# Nowcasting PIB : Imputation des données non encore publiées

Marion CABROL[1], Kevin FERNANDES[1]
Michel MARTINEZ[2]

[1] Core data science team - Société Générale CIB
[2] Chef économiste zone euro - Société Générale CIB

31 mars 2022

# Contexte : "Nowcasting" de PIB français

- **Anticipation de la position de l'économie française (PIB)**

  Modèle de "Nowcasting" ("now"+"forecasting") : Aptitude d'un modèle à fournir des prédictions immédiates

  Prédiction : À tout moment au cours du trimestre jusqu'à la première estimation du PIB

- **Obstacles :** Modéliser en tenant compte d'un flux de données

  Conforme à la réalité

  Soumis à de nombreuses contraintes (e.g., "ragged-edge", révisions)

- **Utilisation des méthodes d'apprentissage statistique peu fréquentes**

  Utilisation répandue de méthodes économétriques (e.g., DFM)

  Résultats probants des modèles d'apprentissage statistique ("machine learning") dans de nombreux domaines

  + Adaptation : Volume de données
  − Contrainte : Ensemble prédéfini de variables explicatives requis

# Notre contribution

**Modélisation basée sur la création d'un flux réaliste de 634 variables macro-économiques et financières**

(période valorisée, valeur) $\Rightarrow$ (période valorisée, **date de publication**, valeur)

- **Solutions concrètes** de traitement du flux de données :
  - Prise en compte des révisions, changements de nomenclature, mixtes de fréquence etc.
  - Réadaptation des traitements de séries temporelles (e.g., révisions, double indexation temporelle) : Traitements classiques obsolètes

## Modélisation en deux temps

1. **Imputation des variables non encore publiées :** Prévoir les données non encore publiées
2. **Prédiction du PIB :** Prédire le PIB à l'aide d'un jeu de données complet (variables déjà publiées et variables imputées à l'étape 1).

- Avantages d'une modélisation en 2 temps :
  - Utilisable avec des modèles récents ("machine learning")
  - Modèle unique quelque soit l'information disponible
  - Prise en compte des interactions entre variables macro-économiques

# Notre contribution

- **Contribution passée d'une modélisation en 2 temps :**

Miller et al. (1996) [6], de Zheng et al. (2006) [10] et de Bouwman et al. (2011) [2] :

  - Première étape d'imputation des travaux précédents basée sur des méhodes de **prévisions temporelles autoregressives** :
    - AR, VAR, "Naïve random walk in growth rate"

- **Contribution actuelle :**

**Quelle méthode d'imputation est la plus appropriée ?**

  - Recherches liées à l'apprentissage statistique (ou **"machine learning"**), y compris l'apprentissage profond (ou **"deep learning"**)
  - Recueils relatifs à l'imputation de données manquantes
  - Méthodes de prévisions temporelles

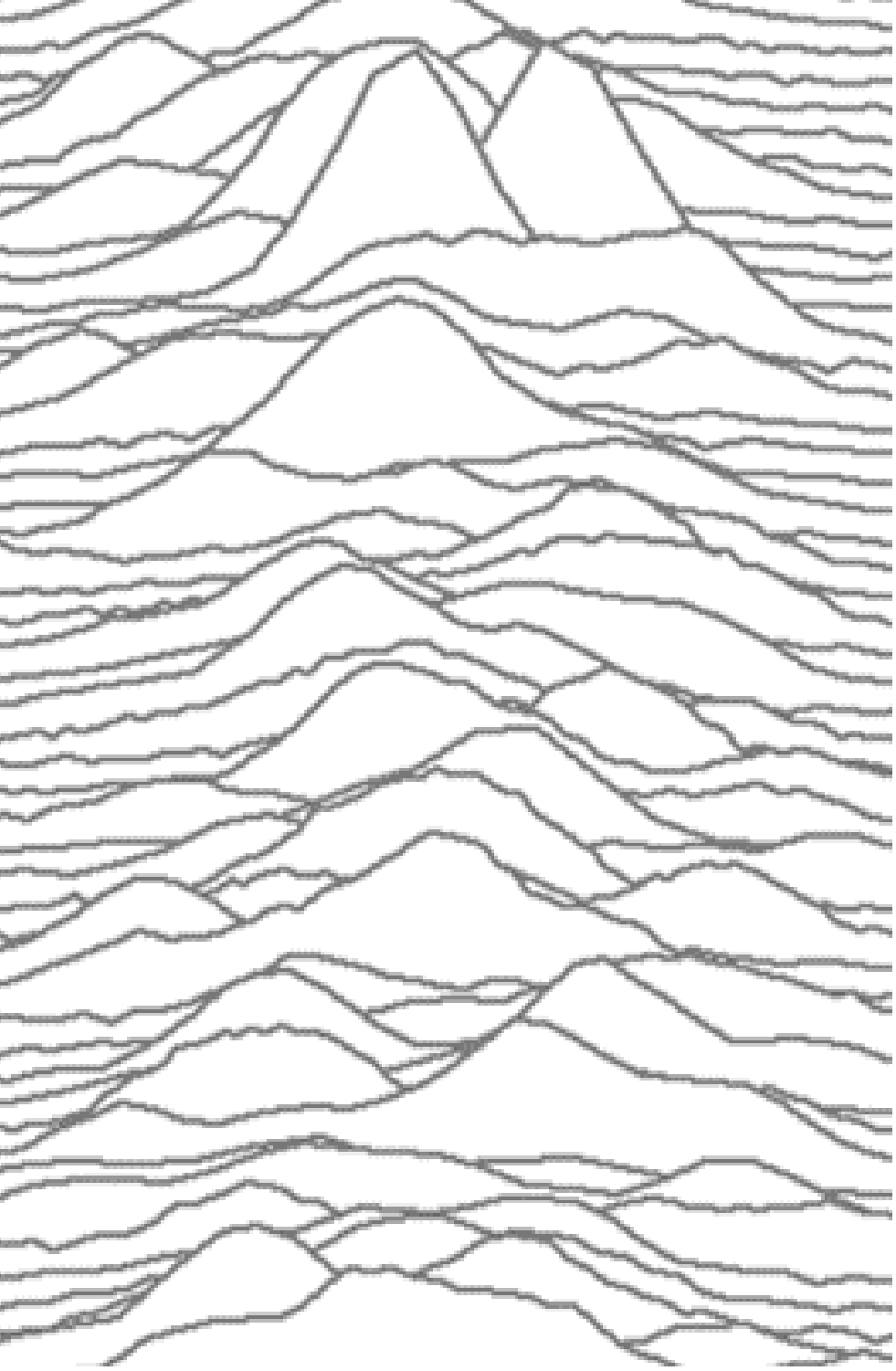
**Optimisation d'une modélisation en 2 temps :**

Le modèle devient un paramètre à optimiser.

# Notre contribution

## Améliorations

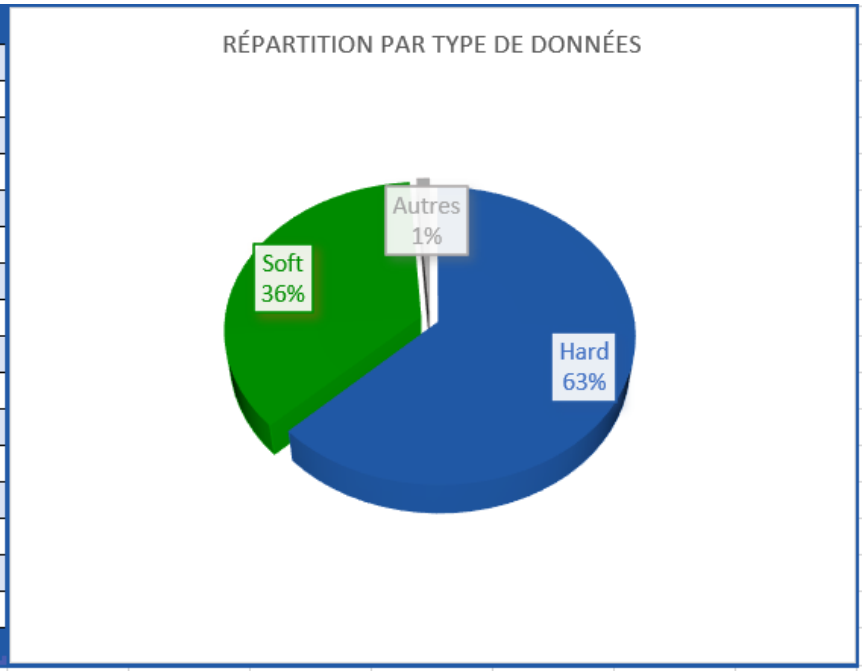
- RMSFE de prédiction du PIB français : diminué en utilisant des méthodes d'imputation différentes de méthodes de prévision temporelle autoregressive (recherches passées).
- L'**IKNN** est un candidat d'imputation optimal dans un contexte "business as usual".
- L'**imputation monotone** sur-performe dans un début de crise covid-19.
- RMSFE multipié par 10 en contexte de début de crise covid-19.



# 1. CREATION ET TRAITEMENT DU FLUX DE DONNEES

# 1. Création et traitement du flux de données

| Catégorie | Nombre de variables |
|---|---|
| Enquêtes auprès des entreprises | 217 |
| Comptes nationaux | 208 |
| Commerce | 59 |
| Marché du travail | 41 |
| Production industrielle | 31 |
| Consommation | 15 |
| Enquêtes auprès des consommateurs | 13 |
| Construction | 12 |
| Revenus et investissements | 10 |
| Sectoriel | 10 |
| Finance | 7 |
| Masse monétaire | 4 |
| IPC (indice des prix à la consommation) | 3 |
| Productivité | 2 |
| Comptes relatifs au gouvernement | 1 |
| IPP (indice des prix à la production) | 1 |
| **Total** | **634** |



# 1. Création et traitement du flux de données

**1. Reconstruction de la chronologie des publications passées**
(période valorisée, valeur) -> (période valorisée, date de publication, valeur)

**Création**

**2. Détection de changement de nomenclature et rétropropagation de la nomenclature actuelle**

**3. Détection et suppression de tendances dans nos séries.**

**4. Incorporation de variables décalées temporellement**

**5. Filtres temporels**

**6. Agrégation mixte de fréquence**

**Traitement**

# 1. Exemple factice de flux de données

**Triplet (période valorisée, date de publication, valeur)**

| Trimestre | Date de disponibilité | PMI construction 1er mois | PMI construction 2ème mois | PMI construction 3ème mois | Consommation ménages -textiles - trimestre précédent | Ventes de véhicules à moteur | Climat des affaires 1er mois | Climat des affaires 2ème mois |
|---|---|---|---|---|---|---|---|---|
| Q3 2019 | 2019-10-23 | -0.23198281 | -0.28821969 | 0.054068937 | 0.053352713 | 0.501582275 | 0.194745065 | -0.142081073 |
| | 2019-10-25 | -0.23198281 | -0.28821969 | 0.054068937 | 0.053352713 | 0.501582275 | 0.194745065 | -0.142081073 |
| | 2019-10-29 | -0.23198281 | -0.28821969 | 0.054068937 | 0.053352713 | 0.501582275 | 0.194745065 | -0.142081073 |
| | 2019-10-30 | -0.23198281 | -0.28821969 | 0.054068937 | 0.053352713 | 0.501582275 | 0.194745065 | -0.142081073 |
| Q4 2019 | 2019-10-01 | | | | -0.302332043 | | | |
| | 2019-10-04 | | | | -0.302332043 | | | |
| | 2019-10-07 | | | | -0.302332043 | | | |
| | 2019-10-08 | | | | -0.302332043 | | | |
| | 2019-10-09 | | | | -0.302332043 | | | |
| | 2019-10-10 | | | | -0.302332043 | | | |
| | 2019-10-15 | | | | -0.302332043 | | | |
| | 2019-10-22 | | | | -0.302332043 | | | |
| | 2019-10-23 | | | | -0.302332043 | | | |
| | 2019-10-24 | | | | -0.302332043 | | 0.065493102 | |
| | 2019-10-25 | | | | -0.302332043 | | 0.065493102 | |
| | 2019-10-29 | | | | -0.302332043 | | 0.065493102 | |
| | 2019-10-30 | | | | -0.302332043 | | 0.065493102 | |
| | 2019-10-31 | | | | -0.302332043 | | 0.065493102 | |
| | 2019-11-01 | -0.733100539 | | | -0.302332043 | | 0.065493102 | |
| | 2019-11-04 | -0.733100539 | | | -0.302332043 | | 0.065493102 | |
| | 2019-11-05 | -0.733100539 | | | -0.302332043 | | 0.065493102 | |
| | 2019-11-07 | -0.733100539 | | | -0.302332043 | | 0.065493102 | |
| | 2019-11-08 | -0.733100539 | | | -0.302332043 | | 0.065493102 | |
| | 2019-11-12 | -0.733100539 | | | -0.302332043 | | 0.065493102 | |
| | 2019-11-13 | -0.733100539 | | | -0.302332043 | | 0.065493102 | |
| | 2019-11-14 | -0.733100539 | | | -0.302332043 | | 0.065493102 | |
| | 2019-11-20 | -0.733100539 | | | -0.302332043 | | 0.065493102 | |
| | 2019-11-21 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-11-22 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-11-26 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-11-27 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-11-28 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-11-29 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-02 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-03 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-04 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-05 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-06 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-10 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-11 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-12 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-19 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-20 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-23 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-24 | -0.733100539 | -0.014471138 | -0.72426521 | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-25 | -0.733100539 | -0.014471138 | -0.72426521 | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-27 | -0.733100539 | -0.014471138 | -0.72426521 | -0.302332043 | | 0.065493102 | -0.088057963 |

# 1. Création du flux de données

**1. Reconstruction de la chronologie des publications passées**
(période valorisée, valeur) -> (période valorisée, date de publication, valeur)

Evaluation des règles de publication basées sur une période de référence → Projection sur tout l'historique

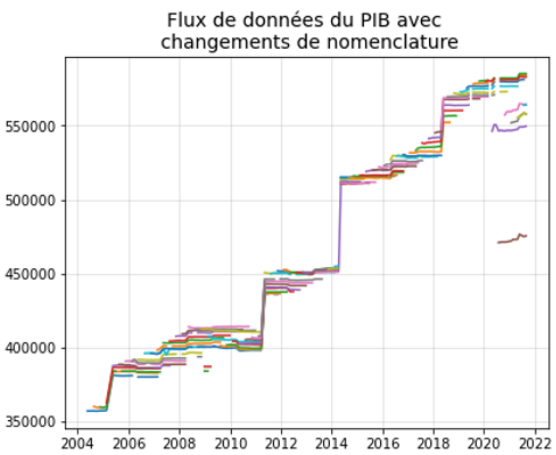
# 1. Traitement du flux de données

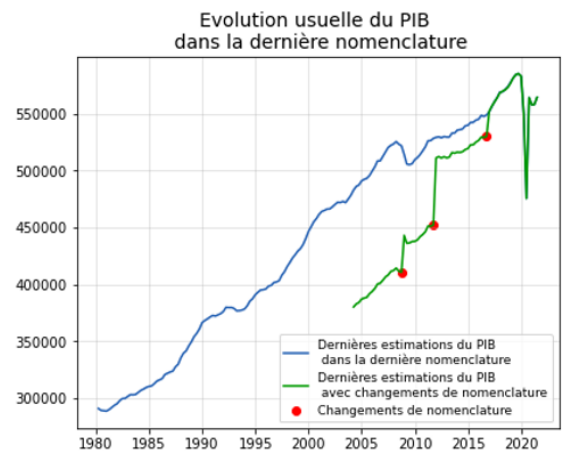
**2. Détection de changement de nomenclature et rétropropagation de la nomenclature actuelle**

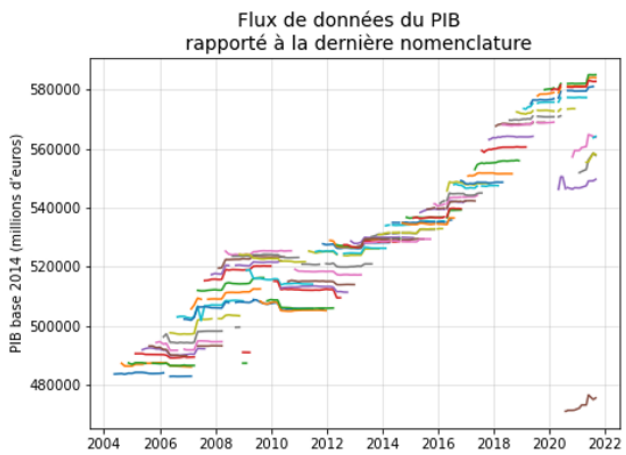
- Série de référence
**dernière nomenclature** avec valeur finale
- Flux de données à traiter :
avec **changement de nomenclature** et **révisions**

→

- **Détection d'anomalies** sur chaque période
adaptation de l'algorithme de Tolvi (2001)
- **Rétropagation**



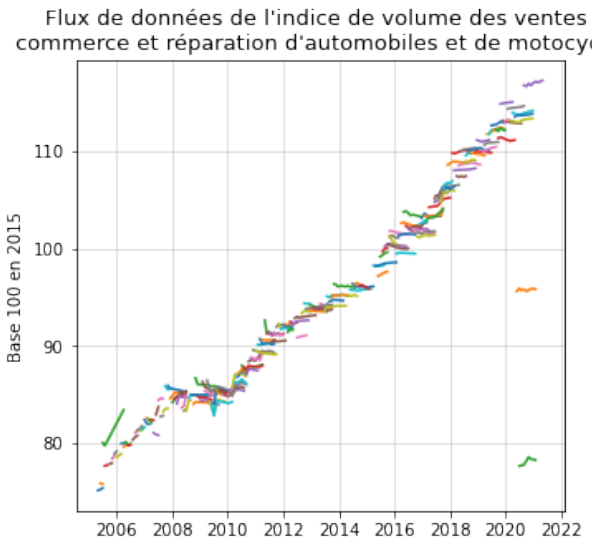



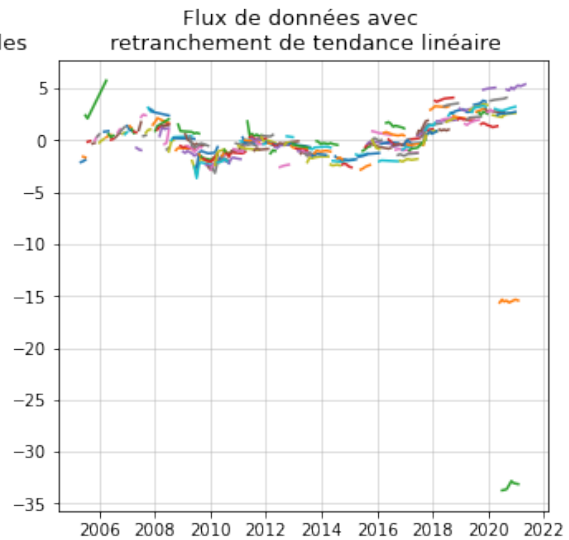

Tolvi, J. (2001). Outliers in eleven finnish macroeconomic time series. Finnish Economic Papers, 14(1), 14-32.

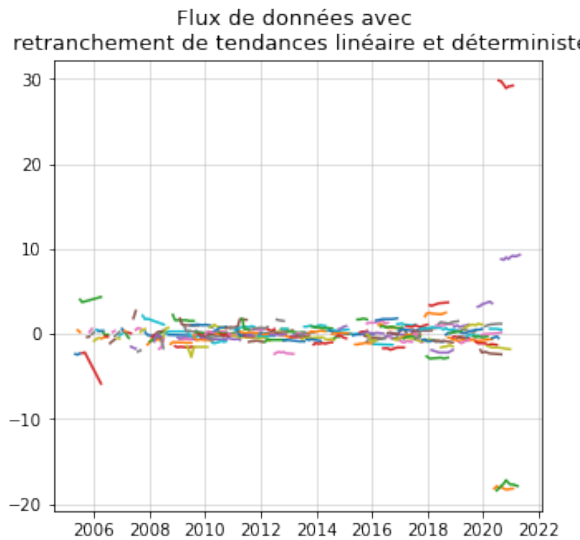
# 1. Traitement du flux de données

**3. Détection et suppression de tendances dans nos séries.**

| | | |
|---|---|---|
| - Série de référence **dernières estimations** - Flux de données à traiter : Flux des **différentes corrections** | → Analyse de présence des **tendances** (déterministes ou stochastiques) | → Application des traitements de tendance sur chaque série de **manière séquentielle** |



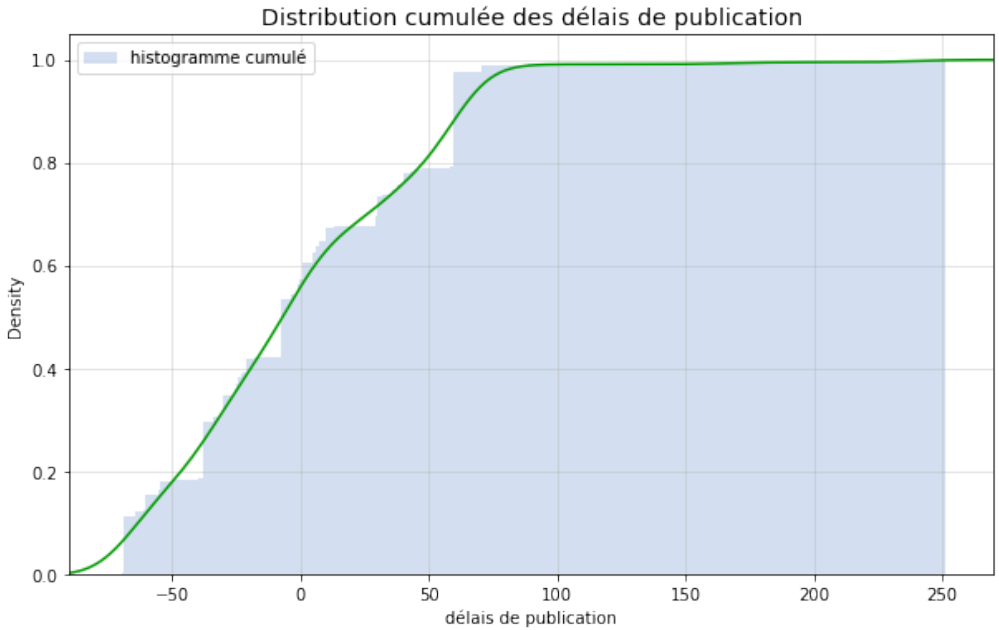

Zivot, E., & Wang, J. (2006). Unit root tests. Modeling Financial Time Series with s-plus®, 111-139.

# 1. Traitement du flux de données

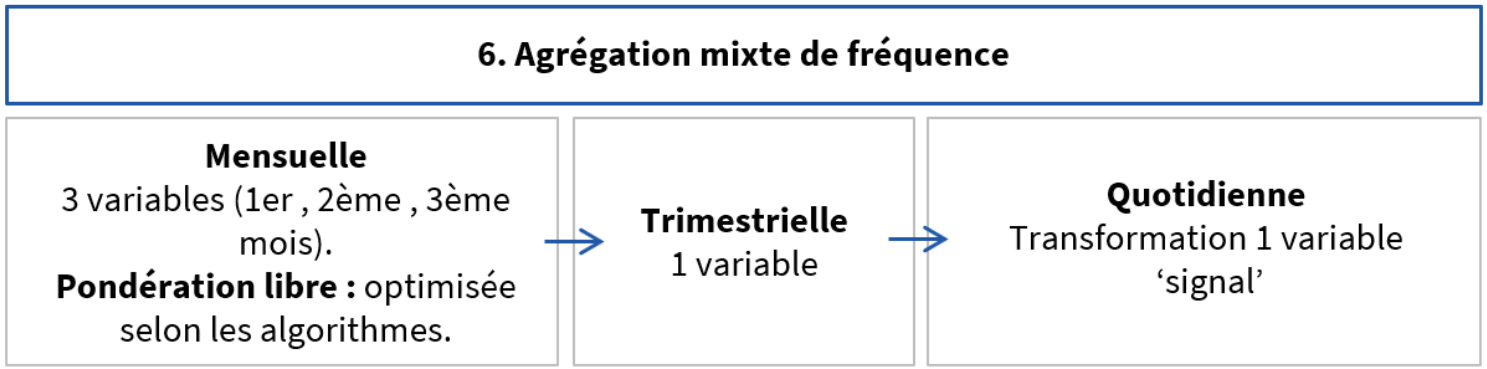
**4. Incorporation de variables décalées temporellement**

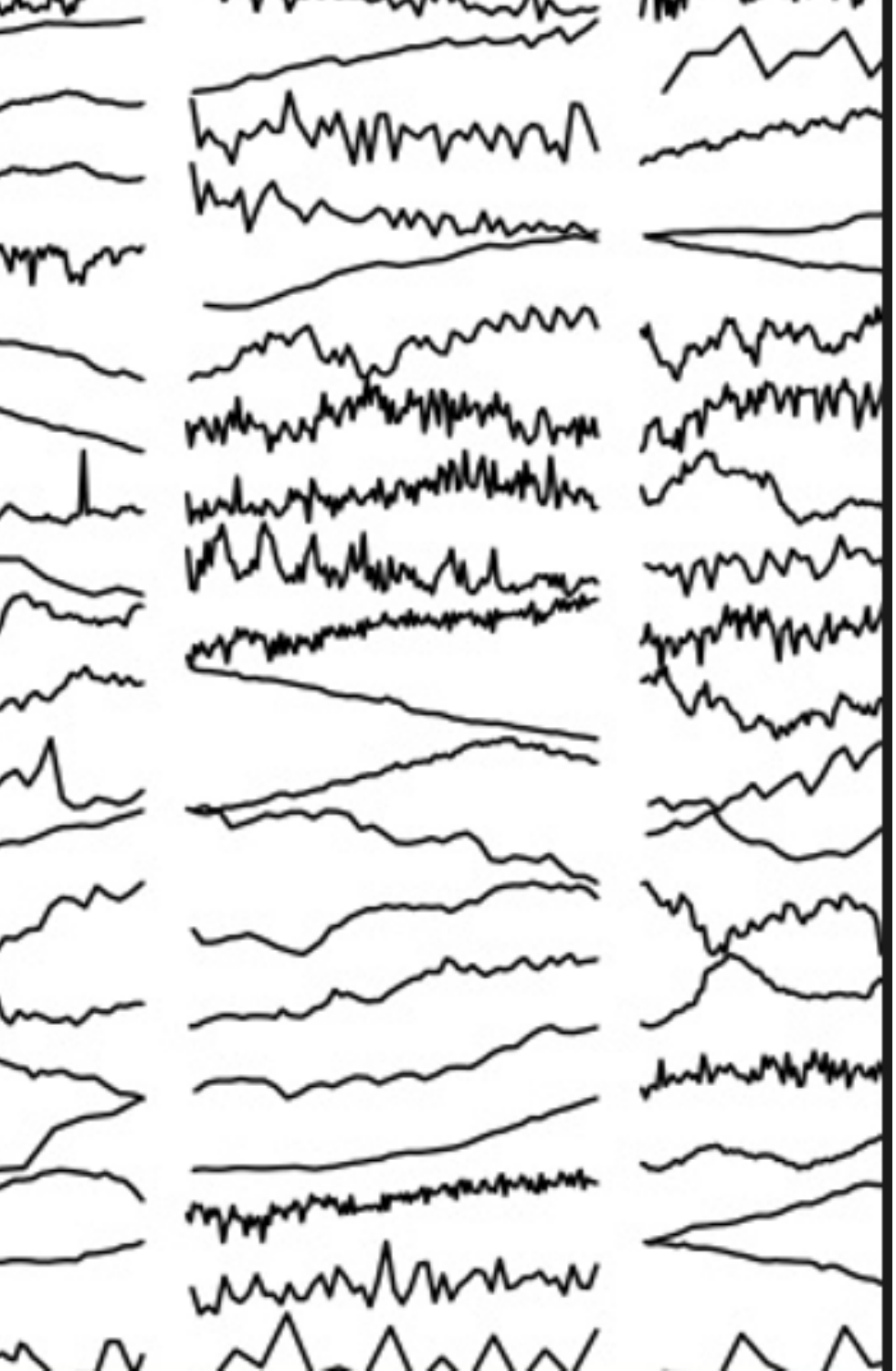
↓

**5. Filtres temporels**



# 1. Traitement du flux de données





# 2. APPROCHE DE MODELISATION EN DEUX ETAPES

# 2. Approche de modélisation en deux temps

1. **Imputation des variables non encore publiées :**
   Prévoir les données non encore publiées
   - Ensemble de variables explicatives rendu complet à tout moment.
2. **Prédiction du PIB :**
   Prédire le PIB à l'aide :
   - Des variables déjà publiées
   - Et des variables explicatives imputées à l'étape précédente.

# 2. Imputation des variables non encore publiées

Illustration factice de la nécéssité de la première étape

| Trimestre | date de disponibilité | PMI construction 1er mois | PMI construction 2ème mois | PMI construction 3ème mois | Consommation ménages -textiles - trimestre précédent | Ventes de véhicules à moteur | Climat des affaires 1er mois | Climat des affaires 2ème mois |
|---|---|---|---|---|---|---|---|---|
| Q3 2019 | 2019-10-23 | -0.23198281 | -0.28821969 | 0.054068937 | 0.053352713 | 0.501582275 | 0.194745065 | -0.142081073 |
| | 2019-10-25 | -0.23198281 | -0.28821969 | 0.054068937 | 0.053352713 | 0.501582275 | 0.194745065 | -0.142081073 |
| | 2019-10-29 | -0.23198281 | -0.28821969 | 0.054068937 | 0.053352713 | 0.501582275 | 0.194745065 | -0.142081073 |
| | 2019-10-30 | -0.23198281 | -0.28821969 | 0.054068937 | 0.053352713 | 0.501582275 | 0.194745065 | -0.142081073 |
| | 2019-10-01 | | | | -0.302332043 | | | |
| | 2019-10-04 | | | | -0.302332043 | | | |
| | 2019-10-07 | | | | -0.302332043 | | | |
| | 2019-10-08 | | | | -0.302332043 | | | |
| | 2019-10-09 | | | | -0.302332043 | | | |
| | 2019-10-10 | | | | -0.302332043 | | | |
| | 2019-10-15 | | | | -0.302332043 | | | |
| | 2019-10-22 | | | | -0.302332043 | | | |
| | 2019-10-23 | | | | -0.302332043 | | | |
| | 2019-10-24 | | | | -0.302332043 | | 0.065493102 | |
| | 2019-10-25 | | | | -0.302332043 | | 0.065493102 | |
| | 2019-10-28 | | | | -0.302332043 | | 0.065493102 | |
| | 2019-10-30 | | | | -0.302332043 | | 0.065493102 | |
| | 2019-10-31 | | | | -0.302332043 | | 0.065493102 | |
| | 2019-11-01 | -0.733100539 | | | -0.302332043 | | 0.065493102 | |
| | 2019-11-04 | -0.733100539 | | | -0.302332043 | | 0.065493102 | |
| | 2019-11-05 | -0.733100539 | | | -0.302332043 | | 0.065493102 | |
| | 2019-11-07 | -0.733100539 | | | -0.302332043 | | 0.065493102 | |
| | 2019-11-08 | -0.733100539 | | | -0.302332043 | | 0.065493102 | |
| | 2019-11-12 | -0.733100539 | | | -0.302332043 | | 0.065493102 | |
| | 2019-11-13 | -0.733100539 | | | -0.302332043 | | 0.065493102 | |
| | 2019-11-14 | -0.733100539 | | | -0.302332043 | | 0.065493102 | |
| | 2019-11-20 | -0.733100539 | | | -0.302332043 | | 0.065493102 | |
| Q4 2019 | 2019-11-21 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-11-22 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-11-26 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-11-27 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-11-28 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-11-29 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-02 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-03 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-04 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-05 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-06 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-10 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-11 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-12 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-19 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-20 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-23 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-24 | -0.733100539 | -0.014471138 | -0.72426521 | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-25 | -0.733100539 | -0.014471138 | -0.72426521 | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-27 | -0.733100539 | -0.014471138 | -0.72426521 | -0.302332043 | | 0.065493102 | -0.088057963 |

Zone nécessitant une étape d'imputation

Imputation univariée

Imputation multivariée

# 2. Imputation des variables non encore publiées

## Illustration factice de la nécéssité de la première étape

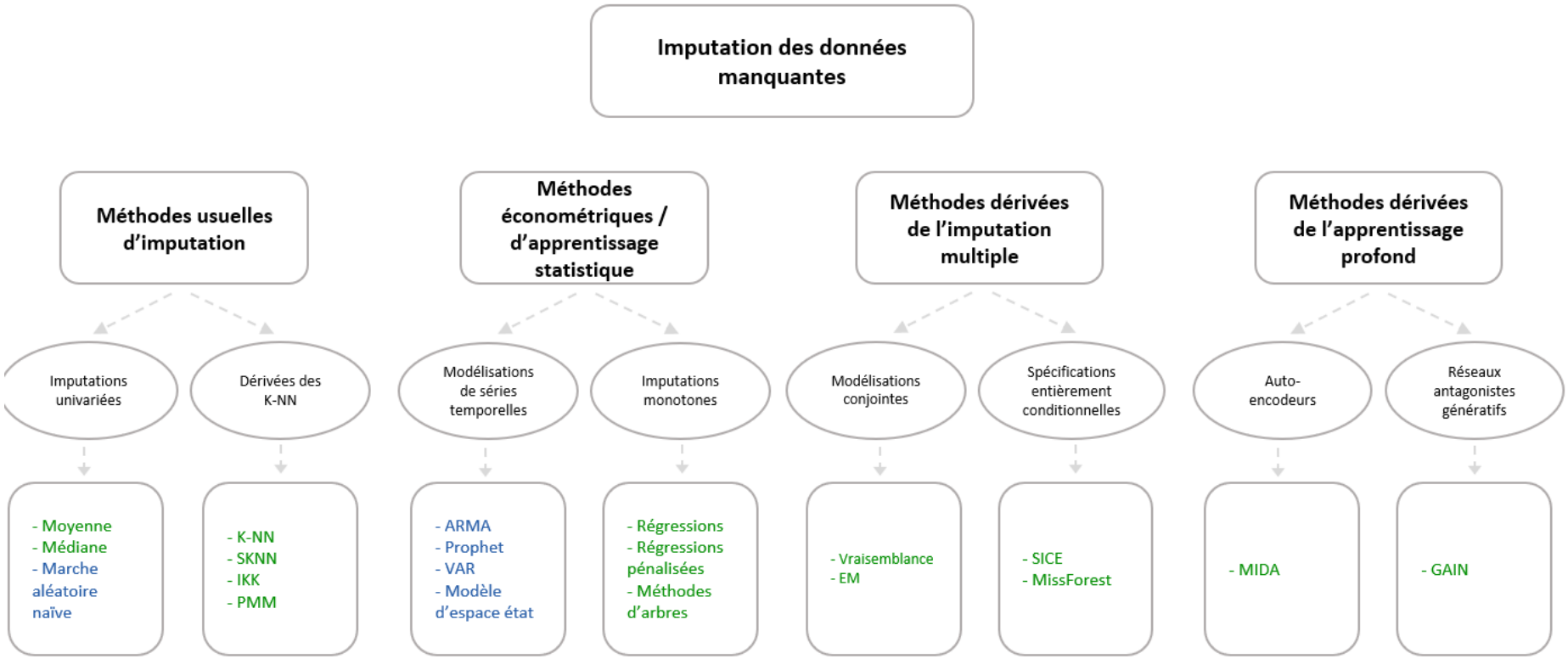
| Trimestre | date de disponibilité | PMI construction 1er mois | PMI construction 2ème mois | PMI construction 3ème mois | Consommation ménages -textiles - trimestre précédent | Ventes de véhicules à moteur | Climat des affaires 1er mois | Climat des affaires 2ème mois |
|---|---|---|---|---|---|---|---|---|
| Q3 2019 | 2019-10-23 | -0.23198281 | -0.28821969 | 0.054068937 | 0.053352713 | 0.501582275 | 0.194745065 | -0.142081073 |
| | 2019-10-25 | -0.23198281 | -0.28821969 | 0.054068937 | 0.053352713 | 0.501582275 | 0.194745065 | -0.142081073 |
| | 2019-10-29 | -0.23198281 | -0.28821969 | 0.054068937 | 0.053352713 | 0.501582275 | 0.194745065 | -0.142081073 |
| | 2019-10-30 | -0.23198281 | -0.28821969 | 0.054068937 | 0.053352713 | 0.501582275 | 0.194745065 | -0.142081073 |
| | 2019-10-01 | | | | -0.302332043 | | | |
| | 2019-10-04 | | | | -0.302332043 | | | |
| | 2019-10-07 | | | | -0.302332043 | | | |
| | 2019-10-08 | | | | -0.302332043 | | | |
| | 2019-10-09 | | | | -0.302332043 | | | |
| | 2019-10-10 | | | | -0.302332043 | | | |
| | 2019-10-15 | | | | -0.302332043 | | | |
| | 2019-10-22 | | | | -0.302332043 | | | |
| | 2019-10-23 | | | | -0.302332043 | | | |
| | 2019-10-24 | | | | -0.302332043 | | 0.065493102 | |
| | 2019-10-25 | | | | -0.302332043 | | 0.065493102 | |
| | 2019-10-29 | | | | -0.302332043 | | 0.065493102 | |
| | 2019-10-30 | | | | -0.302332043 | | 0.065493102 | |
| | 2019-10-31 | | | | -0.302332043 | | 0.065493102 | |
| | 2019-11-01 | -0.733100539 | | | -0.302332043 | | 0.065493102 | |
| | 2019-11-04 | -0.733100539 | | | -0.302332043 | | 0.065493102 | |
| | 2019-11-05 | -0.733100539 | | | -0.302332043 | | 0.065493102 | |
| | 2019-11-07 | -0.733100539 | | | -0.302332043 | | 0.065493102 | |
| | 2019-11-08 | -0.733100539 | | | -0.302332043 | | 0.065493102 | |
| | 2019-11-12 | -0.733100539 | | | -0.302332043 | | 0.065493102 | |
| | 2019-11-13 | -0.733100539 | | | -0.302332043 | | 0.065493102 | |
| | 2019-11-14 | -0.733100539 | | | -0.302332043 | | 0.065493102 | |
| | 2019-11-20 | -0.733100539 | | | -0.302332043 | | 0.065493102 | |
| Q4 2019 | 2019-11-21 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-11-22 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-11-26 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-11-27 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-11-28 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-11-29 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-02 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-03 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-04 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-05 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-06 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-10 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-11 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-12 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-19 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-20 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-23 | -0.733100539 | -0.014471138 | | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-24 | -0.733100539 | -0.014471138 | -0.72426521 | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-25 | -0.733100539 | -0.014471138 | -0.72426521 | -0.302332043 | | 0.065493102 | -0.088057963 |
| | 2019-12-27 | -0.733100539 | -0.014471138 | -0.72426521 | -0.302332043 | | 0.065493102 | -0.088057963 |

Zone nécessitant une étape d'imputation

Imputation univariée

Imputation multivariée

# 2. Imputation des variables non encore publiées





Bertsimas, D., Pawlowski, C., & Zhuo, Y. D. (2017). From predictive methods to missing data imputation : an optimization approach. The Journal of Machine Learning Research

Pereira, R. C., Santos, M. S. & Rodrigues, P. P.(2020). Reviewing autoencoders for missing data imputation : Technical trends, applications and outcomes. Journal of Artificial Intelligence Research

Van Buuren, S. (2018). Flexible imputation of missing data. CRC press.

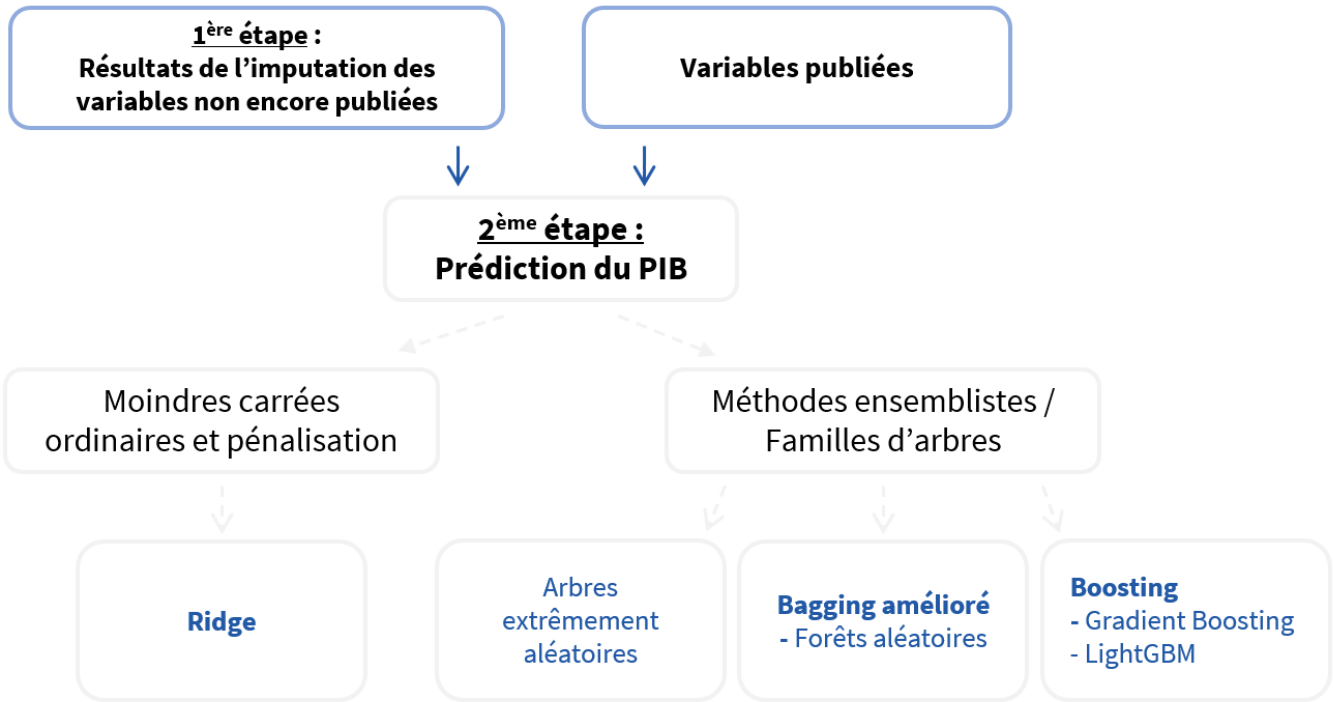
# 2. Prédiction du PIB

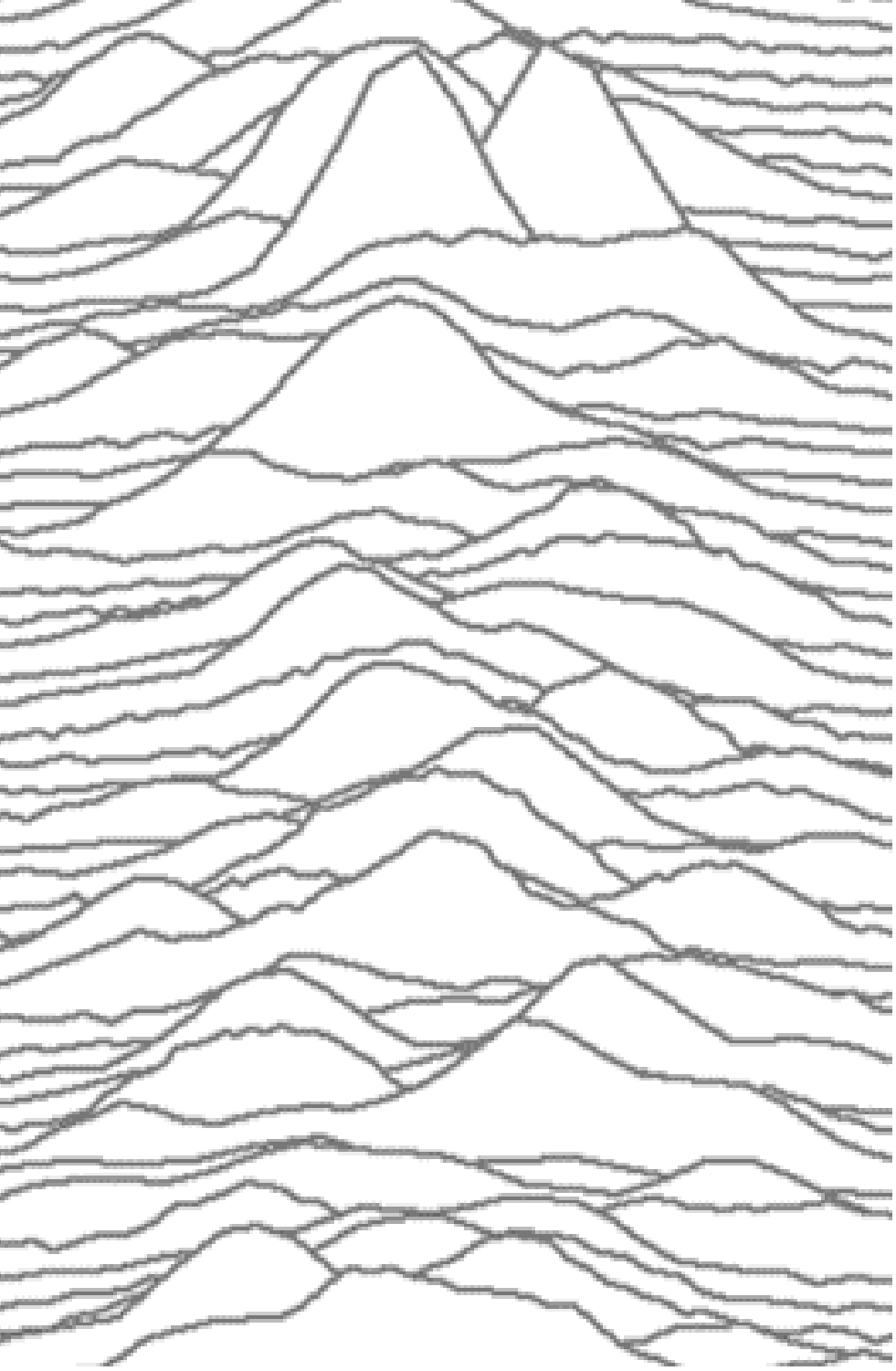
**Optimisation du paramètre 'modèle de prédiction'**

Relations de nature différente

Diverses agrégations de sous modèles

Méthodes distinctes d'échantillonnage de données etc.



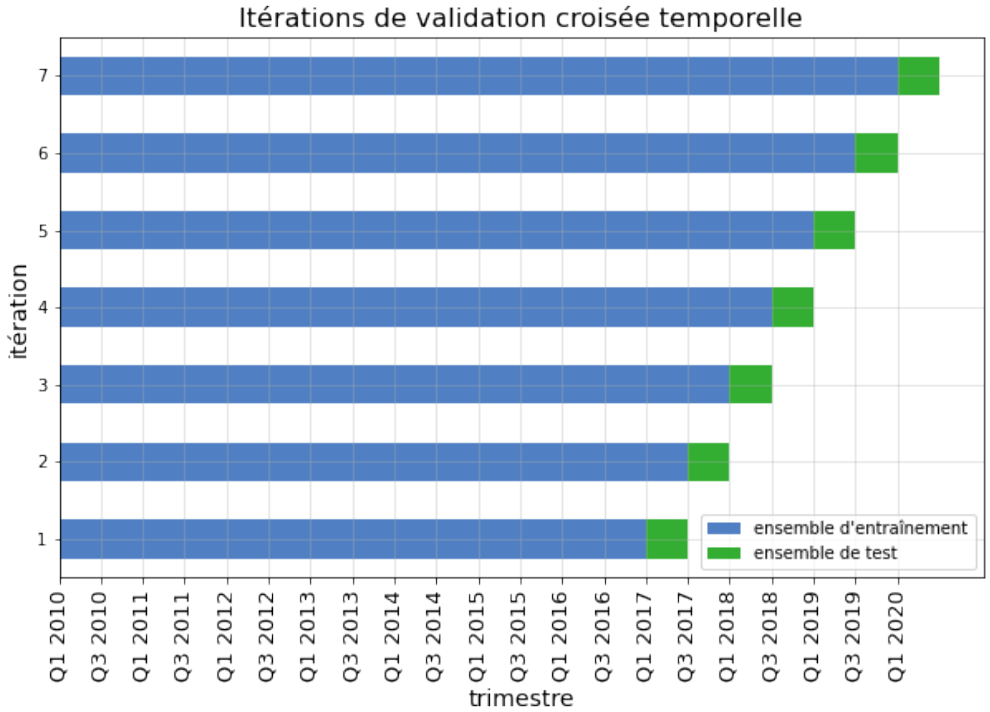


# 3. RESULTATS

# 3. Résultats : Procédure d'évaluation

1. **Métriques intrinsèques**
   Mesures de performance d'imputation des variables non encore publiées
   - MAE, RMSE, etc.

2. **Métriques extrinsèques**
   Mesures de performance de prédiction du PIB
   - MAFE, RMSFE, F-Signe (récession vs croissance), etc.



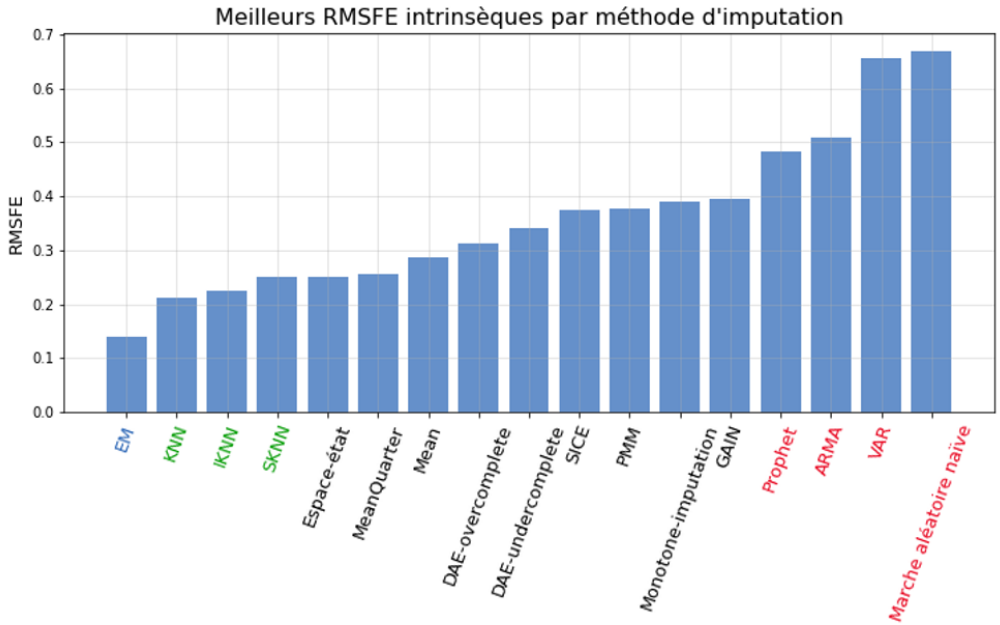


Bertsimas, D., Pawlowski, C., & Zhuo, Y. D. (2017). From predictive methods to missing data imputation : an optimization approach. The Journal of Machine Learning Research

Cerqueira, V., Torgo, L., Smailovic, J., & Mozetic, I. (2017, October). A comparative study of performance estimation methods for time series forecasting. In 2017 IEEE international conference on data science and advanced analytics (DSAA) (pp. 529-538). IEEE.

## Performances intrinsèques

| Méthode d'imputation | Famille | Méthode de sélection | Nombre de variables | Mafe | Rmsfe |
|---|---|---|---|---|---|
| EM | JM | score | 150 | **0.095** | **0.138** |
| EM | JM | corrélation | 150 | 0.145 | 0.190 |
| KNN | KNN | corrélation | 30 | 0.210 | 0.211 |
| IKNN | KNN | corrélation | 30 | 0.224 | 0.224 |
| SKNN | KNN | corrélation | 30 | 0.249 | 0.250 |



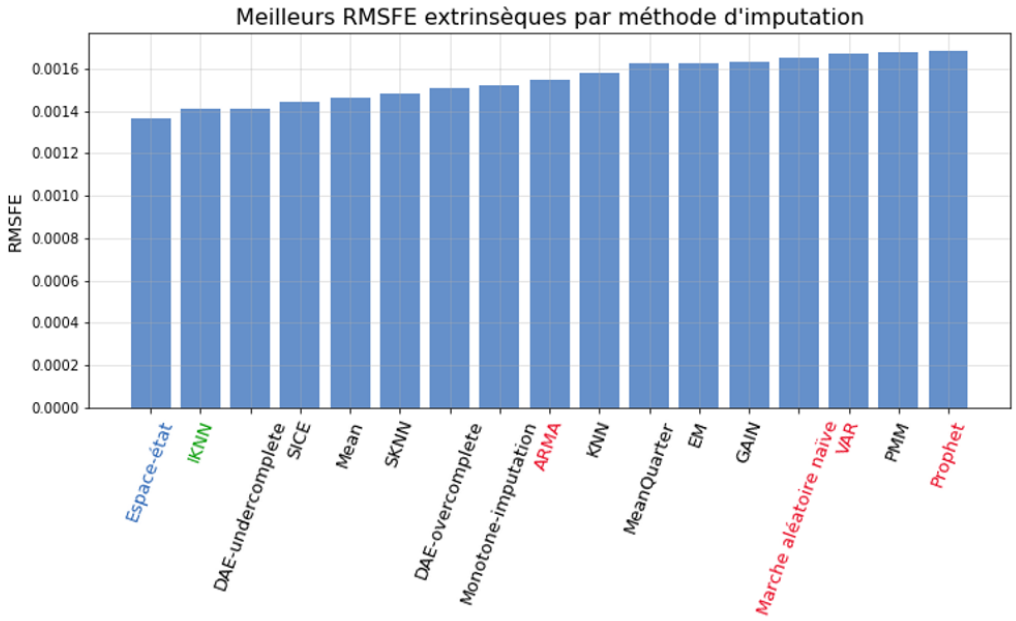

### Remarques

- Sur-performance des méthodes de la famille des K-NN
- Sous-performance des méthodes de prévisions temporelles

# 3. Résultats : Hors crise covid-19

## Performances extrinsèques

| Méthode d'imputation | Famille | Méthode de prédiction | Méthode de sélection | Nombre de variables | mafe | rmsfe | signe |
|---|---|---|---|---|---|---|---|
| Espace-état | Séries temporelles | Arbres extrêmement aléatoires | Score | 150 | **0.00135** | **0.00137** | 0.83 |
| IKNN | KNN | Forêt aléatoire | Corrélation | 150 | 0.00139 | 0.00141 | **0.91** |
| DAE sur-complet | DAE | Forêt aléatoire | Score | 150 | 0.00139 | 0.00142 | 0.83 |
| IKNN | KNN | Forêt aléatoire | Score | 150 | 0.00143 | 0.00143 | 0.83 |
| SICE | FCS | Forêt aléatoire | Corrélation | 150 | 0.00143 | 0.00145 | 0.85 |

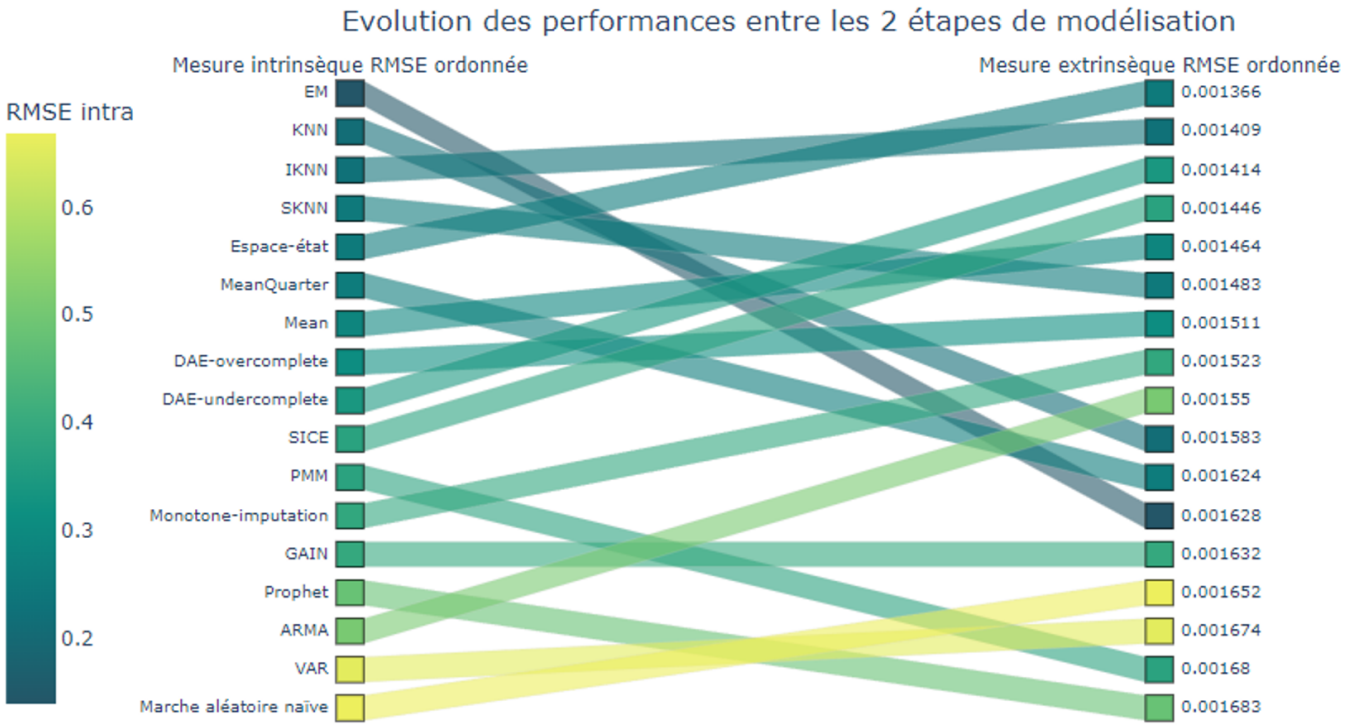



## Remarques

- Prééminence de la modélisation espace-état et IKNN
  Signe de prédiction IKNN > signe de prédiction espace-état
- Sous-performance des méthodes de prévisions temporelles

# 3. Résultats : Hors crise covid-19

**Evolution de la mesure de performance des modèles d'imputation au cours des 2 étapes de modélisation**

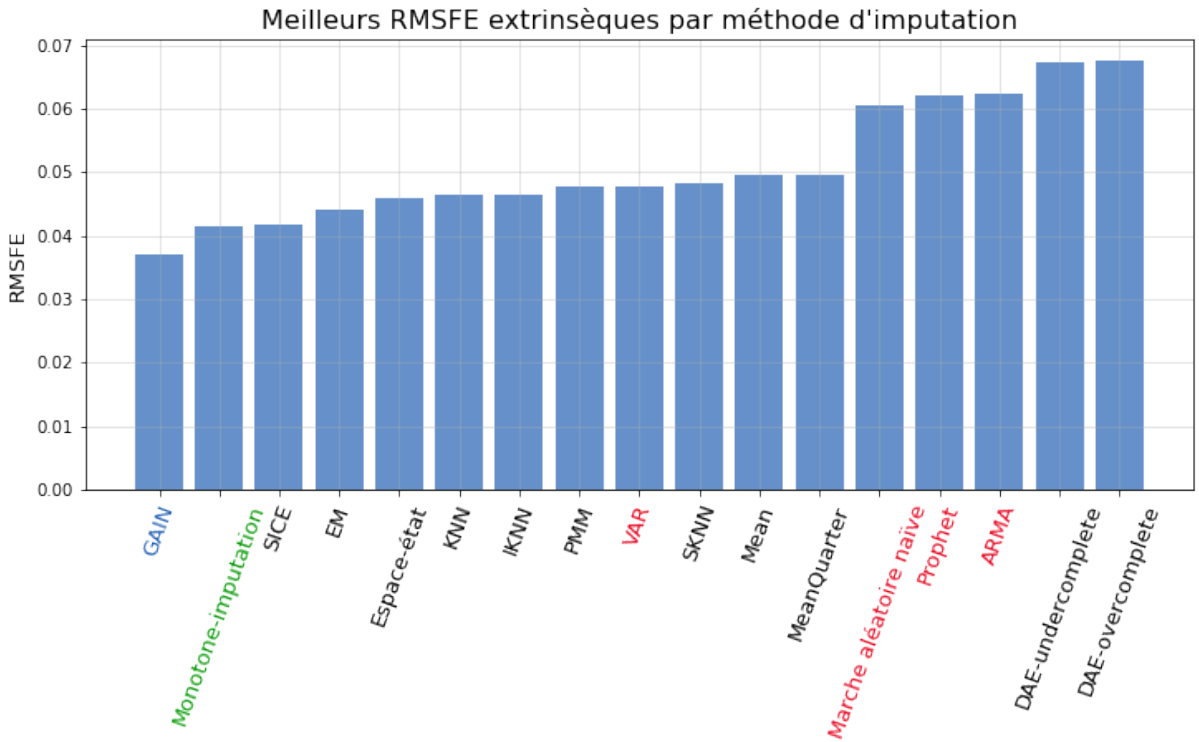



## Remarques

- Modèles à considérer : IKNN et modélisation espace-état

# 3. Résultats : Début crise covid-19

## Performances extrinsèques



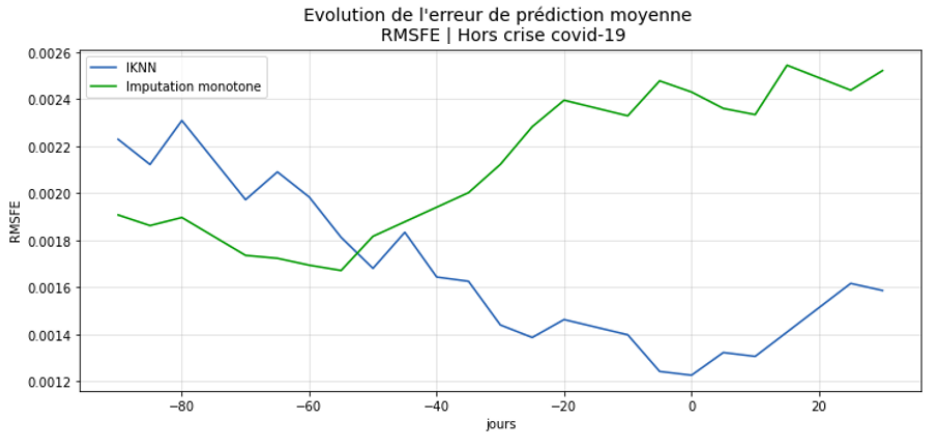

### Remarques

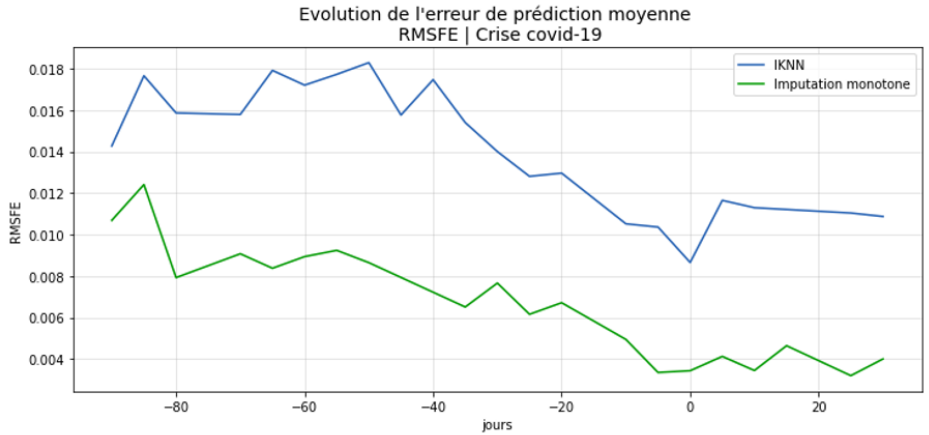
- Facteur multiplicatif de 10 entre les performances covid-19 et hors covid-19
- Changement notable dans le classement des méthodes d'imputation

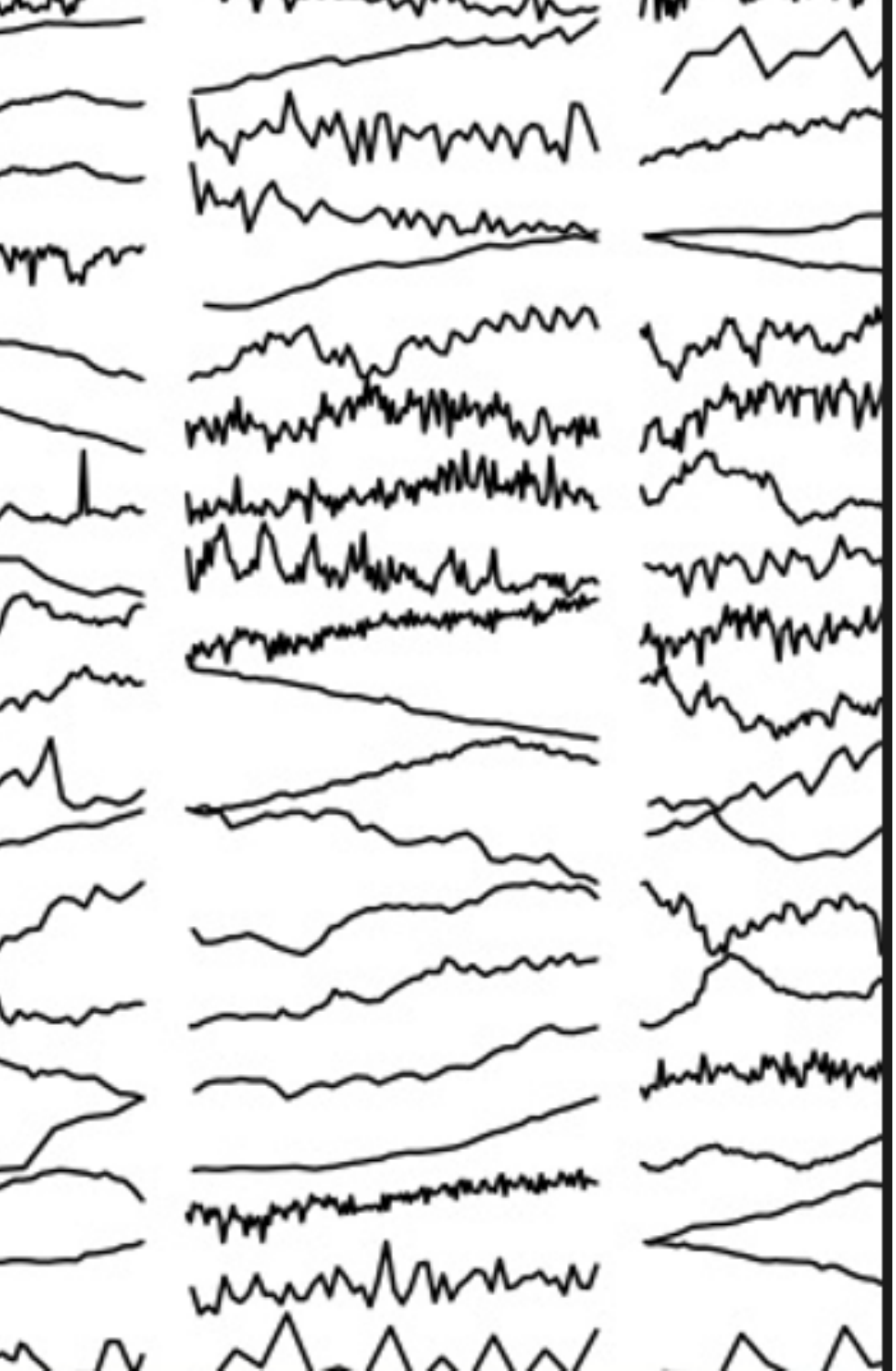
**Prééminence du K-NN itératif et de l'imputation monotone**

## Modèle d'intérêt par période

- Période "Business as usual" :
  - Imputation IKNN
  - Méthode de prédiction de forêt aléatoire
- Période début covid-19 :
  - Imputation monotone
  - Méthode de prédiction de régression Ridge





# CONCLUSION ET VOIES D'AMELIORATION

# Conclusion et voies d'amélioration

## Bilan

- Méthodes d'imputation introduites surperforment les méthodes de prévisions temporelles (**IKNN** et l'**imputation monotone**)

## Voies d'amélioration

- Ajout de données alternatives (Ferrara et al., 2020 [5])
- Estimation des variables non encore publiées :
  - Surpondérer les variables explicatives issues de la même famille économique.
- Plus grande étendue du nombre de variables sélectionnées :
  - Davantage étudier la sensibilité au paramètre du nombre de variables sélectionnées
- Incorporer des modèles qui imputent et régressent simultanément :
  - Modèles BRITS (Cao et al., 2018 [3])

---

Cao, W., Wang, D., Li, J., Zhou, H., Li, L., & Li, Y. (2018). Brits : Bidirectional recurrent imputation for time series. Advances in neural information processing systems, 31.
Ferrara, L., & Simoni, A. (2020). When are Google data useful to nowcast GDP ? An approach via pre-selection and shrinkage. arXiv preprint arXiv :2007.00273.

# Références


Bertsimas, D., Pawlowski, C., & Zhuo, Y. D. (2017). From predictive methods to missing data imputation : an optimization approach. The Journal of Machine Learning Research

Bouwman, K. E., & Jacobs, J. P. (2011). Forecasting with real-time macroeconomic data : the ragged-edge problem and revisions. Journal of Macroeconomics

Cao, W., Wang, D., Li, J., Zhou, H., Li, L., & Li, Y. (2018). Brits : Bidirectional recurrent imputation for time series. Advances in neural information processing systems, 31.

Cerqueira, V., Torgo, L., Smailovic, J., & Mozetic, I. (2017, October). A comparative study of performance estimation methods for time series forecasting. In 2017 IEEE international conference on data science and advanced analytics (DSAA) (pp. 529-538). IEEE.

Ferrara, L., & Simoni, A. (2020). When are Google data useful to nowcast GDP ? An approach via pre-selection and shrinkage. arXiv preprint arXiv :2007.00273.

Miller, P. J., & Chin, D. M. (1996). Using monthly data to improve quarterly model forecasts. Federal Reserve Bank of Minneapolis Quarterly Review

Pereira, R. C., Santos, M. S. & Rodrigues, P. P.(2020). Reviewing autoencoders for missing data imputation : Technical trends, applications and outcomes. Journal of Artificial Intelligence Research

Tolvi, J. (2001). Outliers in eleven finnish macroeconomic time series. Finnish Economic Papers, 14(1), 14-32.

Van Buuren, S. (2018). Flexible imputation of missing data. CRC press.

Zheng, I. Y., & Rossiter, J. (2006). Using monthly indicators to predict quarterly GDP.

Zivot, E., & Wang, J. (2006). Unit root tests. Modeling Financial Time Series with s-plus®, 111-139.

# Nowcasting PIB : Imputation des données non encore publiées

Marion CABROL[1], Kevin FERNANDES[1]
Michel MARTINEZ[2]

[1] Core data science team - Société Générale CIB
[2] Chef économiste zone euro - Société Générale CIB

31 mars 2022