

IMOSE :

un Institut à l'interface des mathématiques

Christophe Chalons, Laurent Dumas

Laboratoire de Mathématiques de Versailles
Université de Versailles

petit déjeuner 'Maths et IA', 29 mars 2019



(Institut pour la Modélisation et l'Optimisation
des Systèmes et des Energies)

- Plateforme hébergée par l'Université de Versailles (UVSQ)
- Fondé en septembre 2015 par C. Chalons et L. Dumas
- Un ingénieur de recherche (post doc) à temps plein depuis Octobre 2016 (F. Blachère puis F. Omnes)
- Site web : www.imose.fr

Principaux objectifs

- Transférer les connaissances et l'approche mathématique dans différents domaines appliqués.
- Donner une réponse adaptée aux TPE/PME.

Expertise

Modélisation, simulation numérique, optimisation, statistiques et... IA

Collaborations récentes ou en cours

Grands groupes :



Agences :



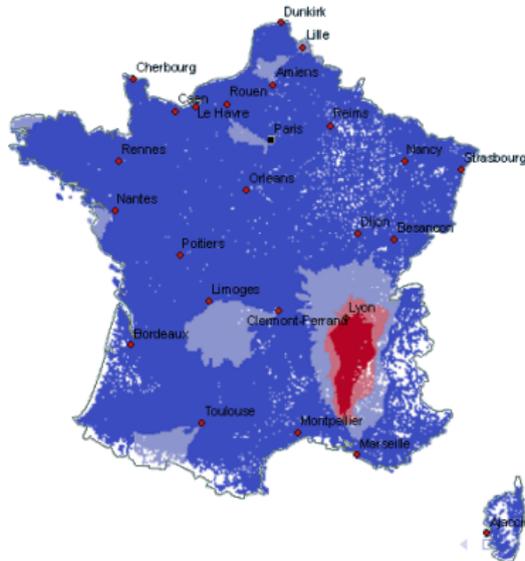
PME/ Start'up :



Forecasting seasonal epidemics

Collaboration with the SME Offisanté

The objective was to model seasonal epidemics propagation, with the use of medical drugs sell data, collected in the 7000 Offisanté drugstore clients.



Forecasting seasonal epidemics

Collaboration with the SME Offisanté

The numerical model, based on a coupled set of PDEs, has lead to a quantitative prediction of seasonal epidemics for the next five days from the Offisanté data.

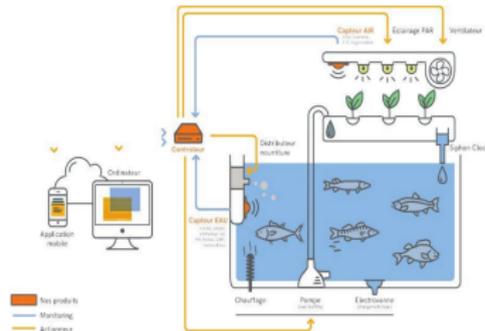


The results have been tested and validated on many real cases and a first 'epidemic forecast' broadcasting has already been done.

Aquaponics models

Collaboration with the start'up Selfeden

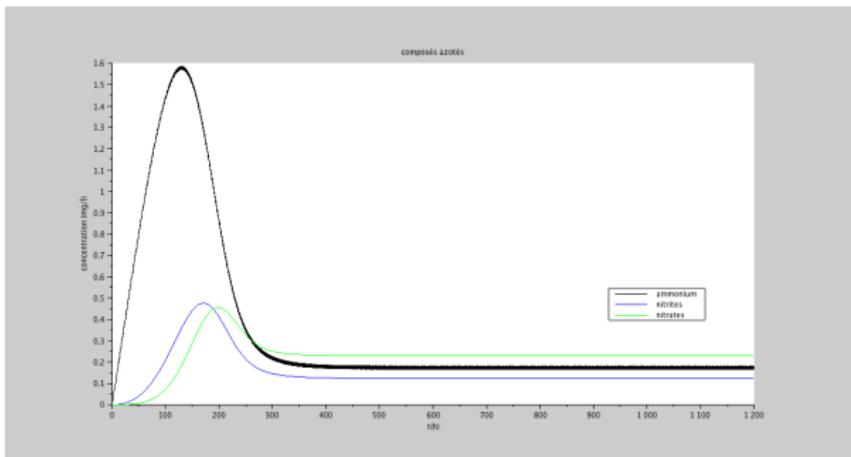
Aquaponics refers to any system that combines conventional aquaculture (raising aquatic animals such as fish) with hydroponics (cultivating plants in water) in a symbiotic environment.



The objective was to develop a mathematical model for the observation of the coupled system fish + house plants with respect to various parameters.

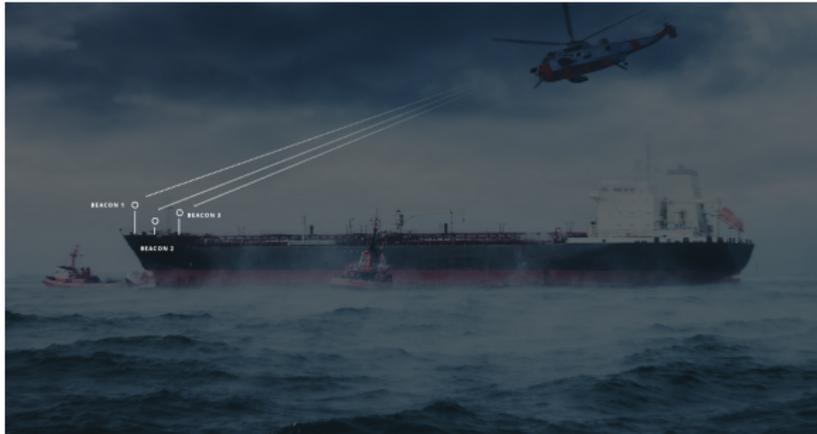
Collaboration with the start'up Selfeden

The coupled ODE system that has been developed models time evolution of the main chemical components of the azota cycle (ammonium, nitrites, nitrata) into water.



Collaboration with the SME INTERNEST

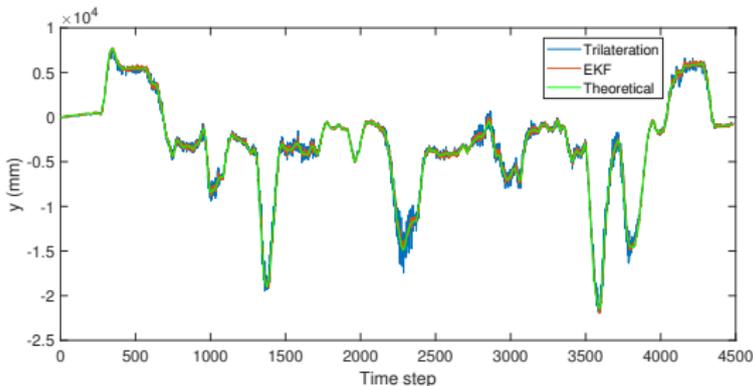
Interneest has developed a robust technology based on ultra sounds measurements to secure drone landing in hostile conditions.



The objective was to find the best approach for drone trajectory tracking.

Collaboration with the SME INTERNEST

Various data assimilation have been tested on realistic flight to find the best solution for drone trajectory tracking.



An adapted version of an Extended Kalman Filter has given the best results in a very noisy context.

Petit-déjeuner “Maths et IA”

Christophe Chalons, Laurent Dumas, Maëlle Nodet,
Florian Omnès, Ider Tseveendorj

29/03/2019

The logo for IMQSE is a horizontal rectangle with a green-to-teal gradient. The text "IMQSE" is written in white, bold, sans-serif capital letters.

IMQSE

The logo for LMV consists of the letters "LMV" in a large, bold, black sans-serif font. Below it, the text "Laboratoire de mathématiques de Versailles - CNRS UMR 8100" is written in a smaller, black sans-serif font.

LMV
Laboratoire de mathématiques
de Versailles - CNRS UMR 8100



Plan

- ① Intelligence artificielle : tentative de définition et exemples
- ② Apprentissage automatique par réseaux de neurones
- ③ Logiciel CPP-NN
- ④ Exemple d'utilisation : Data Challenge Engie

Tentative de définition

- Tentative de faire réaliser par des ordinateurs des tâches qui étaient réalisées par les humains

Tentative de définition

- Tentative de faire réaliser par des ordinateurs des tâches qui étaient réalisées par les humains
- L'IA fait apparaître de nouvelles tâches, de nouveaux usages

Grandes classes de problèmes d'apprentissage supervisé

- Classification

Grandes classes de problèmes d'apprentissage supervisé

- Classification
- Régression

Grandes classes de problèmes d'apprentissage supervisé

- Classification
- Régression
- **Transcription**

Grandes classes de problèmes d'apprentissage supervisé

- Classification
- Régression
- Transcription
- **Traduction**

Grandes classes de problèmes d'apprentissage supervisé

- Classification
- Régression
- Transcription
- Traduction
- Détection d'anomalies

Grandes classes de problèmes d'apprentissage supervisé

- Classification
- Régression
- Transcription
- Traduction
- Détection d'anomalies
- Synthèse de données similaires

Quelques exemples

- Classification d'e-mails, rejet des SPAMS



Quelques exemples

- Classification d'e-mails, rejet des SPAMS



- Proposer des postes adaptés au profil des candidats



Quelques exemples

- Classification d'e-mails, rejet des SPAMS



- Proposer des postes adaptés au profil des candidats



- Proposer des recommandations de produits ciblées



Quelques exemples

- Classification d'e-mails, rejet des SPAMS



- Proposer des postes adaptés au profil des candidats



- Proposer des recommandations de produits ciblées



- Traitement d'imagerie médicale : aide au diagnostic, diagnostic



Quelques exemples

- Classification d'e-mails, rejet des SPAMS



- Proposer des postes adaptés au profil des candidats



- Proposer des recommandations de produits ciblées



- Traitement d'imagerie médicale : aide au diagnostic, diagnostic



- Chatbots médicaux : Ada Health, Babylon Health, Your.MD, etc.

Position du problème

$x^{(\text{train})} \longrightarrow y^{(\text{train})}$	Connu
$x^{(\text{test})} \longrightarrow y^{(\text{test})}$	À déterminer

Position du problème

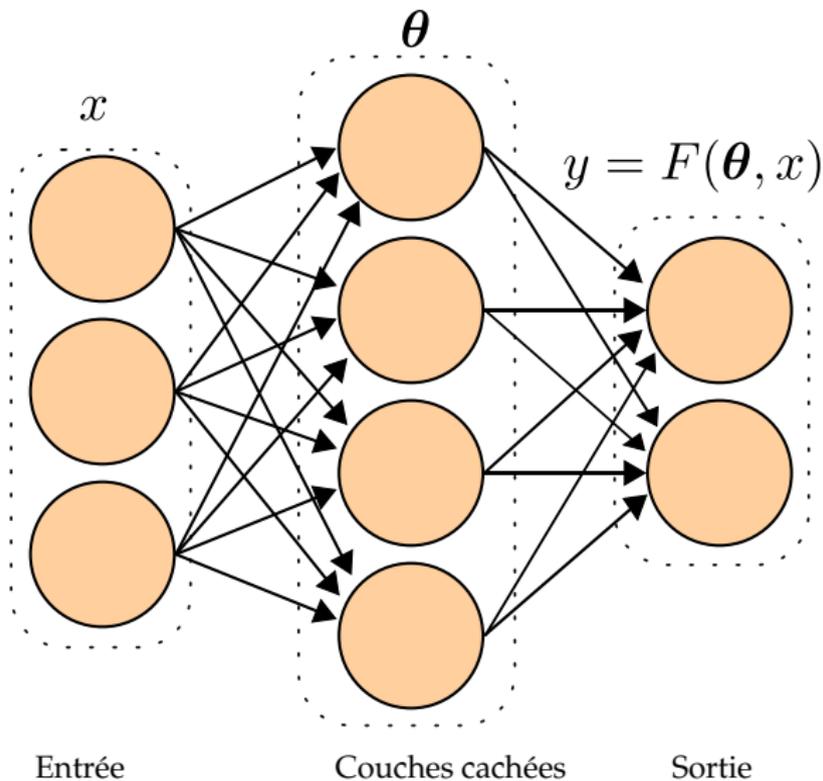
$x^{(\text{train})} \longrightarrow y^{(\text{train})}$	Connu
$x^{(\text{test})} \longrightarrow y^{(\text{test})} = F(\theta, x^{(\text{test})})$	À déterminer

où

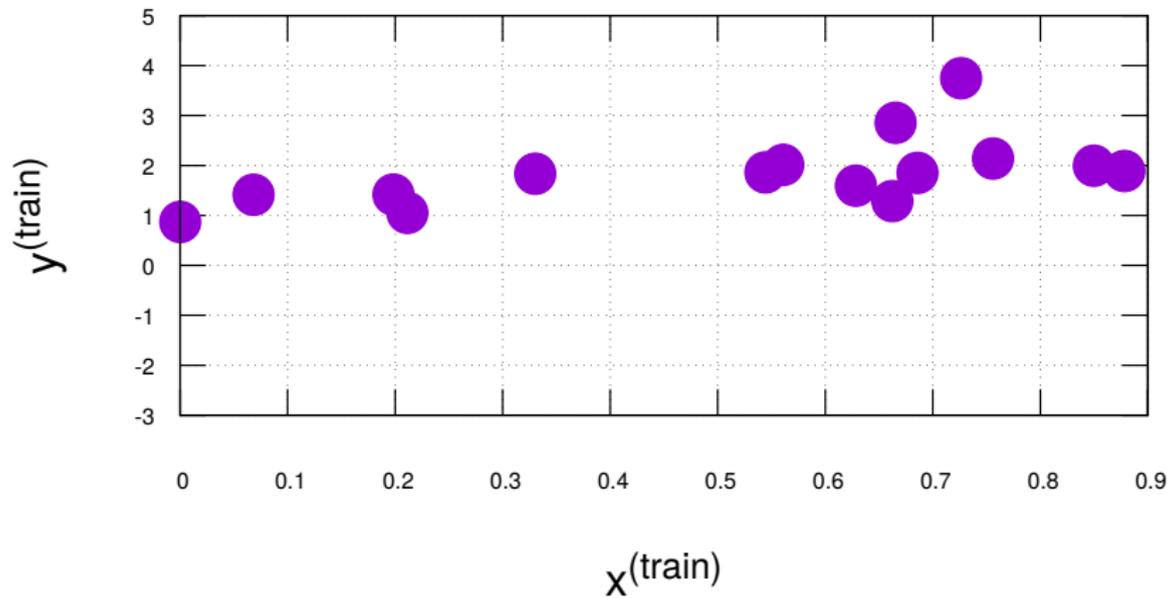
$$x \mapsto F(\theta, x)$$

représente le modèle par réseau de neurones.

Qu'est-ce qu'un réseau de neurones ?



Exemple de données en 1D



Fonction coût

Définition

La fonction coût mesure l'écart entre les données réelles et les données prédites.

$$J(\boldsymbol{\theta}) = L(F(\boldsymbol{\theta}, \mathbf{x}^{(\text{train})}), \mathbf{y}^{(\text{train})})$$

Exemples de fonctions coût

- “Mean square error”

$$\text{MSE}^{(\text{train})}(\boldsymbol{\theta}) = \frac{1}{n^{(\text{train})}} \sum_{i=1}^{n^{(\text{train})}} \left\| F(\boldsymbol{\theta}, \mathbf{x}_i^{(\text{train})}) - y_i^{(\text{train})} \right\|^2$$

- “Mean absolute error”

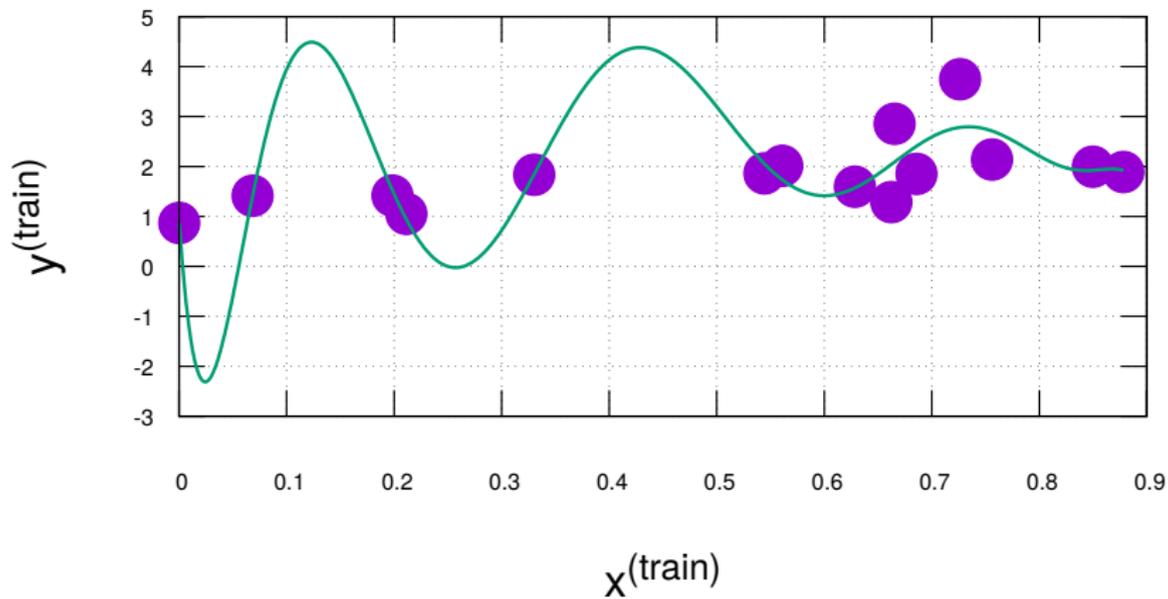
$$\text{MAE}^{(\text{train})}(\boldsymbol{\theta}) = \frac{1}{n^{(\text{train})}} \sum_{i=1}^{n^{(\text{train})}} \left\| F(\boldsymbol{\theta}, \mathbf{x}_i^{(\text{train})}) - y_i^{(\text{train})} \right\|$$

Objectifs en apprentissage par réseaux de neurones

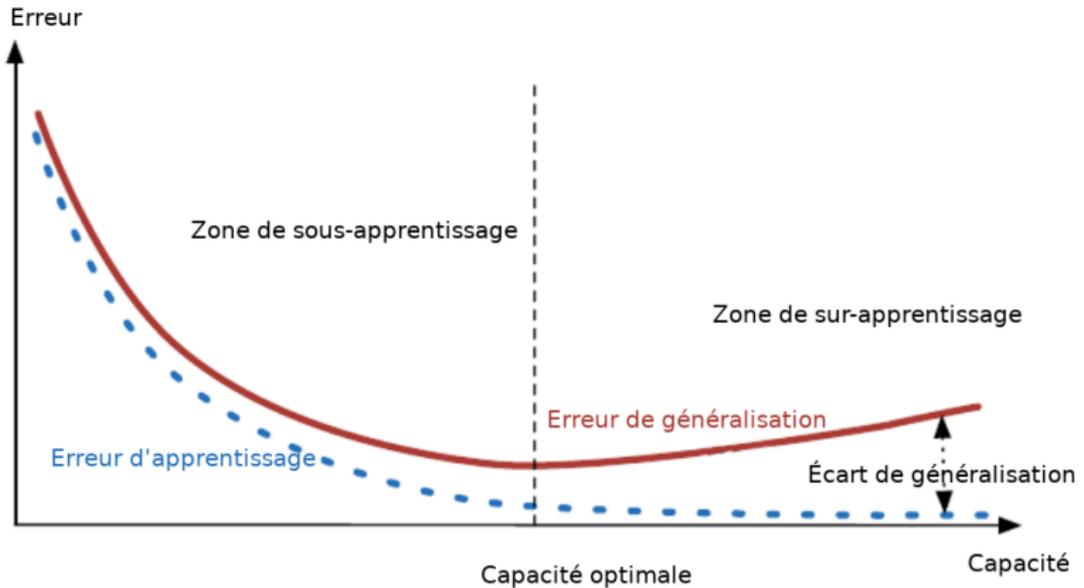
On souhaite obtenir un réseau de neurones qui :

- Minimise la fonction coût $\theta \mapsto J(\theta)$
- Évite le sur-apprentissage

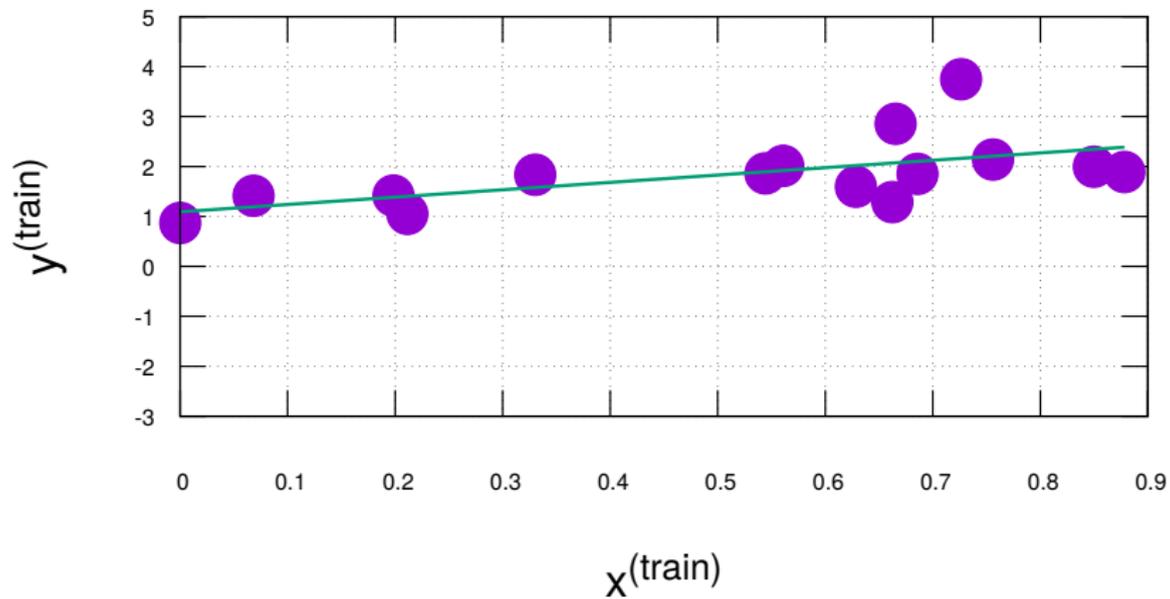
Exemple de sur-apprentissage



Capacité du modèle et sur-entraînement



Modèle plus adapté aux données



Logiciel CPP-NN

Mise en œuvre



Logiciel CPP-NN

Mise en œuvre

- Calcul de la sensibilité de la fonction coût par rapport aux paramètres θ du réseau de neurones



Logiciel CPP-NN

Mise en œuvre

- Calcul de la sensibilité de la fonction coût par rapport aux paramètres θ du réseau de neurones
- Méthode de gradient stochastique



Logiciel CPP-NN

Mise en œuvre

- Calcul de la sensibilité de la fonction coût par rapport aux paramètres θ du réseau de neurones
- Méthode de gradient stochastique
- Validation sur un jeu de données connu



Data Challenge Engie

On dispose de 75 données physiques mesurées à 600000 instants différents sur quatre éoliennes.

Objectif

Prédire précisément la production électrique connaissant ces mesures.



Data Challenge Engie

On dispose de 75 données physiques mesurées à 600000 instants différents sur quatre éoliennes.

Objectif

Prédire précisément la production électrique connaissant ces mesures.

- 1 Lecture des fichiers de configuration ;
- 2 Lecture des données ;
- 3 Apprentissage par réseau dense ;
- 4 Écriture des données prédites ;
- 5 Calcul du score sur $y^{(\text{test})}$ et $F(\theta, x_i^{(\text{test})})$

Résultats

$$\text{MAE}^{(\text{test})}(\boldsymbol{\theta}) = \frac{1}{n^{(\text{test})}} \sum_{i=1}^{n^{(\text{test})}} \left| F\left(\boldsymbol{\theta}, x_i^{(\text{test})}\right) - y_i^{(\text{test})} \right|$$

Méthode	Score MAE (kW)
XGBoost	24
CPP-NN	39.6
Benchmark Engie (1 variable)	75

Utilisations possibles de CPP-NN

- Problèmes de régression
- Problèmes de classification
- Problèmes dépendant du temps : utilisation d'observations à plusieurs instants successifs
- *etc.*