

Module 06 - Séries chronologiques : synthèse

MQT-2101 - Analyse et modélisation des données

Objectifs du module

- Préparer les données et définir la cible.
- Comparer aux références sur un futur simulé.
- Lire erreurs et incertitude.
- Choisir et suivre une méthode.

1. Définir une prévision utile

Quelle information faut-il pour préparer le suivi énergétique du prochain trimestre?

Un tableau ordonné dans le temps

TRANSVERSAL

Plusieurs unités, une période

CHRONOLOGIQUE

Une quantité, plusieurs dates

Le cas guidé

72

mois, de 2020 à 2025

1

bâtiment fictif

kWh

consommation mensuelle d'électricité

Données simulées. Une ligne représente un mois; aucun bâtiment réel n'est décrit.

Origine, cible et horizon

ORIGINE T

Dernière information disponible

HORIZON H

Nombre de périodes à prévoir

$\hat{y}_{T+h|T}$: prévision de y_{T+h} produite à l'origine T .

Fréquence et granularité

FRÉQUENCE

Jour, semaine, mois

GRANULARITÉ

Succursale, région, produit

L'expertise complète les mesures

OPÉRATIONS

Fermeture, chauffage, panne

GESTION

Occupation, horaires, travaux

Le processus est une boucle

1

Cadrer et préparer

2

Comparer et
choisir

3

Prévoir et suivre

Préciser la décision

ÉNERGIE

Consommation et horizon mensuel

BUDGET

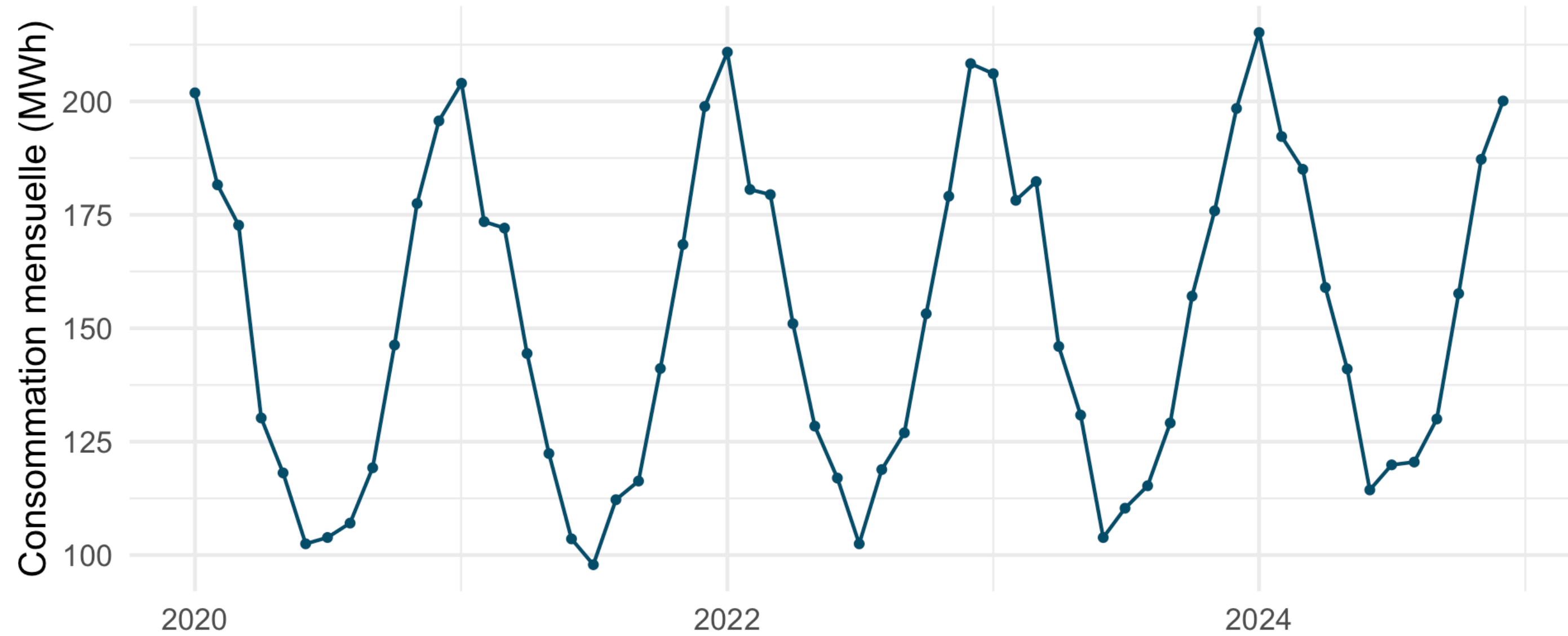
Tarifs et postes de facturation

Le coût d'une sous-prévision peut différer du coût d'une sur-prévision.

2. Lire les composantes d'une série

La hausse vient-elle du temps ou du calendrier?

Respecter les axes



Explorer ici 2020 à 2024. Garder 2025 pour la validation.

Niveau et tendance

NIVEAU

Valeur de référence locale

TENDANCE

Évolution persistante

Saisonnalité, cycle et reste

Calendrier

motif de période connue

Cycle

fluctuation sans période fixe

Reste

variation non expliquée

Additif ou multiplicatif?

ADDITIF

$$y_t = T_t + S_t + R_t$$

MULTIPLICATIF

$$y_t = T_t \times S_t \times R_t$$

T_t regroupe ici niveau, tendance et cycle.

Montant ou pourcentage?

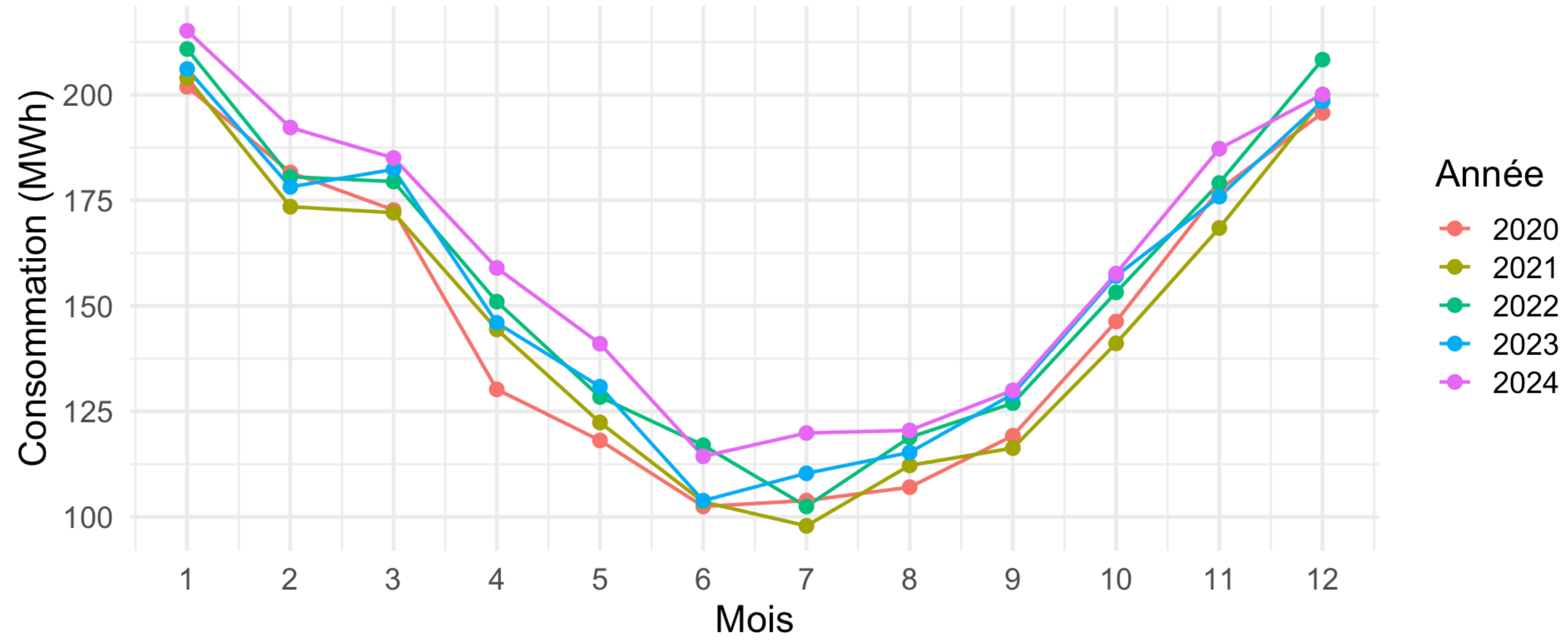
+20 000 KWH

même amplitude à deux niveaux

+10 %

amplitude croissante avec le niveau

Comparer les mêmes mois



Zoomer et changer d'échelle

```
1 zoom <- apprentissage |> filter(date >= as.Date("2023-01-01"))  
2 graphique_serie(zoom)  
3 graphique_serie(apprentissage) + scale_y_log10()
```

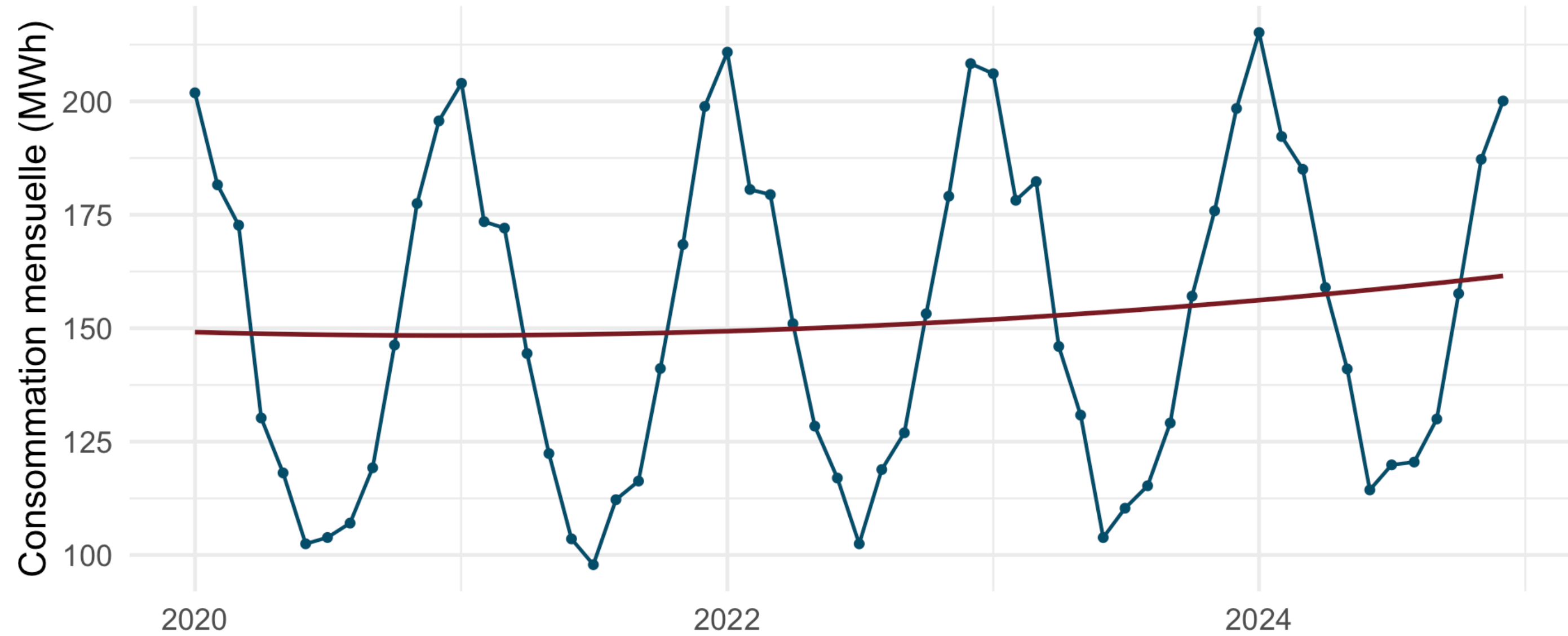


L'objet ts et la fenêtre

```
1 serie <- ts(apprentissage$consommation_kwh, start = c(2020, 1), frequency = 12)
2 zoom_ts <- window(serie, start = c(2023, 1), end = c(2024, 12))
```



Une tendance descriptive



3. Préparer les données temporelles

Un mois absent peut-il être traité comme un mois observé?

Quatre vérifications

Calendrier

doublons et périodes absentes

Mesure

NA, zéro, unités

Contexte

extrêmes et ruptures

Mensuel ne signifie pas trente jours

```
1 dates_attendues <- seq.Date(min(apprentissage$date),  
2                             max(apprentissage$date), by = "month")  
3 setdiff(dates_attendues, apprentissage$date)  
4 anyDuplicated(apprentissage$date)
```



Valeur manquante ou période absente

DATE PRÉSENTE, NA

La mesure manque

DATE ABSENTE

Le calendrier a un trou

Retirer une ligne ne rend pas les périodes voisines consécutives.

Imputer exige une règle

HISTORIQUE

estimer avec l'information déjà disponible

SENSIBILITÉ

comparer plusieurs traitements

Espacement irrégulier

ÉVÉNEMENTS

arrivées à des heures variables

AGRÉGATION

compte par jour ou semaine

Extrême : erreur ou événement?

ERREUR VÉRIFIÉE

corriger et documenter

ÉVÉNEMENT RÉEL

conserver ou modéliser le contexte

Sans justification, ne pas supprimer automatiquement.

Combien d'années utiliser?

HISTORIQUE LONG

davantage de répétitions

HISTORIQUE RÉCENT

meilleure pertinence après une rupture

Les variables externes arrivent-elles à temps?

CONNUES À L'ORIGINE

calendrier, fermeture planifiée

INCONNUES

prix réalisé, demande future

4. Comparer des prévisions de référence

Une méthode élaborée fait-elle mieux qu'une règle simple?

Bien ajuster n'est pas bien prévoir

APPRENTISSAGE

paramètres et structure

VALIDATION

erreurs sur le futur simulé

La séparation du cas guidé

60

mois d'apprentissage : 2020 à 2024

12

mois de validation : 2025

T

origine : décembre 2024

Le naïf à origine fixe

$$\hat{y}_{T+h|T} = y_T, \quad h \geq 1.$$

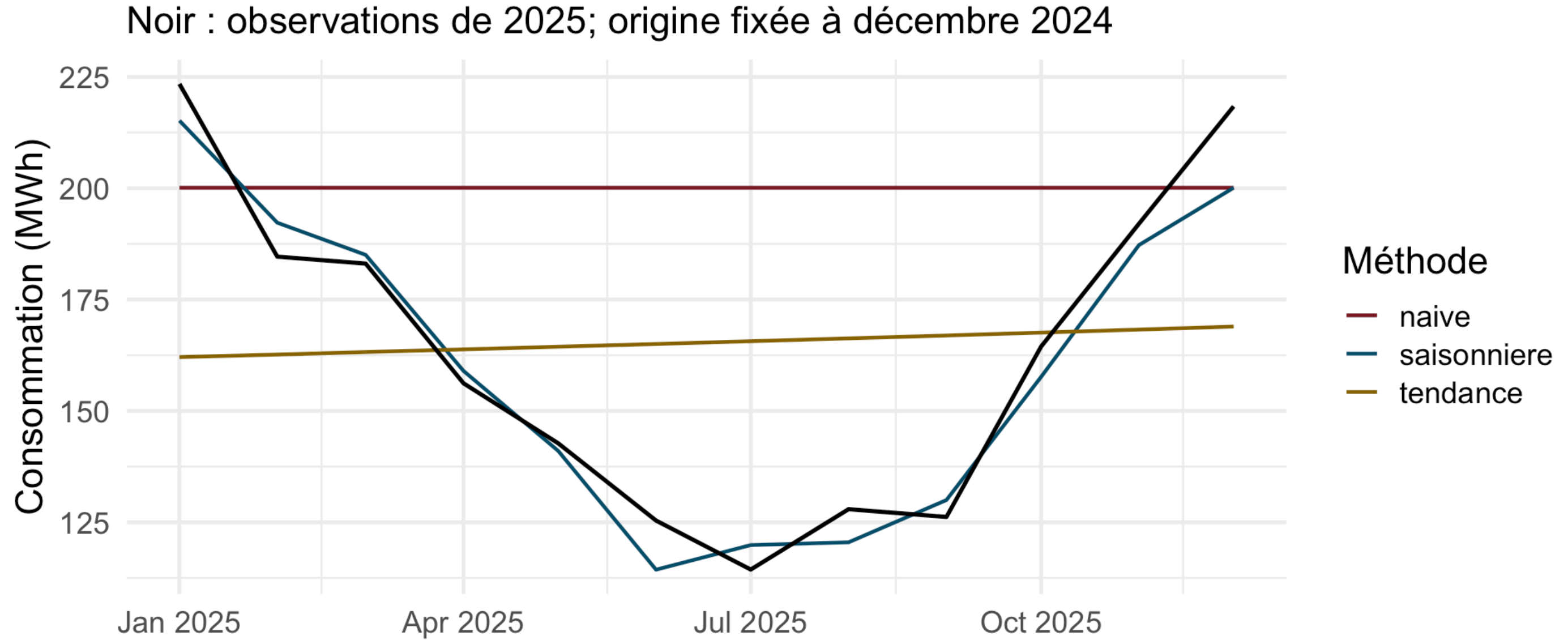
Une origine en décembre 2024. La dernière consommation mensuelle connue est répétée pour tous les mois de 2025.

Le naïf saisonnier

$$\hat{y}_{T+h|T} = y_{T+h-m}, \quad 1 \leq h \leq m.$$

$m = 12$: reprendre les douze mois de la dernière année.

Comparer sur les mêmes mois



Du code à la référence

```
1 refs <- prevoir_references(apprentissage$consommation_kwh, horizon = 12, periode = 12)
2 validation |> mutate(naive = refs$naive,
3                     saisonniere = refs$saisonniere)
```



Choisir la période selon l'horizon

UN MOIS

répéter plusieurs origines à $h = 1$

UN AN

évaluer aussi les horizons jusqu'à 12

Réajuster avant la vraie prévision

1

Évaluer sur 2025

2

Fixer le choix

3

Réajuster sur 2020
à 2025

4

Prévoir 2026

5. Mesurer et lire les erreurs

Une erreur moyenne proche de zéro suffit-elle?

Une convention de signe

$$e_{T+h} = y_{T+h} - \hat{y}_{T+h|T}.$$

E > 0

sous-prévision

E < 0

sur-prévision

ME, MAE et RMSE

$$\text{ME} = \frac{1}{n} \sum e_i, \quad \text{MAE} = \frac{1}{n} \sum |e_i|.$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum e_i^2}.$$

Même unité que la cible. MAE est aussi appelée MAD.

Les erreurs peuvent s'annuler

-20, +20

deux erreurs, en unités

ME = 0

aucun biais moyen ici

MAE = 20

erreur absolue persistante

RMSE = 20 également dans cet exemple.

Le MAPE et ses limites

$$\text{MAPE} = \frac{100}{n} \sum \left| \frac{e_i}{y_i} \right|.$$

Zéro : indéfini. Valeur proche de zéro : instabilité. Zéro arbitraire : pourcentage inadapté.

La MASE : une échelle d'apprentissage

$$q_m = \frac{1}{T - m} \sum_{t=m+1}^T |y_t - y_{t-m}|, \quad \text{MASE} = \frac{\text{MAE}_{\text{validation}}}{q_m}.$$

$m = 1$: référence naïve. $m = 12$: référence saisonnière mensuelle.

Que signifie $MASE < 1$?

INTERPRÉTATION

MAE plus petite que q_m

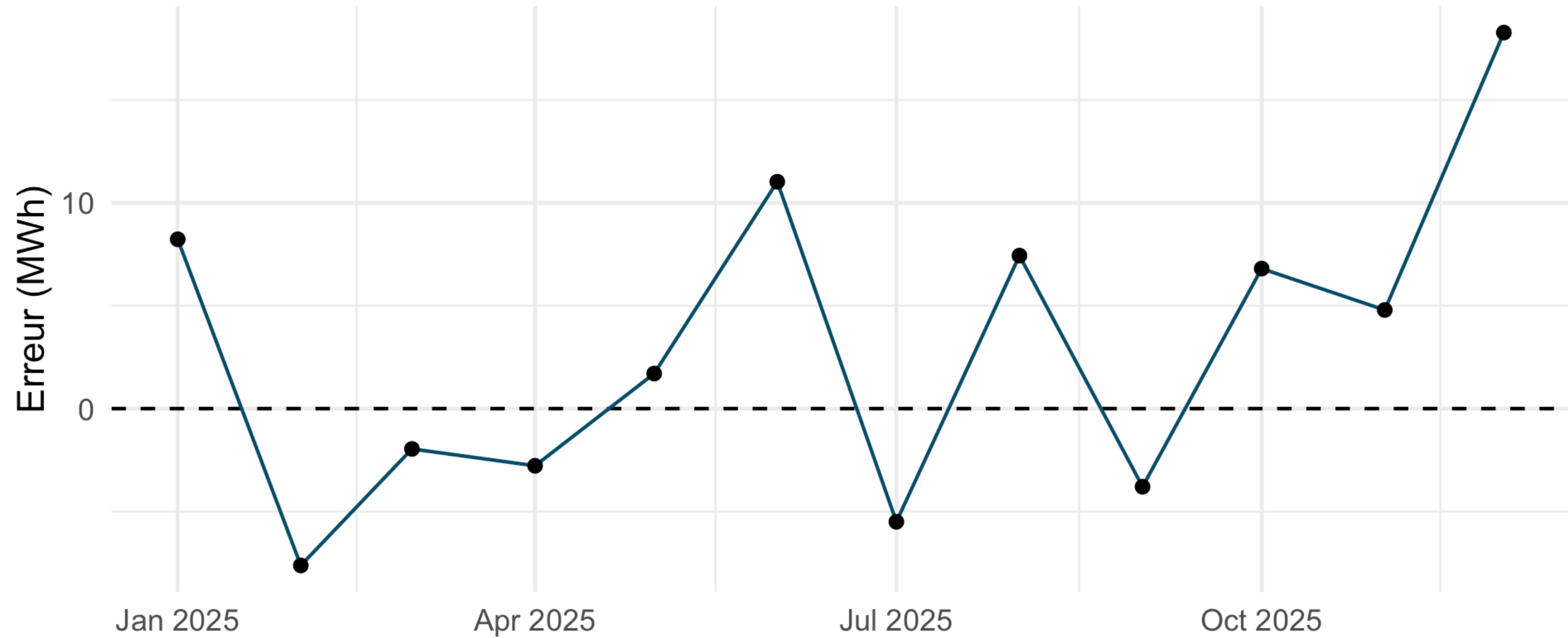
LIMITE

pas une victoire garantie sur la validation

Le résultat du cas guidé

Méthode	ME (kWh)	MAE (kWh)	RMSE (kWh)	MAPE (%)	MASE
naive	-36 878,67	43 810,17	51 298,07	31,69	6,71
saisonniere	3 052,50	6 662,33	7 987,13	4,11	1,02
tendance	-2 172,18	31 573,64	35 965,65	20,50	4,84

Lire le temps et la distribution



6. Communiquer l'incertitude

Une seule valeur suffit-elle au suivi énergétique?

Un point et une plage

POINT

Une valeur prévue

INTERVALLE

Une plage avec un niveau nominal

Interpréter 95 %

Sur des répétitions comparables, une procédure calibrée à 95 % couvre la future observation dans environ 95 % des cas.

Une approximation normale

$$\hat{y}_{T+h|T} \pm 1,96 \hat{\sigma}_h.$$

Pour 95 %, si l'approximation normale et l'écart-type prédictif sont appropriés.

Plus loin, plus incertain?

NIVEAU NOMINAL PLUS ÉLEVÉ

intervalle généralement plus large

HORIZON PLUS LOINTAIN

incertitude souvent accrue

Lire un cône de prédiction

Prévisions et intervalles nominaux

Illustration ETS(A,A,N), ajustée uniquement sur 2020 à 2024



À $h = 12$, la borne à 95 % devient négative : limite du modèle pour une consommation d'énergie.

Lire les trois lettres ETS

E

type d'erreur

T

type de tendance

S

type de saisonnalité

A : additif. M : multiplicatif. N : absence de tendance ou saisonnalité. Z : choix automatique dans `forecast::ets()`.

Comparer des cônes sans choisir à l'œil

ANN, AAN, AAA

trois structures illustratives

VALIDATION

comparer erreurs et couverture

Relier l'intervalle à une décision

CAPACITÉ

risque de dépassement

BUDGET

marge et scénario défavorable

Complément A - Validation à plusieurs origines

Le choix reste-t-il bon à un mois et à six mois?

Faire avancer l'origine

1

Déc. 2024 → janv.
2025

2

Janv. 2025 → févr.
2025

3

Févr. 2025 → mars
2025

Fixe ou mobile?

FIXE

Douze horizons, une seule origine

MOBILE

Un même horizon, plusieurs origines

Combien de comparaisons?

$h = 1$

12 origines en 2025

$h = 6$

7 origines admissibles

$h = 12$

1 origine admissible

Le geste à chaque origine

```
1 origine <- 60
2 h <- 6
3 historique <- electricite$consommation_kwh[seq_len(origine)]
4 refs <- prevoir_references(historique, horizon = h)
5 erreur <- electricite$consommation_kwh[origine + h] - refs$naive[h]
```



Comparer par horizon

h	methode	n	ME	MAE	RMSE
1	naive	12	1523.3	19989.0	23242.7
1	saisonniere	12	3052.5	6662.3	7987.1
6	naive	7	-20973.9	63994.1	70792.8
6	saisonniere	7	5579.0	8235.6	9450.3

Fenêtre croissante ou glissante

CROISSANTE

conserver tout le passé disponible

GLISSANTE

conserver un passé récent fixé

Réévaluer sans recycler le test

1

Fixer les
candidats

2

Comparer
plusieurs origines

3

Retenir une
méthode

4

Suivre les
nouvelles erreurs

Choisir selon la décision

BUDGET

horizon court et actualisation fréquente

APPROVISIONNEMENT

horizon imposé par le délai

Complément B - Familles de méthodes

Faut-il plus de complexité pour améliorer la décision?

Structure explicite ou adaptation locale

STRUCTURE EXPLICITE

forme et paramètres d'un modèle

ADAPTATION LOCALE

mémoire et pondération du passé

Motif global et mémoire locale

GLOBAL DANS UNE SÉRIE

une structure sur toute la période

LOCAL DANS UNE SÉRIE

une structure qui évolue

Extrapoler ou utiliser des entrées?

HISTOIRE DE LA CIBLE

naïf, lissage, ARIMA

INFORMATIONS EXTERNES

calendrier, météo, occupation

Disponible au moment de prévoir

CALENDRIER FUTUR

connu à l'origine

PRIX RÉALISÉ FUTUR

à prévoir ou scénariser

Manuel, automatisé, surveillé

AUTOMATISER

calcul et actualisation

SURVEILLER

erreurs, ruptures, qualité

Combiner les prévisions

$$\hat{y}^{\text{comb}} = w\hat{y}^{(1)} + (1 - w)\hat{y}^{(2)}, \quad 0 \leq w \leq 1.$$

Exemple simple : moyenne à poids égaux, fixée avant la comparaison.

Diversifier les erreurs

$$\text{Var}(we_1 + (1 - w)e_2) = w^2 \text{Var}(e_1) + (1 - w)^2 \text{Var}(e_2) + 2w(1 - w) \text{Cov}(e_1, e_2).$$

Erreurs peu corrélées : diversification possible. Aucun gain garanti.

Corriger en deux étapes

$$\hat{y}^{\text{final}} = \hat{y}^{\text{base}} + \hat{e}^{\text{base}}$$

Modèle de base → prévision de ses erreurs → correction

La suite du cours

Module 7

lissage et comparaison

Module 8

tendance et saisonnalité

Module 9

variables externes et autocorrélation

Produire un rapport reproductible

Une question, un calendrier vérifié, des graphiques, un protocole temporel, des erreurs et des limites.

Référence

Hyndman, R. J., et Athanasopoulos, G. (2021). Forecasting: Principles and Practice, 3e édition.

Lectures ciblées et documentation