Applied Statistics in the Climate Sciences
A mild overview

Alexis Hannart
12 September 2019
Outline

- General considerations
- Statistical Methods & Illustrations
- Conclusion
# Mathematics of the Climate and Environment

## CliMathParis2019 Summer School Planning

<table>
<thead>
<tr>
<th>Time</th>
<th>Sun 08 Sept</th>
<th>Mon 09 Sept</th>
<th>Tue 10 Sept</th>
<th>Wed 11 Sept</th>
<th>Thurs 12 Sept</th>
<th>Fri 13 Sept</th>
<th>Sat 14 Sept</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00</td>
<td>Le Treut: Phys. of Climate I</td>
<td>Bocquet: Data Assim I</td>
<td>Djikstra: Math Methods III</td>
<td>Bruna: Big Data I</td>
<td>Dubos: Num. Methods I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:20</td>
<td>Break</td>
<td>Break</td>
<td>Break</td>
<td>Break</td>
<td>Break</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:40</td>
<td>ARRIVAL IN CARGÈSE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12:00</td>
<td>Lunch (IESC)</td>
<td>Lunch (IESC)</td>
<td>Lunch (IESC)</td>
<td>Lunch (IESC)</td>
<td>Lunch (IESC)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14:00</td>
<td>Ghil: Math Methods I</td>
<td>Brovkin: Climate-Ecology I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15:20</td>
<td>Break</td>
<td>Break</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15:40</td>
<td>Ghil: Math Methods II</td>
<td>Brovkin: Climate-Ecology II</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17:00</td>
<td>Poster Session</td>
<td>Poster Session</td>
<td>Poster Session</td>
<td>Poster Session</td>
<td>Poster Session</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18:00</td>
<td>Poster Session</td>
<td>Poster Session</td>
<td>Poster Session</td>
<td>Poster Session</td>
<td>Poster Session</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Departure from CARGÈSE
Randomness and determinism

- Is the climate system deterministic or probabilistic?
Randomness and determinism

- Is the climate system deterministic or probabilistic?
- Is this question nonsense?
Randomness and determinism
Randomness and determinism

Head or tail?  
Is it random?
Generating randomness: flipping a coin

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

Figure 2: Coordinates of Precessing Coin.

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

Diaconis et al. 2007
Generating randomness
Generating randomness

‘Middle-square’ algorithm

\[
\begin{align*}
\text{seed} & \rightarrow 675248 \\
\text{seed}^2 & \rightarrow 455959861504 \\
\downarrow & \\
959861 & \rightarrow \text{output}
\end{align*}
\]
Generating randomness

‘Middle-square’ algorithm

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

‘Middle-square’ algorithm

\[ \begin{align*}
\text{seed} & \quad \rightarrow \quad 675248 \\
675248^2 & \quad \rightarrow \quad 455959861504 \\
\text{output} & \quad \rightarrow \quad 959861
\end{align*} \]

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.
Illustration: Lorenz 96 model

Numerical model

\[ \frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F \]
Numerical model

\[ \frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F \]

Probabilistic model

\[ x = (x_{i,t})_{i=1,\ldots,40 \ t=1,\ldots,T} \]
\[ p(x) = \mathcal{N}(\mu, \Sigma) \]
Illustration: midlatitude temperature

Observations (ERAint)
surface temperature anomaly, 45N, year 2016
Outline

1. General considerations
2. Statistical Methods & Illustrations
3. Conclusion
Theory and Application

- Mathematics
Theory and Application

- Mathematics
- Applied Mathematics
Theory and Application

- Mathematics
- Applied Mathematics
- Statistics
Theory and Application

- Mathematics
- Applied Mathematics
- Statistics
- Applied Statistics
Theory and Application

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

- Mathematics
- Applied Mathematics
- Statistics

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

- Mathematics
- Applied Mathematics
- 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 computational 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
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.

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.

Theory and Application

- Mathematics
- Applied Mathematics
- Statistics
- Applied Statistics
- Application of Applied Statistics to theoretical problems
- Application of Applied Statistics to applied problems
## Theory and Application

<table>
<thead>
<tr>
<th>Theory</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mathematics</td>
<td>Application of Applied Statistics to theoretical problems</td>
</tr>
<tr>
<td>Applied Mathematics</td>
<td>Application of Applied Statistics to applied problems</td>
</tr>
<tr>
<td>Statistics</td>
<td></td>
</tr>
<tr>
<td>Applied Statistics</td>
<td></td>
</tr>
</tbody>
</table>
Theory and 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

- Linear Regression
- Gaussian Processes
- Deep Learning
- Causal Inference
- Hidden Markov Models
- Extreme Value Analysis
Areas of applications in climate science
Areas of applications in climate science
Outline

- General considerations
- Statistical Methods & Illustrations
- Conclusion
Basic principles

\[ p(x \mid \theta) \]
Basic principles

\[ p(x \mid \theta) \]

\[ \hat{\theta} = \operatorname{argmax}_\theta \log p(x \mid \theta) \]
Basic principles

\[ p(x \mid \theta) \]

\[ p(\theta) \]
Basic principles

\[ p(x \mid \theta) \]

\[ p(\theta) \]

\[ p(\theta \mid x) \propto p(x \mid \theta) \cdot p(\theta) \]
Basic principles

\[ p(x \mid \theta) \]

\[ p(\theta) \]

\[ p(\theta \mid x) \propto p(x \mid \theta) \cdot p(\theta) \]

\[ \hat{\theta} = \arg\max_\theta p(\theta \mid x) \]
Basic principles

\[ p(x \mid \theta) \]
\[ p(\theta) \]
\[ p(\theta \mid x) \propto p(x \mid \theta) \cdot p(\theta) \]
\[ \hat{\theta} = \text{argmax}_\theta p(\theta \mid x) \]
\[ \hat{\theta} = \mathbb{E}(\theta \mid x) \]
Basic principles

\[ p(x \mid \theta) \]

\[ p(\theta) \]

\[ p(\theta \mid x) \propto p(x \mid \theta) \cdot p(\theta) \]

\[ \hat{\theta} = \arg\max_{\theta} p(\theta \mid x) \]

\[ \hat{\theta} = \mathbb{E}(\theta \mid x) \]

\[ \hat{\theta} = \arg\max_{\theta^*} \mathbb{E}(C(\theta, \theta^*) \mid x) \]
Linear regression

\[ y = x\beta + \varepsilon \]

\[ p(y \mid x, \beta) = \mathcal{N}(x\beta, \sigma^2 I) \]

\[ \hat{\beta} = (x'x)^{-1}(x'y) \]
Linear regression

Impact variable \quad \text{Climate variables}

\[ y = \mathbf{x} \beta + \varepsilon \]

\[ p(y \mid \mathbf{x}, \beta) = \mathcal{N}(\mathbf{x} \beta, \sigma^2 \mathbf{I}) \]
Observations: Yields by crop, year and country.
Data

Observations:
Yields by crop, year and country (FAO).

Observations:
Growing season temperature by crop, year and country (HadCRUT).

yield (maize)
growing season temperature (maize)
Model

log(yield)_{cit} = \beta_{1c} T_{cit} \times I_c + \beta_{2c} T_{cit}^2 \times I_c + \beta_{3c} P_{cit} \times I_c + \beta_{4c} P_{cit}^2 \times I_c + \beta_{5c} t \times I_c \times I_i + \beta_{6c} t^2 \times I_c \times I_i + \text{Irr}_{it} \times I_c + \text{Fert}_{it} \times 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 \rightarrow \begin{equation}
y = \mathbf{x} \beta + \varepsilon
\end{equation}

\begin{equation}
p(y \mid \mathbf{x}, \beta) = \mathcal{N}(\mathbf{x} \beta, \sigma^2 \mathbf{I})
\end{equation}
Basins
Data
Results
Skill
Attribution

Evidencing the causal influence of external factors
Magnitude of the warming rate

- 1° C / 10,000 years (typical natural variation)
- 1° C / millenium (fast natural variation)
- 1° C / century (observed recent variation)

- Fewly compatible with natural variations
- Quite compatible with anthropogenic forcings
Magnitude of the warming rate

Typical variations

- Rate of change: $1^\circ$ C / 10,000 years
Magnitude of the warming rate

Last deglaciation

- Rate of change:
  - 1°C / 1000 years
Magnitude of the warming rate

Last century
- Rate of change: 1°C / 100 years
Conventional method for attributing trends

\[ y = x\beta + \nu \]

Observations \rightarrow GCM patterns \rightarrow 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 \rightarrow GCM patterns \rightarrow 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

\[
\begin{align*}
    y &= \mathbf{x} \beta + \nu \\
    \text{Var}(\nu) &= \Sigma \\
    \mathbf{x} &= (x_1, \ldots, x_p)
\end{align*}
\]

Inference: projection of the data

\[
\begin{align*}
    T_y &= T \mathbf{x} \beta + T \nu \\
    T \Sigma T' &= I \\
    \hat{\beta} &= (\mathbf{x}' \Sigma^{-1} \mathbf{x})^{-1} (\mathbf{x}' \Sigma^{-1} y)
\end{align*}
\]
Optimal projection
Two steps approach
Integrated approach

Hannart 2016
Integrated likelihood

\[
\begin{align*}
\hat{\alpha} &= \arg\max_{\alpha \in [0,1]} \{ \log \ell(\alpha) \} \\
\hat{\beta} &= (x' \Sigma_{\hat{\alpha}}^{-1} x)^{-1} (x' \Sigma_{\hat{\alpha}}^{-1} y) \\
-2 \log \ell(\alpha) &= \phi \left( \frac{r}{1-\alpha} + 1 \right) - \phi \left( \frac{ar}{1-\alpha} \right) - n \left( \frac{r}{1-\alpha} + n + 2 \right) \log \left( \frac{1-\alpha}{r} + 1 \right) \\
&\quad + \left( \frac{r}{1-\alpha} + n + 2 \right) \log |\Sigma_{\alpha}| - \left( \frac{ar}{1-\alpha} + n + 1 \right) \log |\Delta| \\
&\quad + \left( \frac{r}{1-\alpha} + n + 2 \right) \log \left\{ 1 + \frac{(1-\alpha)n}{r} F_{\alpha} \right\}
\end{align*}
\]
Skill

MSE of the estimator

Reliability of the confidence interval

MSE

r

ROF  IOF1  IOF2  IOF3
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.
Idea of “upstream prediction”

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

- Idea: wind farm + “integrated forecasting network”
Data
\[ \mathbf{x} = (x_{s,t-\tau})_{s \in S, \tau = 0,1,\ldots,T} \]
Notation

\[ \mathbf{x} = (x_{s, t-\tau})_{s \in S, \tau = 0, 1, \ldots, T} \]

assumed to be a multivariate Gaussian with covariance \( \Sigma \)
Notation

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

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

regularized estimate of $\Sigma$
The prediction follows:

\[
\mathbb{E}(x_0 \mid x_1) = x_1 \beta \quad \beta = \Sigma_{11}^{-1} \Sigma_{10}
\]
Correlogram of $\Sigma$

The correlogram reflects the non-separability of the dynamic of the flow.
Covariance regularization: low rank representation

\[ \Sigma = \text{Var}(\mathbf{x}) = \mathbf{V}_r \Delta_r \mathbf{V}_r' + \lambda \mathbf{I} \rightarrow \text{nugget} \]

r basis functions are retained (r<<p)
Covariance regularization: low rank representation

\[ \Sigma = \text{Var}(x) = V_r \Delta_r V_r' + \lambda I \rightarrow \text{nugget} \]

- \( r \) basis functions are retained \((r<<p)\)
- Ad hoc wave propagation basis functions probably exist
- But use of EOFs
Eigenvectors

The eigenvectors of the covariance $\Sigma$ are dynamic maps. EOFs also exhibit wave-like moving patterns.
Skill of covariance estimation

time

truth

emulator
Estimated weights

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

t-6  t-5  t-4  t-3  t-2  t-1
Skill of prediction

~85% mse reduction
Interpolating temperature missing values

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

Context of this research work:
- 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

\[ \mathbf{x} = (x_{s,t-\tau})_{s \in \mathcal{S}, \tau = 0, 1, ..., T} \]

assumed to be a multivariate Gaussian with covariance \( \Sigma \)
• #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.
ReLU function (Rectified Linear Unit)

\[ \sigma(u) = \max(u, 0) \]
Deep learning

\[ y = \sigma ( \mathbf{W}_d \sigma ( \mathbf{W}_{d-1} \sigma ( \ldots \sigma ( \mathbf{W}_1 \mathbf{x} ) ) )) + \varepsilon \]
$y = \sigma \left( W_d \sigma \left( W_{d-1} \sigma \left( \ldots \sigma \left( W_1 x \right) \right) \right) \right) + \varepsilon$
Deep learning

\[ y = \sigma(\mathbf{W}_d \sigma(\mathbf{W}_{d-1} \sigma(\ldots \sigma(\mathbf{W}_1 \mathbf{x})))) + \varepsilon = \phi(\mathbf{x}, \mathbf{W}) + \varepsilon \]

\[ p(y \mid \mathbf{x}, \mathbf{W}) = \mathcal{N}(y \mid \phi(\mathbf{x}, \mathbf{W}), \lambda \mathbf{I}) \]
Deep learning

\[ y = \sigma \left( W_d \sigma \left( W_{d-1} \sigma \left( \ldots \sigma \left( W_1 x \right) \right) \right) \right) + \varepsilon = \phi(x, W) + \varepsilon \]

\[ p(y \mid x, W) = \mathcal{N}(y \mid \phi(x, W), \lambda I) \]

\[ \hat{W} = \arg\max_W \log p(y \mid x, W) \]

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

Encoders

Decoders
Hurricane attribution

8 September 2017 06.00pm GMT
Temporal plot of tropical cyclones occurrences

Spatial plot of tropical cyclones tracks
Individual trajectories
Individual trajectories: dimension reduction
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

- The "Cloud Brain"
  - Can the 140M outputs from 1 year of 9k Cloud Resolving Models...
    - {solutions of accurate radiative transfer & explicit CRM equations}
  - 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.

Geophysical Research Letters
Could machine learning break the convection parameterization deadlock?
Some encouraging early results

Rasp et al. 2018
A promising way forward

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

Sugar
Dusting of very fine clouds, little evidence of self-organization

Flower
Large-scale stratiform cloud features appearing in bouquets, well separated from each other.

Fish
Large-scale skeletal networks of clouds separated from other cloud forms.

Gravel
Meso-beta lines or arcs defining randomly interacting cells with intermediate granularity.
Understanding clouds from satellite images
Anybody interested?
Outline

- Linear Regression
- Gaussian Processes
- Deep Learning
- Causal Inference
- Hidden Markov Models
- Extreme Value Analysis
Data Assimilation: hybrid approach stat + physical models

Observations: multiple sensors

State vector: 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

**Observations:** multiple sensors

**State vector:** atmospheric model

**Goal:** 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

- Dynamic equation:
  \[ X_{t+1} = M(X_t, F_t) + v_t \]
- Observational equation:
  \[ Y_t = H(X_t) + w_t \]
- \( v_t \) and \( w_t \) Gaussian error terms with covariance \( Q \) and \( R \);
- \( M \) is the model with \( F_t \) external forcing;
- \( H \) is the observation operator.

Solution: Gaussian linear approximation

- Propagation equation:
  \[ x_{t+1}^f = Mx_t^a \]
  \[ P_{t+1}^f = MP_t^a M' + Q \]
- Update equation:
  \[ x_t^a = x_t^f + K(y_t - Hx_t^f) \]
  \[ P_t^a = (I - KH)P_t^f \]
  \[ K = P_t^f H'(HP_t^f H' + R)^{-1} \]
The likelihood is a by-product of data assimilation

Solution: Gaussian linear approximation

- Propagation equation:
  \[
  x_{t+1}^f = M(x_t^a)
  \]
  \[
  P_{t+1}^f = V(x_{t+1}^f)
  \]

- Update equation:
  \[
  x_t^a = x_t^f + K(y_t - Hx_t^f)
  \]
  \[
  P_t^a = (I - KH)P_t^f
  \]
  \[
  K = P_t^f H'(HP_t^f H' + R)^{-1}
  \]

By-product: PDF of observation \( y \)

- Likelihood equation:
  \[
  -\log p(y) = \sum_{t=0}^{T} \frac{1}{2} \log |\Sigma_t| + \frac{1}{2} d_t' \Sigma_t^{-1} d_t
  \]

with:
  \[
  d_t = y_t - Hx_t^f
  \]
  \[
  \Sigma_t = HP_t^f H' + R
  \]
The likelihood is a by-product of data assimilation

- **Likelihood:**

\[- \log p(y) = \sum_{t=0}^{T} \frac{1}{2} \log |\Sigma_t| + \frac{1}{2} d'_t \Sigma_t^{-1} d_t\]

with:

- \(d_t = y_t - Hx^f_t\)
- \(\Sigma_t = HP^f_t H' + R\)

Accounts for observational noise and inhomogeneity

Spatial-temporal-variable aggregation
Test in the forced Lorenz model

A trajectory in the state space

\[
\begin{align*}
\frac{dx}{dt} &= \sigma(y - x) + \lambda_i \cos \theta_i \\
\frac{dy}{dt} &= \rho x - y - xz + \lambda_i \sin \theta_i \\
\frac{dz}{dt} &= xy - \beta z
\end{align*}
\]
Test in the forced Lorenz model

Model 1 ($\lambda = 40$)

Model 2 ($\lambda = 0$)

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.
The Gaussian assumption is actually an approximation, hence the obtained value of marginal likelihood is also an approximation.

Different assimilation methods are possible, yielding different values:
- EnKF
- IEnKS
- En-4D-Var

Which DA method provides the best estimate?

Comparison to the “true” value (very accurate approximation)
- Gauss Hermite Quadrature,
- Monte Carlo with large sample ($n = 10^6$).
Results

- In case the model is correct:

- In case the model is not correct:

EnKF gives acceptable results.
Test in the forced Lorenz model: parameter estimation

Marginal likelihood maximization

Large dependence to method/assumptions

May still be useful if the same method is used.
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
- Deep Learning
- Causal Inference
- Extreme Value Analysis
- Hidden Markov Models
Estimating the probability of a rare event

- The easiest way: empirical frequency

$\hat{p} = \frac{1}{N} \sum_{i=1}^{N} 1\{x_i > u\}$
Estimating the probability of a rare event

- The easiest way: empirical frequency

\[ \hat{p} = \frac{1}{N} \sum_{i=1}^{N} 1\{x_i > u\} \]

- A very noisy estimate:

\[ RE = \frac{1}{\sqrt{\hat{p}N}} \]

To estimate a return time of ~5000 years with a 10% error, one needs 500,000 years of simulation.
Estimating the probability of a rare event

- The easiest way: empirical frequency

\[ \hat{p} = \frac{1}{N} \sum_{i=1}^{N} 1\{x_i > u\} \]

- What if the threshold is never exceeded in the dataset?

<table>
<thead>
<tr>
<th>run</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14.8464</td>
</tr>
<tr>
<td>2</td>
<td>12.1998</td>
</tr>
<tr>
<td>3</td>
<td>18.4397</td>
</tr>
<tr>
<td>4</td>
<td>14.9055</td>
</tr>
<tr>
<td>5</td>
<td>17.1043</td>
</tr>
<tr>
<td>6</td>
<td>16.3552</td>
</tr>
<tr>
<td>7</td>
<td>14.5039</td>
</tr>
<tr>
<td>8</td>
<td>13.0704</td>
</tr>
<tr>
<td>9</td>
<td>15.2483</td>
</tr>
<tr>
<td>10</td>
<td>12.8561</td>
</tr>
<tr>
<td>11</td>
<td>18.8878</td>
</tr>
<tr>
<td>12</td>
<td>23.8842</td>
</tr>
<tr>
<td>13</td>
<td>13.0359</td>
</tr>
<tr>
<td>14</td>
<td>22.1881</td>
</tr>
<tr>
<td>15</td>
<td>18.5205</td>
</tr>
<tr>
<td>16</td>
<td>18.0352</td>
</tr>
</tbody>
</table>

... to N
PDF and return level plot

Probability 1/100

Probability 1/10,000
Estimation strategy #1: the histogram

\[ \hat{p} = \frac{1}{N} \sum_{i=1}^{N} 1_{\{x_i > u\}} \]

- Most straightforward estimator
Estimation strategy #1 with $N = 10,000$
Estimation strategy #1 with $N = 100$

often hopeless (when $Np < 1$)
Estimation strategy #2: the kernel estimator

\[ \hat{p} = \frac{1}{N} \sum_{i=1}^{N} K \left( \frac{x_i - u}{h} \right) \]

- K is the so-called kernel:

\[ K(x) = \int_{x}^{\infty} N(z) \, dz \]

- It basically smoothes the estimation.
Estimation strategy #2 with N = 100
Estimation strategy #2 with N = 100
Estimation strategy #3 : the GEV estimator

\[ \hat{p} = 1 - \exp \left( - \left( 1 + \hat{\xi} \frac{u - \hat{\mu}}{\hat{\sigma}} \right)^{-\frac{1}{\hat{\xi}}} \right) \]

- The above function is the so-called GEV distribution.
  - A parametric class of PDF,
  - Three parameters: location, scale, shape,
  - Wide variety of tail behaviours,
  - Good theoretical properties to represent block-maximum values.

- It basically allows for extrapolation when \( N_p > 1 \) and also regularizes for \( N_p \sim 1 \).
Estimation strategy #3 with N = 100
Estimation strategy #3 with N = 100
Streamflow, Rouge River

Streamflow time series

Streamflow histogram
Streamflow, Rouge River

Streamflow time series

Streamflow histogram
Hydropower generation
Hydropower generation

Hydroelectric Capacity (MW)
- 10-49
- 50-99
- 100-199
- 200-499
- 500-999
- 1000+

Total Generation
212.3 TW.h

- Hydro: 95%
- Wind: 4%
- Biomass / Geothermal: 1%
- Petroleum: <1%
- Natural Gas: <0.1%
- Solar: <0.1%
Bell Falls, Rouge River
A **spillway** is a structure used to release the surplus of flow from a dam into a downstream area.
Bell Falls, April 2019
Fears of failure of Chute-Bell dam prompt evacuations in Quebec

04/26/2019

By Elizabeth Ingram
Content Director

The government of Quebec has issued a risk of dam failure alert related to the Rouge River downstream of Chute-Bell dam.

The government has directed people in the affected area to evacuate immediately, effective yesterday afternoon.

The dam impounds water for a 10-MW hydroelectric powerhouse, and water has been overtopping it due to a high flow rate in the river. The run-of-river Chute-Bell facility contains two turbine-generator units and was commissioned in 1915.
### DAM SAFETY

21. Subject to sections 21.1, 22 and 24, every dam must be able to withstand any of the following safety check floods, taking into account the highest dam failure consequence category in flood conditions:

<table>
<thead>
<tr>
<th>Highest dam failure consequence category</th>
<th>Safety Check</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flood in flood conditions</td>
<td></td>
</tr>
<tr>
<td>Very low or low</td>
<td>Centennial*</td>
</tr>
<tr>
<td></td>
<td>(1: 100 years)</td>
</tr>
<tr>
<td>Moderate or high</td>
<td>Millennial*</td>
</tr>
<tr>
<td></td>
<td>(1: 1,000 years)</td>
</tr>
<tr>
<td>Very high</td>
<td>Decamillennial*</td>
</tr>
<tr>
<td></td>
<td>(1: 10,000 years)</td>
</tr>
<tr>
<td>Severe</td>
<td>Probable maximum flood</td>
</tr>
</tbody>
</table>

* Safety check floods expressed according to their recurrence interval.

O.C. 300-2002, s. 21; O.C. 901-2014, s. 2.
Question asked by hydropower companies and regulating bodies

What is the value of the decamillenial flow?

- Civil engineers designing the spillway will build one spillway. The answer is requested to be a single value.
Problem

Estimate a high quantile for a given low probability

Estimate a low probability for a given high threshold
Data: annual maxima
Model: univariate Generalized Extreme Value distribution

\[ \mathbf{x} = (x_1, x_2, \ldots, x_n) \text{ i.i.d.} \]

\[ p(\mathbf{x} | \theta) = \prod_{t=1}^{n} \text{GEV}(x_t | \theta) \]

\[ \theta = (\mu, \sigma, \xi) \]
Inference: maximum likelihood
Can we come up with a better estimator?
Bayesian estimation: a brief overview

\[ \mathbf{x} = (x_1, x_2, \ldots, x_n) \]

\[ p(\mathbf{x} \mid \theta) = \ldots \]

\[ \pi(\theta) = \ldots \]

\[ p(\theta \mid \mathbf{x}) \propto p(\mathbf{x} \mid \theta) \pi(\theta) \]
Bayesian estimation: a brief overview

\[ \mathbf{x} = (x_1, x_2, \ldots, x_n) \]

\[ p(\mathbf{x} \mid \theta) = \ldots \]

\[ \pi(\theta) = \ldots \]

\[ p(\theta \mid \mathbf{x}) \propto p(\mathbf{x} \mid \theta) \pi(\theta) \]

\[ (\theta_1, \theta_2, \ldots, \theta_N) \quad \text{MCMC simulations} \]
Bayesian estimation: a brief overview

\[ C(\theta, \theta^*) = \ldots \]

\[ C(\theta^* \mid x) = \mathbb{E}_{\theta \mid x}(C(\theta, \theta^*)) = \int_{\theta} C(\theta, \theta^*) p(\theta \mid x) \, d\theta \]

\[ \hat{\theta} = \text{argmin}_{\theta^*} C(\theta^* \mid x) \]
Bayesian estimation: a brief overview

\[ C(\theta, \theta^*) = \{\theta - \theta^*\}^2 \]

\[ \hat{\theta}_{\text{MMSE}} = \int_\theta \theta p(\theta | x) \, d\theta \sim \frac{1}{N} \sum_{i=1}^{N} \theta_i \]
Bayesian estimation: a brief overview

\[ C(\theta, \theta^*) = 1 \{ \theta \neq \theta^* \} \]
\[ \hat{\theta}_{\text{MAP}} = \arg\max_\theta p(\theta | x) \simeq \text{mode}(\theta_1, \ldots, \theta_N) \]
Model: univariate Generalized Extreme Value distribution

\[ x = (x_1, x_2, ..., x_n) \text{ i.i.d.} \]

\[ p(x \mid \theta) = \prod_{t=1}^{n} \text{GEV}(x_t \mid \theta) \]

\[ \theta = (\mu, \sigma, \xi) \]

\[ \pi(\theta) \propto \sigma^{-1} \]
Model: univariate Generalized Extreme Value distribution

\[ x = (x_1, x_2, \ldots, x_n) \text{ i.i.d.} \]

\[ p(x \mid \theta) = \prod_{t=1}^{n} \text{GEV}(x_t \mid \theta) \]

\[ \theta = (\mu, \sigma, \xi) \]

\[ \pi(\theta) \propto \sigma^{-1} \]

Northrop and Attalides 2015
MCMC simulation of the posterior distribution
Attempt 1: conventional cost function and estimator

\[ C(\theta, \theta^*) = (\theta - \theta^*)^2 \]

\[ \hat{\theta}_{\text{MMSE}} = \int_\theta \theta p(\theta | x) d\theta \approx \frac{1}{N} \sum_{i=1}^{N} \theta_i \]
Attempt 2: conventional cost function and estimator

\[ C(\theta, \theta^*) = 1 \{\theta \neq \theta^*\} \]

\[ \hat{\theta}_{\text{MAP}} = \arg\max_{\theta} p(\theta \mid x) \simeq \text{mode}(\theta_1, \ldots, \theta_N) \]
Attempt 3: new cost function and estimator

\[ p = 10^{-4} \]
\[ C(\theta, \theta^*) = \left\{ F^{-1}(p \mid \theta) - F^{-1}(p \mid \theta^*) \right\}^2 \]

\( \hat{\theta} \) verifies:
\[ F^{-1}(p \mid \hat{\theta}) = \int_{\theta} F^{-1}(p \mid \theta) p(\theta \mid x) \, d\theta \approx \frac{1}{N} \sum_{i=1}^{N} F^{-1}(p \mid \theta_i) \]

E.g.:
\[ \hat{F}^{-1}(p \mid x) = \int_{\theta} F^{-1}(p \mid \theta) p(\theta \mid x) \, d\theta \approx \frac{1}{N} \sum_{i=1}^{N} F^{-1}(p \mid \theta_i) \]
Attempt 4: new cost function and estimator

\[ p = 10^{-4} \]

\[ C(\theta, \theta^*) = 1 \{ F^{-1}(p \mid \theta) \neq F^{-1}(p \mid \theta^*) \} \]

\[ \hat{\theta} = \arg\max_{\theta} p \left( F^{-1}(p \mid \theta) \mid x \right) \]

e.g:

\[ \hat{F}^{-1}(p \mid x) \simeq \text{mode} \left( F^{-1}(p \mid \theta_1), \ldots, F^{-1}(p \mid \theta_N) \right) \]
Attempt 5: new cost function and estimator

\[ u = \text{fixed threshold (e.g. 2019 record value)} \]

\[ C(\theta, \theta^*) = \left\{ F(u \mid \theta) - F(u \mid \theta^*) \right\}^2 \]

\[ \hat{F}(u \mid \mathbf{x}) = \int_{\theta} F(u \mid \theta) p(\theta \mid \mathbf{x}) \, d\theta \sim \frac{1}{N} \sum_{i=1}^{N} F(u \mid \theta_i) \]
Attempt 6: new cost function and estimator

\[ u = \text{fixed threshold (e.g. 2019 record value)} \]
\[ C(\theta, \theta^*) = 1 \{ F(u | \theta) \neq F(u | \theta^*) \} \]

\[ \hat{F}(u | \bm{x}) \asymp \text{mode} (F(u | \theta_1), \ldots, F(u | \theta_N)) \]
Illustration on Rouge River
Illustration on Rouge River

- Estimator #1
- Estimator #2

Graph showing the cost of infrastructure as a function of flood level (m3s-1).
Properties and performance of estimators

\[ \mathbb{E}_x \left( \hat{F}^{-1}(p \mid x) \right) = \int_x \hat{F}^{-1}(p \mid x) p(x \mid \theta) \, dx = F^{-1}(p \mid \theta) \quad ? \]

\[ \mathbf{V}_x \left( \hat{F}^{-1}(p \mid x) \right) \to 0 \text{ as } n \to \infty \quad ? \]

\[ \text{MSE} = \mathbb{E}_x \left( \left\{ \hat{F}^{-1}(p \mid x) - F^{-1}(p \mid \theta) \right\}^2 \right) \text{ minimal?} \]
Simulation testbed results for $p = 10^{-2}$ and $x_i < 0$
Simulation testbed results for $p = 10^{-4}$ and $x_i < 0$
Simulation testbed results for $p = 10^{-2}$ and $x_i > 0$
Simulation testbed results for $p = 10^{-4}$ and $x_i > 0$
Even for a single parametric model, several different point estimators of high quantiles and low probabilities can be proposed.

Within a Bayesian approach, such estimators can be obtained by choosing alternative cost functions that are ad-hoc to the problem.

The conventional estimator of a high quantile (inverse CDF evaluated at $p$ with MLE of theta) is not necessarily the best solution. Neither is the intuitive solution of the « posterior predictive ».

Instead, the « MAP quantile » estimator appears to consistently perform best, based on simulation results.

More simulation and theoretical grounding for these estimators is needed.
Outline

- General considerations
- Statistical Methods & Illustrations
- Conclusion
Thank you
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

7th International Workshop on Climate Informatics
September 20-22, 2017

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-occurrence, and, instead, enable increased...
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

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.
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.