# Apprentissage de systèmes dynamiques à partir d'observations

AVERTY Tristan, Master 1 Mathématiques et Intéractions, La Rochelle Université

***« Il est possible d'apprendre des dynamiques très complexes avec seulement des informations partielles »***

## ❶ Objectifs

- **Prévoir** l'évolution de processus complexes et non-linéaires à partir d'observations ;
- Palier le **manque d'informations** lié à l'observation ;
- Utiliser un **réseau de neurones** pour estimer le terme évolutif du processus ;
- **Évaluer la méthode** sur des équations de dynamiques des fluides et sur des expérimentations climatiques ;
- Obtenir des résultats **sans imposer d'*a priori*** sur la forme de l'équation étudiée.

## ❷ Problème

- $\Omega \subset \mathbb{R}^d$ : **domaine spatial** d'étude ;
- **Processus spatio-temporel** :

$$\begin{aligned} X : \mathbb{R}_+ \times \Omega &\longrightarrow \Omega \\ (t, x) &\longmapsto X_t \end{aligned}$$

- $X_t$ : **état** du système à un instant $t$ ;
- $X_t$ suit une **équation d'évolution**

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = F(X_t) \tag{1}$$

avec $F$ non linéaire et complexe ;

- **Enjeu :** « trouver » ce $F$.

## ❸ Introduction & notions

- $\{F_\theta\}$ : ensemble de fonction admissibles ;
- Chercher $\theta$ tel que la solution $X^\theta$ de

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = F_\theta(X_t) \tag{2}$$

soit en accord avec les données mesurées.

- On observe rarement l'état du système $X_t$ complet (**observations**) ;
- $\mathcal{H}$ : **opérateur d'observation** supposé connu, fixé et différentiable ;
- $Y_t$ : **informations observées** à partir de l'état $X_t$ ;
- $X_0$ : **condition initiale** de notre processus.

On a donc le problème ci-dessous :

$$\begin{cases} X(0, x) = X_0 \\ \dfrac{\mathrm{d}X_t}{\mathrm{d}t} = F(X_t) \\ Y_t = \mathcal{H}(X_t) \end{cases} \tag{3}$$

## ❹ Contrôle optimal

- Lien entre réseaux résiduels et systèmes dynamiques ;
- **Variables de contrôles** correspondent aux paramètres du réseau ;
- **Fonction coût** :

$$\mathcal{J}\left(Y, \widetilde{Y}\right) = \int_0^T \left\| Y_t - \widetilde{Y}_t \right\|^2_{L^2(\Omega)} \mathrm{d}t$$

  - $Y$ : **observations** du système ;
  - $\widetilde{Y}$ : **sortie** du système.

- $X_t$ vérifie (2) donc on a le problème

$$\min_\theta \mathbb{E}_{Y \in \text{Dataset}} \left[\mathcal{J}\left(Y, \mathcal{H}(X)\right)\right]$$
$$\text{s.c} \begin{cases} \dfrac{\mathrm{d}X_t}{\mathrm{d}t} = F_\theta(X_t) \\ X_0 = g_\theta\left(Y_{-k}, \breve{X}_0\right) \end{cases} \tag{4}$$

  - $F_\theta$ : **trajectoire** de $X$ ;
  - $g_\theta$ : **condition initiale** ;
  - $Y_{-k} = \{Y_{-k+1}, \ldots, Y_0\}$.

- On note $X^\theta$ la **solution du problème** (4).

## ❺ Calcul du gradient par état adjoint

- Méthode de **descente de gradient** :

$$\text{Dériver } \theta \longmapsto \mathbb{E}_Y\left[\mathcal{J}\left(Y, \mathcal{H}\left(X^\theta\right)\right)\right] \iff \text{Calculer } \frac{\partial X^\theta}{\partial \theta} \text{ (ce qui est difficile)}$$

- Définir le **lagrangien**

$$\mathcal{L}(X, \lambda, \mu, \theta) = \mathcal{J}(X) + \int_0^T \left\langle \lambda_t, \frac{\mathrm{d}X_t}{\mathrm{d}t} - F_\theta(X_t) \right\rangle \mathrm{d}t + \langle \mu, X_0 - g_\theta \rangle \tag{5}$$

- **Équation de l'état adjoint** :

$$\partial_\theta \mathcal{J}(X^\theta) = -\int_0^T \left\langle \lambda_t, \partial_\theta F_\theta\left(X_t^\theta\right) \right\rangle \mathrm{d}t + \langle \lambda_0, \partial_\theta g_\theta \rangle \tag{6}$$

où $\lambda$ est solution de

$$\partial_t \lambda_t = A_t \lambda_t + B_t \tag{7}$$

résolu *backward* en prenant $\lambda_T = 0$ avec

$$\begin{aligned} A_t &= -\left(\partial_X F_\theta\left(X_t^\theta\right)\right)^* \\ B_t &= 2\left(\partial_X \mathcal{H}\left(X_t^\theta\right)\right)^* \left(\mathcal{H}\left(X_t^\theta\right) - Y_t\right) \end{aligned}$$

## ❻ Algorithme principal

**Input:**
- Échantillons $\left\{\left(Y_{-k}, \breve{X}_0\right), Y_\ell\right\}$
- Paramètre initial $\theta$

**while** *ne converge pas* **do**
  Sélectionner $\left\{\left(Y_{-k}, \breve{X}_0\right), Y_\ell\right\}$
  **if** *état initial complètement observé* **then**
    $X_0 \leftarrow \breve{X}_0$
  **end**
  **else**
    $X_0 \leftarrow g_\theta\left(Y_{-k}, \breve{X}_0\right)$
  **end**
  Résoudre *forward*
  $\dfrac{\mathrm{d}X_t}{\mathrm{d}t} = F_\theta(X_t), X(0) = X_0, t \in [0, \ell]$
  Résoudre *backward*
  $\dfrac{\mathrm{d}\lambda_t}{\mathrm{d}t} = A_t \lambda_t + B_t, \lambda_\ell = 0, t \in [0, \ell]$
  Calculer le gradient $\partial_\theta \mathcal{J}\left(X^\theta\right)$
  Mettre à jour $\theta$ dans la direction de plus forte pente
**end**
**Output:** Paramètre appris $\theta$

## ❼ Barré de Saint-Venant

- **Discrétisation spatiale** : $80 \times 80$ (1600 acquisitions) ;
- **Train / test** : 37,5 % / 62,5 % ;
- **État complet** comme entrée initiale ;
- Modèle capable de prévoir les observations en **temps long** ;
- Prédit l'**état complet** $X_t$ et pas seulement les observations $Y_t$ (présence d'états cachés).

## ❽ Glorys2v4

- Température de la surface de la mer ;
- **Pas de temps** : 1 jour ;
- **Discrétisation spatiale** : sous-régions de $64 \times 64$ ;
- **Train / test** : 93,8 % / 6,2 % ;
- Entrée : état complet **non disponible** ;
- Problème de **conditions aux bords** : travail sur une petite zone alors que le modèle fonctionne sur une plus grande zone ;

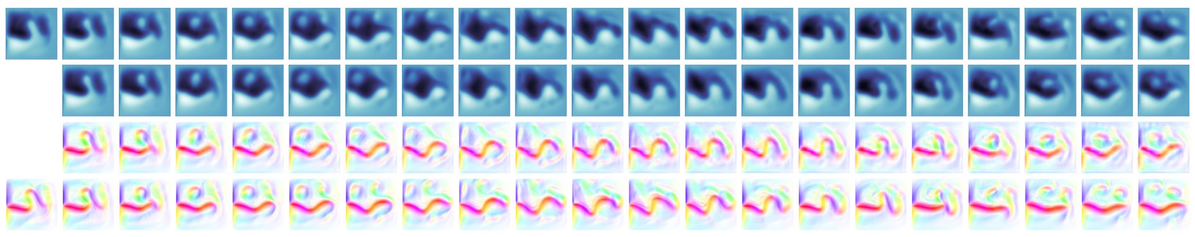

FIGURE 1 : Prédiction du modèle de Barré de Saint-Venant. De haut en bas : entrée et observations cibles, sortie du modèle, état caché estimé, et état caché de la vérité terrain

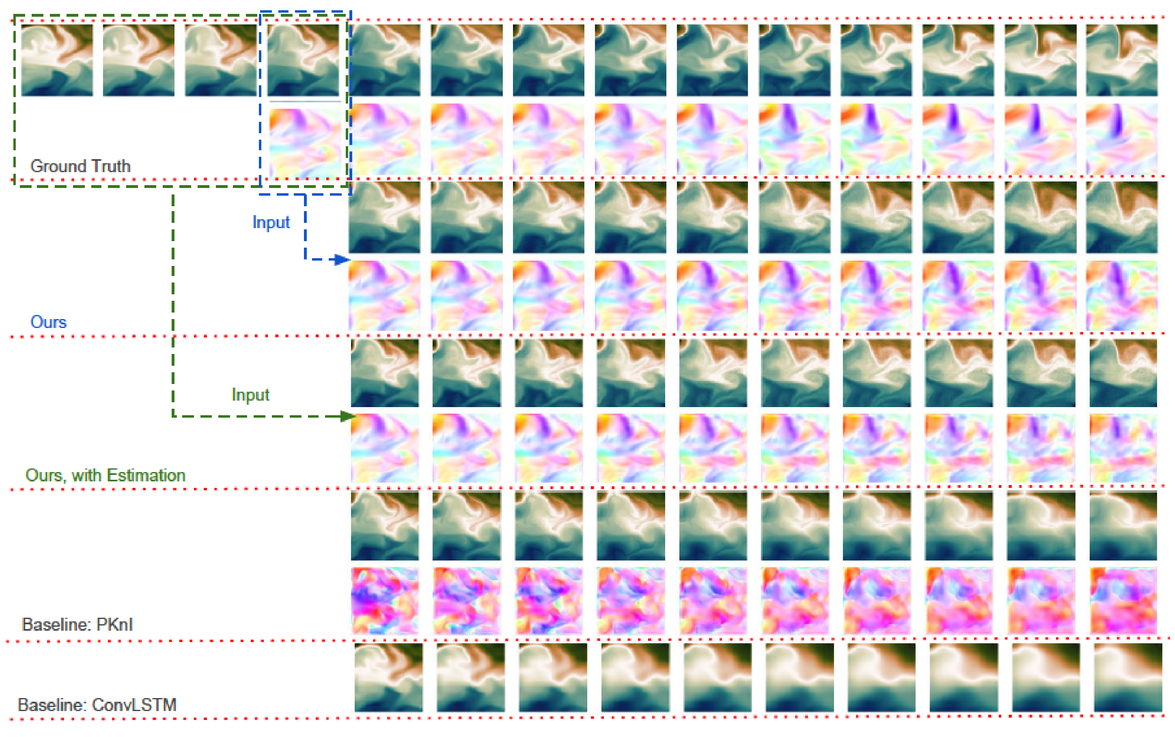


FIGURE 2 : Prédiction du modèle Glorys2v4. De haut en bas : entrée et observations cibles ainsi que l'état caché de la vérité terrain associé, les sorties du modèle, la variante du modèle lorsque les conditions initiales sont estimées à partir des observations, les sorties du PKnl et du ConvLSTM.

## ❾ Détails de l'implémentation

- **Estimation** de $\mathcal{J}$ : **sélection** d'un petit sous-ensemble des données et **optimisation** par descente de gradient stochastique ;
- $F_\theta$ : **réseau résiduel standard** avec 2 couches de sous-échantillonnages et 6 blocs résiduels ;
- **Discrétisation** de l'équation *forward* : simple **schéma d'Euler** avec
  - $\Delta t = 1/3$ ;
  - $\Delta x$ imposé par les données.
- **Initialisation orthogonale** des poids du réseau résiduel ;
- **Taux d'apprentissage** : $10^{-5}$ ;
- Bibliothèque Deep Learning : **Pytorch**.

## ❿ Comparaisons

Résultats quantitatifs avec Glorys2v4 :

| ***Modèles*** | $K = 5$ | $K = 10$ |
|---|---|---|
| Modèle présenté | 0.124 | 0.231 |
| Modèle présenté (avec estimations) | **0.113** | **0.209** |
| PKnl | 0.145 | 0.250 |
| ConvLSTM | 0.137 | 0.224 |

TABLE 1 : Erreur quadratique moyenne entre les observations prédites et la vérité terrain pour différents horizons $K$ (le plus faible est le mieux).

où

- PKnl est un modèle de deep-learning intégrant la **connaissance physique** ;
- ConvLSTM est un modèle de *long short-term memory* utilisant des **transitions convolutionnels**.

## ⓫ Conclusion

- **Prédiction de l'évolution** d'un système **complexe** et **non linéaire** et lorsque l'état n'est **pas entièrement observé** ;
- **Estimation** du paramètre d'évolution avec un réseau de neurones ;
- **Algorithme d'apprentissage** : théorie du **contrôle optimal** ;
- Méthode capable
  - de produire des prévisions de haute qualité à **différents horizons** ;
  - d'apprendre avec une bonne précision la **dynamique des états sous-jacents**

MAIS :

- On pourrait **trouver directement** $T$ tel que $X_{t+\Delta t} = T(X_t)$ sans voir $X$ comme résultant d'une équation différentielle ;
- **Discrétisation explicite** : limitations de la classe d'équations (équations raides nécessitent des **méthodes implicites**).
- Que se passerait-il si l'opérateur $\mathcal{H}$ était **plus complexe** ?

## Référence

[1] I. Ayed E. de Bézenac.
*Learning Dynamical Systems from Partial Observations.*

La Rochelle Université

MATHEMATIQUES IMAGE APPLICATIONS UNIVERSITÉ DE LA ROCHELLE

Poster réalisé le 21 avril 2019