



Time Series

HES-SO 6 Décembre 2024



Agenda

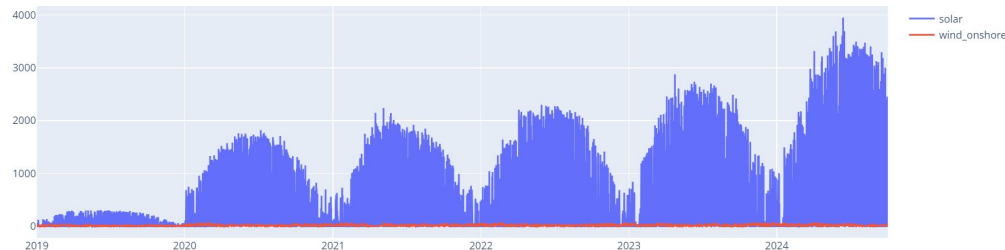
- Introduction
- Caractéristiques des séries temporelles
 - Définitions générales et importance de ce type des données
 - Granularité, Trend, Saisonnalité, Stationnarité
 - Demo d'analyse exploratoire sur time series énergétiques publiques (ENTSOe)
- Pause
- Modélisation
 - Use cases : detection d'anomalies, forecasts, clustering
 - Particularités d'entraînement sur le time series (cross validation, temporal leakage)
 - Points d'attention
 - Demo forecast demande électrique
- Feedback

Séries temporelles

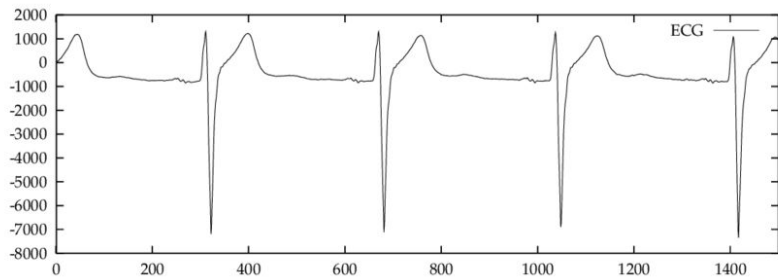
Série temporelle

Une série temporelle, ou série chronologique, est une suite de valeurs numériques représentant l'évolution d'une quantité spécifique au cours du temps.

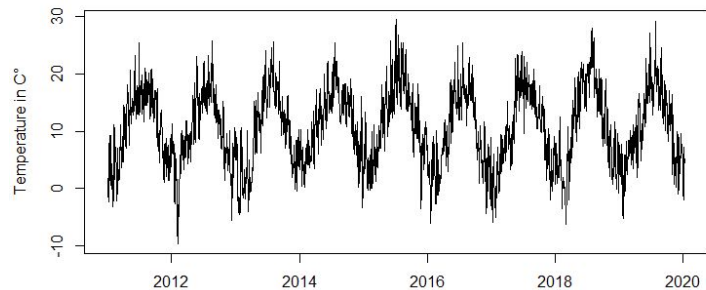
Solar & Wind Generation [MW]



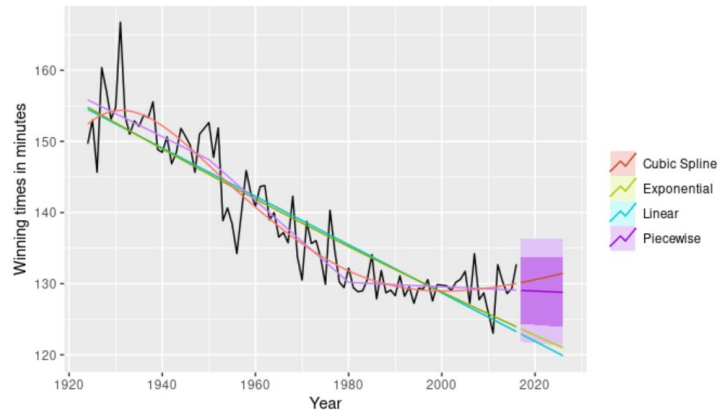
		dt	solar	wind_onshore
10	2019-01-01 10:00:00		27.0	1.0
11	2019-01-01 11:00:00		38.0	1.0
12	2019-01-01 12:00:00		42.0	1.0
13	2019-01-01 13:00:00		38.0	2.0
14	2019-01-01 14:00:00		25.0	2.0



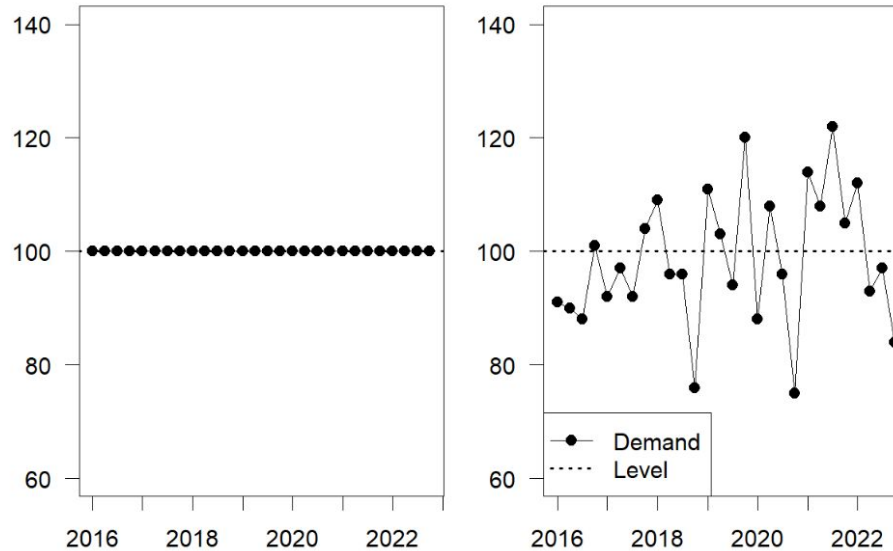
Daily Temperatur



Boston Marathon



Niveau



La valeur moyenne générale autour de laquelle les données fluctuent au fil du temps

Trend

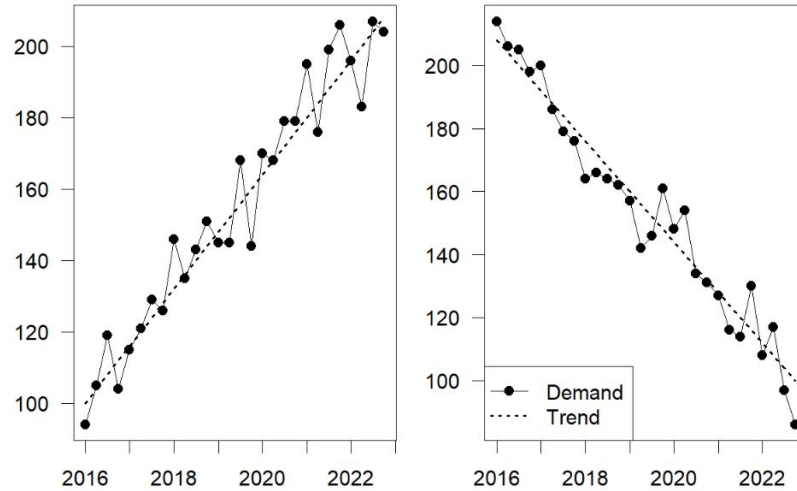
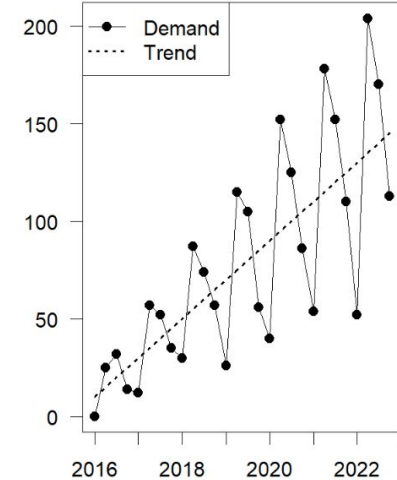
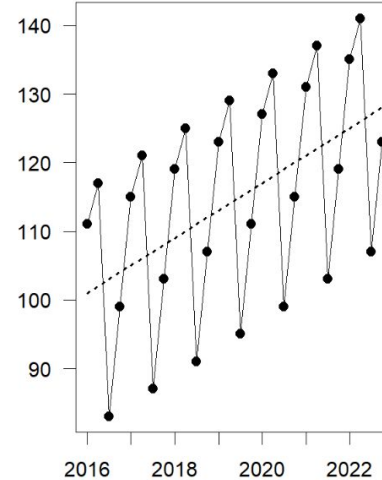


Figure 6.2: Time series with an increasing (left) and decreasing (right) trend

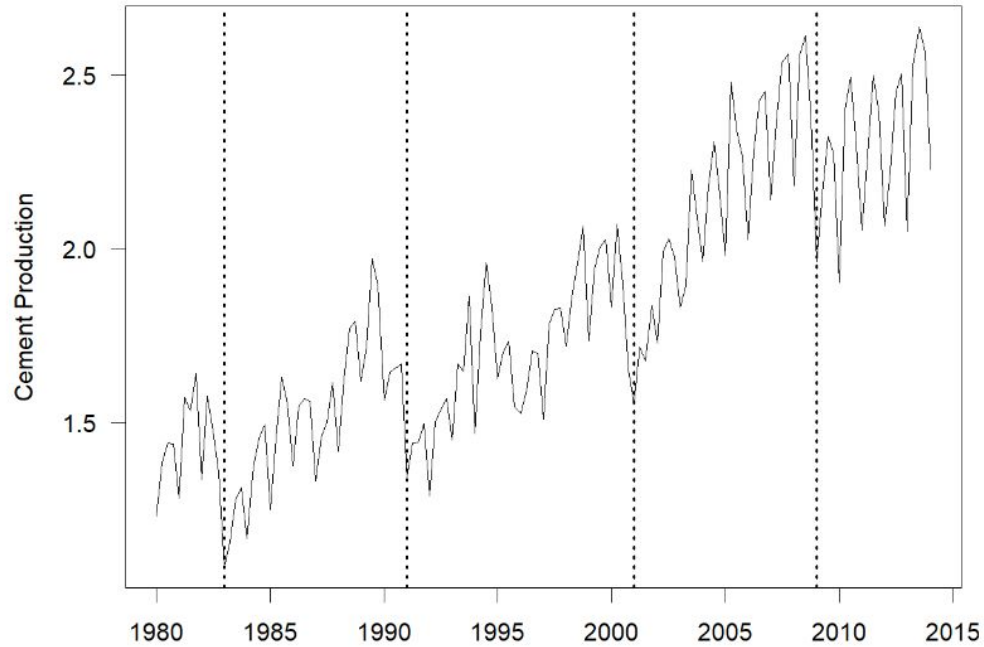
La direction à long terme que prend la série temporelle (les timeseries venant de la vie réelle peuvent avoir plusieurs trends)

Saisonnalité

Désigne des fluctuations régulières et récurrentes dans la série, généralement associées à des périodes spécifiques (jour, mois, trimestre, année). Ces variations se reproduisent de manière systématique. La saisonnalité peut être additive ou multiplicative par rapport à la tendance



Cycles



Les éléments cycliques reviennent dans le temps mais pas avec des intervalles fixes. Les causes des cycles sont souvent économiques, politiques ou facteurs sociaux

Decomposition Trend - Saisonnalité - Bruit



Decomposition Additive

$$Y(t) = \text{Niveau} + \text{Trend} + \text{Saisonnalité} + \text{Bruit}$$

Decomposition Multiplicative

$$Y(t) = \text{Niveau} * \text{Trend} * \text{Saisonnalité} * \text{Bruit}$$

Les timeseries de la vie réelle peuvent être plus compliquées !



Stationnarité

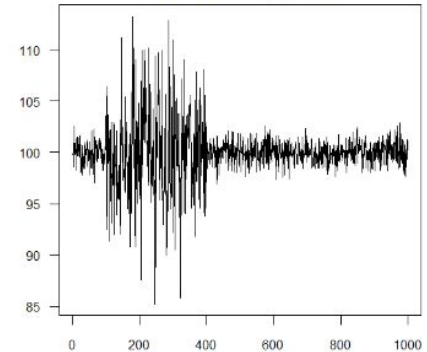
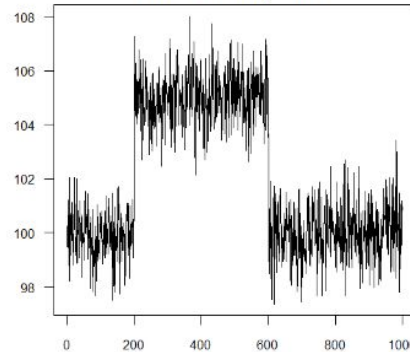
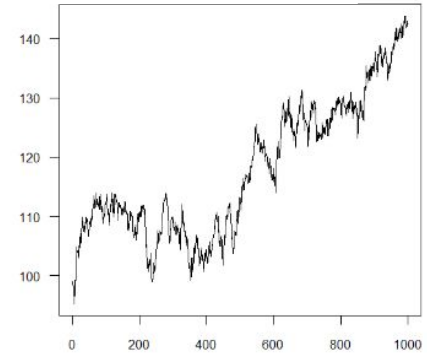
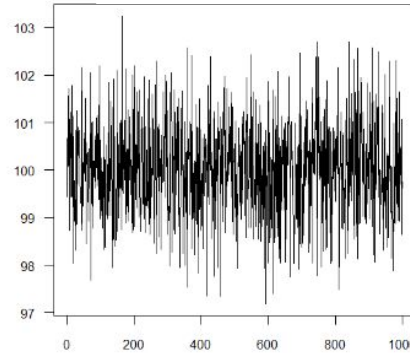
Une **série temporelle stationnaire** est une série dont les propriétés statistiques (moyenne, variance, et covariance) sont constantes au fil du temps

Des tests existent pour s'assurer de la stationnarité d'une time série

- Augmented Dickey-Fuller test (ADF Test)
- Kwiatkowski-Phillips-Schmidt-Shin test (KPSS test)

La stationnarité est nécessaire pour les modèles de prévision linéaires (famille ARIMA)

Une manière pour faire devenir une time série stationnaire est effectuée la différence d'une valeur par rapport à l'élément précédent



Autocorrelation



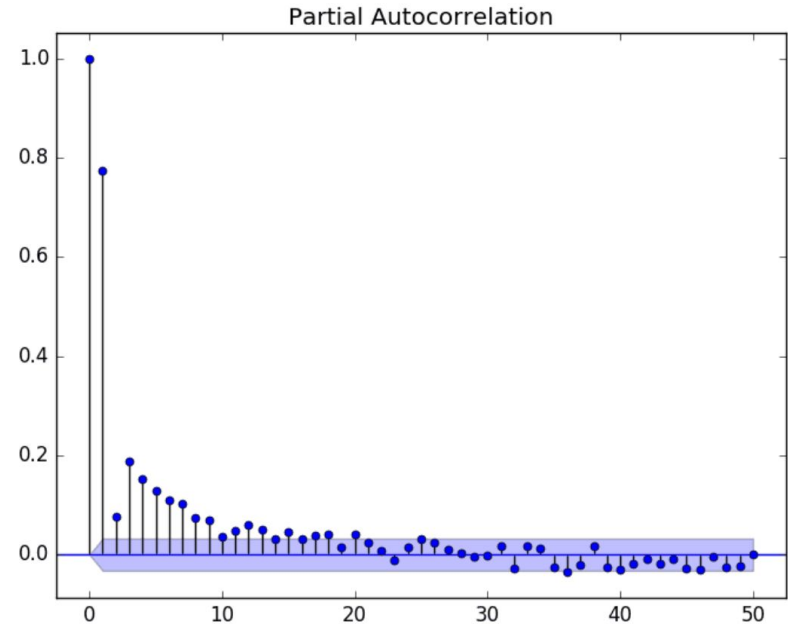
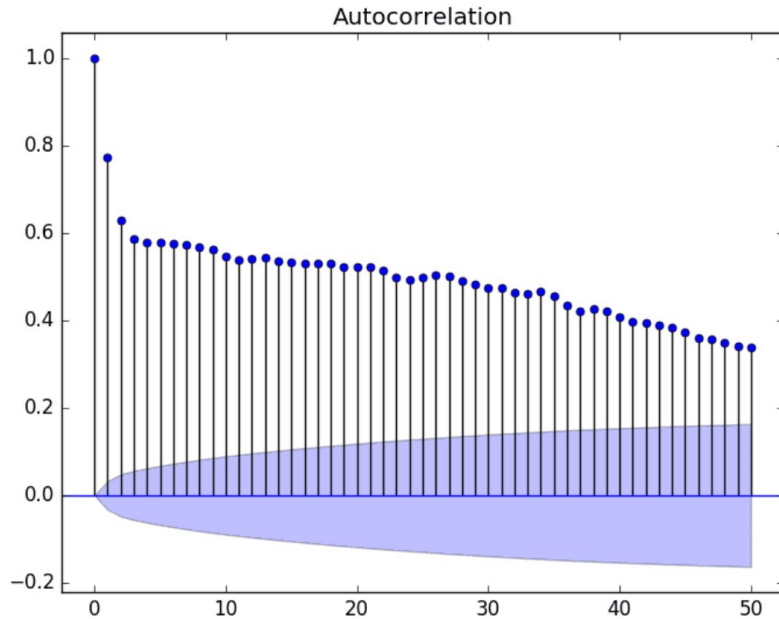
L'**autocorrélation** dans les séries temporelles mesure la corrélation entre les valeurs de la série à différents moments dans le temps. Autrement dit, elle examine dans quelle mesure une valeur de la série est liée à une ou plusieurs valeurs passées de la même série. Cela nous aide à identifier des schémas ou des tendances répétées, comme des cycles ou des saisonnalités, en nous montrant si les valeurs passées influencent directement les valeurs futures.

Autocorrélation (ACF - Autocorrelation Function): Il s'agit de la corrélation d'une série temporelle avec elle-même à différents décalages, appelés "lags"

Autocorrélation partielle (PACF - Partial Autocorrelation Function) : L'autocorrélation partielle, quant à elle, mesure la corrélation entre la série et ses lags en **éliminant les effets intermédiaires**

L'étude de l'autocorrélation et de l'autocorrélation partielle est essentiel pour la compréhension du phénomène

Autocorrelation

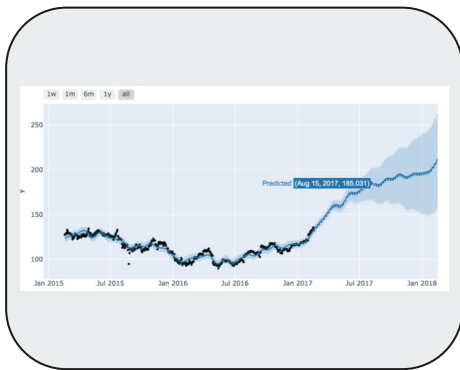


Demo:

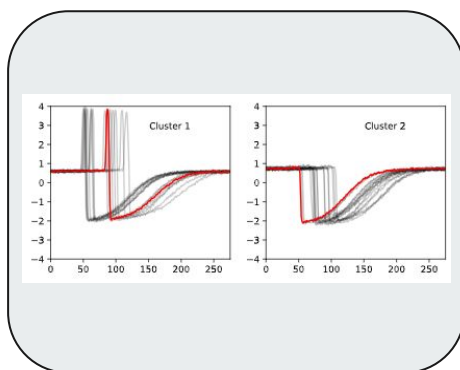
Analyse pratique des timeseries

Modélisation

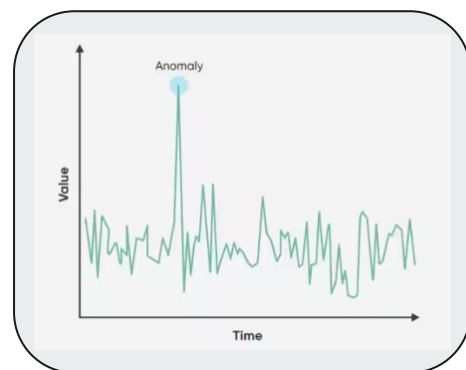
Modelisation



Forecasting



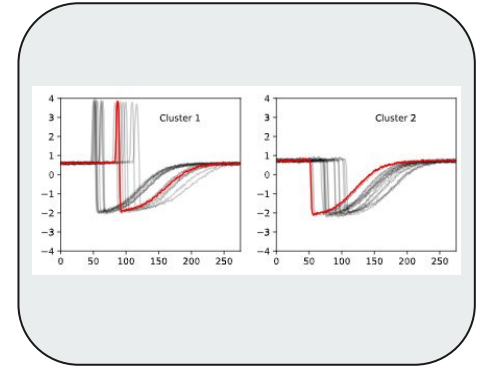
Clustering



Anomaly Detection

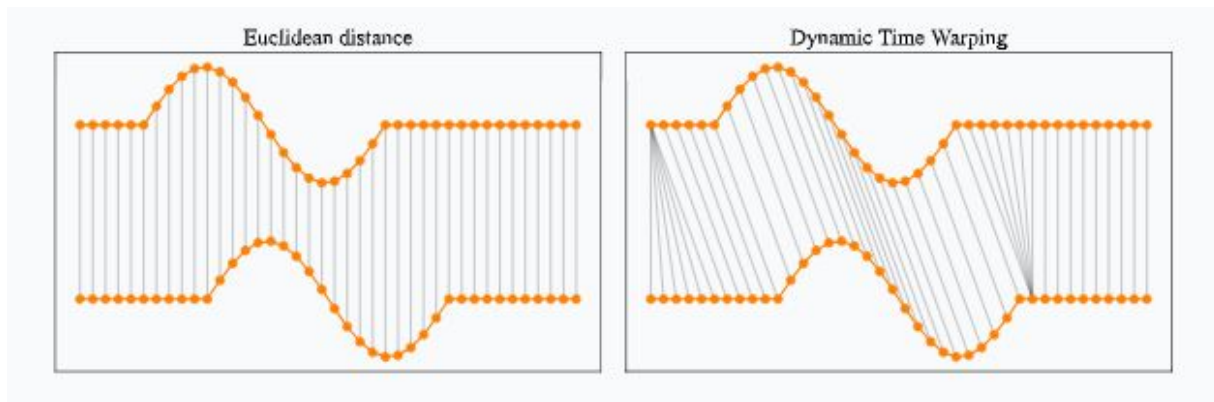
Clustering

- **Définition** : Le clustering de séries temporelles est une méthode qui regroupe des séries temporelles similaires en fonction de leurs caractéristiques (tendances, formes, motifs, etc.). Cela permet d'identifier des patterns communs entre différentes séries, même si elles représentent des systèmes différents.
- **Exemple** : Dans un environnement industriel, plusieurs capteurs enregistrent des signaux de machines différentes. Le clustering peut regrouper les machines qui ont des comportements similaires, facilitant ainsi la maintenance et l'analyse des performances des équipements. Un autre exemple peut être lié aux pattern de consommation énergétique des clients industriels qui peuvent identifier des jours similaires ou des “jours type”



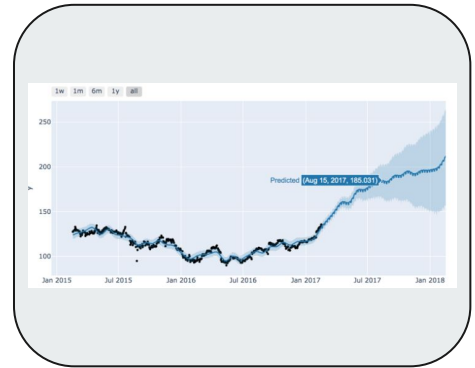
Clustering

- Applications massives dans le traitement du son, le but c'est d'identifier des similarités entre différentes séries temporelles ou entre différentes parties de la même série temporelle
- Technique non-supervisionné
- Particulièrement complexe dans le cas des time series car le clustering dépend principalement des mesures de **distance**: quoi utiliser pour mesurer la distance entre deux séries temporelles ? Comment capturer correctement la similarité de deux timeseries qui ont un "décalage temporel" ?
- Souvent on peut appliquer des filtres ou des opérations d'extraction des features pour simplifier le processus

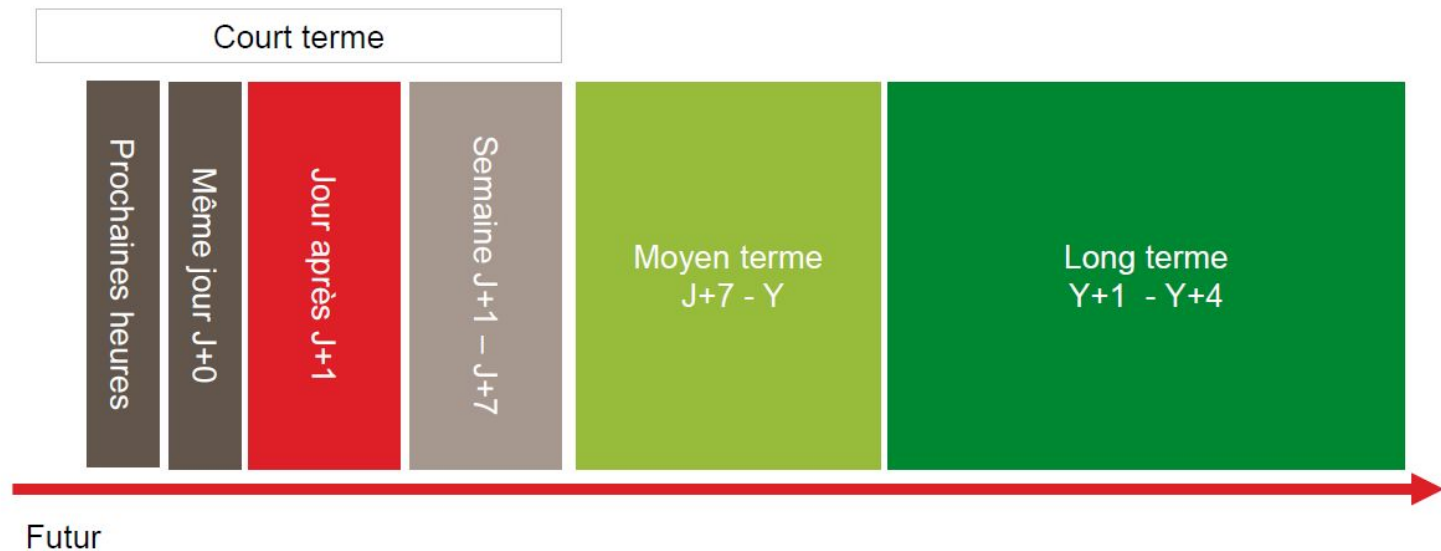


Forecasts

- **Définition** : Le forecasting consiste à utiliser des données passées pour prédire des valeurs futures d'une série temporelle. Il repose souvent sur des méthodes statistiques ou des modèles d'apprentissage automatique pour identifier des tendances et des patterns dans les données historiques.
- **Exemple** : Prédire les ventes futures d'un produit dans un supermarché en fonction des ventes passées. Par exemple, si un magasin observe une hausse des ventes de certains produits avant les fêtes de fin d'année, le forecasting peut anticiper cette tendance pour ajuster les stocks.



Horizons de prévisions



Forecasting



Le monde de l'énergie nécessite des prévisions dans plusieurs sous domaines:

- Approvisionnement
- Achats/Ventes d'énergie
- Optimisation des assets de production / flexibilité
- Gestion de l'infrastructure électrique
- Prévisions de surcharge et décisions de renforcement
- Optimisation énergétique
- Optimisation des consommations industrielles
- Décisions d'optimisation de chauffage
- Décision d'investissement dans de ouvrages renouvelables

Models



Forecasting models

Naïfs

- Persistence models
- Expert judgment

Statistiques

- Arima
- Sarima
- ETS
- Vector Autoregression

ML

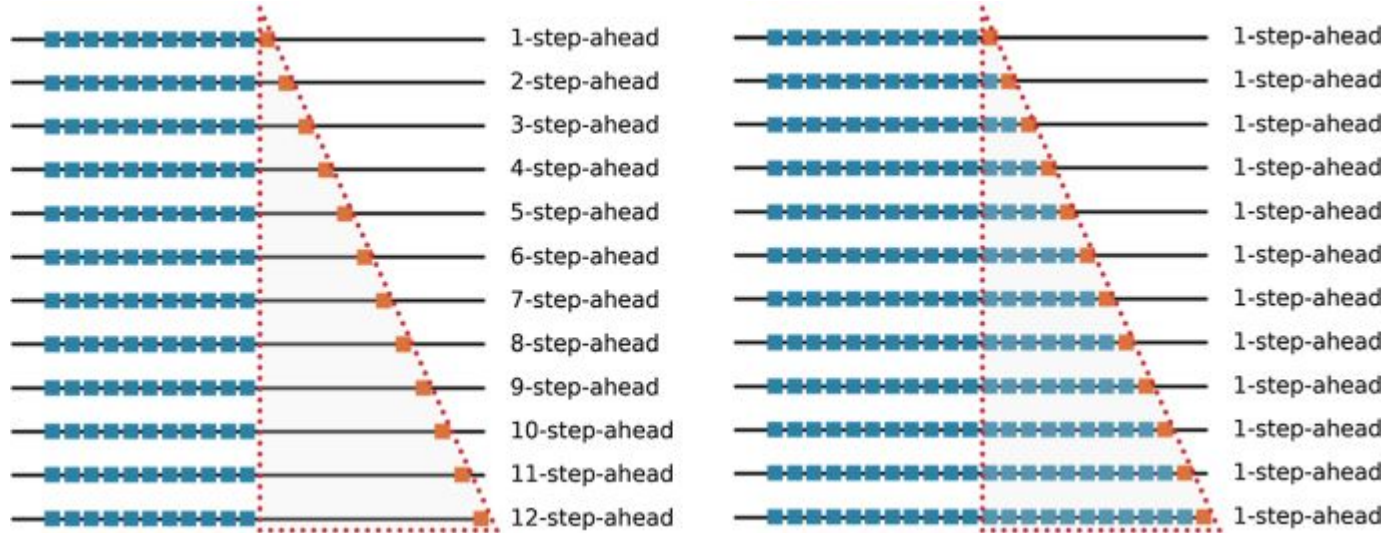
- Linear Regression
- KNN
- Tree-based models (Decision Tree, Random Forest Boosted models)

Deep Learning

- LSTM
- RNN
- Transformers

Hybrid

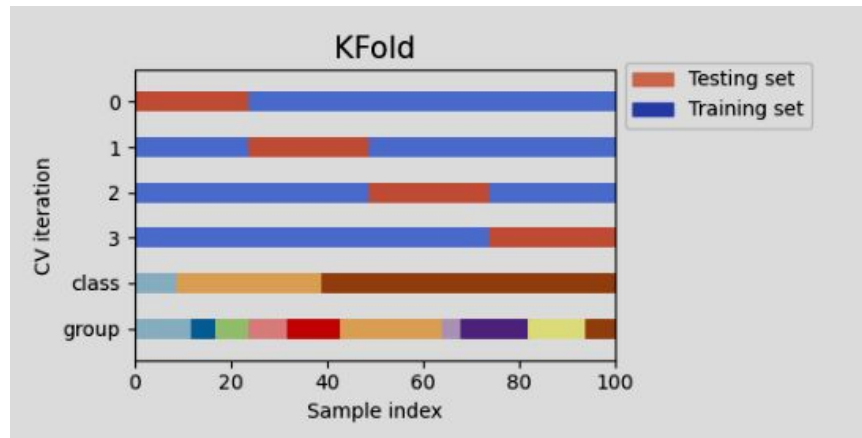
Point forecast



Certains modèles effectuent un forecast ponctuel, d'autres une série de points futurs. Dans le premier cas on peut utiliser d'une manière sequentielle les resultats pour fournir une série de forecasts

Machine learning sur time series: points d'attention

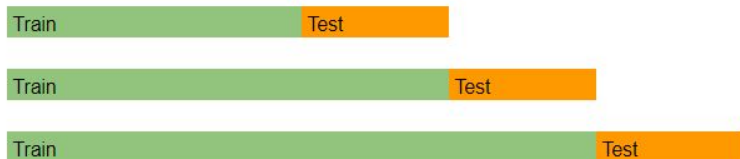
La **cross-validation** est une méthode statistique utilisée pour évaluer la performance d'un modèle en apprentissage automatique. Elle consiste à diviser l'ensemble des données en plusieurs sous-ensembles, appelés *folds*. À chaque itération, le modèle est entraîné sur une partie des données ($k-1$ folds) et testé sur la partie restante (1 fold). Ce processus est répété k fois, et les performances sont ensuite moyennées pour obtenir une estimation plus robuste, limitant les risques de sur apprentissage (*overfitting*) et améliorer la capacité du modèle à généraliser



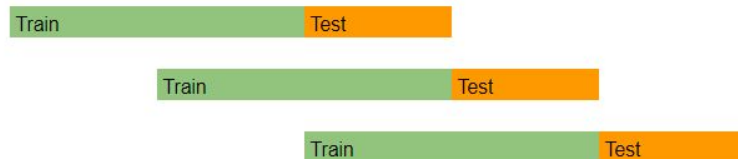
L'ordre n'as pas d'importance?

Time Series Cross validations

Expanding Window Cross Validation

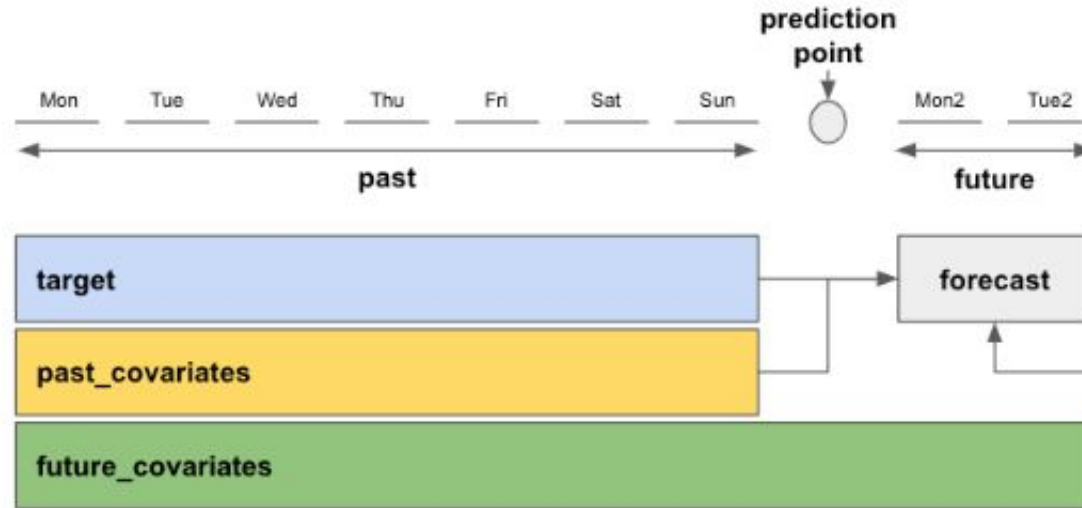


Sliding Window Cross Validation



Plusieurs techniques de cross-validation sont présentés pour les timeseries et leur choix est souvent faite en relation au problème en question, toutefois l'utilisation de techniques de cross validation classique sont déconseillés à cause du risque d'utiliser des informations “futures de la time série”. Expanding window and Sliding window sont à préférer.

Connaissance du passé ET Connaissance du futur



Dans un exercice de forecast on connaît des quantités décrivant le passé et d'autres décrivant le futur. L'utilisation d'une combinaison des deux est souvent présent, selon l'impact des deux familles sur les prévisions.

Avant de lancer un modèle de deep learning contre les timeseries



- 1) Est-ce-qu'il y a eu un changement important de comportement dans ma time série que j'utilise pour entraîner les modèles ?
- 2) Est-ce-qu'on a raison de penser que le futur ne suivra pas les mêmes logiques que le passé ?
- 3) C'est quoi l'état de l'art sur ces modèles ?

Vers la mise en production



- 1) Dialoguer avec les ML Engineers/DevOps pour s'assurer d'avoir les bons composants pour la mise en production
- 2) Mettre en place les mesure pour s'assurer de la qualité des inputs
- 3) Prévoir un monitoring des performances des modèles
- 4) Prévoir une stratégie d'entraînement en production



Ressources

Repository avec cette présentation et les démos:

https://github.com/PatrizioCanzi/hes_timeseries

Ressources sur time series

<https://github.com/Nixtla/fpp3-python>

<https://otexts.com/fpp3/>

<https://dfep.netlify.app/>

https://tslearn.readthedocs.io/en/stable/user_guide/clustering.html

https://unit8co.github.io/darts/generated_api/darts.timeseries.html

https://scikit-learn.org/stable/modules/cross_validation.html