

## 11èmes Journées Informatique IN2P3/IRFU

### InTheArt : intelligence artificielle

Valérie Gautard (IRFU/DEDIP) et Edith Le Floch (DRF/IBFJ/CNRGH)

## Contexte :

### **Au CNRGH :** étude du lien entre le génome et les maladies

- Approche classique : association statistique entre chaque variation *unique* du génome et la maladie
- Génome humain : 3 milliards de nucléotides, 3 milliards de variations potentielles + **facteurs épigénétiques**
- Aujourd'hui : résultats peu corrélés avec l'observation de la transmission familiale de certaines maladies

### **A l'IRFU :** expertise en Machine Learning dans des collaborations internationales (CERN, USA, etc.)

- DAp (astrophysique) et DEDIP (détecteurs et info) : reconnaissance d'images
- DPhP (phys. des particules) : algorithmes pour l'extraction de signaux faibles dans des environnements très bruités

Idée: combiner les effets de - plusieurs - variations génétiques et de plusieurs types de données génomiques grâce aux techniques de Machine Learning afin de mieux comprendre certaines maladies.

## Objectifs :

➔ **Organisation d'une série de séminaires** qui ont pour objectif d'aider à la compréhension des différents problèmes « théoriques », et permettre l'échange et la définition d'un langage commun dans le but de les résoudre.

➔ **Constitution d'un groupe (InTheArt)**

➔ **Projet Transverse à la DRF au CEA**, en s'appuyant sur notre expérience et celles des experts des domaines de nos instituts, IRFU, CNRGH et au delà LSCE, IPhT, Neurospin, DEN, DRT et pers. de univ. Paris Sud

**Edith Le Floch – CNRGH – Nov. 2017**

Can we predict the phenotype of an individual from DNA ?

**Özgur Sahin – DPhP – Jan. 2018**

Machine learning techniques at the LHC experiment

**Davide Feranda – LSCE – Avr. 2018**

New dynamical systems tools to study atmospheric flows

**David Rousseau – LAL – Mai 2018**

Advances in Machine learning in High Energy Physics


**Etienne Thevenot – LIST – Juil. 2018**

Omics data processing and analysis for high-throughput phenotyping

**Michalis Vazirgiannis, DASCIM, Ecole Polytechnique – Sep. 2018**

Graph Degeneracy for social nets and text mining

## Thèmes :

-  Biologie, app. médical
-  Machine Learning, informatique
-  Mathématique
-  Autres applications
-  Chimie
-  Physique

## Un noyau « dur » de ~20 personnes

- ▶ LSCE (labo mixte), CNRGH, Irfu, DEN, DRT
- ▶ + personnes de l'université Paris Sud, etc.
- ▶ Feuille de route :
  - Liste de communautés identifiées
  - Présentations pédagogiques
  - Séminaires
  - Projets concrets

## Un groupe plus élargi ~50 personnes

### Groupe d'Echange sur les Technologies et Applications de l'Intelligence Artificielle à la DRF

#### InTheArt

Proposition de texte, destinée à jeter les fondations d'une organisation, suite au premier CR de réunion et à des discussions qui ont suivi entre [Sjava](#), Dominique, Geoffrey, [Ozgür](#), Valérie, Edith, *liste à compléter....*

#### Liste des communautés identifiées:

Physique des particules  
Génétique  
Astrophysique  
Sciences du Climat  
Autres.

#### Proposition d'organisation pour initier un travail de longue haleine:

- Nous voulons privilégier un travail *concret* avant d'aborder des notions théoriques avancées.

#### Page Web InTheArt.

Pour organiser une mémoire à notre travail, ouvrir et organiser une page Web sur les serveurs internes du CEA, documentant les séminaires, nos échanges et nos codes.

#### Séminaires sur des projets ambitieux sur le thème du Machine Learning ou Intelligence Artificielle.

Pas forcément souvent. ?Deux à quatre? par an?

J'ai un nom qui pourrait vous intéresser. Je garde le secret pour le moment ... :-)

#### Mini-Projets « cadre de communauté »:

Chaque communauté amène à la disposition des participants un mini projet d'étude, base de discussion et illustration de l'usage qu'elle pratique dans sa technologie, Comprenant

- Une mini présentation, motivation, objectif scientifique, technologies aujourd'hui envisagées.
- Un jeu de données sur lequel travailler.

Le format de donnée doit être facilement lisible, ou être accompagné des logiciels open-source permettant l'usage de ces données. Ce format doit être autant que possible consensuel dans la communauté de travail représentée (Exemple, Physique corpusculaire: Fichier ASCII, ou Root)

- Un premier code, mettant en oeuvre sur ces données une technique de machine Learning ou d'intelligence Artificielle.

On en fait une démonstration. On se comprend et on en discute.

#### Présentations pédagogiques:

- On ne parle pas de recherches, mais de cours d'initiation.

- On peut aussi envisager de projeter un cours disponible en ligne (Cours du MIT, autres) et d'ouvrir une discussion après.

- **Stage de 5 mois**

- ▶ LSCE et IRFU, dir. Davide Faranda et Valérie Gautard
- ▶ Techniques de *Machine Learning* pour les systèmes chaotiques
- ▶ Stagiaire : Adnane Hamid

- **Proposition de thèse**

- ▶ IRFU, dir Viatcheslav Sharyy
- ▶ Intelligence artificielle sur un détecteur gamma pour l'imagerie médicale

- **Site internet :**

<https://indico.in2p3.fr/event/17858/page/1967-inheart>

**PROPOSITION DE STAGE**

**MACHINE LEARNING TECHNIQUES FOR CHAOTIC SYSTEMS**

Valerie Gautard, Davide Faranda, Mathieu Vrac, Pascal Yiou

Chaos is an intrinsic component of many complex systems ranging from molecular dynamics to astronomy and atmospheric physics. In 1963 Lorenz [1] discovered that the behavior of a system becomes rapidly unpredictable because of the ignorance of its exact initial state. He proved that this holds true for the weather forecasts whose validity is theoretically limited to about 14 days [2].

However, in a series of recent papers [3,4], it has been shown that machine learning techniques can be used to extend the predictability of chaotic systems well beyond the predictability limit. This opens the possibility of improving weather forecast and the ability to describe the climate system.

The goal of this stage i) to reproduce with a suitable numerical code the results presented in [3] and [4] ii) to understand the limits of those techniques e.g. in reproducing extreme situations iii) to apply the method on climate data, namely the sea-level pressure fields over the North Atlantic for the period 1948-2018, issued from NCEP database, using as training set the first period of the available dataset and as verification the rest part of the data. The candidate will then compare the capability of the

COMMISSARIAT À L'ÉNERGIE ATOMIQUE ET AUX ÉNERGIES ALTERNATIVES (CEA)



**INTELLIGENCE ARTIFICIELLE SUR UN DÉTECTEUR GAMMA POUR L'IMAGERIE TEP HAUTE RÉOLUTION**

**DESCRIPTION**

La Tomographie par Emission de Positrons (TEP) est une technique d'imagerie utilisée largement dans le traitement de cancer et dans les recherches neurobiologiques. Il s'agit de détecter deux photons de 511 keV produits par l'annihilation d'un positon issus d'un traceur biochimique fixé dans les tissus. Elle image ainsi l'activité biologique des organes.

La détermination précise du point d'annihilation

**ARTIFICIAL INTELLIGENCE FOR A GAMMA-DETECTOR FOR HIGH RESOLUTION PET IMAGING**

**DESCRIPTION**

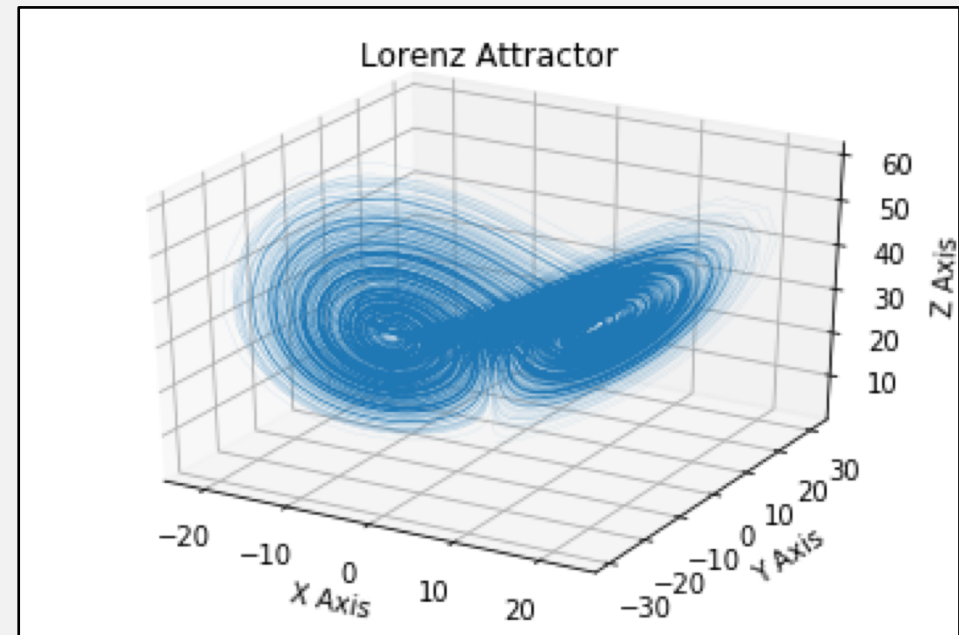
Positron emission tomography (PET) is a nuclear imaging technique widely used in oncology and neurobiological research. Decay of the radioactive tracer emits positrons, which annihilate in the nearby tissue and emits two back-to-back 511 keV photons. These photons are used to reconstruct the biological activity in the body.

The precise determination of the position of the

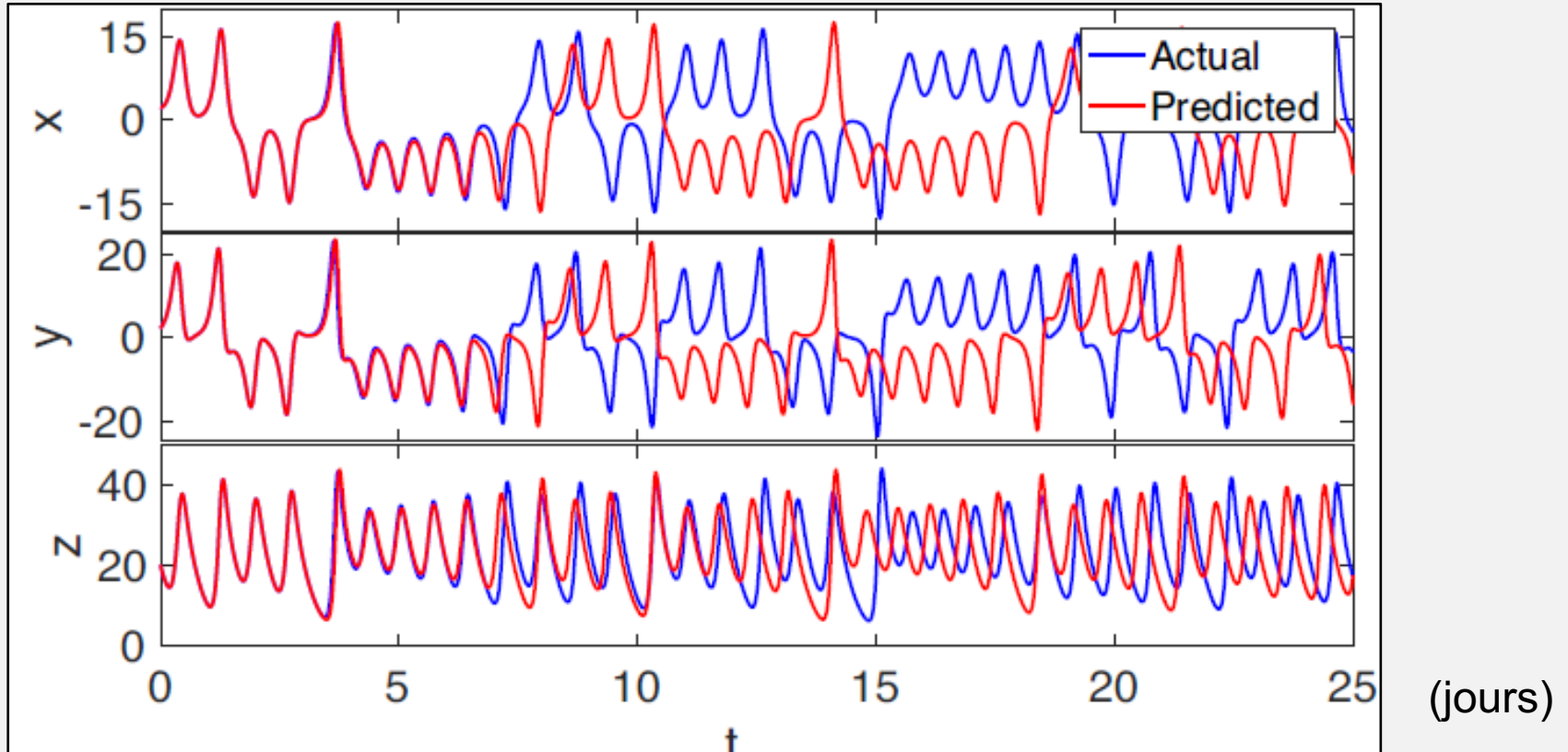
## Prédiction de systèmes chaotiques

Pathak et Al, 2017

- **Objectif : prédire la météo ou le climat**
  - ▶ Beaucoup de données enregistrées mais avec des erreurs de mesure
  - ▶ Forte sensibilité aux conditions initiales
  - ▶ Apport du Machine Learning ?
  - ▶ Test sur des cellules convectives
  
- **Retrouver les résultats de l'article**
  - ▶ Machine Learning
  - ▶ Attracteur de Lorenz : toy model de l'atmosphère



## Les résultats de l'article

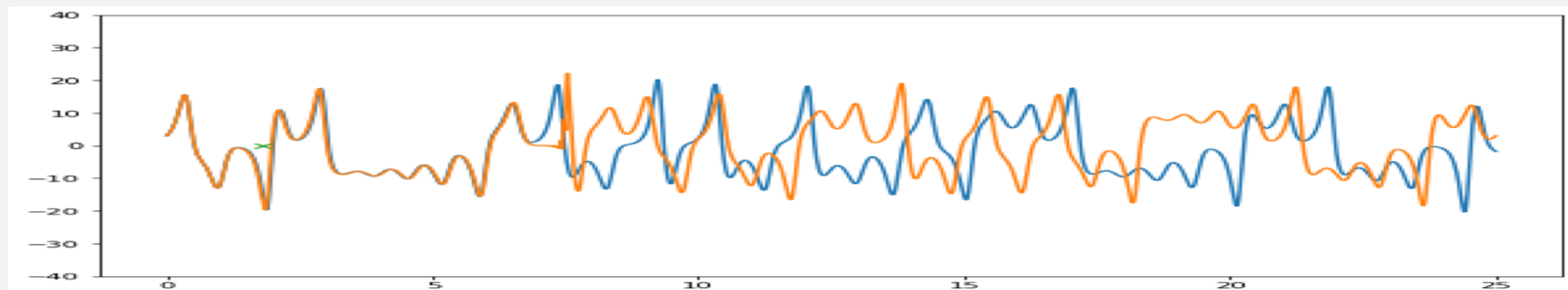


Approximation de la trajectoire de l'attracteur de Lorenz grâce au réservoir computing.

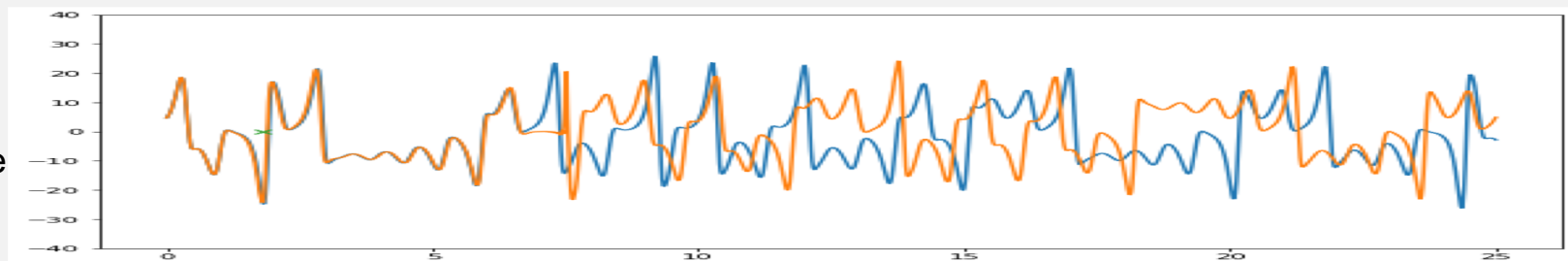
Source: Pathak et al 2017.

## Résultats obtenus :

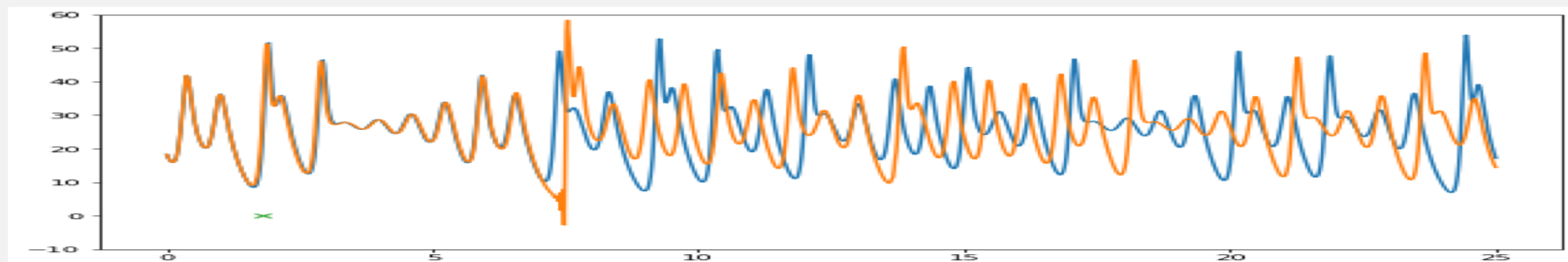
X  
Convection



Y  
Température



Z  
Symétrie



(jours)



- **Prochainement** : formation Deep Learning (*Geoffrey Daniel, Irfu/Dap*)
- **Suivi de Mooc** :
  - ▶ Stanford university :
    - <https://www.coursera.org/learn/machine-learning>
    - <https://www.coursera.org/specializations/deep-learning>
  - ▶ MIT :
    - <https://www.youtube.com/watch?v=uXt8qF2Zzfo>
    - [https://www.youtube.com/watch?v=VrMHA3yX\\_QI](https://www.youtube.com/watch?v=VrMHA3yX_QI)

- Constitution d'un groupe de travail d'une vingtaine de personnes de divers instituts du CEA et quelques personnes du CNRS et de l'université Paris Sud
- N'hésitez à nous contacter ! (*valerie.gautard@cea.fr*)
- Etude de cas concrets
- Développer et renforcer le groupe :
  - Implication de généticiens, mathématiciens, statisticiens, physiciens mais manque de chimistes et de biologistes
  - Nécessité d'intégrer des étudiants