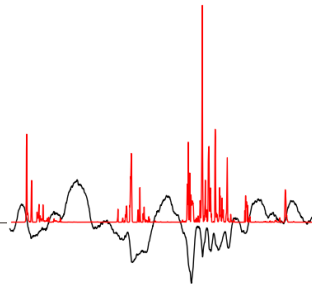


MACHINE LEARNING FOR ENERGY AND CLIMATE

< Introduction >

Bruno Deremble & Alexis Tantet

Fall 2022



PRACTICAL INFORMATION

Schedule

- ▶ Classes are Monday morning 8h30-12h45
- ▶ Coding test + seminar: November 14th
- ▶ Project presentation: December 12th



Theory, tutorial and project



For the first classes we will have

- ▶ 2×45 min to explain concepts and derive important results
- ▶ 2h tutorial session

Towards the end: more time spent on tutorial and less on theory.
At the end: full time on projects.



CONTENT OF THE CLASS

One objective of the class is to derive key results of statistical learning. We will use low dimensional systems to illustrate new concepts.

Prerequisite

- ▶ Elements of probability theory to understand uncertainty
- ▶ Linear algebra notation to handle multi-dimensional systems
- ▶ Calculus to do the actual derivation

At the beginning of each class: Wooclap quiz. ~ 5 multiple choice questions on the previous class (should last 5-10 min). This will count for 25% of the final grade.

TUTORIAL

Tutorials are meant to acquire a deep understanding of the formalism and to apply to real data sets (related to climate and to the environment).

- ▶ Follow the literate programming and literate computing wisdom
- ▶ **Develop methods from scratch** (e.g. neural networks) to understand every detail of the algorithms
- ▶ Existing library: Scikit-learn
- ▶ if the first questions are too easy for you: spend more time on the *Optional questions*



Prerequisite

- ▶ Basic coding skills in python

PROJECT

This is a project based class.

During the class, you will manipulate several data sets related to climate and energy. We will also provide a collection of scientific articles that deploy machine learning algorithms to tackle new problems. We will ask you to

- ▶ Form groups of 2 students
- ▶ **November 7th**: project discussion (5 min presentation + questions): present the problematic, the data set, the method and some references.
- ▶ **December 12th**: Final presentation + Notebook
- ▶ Final grade: 25% Participation, 25% coding test, 50% project presentation

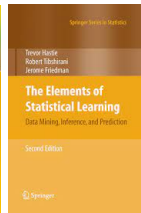
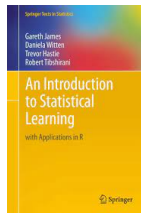
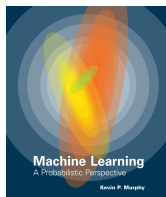
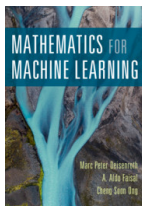


SYLLABUS

1. Introduction
2. Bias-variance, regularization and validation
3. Classification
4. Unsupervised learning
5. Ensemble methods for prediction and regularization
6. Neural network (I)
7. Neural network (II)
8. Project
9. Project

- PC1. Python, linear regression, sci-kit learn
- PC2. Regularization and choice of parameters
- PC3. Tree methods
- PC4. PCA, k-means, data choice
- PC5. Bootstrap
- PC6. Project discussion
- PC7. Neural network
- PC8. Coding test
- PC9. Project

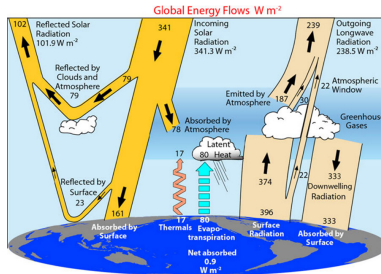
REFERENCES



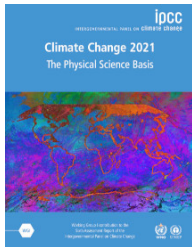
- ▶ M. P. Deisenroth, A. A. Faisal, and C. S. Ong. *Mathematics for machine learning*. Cambridge University Press, 2020. URL: <http://mml-book.github.io>
- ▶ K. P. Murphy. *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012. URL: probml.ai
- ▶ G. James et al. *An introduction to statistical learning*. Springer, 2013. URL: <http://statlearning.com>
- ▶ T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning*. 2009. URL: <http://web.stanford.edu/~hastie/ElemStatLearn>

CLIMATE & ENERGY

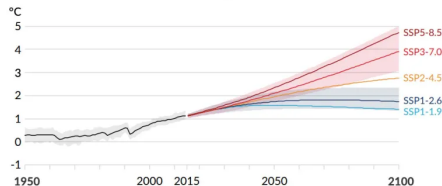
CLIMATE PREDICTION



Earth radiative budget, Credit: Kevin Trenberth et. al UCAR



a) Global surface temperature change relative to 1850-1900



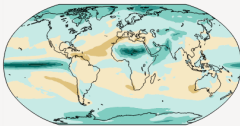
Source: IPCC, AR6

FROM GLOBAL TO REGIONAL PREDICTION

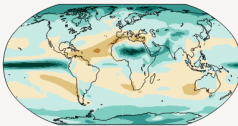
c) Annual mean precipitation change (%) relative to 1850-1900

Precipitation is projected to increase over high latitudes, the equatorial Pacific and parts of the monsoon regions, but decrease over parts of the subtropics and in limited areas of the tropics.

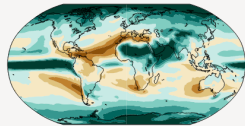
Simulated change at 1.5 °C global warming



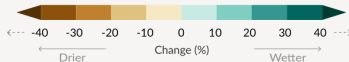
Simulated change at 2 °C global warming



Simulated change at 4 °C global warming



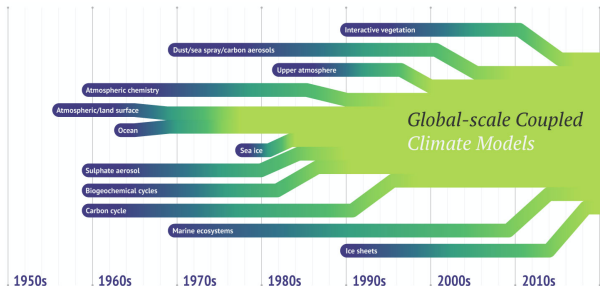
Relatively small absolute changes may appear as large % changes in regions with dry baseline conditions



Source: IPCC, AR6

- ▶ Although the global temperature trend is relatively easy to predict (based on simple radiative calculation), the regional impacts are much more complicated to anticipate.
- ▶ The IPCC uses models for most sub-components of the earth. The interaction between all components is key for a good representation of the climate.

INSIDE A CLIMATE MODEL



Climate models are increasingly complex and the available computing resources may not allow us to keep on running these tools. The scientific community is now exploring new avenues to use the massive amount of data that has already been produced and propose new ways to do reliable forecasts.

WEATHER FORECASTING

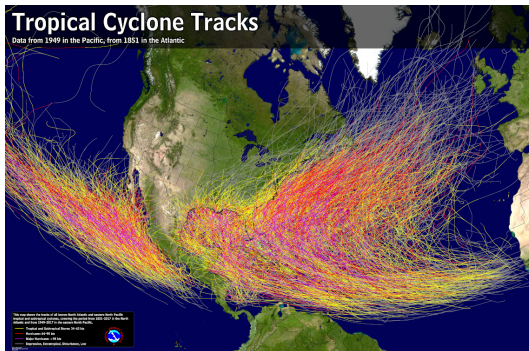
Long term weather forecast is limited due to

- ▶ Our limited knowledge of the initial conditions (observations)
- ▶ We use approximations in the equation of evolution
- ▶ We don't have enough computing power
- ▶ The weather is a chaotic system

Machine learning applications

- ▶ Data-driven approaches directly learn from the best available observations and could potentially produce better forecasts. (Rasp et. al 2020)

TROPICAL CYCLONE



Source: NHC

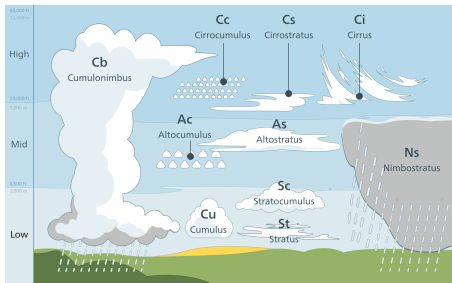
There are on average 10 tropical storms per year in the Atlantic It is easy to predict their track when they are formed

Machine learning applications

- Use statistical algorithm to predict the intensity and landfall

CLOUDS

- ▶ Still a big uncertainty for climate (warming and cooling effect)
- ▶ Important for solar energy production



Machine learning applications

- ▶ Microphysics parameterization
- ▶ Shallow and deep convection
- ▶ Identify cloud pattern with satellite images

AIR-SEA INTERACTION



The air-sea interface is a highly turbulent zone and we still use many approximations to describe how the ocean and the atmosphere exchange heat and momentum.

Machine learning applications

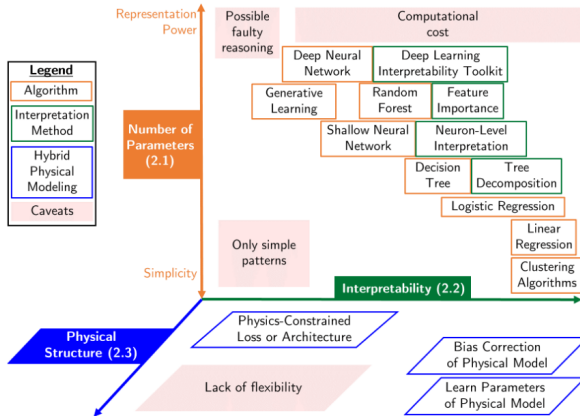
- ▶ Crunching data from observations or super-high resolution models could help us uncover new laws of physics.

REDUCE OUR ENERGY FOOTPRINT

- ▶ Use machine learning to detect trends
- ▶ Energy efficient buildings

PHYSICS-DRIVEN APPROACH

- Use the laws of physics to guide the construction of your model



Source: Beucler et. al 2021

SUMMARY

Machine learning is a tool that can help us reduce the complexity of a system by detecting recurring patterns. We can use it to

- ▶ Better understand the law of evolution of the system
- ▶ Forecast the future evolution
- ▶ Propose an alternative to traditional methods (cost reduction)

MACHINE LEARNING