

# Machine Learning à l'IPSL

J. Brajard, L. Barthes, O. Bouchet, R. Hourani, C. Mallet, E. de Bezenac, O. Schwander, S. Thiria

LOCEAN-SU

13 Mars 2018

# What is machine learning?

## A machine

$$y = \mathcal{M}(x, \theta)$$

- $x$  : input
- $y$  : output
- $\mathcal{M}$  : a model (named "machine")
- $\theta$  : parameters of the model  $\mathcal{M}$ .

**Machine learning** consists in optimizing  $\theta$  using a set of data. This is the training process.

# Example 1 : Computer Vision

**Classification**



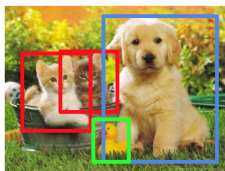
CAT

**Classification  
+ Localization**



CAT

**Object Detection**



CAT, DOG, DUCK

**Instance  
Segmentation**



CAT, DOG, DUCK

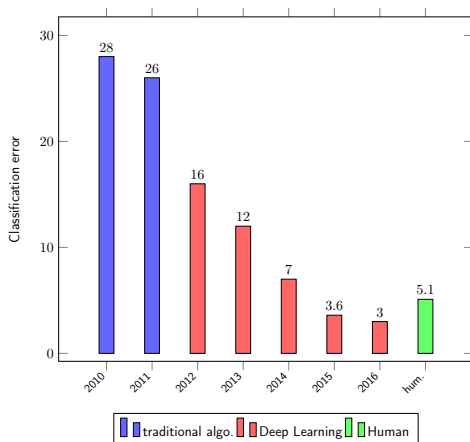
Single object

Multiple objects

*Li, Karpathy and Johnson, 2016, Stanford CS231n course*

# Example 1 : Computer Vision

## Image Net Large Scale Recognition Challenge (ILSVRC)

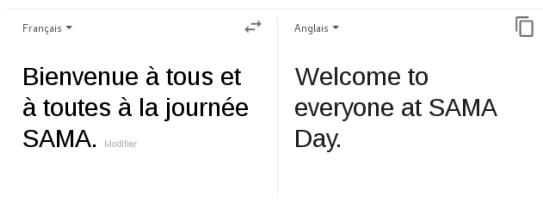


Deep learning architectures was based on Convolutional Neural Networks (CNN).



# Machine Translation

Objective : translate a text from a language to another.



- Oct. 2013 : Pioneering scientific paper about neural networks in machine translation (Kalchbrenner, N., and Blunsom, P).
- 2016 : Neural machine translation outperform traditional approaches on public benchmarks
- 2017 : Major systems (Google, Systran, WIPO, Microsoft) switch to neural machine translation (using deep recurrent neural networks)

# Playing Games

- 1997 : Deep Blue defeats Kasparov at Chess.
- 2016 : AlphaGo's victory again Lee Sedol at Go.
- 2017 : AlphaGo Zero learns how to play Go only by playing against itself. It outperformed previous AlphaGo version (Reinforcement learning)
- 2017 : DeepStack beats professional human poker players.





CHRIS KONDAL/BALL

Many of the latest climate models seek to increase the detail in simulations of cloud structure.

ARTIFICIAL INTELLIGENCE

## Machine learning tapped to improve climate forecasts

*The approach helps to identify atmospheric processes and rank climate models by quality.*

*Jones, Nature 548, 379–380 (24 August 2017)*

# Some problems to be addressed by machine learning ?

A non-exhaustive list :

- Forecasting
- Nowcasting
- Model aggregation
- Feature identification
- Parameter estimation
- Quality control
- ...

## Contexte institutionnel à l'IPSL (et au-delà)

- Projet européen de recherche et d'innovation (RIA) sur l'Intelligence Artificiel.

Importance des enjeux sociétaux et économiques, et notamment construction d'un volet "Climat"

<http://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/ict-26-2018-2020.html>

## Contexte institutionnel à l'IPSL (et au-delà)

- Projet européen de recherche et d'innovation (RIA) sur l'Intelligence Artificiel.

Importance des enjeux sociétaux et économiques, et notamment construction d'un volet "Climat"

<http://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/ict-26-2018-2020.html>

- Projets "Make ou planet great again", lauréat Pr V. Balaji (Automn 2018).

Utiliser des simulations hautes résolutions afin de mieux représenter l'effet des nuages dans les modèles de climat.

## Contexte institutionnel à l'IPSL (et au-delà)

- Projet européen de recherche et d'innovation (RIA) sur l'Intelligence Artificiel.

Importance des enjeux sociétaux et économiques, et notamment construction d'un volet "Climat"

<http://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/ict-26-2018-2020.html>

- Projets "Make ou planet great again", lauréat Pr V. Balaji (Automn 2018).

Utiliser des simulations hautes résolutions afin de mieux représenter l'effet des nuages dans les modèles de climat.

- Projet "Extreme earth" (porté par ECMWF)

Représentation et prédiction d'évènements extrêmes (avec une composante Big Data, Deep learning)

[www.extremearth.eu](http://www.extremearth.eu)

## Contexte institutionnel à l'IPSL (et au-delà)

- Projet européen de recherche et d'innovation (RIA) sur l'Intelligence Artificiel.

Importance des enjeux sociétaux et économiques, et notamment construction d'un volet "Climat"

<http://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/ict-26-2018-2020.html>

- Projets "Make ou planet great again", lauréat Pr V. Balaji (Automn 2018).

Utiliser des simulations hautes résolutions afin de mieux représenter l'effet des nuages dans les modèles de climat.

- Projet "Extreme earth" (porté par ECMWF)

Représentation et prédiction d'évènements extrêmes (avec une composante Big Data, Deep learning)

[www.extremearth.eu](http://www.extremearth.eu)

- Plus local : groupe de travail "SAMA-Machine Learning"



# La formation en Machine Learning à l'IPSL

## Master TRIED depuis 1998 (28 inscrits en 2017)

- Exploration de données multidimensionnelles
- Apprentissage statistique
- Réseaux de neurones
- Reconnaissance des Formes
- Projets d'initiation à la recherche en science des données à L'IPSL

# La formation en Machine Learning à l'IPSL

## Master TRIED depuis 1998 (28 inscrits en 2017)

- Exploration de données multidimensionnelles
- Apprentissage statistique
- Réseaux de neurones
- Reconnaissance des Formes
- Projets d'initiation à la recherche en science des données à L'IPSL

université  
PARIS-SACLAY

## Portail e-learning de l'IPSL

<https://claroline.locean-ipsl.upmc.fr>

- cours Perceptron multicouche, cartes auto-organisatrices
- Travaux pratiques en science de l'environnement
- Études de cas en télédétection
- Notebook Ipython
- Module Apprentissage Statistique Ecole Doctorale de l'IPSL

# Le groupe de travail SAMA-Machine learning

- Existe depuis Octobre 2017
- Réunit des chercheurs, ingénieurs de l'IPSL, du LIP6 et de l'ENSIEE.



- Réunion une fois par mois
- Un (ou deux) présentations scientifiques par réunion
- Définition collectivement de sujets de recherche, de stages, ...

# Some problems to be addressed by machine learning ?

A non-exhaustive list :

- Forecasting
- Nowcasting
- Model aggregation
- Feature identification
- Parameter estimation
- Quality control

# Some problems to be addressed by machine learning ?

A non-exhaustive list :

- Forecasting
  - ▶ Sea surface temperature advection
- Nowcasting
  - ▶ Rain cells nowcasting from radar data
  - ▶ Total suspended matter (sea surface satellite images)
- Model aggregation
  - ▶ CMIP5 model ranking on senegal upwelling representation.
- Feature identification
  - ▶ Eddy detection over Mediterranean Sea.
  - ▶ Data-driven clustering of rain events
- Parameter estimation
  - ▶ Completion (inpainting) of satellite data under clouds.
  - ▶ Parametrization of sub-scale processes in numerical models.
  - ▶ Phytoplankton type estimation from satellite images.
- Quality control

# Some problems to be addressed by machine learning ?

A non-exhaustive list :

- Forecasting
  - ▶ Sea surface temperature advection
- Nowcasting
  - ▶ Rain cells nowcasting from radar data
  - ▶ Total suspended matter (sea surface satellite images)
- Model aggregation
  - ▶ CMIP5 model ranking on senegal upwelling representation.
- Feature identification
  - ▶ Eddy detection over Mediterranean Sea.
  - ▶ Data-driven clustering of rain events
- Parameter estimation
  - ▶ Completion (inpainting) of satellite data under clouds.
  - ▶ Parametrization of sub-scale processes in numerical models.
  - ▶ Phytoplankton type estimation from satellite images.
- Quality control

# Some problems to be addressed by machine learning ?

A non-exhaustive list :

- Forecasting
  - ▶ Sea surface temperature advection
- Nowcasting
  - ▶ Rain cells nowcasting from radar data
  - ▶ Total suspended matter (sea surface satellite images)
- Model aggregation
  - ▶ CMIP5 model ranking on senegal upwelling representation.
- Feature identification
  - ▶ Eddy detection over Mediterranean Sea.
  - ▶ Data-driven clustering of rain events
- Parameter estimation
  - ▶ Completion (inpainting) of satellite data under clouds.
  - ▶ Parametrization of sub-scale processes in numerical models.
  - ▶ Phytoplankton type estimation from satellite images.
- Quality control

# Some problems to be addressed by machine learning ?

A non-exhaustive list :

- Forecasting
  - ▶ Sea surface temperature advection
- Nowcasting
  - ▶ Rain cells nowcasting from radar data
  - ▶ Total suspended matter (sea surface satellite images)
- Model aggregation
  - ▶ CMIP5 model ranking on senegal upwelling representation.
- Feature identification
  - ▶ Eddy detection over Mediterranean Sea.
  - ▶ Data-driven clustering of rain events
- Parameter estimation
  - ▶ Completion (inpainting) of satellite data under clouds.
  - ▶ **Parametrization of sub-scale processes in numerical models.**
  - ▶ Phytoplankton type estimation from satellite images.
- Quality control



# Some problems to be addressed by machine learning ?

A non-exhaustive list :

- Forecasting
  - ▶ Sea surface temperature advection
- Nowcasting
  - ▶ Rain cells nowcasting from radar data
  - ▶ Total suspended matter (sea surface satellite images)
- Model aggregation
  - ▶ CMIP5 model ranking on senegal upwelling representation.
- Feature identification
  - ▶ Eddy detection over Mediterranean Sea.
  - ▶ Data-driven clustering of rain events
- Parameter estimation
  - ▶ Completion (inpainting) of satellite data under clouds.
  - ▶ Parametrization of sub-scale processes in numerical models.
  - ▶ Phytoplankton type estimation from satellite images.
- Quality control

# Data-driven clustering of rain events

**Objective** : Selection of an optimal number of indicators to classify rain events and infer non-observable parameters

**Model** : genetic algorithm, self-organizing map.

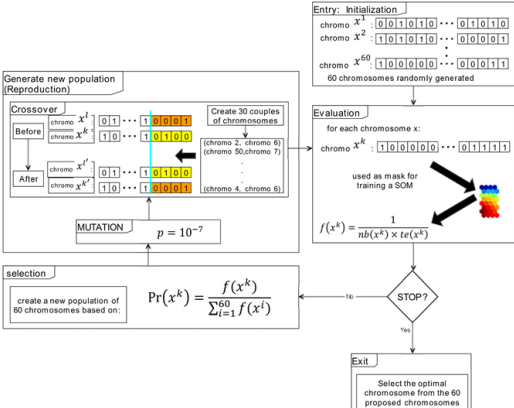
*Dilmi et al. 2017*

# Data-driven clustering of rain events

**Objective :** Selection of an optimal number of indicators to classify rain events and infer non-observable parameters

**Model :** genetic algorithm, self-organizing map.

- 1 Select S indicators among N candidates for a optimal clustering of rain events (N=23)



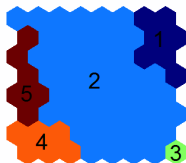
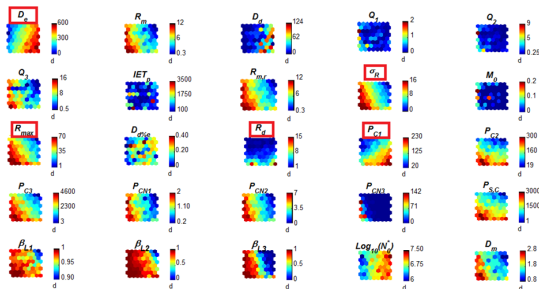
# Data-driven clustering of rain events

**Objective** : Selection of an optimal number of indicators to classify rain events and infer non-observable parameters

**Model** : genetic algorithm, self-organizing map.

1 Select  $S$  indicators among  $N$  candidates for a optimal clustering of rain events ( $N=23$ )

2 Result : coherent clustering of rain events using a small number of parameters ( $S=5$ )

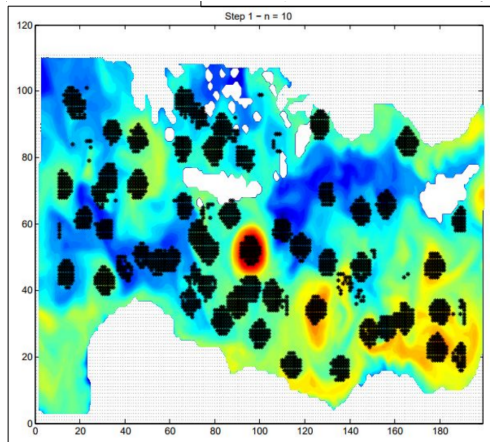


*Dilmi et al. 2017*

# Eddy detection

**Objective** : Eddy detection from satellite image

**Model** : Deep convolutive network



*stage Luiz André Schiaveto Neto*

# Detection of phytoplankton groups from satellite images

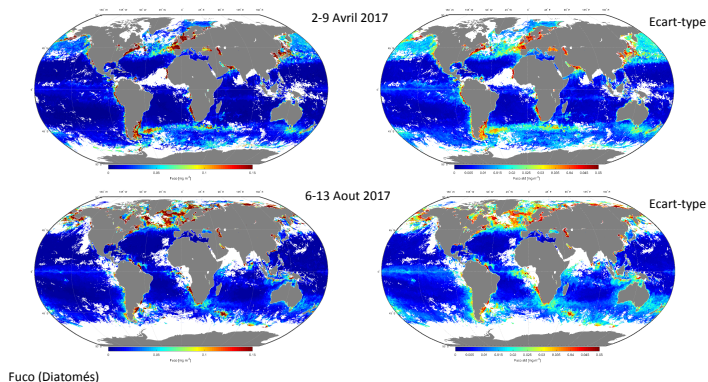
**Objective** : Estimate some pigments of phytoplankton from satellite ocean color images

**Model** : Self-organizing map with incomplete data

# Detection of phytoplankton groups from satellite images

**Objective** : Estimate some pigments of phytoplankton from satellite ocean color images

**Model** : Self-organizing map with incomplete data



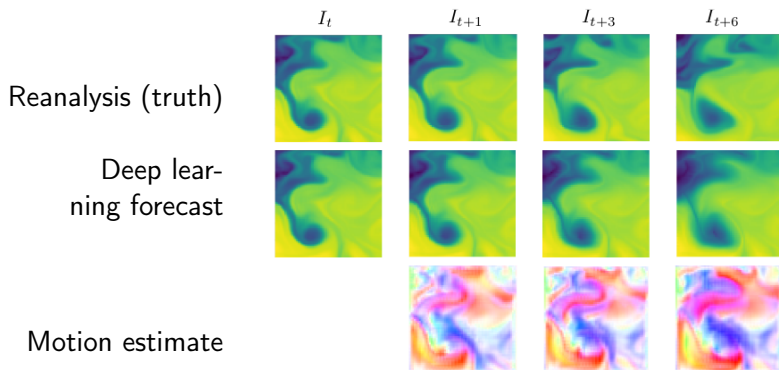
Training set : 5500 in-situ data (MAREDAT global database)

*Hourani et al.*

# Sea Surface Temperature forecast (SST)

**Objective** : predict daily SST images from past SST images.

**Model** : Combination of a convolutive neural network model for motion estimate and a warping scheme to advect the input image.

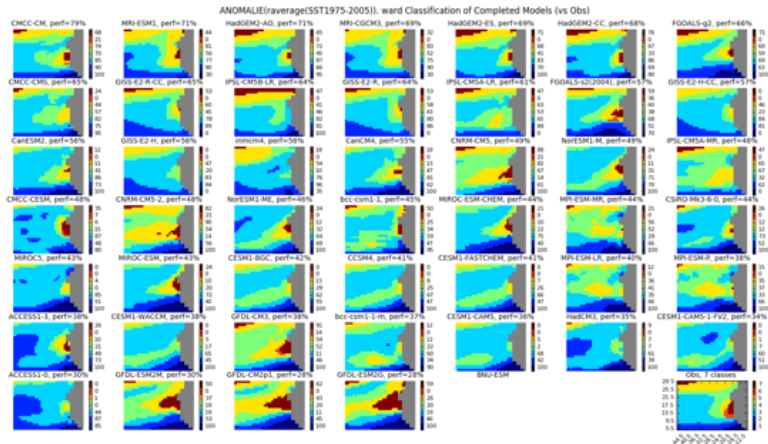




# Evaluate CMIP5 models about upwelling representation

**Objective :** Compare, evaluate (and combine) climate model outputs.

**Model :** Clustering using Self-Organizing-Maps.

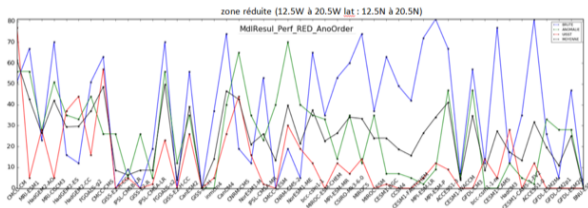
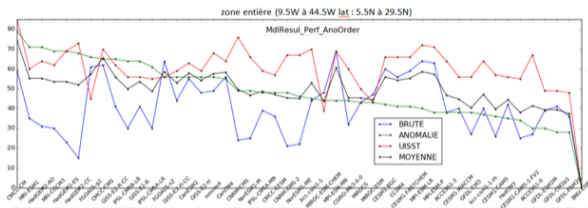


# Evaluate CMIP5 models about upwelling representation

**Objective :** Compare, evaluate (and combine) climate model outputs.

**Model :** Clustering using Self-Organizing-Maps.

Performances globales des Modèles triés sur la performance globale en Anomalie sur la zone entière



# Les besoins pour le futur

- Renforcer les liens entre formations en science de l'environnement et les formations en Machine Learning
  - ▶ Stages pluri-disciplinaires

# Les besoins pour le futur

- Renforcer les liens entre formations en science de l'environnement et les formations en Machine Learning
  - ▶ Stages pluri-disciplinaires
- Mieux cibler les grandes questions scientifiques liées à l'apprentissage automatique.

# Les besoins pour le futur

- Renforcer les liens entre formations en science de l'environnement et les formations en Machine Learning
  - ▶ Stages pluri-disciplinaires
- Mieux cibler les grandes questions scientifiques liées à l'apprentissage automatique.
- Améliorer la visibilité d'apprentissage automatique à l'IPSL.

# Les besoins pour le futur

- Renforcer les liens entre formations en science de l'environnement et les formations en Machine Learning
  - ▶ Stages pluri-disciplinaires
- Mieux cibler les grandes questions scientifiques liées à l'apprentissage automatique.
- Améliorer la visibilité d'apprentissage automatique à l'IPSL.
- Bénéficier d'une infrastructure technique permettant la mise aux point des algorithmes d'apprentissage (apprentissage profond, ...).

# Les besoins pour le futur

- Renforcer les liens entre formations en science de l'environnement et les formations en Machine Learning
  - ▶ Stages pluri-disciplinaires
- Mieux cibler les grandes questions scientifiques liées à l'apprentissage automatique.
- Améliorer la visibilité d'apprentissage automatique à l'IPSL.
- Bénéficier d'une infrastructure technique permettant la mise aux point des algorithmes d'apprentissage (apprentissage profond, ...).
- Structurer la recherche en apprentissage automatique sur les thèmes du climat et des sciences de l'environnement en lien avec d'autres équipes (nationales et internationales).
  - ▶ AMS : Committee on Artificial Intelligence Applications to Environmental Science
  - ▶ ...