



Séries temporelles

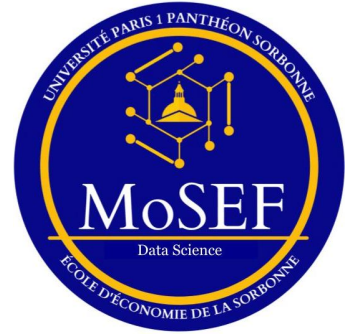
Méthodes ML et DL

Introduction

Intervenant

Guillaume Hochard

1.1. Introduction



De quoi parle-t-on ?

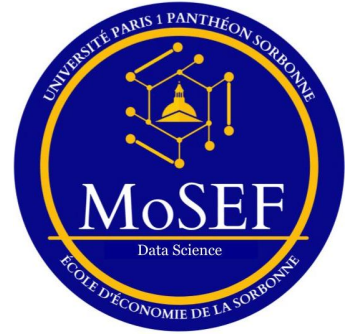
Il s'agit des **méthodes de prévision de séries temporelles**, techniques utilisées pour analyser, manipuler, traiter et prévoir sur ce type de données.

Comme tout sous-domaine de la data science (computer vision, NLP), maîtriser ces méthodes constitue une **compétence à part** qui nécessite une formation particulière !

Les grandes étapes d'un projet de prévision

- L'analyse exploratoire
- La création de variables explicatives
- La modélisation
- L'évaluation de la modélisation

1.2 Objectifs du cours



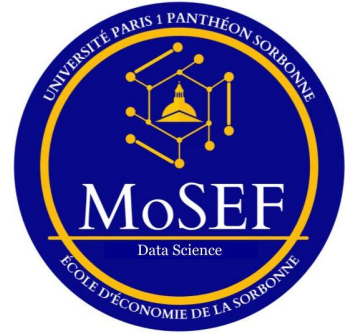
L'objectif de ce cours est de vous donner les **clefs nécessaires** pour traiter un cas d'usage de **prévision de série temporelles**. C'est un module **incontournable** pour tout data scientist voulant aborder ce type de cas d'usage !

Avec ce module, vous pourrez

- Comprendre les différentes **étapes d'un projet de prévision**
- Savoir **évaluer un modèle** de prévision
- Connaitre les **modèles à l'état de l'art** de machine learning et de deep learning pour la prévision

NB : Ce module ne traite pas des méthodes statistiques de prévision. Ces méthodes sont un bon prérequis mais en aucun cas obligatoire à la bonne compréhension du cours.

1.3. Qu'est-ce qu'une série temporelle ?



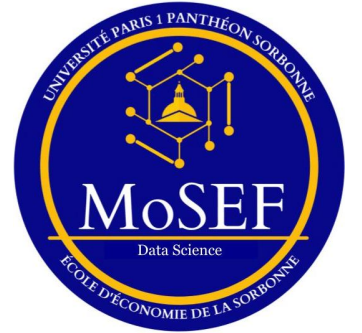
Définition

- Une série temporelle (ou série chronologique) est une suite de valeurs numériques représentant l'évolution d'une quantité spécifique au cours du temps.
- Les données peuvent être acquises ponctuellement, **ponctuellement**, à des **intervalles de temps réguliers** ou à des **intervalles de temps irréguliers**.





Une série temporelle peut être :

- **Univariée** (on cherche à prédire la série à partir de son historique et de régresseurs externes)
- **Multivariée** (on peut soit prédire une seule composante de cette série à partir des autres composantes ou prédire l'ensemble des composantes)

1.4. Cas d'usages



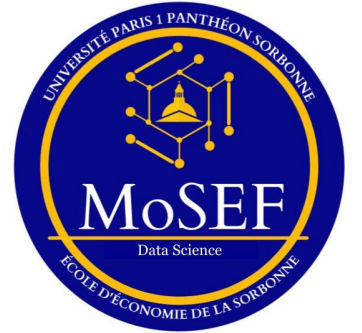
Les secteurs d'activité regroupant le plus de cas d'usages sont :

- Transports 
- Supply chain  
- Énergie 
- Retail 
- Santé 

Quelques exemples de cas d'usages les plus courants :

- Prédiction de la demande
- Prédiction de consommation d'énergie
- Prédiction d'affluence en gare, en magasin
- ...

1.5. Un peu d'histoire...



1676

1850

1927

1969

1970



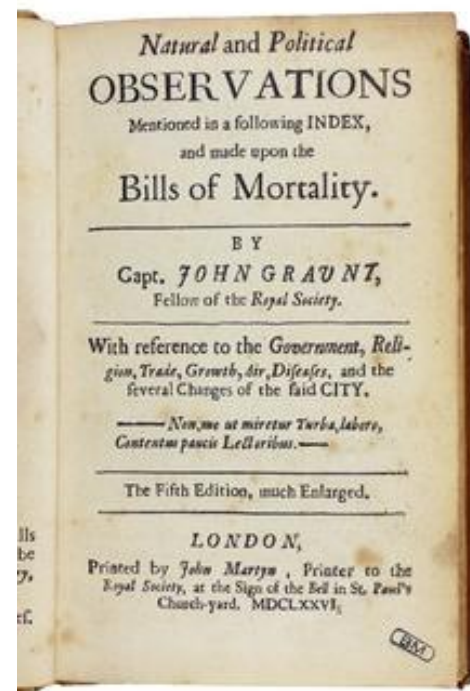
John Graunt

Robert FitzRoy

Udny Yule

Bates & Granger

Georges Box



The Combination of Forecasts

J. M. BATES and C. W. J. GRANGER

Department of Economics, University of Nottingham

Two separate sets of forecasts of airline passenger data have been combined to form a composite set of forecasts. The main conclusion is that the composite set of forecasts can yield lower mean-square error than either of the original forecasts. Past errors of each of the original forecasts are used to determine the weights to attach to these two original forecasts in forming the combined forecasts, and different methods of deriving these weights are examined.



1.6. Modélisation vs Prédiction



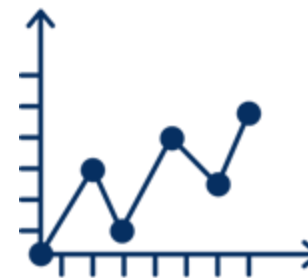
Deux procédés distincts, qui répondent à des enjeux différents



Expliquer

Modélisation

Comprendre le passé
Décrire la série temporelle



Prévoir

Prédiction

Prédire les futures valeurs
de la série



Décider

Prédiction probabiliste

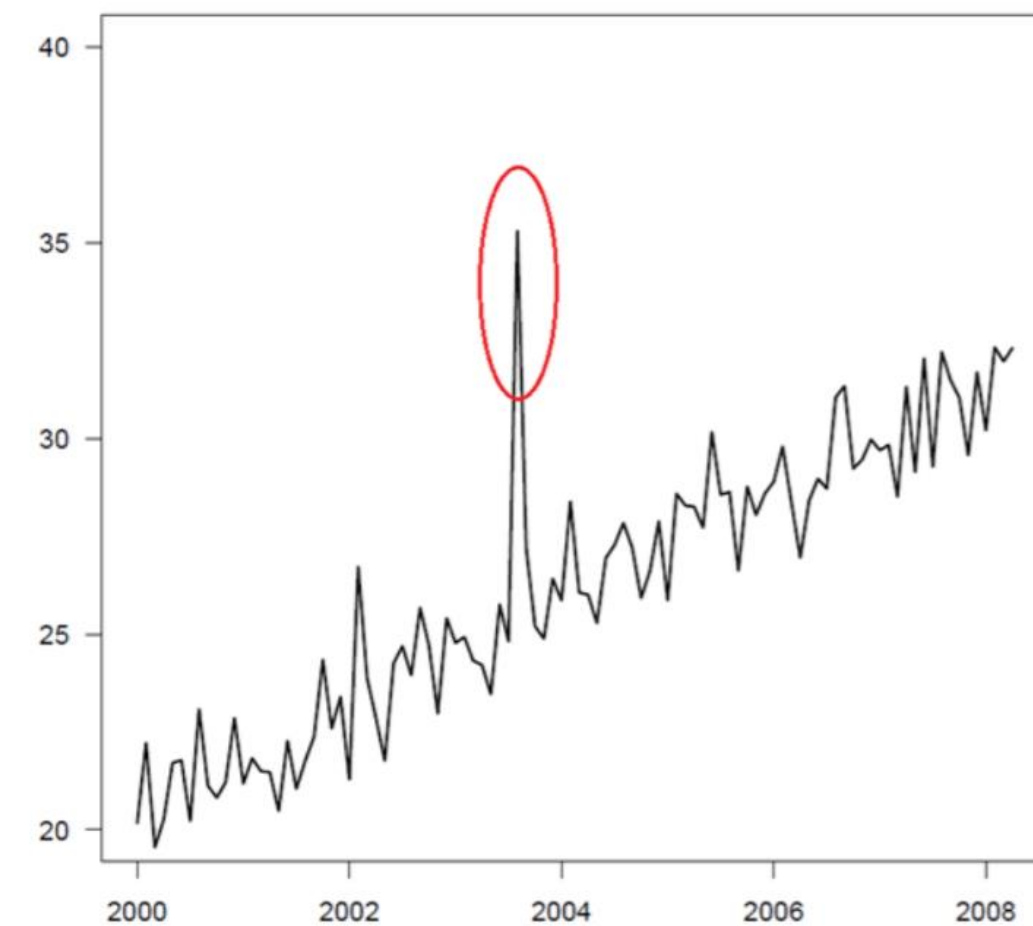
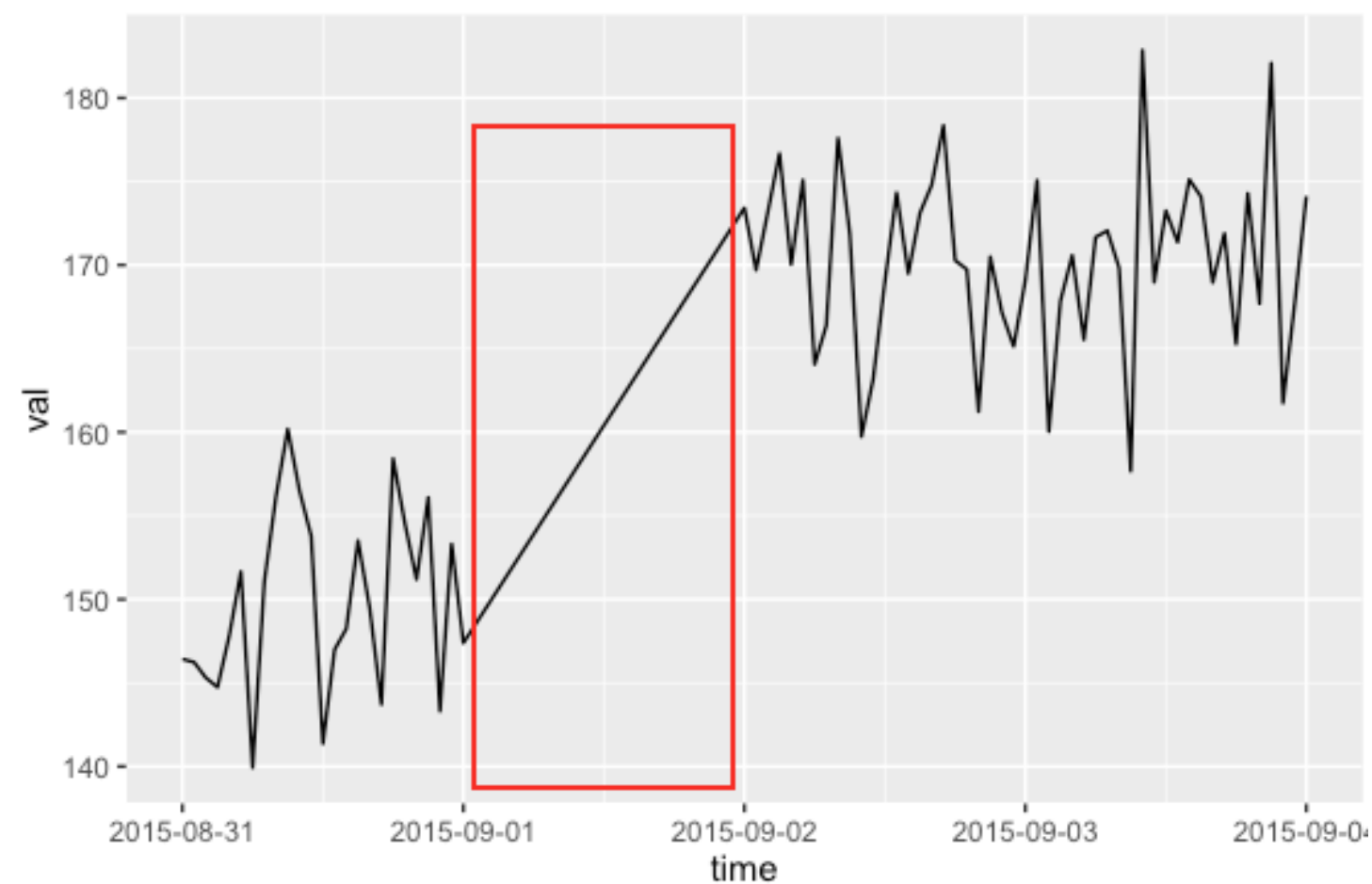
Optimiser la prise de décision
en gérant l'incertitude

1.7. Data vizualisation 1/2



La visualisation de données est incontournable en séries temporelles

- Cela permet de comprendre la structure de la donnée
- Par exemple : Les données sont-elles en qualité, y'a-t-il des outliers ?
- Ou encore : avons-nous affaire à de la demande régulièrement ou intermittente?

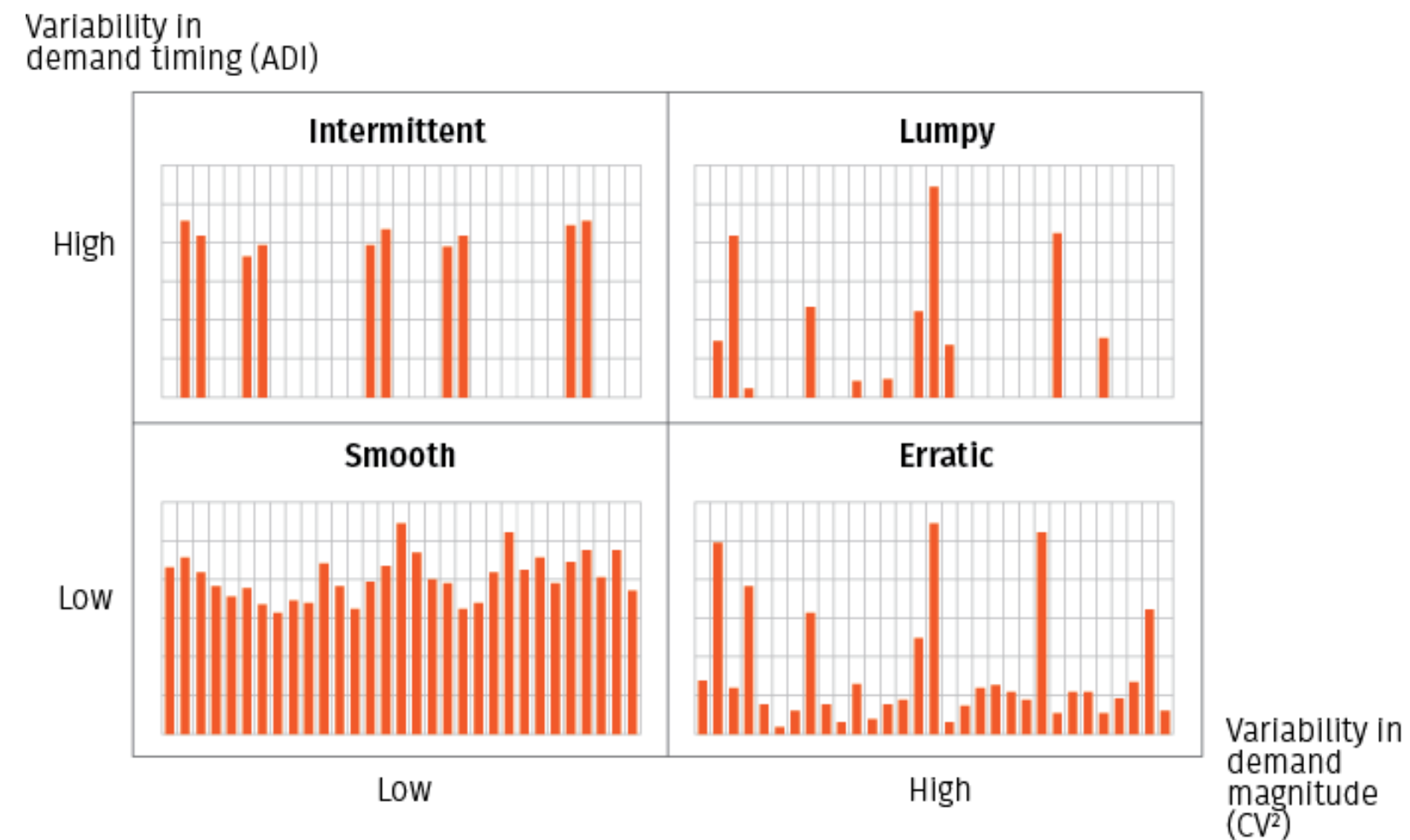


1.7. Data visualisation 1/2



La visualisation de données est incontournable en séries temporelles

- Cela permet de comprendre la structure de la donnée
- Par exemple : Les données sont-elles en qualité, y'a-t-il des outliers ?
- Ou encore : avons-nous affaire à de la demande régulièrement ou intermittente?

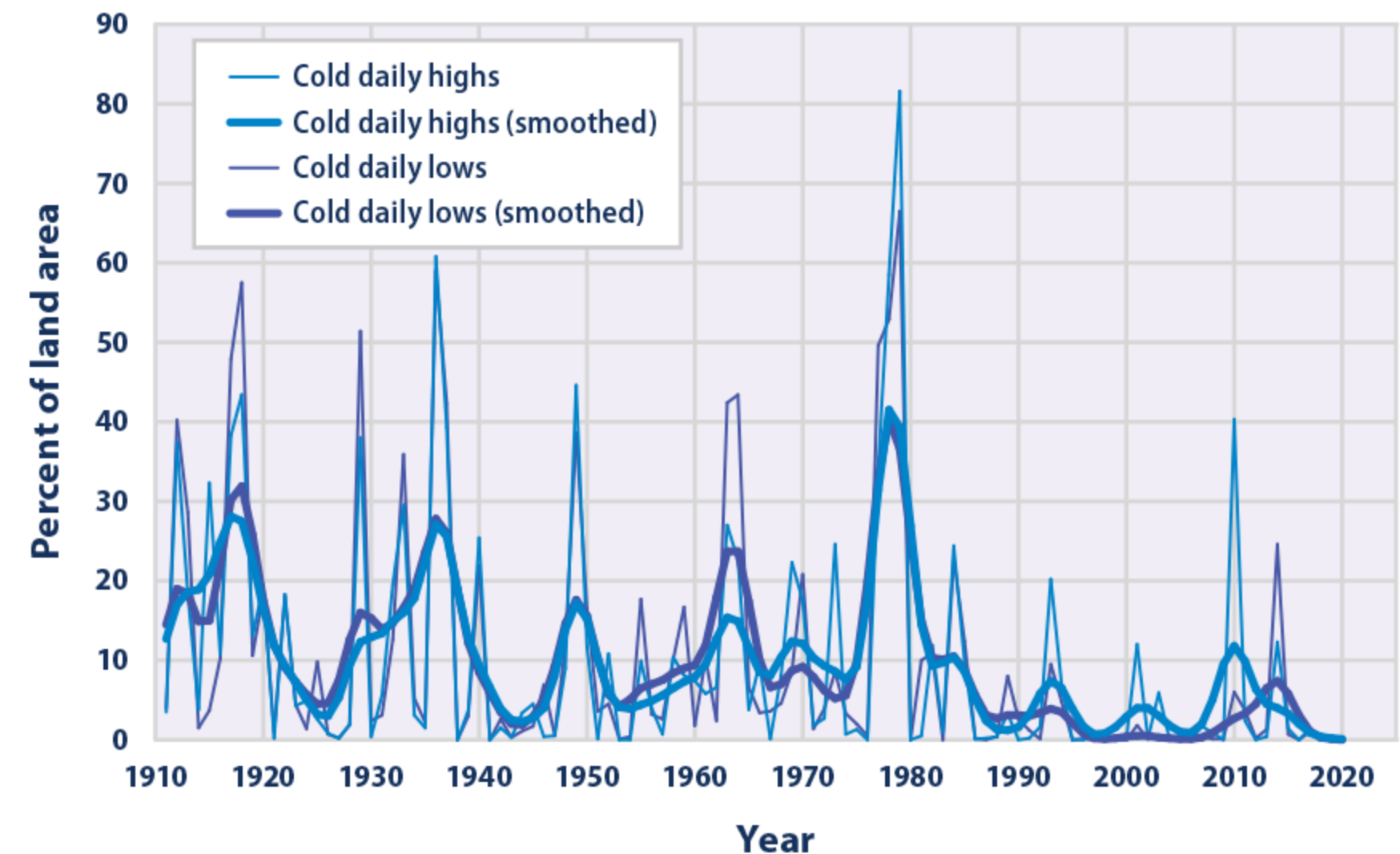


1.7. Data vizualisation 2/2

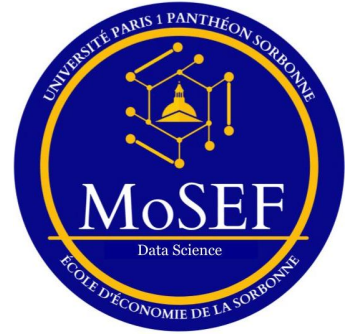


Parmi les représentations les plus utiles, on peut citer :

- Line plots

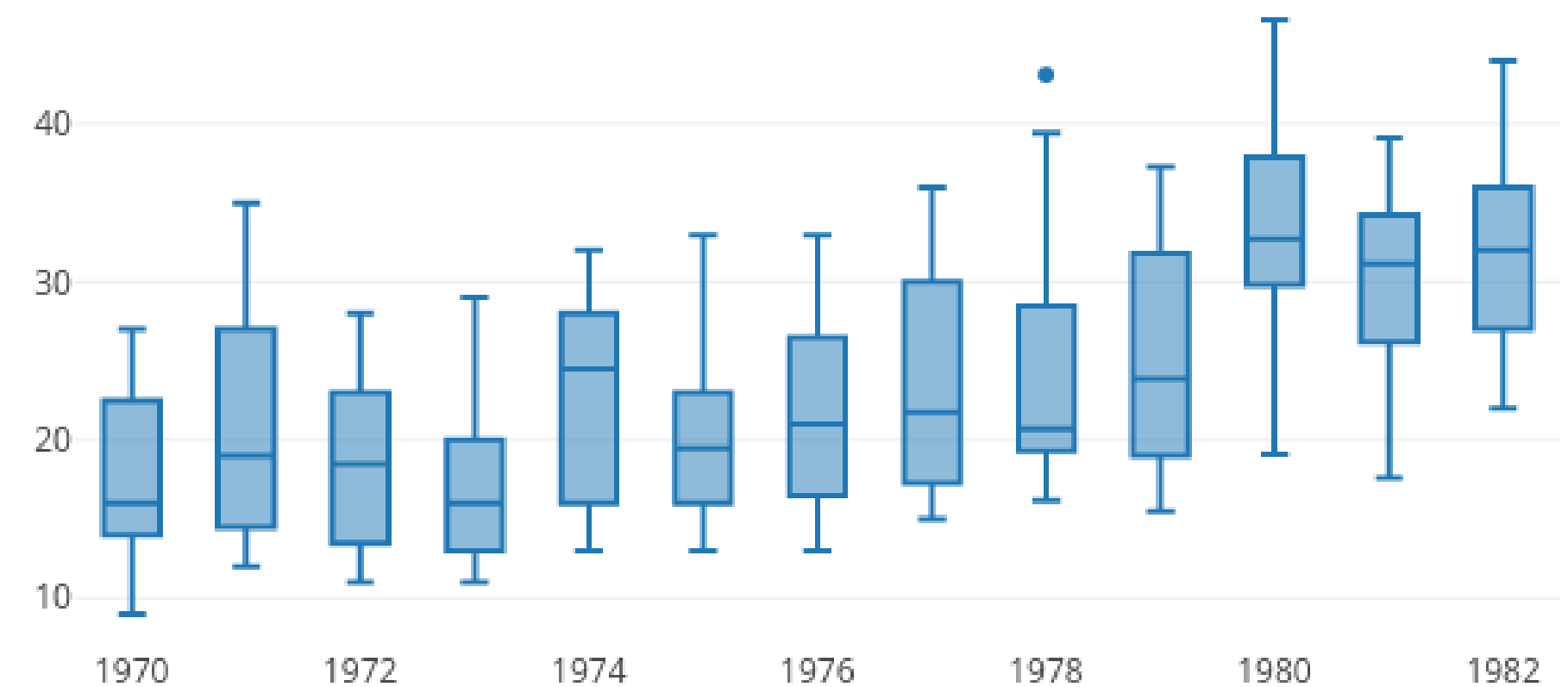


1.7. Data vizualisation 2/2



Parmi les représentations les plus utiles, on peut citer :

- Line plots
- Box plots

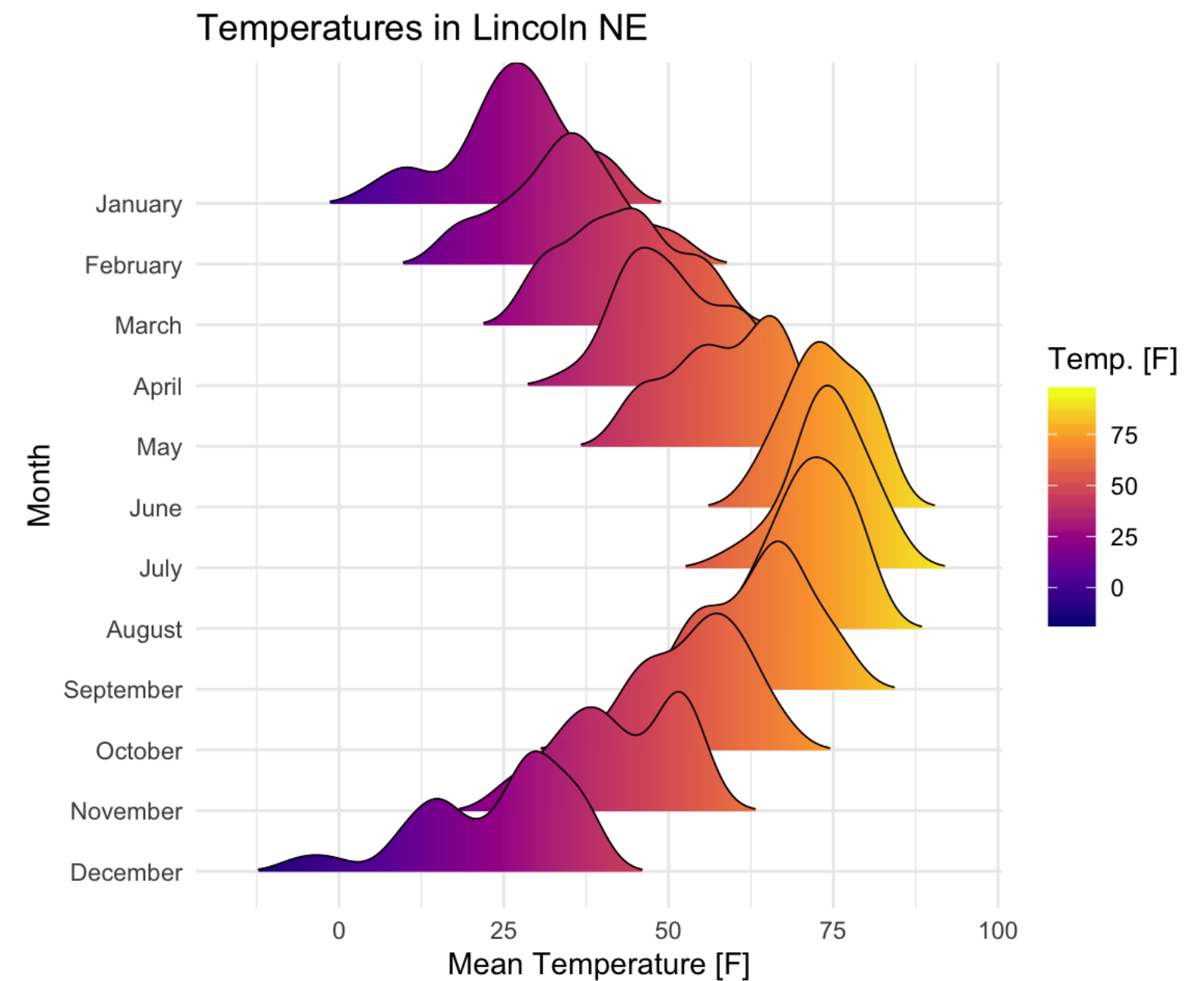


1.7. Data vizualisation 2/2

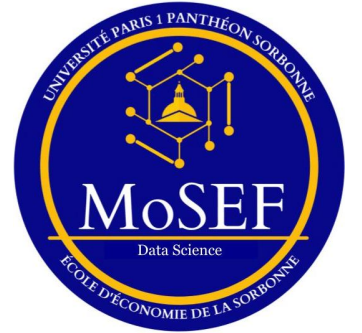


Parmi les représentations les plus utiles, on peut citer :

- Line plots
- Box plots
- Histogrammes ou density plots

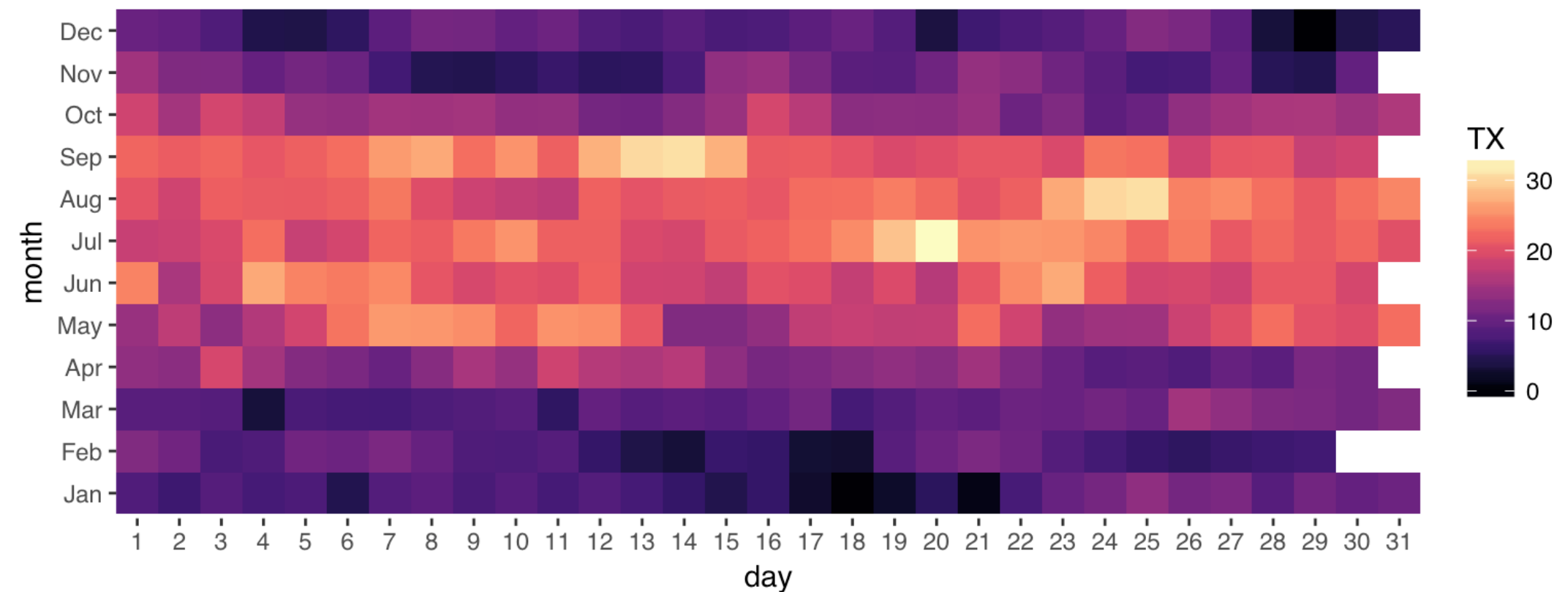


1.7. Data vizualisation 2/2

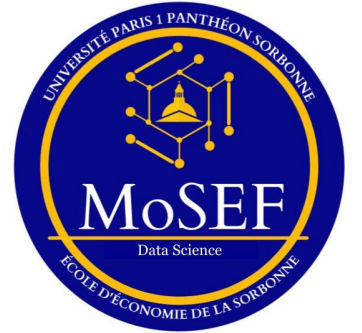


Parmi les représentations les plus utiles, on peut citer :

- Line plots
- Box plots
- Histogrammes ou density plots
- Heatmaps

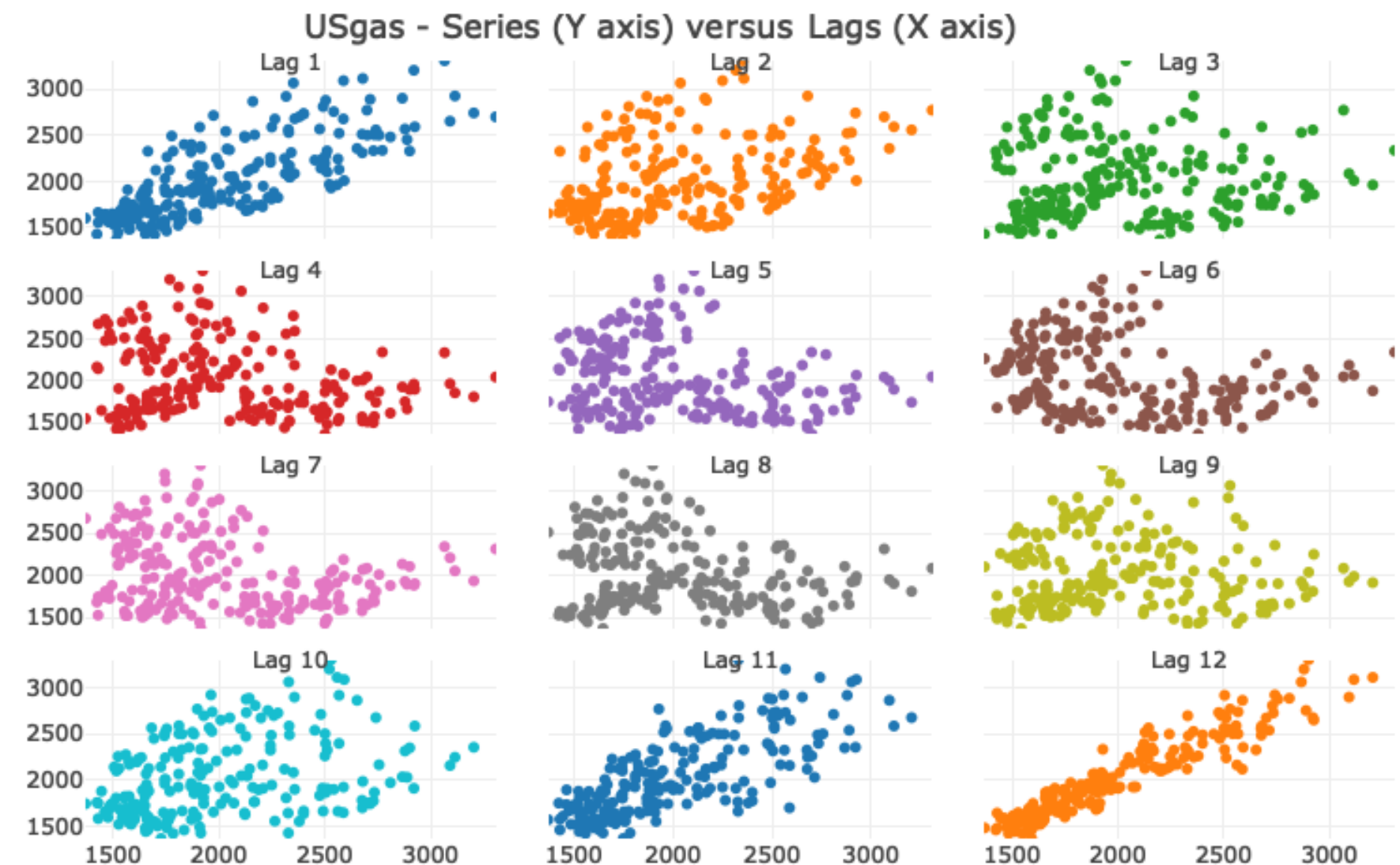


1.7. Data vizualisation 2/2



Parmi les représentations les plus utiles, on peut citer :

- Line plots
- Box plots
- Histogrammes ou density plots
- Heatmaps
- Lags scatter plots



1.8. Time to practice



TP1 Dataviz

Rendez-vous sans plus attendre sur Python pour la mise en pratique !





Séries temporelles

Méthodes ML et DL

Introduction à la prévision

Intervenant

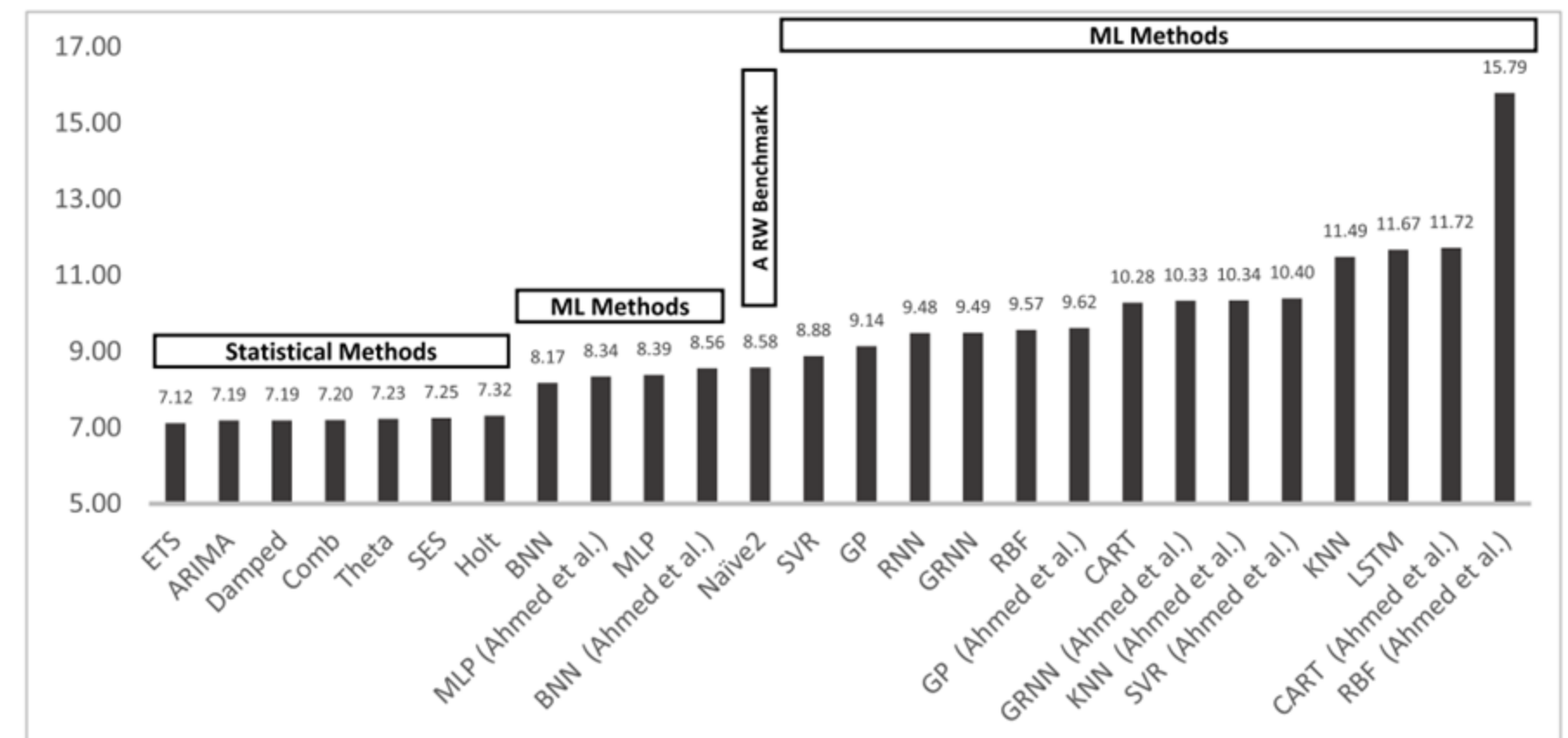
Guillaume Hochard

2.1. Etat de l'art 1/2



Un changement de paradigme récent

- Opposition historique entre les prévisions statistiques et les prévisions ML
- Domination historique des méthodes statistiques (sans régresseurs externes)



(Makridakis et al., 2018)

2.1. Etat de l'art 1/2



Un changement de paradigme récent

- Opposition historique entre les prévisions statistiques et les prévisions ML
- Domination historique des méthodes statistiques (sans régresseurs externes)

- En 2018, les résultats de la compétition de forecasting M4 bouleversent ce paradigme

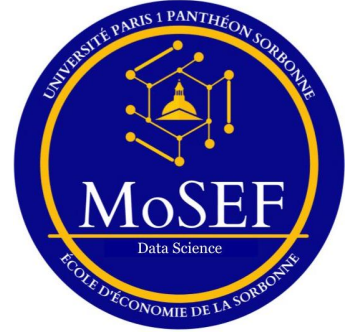
- Les approches gagnantes sont des approches hybrides

- Le grand vainqueur est ES-RNN : réseau de neurones récurrent qui prend en paramètre des coefficient de lissage exponentiel (ES)

Type	Author(s)	Affiliation	Rank ^a	% improvement of method over the benchmark	
Hybrid	Smyl, S.	Uber Technologies	1	9.4%	8.6%
Combination	Montero-Manso, P., Talagala, T., Hyndman, R. J. & Athanasopoulos, G.	University of A Coruña & Monash University	2	6.6%	6.7%
Combination	Pawlikowski, M., Chorowska, A. & Yanchuk, O.	ProLogistica Soft	3	5.7%	6.3%
Combination	Jaganathan, S. & Prakash, P.	Individual	4	6.8%	6.2%
Combination	Fiorucci, J. A. & Louzada, F.	University of Brasilia & University of São Paulo	5	5.7%	6.1%
Combination	Petropoulos, F. & Svetunkov, I.	University of Bath & Lancaster University	6	5.3%	5.6%
Combination	Shaub, D.	Harvard Extension School	7	4.3%	4.2%
Statistical	Legaki, N. Z. & Koutsouri, K.	National Technical University of Athens	8	4.5%	4.1%
Combination	Doornik, J., Castle, J. & Hendry, D.	University of Oxford	9	5.0%	3.7%
Combination	Pedregal, D.J., Trapero, J. R., Villegas, M. A. & Madrigal, J. J.	University of Castilla-La Mancha	10	3.5%	3.2%
Statistical	Spiliotis, E. & Assimakopoulos, V.	National Technical University of Athens	11	3.2%	2.7%

M4 competition results (October 2018)

2.1. Etat de l'art 1/2

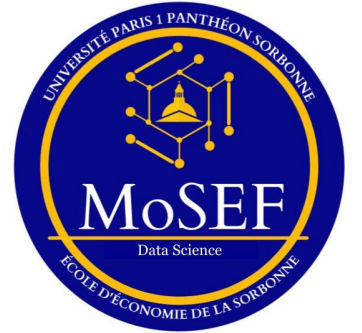


Un changement de paradigme récent

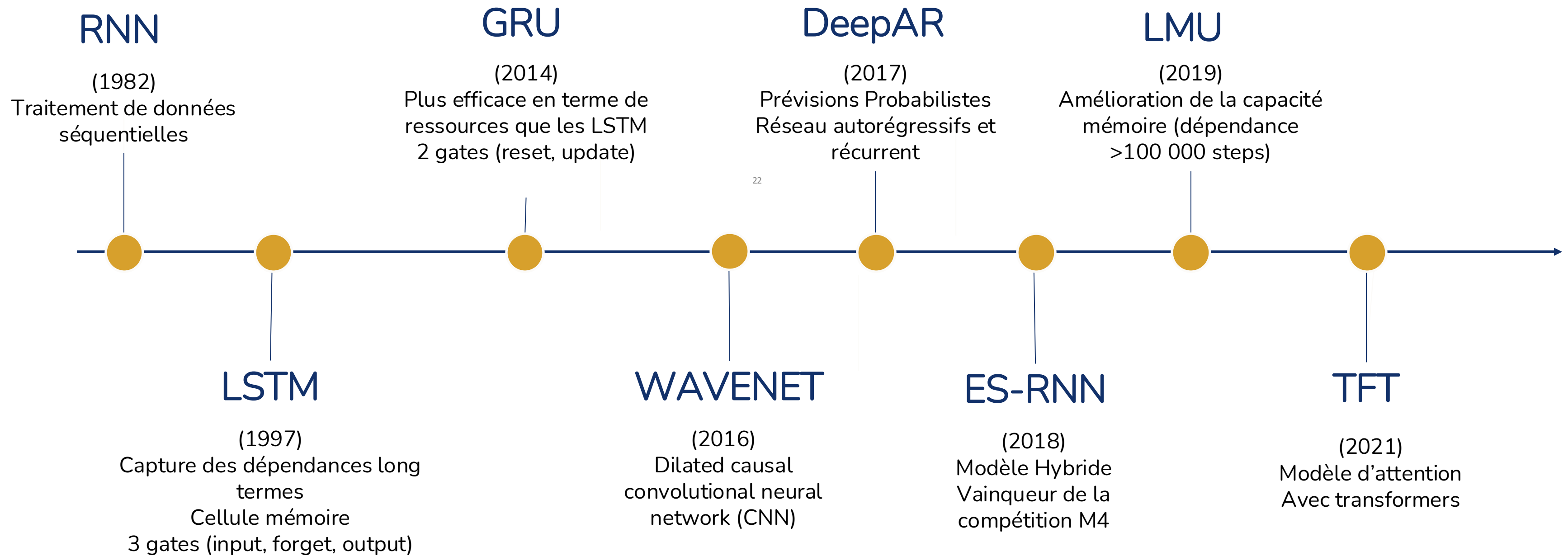
- Opposition historique entre les prévisions statistiques et les prévisions ML
- **Domination historique des méthodes statistiques** (sans régresseurs externes)
- En 2018, les résultats de la compétition de forecasting M4 bouleversent ce paradigme
- Les approches gagnantes sont des approches hybrides
- Le grand vainqueur est ES-RNN : réseau de neurones récurrent qui prend en paramètre des coefficient de lissage exponentiel (ES)
- En 2020, nouvelle compétition : M5, sur des données retail (Walmart)
- LightGBM (Boosting) gagne la compétition



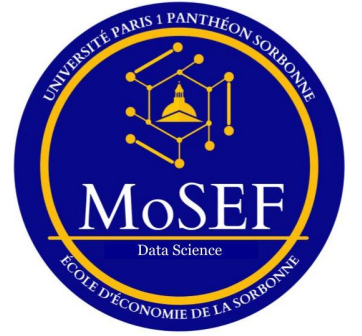
2.1. Etat de l'art 2/2



Une recherche très active en Deep Learning



2.2. Concepts fondamentaux 1/2



Concepts importants pour cadrer un projet de prévision

Définir ce qui est nécessaire et disponible pour réaliser une prédiction

- **Entrées** : données historiques fournies au modèle pour réaliser une prévision
- **Sorties** : prévision pour pas de temps futur non inclus dans les données d'entrée

Identifier et localiser les variables clefs / enrichir les données

- **Données endogènes** : données d'entrées qui sont influencées par d'autres données d'entrée dans le système (ex : capteur de température / capteur de pression)
- **Données exogènes** : données d'entrée qui ne sont pas influencées par d'autres variables du système (ex : météo)

Préciser la cible souhaitée

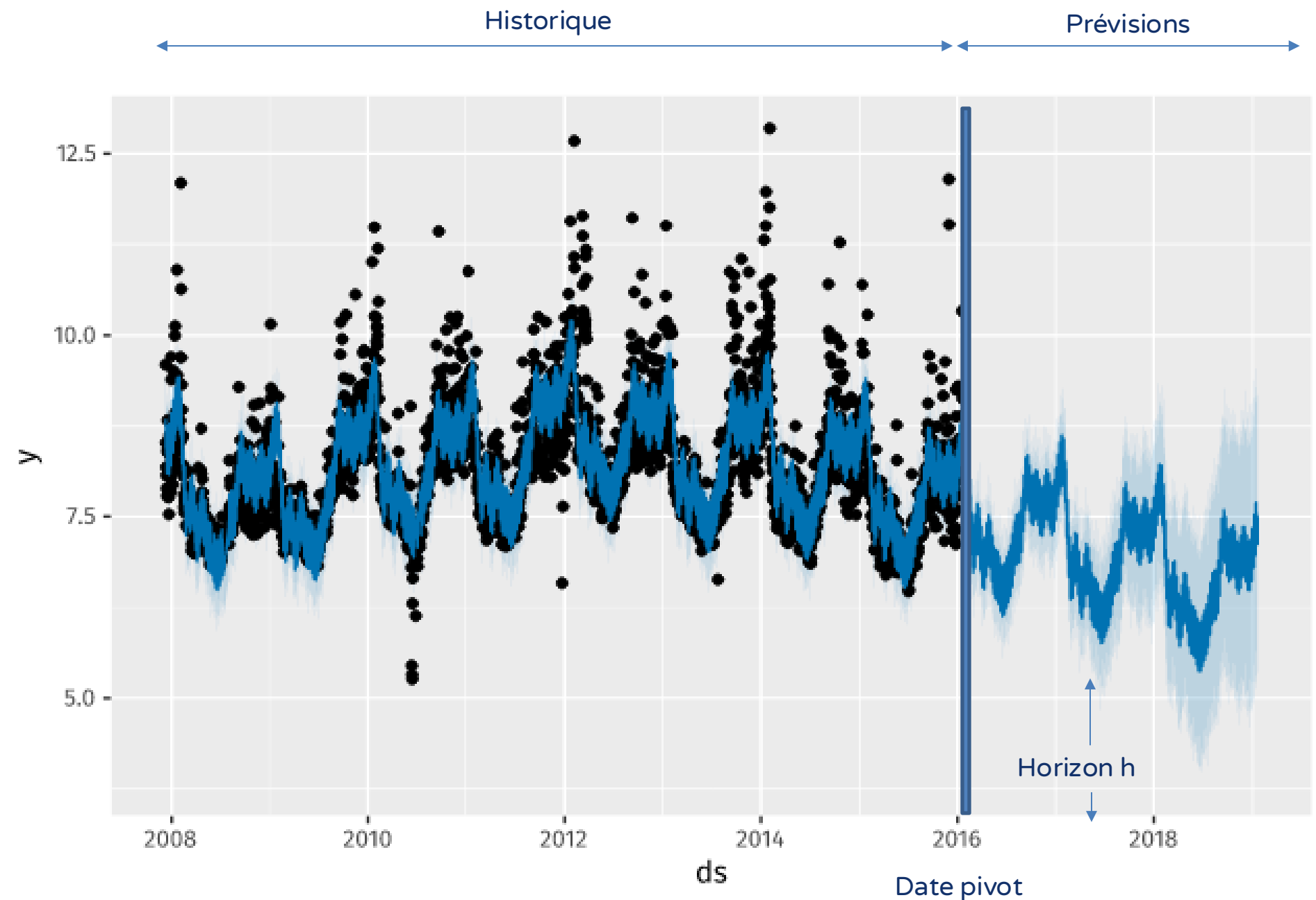
- **Entrées / sorties univariées ou multivariées** : doit-on prédire un scalaire ou un vecteur?
- **Single-step ou multi-step** : prévision requise pour un ou plusieurs prochains pas de temps

2.2. Concepts fondamentaux 2/2

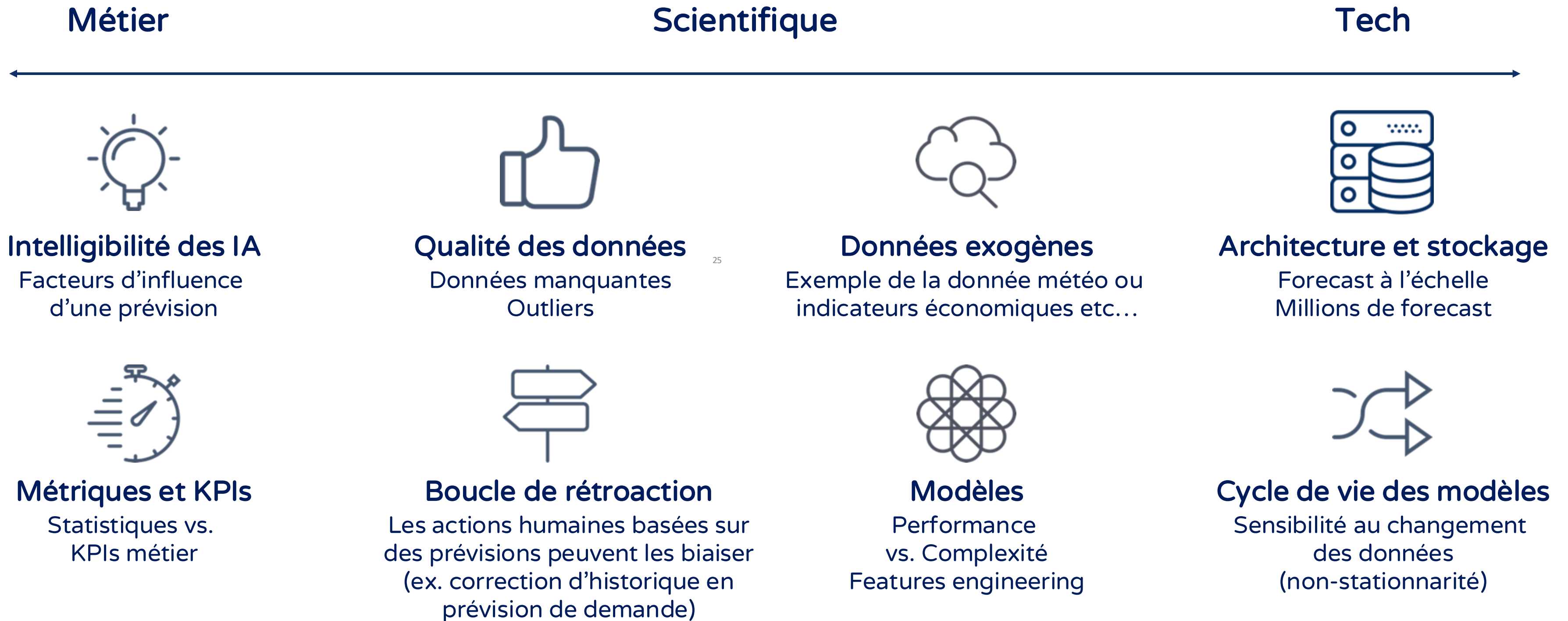


Quelques définitions

- **Historique** : données historiques disponibles qui *seront préparées* pour être fournies au modèle pour réaliser une prévision (l'historique disponible n'est pas nécessairement le jeu d'entraînement du modèle)
- **Date pivot** : date séparant les données historiques des horizons futurs à prévoir (aussi appelée *cutoff*)
- **Horizon** : pas de temps après la date pivot où est réalisée une prévision (aussi appelé *leadtime*)



2.3. Challenges majeurs d'un projet de prévision



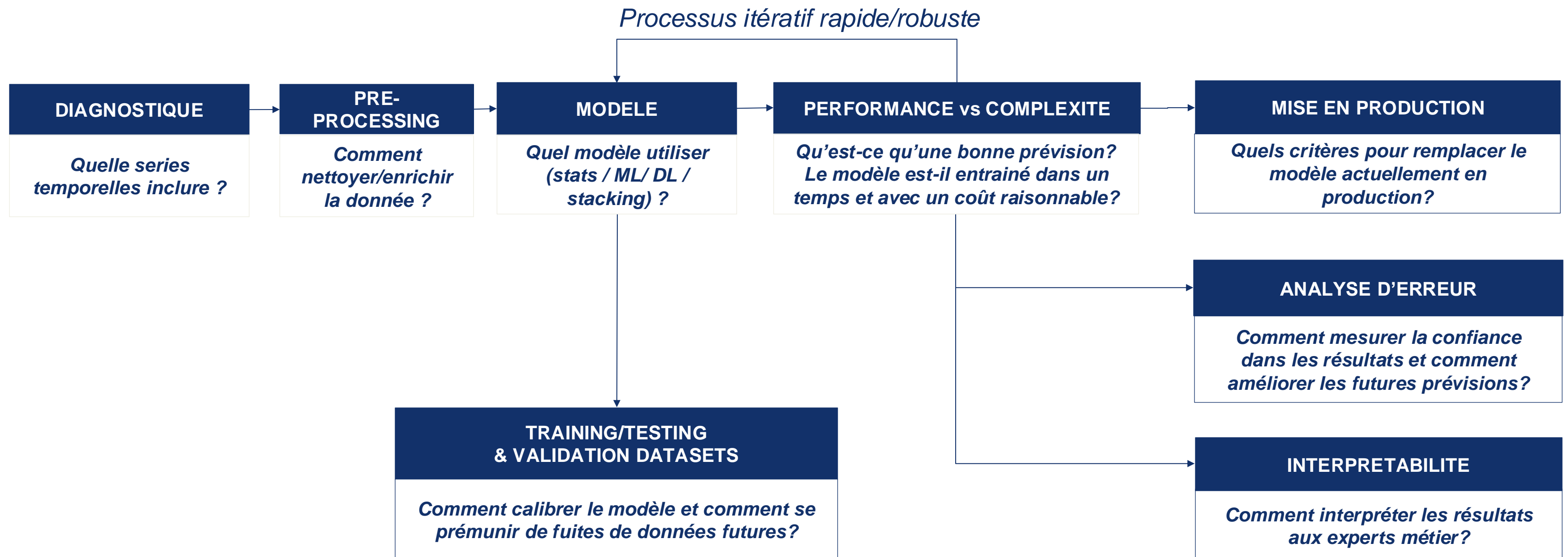
2.4. Méthodologie d'un projet de prévision 1/2



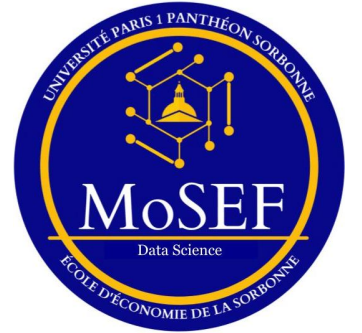
1. Cadrage

2. Modélisation

3. Exploitation



2.4. Méthodologie d'un projet de prévision 2/2



Autre cadre théorique reproductible pour aborder les questions de prévision

1. Définir l'objectif du forecast : que prédit-on ?

- Cadrage du projet : granularité temporelle, granularité des produits, leadtime

2. Evaluer le coût des erreurs de forecast

- Quelles sont les conséquences financières d'une sous-estimation ou d'une surestimation ?

3. Préciser judicieusement la fonction de perte

- Ai-je besoin d'une fonction de coût symétrique ou asymétrique : qu'est-ce qui compte vraiment en termes d'objectifs commerciaux ?

4. Préciser l'horizon de prévision : jusqu'où faut-il prévoir ?

- La précision varie selon l'horizon : plus l'horizon est éloigné, plus le degré de confiance dans la prévision est faible

5. Choix des features

- Cartographier les données disponibles (internes/externes) qui pourraient être utilisées comme prédicteurs et décrire leur source, leur qualité, leur fréquence et leur disponibilité historique. Limiter le nombre de prédicteurs et évaluer leur impact sur la précision des prévisions

6. Choisir et comparer les méthodes de prévision

- Sélectionner les techniques en fonction de l'objectif et les étalonner

7. Présenter à la fois les prévisions et leurs incertitudes

- Les intervalles de confiance des prévisions sont aussi importants que la valeur ponctuelle prévue

8. Choisir la métrique d'évaluation

- Une métrique statistique est essentielle pour les modèles de benchmarking. A chaque cas d'usage sa métrique. L'intégration des informations commerciales dans la métrique est incontournable

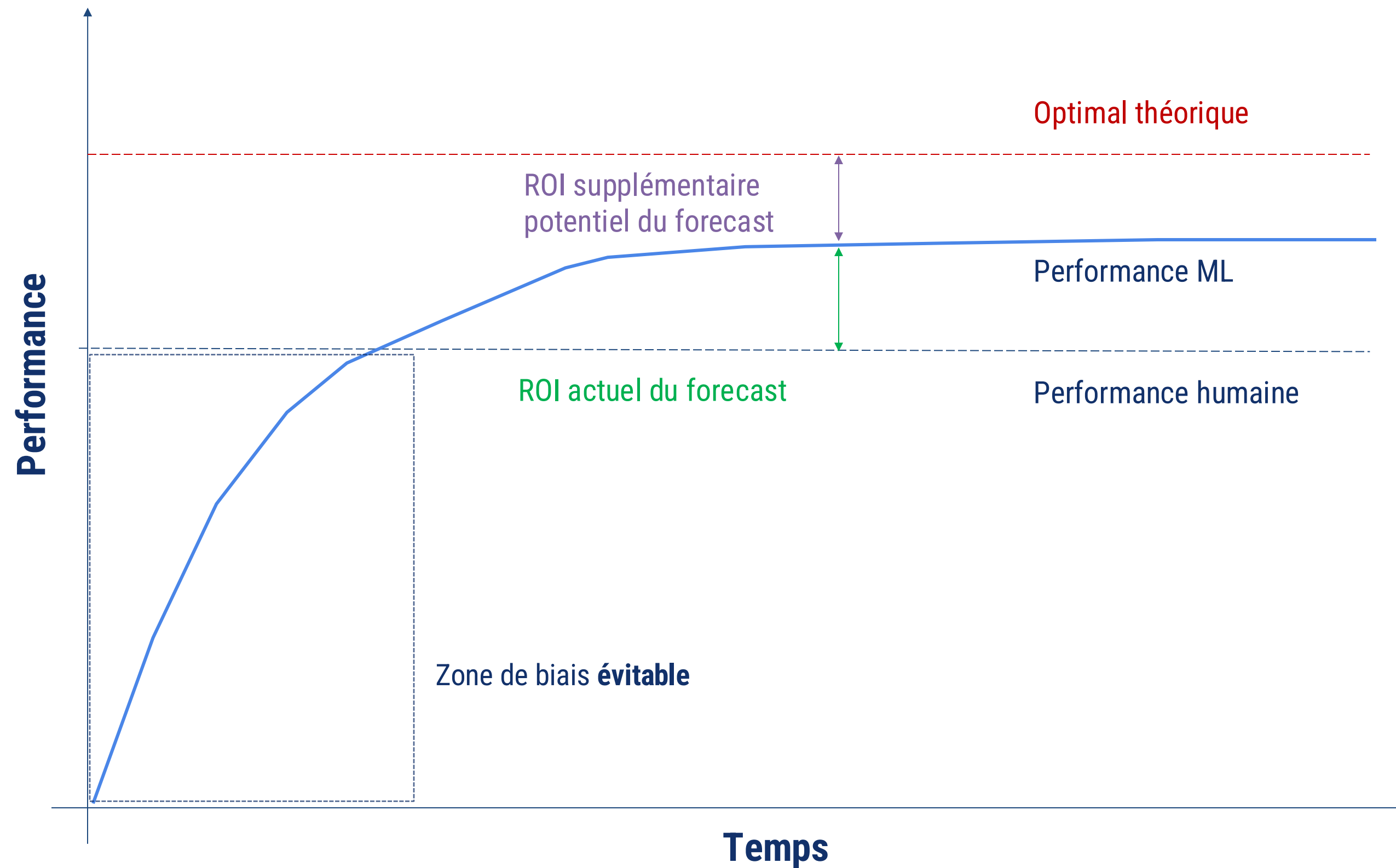
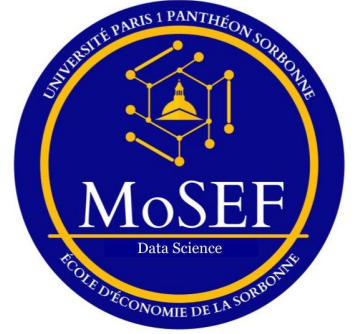
9. Effectuer des tests Out-of-sample

- Une méthodologie solide permet d'obtenir des résultats solides, fiables et plus précis. Les tests hors échantillon améliorent la robustesse des modèles

10. Itérer et définir un ROI cible

- Toujours essayer d'améliorer et de surveiller votre modèle. Définir les paramètres de dérive et les KPI. Définir des critères pour remplacer un modèle en production

2.5. Mesurer le ROI d'un projet



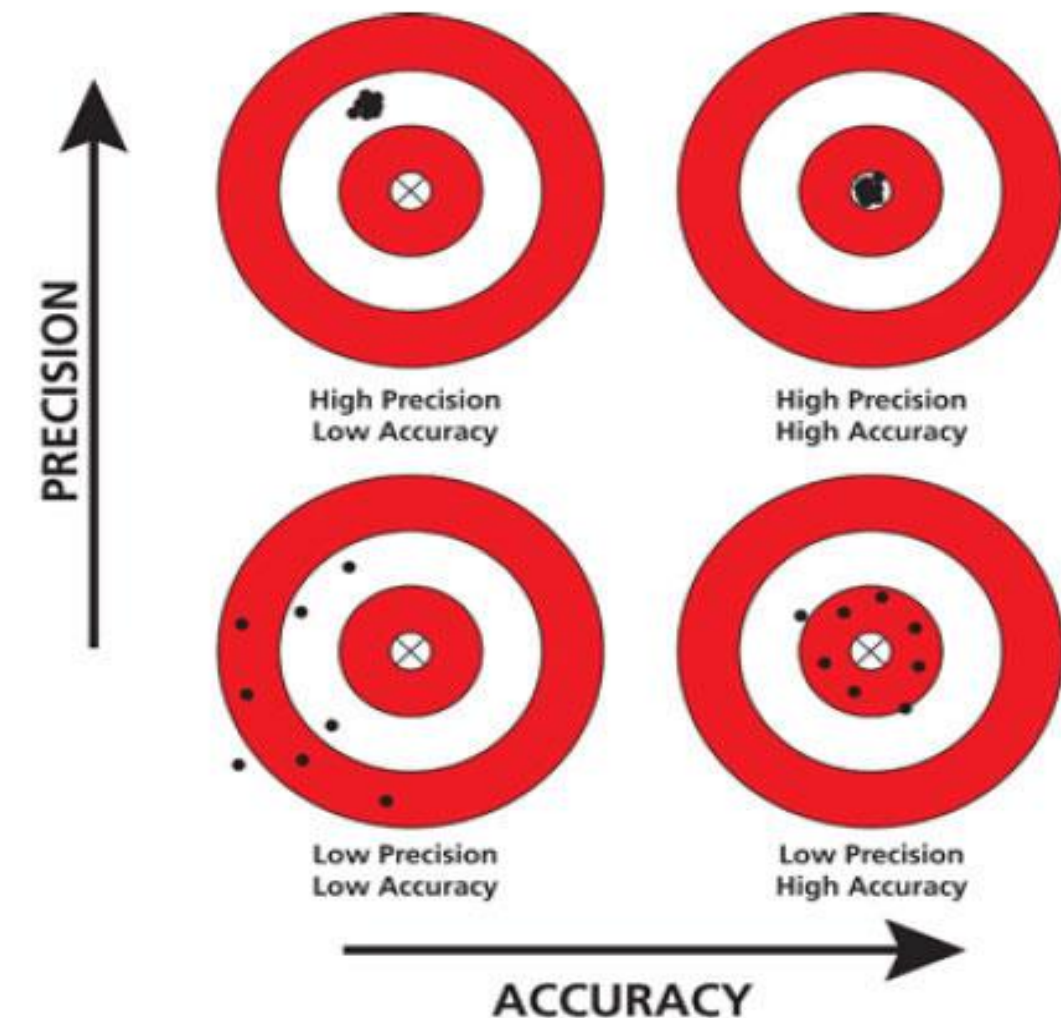
2.6. Métriques d'évaluation 1/7

La performance d'un modèle de prévision doit prendre en compte deux concepts

- La précision et l'accuracy
- Historiquement, on peut apprécier l'importance de ce concept



Bataille d'Azincourt, Paris, [BnF](#), département des Manuscrits.

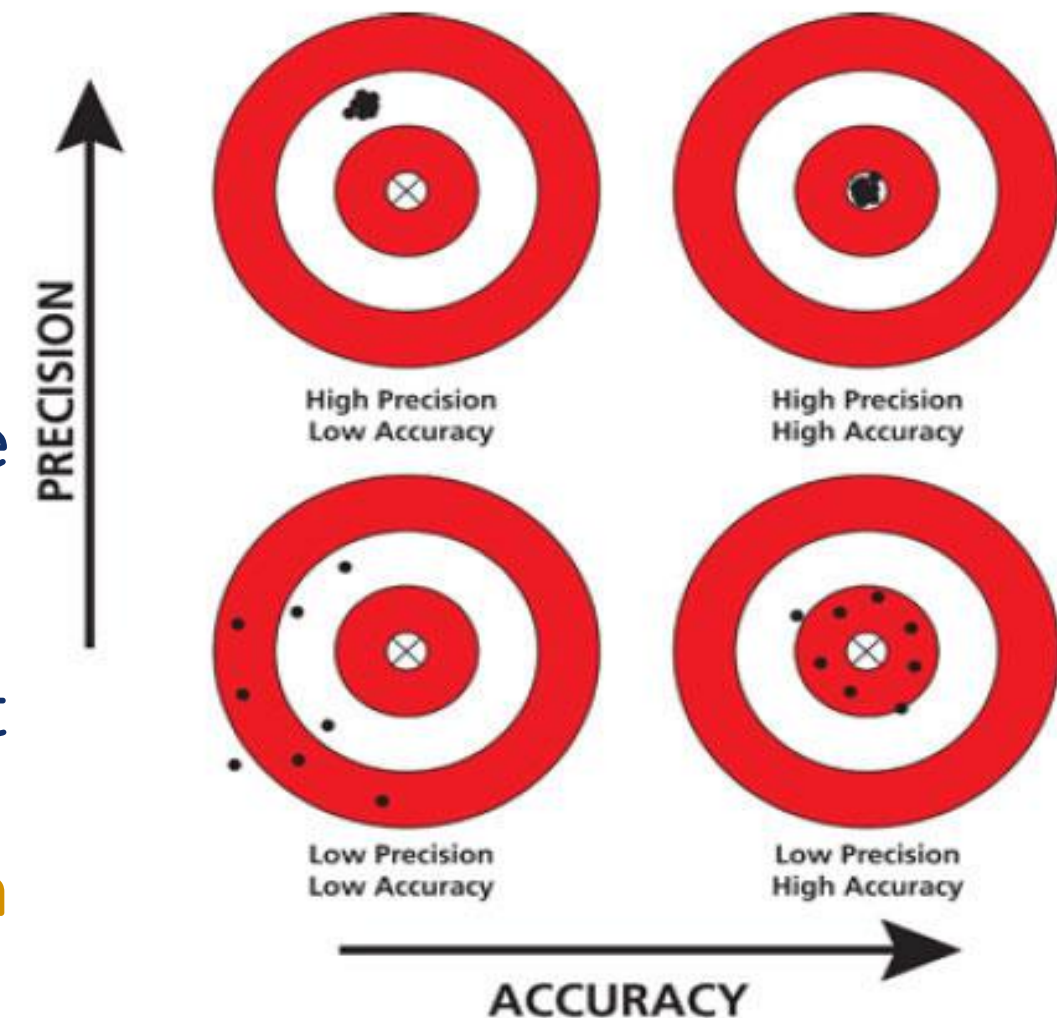


2.6. Métriques d'évaluation 1/7



La performance d'un modèle de prévision doit prendre en compte deux concepts

- La précision et l'accuracy
- Historiquement, on peut apprécier l'importance de ce concept
- La précision mesure la **dispersion des mesures** (variance)
- L'accuracy mesure quand à elle **l'écart au centre de la cible** (biais)
- La **Forecast accuracy** (biais) **n'est pas la seule métrique à considérer**
- La **Forecast precision** (variance) correspondant à un intervalle de confiance doit être aussi prise en compte.
- La plupart des métriques sont des indicateurs de bias (forecast ponctuel => mesure de biais)
- La précision est une métrique particulièrement importante **pour un forecast probabiliste**



2.6. Métriques d'évaluation 2/7



Métriques de performance absolue
Elles permettent une lecture directe de l'erreur du modèle

Abréviation	Nom	Formule	Commentaire
	Erreur	$\hat{y}_i - y_i$	Erreur positive en cas de sur-estimation.
	Biais	$\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)$	Erreur moyenne
MSE	Mean Square Error	$\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2$	Régression standard, sensible aux outliers
RMSE	Root Mean Square Error	$\sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2}$	Interprétable et comparable à la cible
MLSE	Mean Square Log Error	$\frac{1}{N} \sum_{i=1}^N (\ln(1 + \hat{y}_i) - \ln(1 + y_i))^2$	Intéressant dans certains cas (valeurs extrêmes)

2.6. Métriques d'évaluation 3/7



Métriques de performance relative
Elles permettent une interprétation plus intelligible de la performance du modèle

Abréviation	Nom	Formule	Commentaire
MAPE	Mean Average Percentage Error	$\frac{100}{N} \sum_{i=1}^N \frac{ \hat{y}_i - y_i }{y_i}$	Attention à cette métrique !
WMAPE	Weighted Mean Average Percentage Error	$100 * \frac{\sum_{i=1}^N \hat{y}_i - y_i }{\sum_{i=1}^N y_i }$	Meilleure alternative à la MAPE, pondération par la cible
MAE%	Scaled Mean Absolute Error	$\frac{\sum_{i=1}^N \hat{y}_i - y_i }{\sum_{i=1}^N y_i}$	Mesure relative de la MAE en divisant par la quantité moyenne
RMSE%	Scaled Root Mean Square Error	$\frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^N y_i}}$	Mesure relative de la RMSE en divisant par la quantité moyenne

2.6. Métriques d'évaluation 4/7



La MAPE : un faux ami à éviter

Avantages de cette métrique

- Exprimée en %, indépendant de l'échelle
- Facile à expliquer aux parties prenantes

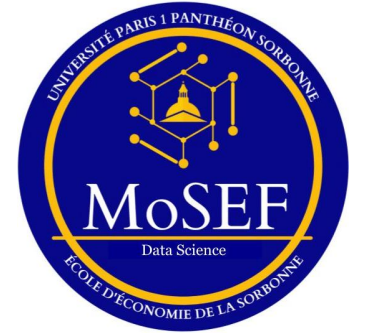
Désavantages de cette métrique

- Valeurs indéfinies lorsque les valeurs réelles sont nulles et valeurs extrêmes lorsque les valeurs réelles sont très proches de zéro.
- Asymétrie et pénalise davantage les erreurs positives* (lorsque les prévisions sont plus élevées que les valeurs réelles) que les erreurs négatives. Ainsi, on favorisera les modèles qui sous-estiment plutôt que ceux qui sur-estiment. (*voir définition erreur au slide 2/7)
- Suppose que l'unité de mesure de la variable a une valeur zéro significative.
- N'est pas partout différentiable, ce qui peut poser des problèmes lorsqu'on l'utilise comme critère d'optimisation.



Stephan Kolassa & Roland Martin, 2011. "[Percentage Errors Can Ruin Your Day \(and Rolling the Dice Shows How\)](#)," *Foresight: The International Journal of Applied Forecasting*, International Institute of Forecasters, issue 23, pages 21-27, Fall.

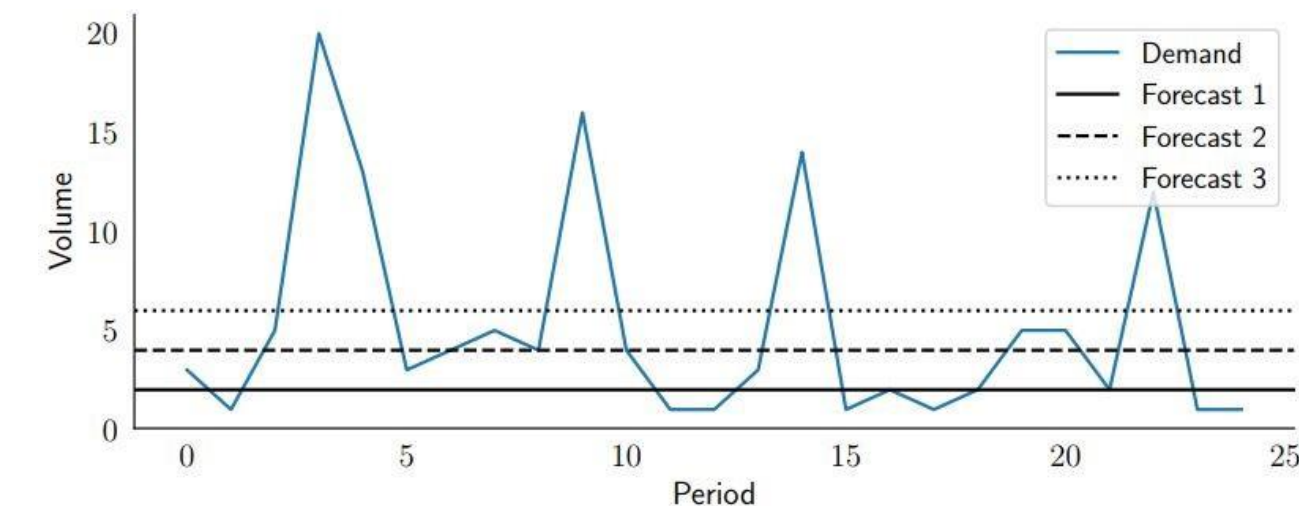
2.6. Métriques d'évaluation 5/7



L'importance du choix de la métrique : illustration

Si je compare plusieurs modèles de prévision, quel modèle choisir?

- Forecast 1 : faible quantité
- Forecast 2 : demande médiane
- Forecast 3 : demande moyenne
- Le choix de la métrique impacte sur le choix du modèle
- En fonction de la métrique, des objectifs différents :
- RMSE : cherche à produire un **forecast correct en moyenne**, non biaisé
- MAE : cherche à produire un **forecast médian**



	Forecast 1	Forecast 2	Forecast 3
Bias	-3.9	-1.9	0.1
MAPE	64%	109%	180%
MAE	4.4	4.1	4.8
RMSE	7.1	6.2	5.9

2.6. Métriques d'évaluation 6/7

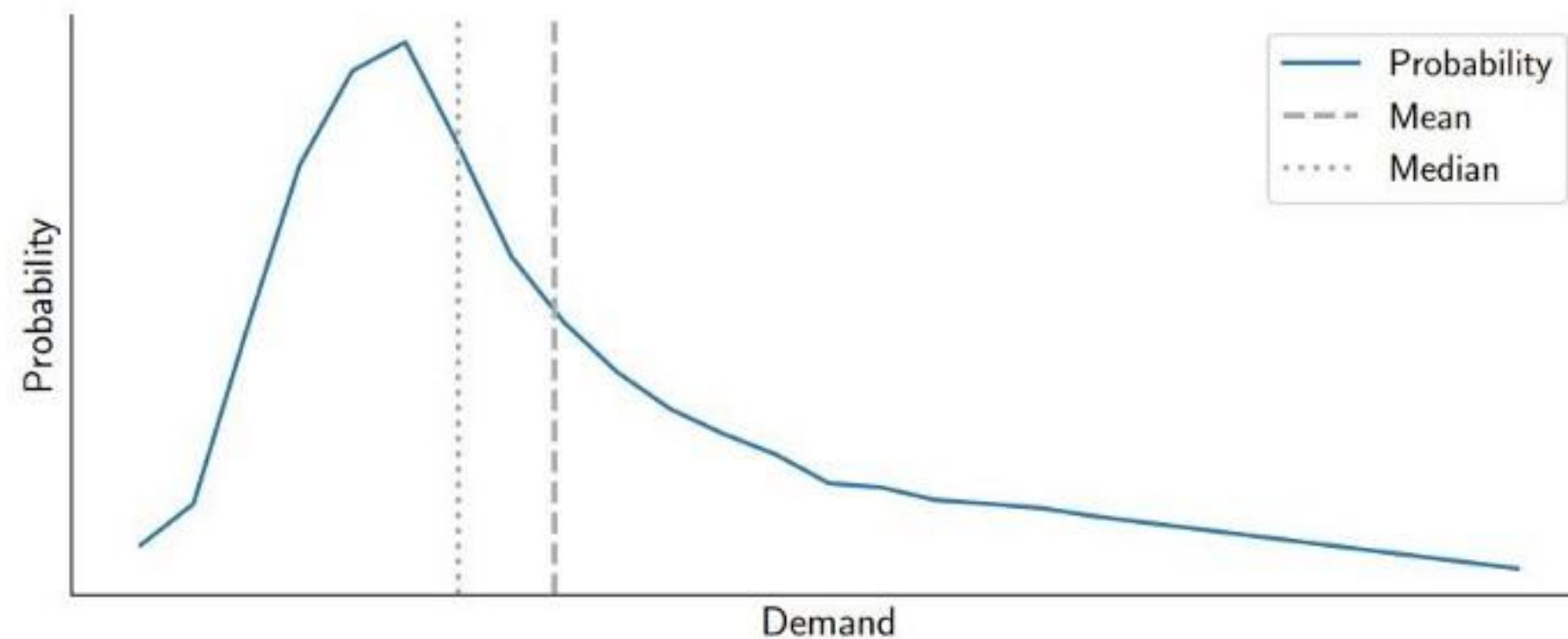


L'importance du choix de la métrique : illustration

Dans les cas d'usage de prévision, la quantité à prédire est rarement distribuée selon une loi normale

Phénomène courant en *supply chain*, la demande peut être soumise à de forts pics lors de promotion ou de commandes en gros

Dans cet exemple, la demande moyenne est supérieure à la demande médiane



2.6. Métriques d'évaluation 6/7



L'importance du choix de la métrique : illustration

Dans les cas d'usage de prévision, la quantité à prédire est rarement distribuée selon une loi normale

Phénomène courant en *supply chain*, la demande peut être soumise à de forts pics lors de promotion ou de commandes en gros

Dans cet exemple, la demande moyenne est supérieure à la demande médiane

Cela signifie qu'une prévision qui minimise la MAE entraînera un biais. En comparaison, une prévision minimisant la RMSE n'entraînera pas de biais (puisque'elle vise la moyenne). C'est certainement la principale faiblesse de MAE.

Le RMSE accorde une plus grande importance aux erreurs les plus élevées. Cela a un coût : une sensibilité aux valeurs extrêmes.

Important en supply chain car nombreuses valeurs extrêmes dues à des pics de demande (marketing, promotions, offres spéciales).

2.7. Métriques d'évaluation 7/7



L'importance du choix de la métrique : synthèse

Si je compare plusieurs modèles de prévision, quel modèle choisir?

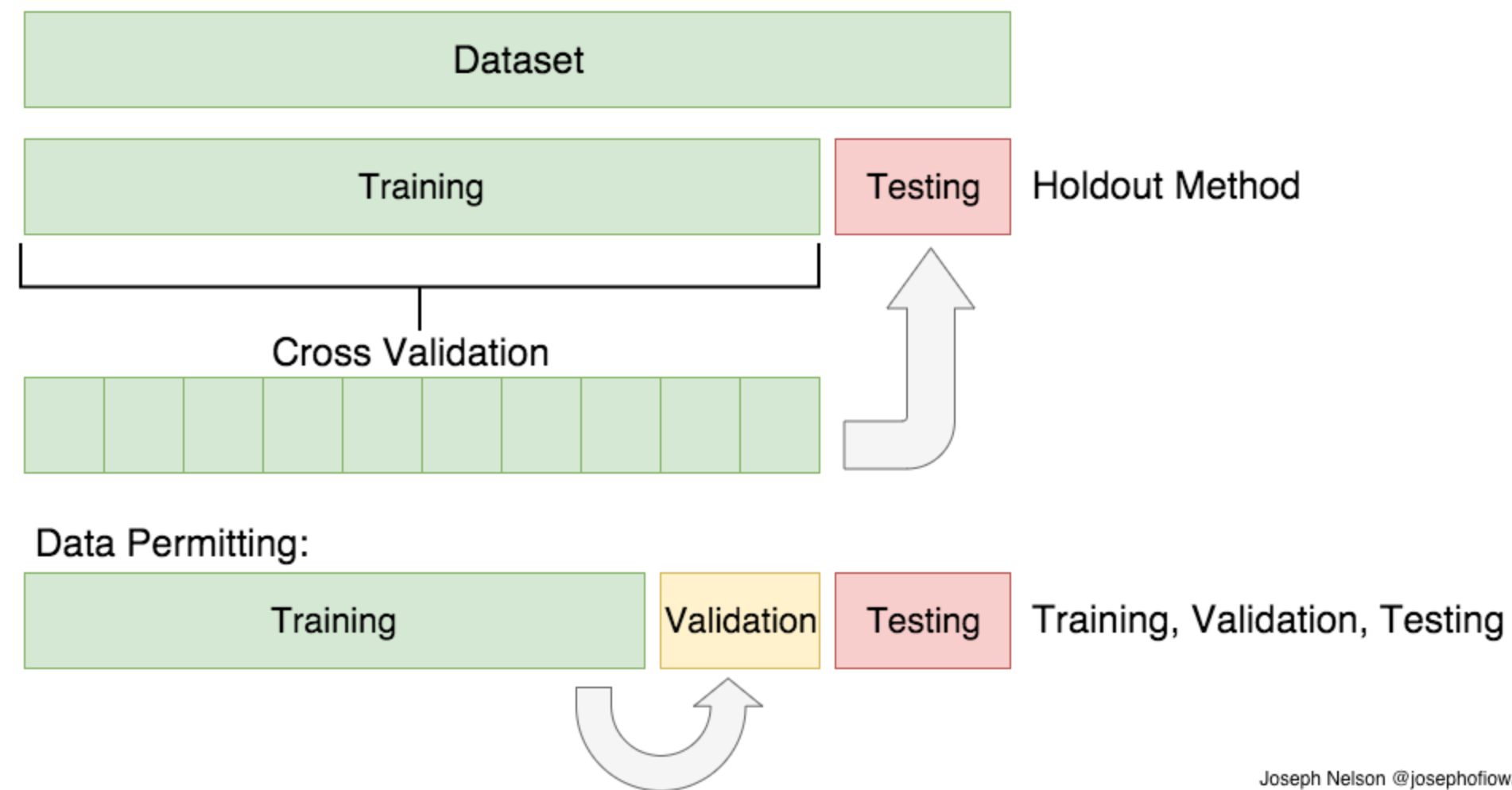
- Le choix de la métrique impacte sur le choix du modèle
- Le choix de la métrique doit être fait en fonction de critères métiers
- Si l'utilisation de la MAE comme KPI entraîne un biais élevé, il est préférable d'utiliser la RMSE.
- Si l'ensemble de données contient de nombreuses valeurs extrêmes, ce qui entraîne une prévision biaisée, il est préférable d'utiliser la MAE.
- Tout dépend du contexte métier, si vous souhaitez « être bon » sur les valeurs extrêmes ou pas : est-ce important pour le business?
- Conclusion : toujours valider avec le métier la métrique à utiliser!

2.8. Evaluer un modèle de prévision 1/3



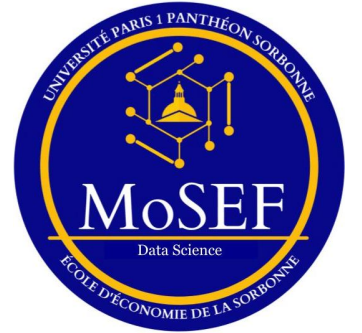
Méthodologie d'évaluation

- En machine learning, nécessaire de disposer d'un jeu d'entraînement et de test pour évaluer un modèle
- Séparation train et test classique: non adaptée, **il faut respecter l'ordre chronologique des series pour ne pas risquer une fuite de données (future leakage)**



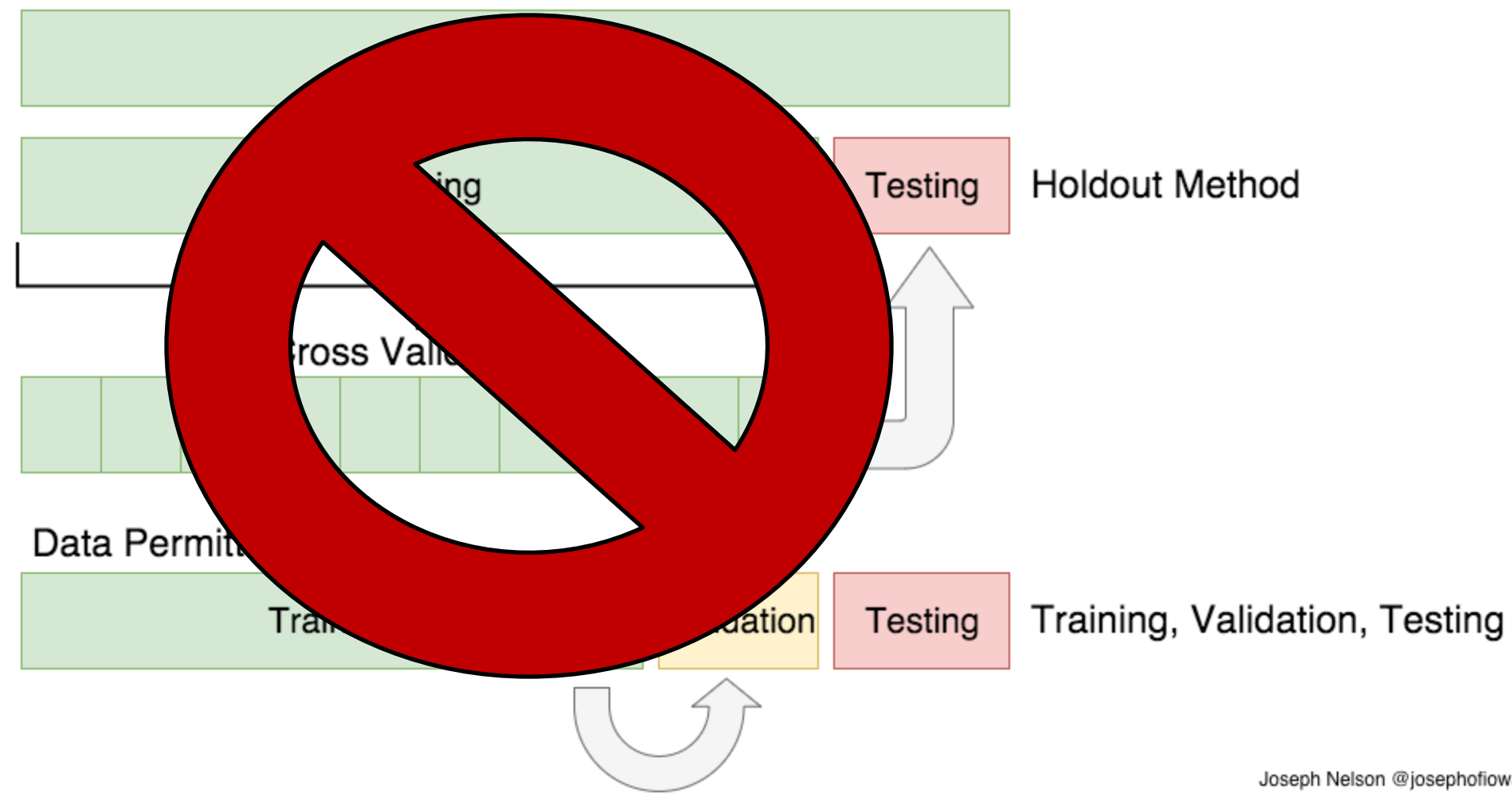
Joseph Nelson @josephofiowa

2.8. Evaluer un modèle de prévision 1/3



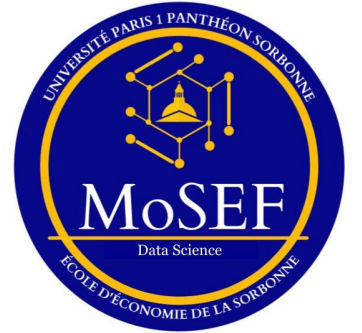
Méthodologie d'évaluation

- En machine learning, nécessaire de disposer d'un jeu d'entraînement et de test pour évaluer un modèle
- Séparation train et test classique: non adaptée, **il faut respecter l'ordre chronologique des series pour ne pas risquer une fuite de données (future leakage)**



Joseph Nelson @josephofiowa

2.8. Evaluer un modèle de prévision 1/3



Méthodologie d'évaluation

- En machine learning, nécessaire de disposer d'un jeu d'entraînement et de test pour évaluer un modèle
- Séparation train et test classique: non adaptée, **il faut respecter l'ordre chronologique des series pour ne pas risquer une fuite de données (future leakage)**

L'entraînement doit toujours s'effectuer sur un **ensemble de données qui sont plus anciennes que les données du test.**

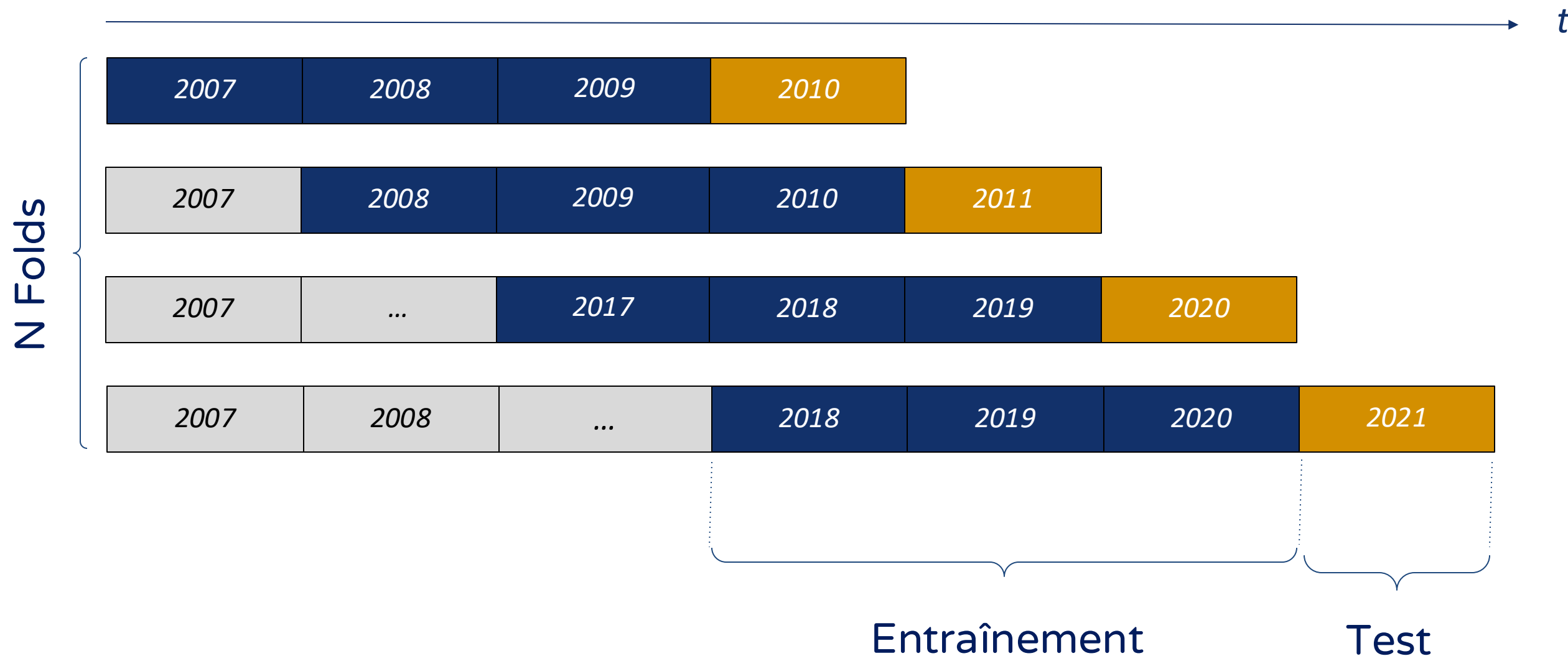
Deux approches principales pour tester les modèles de prévision:

- **Sliding window** : fenêtre de taille fixe pour l'entraînement. Permet un entraînement plus rapide, surtout face à un gros volume de données.
- **Expanding window** : utilise de plus en plus de données d'entraînement, tout en maintenant la taille de la fenêtre de test fixe. Cette dernière approche est particulièrement utile si la quantité de données à utiliser est limitée.

2.8. Evaluer un modèle de prévision 2/3



Méthode par fenêtre glissante (sliding window)



Entraînement (en bleu)

- Apprentissage du modèle en utilisant les données historiques
- Choix des paramètres du modèle avec un algorithme d'optimisation

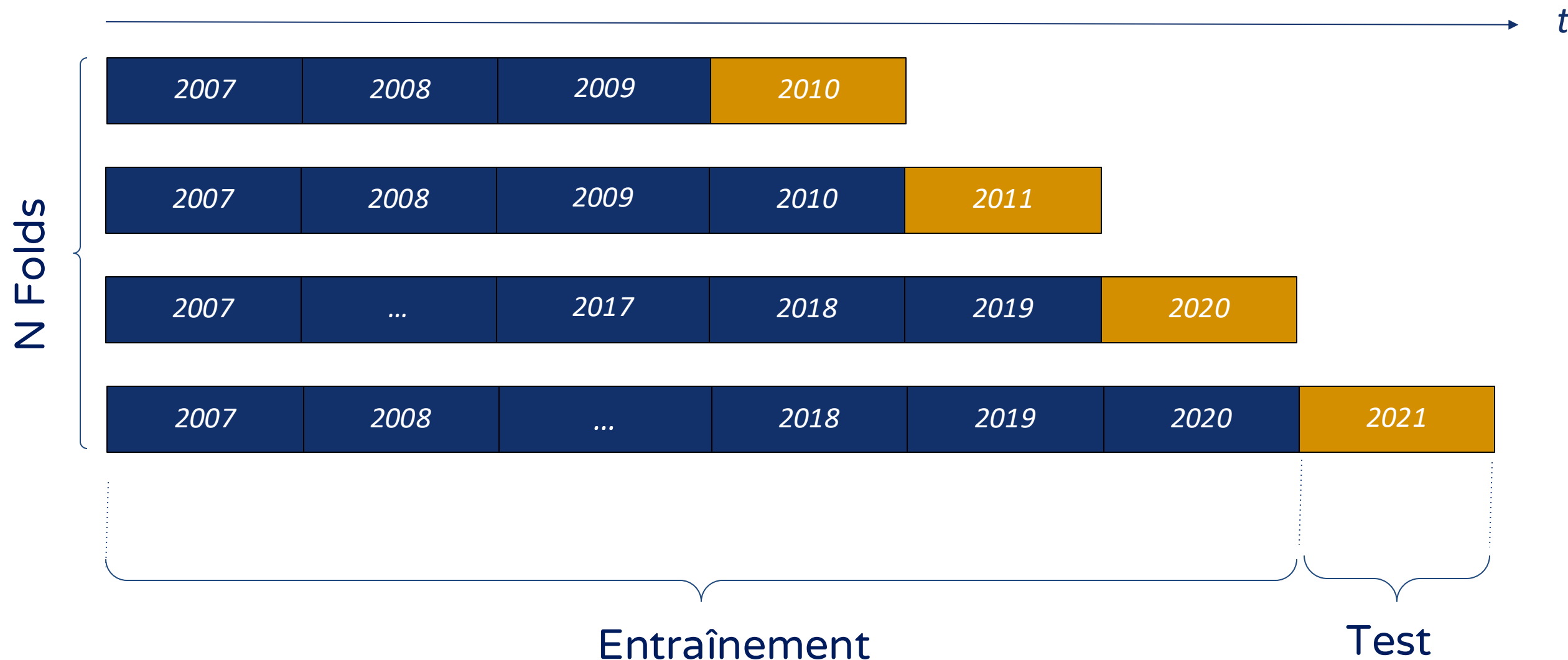
Test (en jaune)

- Comparaison de la prédiction avec le réel
- Calcul des métriques de performance (MAE, RMSE, ...)

2.8. Evaluer un modèle de prévision 3/3



Méthode par fenêtre croissante (expanding window)



Entraînement (en bleu)

- Apprentissage du modèle en utilisant les données historiques
- Choix des paramètres du modèle avec un algorithme d'optimisation

Test (en jaune)

- Comparaison de la prédiction avec le réel
- Calcul des métriques de performance (MAE, RMSE, ...)

2.9. Benchmarking & modèles naïfs 1/3



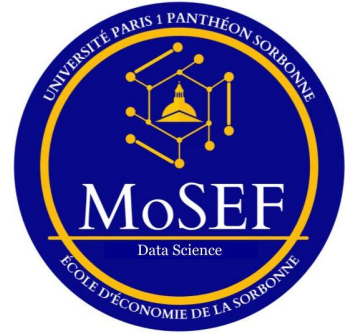
Un modèle de prévision doit être évalué selon plusieurs critères

- Métrique appropriée au **cas métier**
- Performance à mesurer pour **les différents horizons de prévisions**
- On constate souvent une **degradation de la performance** pour des horizons de plus en plus lointains

Benchmarking

- La performance doit être **comparée à des modèles naïfs**
- A quoi bon développer un modèle complexe s'il n'est pas meilleur qu'une moyenne mobile ou un modèle naïf saisonnier?

2.9. Benchmarking & modèles naïfs 2/3



Modèles naïfs

- Modèle naïf simple

$$y_{T+h|T} = y_T$$

- Modèle naïf saisonnier

$$y_{T+h|T} = y_{T+h-m(k+1)}$$

où m = période saisonnière et k est la partie entière de $(h-1)/m$ (c'est-à-dire le nombre d'années complètes dans la période de prévision précédant le temps $T+h$)

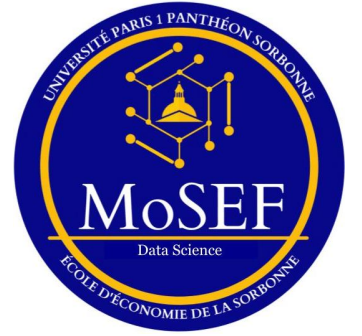
- Moyenne (ou moyenne mobile, avec $m < T$)

$$y_{T+h|T} = \bar{y} = (y_1 + \dots + y_T)/T$$

- Dérive

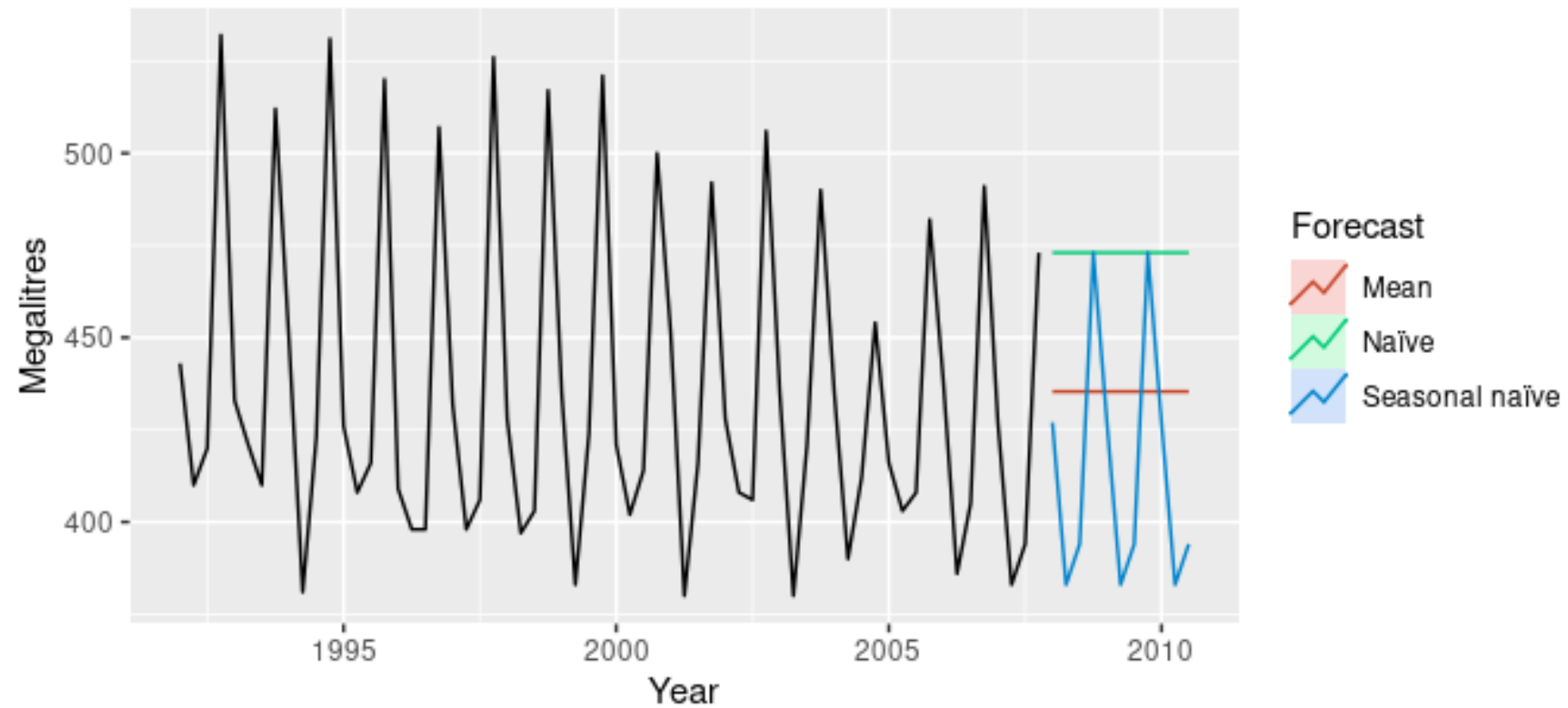
$$y_{(T+h|T)} = y_T + \frac{h}{T-1} \sum_{t=2}^T (y_t - y_{t-1}) = y_T + h \left(\frac{y_T - y_1}{T-1} \right)$$

2.9. Benchmarking & modèles naïfs 3/3



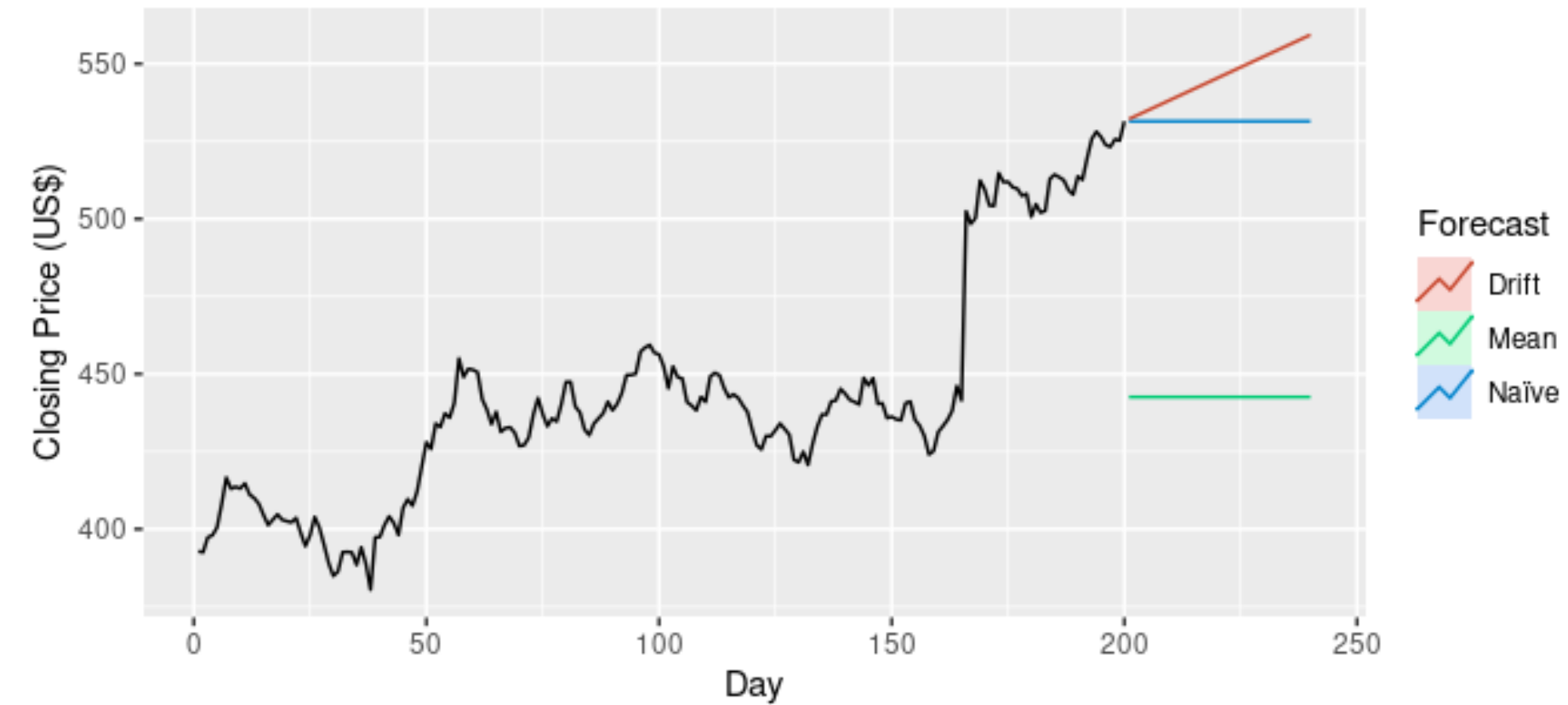
Modèles naïfs (illustration)

Forecasts for quarterly beer production



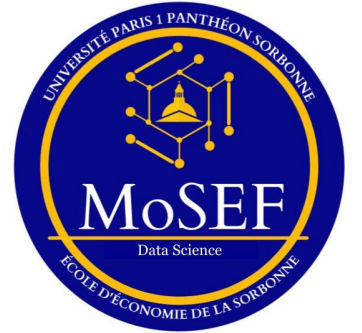
Moyenne, naïf et naïf saisonnier

Google stock (daily ending 6 Dec 2013)



Moyenne, naïf et dérive

2.10. Time to practice



TP2 EDA

Rendez-vous sans plus attendre sur Python pour la mise en pratique !

