

# PFIA 2023

Optimisation du pilotage industriel d'une piscine publique grâce à la modélisation dirigée par les données : proposition d'un jeu de données applicatif

---

François Gauthier-Clerc, Purecontrol, IMT Atlantique, LS2N, France

Hoel Le Capitaine, Polytech Nantes, LS2N, France

Fabien Claveau, IMT Atlantique, LS2N, France

Philippe Chevrel, IMT Atlantique, LS2N, France

Juillet 2023

purecontrol



1. Contexte scientifique et industriel
2. Le système Piscine
3. Les limites des modélisations dirigées par les données
4. Proposition d'un banc d'essai
5. Conclusion et futures directions

## Contexte scientifique et industriel

---

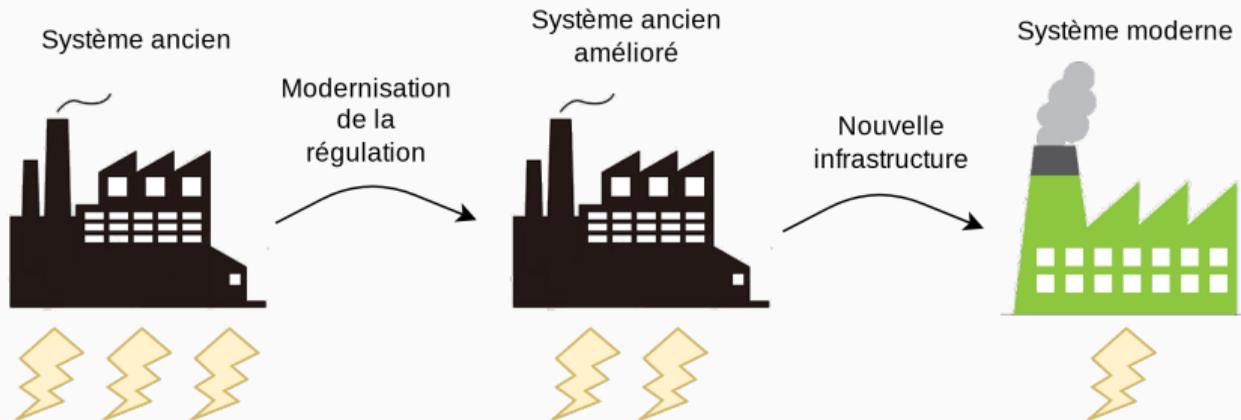
Notre société fait face à d'importants défis:

- Dérèglement climatique.
- Augmentation de la population mondiale.
- Raréfaction des matières premières.

Les solutions envisagées passent par:

- Transition énergétique, implique des sources moins pilotage donc augmentation de la volatilité des prix de l'énergie.
- Modernisation des postes de dépenses en énergies, demande des manières première pour réaliser les travaux.

# Modernisation des réglementations industriels



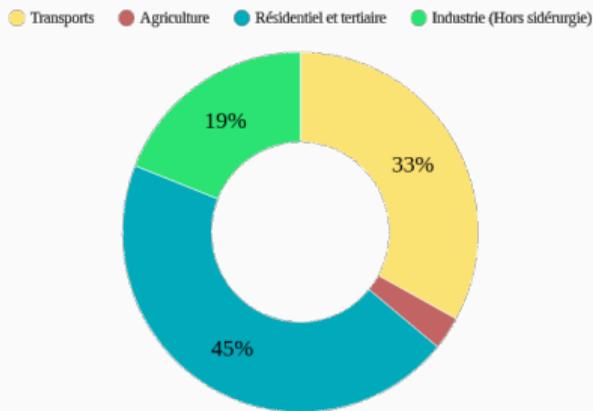
Modernisation des réglementations industrielles:

- Réduit l'impact énergétique
- Besoin en matière première nul
- Offre une solution à volatilité des prix.

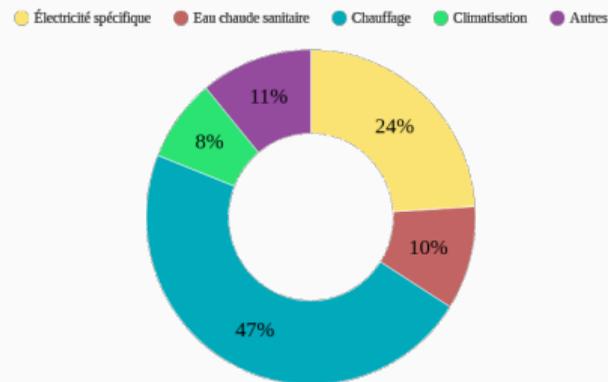
⇒ Le monde du bâtiment et de la piscine publique est un bon exemple d'application.

# Secteur du bâtiment

Le secteur du bâtiment et tertiaire représente 45% de la consommation énergétique en France dont 47% en chauffage [1].

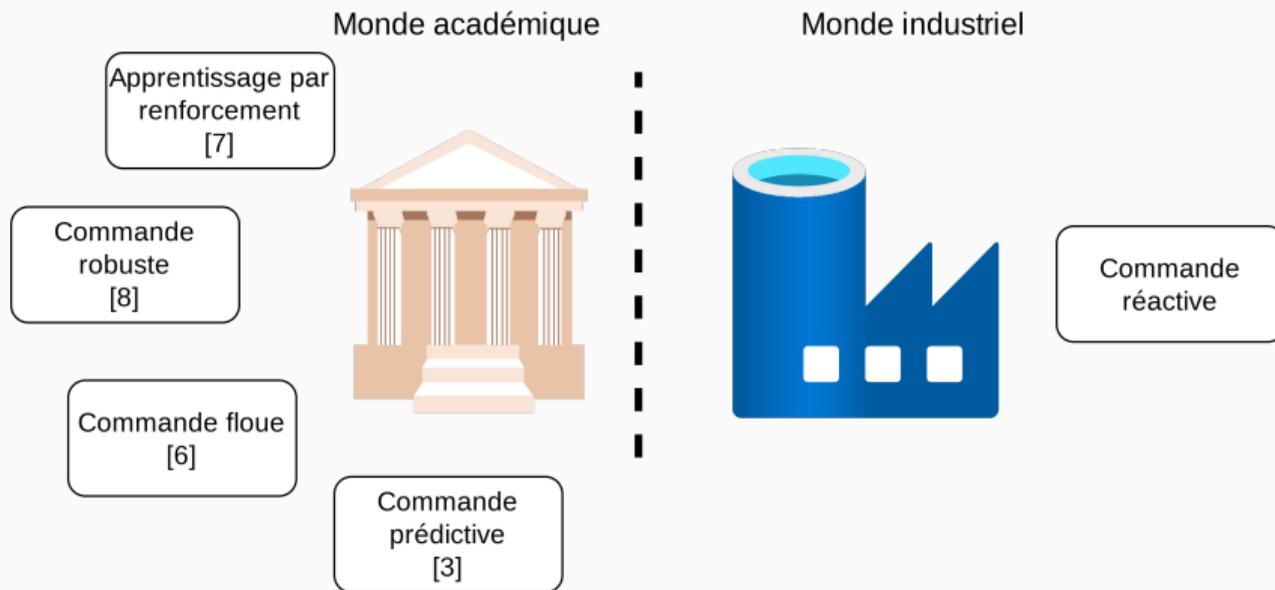


**Figure 1:** Répartition de la consommation d'énergie par secteur en France (2018).



**Figure 2:** Répartition de la consommation par application dans le bâtiment en France (2018).

⇒ Un **complexe aquatique** consomme plus de **10 fois** plus d'énergie qu'un bâtiment classique [2].



⇒ 5% et 20% d'énergie peuvent être économisés avec des approches modernes [3], [4].

**Solutions actuellement trop coûteuses**

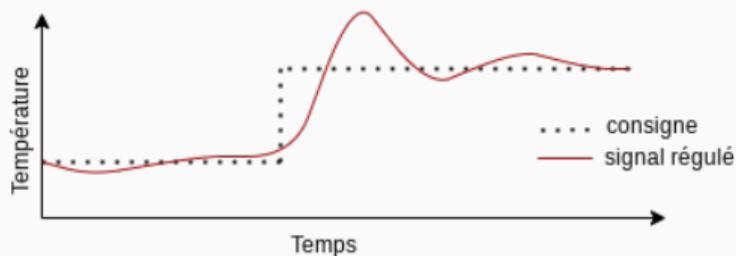
Une multitude de méthodes existent pour réduire l'impact carbone des systèmes de chauffage, parmi eux :

- Planification avancées [5].
- Contrôleur flou [6].
- Apprentissage par renforcement [7]
- Commande robuste [8]
- Model predictive Control (MPC) [3]
- Et d'autres...

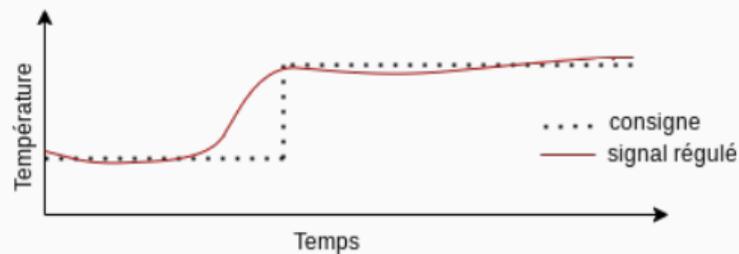
⇒ **Ce travail se concentrera sur la méthode MPC**

# Pilotages industriels souvent sous optimaux

Comparaison entre une régulation PI et une régulation avancées:



**Figure 3:** Illustration d'une régulation PI



**Figure 4:** Illustration d'une régulation avancée

Les solutions expertes peuvent permettre:

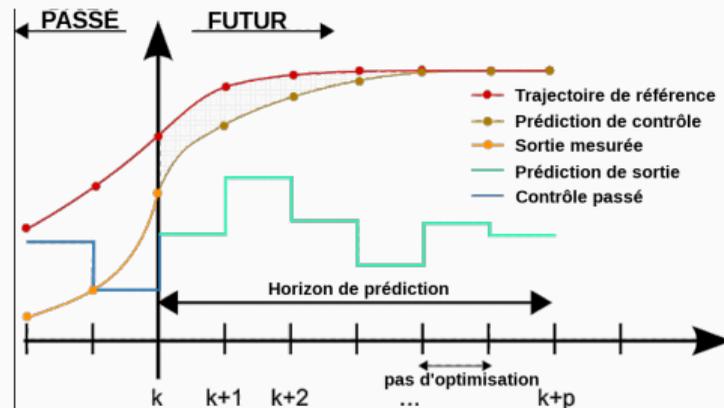
- Capacité d'anticipation (contrôleur prédictif).
- Adaptation à l'incertitude (contrôleur robuste/stochastique).
- Notion d'optimalité économique (contrôleur économique).

# Model Predictive Control

Planification basée modèle sur un horizon glissant.

$$u^*[t+1:t+H|t] = \operatorname{argmin} \sum_{k=1}^H \|y_k - y_{ref}\|_2^2 + V(y_H)$$

$$\text{avec} \begin{cases} x_k = f(x_{k-1}, u_{k-1}), \forall k \in [0, H] \\ x_0 = x[t] \\ x_k \in \mathcal{X}, u_k \in \mathcal{U} \end{cases}$$



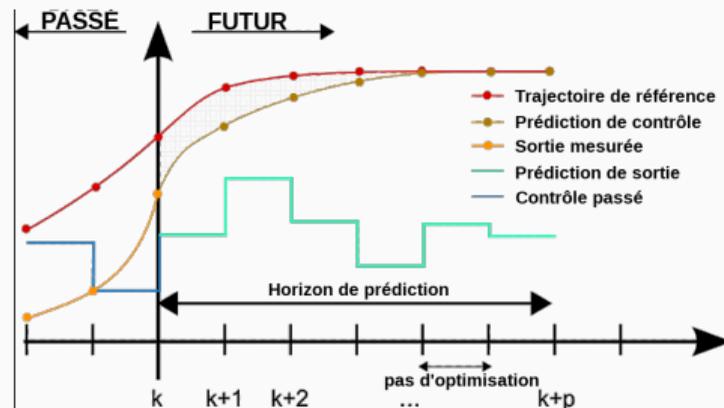
Avec  $H$ , l'horizon de prédiction,  $x$  l'état du système,  $y$  l'observation du système (signal à réguler),  $u$  variable de contrôle.

# Model Predictive Control

Planification basée modèle sur un **horizon glissant**.

Utilisation en boucle fermée:

- 1: **for**  $t = 1$  to  $\infty$  **do**
- 2: Estimation de l'état du système  $x[t]$
- 3: Récupération des prédictions  $T_{min}[t : t + H]$
- 4:  $u^*[t] = \text{planning\_MPC}(x[t], T_{min}[t : t + H])$
- 5: Application de  $u[t + 1|t]$
- 6: **end for**



Avec  $H$ , l'horizon de prédiction,  $x$  l'état du système,  $y$  l'observation du système (signal à réguler),  $u$  variable de contrôle.

La réalisation de notre planification génère un besoin de modèle.

Le modèle doit être :

- Relativement précis
- Avec une grande plage de validité
- Utilisable par des solveurs numériques.

$$\operatorname{argmin} \sum_{k=1}^H l(y_k - y_{ref}) + V(y_H)$$

avec

$$\begin{cases} x_k = f(x_{k-1}, u_{k-1}), \forall k \in [0, H] \\ x_0 = x[t] \\ H(x_k, u_k) \leq 0, G(x_k, u_k) = 0 \\ x_k \in \mathcal{X}, u_k \in \mathcal{U} \end{cases}$$

⇒ La formulation d'un modèle est coûteuse !

# Différentes approches de modélisation (*system identification*)

Il existe différentes philosophies pour modéliser un système:

## Modèle boîte blanche:

- loi physique
- Modélisation continue

Exemple:

$$c \frac{\partial \bar{A}_z}{\partial \bar{t}} = \nabla^2 \bar{A}_z + dp \frac{1}{\bar{x}} \left[ \left( \frac{\partial \bar{A}_z}{\partial \bar{x}} \right) \left( \frac{\partial \bar{B}_z}{\partial y} \right) - \left( \frac{\partial \bar{A}_z}{\partial y} \right) \left( \frac{\partial \bar{B}_z}{\partial \bar{x}} \right) \right]$$

$$c \frac{\partial \bar{B}_z}{\partial t} = \nabla^2 \bar{B}_z + \frac{p}{d} \frac{1}{\bar{x}} \left[ \left( \frac{\partial \bar{A}_z}{\partial y} \right) \frac{\partial}{\partial \bar{x}} \nabla^2 \bar{A}_z - \left( \frac{\partial \bar{A}_z}{\partial \bar{x}} \right) \frac{\partial}{\partial y} \nabla^2 \bar{A}_z \right]$$

Figure 5: Équation différentielle

## Modèle boîte grise:

- Inspiration physique.
- Approche hybride données/modèles

Exemple:

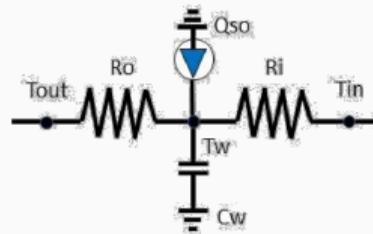


Figure 6: Modèle RC

## Modèle boîte noire:

- Modélisation dirigée par la données
- Modélisation à partir des observations.

Exemple:

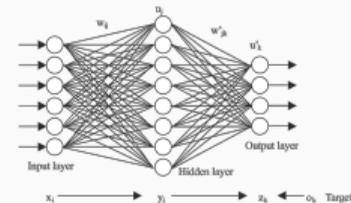


Figure 7: Fonction paramétrique

# Modélisation boîte noire

Il existe de nombreuses méthodes en théorie du contrôle pour l'identification en boîte noire :

$$\begin{cases} x[k+1] = Ax[k] + Bu[k] \\ y[k] = Cx[k] + Du[k] \end{cases}$$

Espace d'état Linéaire (LSS)

$$A(q^{-1})y[k] = B(q^{-1})u[k]$$

Auto Regressive Exogenous (ARX)

$$\begin{cases} x[k+1] = f_{\theta}(x[k], u[k]) \\ y[k] = g_{\theta}(x[k], u[k]) \\ x[k] = h_{\theta}(x[k], x[k-1], \dots, u[k], u[k-1], \dots) \end{cases}$$

Deep State space encoder (DESS)

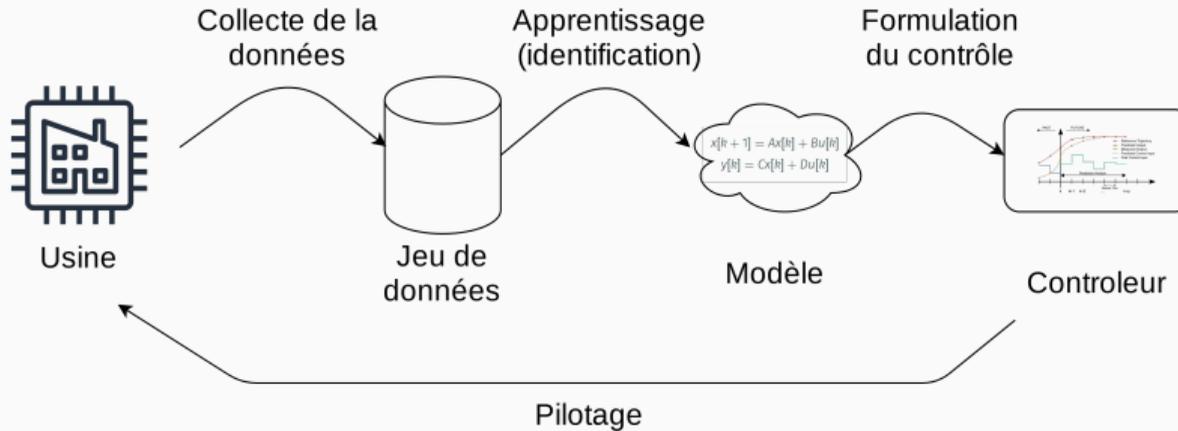
$$y[k] = f_{\theta}(y[k], y[k-1], \dots, u[k], u[k-1], \dots)$$

Non Linear Auto Regressive Exogenous (NLARX)

**Figure 8:** Illustration de quelques modèles de type boîte noire.

Pour plus d'informations  $\Rightarrow$  [9]

# Proposition d'une stratégie de modernisation



Cette approche offre les avantages suivants:

- Adaptation rapide aux spécificités.
- Réduction du besoin en expertise.
- Amélioration de l'efficacité du système à réguler.

⇒ Cette idée est retrouvée sous différentes appellations : *Learning based Model Predictive Control*, *Data-driven model predictive control*, *MPC based Reinforcement learning*.

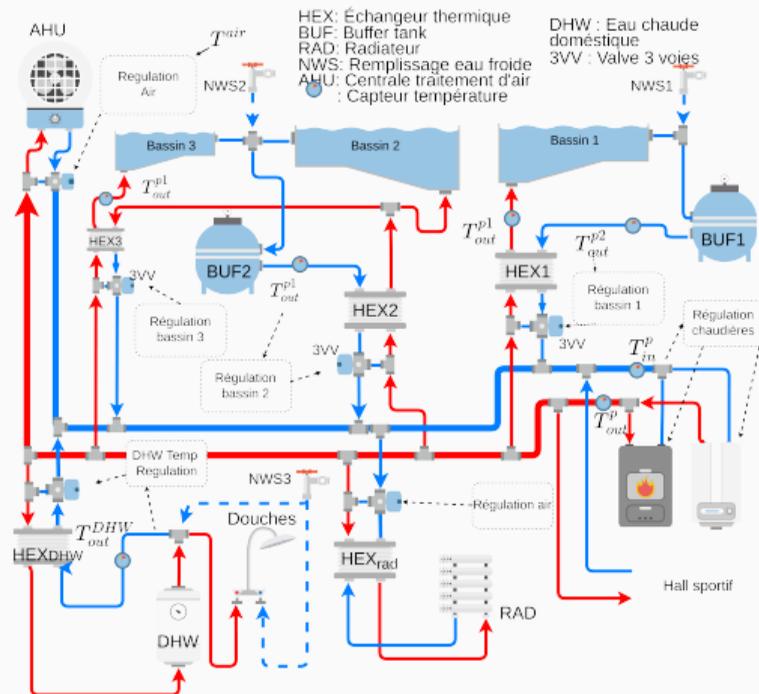
# Le système Piscine

---

# Le système de chauffage dans une piscine

Caractéristique du circuit de chaleur:

- Deux chaudières bois et gaz (1,5 mégawatt).
- Deux grands bassins.
- Un système de régulation de l'air.
- Eau chaude sanitaire.
- Hall sportif.
- SPA.



Objectif : Améliorer la régulation des deux bassins

Pilotage par ouverture de la **vanne 3 voies**.

Contrainte sur une température minimale.

Perturbation par l'ajout d'eau froide.

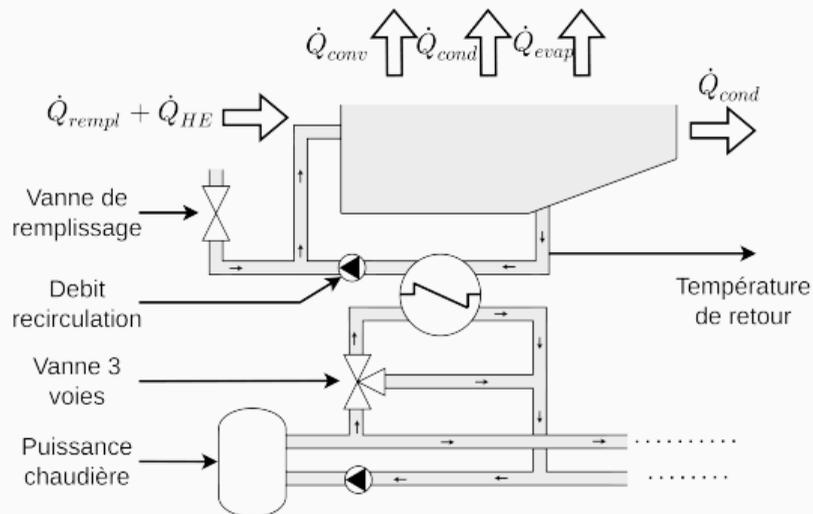


Figure 9: Sous-système d'un bassin

# Illustration des régulations

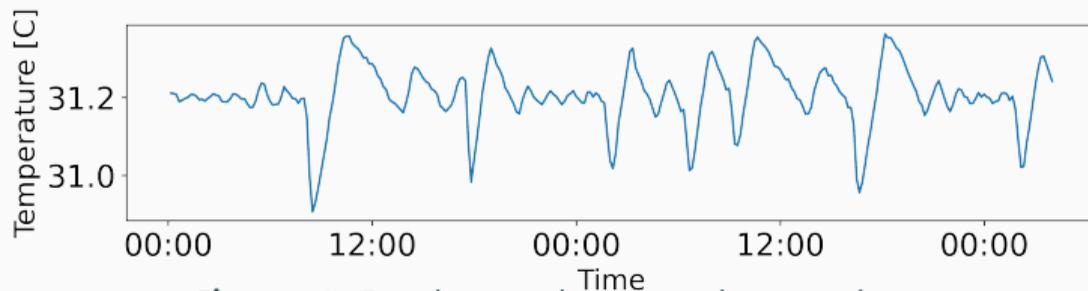


Figure 10: Bassin natation - consigne statique

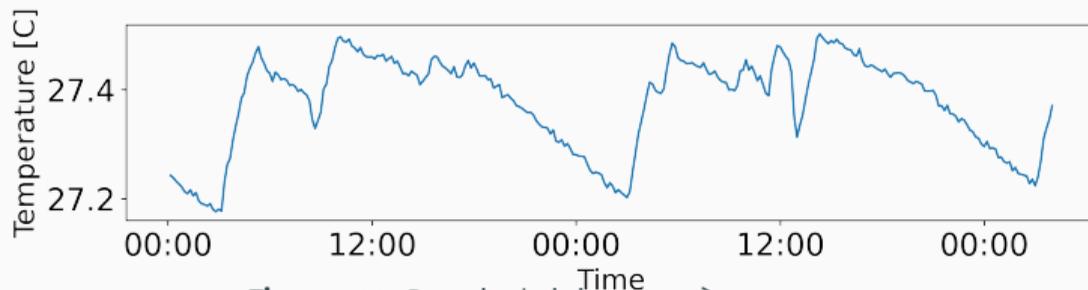


Figure 11: Bassin loisir - système expert

# Les limites des modélisations dirigées par les données

---

Données en opérations  $\Rightarrow$  Faible distribution pour l'apprentissage.

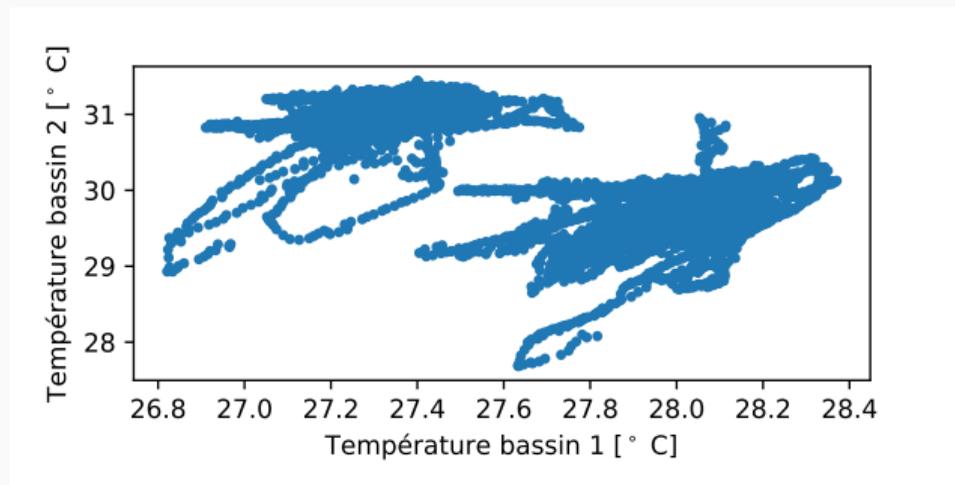
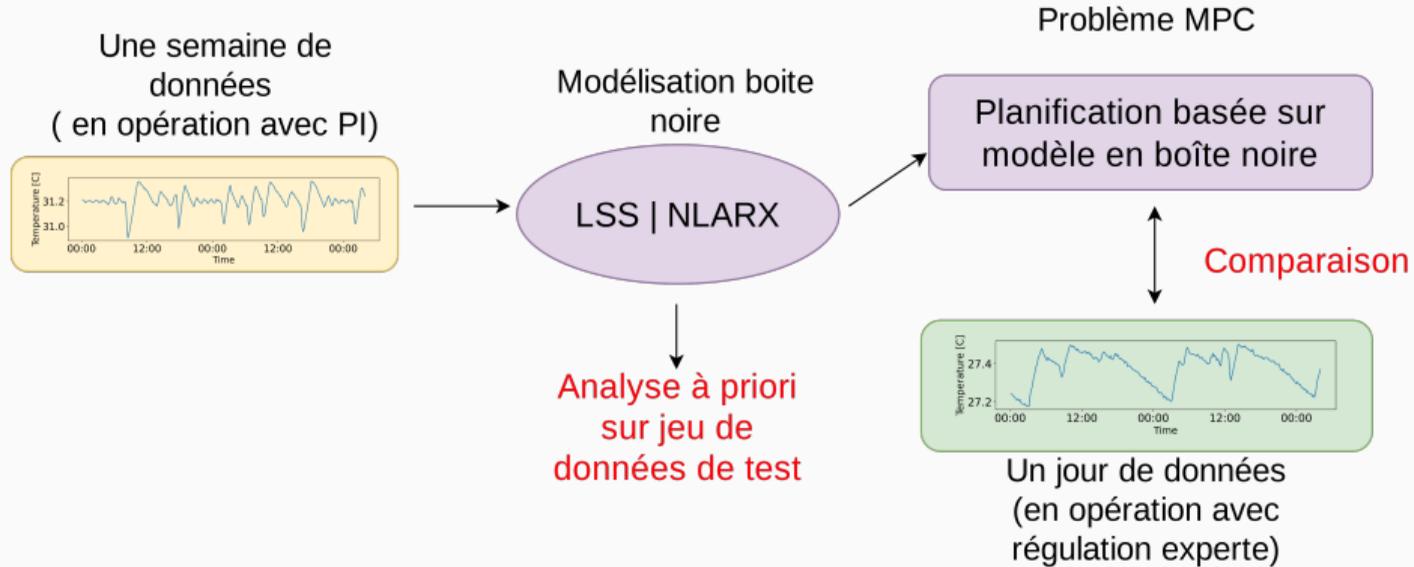


Figure 12: Plan de phase de notre jeu de données d'un système en opération.

Les modèles à haute expressivité admettent des limites dans ce régime de fonctionnement.

# Exemple d'application



Nous utiliserons pour l'entraînement des données avec pilotage PI et pour la comparaison de planification avec un pilotage expert.

## Exemple d'application

Modèle linéaire à espace d'état (LSS) :

$$\begin{cases} x[k+1] = Ax[k] + Bu[k] \\ y[k] = Cx[k] + Du[k] \end{cases}$$

Modèle NLARX avec réseau perceptron multicouche :

$$y[k] = f_{\theta}(y[k-1], \dots, y[k-n_a], \\ u[k-1], \dots, u[k-n_b])$$

$$u^* = \operatorname{argmin}_{u \in [0,1]^H} \sum_{k=0}^{H-1} \frac{u_1[k]}{H} \quad (1)$$

$$\text{avec } \begin{cases} y[k+1] = f_{\theta}(y[k], u_1[k], u_2[k]) \\ y[0] = T_{deb}, y[H-1] \geq T_{fin} \end{cases}$$

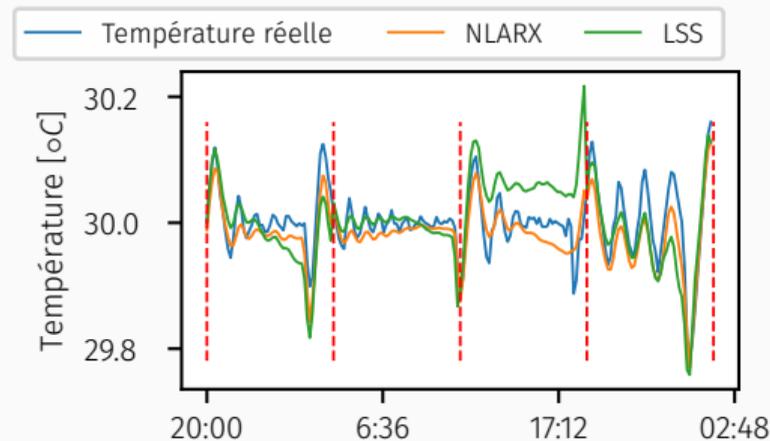
Formulation du problème de planification

L'utilisation unique de données réelles, nous impose une formulation de la planification qui cherche à reproduire le pilotage expert dans ces conditions aux limites.

# Résultats sur les données de test

Analyse de la prédiction sur 8 heures:

- Modèle linéaire : 0,08
- Modèle non-linéaire : 0,04



**Figure 13:** Prédiction sur un horizon de 8 heures avec des données de test (bassin loisir).

⇒ Le modèle non linéaire réduit de moitié les erreurs de prédiction sur un horizon de 8 heures.

# Test de planification

Comparaison avec le pilotage expert et nos planification basée données.

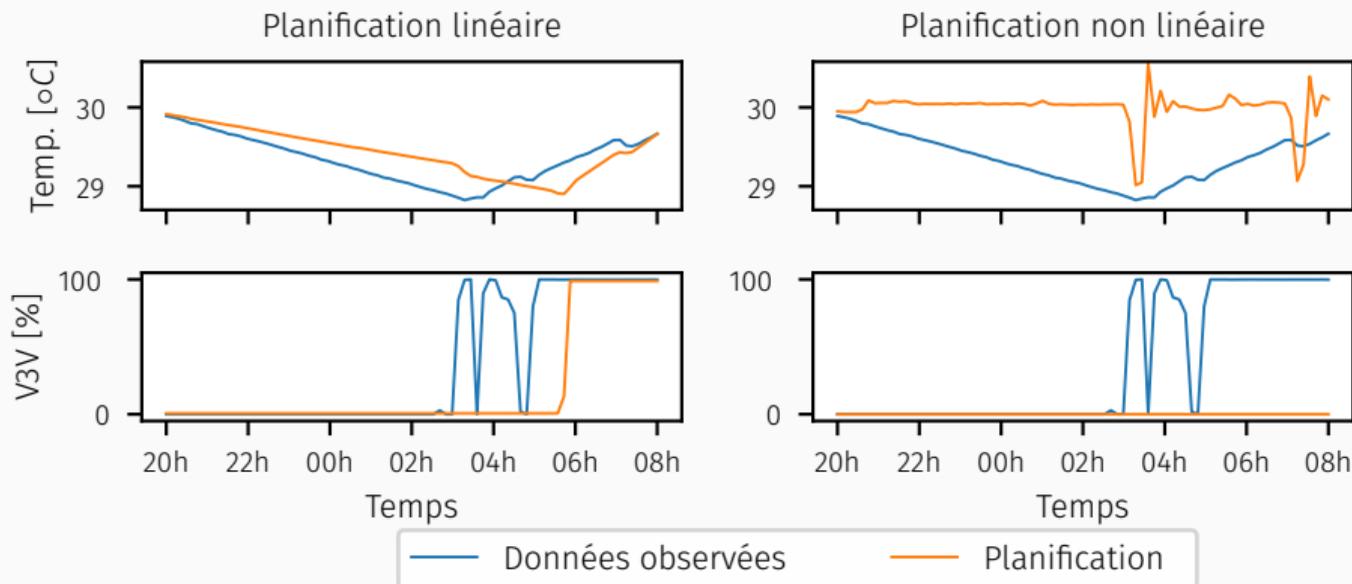


Figure 14: Comparaison données-planification

## Remarques:

- Le modèle linéaire est cohérent malgré des biais.
- Le modèle non-linéaire a sur-appris et donne des prédictions incohérentes.

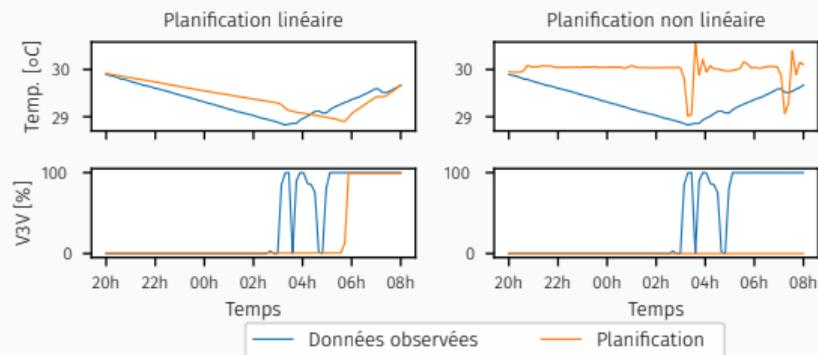


Figure 15: Comparaison données-planification

⇒ Malgré des biais pouvant impacter la qualité du pilotage, le modèle linéaire reste à privilégier.

## Proposition d'un banc d'essai

---

# banc d'essai de identification boite noire - Piscine publique

Proposition d'un banc d'essai pour l'identification en boîte noire.

- Des **données réelles** d'un centre aquatique en opération.
- Formulation de la modélisation en **Multiple-Input-Multiple-Output (MIMO)**
- **Critères de performance** pour une application en contrôle.



Figure 16: Photo du hall des bassins

⇒ Toutes les ressources sont disponibles :

<https://benchmark-datadriven-sysid.purecontrol.com/>

# Formulation MIMO du problème

Le banc d'essai est formulé sous la forme d'un problème d'identification MIMO pour les températures des deux bassins principaux.

- Sélection de données d'entrée le plus générique possible.
- Aucune augmentation de données n'est réalisée.

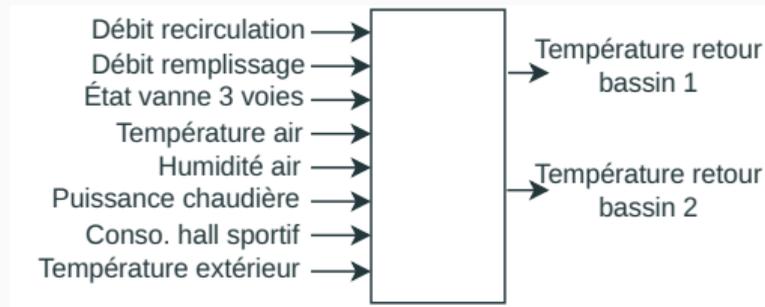


Figure 17: La formulation MIMO

# La section de données

Un total de 8 mois de données, deux régimes de fonctionnement et un taux de sous-échantillonnage de 10 minutes.



Figure 18: Chronologie de la base de données

On divise les données en trois parties:

- Jeu de données d'apprentissage
- Jeu de données de test
- Jeu de données d'extrapolation

# Rappel des données

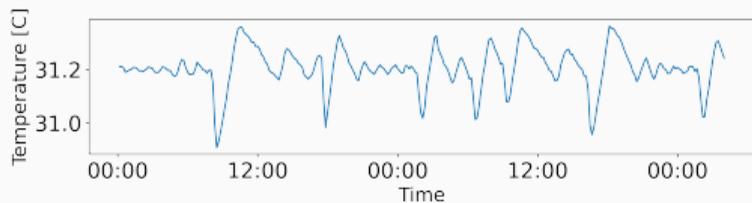


Figure 19: Bassin natation - consigne statique

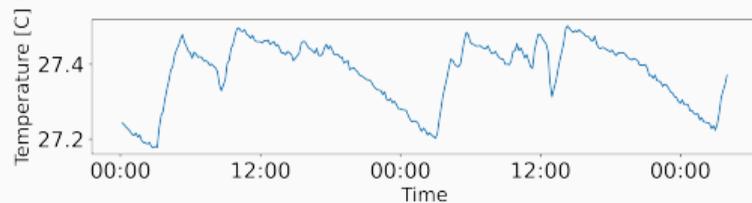


Figure 20: Bassin loisir - système expert

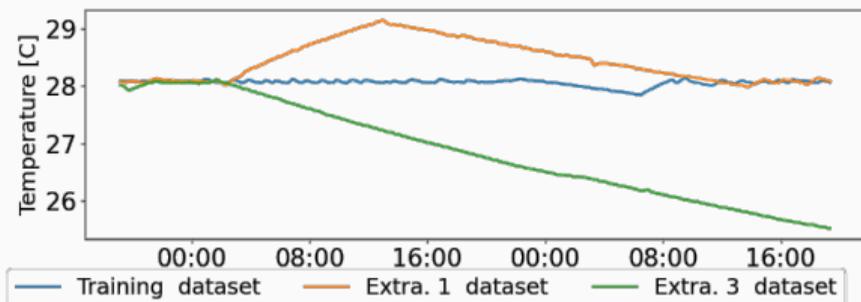


Figure 21: Bassin natation - Comparaison des distributions

# Critères de précision

Critères de prédiction définis sur l'horizon du contrôleur prédictif de 8 heures.

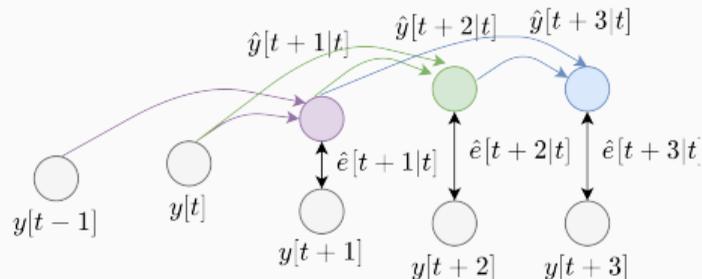
La précision de l'horizon de prédiction:

$$\mathcal{L}(l, J) = \frac{1}{J+l-1} \sum_{i=l}^J \sqrt{\sum_{k=0}^K \frac{\|\hat{y}[i+k|k] - y[i+k]\|_2^2}{K}} \quad (2)$$

Avec  $K$ , le nombre d'exemples du jeu de données.

Critères sélectionnés:

- $\mathcal{L}(1, H), \mathcal{L}(1, H/4), \mathcal{L}(3H/4, H/4)$ , précision globale, précision court terme, précision long terme.
- $\text{Si } \mathcal{L}(1, H)$ , Précision globale pour le  $i^{\text{eme}}$  scénario.



**Figure 22:** Illustration d'une prédiction d'un horizon.

# Modèles sélectionnés

Deux modèles sélectionnés pour initier le banc de test:

Espace d'état linéaire (LSS):

$$\begin{aligned}x[k + 1] &= Ax[k] + Bu[k] \\ y[k] &= Cx[k] + Du[k]\end{aligned}\quad (3)$$

Identifié avec **Matlab** et la méthode des sous espaces.

Modèle auto-regressif non linéaire (NLARX):

$$\begin{aligned}y[k + 1] &= f_{\theta}(y[k], \dots, y[k - n], \\ &u[k], \dots, u[k - n])\end{aligned}\quad (4)$$

Identifié avec python et la librairie **Tensorflow**.

Utilisation d'une fonction de perte sur l'horizon de prédiction pour le modèle non linéaire (méthode **MPC-Relevant-Information** approach [10]).

# Résultats des modèles

	$\mathcal{L}(1, H)$	$\mathcal{L}(1, H/4)$	$\mathcal{L}(3H/4, H)$	S1 $\mathcal{L}(1, H)$	S2 $\mathcal{L}(1, H)$	S3 $\mathcal{L}(1, H)$	S4 $\mathcal{L}(1, H)$
LSS	0.32	0.11	0.52	0.29	<b>0.49</b>	<b>0.88</b>	<b>0.36</b>
NLARX	<b>0.23</b>	<b>0.10</b>	<b>0.31</b>	<b>0.22</b>	0.60	1.24	1.45

Table 1: Résultats des modèles

On observe les éléments suivants:

- NLARX - meilleure précision en interpolation.
- LSS - meilleure précision en extrapolation.
- Résultats équivalents à court terme.

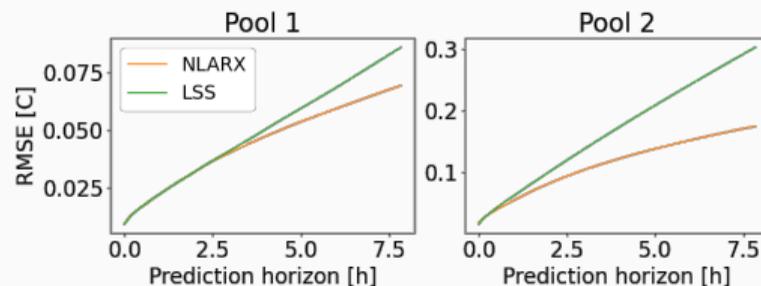
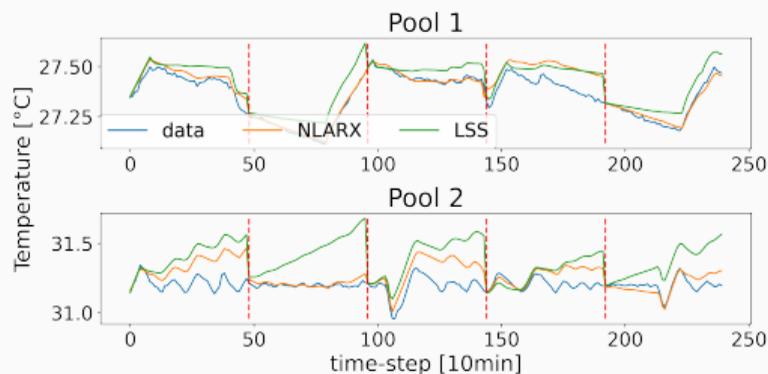
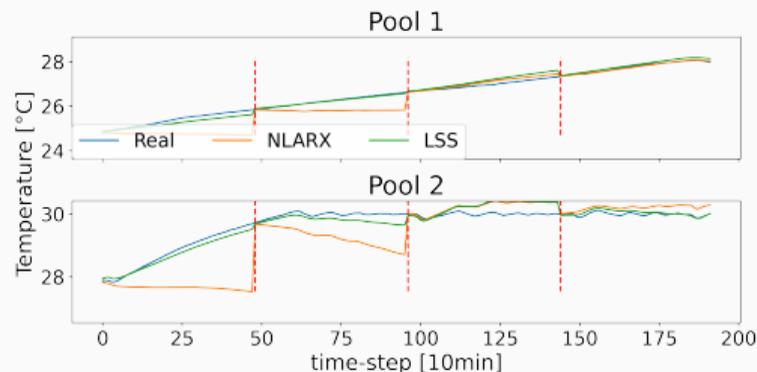


Figure 23: Courbe de précision

Quelques prédictions sur l'horizon de prédiction visé (8 heures):



**Figure 24:** Plusieurs prédictions avec les données de test



**Figure 25:** Plusieurs prédictions en extrapolation

Les éléments suivant sont disponibles:

- Les **données originales** brutes de tout traitement.
- Les **données formalisées** .
- Des scripts en python et matlab, (charger les données, évaluer les modèles).

⇒ Toutes les ressources à l'adresse :

*<https://benchmark-datadriven-sysid.purecontrol.com/>*

## Conclusion et futures directions

---

Points clés:

- Présentation du besoin industriel
- Banc d'essai pour l'identification de système
- Résultats avec deux modèles en boîte noire.

Opportunités de recherche en modélisation:

- Réseaux de neurones informés par la physique (*Physics guided Neural network*) [11].
- Modélisation hybride, résiduelle [12].
- Estimation de l'incertitude épistémique des modèles. [13]

Merci pour votre attention

Question ?

## References

---

- [1] ADEME, *Climat air et energie*, Accédé : 5-3-2023, 2018. [Online]. Available: <https://centre.ademe.fr/sites/default/files/climat-air-energie-chiffres-cles.pdf>.
- [2] *Rénovation énergétique des bâtiments publics : Un enjeu majeur pour les collectivités*, Apr. 2023. [Online]. Available: <https://www.banquedesterritoires.fr/renovation-energetique-des-batiments-publics-un-enjeu-majeur>.
- [3] J. Drgoña, J. Arroyo, I. C. Figueroa, et al., “All you need to know about model predictive control for buildings,” *Annual Reviews in Control*, vol. 50, pp. 190–232, May 2020, ISSN: 13675788. DOI: [10.1016/j.arcontrol.2020.09.001](https://doi.org/10.1016/j.arcontrol.2020.09.001).

- [4] J. P. D. Marín and J. R. García-Cascales, “Dynamic simulation model and empirical validation for estimating thermal energy demand in indoor swimming pools,” *Energy Efficiency*, vol. 13, pp. 955–970, 2020.
- [5] M. F. Haniff, H. Selamat, R. Yusof, S. Buyamin, and F. S. Ismail, “Review of hvac scheduling techniques for buildings towards energy-efficient and cost-effective operations,” *Renewable and Sustainable Energy Reviews*, vol. 27, pp. 94–103, 2013, ISSN: 13640321. DOI: [10.1016/j.rser.2013.06.041](https://doi.org/10.1016/j.rser.2013.06.041).
- [6] Y. Dong, H. Yonghong, and X. Gaohong, “Design of indoor swimming pool water temperature control system based on fuzzy controller and smith predictor,” in *Proceedings of 2011 International Conference on Electronic & Mechanical Engineering and Information Technology*, vol. 9, 2011, pp. 4678–4681. DOI: [10.1109/EMEIT.2011.6024079](https://doi.org/10.1109/EMEIT.2011.6024079).

- [7] Y. Chen, L. K. Norford, H. W. Samuelson, and A. Malkawi, “Optimal control of hvac and window systems for natural ventilation through reinforcement learning,” *Energy and Buildings*, vol. 169, pp. 195–205, 2018, ISSN: 0378-7788. DOI: <https://doi.org/10.1016/j.enbuild.2018.03.051>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0378778818302184>.
- [8] M. Tanaskovic, D. Sturzenegger, R. Smith, and M. Morari, “Robust adaptive model predictive building climate control,” *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 1871–1876, 2017, 20th IFAC World Congress, ISSN: 2405-8963. DOI: <https://doi.org/10.1016/j.ifacol.2017.08.257>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2405896317305700>.

- [9] J. Schoukens and L. Ljung, “Nonlinear system identification: A user-oriented roadmap,” Feb. 2019. [Online]. Available: <http://arxiv.org/abs/1902.00683>.
- [10] R. Gopaluni, R. Patwardhan, and S. Shah, “Mpc relevant identification—tuning the noise model,” *Journal of Process Control*, vol. 14, no. 6, pp. 699–714, 2004, ISSN: 0959-1524.
- [11] F. Djeumou, C. Neary, E. Goubault, S. Putot, and U. Topcu, *Neural networks with physics-informed architectures and constraints for dynamical systems modeling*, 2022. arXiv: [2109.06407 \[cs.LG\]](https://arxiv.org/abs/2109.06407).
- [12] G. Zhang, “Time series forecasting using a hybrid arima and neural network model,” *Neurocomputing*, vol. 50, pp. 159–175, 2003, ISSN: 0925-2312. DOI: [https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0). [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0925231201007020>.

- [13] X. Zhou, H. Liu, F. Pourpanah, T. Zeng, and X. Wang, “A survey on epistemic (model) uncertainty in supervised learning: Recent advances and applications,” *Neurocomputing*, vol. 489, pp. 449–465, 2022, ISSN: 0925-2312. DOI: <https://doi.org/10.1016/j.neucom.2021.10.119>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0925231221019068>.