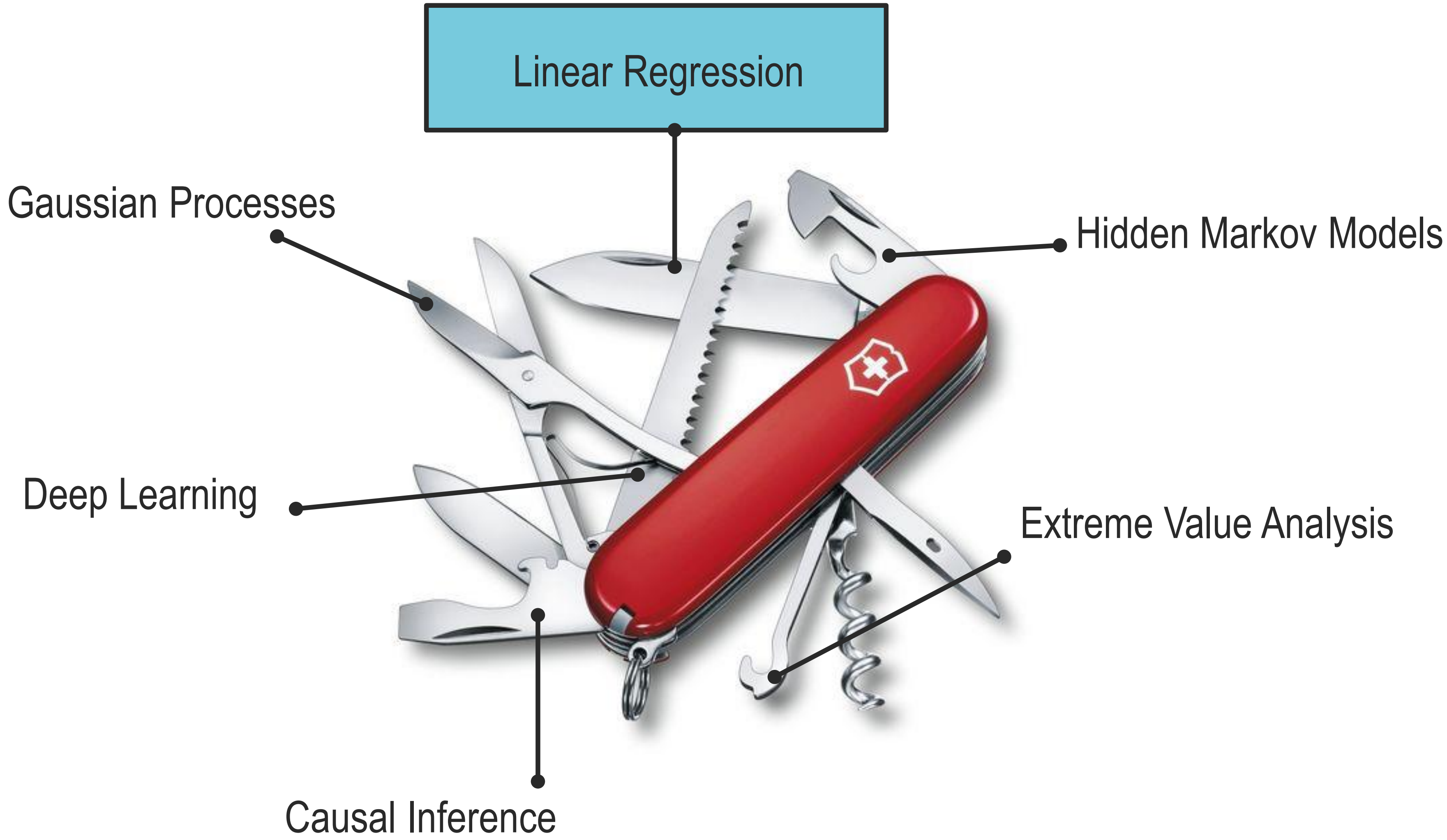
$$\widehat{\theta} = \mathbb{E}(\theta \mid x)$$

$$\widehat{\theta} = \text{argmax}_{\theta *}\; \mathbb{E}(\mathcal{C}(\theta, \theta^*) \mid x)$$

Outline


Linear Regression
Gaussian Processes
Hidden Markov Models
Deep Learning
Extreme Value Analysis
Causal Inference

Outline



Linear Regression
Gaussian Processes
Hidden Markov Models
Deep Learning
Extreme Value Analysis
Causal Inference

Linear regression

$$y = \boldsymbol{x}\beta + \varepsilon$$

$$p(y \mid \boldsymbol{x}, \beta) = \mathcal{N}(\boldsymbol{x}\beta, \sigma^2\mathbf{I})$$

$$\widehat{\beta} = (\boldsymbol{x}'\boldsymbol{x})^{-1}(\boldsymbol{x}'y)$$

Linear regression

Impact variable

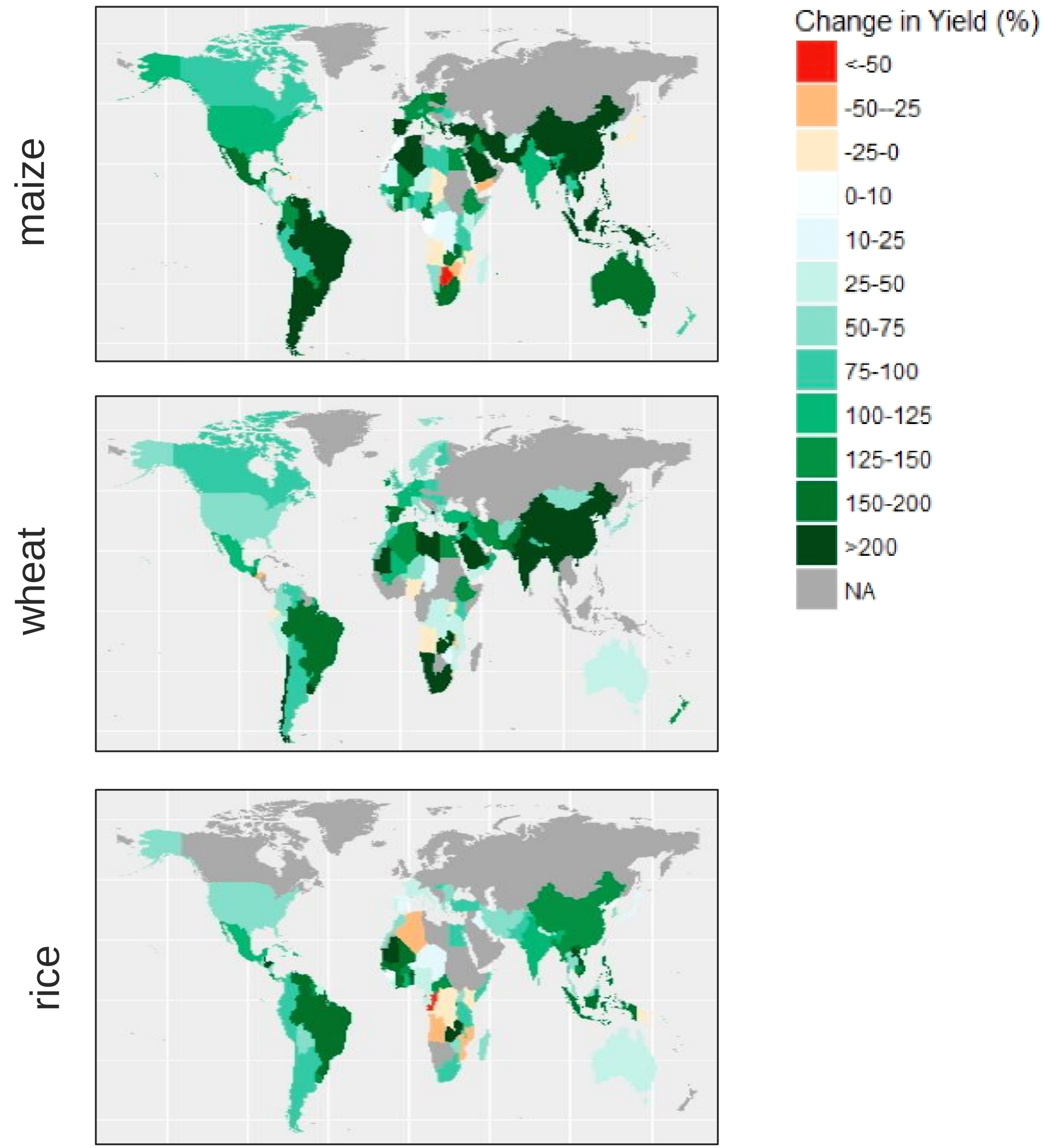
Climate variables

$$y = \boldsymbol{x}\beta + \varepsilon$$

$$p(y \mid \boldsymbol{x}, \beta) = \mathcal{N}(\boldsymbol{x}\beta, \sigma^2 \mathbf{I})$$

Data

<u>Observations:</u>
Yields by crop, year and country.

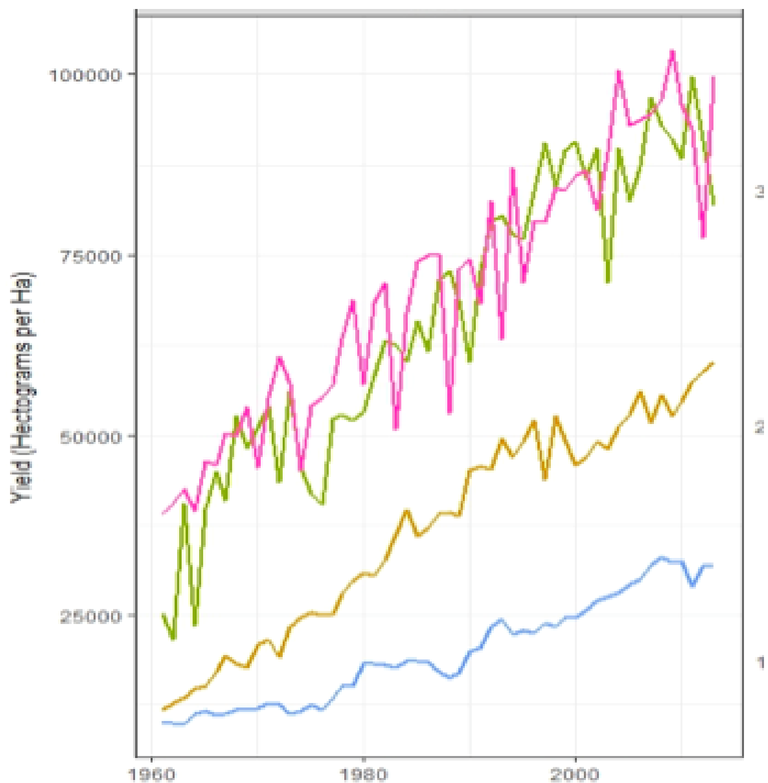


Data

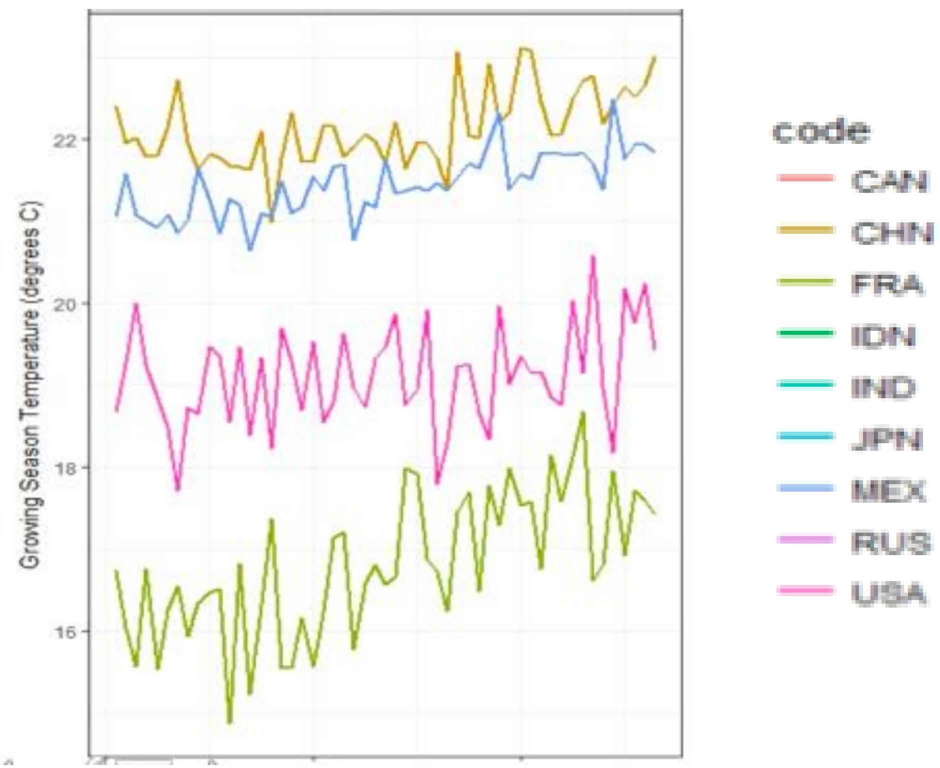
<u>Observations:</u>
Yields by crop, year and country (FAO).

<u>Observations:</u>
Growing season temperature by crop, year and country (HadCRUT).

yield (maize)



growing season temperature (maize)



Model

$$\log(yield)_{cit}$$

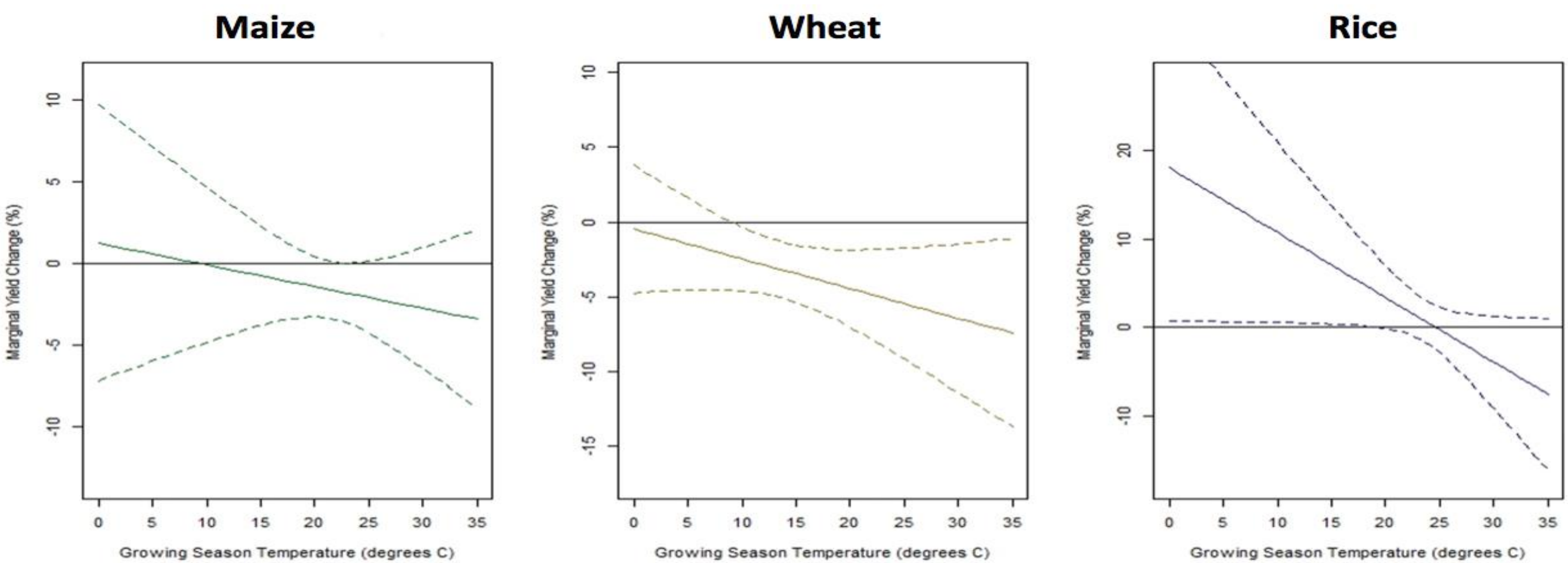
$$= \beta_{1c}T_{cit} * I_c + \beta_{2c}{T_{cit}}^2 * I_c + \beta_{3c}P_{cit} * I_c$$

$$+\beta_{4c}{P_{cit}}^2 * I_c + \beta_{5ci}t * I_c * I_i + \beta_{6ci}\, t^2 * I_c * I_i$$

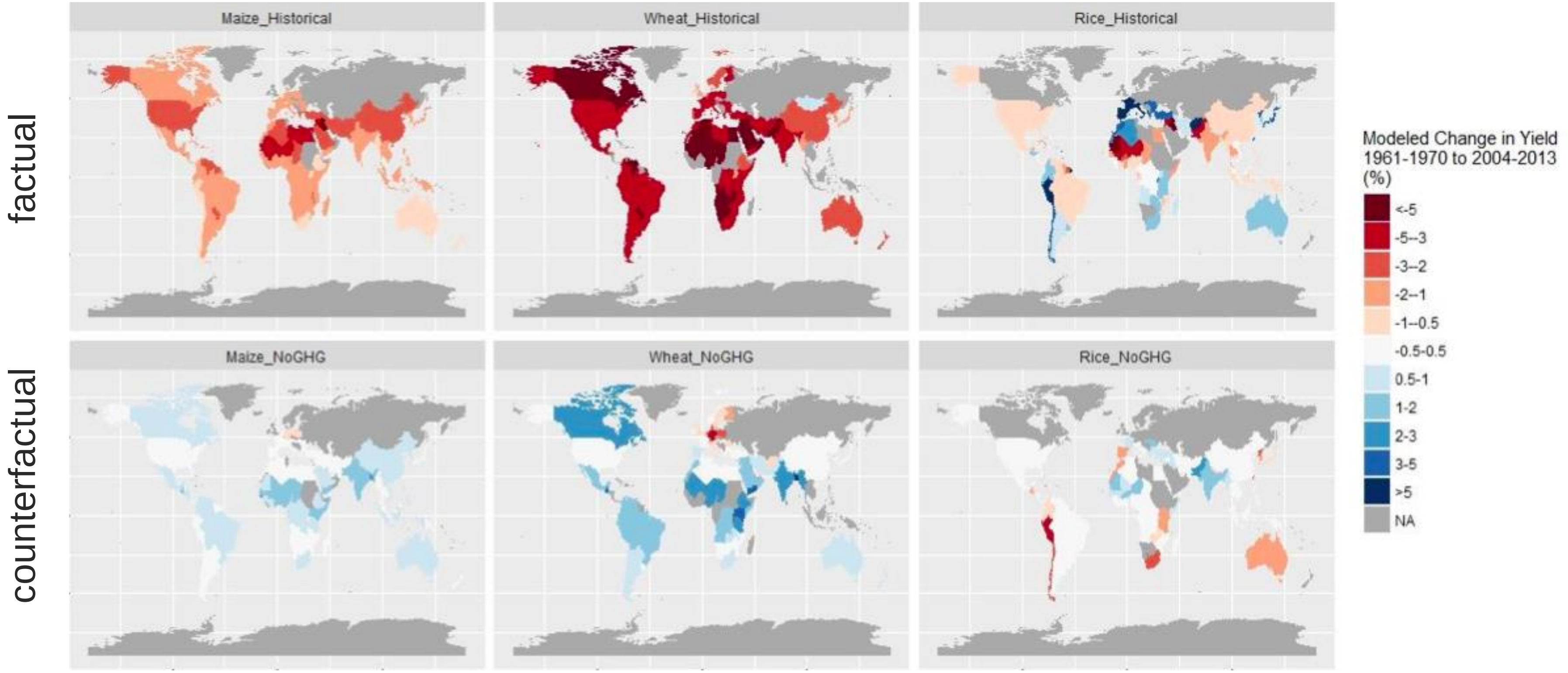
$$+Irr_{it} * I_c + Fert_{it} * I_c + \mu_{ci} + \delta_{tc} + \varepsilon_{cit}$$

Similar to:

- Lobell et al. (2011)
- Moore and Lobell (2015)
- Burney (2014)
- Heft-Neal et al. (2017)



Results



- Factual: Historical
- Counterfactual: No GHG

Linear regression

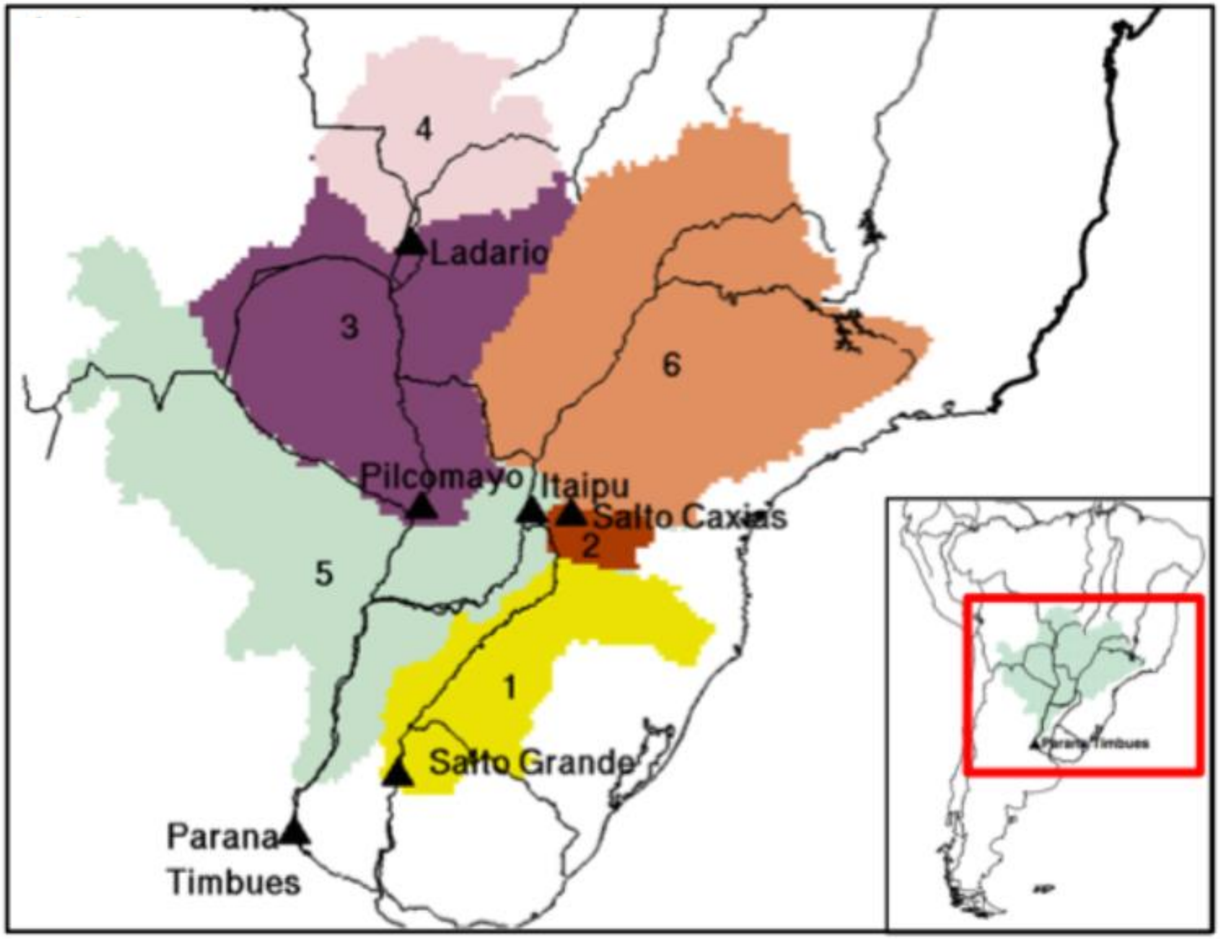
Impact variable

Climate variables

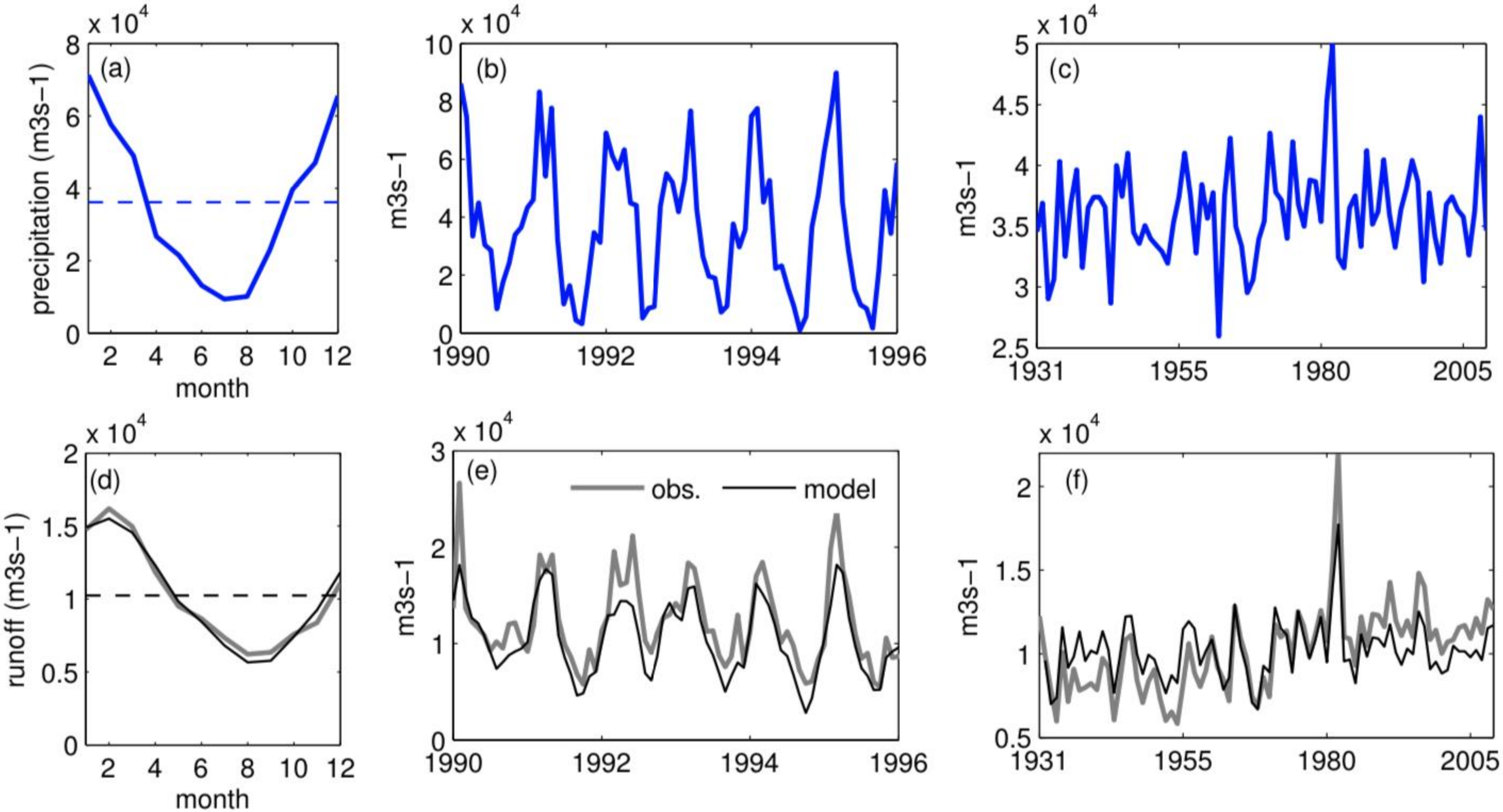
$$y = \boldsymbol{x}\beta + \varepsilon$$

$$p(y \mid \boldsymbol{x}, \beta) = \mathcal{N}(\boldsymbol{x}\beta, \sigma^2\mathbf{I})$$

Basins

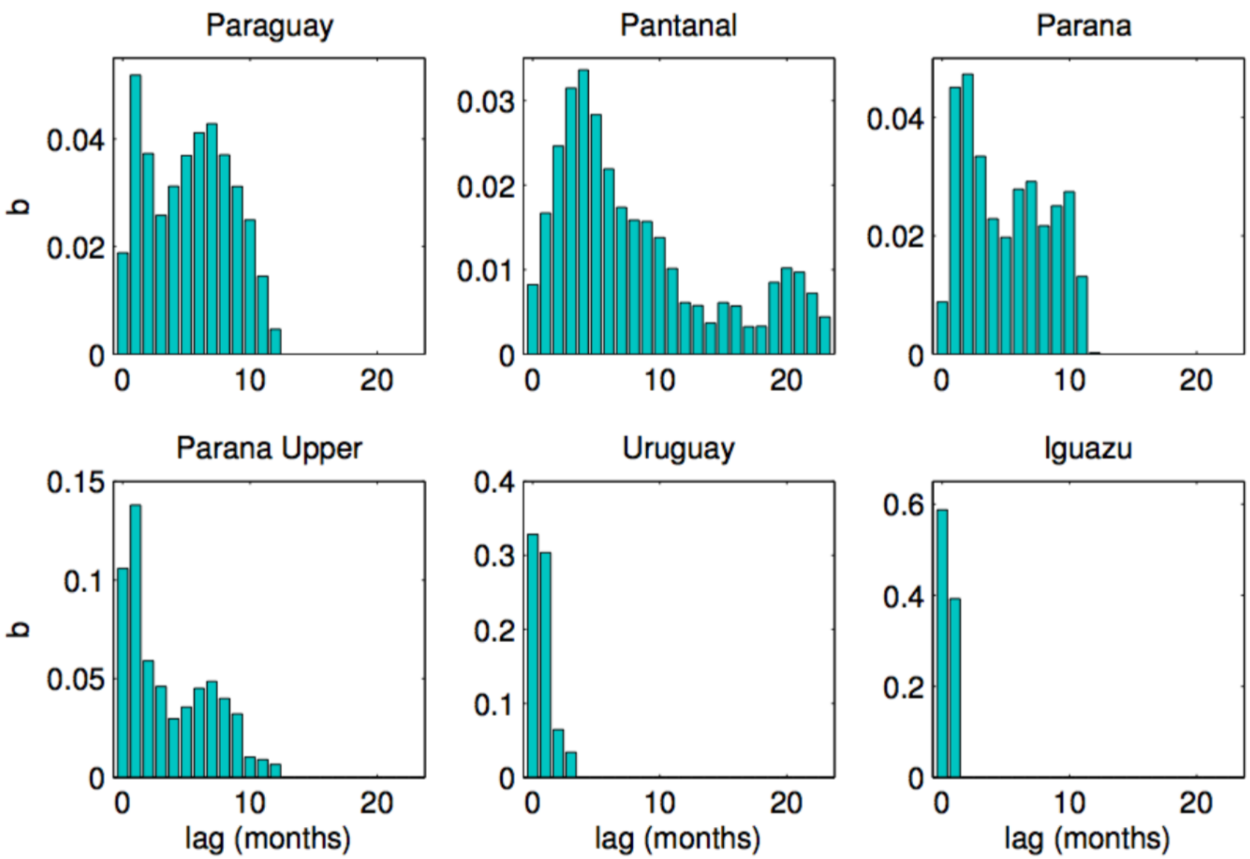

4
Ladario
3
6
Pilcomayo
Itaipu
Salto Caxias
2
5
1
Salto Grande
Parana
Timbues
Parana Timbues

Data



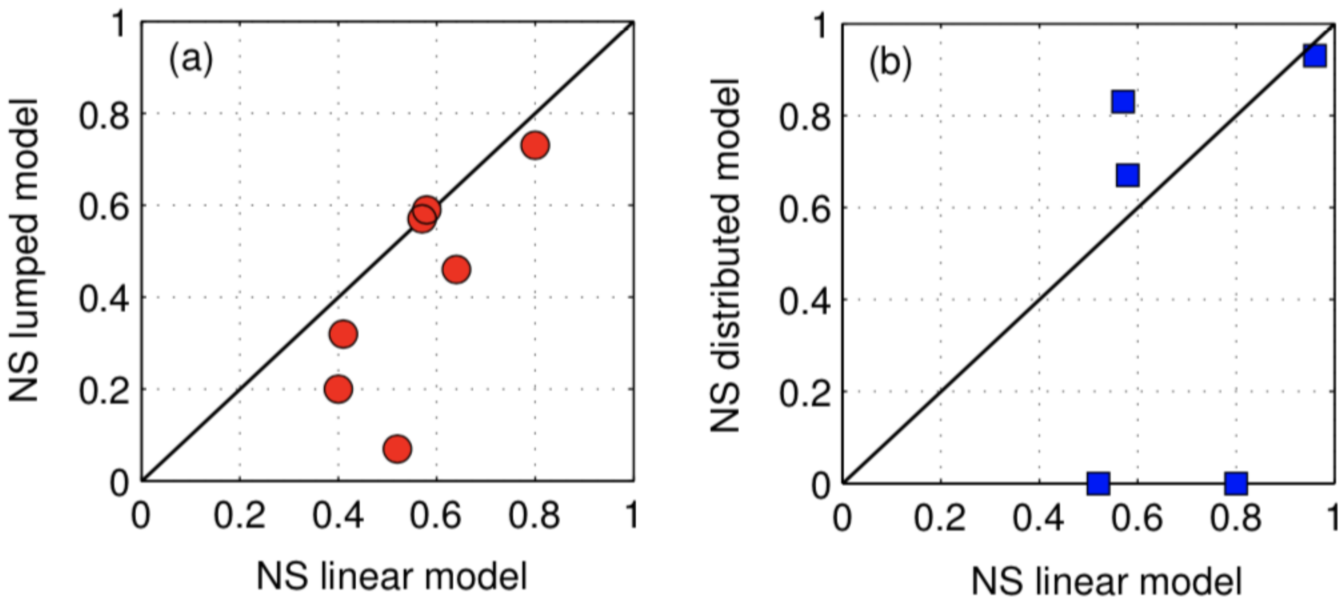
x 10^4
(a)
precipitation (m3s−1)
month
(b)
m3s−1
(c)
m3s−1
(d)
runoff (m3s−1)
month
(e)
obs.
model
m3s−1
(f)
m3s−1
1990
1992
1994
1996
1931
1955
1980
2005

Results



Paraguay
Pantanal
Parana
Parana Upper
Uruguay
Iguazu
b
b
0.04
0.02
0
0.03
0.02
0.01
0.15
0.1
0.05
0.4
0.3
0.2
0.1
0.6
0.4
0.2
0
10
20
lag (months)
lag (months)
lag (months)

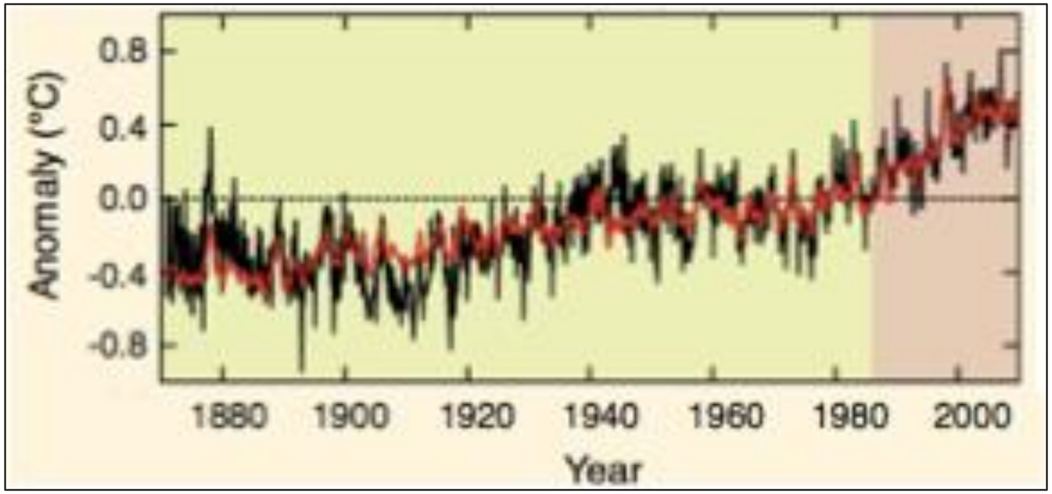
Skill



(a)
NS lumped model
NS linear model
0
0.2
0.4
0.6
0.8
1
(b)
NS distributed model
NS linear model

Attribution

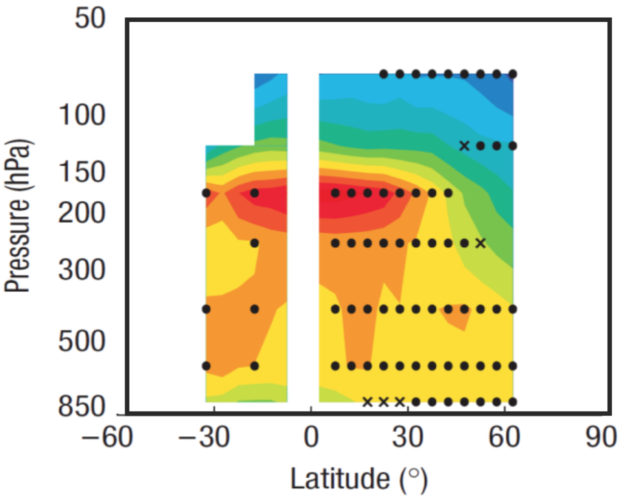
Evidencing the causal influence of external factors



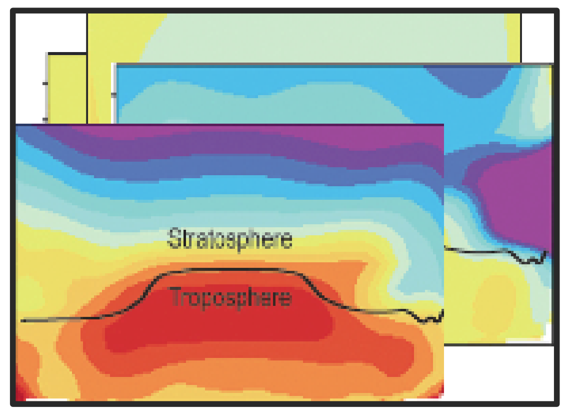
Conventional method for attributing trends

$$y = x\beta + \nu$$

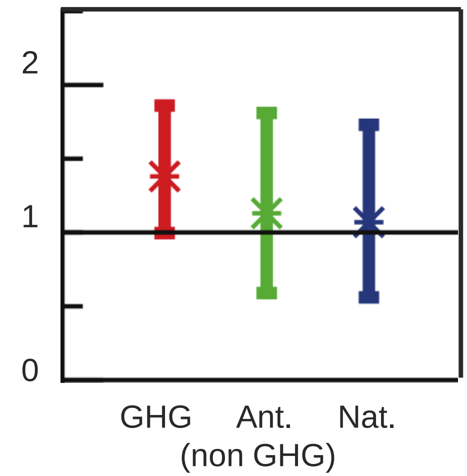
Observations



GCM patterns



Coefficients

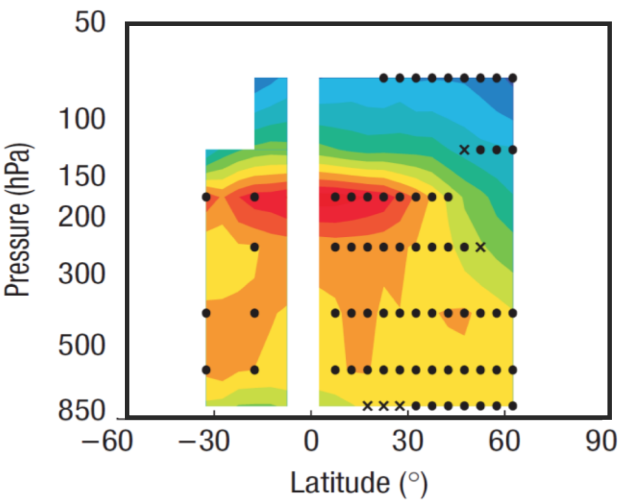


Hasselmann 1993
Hegerl et al. 1996
Allen and Tett 1999
Allen and Stott 2003

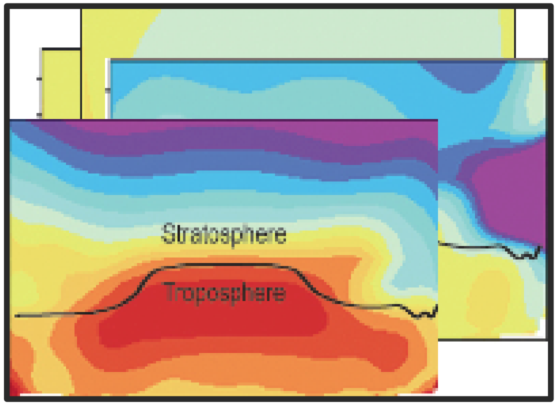
Conventional method for attributing trends

$$y = x\beta + \nu$$

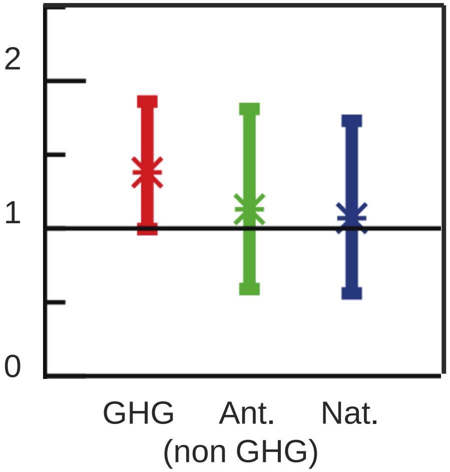
Observations



GCM patterns



Coefficients



Hasselmann 1993
Hegerl et al. 1996
Allen and Tett 1999
Allen and Stott 2003

Ribes et al. 2012
Hannart et al. 2014
Hannart 2016
Katzfuss et al. 2017
Hannart 2018b
More to come.

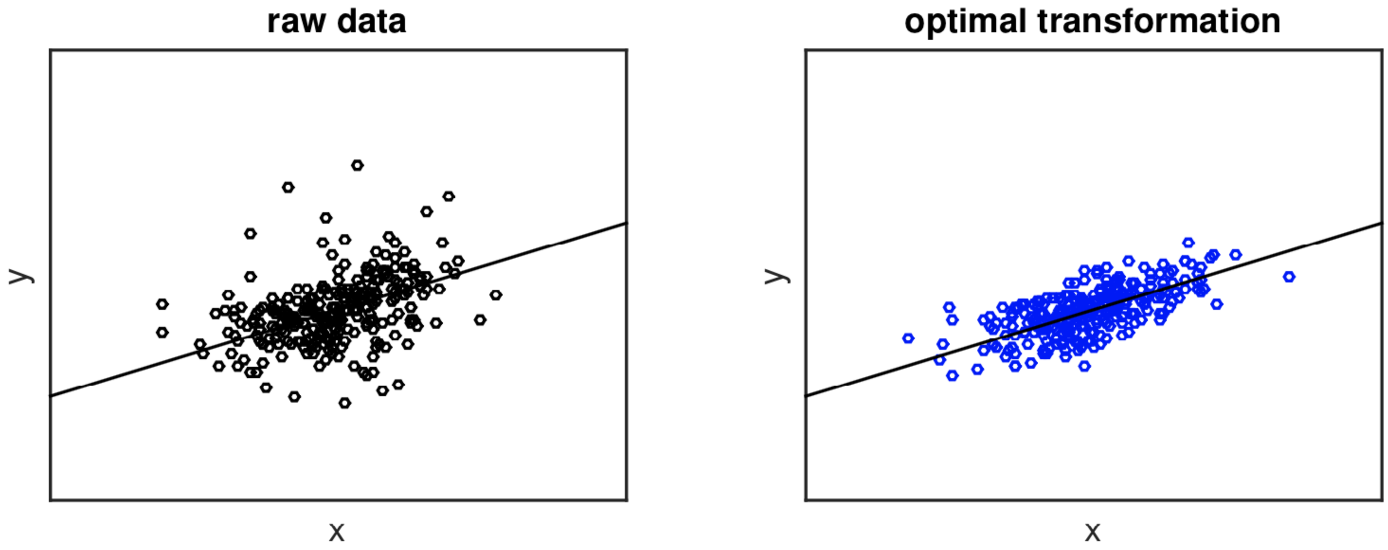
Linear regression model

$$\begin{cases} y = \boldsymbol{x}\beta + \nu \\ \mathrm{Var}(\nu) = \boldsymbol{\Sigma} \\ \boldsymbol{x} = (x_1, \ldots, x_p) \end{cases}$$

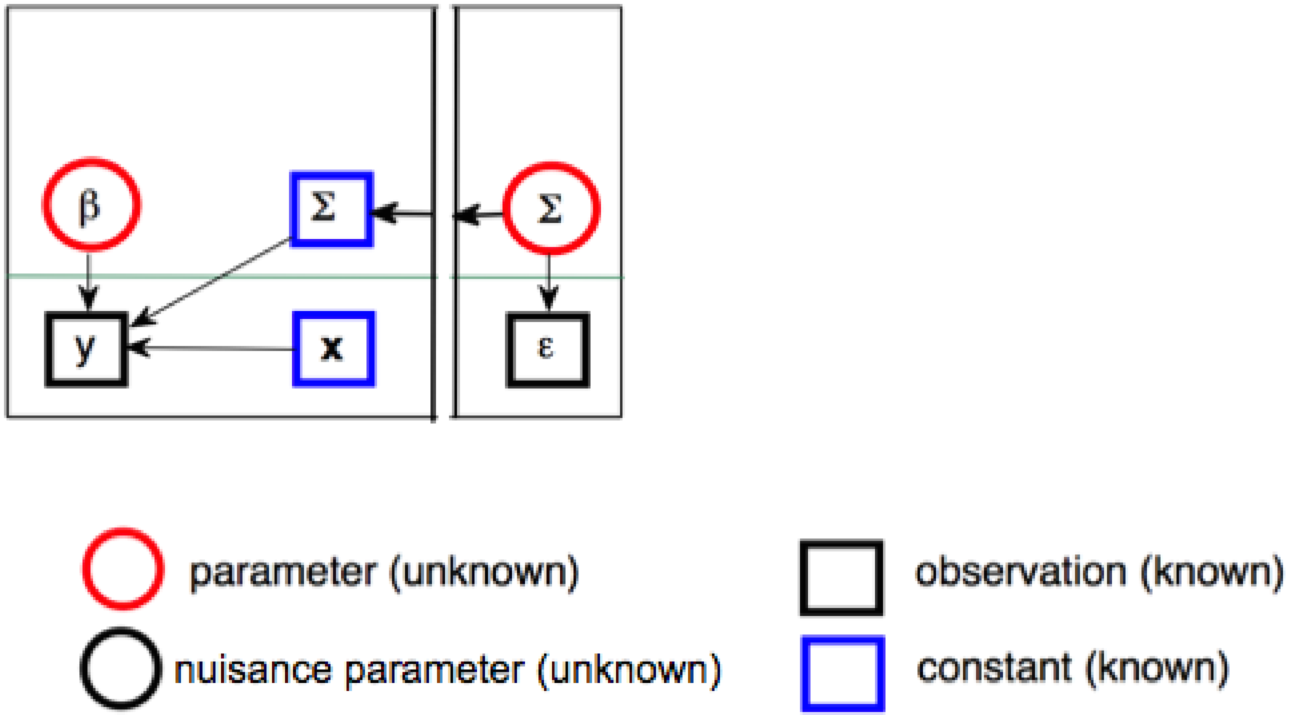
<u>Inference: projection of the data</u>

$$\Rightarrow \begin{cases} \mathbf{T}y = \mathbf{T}\boldsymbol{x}\beta + \mathbf{T}\nu \\ \mathbf{T}\boldsymbol{\Sigma}\mathbf{T}' = \mathbf{I} \\ \widehat{\beta} = (\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x})^{-1}(\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}y) \end{cases}$$

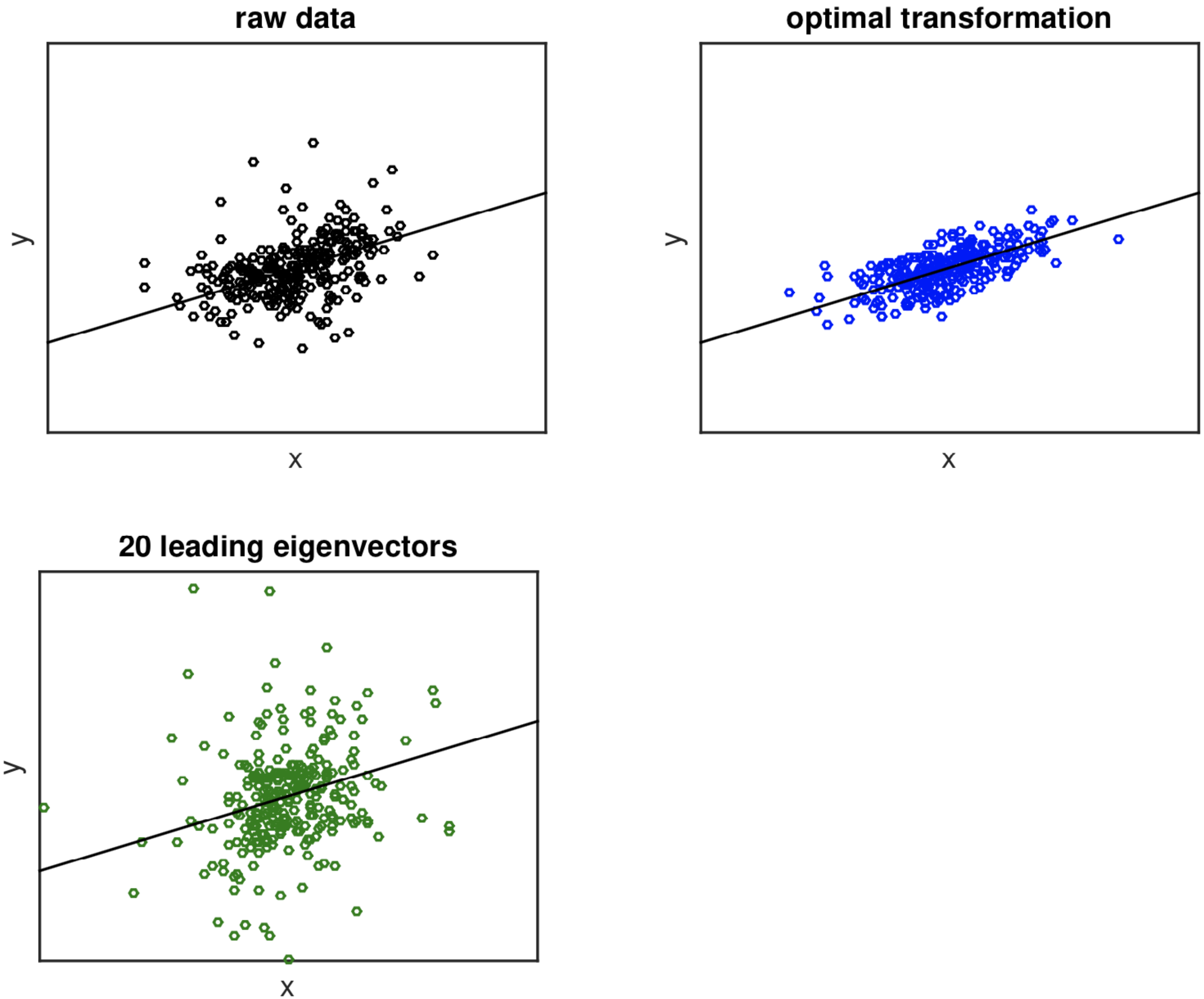
Optimal projection



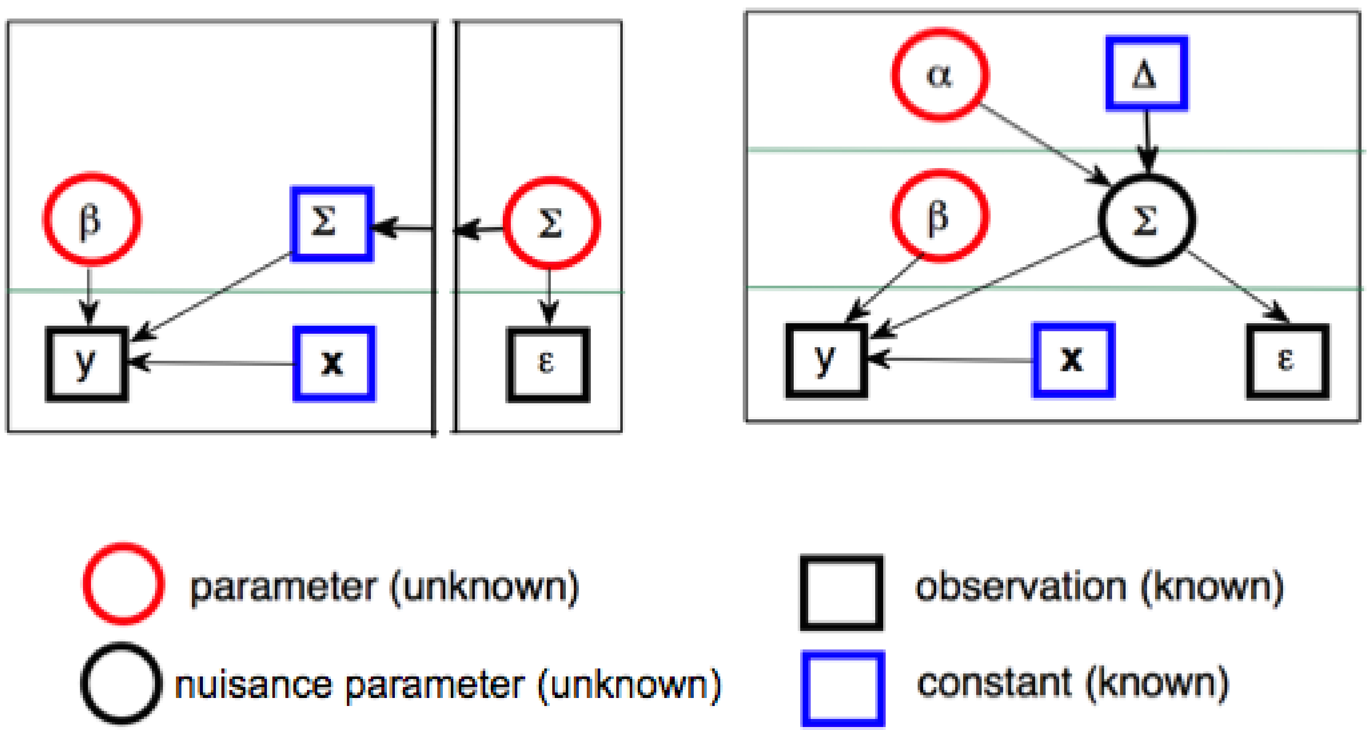
Two steps approach



Illustration



Integrated approach



Hannart 2016

Integrated likelihood

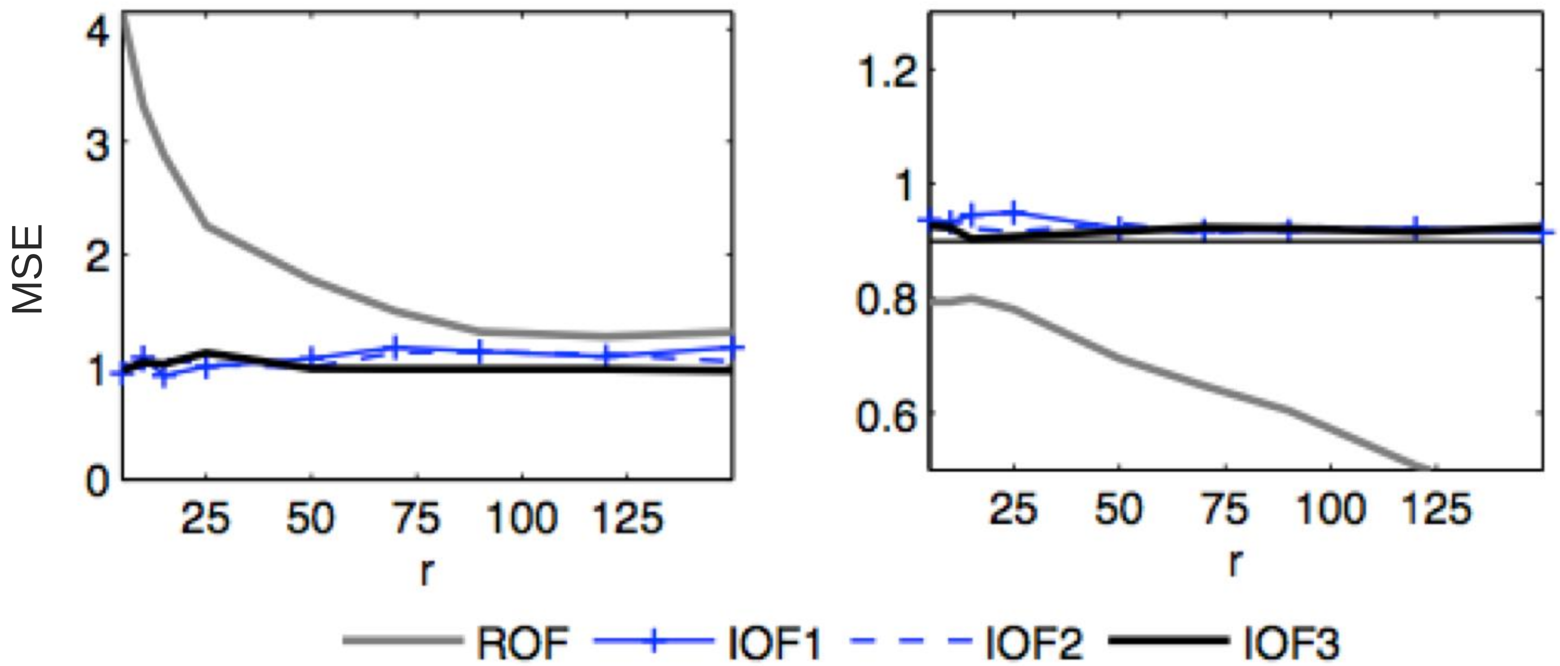
$$\begin{cases} \widehat{\alpha} = \text{argmax}_{\alpha \in [0,1]} \; \{ \log \ell(\alpha) \} \\ \widehat{\beta} = (\boldsymbol{x}' \boldsymbol{\Sigma}_{\widehat{\alpha}}^{-1} \boldsymbol{x})^{-1} (\boldsymbol{x}' \boldsymbol{\Sigma}_{\widehat{\alpha}}^{-1} y) \end{cases}$$

$$\begin{aligned} -2 \log \ell(\alpha) \;&=\; \phi\left(\tfrac{r}{1-\alpha} + 1\right) - \phi\left(\tfrac{\alpha\, r}{1-\alpha}\right) - n(\tfrac{r}{1-\alpha} + n + 2) \log(\tfrac{1-\alpha}{r} + 1) \\ &+\; (\tfrac{r}{1-\alpha} + n + 2) \log |\boldsymbol{\Sigma}_\alpha| - (\tfrac{\alpha\, r}{1-\alpha} + n + 1) \log |\boldsymbol{\Delta}| \\ &+\; (\tfrac{r}{1-\alpha} + n + 2) \log \left\{ 1 + \tfrac{(1-\alpha)\, n}{r} \, \mathrm{F}_\alpha \right\} \end{aligned}$$

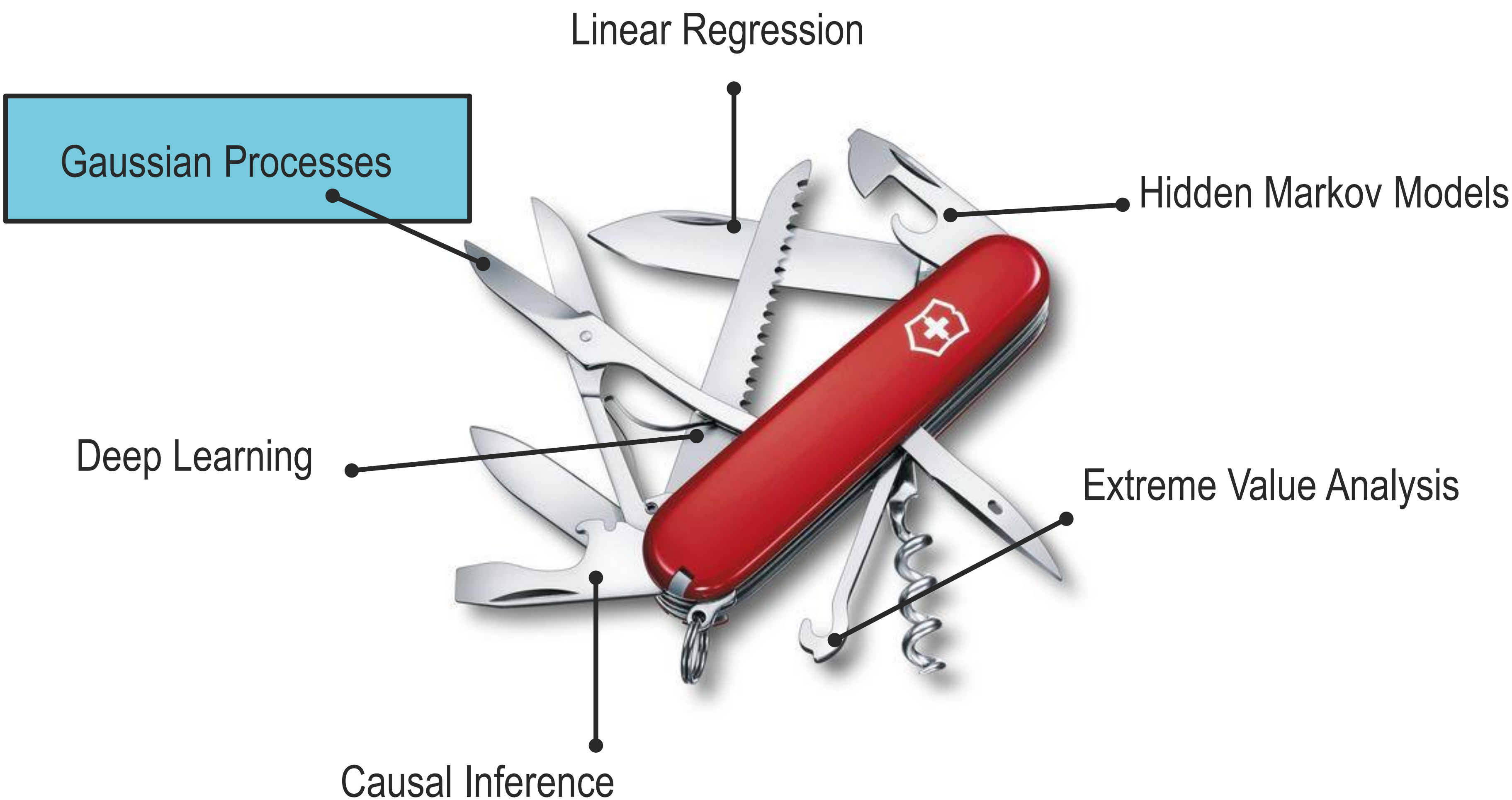
Skill

MSE of the estimator

Reliability of the confidence interval



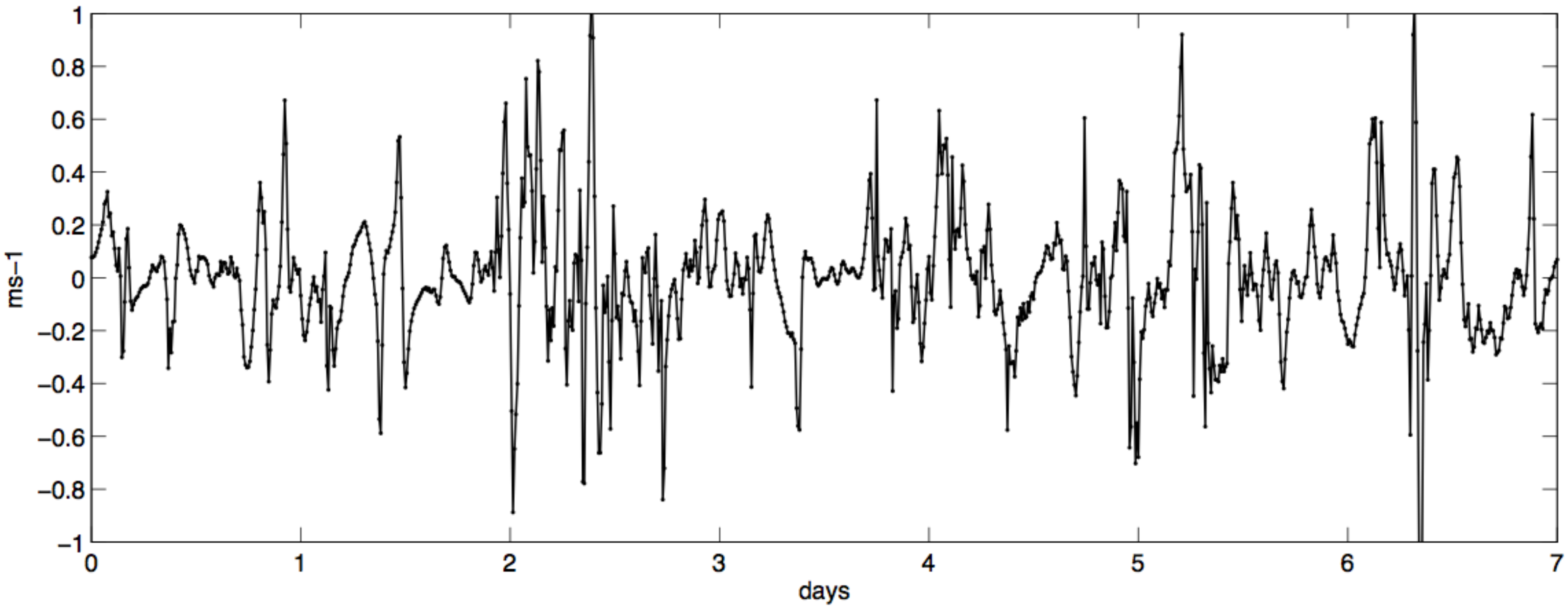
Outline


Linear Regression
Gaussian Processes
Hidden Markov Models
Deep Learning
Extreme Value Analysis
Causal Inference

Wind power generation

Context and motivation

- wind speed (Rawson wind farm): 10’ differentiated series

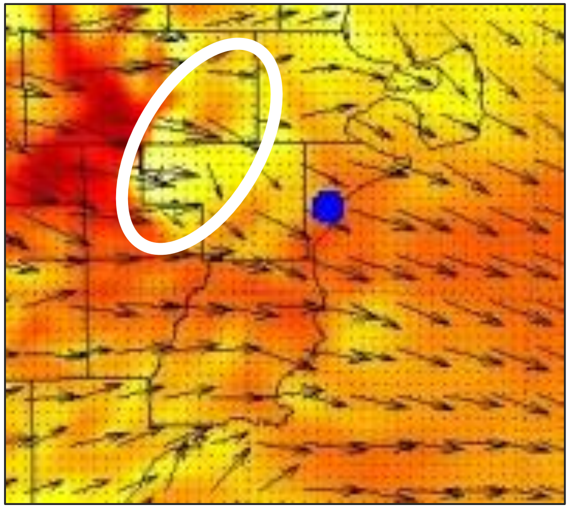


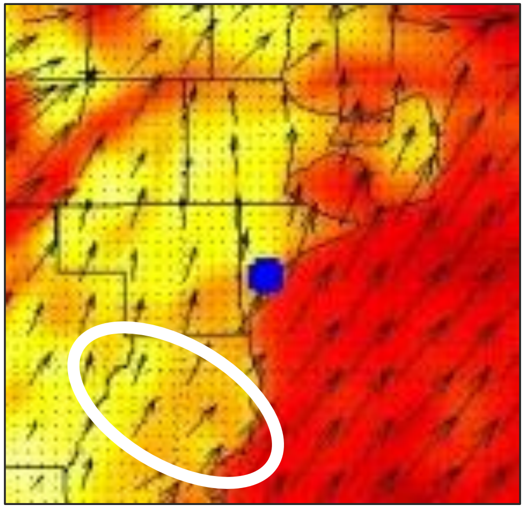
Time dependence structure can be reasonably well modelled e.g. with an autoregressive model of order 2 on the differentiated time series (= ARI(2,1) process)

Some predictivity.

Context and motivation

- Idea of “upstream prediction”

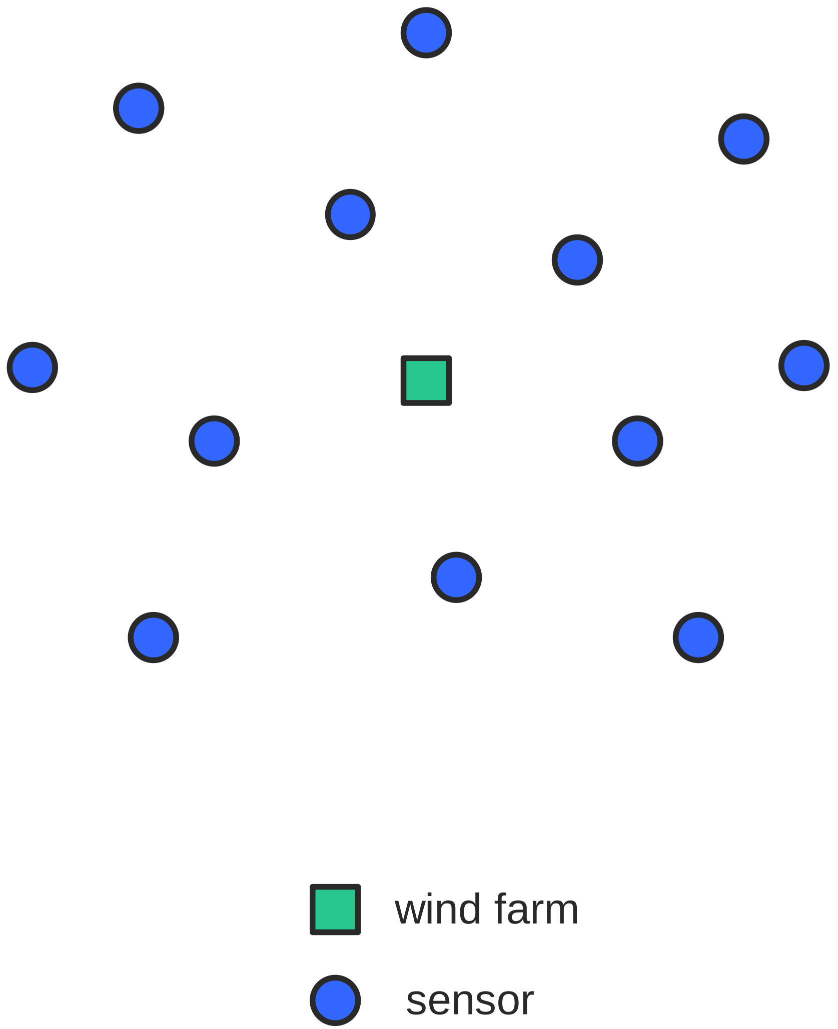




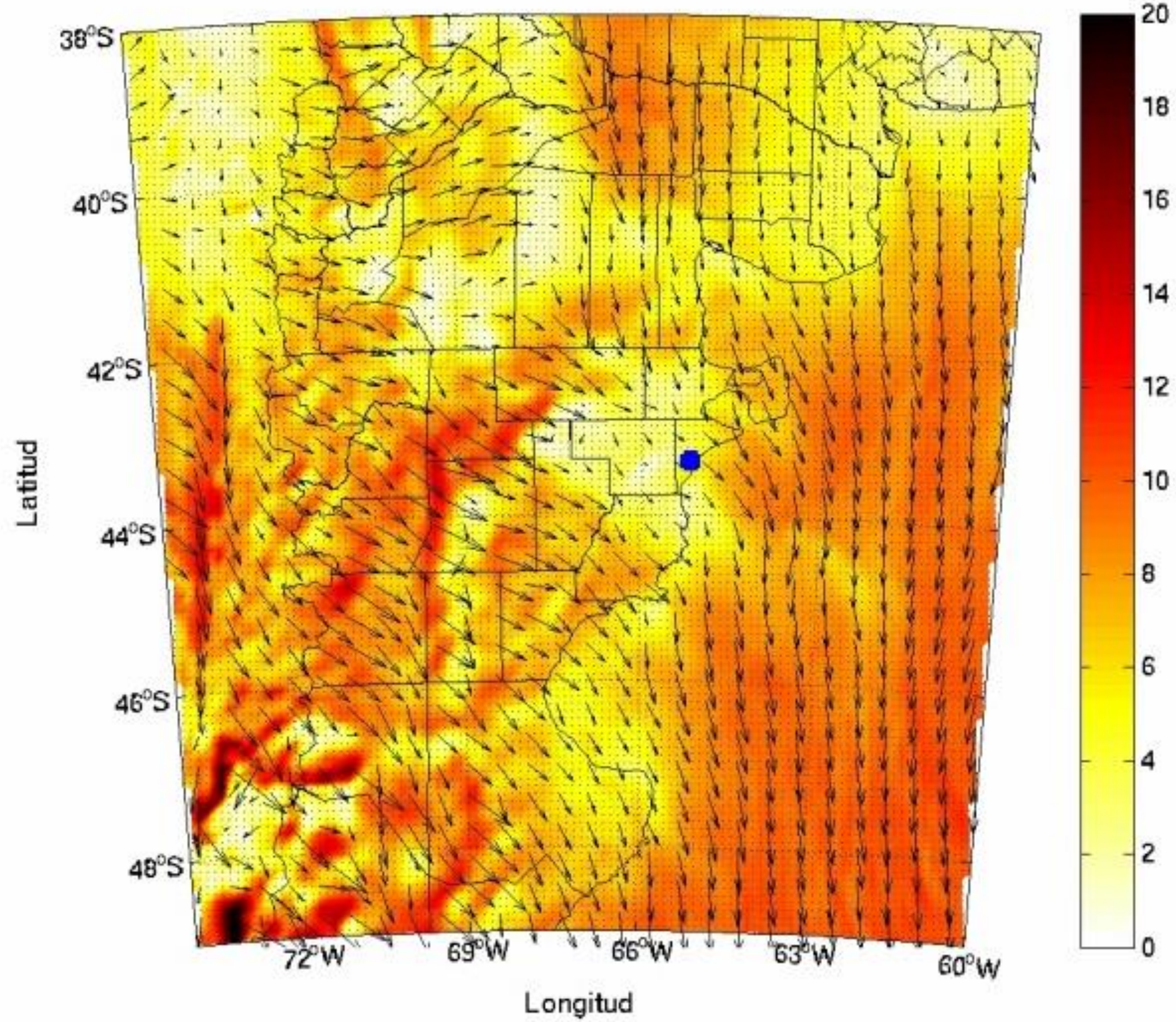
What would be the benefit of leveraging space-time dependence ?

Context and motivation

- Idea: wind farm + “integrated forecasting network”



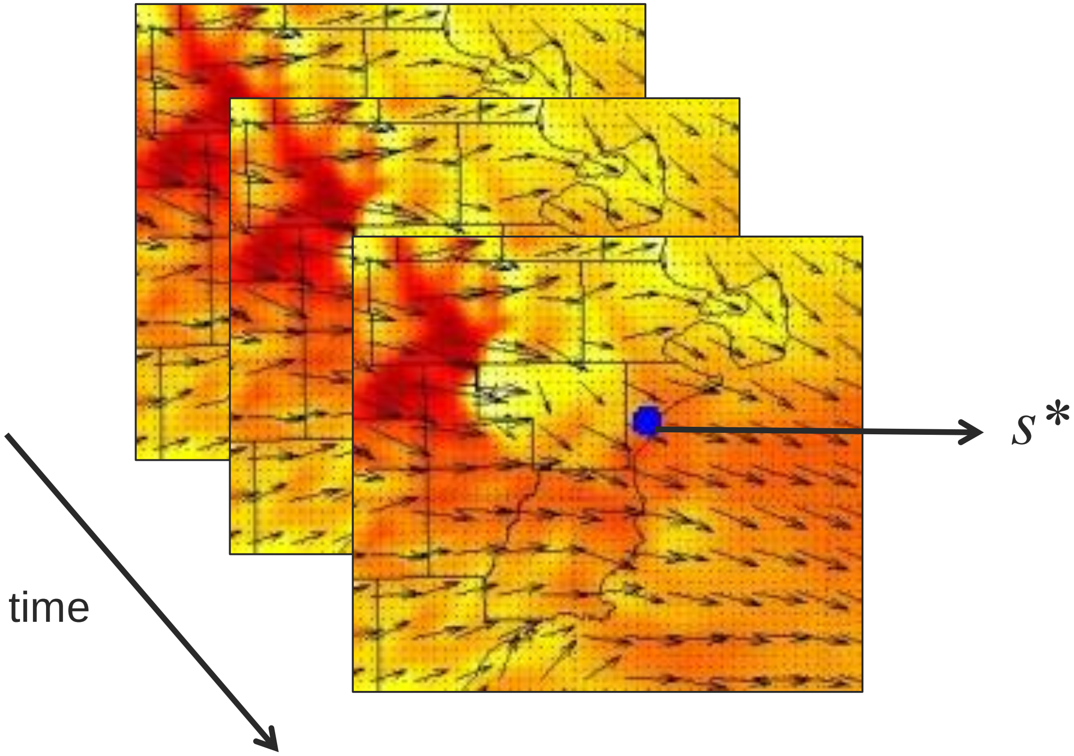
Data



38°S
40°S
42°S
44°S
46°S
48°S
Latitud
72°W
69°W
66°W
63°W
60°W
Longitud
20
18
16
14
12
10
8
6
4
2
0

Notation

$$\boldsymbol{x} = (x_{s,t-\tau})_{s\in\mathcal{S},\,\tau=0,1,\ldots,T}$$



.

Notation

$$\boldsymbol{x} = \left(x_{s,t-\tau}\right)_{s\in\mathcal{S},\,\tau=0,1,\ldots,T}$$

assumed to be a multivariate Gaussian with covariance $\boldsymbol{\Sigma}$

Notation

$$\boldsymbol{x} = (x_{s,t-\tau})_{s \in \mathcal{S},\, \tau=0,1,\ldots,T}$$

assumed to be a multivariate Gaussian mixture with constant covariance $\boldsymbol{\Sigma}$

regularized estimate of Σ

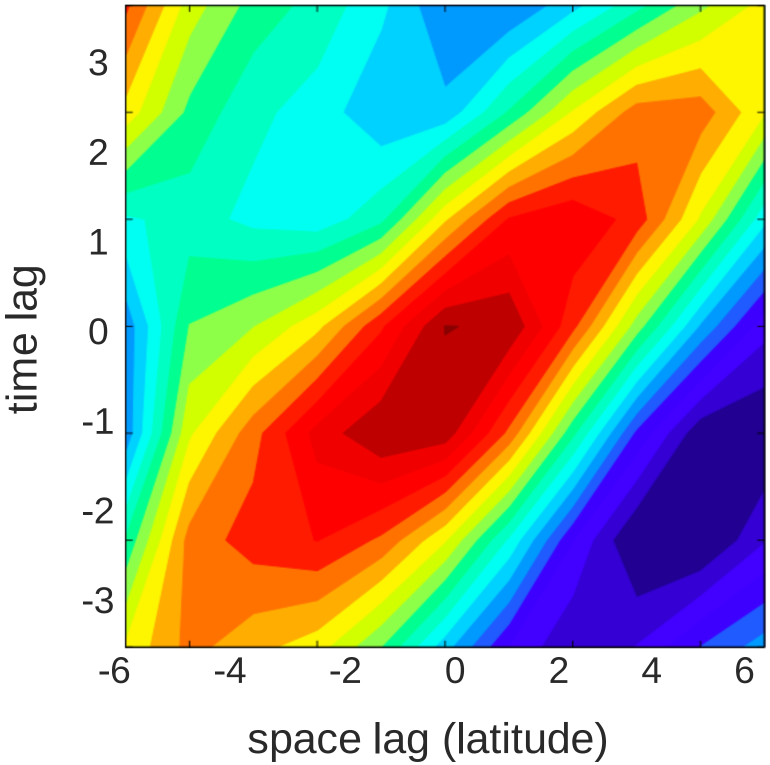
Prediction

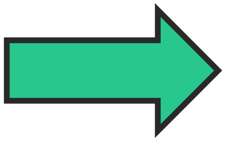
$$\begin{cases} \boldsymbol{x} = (x_{s,t-\tau})_{s\in\mathcal{S},\,\tau=0,1,\ldots,T} \\ \boldsymbol{x}_0 = (x_{s,t})_{s\in\mathcal{S}} \longrightarrow \text{present} \\ \boldsymbol{x}_1 = (x_{s,t-\tau})_{s\in\mathcal{S},\,\tau=1,\ldots,T} \longrightarrow \text{past} \end{cases}$$

The prediction follows:

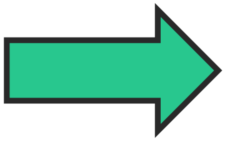
$$\mathbb{E}(\boldsymbol{x}_0 \mid \boldsymbol{x}_1) = \boldsymbol{x}_1\beta \qquad \beta = \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{10}$$

Correlogram of Σ





non separability



reflects the dynamic of the flow

Covariance regularization

Covariance regularization: low rank representation

$$\mathbf{\Sigma} = \mathrm{Var}(\boldsymbol{x}) = \mathbf{V}_r \mathbf{\Delta}_r \mathbf{V}_r' + \lambda \mathbf{I}$$

→ nugget

r basis functions are retained (r<<p)

Covariance regularization

Covariance regularization: low rank representation

$$\mathbf{\Sigma} = \mathrm{Var}(\boldsymbol{x}) = \mathbf{V}_r \mathbf{\Delta}_r \mathbf{V}_r' + \lambda \mathbf{I}$$

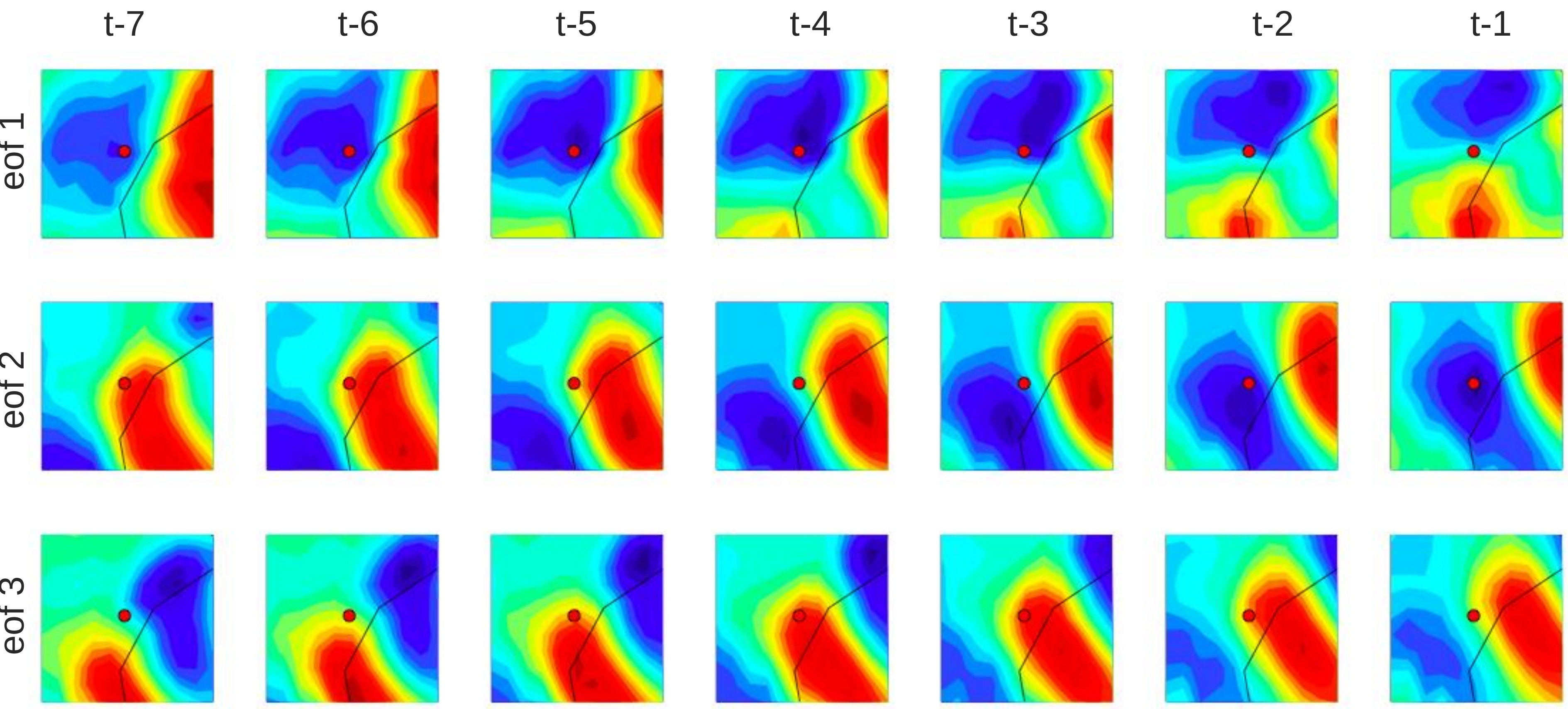
→ nugget

r basis functions are retained (r<<p)

ad hoc wave propagation basis functions probably exist

but use of EOFs

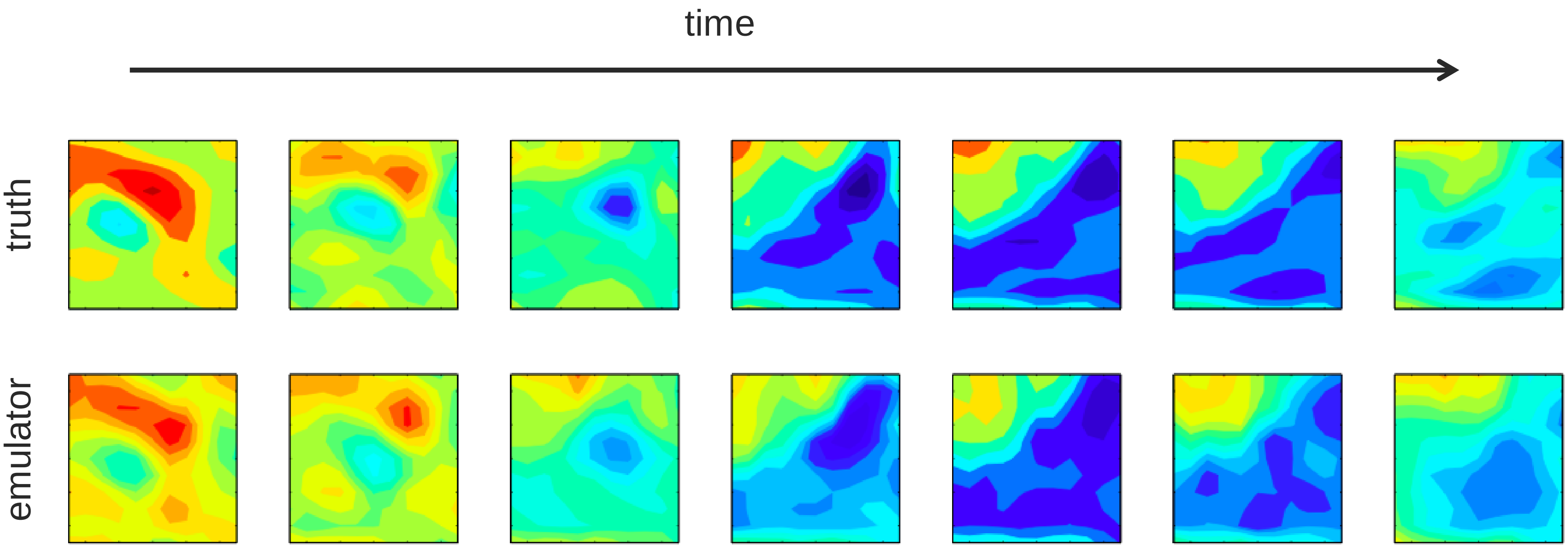
Eigenvectors



the eigenvectors of the covariance Σ are dynamic maps

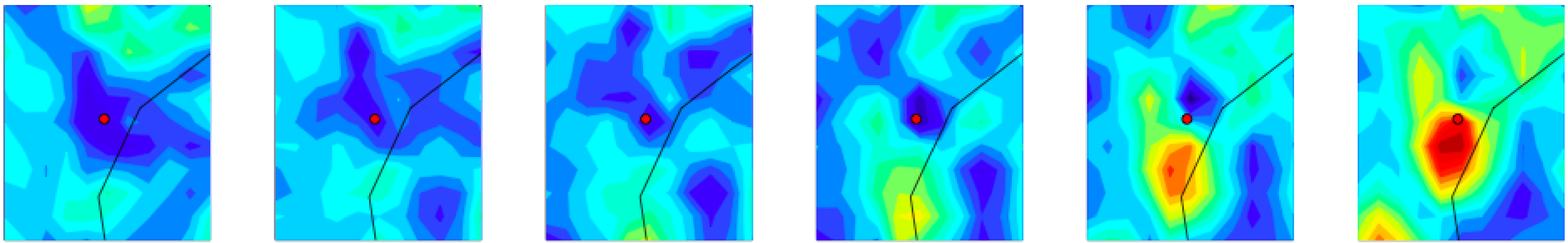
eofs also exhibit wave-like moving patterns

Skill of covariance estimation

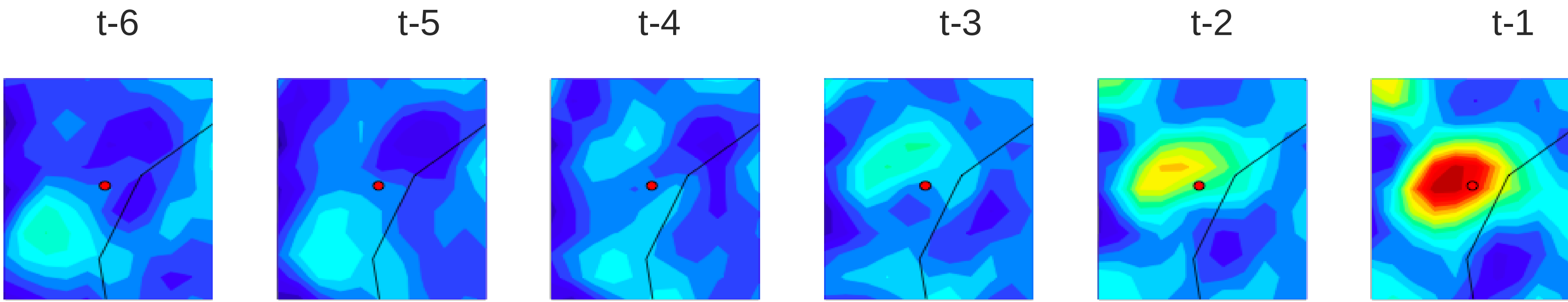


Estimated weights

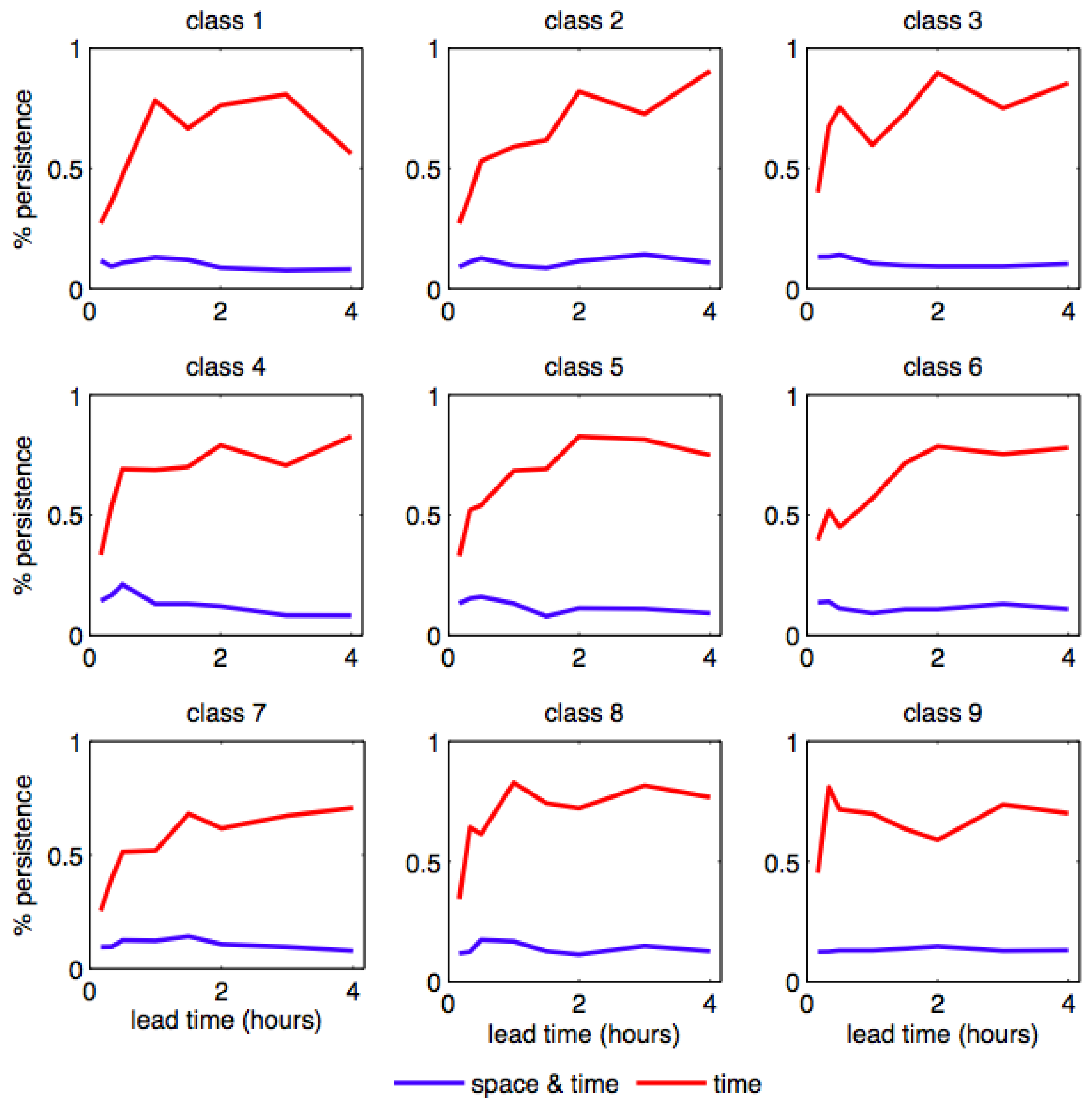
t-6 t-5 t-4 t-3 t-2 t-1



Estimated weights



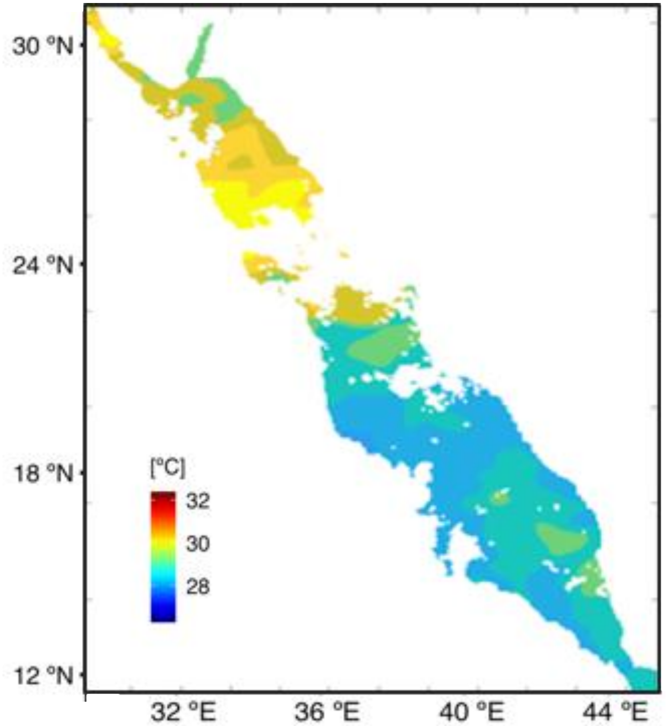
Skill of prediction



~85%
mse
reduction

Interpolating temperature missing values

Daily temperature (SST), Aug 1-31 2010, Red Sea



<u>Context of this research work:</u>

- Data challenge, Extreme Value Analysis 2019
- Co-supervision of the Msc. thesis of F. Baeriswyl, University McGill
- Results presented at Summer School « Mathematics of Climate and the Environment », CNRS / IHP.

Notation

$$\boldsymbol{x} = \left(x_{s,t-\tau}\right)_{s \in \mathcal{S},\, \tau=0,1,\ldots,T}$$

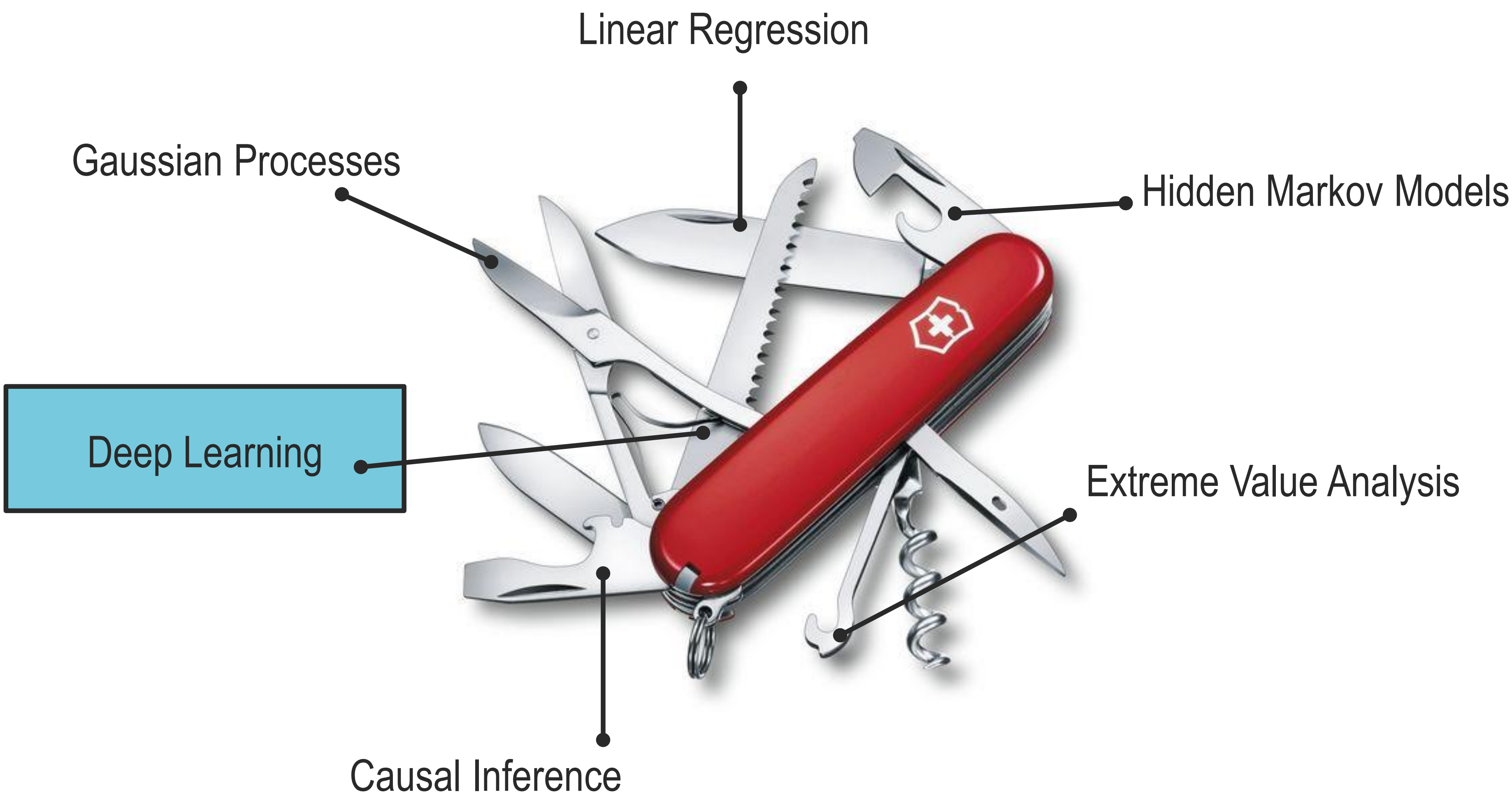
assumed to be a multivariate Gaussian with covariance $\boldsymbol{\Sigma}$

Skill

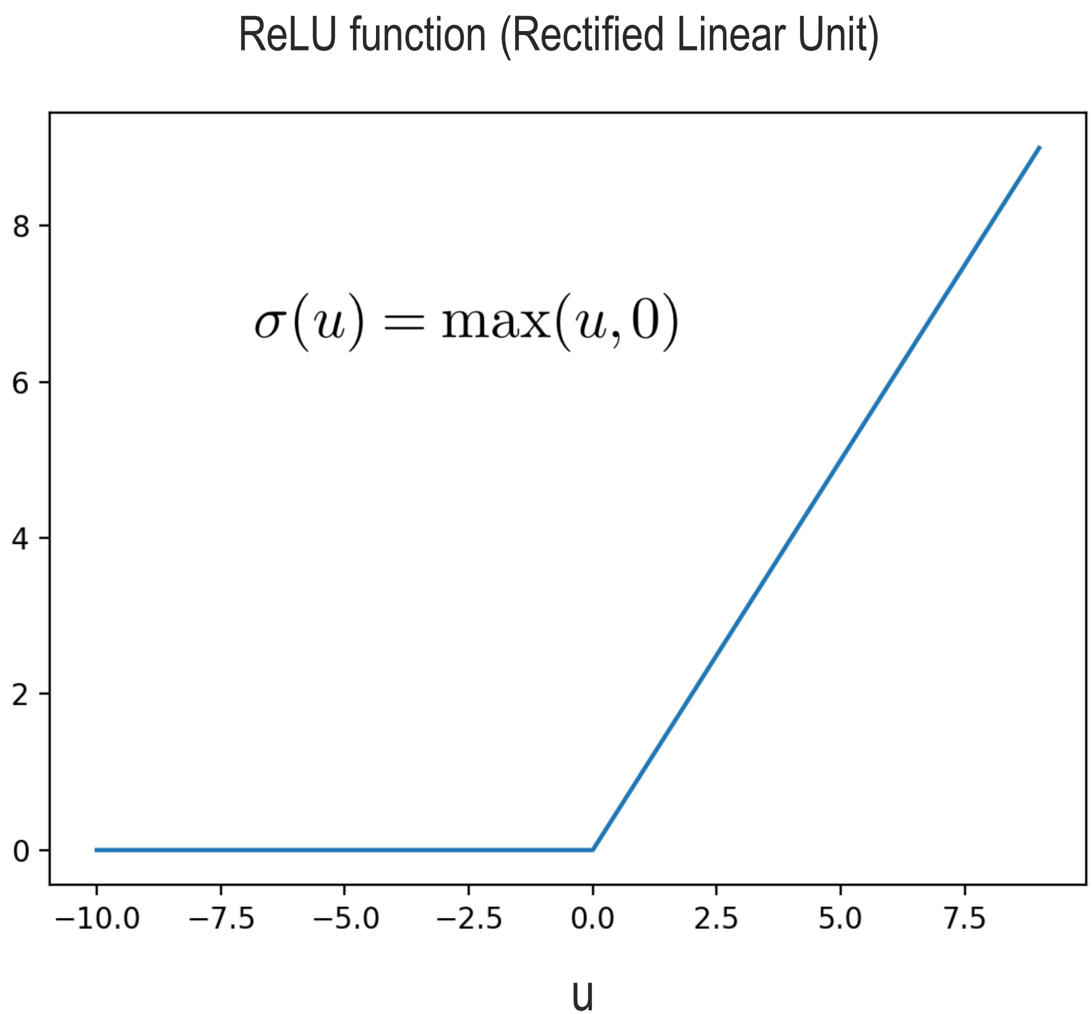
- #1 best ranking: 0.0036
  - Team LC2019.
  - **Poisson equation regularization (~ similar to ours).**
- #2 best ranking: 0.0044
  - Team Rainbow warriors.
  - **Quasiseparable Gaussian process.**
- #3 best ranking: 0.0047
  - Team BlackBox
  - **Deep Learning, convolutional recurrent architecture.**
- about 50 teams participated.

our team →

Outline


Linear Regression
Gaussian Processes
Hidden Markov Models
Deep Learning
Extreme Value Analysis
Causal Inference

Deep learning



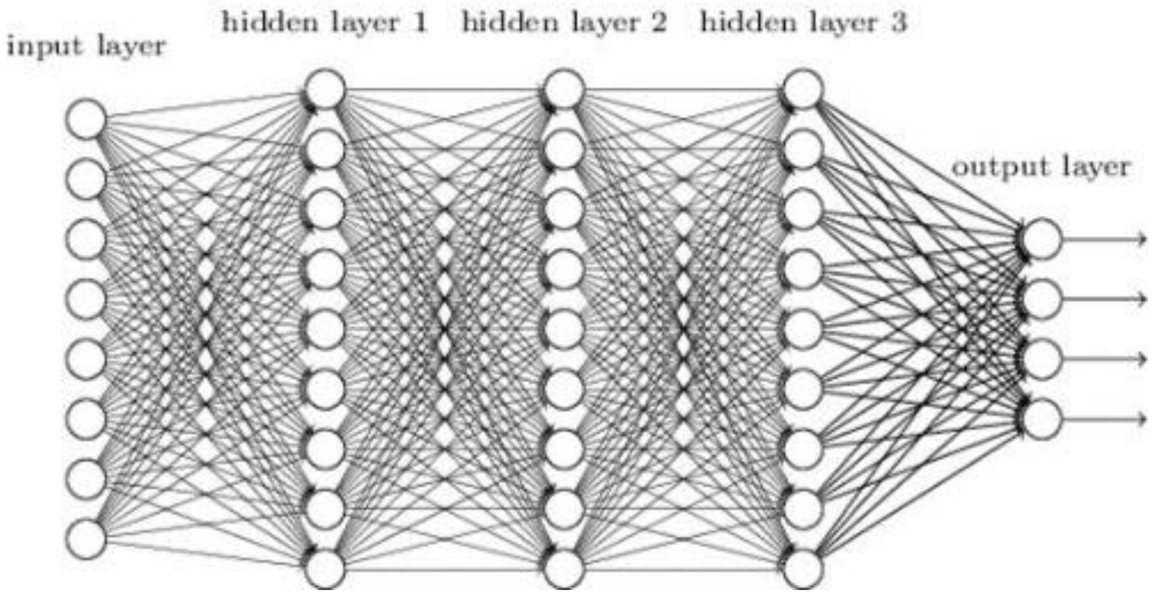
Deep learning

$$y = \sigma\left(\mathbf{W}_d \sigma\left(\mathbf{W}_{d-1} \sigma\left(\ldots \sigma\left(\mathbf{W}_1 \boldsymbol{x}\right)\right)\right)\right) + \varepsilon$$

Deep learning

$$y = \sigma\left(\mathbf{W}_d\,\sigma\left(\mathbf{W}_{d-1}\sigma\left(\ldots\sigma\left(\mathbf{W}_1\boldsymbol{x}\right)\right)\right)\right) + \varepsilon$$

$\boldsymbol{x}$



y

Deep learning

$$y = \sigma\left(\mathbf{W}_d \sigma\left(\mathbf{W}_{d-1} \sigma\left(\ldots \sigma\left(\mathbf{W}_1 \boldsymbol{x}\right)\right)\right)\right) + \varepsilon = \phi(\boldsymbol{x}, \mathbf{W}) + \varepsilon$$

$$p(y \mid \boldsymbol{x}, \mathbf{W}) = \mathcal{N}(y \mid \phi(\boldsymbol{x}, \mathbf{W}), \lambda \mathbf{I})$$

Deep learning

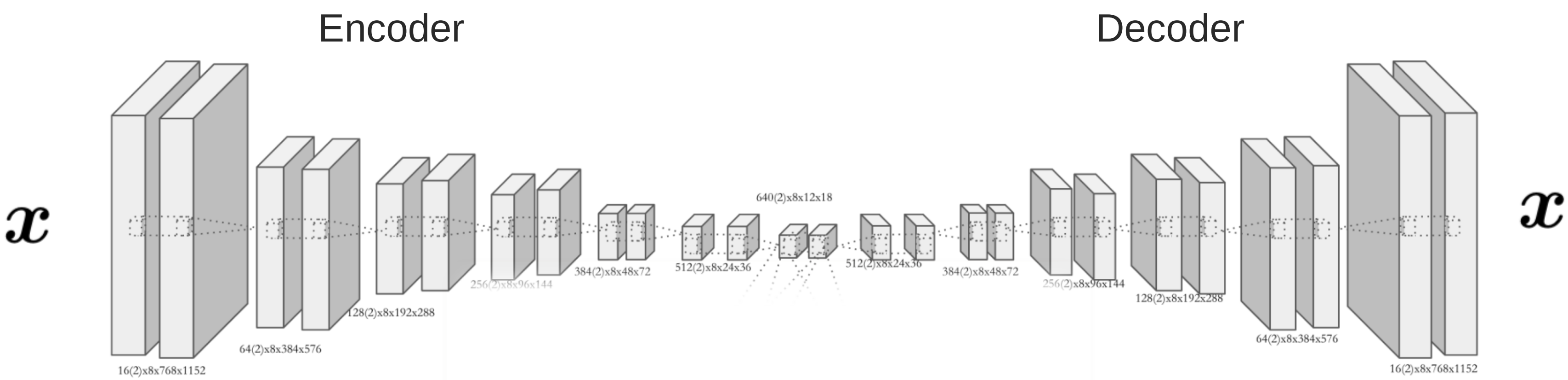
$$y = \sigma\left(\mathbf{W}_d \sigma\left(\mathbf{W}_{d-1} \sigma\left(\ldots \sigma\left(\mathbf{W}_1 \boldsymbol{x}\right)\right)\right)\right) + \varepsilon = \phi(\boldsymbol{x}, \mathbf{W}) + \varepsilon$$

$$p(y \mid \boldsymbol{x}, \mathbf{W}) = \mathcal{N}(y \mid \phi(\boldsymbol{x}, \mathbf{W}), \lambda \mathbf{I})$$

$$\widehat{\mathbf{W}} = \operatorname{argmax}_{\mathbf{W}} \ \log p(y \mid \boldsymbol{x}, \mathbf{W})$$

- High dimensional optimization problem
- Stochastic gradient indecent
- Backpropagation (= chain rule)
- Many tricks

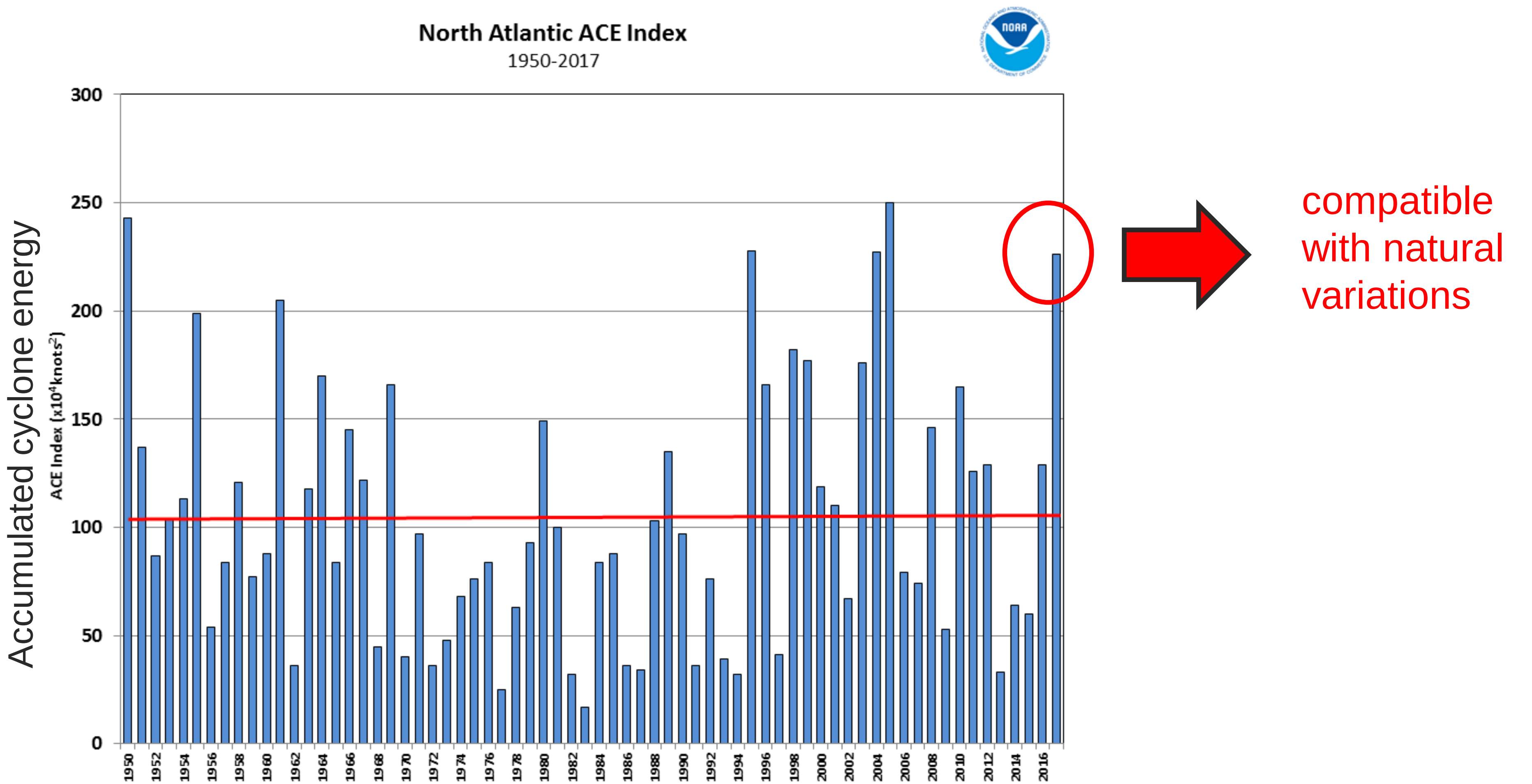
Convolutional Autoencoder



Hurricane attribution

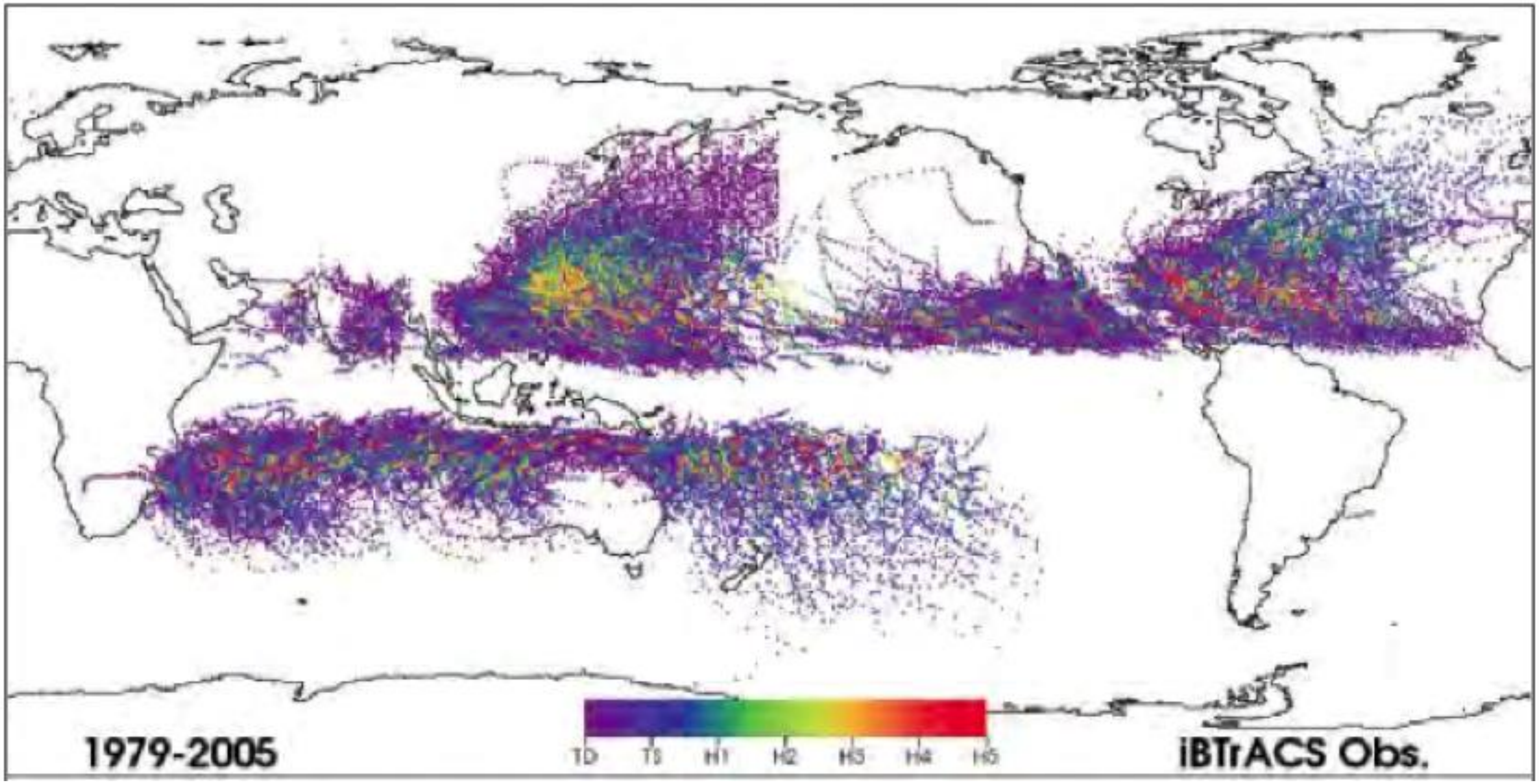
8 September 2017 06.00pm GMT

Temporal plot of tropical cyclones occurrences

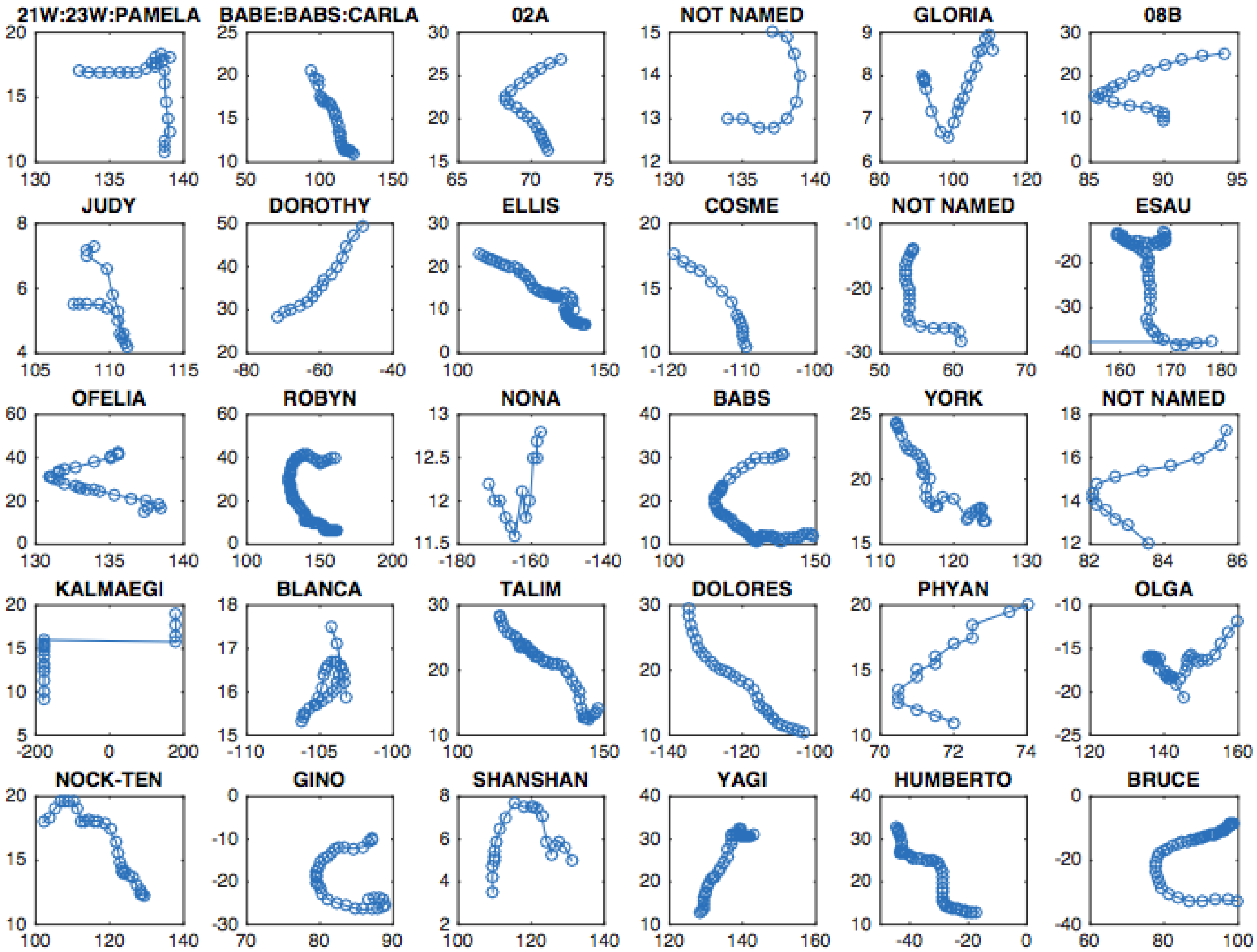


NOAA National Centers for Environmental Information, State of the Climate: Hurricanes and Tropical Storms for Annual 2017, published online January 2018, retrieved on July 27, 2018 from https://www.ncdc.noaa.gov/sotc/tropical-cyclones/201713.

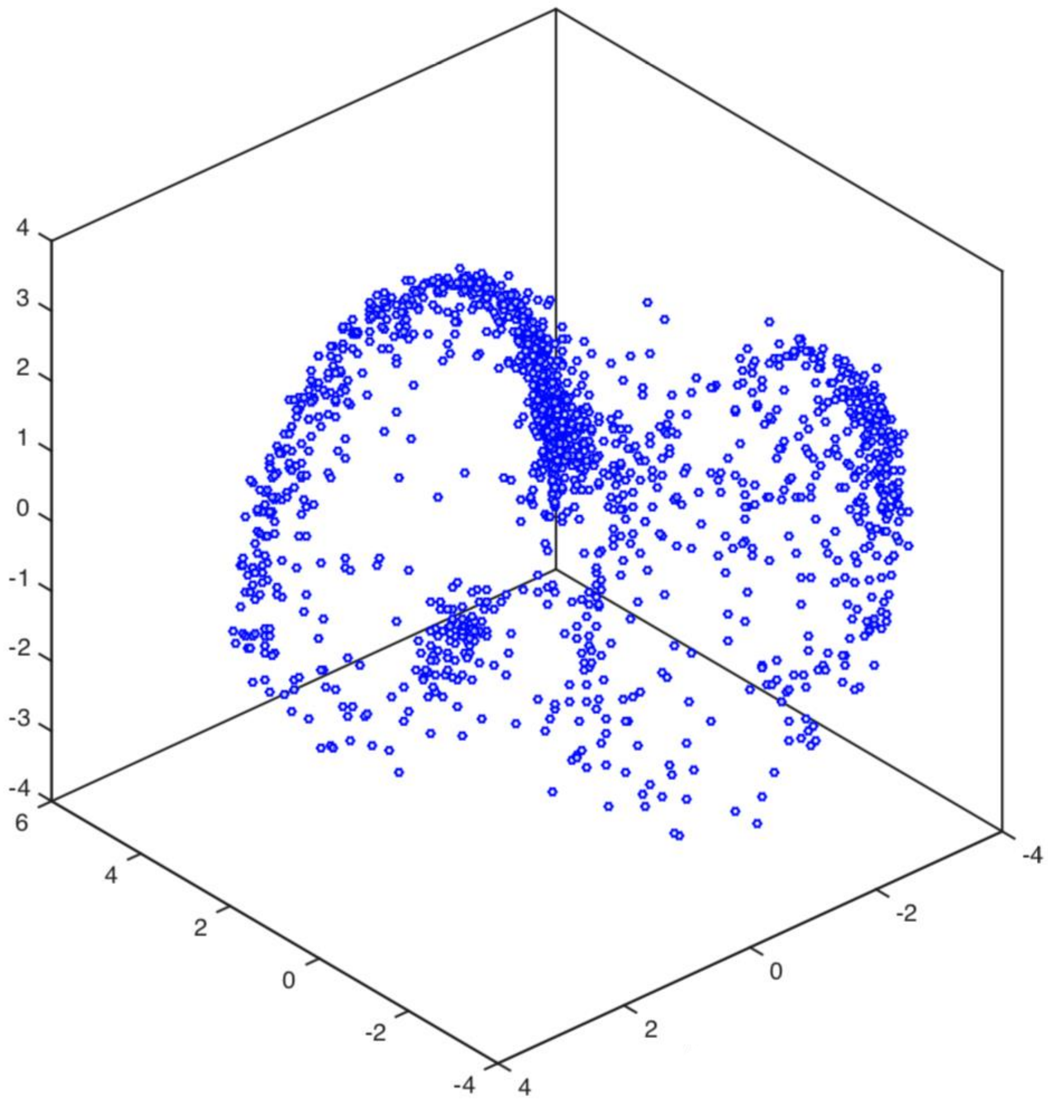
Spatial plot of tropical cyclones tracks



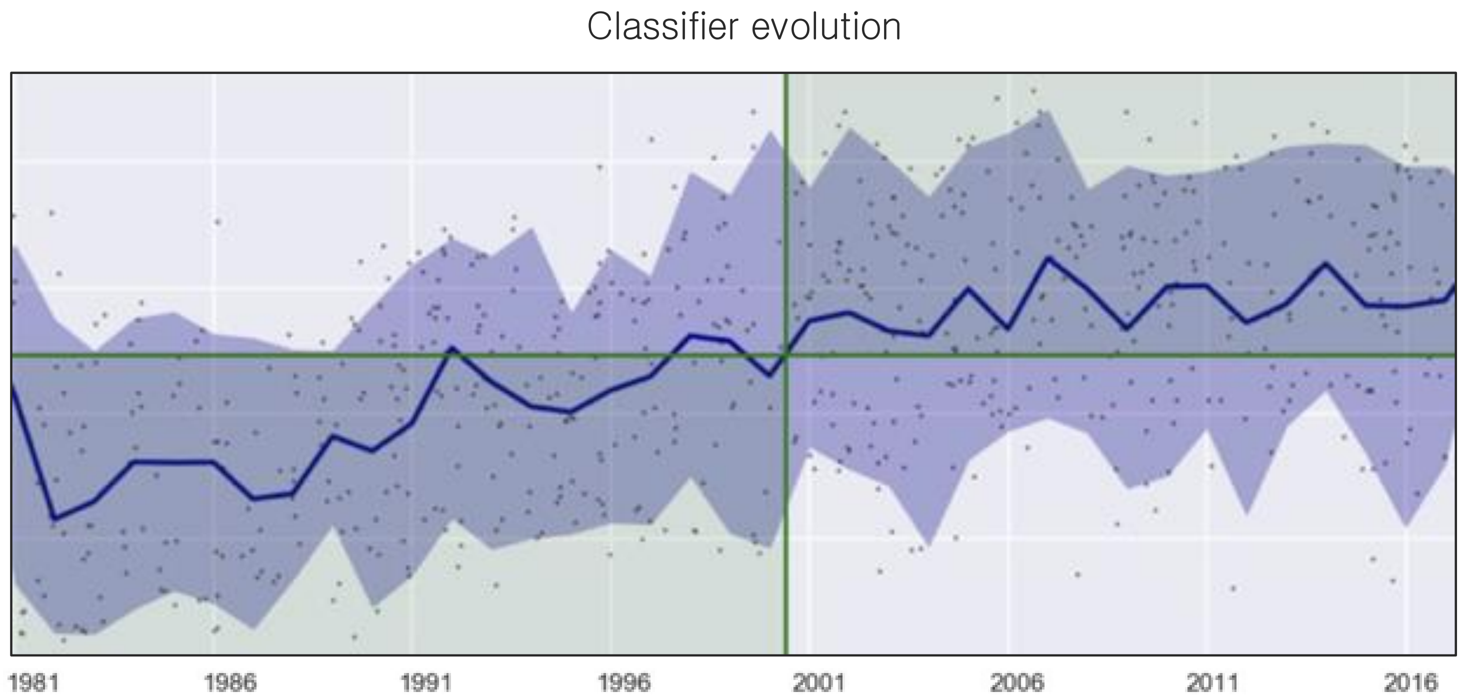
Individual trajectories



Individual trajectories: dimension reduction

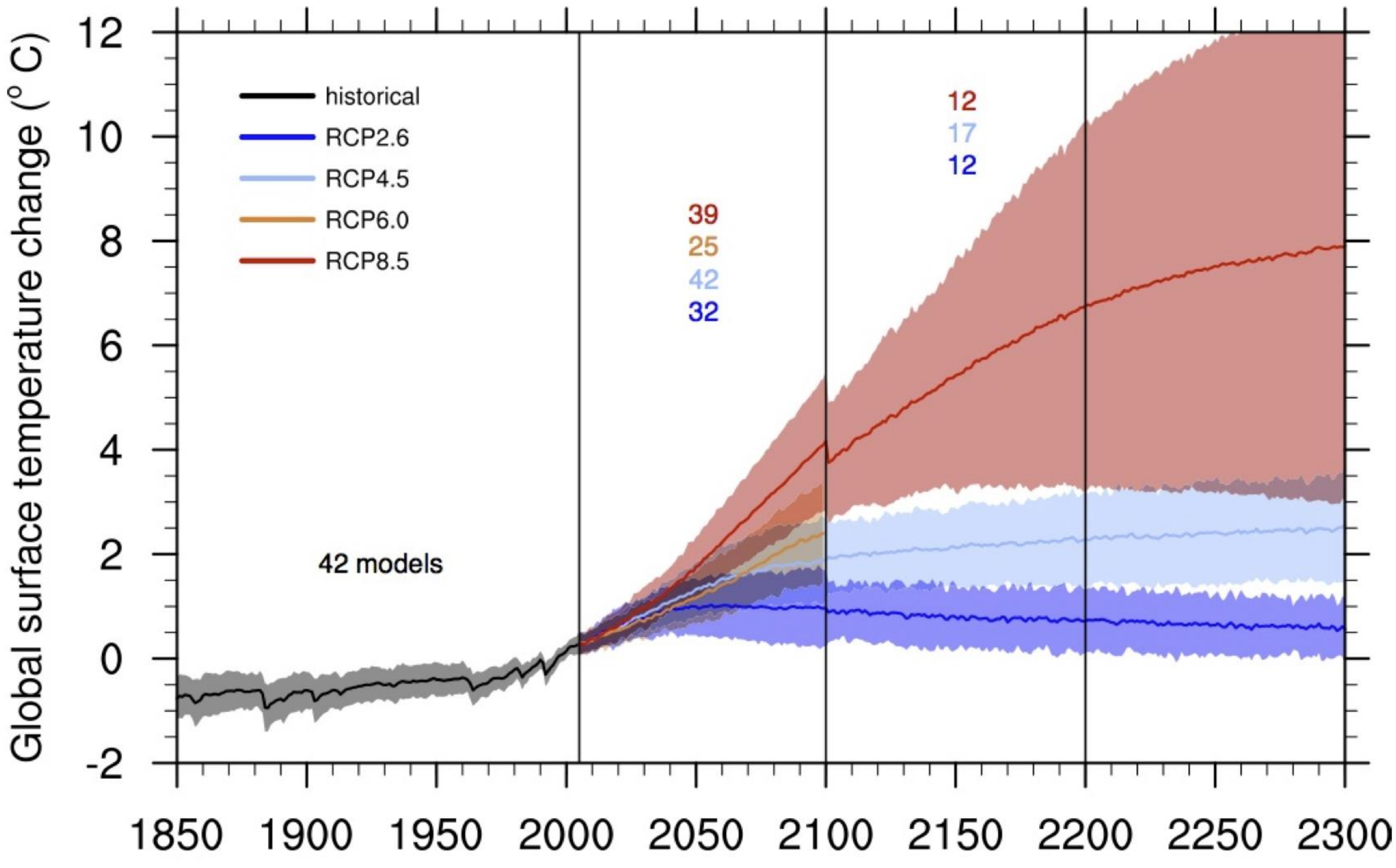


Results – work in progress

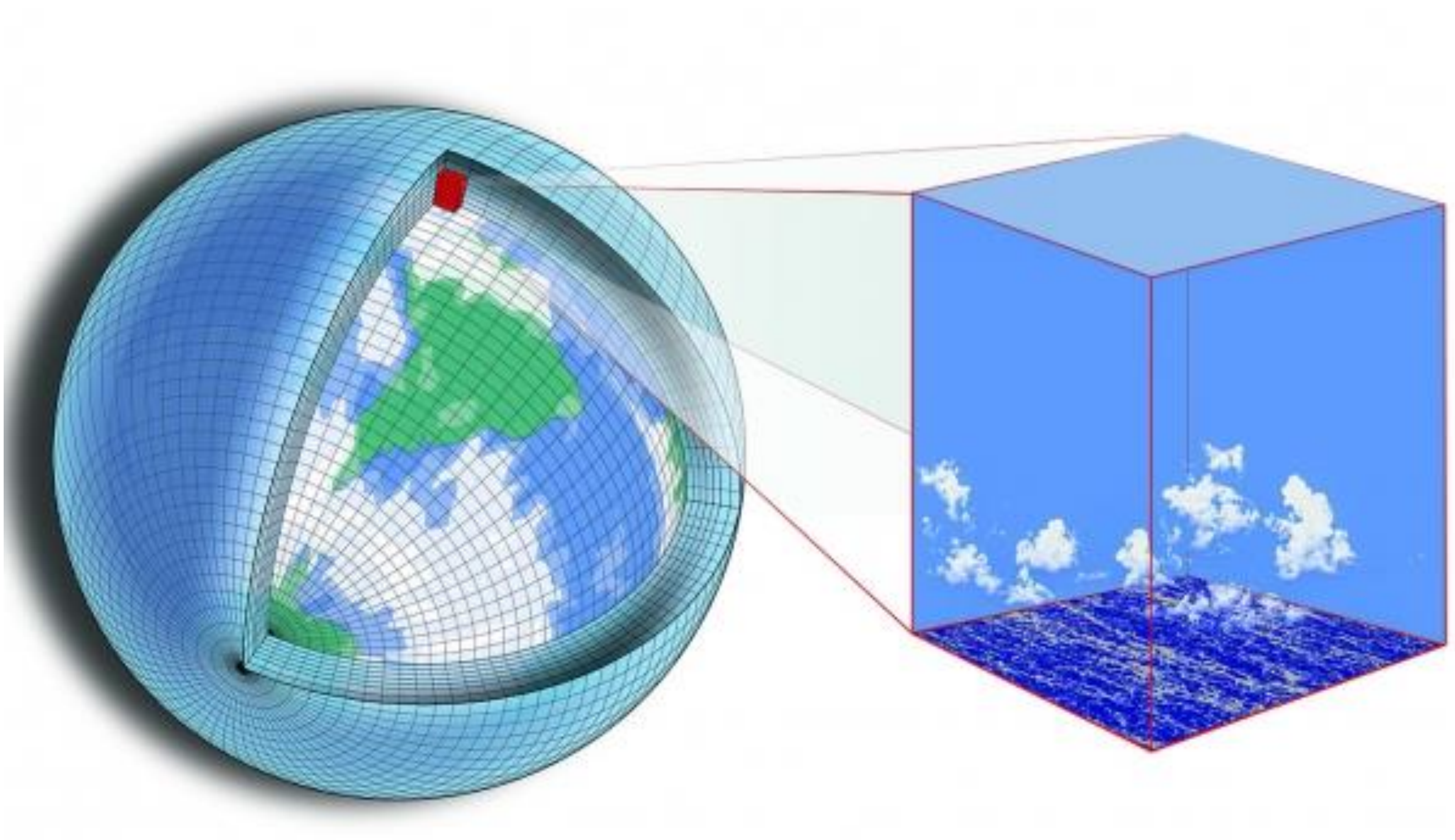


- The probability of hurricanes with z>0.5 has increased by a factor 6.
- Something has changed.
- Work in progress:
  - robustness check & verification on simulations
  - physical interpretation of the classifier

Future projections of climate change



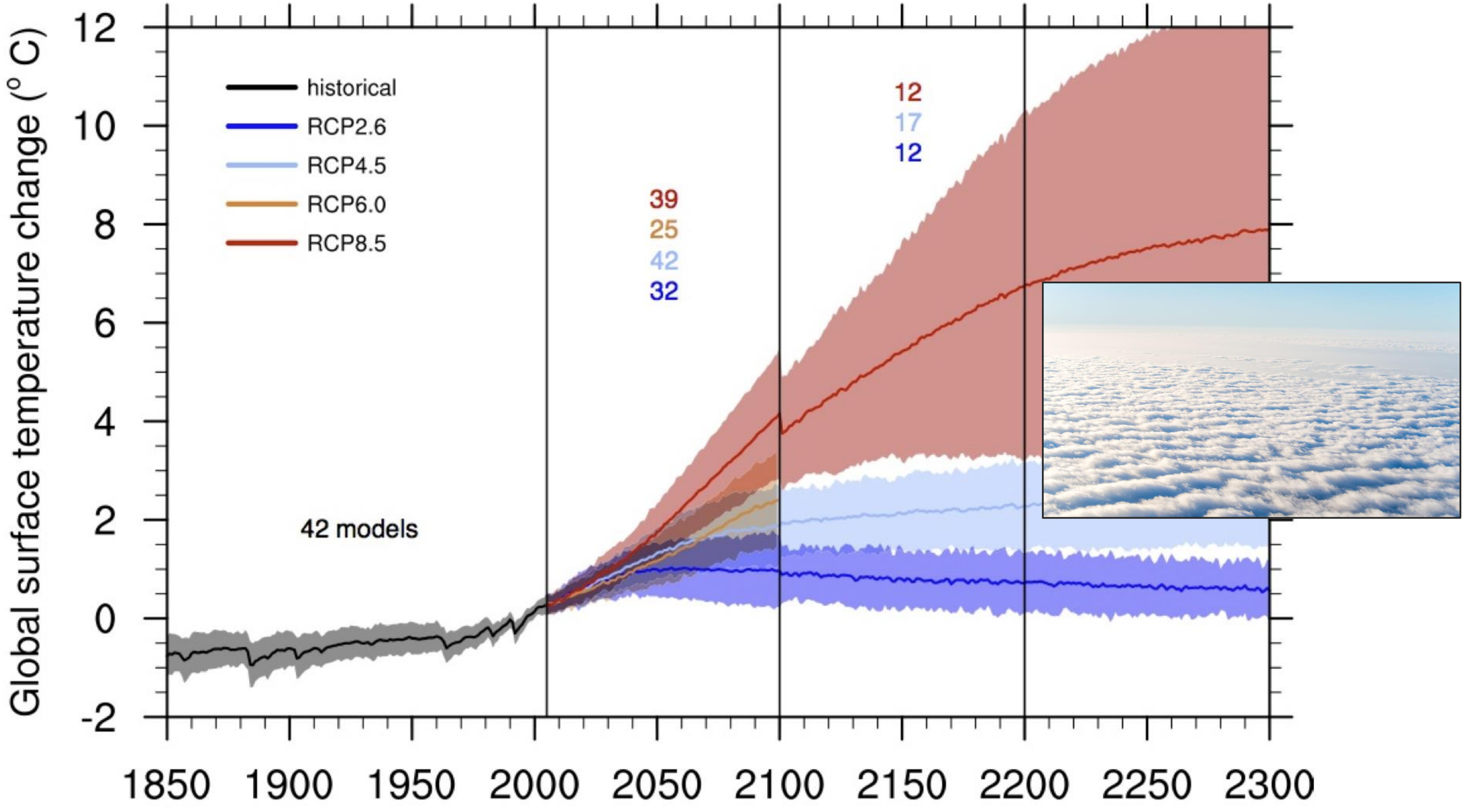
Climate models: subgrid processes



Clouds

Low level clouds: stratocumulus

Stratocumulus response is a major part of uncertainty

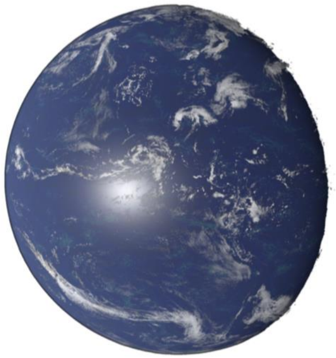


Getting around the computational wall: the AI trick

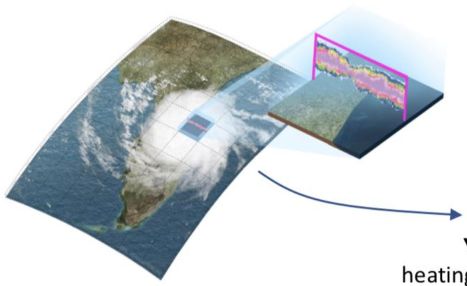
Deep learning can skillfully approximate sub-grid climate model physics harvested from cloud-resolving simulations.

Is deep learning viable for sub-grid parameterization?

Aquaplanet SPCAM testbed
1 year for training, 1 for validation
Globally diverse meteorological regimes

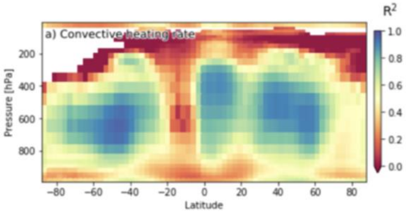


Can the 140M outputs from
1 year of 9k Cloud Resolving Models…
(solutions of accurate radiative transfer & explicit CRM equations)



SuperParameterization

….Be fit by a deep, fully connected network?
Yes, e.g. $R^2 > 0.7$ for mid-tropospheric heating by convection & radiation at 8x512 nodes.



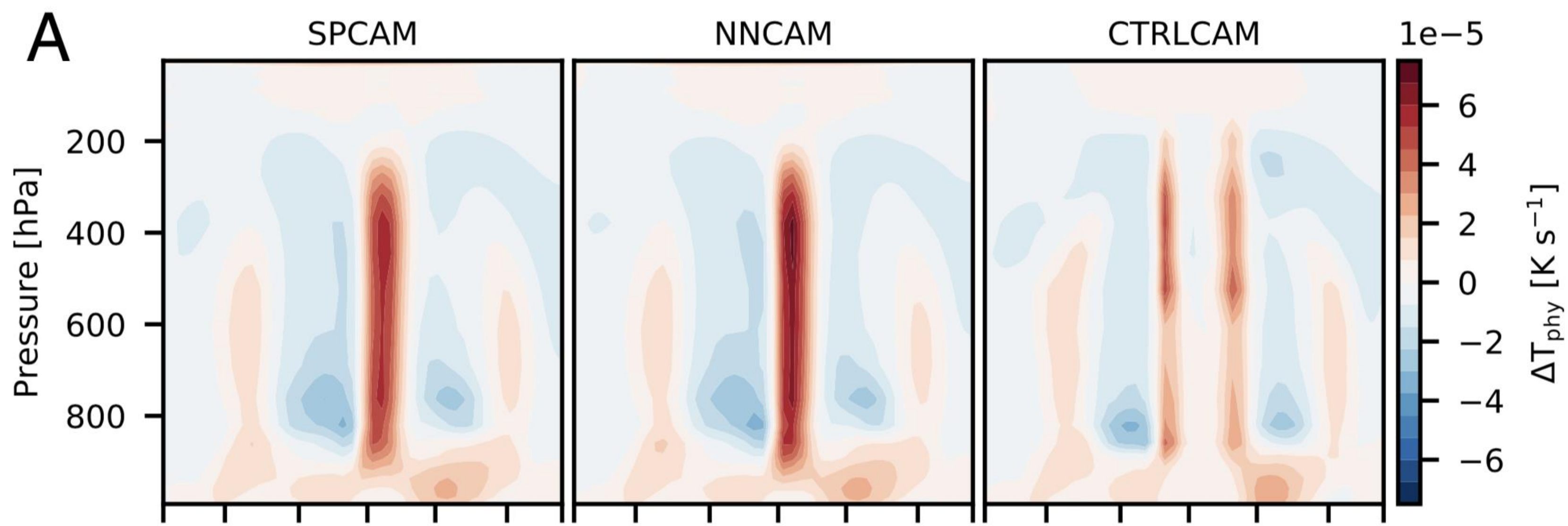
Possibly!
Just 3 months' hi-res sim data is enough for a good fit!

The "Cloud Brain"

Geophysical Research Letters
Could machine learning break the convection parameterization deadlock?
P. Gentine, M. Pritchard, S. Rasp, G. Reinaudi & G. Yacalis. May 2018.

Image credits: Illustration by Tony Gold for IBM systems magazine, ORNL visualization lab.

Some encouraging early results



Rasp et al. 2018

A promising way forward

CliMA
CLIMATE MODELING ALLIANCE

A NEW APPROACH TO CLIMATE MODELING

CLIMATE MACHINE

We are developing the first Earth system model that automatically learns from diverse data sources. Our model will exploit advances in machine learning and data assimilation to learn from observations and from data generated on demand in targeted high-resolution simulations, for example, of clouds or ocean turbulence. This will allow us to reduce and quantify uncertainties in climate predictions.

SCALABLE PLATFORM

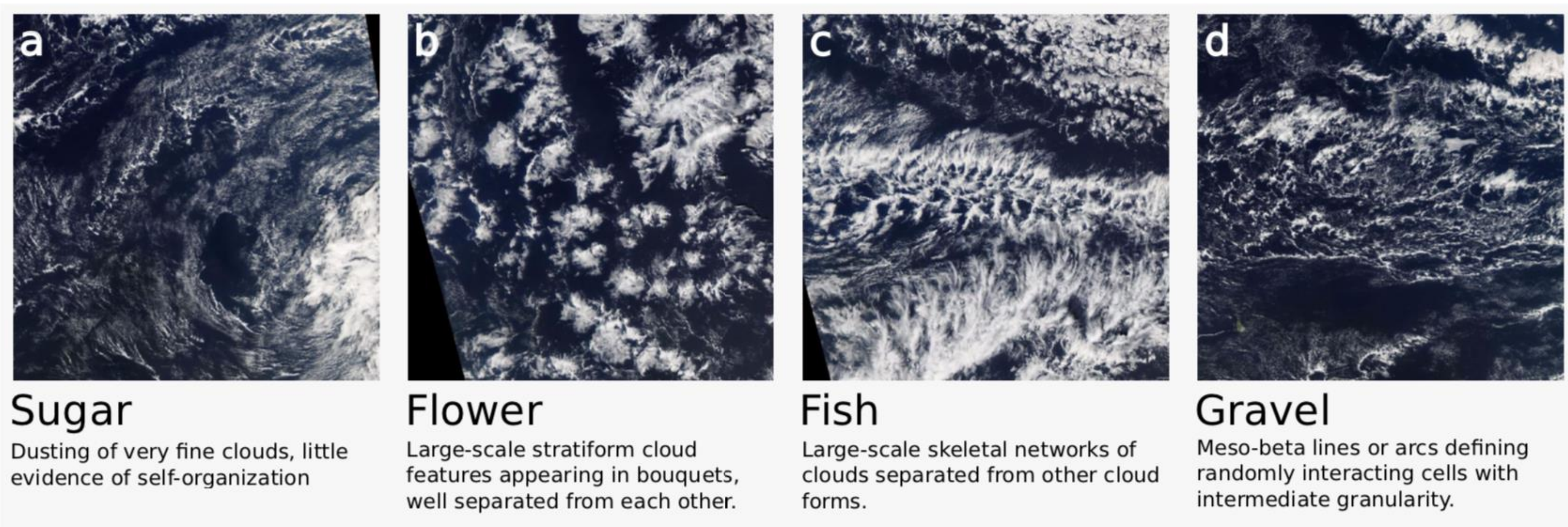
We are engineering a modeling platform that is scalable and built for growth. For processing data and for simulating the Earth system, it will exploit state-of-the-art algorithms to run on the world's fastest supercomputers and on the cloud. It will be scalable to ever finer resolution globally, and its targeted high-resolution simulations will provide detailed local climate information where needed.

OPEN HUB

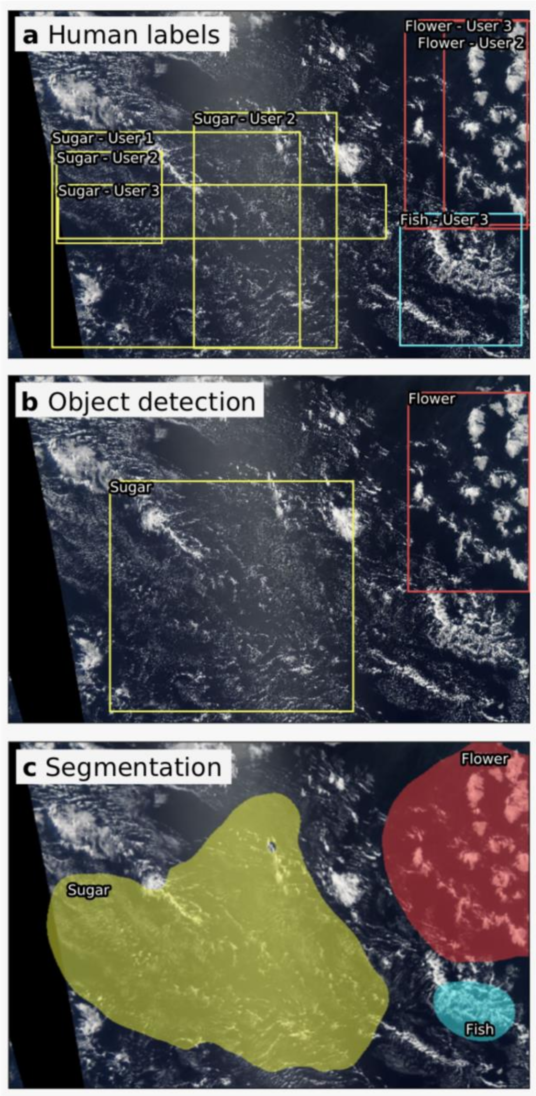
We are committed to transparency and open science principles. Our modeling platform is open source, and our results are available to the public. We will provide interfaces to our modeling platform so that it can become the anchor of an ecosystem of front-end apps. These apps may provide detailed models, for example, of flood risks, risks of extreme heat, crop yields, and other climate impacts.

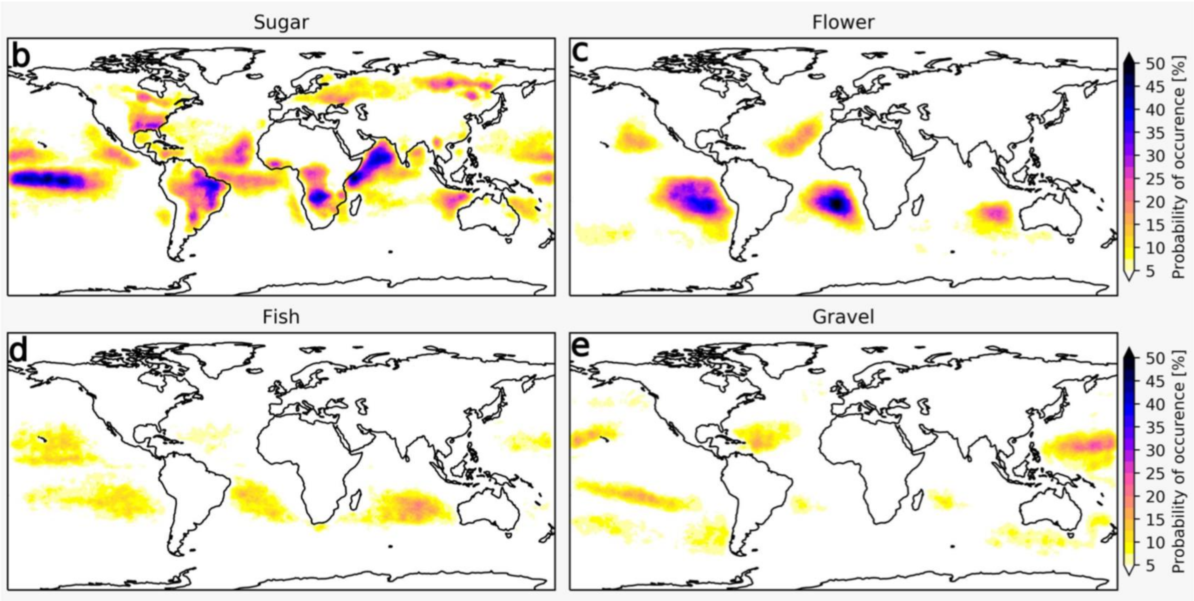
Understanding clouds from satellite images

Rasp et al. 2019

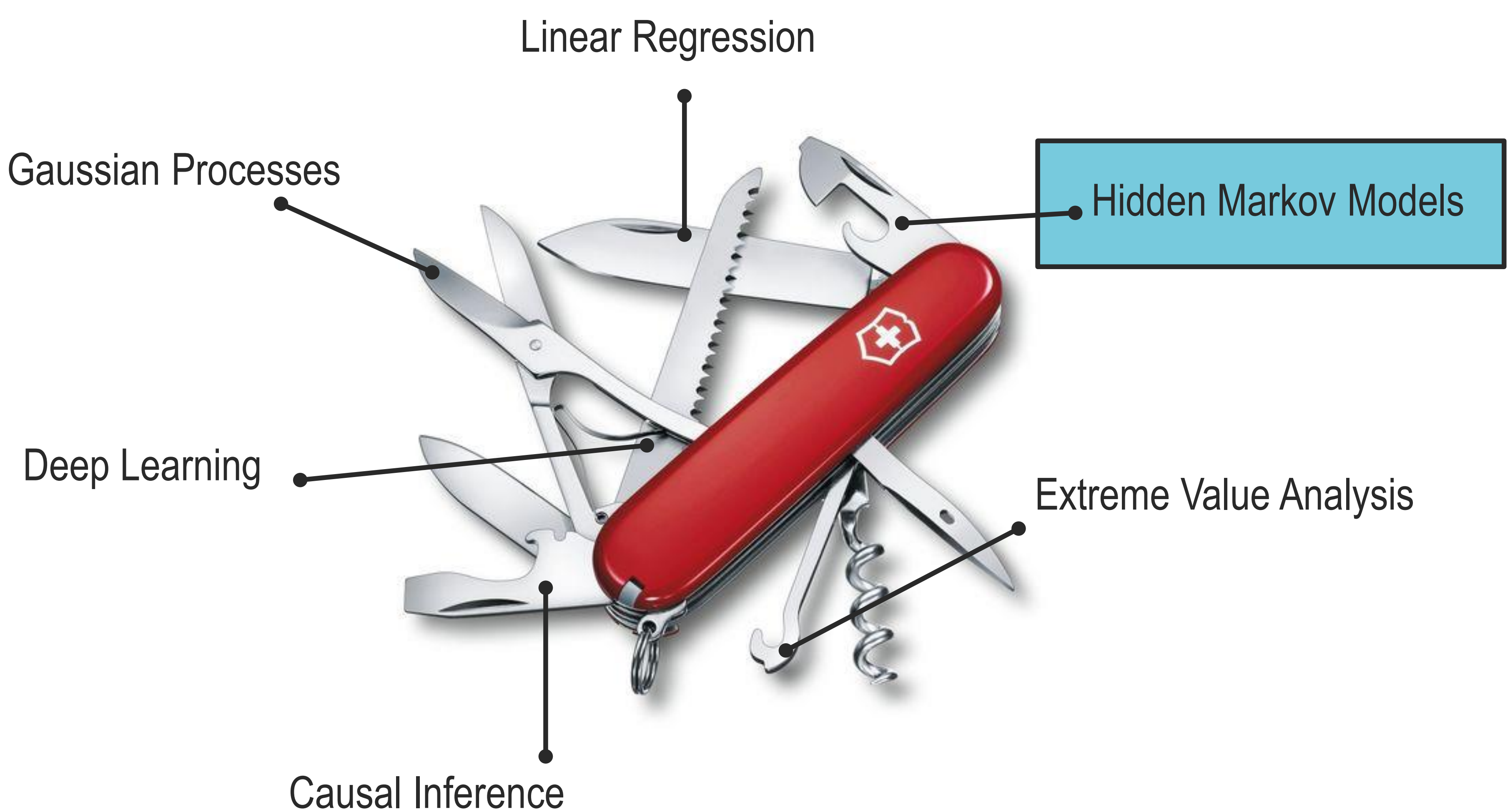


Understanding clouds from satellite images





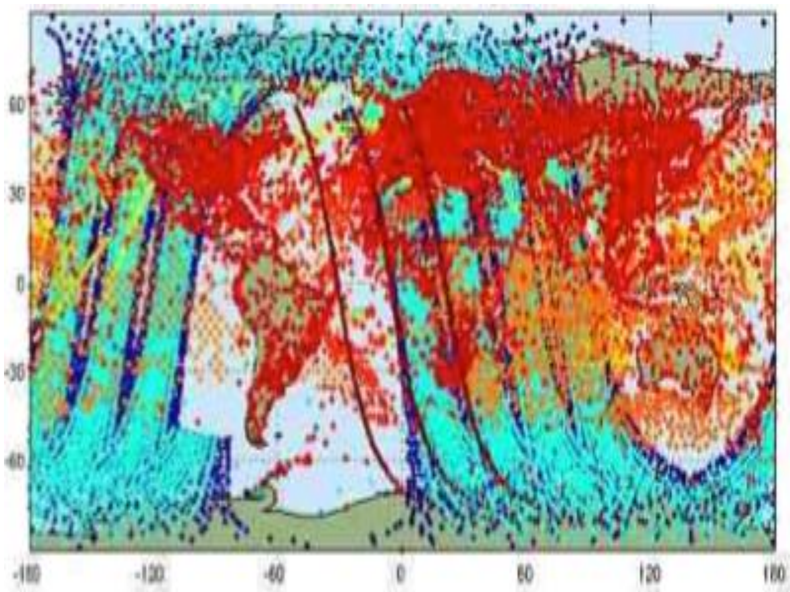
Outline

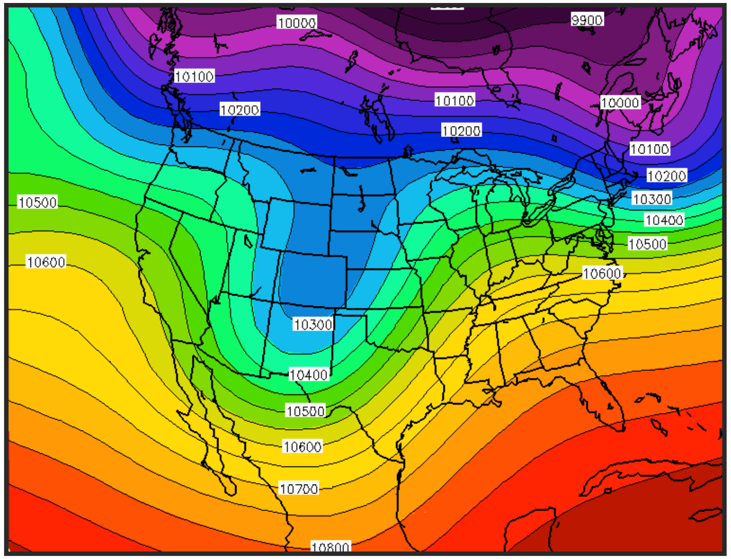

Linear Regression
Gaussian Processes
Hidden Markov Models
Deep Learning
Extreme Value Analysis
Causal Inference

Data Assimilation: hybrid approach stat + physical models

<u>Observations:</u>
multiple
sensors

<u>State vector:</u>
atmospheric
model





Numerical Weather Prediction requires to **initialize** the model every six hours with **new observations**.

Outlook of Data Assimilation

Trend: expansion towards new applications, general framework for interfacing large models and observations.

Examples:

- initialization
- reconstruction
- estimation: model parameters

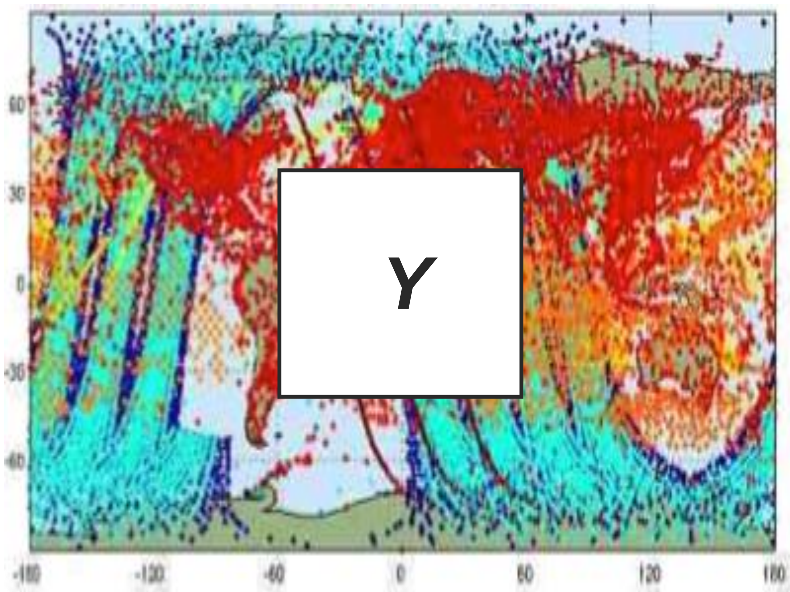
Proposal:

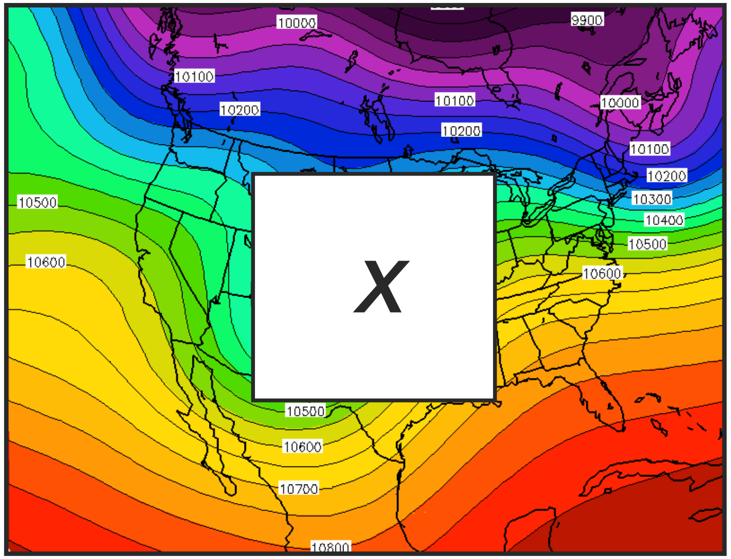
- model evaluation

Outlook of Data Assimilation

<u>Observations:</u>
multiple
sensors

<u>State vector:</u>
atmospheric
model





<u>Goal</u>: deriving the PDF of ***X*** conditional on ***Y*** = ***y***

high dimensional Bayesian update in a HMM

The “primitive equations” of data assimilation

Assumptions: Hidden Markov model

Solution: Gaussian linear approximation

- Dynamic equation:

$$\mathbf{X}_{t+1} = \mathrm{M}(\mathbf{X}_t, \mathbf{F}_t) + \mathbf{v}_t$$

- Observational equation:

$$\mathbf{Y}_t = \mathrm{H}(\mathbf{X}_t) + \mathbf{w}_t$$

- $\boldsymbol{v}_t$ and $\boldsymbol{w}_t$ Gaussian error terms with covariance ***Q*** and ***R***;
- M is the model with $\boldsymbol{F}_t$ external forcing;
- H is the observation operator.

- Propagation equation:

$$\mathbf{x}^f_{t+1} = \mathbf{M}\mathbf{x}^a_t$$

$$\mathbf{P}^f_{t+1} = \mathbf{M}\mathbf{P}^a_t\mathbf{M}' + \mathbf{Q}$$

- Update equation:

$$\mathbf{x}^a_t = \mathbf{x}^f_t + \mathbf{K}(\mathbf{y}_t - \mathbf{H}\mathbf{x}^f_t)$$

$$\mathbf{P}^a_t = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f_t$$

$$\mathbf{K} = \mathbf{P}^f_t\mathbf{H}'(\mathbf{H}\mathbf{P}^f_t\mathbf{H}' + \mathbf{R})^{-1}$$

The likelihood is a by-product of data assimilation

Solution:
Gaussian linear approximation

By-product:
PDF of observation ***y***

- Propagation equation:

$$\mathbf{x}^f_{t+1} = \mathrm{M}(\mathbf{x}^a_t)$$
$$\mathbf{P}^f_{t+1} = \mathrm{V}(\mathbf{x}^f_{t+1})$$

- Update equation:

$$\mathbf{x}^a_t = \mathbf{x}^f_t + \mathbf{K}(\mathbf{y}_t - \mathbf{H}\mathbf{x}^f_t)$$
$$\mathbf{P}^a_t = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f_t$$
$$\mathbf{K} = \mathbf{P}^f_t\mathbf{H}'(\mathbf{H}\mathbf{P}^f_t\mathbf{H}' + \mathbf{R})^{-1}$$

- Likelihood equation:

$$-\log p(\mathbf{y}) = \sum_{t=0}^{T} \frac{1}{2}\log|\boldsymbol{\Sigma}_t| + \frac{1}{2}\mathbf{d}'_t\boldsymbol{\Sigma}_t^{-1}\mathbf{d}_t$$

with:

$$\mathbf{d}_t = \mathbf{y}_t - \mathbf{H}\mathbf{x}^f_t$$
$$\boldsymbol{\Sigma}_t = \mathbf{H}\mathbf{P}^f_t\mathbf{H}' + \mathbf{R}$$

The likelihood is a by-product of data assimilation

- Likelihood :

$$-\log p(\mathbf{y}) = \sum_{t=0}^{T} \frac{1}{2} \log |\mathbf{\Sigma}_t| + \frac{1}{2} \mathbf{d}_t' \mathbf{\Sigma}_t^{-1} \mathbf{d}_t$$

with:

$$\mathbf{d}_t = \mathbf{y}_t - \mathbf{H}\mathbf{x}_t^f$$

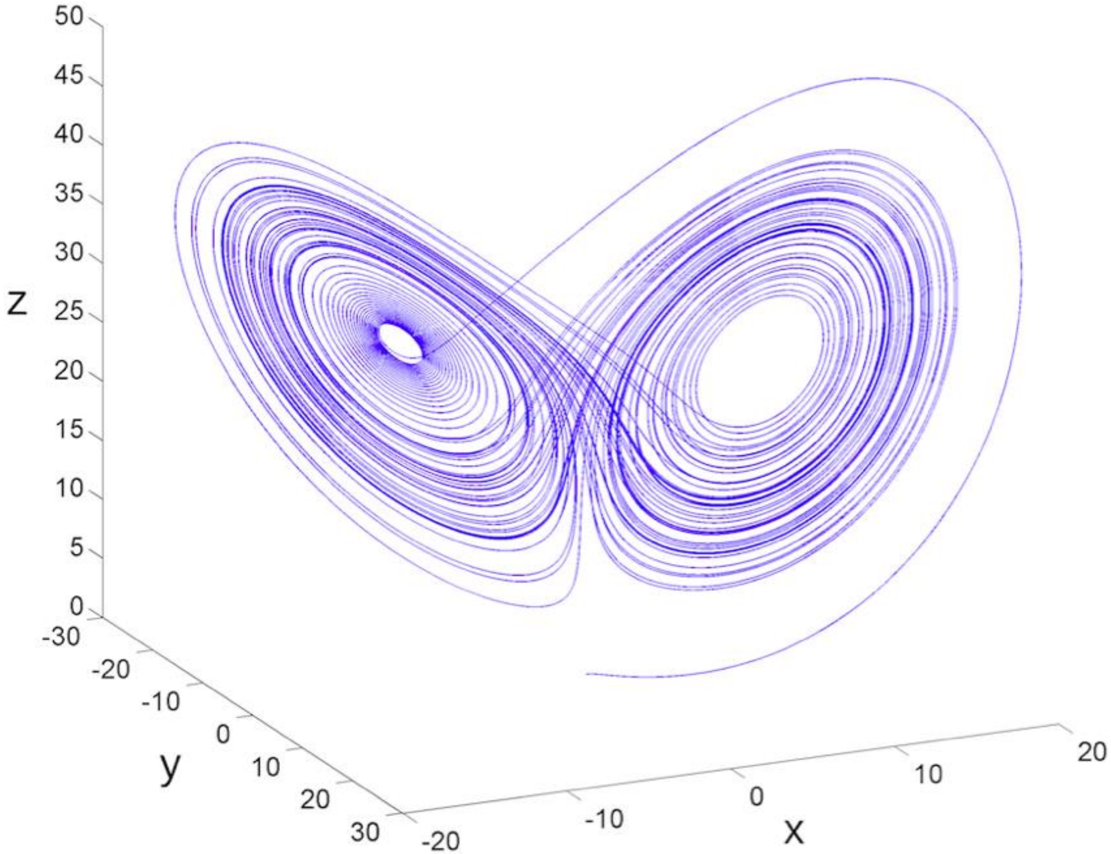
$$\mathbf{\Sigma}_t = \mathbf{H}\mathbf{P}_t^f \mathbf{H}' + \mathbf{R}$$

➡ Accounts for observational noise and inhomogeneity

➡ Spatial-temporal-variable aggregation

Test in the forced Lorenz model

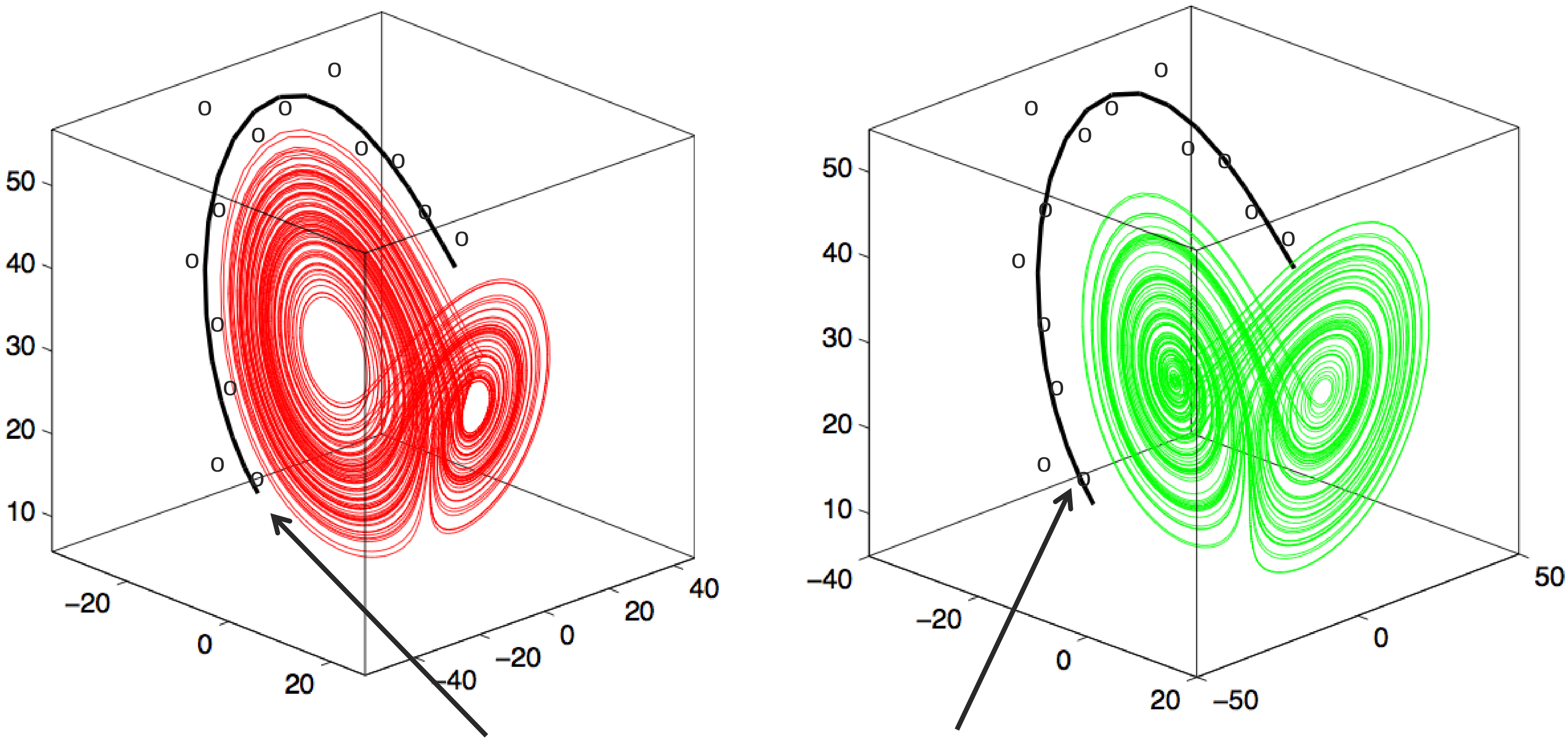
A trajectory in the state space



$$\begin{cases} \dfrac{\mathrm{d}x}{\mathrm{d}t} = \sigma(y-x) + \lambda_i \cos\theta_i \\ \dfrac{\mathrm{d}y}{\mathrm{d}t} = \rho x - y - xz + \lambda_i \sin\theta_i \\ \dfrac{\mathrm{d}z}{\mathrm{d}t} = xy - \beta z \end{cases}$$

Test in the forced Lorenz model

Model 1 (λ = 40) Model 2 (λ = 0)

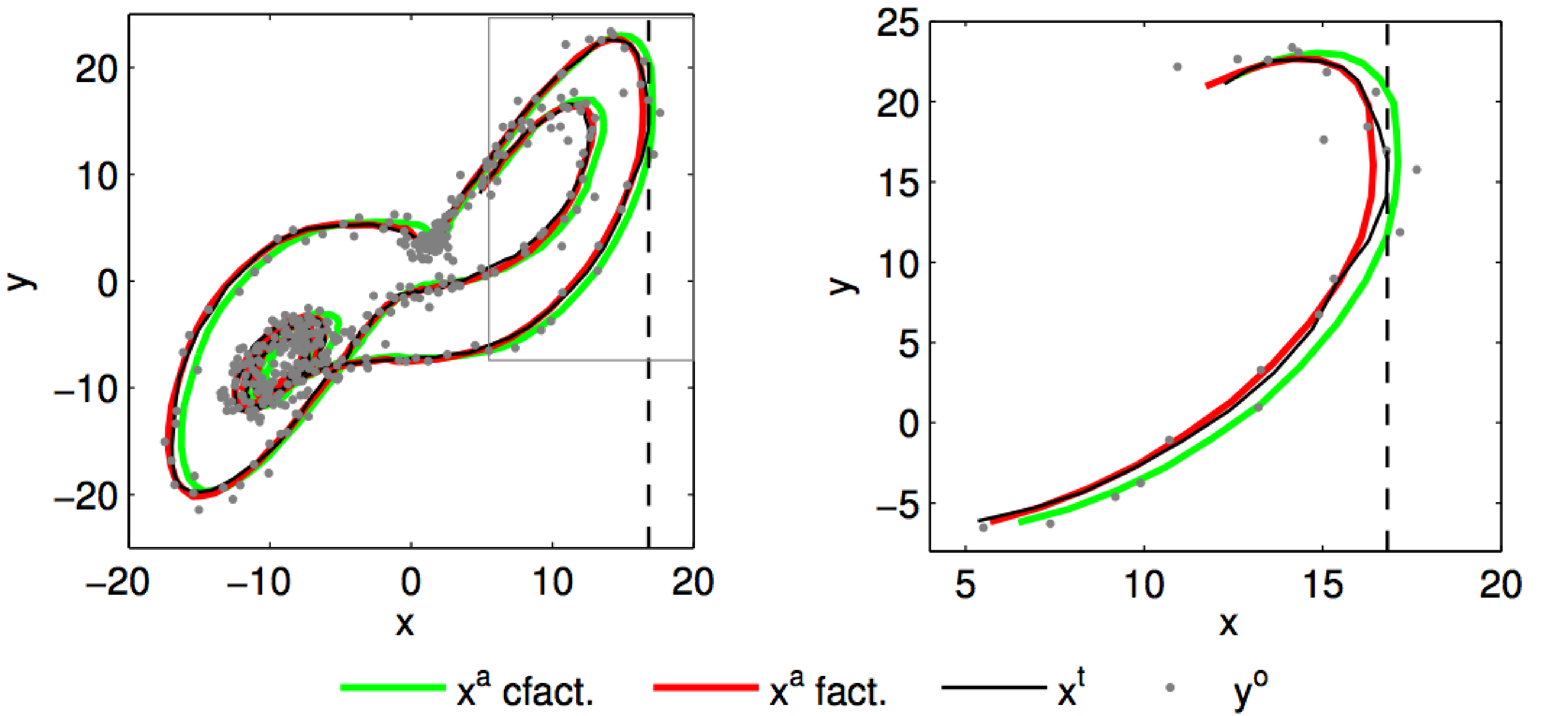


piece of a trajectory + observations

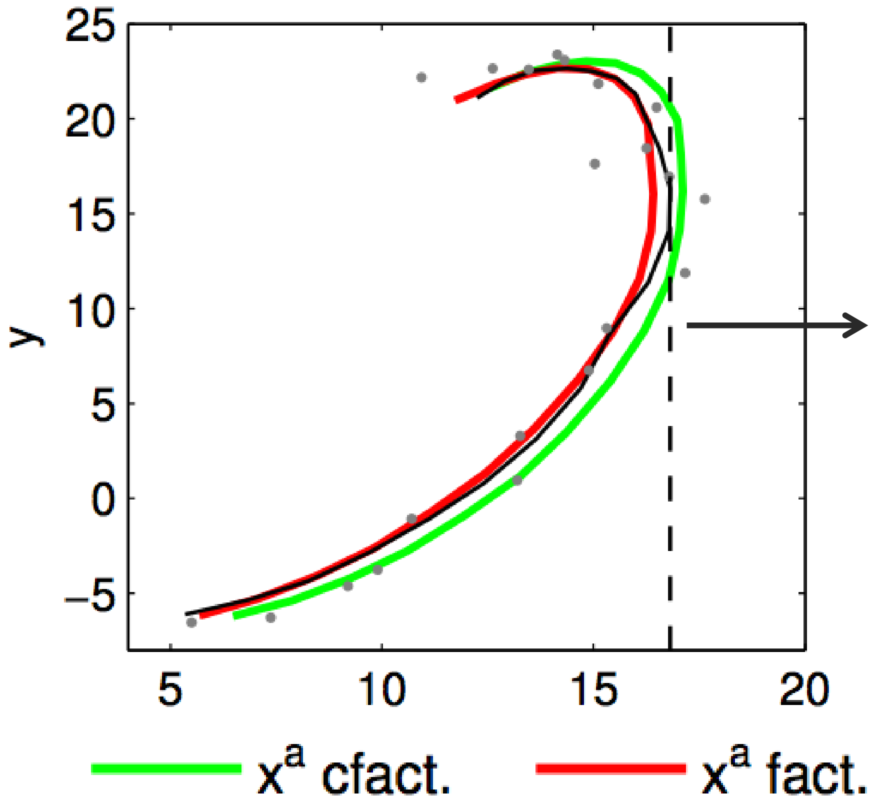
Which is best? (observations come from model 1)

Test in the forced Lorenz model

- Marginal likelihood of the observed trajectory is derived for both models by assimilating observations.



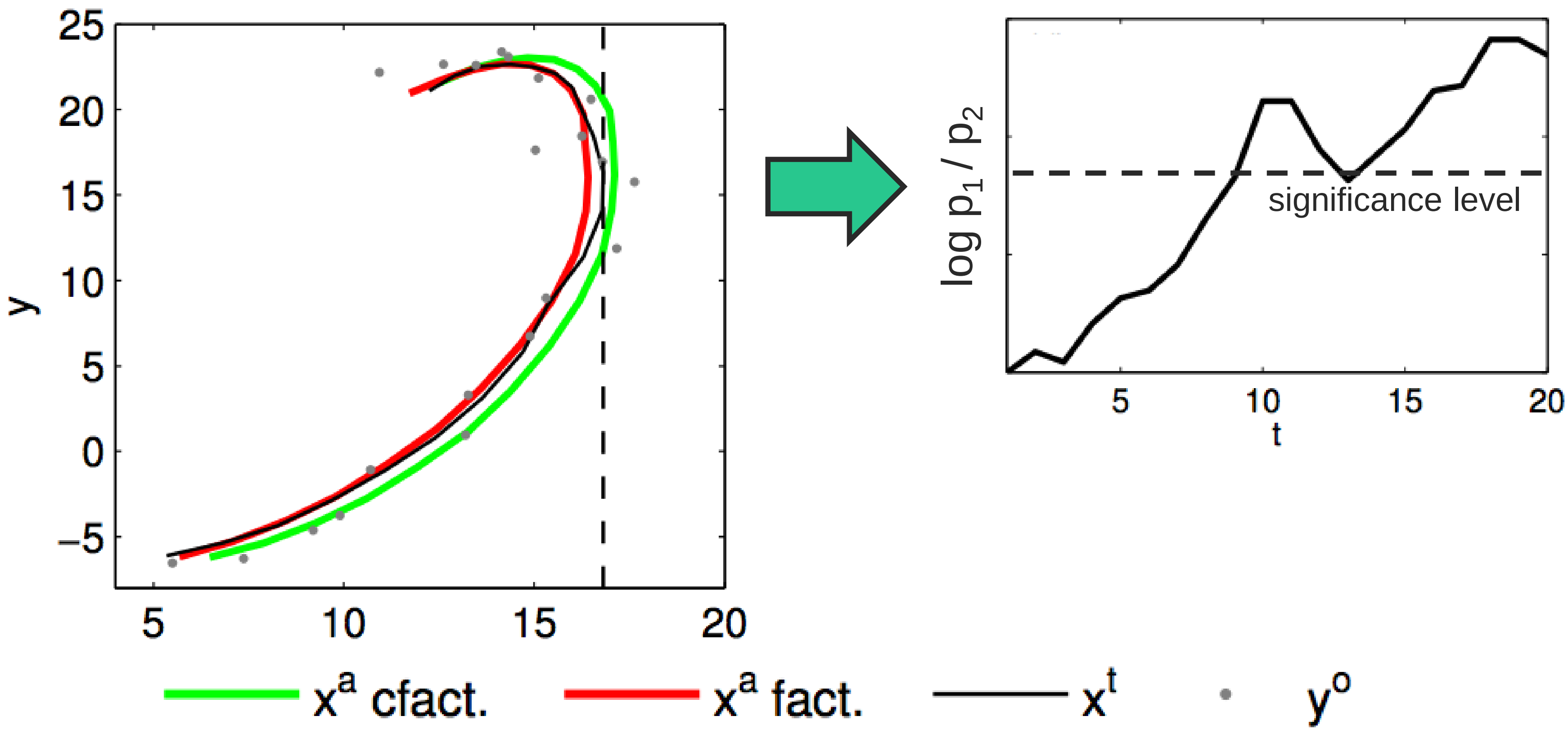
Test in the forced Lorenz model



The reconstruction of the correct model is usually slightly better than the one of the wrong model.

Test in the forced Lorenz model

- The reconstruction of the correct model is usually slightly better than the one of the wrong model.
- Small local differences pile up into a large amount of likelihood difference overall.



Summary

- Marginal likelihood appears to be a possible metric to evaluate the ability of a model to represent a given sequence of observations.

- Data Assimilation appears to be a reasonable solution to compute marginal likelihood.

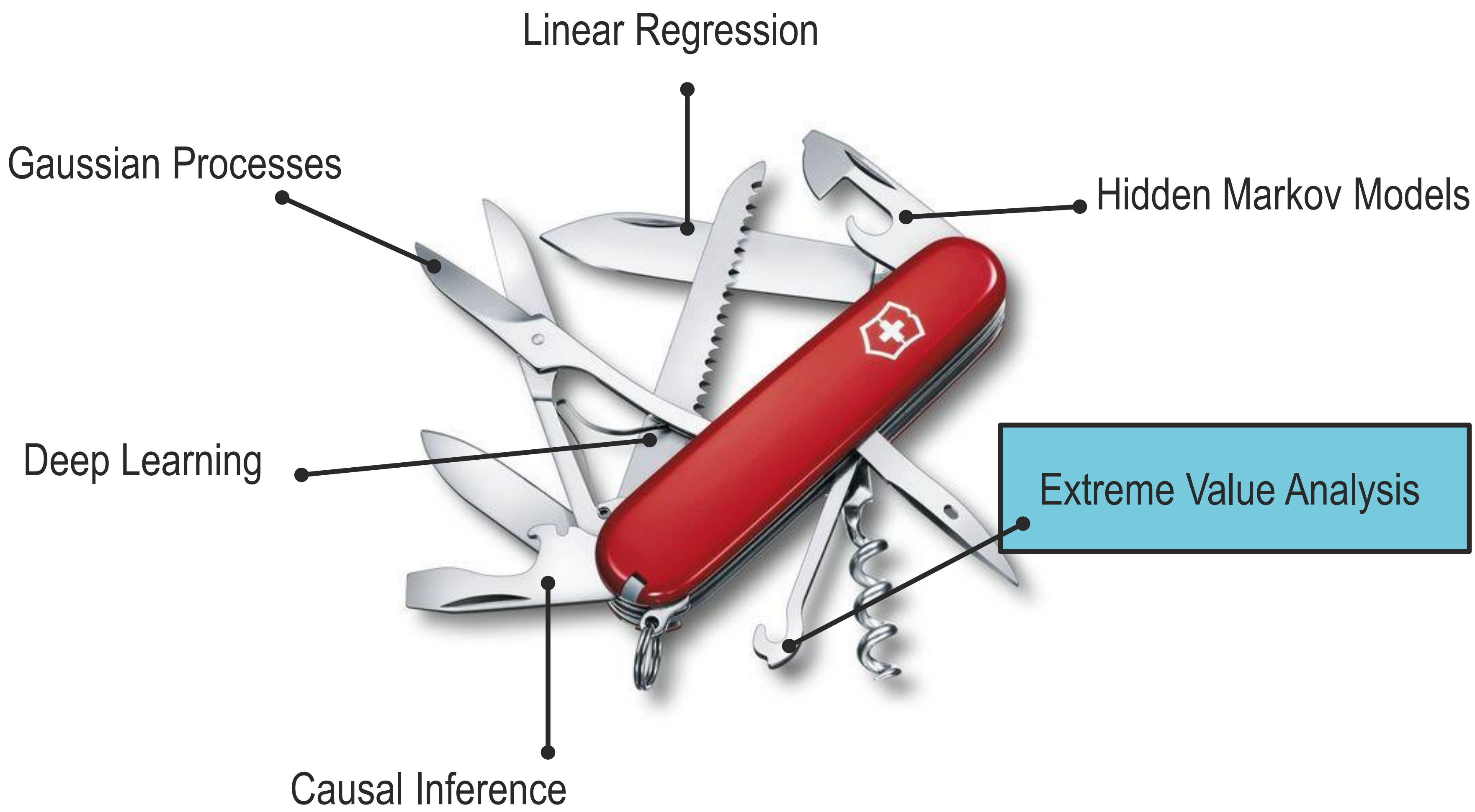
- Offers the advantage to synergize with existing infrastructure and expertise, especially regarding observational error.

- Research under way:
  - Experiments using larger models (ICTP AGCM, WRF)
  - Implementation on real case studies.
  - Theoretical and practical challenges for computing the likelihood (determinant, localization, …)

Outline


Linear Regression
Gaussian Processes
Hidden Markov Models
Deep Learning
Extreme Value Analysis
Causal Inference

Outline



Linear Regression
Gaussian Processes
Hidden Markov Models
Deep Learning
Extreme Value Analysis
Causal Inference

Outline

- General considerations
- Statistical Methods & Illustrations
- Conclusion

Conclusion

- The emergence and improved access to large datasets and increased computational power, transformed the field of applied statistics and computer science.

- New tools are needed to handle large data. The emergence of new tools creates new approaches, applications, products, findings.

- Many areas of climate science and climate services are concerned by these evolutions.

- However, especially in clilmate science, problems for which small data prevails remain many, and are still a very important aspect in applied statistics.

Thank you

World Economic Forum report 2018

WORLD ECONOMIC FORUM

COMMITTED TO IMPROVING THE STATE OF THE WORLD

Fourth Industrial Revolution for the Earth Series

Harnessing Artificial Intelligence for the Earth

Climate change

- Clean power
- Smart transport options
- Sustainable production and consumption
- Sustainable land-use
- Smart cities and homes

Biodiversity and conservation

- Habitat protection and restoration
- Sustainable trade
- Pollution control
- Invasive species and disease control
- Realizing natural capital

Healthy Oceans

- Fishing sustainably
- Preventing pollution
- Protecting habitats
- Protecting species
- Impacts from climate change (including acidification)

Water security

- Water supply
- Catchment control
- Water efficiency
- Adequate sanitation
- Drought planning

Clean air

- Filtering and capture
- Monitoring and prevention
- Early warning
- Clean fuels
- Real-time, integrated, adaptive urban management

Weather and disaster resilience

- Prediction and forecasting
- Early warning systems
- Resilient infrastructure
- Financial instruments
- Resilience planning

Climate Informatics, NCAR, 2014 to present

NCAR UCAR | CISL Computational & Information Systems Lab

Software | Data Portals | User Support | Resources | About Us

Daily Bulletin | Resource Status | Newsroom | Events

CLIMATE INFORMATICS

- Home
- Important Dates
- Registration
- Application for Travel Support
- Paper Submission Guidelines
- Poster Guidelines
- Hackathon

HOME

CI 2017

7th International Workshop on Climate Informatics

September 20-22, 2017

Hosted by the National Center for Atmospheric Research in Boulder, CO

Climate Informatics Workshop

About Climate Informatics

We have greatly increased the volume and diversity of climate data from satellites, environmental sensors and climate models in order to improve our understanding of the climate system. However, this very increase in volume and diversity can make the use of traditional analysis tools impractical and necessitate the need to carry out knowledge discovery from data. Machine learning has made significant impacts in fields ranging from web search to bioinformatics, and the impact of machine learning on climate science could be as profound. However, because the goal of machine learning in climate science is to improve our understanding of the climate system, it is necessary to employ techniques that go beyond simply taking advantage of co-occurence, and, instead, enable increased

AI in weather and climate, Montreal, July 2019

JM07 - Artificial Intelligence and Big data in Weather and Climate Science (IAMAS, IAHS)

Convener: Philippe Roy (Canada, IAMAS)

Co-Conveners: Alexis Hannart (Canada, IAMAS), David Hall (USA, IAMAS), Allen Huang (USA, IAMAS), Ashish Sharma (Australia, IAHS)

Description

Rapid advances in artificial intelligence, combined with the availability of enormous amount of data (termed Big Data) is opening new avenues for climate analysis and climate scenarios. The long awaited promises of AI is now common in many disciplines. Applying AI methods, combined with physical knowledge, can improve climate analysis and provide better climate simulations and climate products, notably for high-impact events, such as floods, wildfires and winds.

Contributions are welcome in the following areas, but not limited to:

- Decision-making tools for climate and weather related hazards;
- Data mining and explorations approaches
- Pattern recognition and classification
- Climate and weather emulators
- Smart-grid and smart cities applications combining AI and weather and climate data
- Novel approaches in the domain of natural hazards using AI methods

Prospective considerations on AI – JASON report

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD-MM-YYYY)* January 2017	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*

4. TITLE AND SUBTITLE Perspectives on Research in Artificial Intelligence and Artificial General Intelligence Relevant to DoD	**5a. CONTRACT NUMBER** **5b. GRANT NUMBER** **5c. PROGRAM ELEMENT NUMBER**
6. AUTHOR(S)	**5d. PROJECT NUMBER** 1316JA01 **5e. TASK NUMBER** PS **5f. WORK UNIT NUMBER**
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The MITRE Corporation JASON Program Office 7515 Colshire Drive, MS T130 McLean, Virginia 22102	**8. PERFORMING ORGANIZATION REPORT NUMBER** JSR-16-Task-003
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) OSD ASDR&E Basic Research Labs 4800 Mark Center Drive Alexandria VA	**10. SPONSOR/MONITOR'S ACRONYM(S)** **11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

12. DISTRIBUTION / AVAILABILITY STATEMENT

Distribution authorized for Public Release

13. SUPPLEMENTARY NOTES

Artificial Intelligence (AI) is conventionally, if loosely, defined as intelligence exhibited by machines. Operationally, it can be defined as those areas of R&D practiced by computer scientists who identify with one or more of the following academic sub-disciplines: Computer Vision, Natural Language Processing (NLP), Robotics (including Human-Robot Interactions), Search and Planning, Multi-agent Systems, Social Media Analysis (including Crowdsourcing), and Knowledge Representation and Reasoning (KRR). The field of Machine Learning (ML) is a foundational basis for AI. While this is not a complete list, it captures the vast majority of AI researchers.

Artificial General Intelligence (AGI) is a research area within AI, small as measured by numbers of researchers or total funding, that seeks to build machines that can successfully perform *any* task that a human might do. Where AI is oriented around specific tasks, AGI seeks general cognitive abilities. On account of this ambitious goal, AGI has high visibility, disproportionate to its size or present level of success, among futurists, science fiction writers, and the public.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Dr. Robin Staffin
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified	UL		**19b. TELEPHONE NUMBER** *(include area code)* 571-372-6460

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Prospective considerations on AI – JASON report

- BD/DL is the mainstream paradigm of AI thus far:
  - Big Data (10^4 - 10^7 examples) combined with Deep Learning,
  - DL is by now a well documented and well accessible expertise.

3.8 Summary of the Big Data Deep Learning "Dogma"

The powerful successes of Big Data / Deep Learning have given it the status of a kind of dogma—a set of principles that, when followed, lead often to unexpectedly powerful successes. A brief summary of these principles might be the following:

- Use deep (where possible, very deep) neural nets. Use convolutional nets, even if you don't know why (that is, even if the underlying problem is not translation invariant).
- Adopt flat numerical data representations, where the input is a vector of reals and the internal representation (for a DNN, the activations) is an even larger number of reals. Avoid the use of more complicated data structures. The model will discover any necessary structure in the data from its flat representation.
- Train with big (*really* big) data. Don't load on model assumptions, but rather learn everything from the data—that is where the truth lies. As an example, don't attempt to hardwire the laws of aerodynamics into an autopilot application. With enough data, it is more efficient to let the DNN discover them on its own.
- An approximate answer is usually good enough. When it works, it is not necessary to understand why or how.

Prospective considerations on AI – JASON report

- BD/DL is probably not the end of the story in IA:
  - Small Data (10^2 - 10^4 examples) is not unfrequent.
  - Explainability / reliability / causality is often requested and yet not particularly amenable to DL.

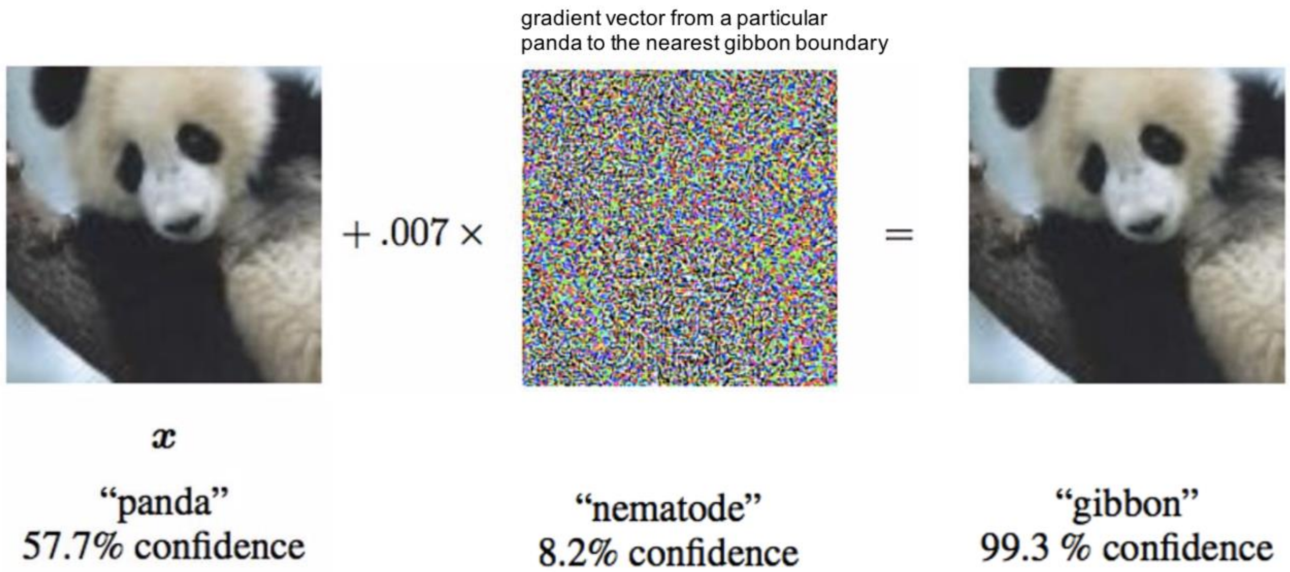


Figure 15: One can get from the panda classification to the gibbon classification by adding what appears to us to be noise. The resulting image looks to us like a panda, but it looks to the DNN like a gibbon, with 99.3% confidence. Source: see footnote [36].

Prospective considerations on AI – JASON report

> **5 AREAS OF RAPID PROGRESS OTHER THAN DEEP LEARNING**
>
> While the “Big Data / Deep Learning dogma”, as summarized above in Section 3.4, has rightly captured the imagination of experts and the lay public alike, there is some danger of its overshadowing some other areas of AI that are advancing rapidly and hold significant future promise, including in DoD applications. In this Chapter, we review what we think are the most important of these.

- Next possible hot topics:
  - Probabilistic graphical models / Bayesian networks, Gaussian processes,
  - Probabilistic generative models / Bayesian priors,
  - Hybridization with other tools (numerical physical models, agent models).

➡ What’s coming next will likely originate from the field of applied statistics.

➡ BD/DL is key, yet a «pure play» BD/DL scientific strategy is arguably risky.