

DÉSAISONNALISER UNE SÉRIE TEMPORELLE



8 - Les Moyennes Mobiles et la qualité de la décomposition

ALAIN QUARTIER-LA-TENTE

Objectifs de cette séquence

Objectifs : définitions et propriétés des moyennes mobiles (MM) de X11.

Après cette séquence, vous saurez :

- Les propriétés des MM
- Identifier les moyennes mobiles utilisées par X11
- Choisir d'autres MM que ceux par défaut

Questions de positionnement

Que signifie une moyenne mobile centrée et symétrique ?

Quelles sont les principales propriétés des moyennes mobiles ?

Quel type de problème affecte le traitement de la fin de série ?

Que représente le gain et le déphasage d'une moyenne mobile ?

Quels indicateurs utilisés pour évaluer la qualité de la décomposition ?

Sommaire

1. Introduction aux moyennes mobiles

1.1 Définitions

1.2 Gain et déphasage

1.3 Construction et propriétés

1.4 Fin de période

1.5 Filtre final X11

2. Choix de la longueur de la MM

3. Diagnostics sur la qualité de la décomposition

Les moyennes mobiles (1/2)

Moyenne mobile d'ordre $p + f + 1$ de coefficients (θ_i) , l'opérateur M défini par :

$$MX_t = \sum_{i=-p}^f \theta_i X_{t+i}$$

Valeur en t remplacée par une moyenne pondérée de p valeurs passées, de la valeur courante et de f valeurs futures.

Moyenne mobile simple (tous les coeffs égaux) d'ordre 3 :

$$MX_t = \frac{1}{3}(X_{t-2} + X_{t-1} + X_t)$$

Sous **R** : `rjd3filters::simple_ma(3)`.

Les moyennes mobiles (2/2)


Si $p = f$, la moyenne mobile est dite *centrée*

Si, de plus $\theta_{-i} = \theta_i$, elle est dite *symétrique*

Remarque : une moyenne d'ordre pair ne peut être centrée et symétrique.

Construire une moyenne centrée et symétrique en combinant 2 moyennes d'ordre pair.

$$\left. \begin{array}{l} M_1 X_t = \frac{1}{4}(X_{t-2} + X_{t-1} + X_t + X_{t+1}) \\ M_2 X_t = \frac{1}{4}(X_{t-1} + X_t + X_{t+1} + X_{t+2}) \end{array} \right\} \rightarrow M_{2 \times 4} = \frac{M_1 + M_2}{2}$$

Quelle serait la moyenne $M_{2 \times 12}$? Sous  :

```
rjd3filters::simple_ma(12, -5) * rjd3filters::simple_ma(12, -6).
```

Pourquoi une moyenne $M_{3 \times 3}$?

Fonction de gain, déphasage

Soit une série $X_t = e^{-i\omega t}$

La transformée de X_t par une MM sera une sinusoïde d'amplitude modifiée et présentant un déphasage :

$$M(X_t) = \rho(\omega)e^{i\varphi(\omega)}X_t$$

$\rho(\omega)$ = *Fonction de gain* de la moyenne mobile

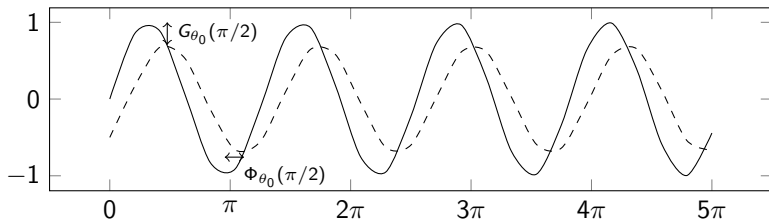
$\varphi(\omega)$ = *Fonction de déphasage* de la moyenne mobile

Une MM symétrique n'introduit pas de déphasage !

Exemple de gain et de déphasage

Lissage $X_t = \sin\left(\frac{\pi t}{3}\right)$ par $M(X_t) = \frac{X_{t-2} + X_{t-1} + X_t}{3}$ asymétrique :

$$G(\omega) = \frac{1 + 2\cos(\omega)}{3} \quad \Gamma(\omega) = -\omega$$



Construction de moyennes mobiles

Propriété évidente, la *linéarité* :

$$\begin{aligned}X_t &= TC_t + S_t + I_t \\MX_t &= M(TC_t) + M(S_t) + M(I_t)\end{aligned}$$

Trois propriétés utiles pour la désaisonnalisation :

- P1 : conservation de la tendance $M(TC_t) = TC_t$
- P2 : élimination de la saisonnalité $M(S_t) = 0$
- P3 : réduction de l'irrégulier, $\mathbb{V}[M(I_t)]$ minimale

P1 : Conservation des tendances

Toute MM **symétrique** qui conserve les constantes (resp. polynômes de degré $2d$) conserve les droites (resp. polynômes de degré $2d+1$)

La série résultante est linéaire par morceaux.

Les MM sont **sensibles aux points extrêmes** !

Exercice : les MM suivantes conservent-elles les droites ?

$$M_1[X_t] = \frac{1}{3}(X_{t-1} + X_t + X_{t+1})$$

$$M_2[X_t] = \frac{1}{4}(X_{t-2} + X_{t-1} + X_t + X_{t+1})$$

$$M_3[X_t] = \frac{1}{4}(X_{t-2} + 2X_{t-1} + 2X_t + 2X_{t+1} + X_{t+2})$$

Qu'en est-il des moyenne $M_{2 \times 12}$ et $M_{3 \times 3}$?

P2 : Elimination de la saisonnalité

Élimination saisonnalités constantes = élimination fonctions périodiques.

À retenir : Il suffit de choisir une MM simple d'ordre de la périodicité.

Exercice : montrer qu'une MM simple d'ordre 12 élimine les saisonnalités stables

Qu'en est-il de la moyenne $M_{2 \times 12}$?

P3 : Réduction du bruit (i.e. l'irrégulier)

On ne peut éliminer la composante irrégulière mais on peut l'atténuer.

Une MM transforme un bruit blanc de variance σ^2 en un bruit de moyenne nulle, autocorrélé et de variance :

$$\sigma^{*2} = \sigma^2 \sum_{i=-p}^f \theta_i^2$$

La quantité $\sum_{i=-p}^f \theta_i^2$ est le *pouvoir réducteur* de la moyenne mobile

Réduire le bruit ➡ **minimiser la somme des carrés des coefficients.**

Quel pouvoir réducteur des moyennes $M_{2 \times 12}$ et $M_{3 \times 3}$?

Le problème des fins de série

Une MM centrée d'ordre $2p+1$ ne peut lisser les p premiers et derniers points

Solution 1 : utiliser des MM asymétriques

Les MM asymétriques de Musgrave permettent de minimiser les révisions (associées à celles d'Henderson) sous certaines contraintes polynomiales

Solution 2 : prolonger la série par prévision et appliquer une MM symétrique

Les prévisions sont une combinaison linéaire du passé, ça reste asymétrique.

Les moyennes mobiles dans X11

- MM centrée d'ordre la périodicité de la série (2X12, 13 termes, pour les séries mensuelles, 2X4 pour séries trim.) : conserve des droites, supprime une saisonnalité stable, faible pouvoir de lissage
- MM de *Macurves*, pour extraire la composante saisonnière : 3X3 (5 termes), 3X5 (7 termes), 3X9 et 3X15
 - ➔ `rjd3x11plus::macurves()`
- MM de *Henderson* (5, 7, 9, 13 ou 23 termes) : conserve des polynômes, bon pouvoir de lissage
 - ➔ MM asymétriques de *Musgrave* en fin de série
 - ➔ `rjd3filters::lp_filter()`

Choix de la longueur des filtres dépend des variations de l'irrégulier par rapport à celles de TC (Henderson) et de S (Macurves)

Idée de la moyenne mobile de Henderson

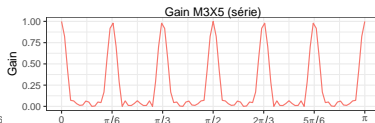
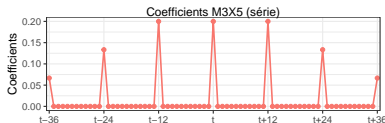
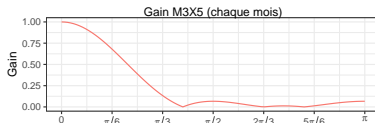
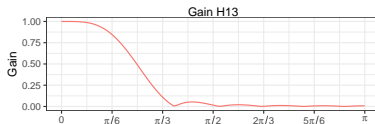
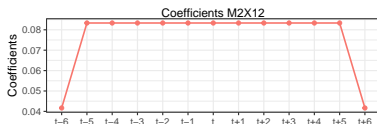
Objectif : obtenir une estimation lisse de la tendance-cycle. Pour M une MM centrée d'ordre $2p + 1$, $X_t = \begin{cases} 1 & \text{si } t = 0 \\ 0 & \text{si } t \neq 0 \end{cases}$ on a :

$$MX_t = \begin{cases} 0 & \text{si } t < -p \text{ ou } t > p \\ \theta_t & \text{si } -p \leq t \leq p \end{cases} \quad \text{lisse si courbe des coefficients lisse}$$

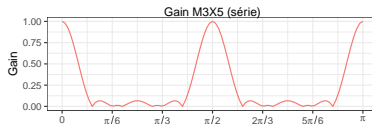
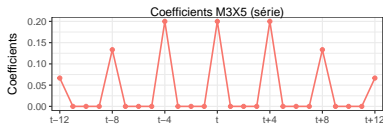
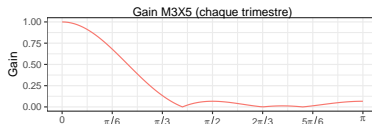
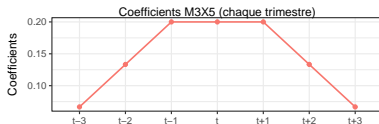
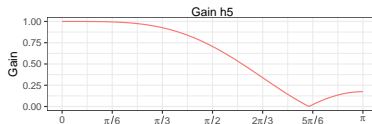
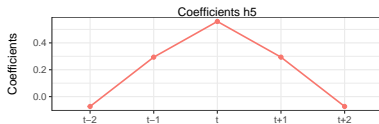
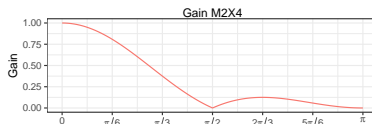
Les MM de Henderson sont solution du programme :

$$\begin{cases} \min_{\theta} \sum (\nabla^3 \theta_i)^2 & \text{critère de "souplesse" de la courbe des coefficients} \\ \sum_{i=-p}^p \theta_i = 1, \sum_{i=-p}^p i\theta_i = 0, \sum_{i=-p}^p i^2\theta_i = 0 \end{cases}$$

Exemples des moyennes mobiles (mens)

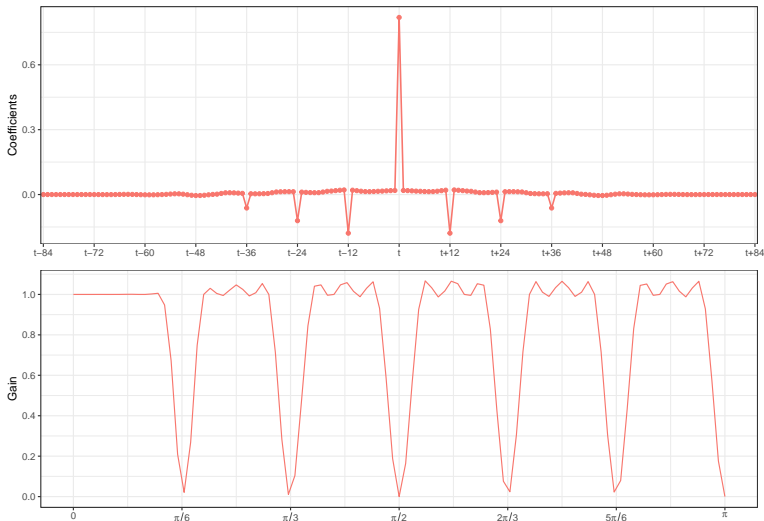


Exemples des moyennes mobiles (trim)



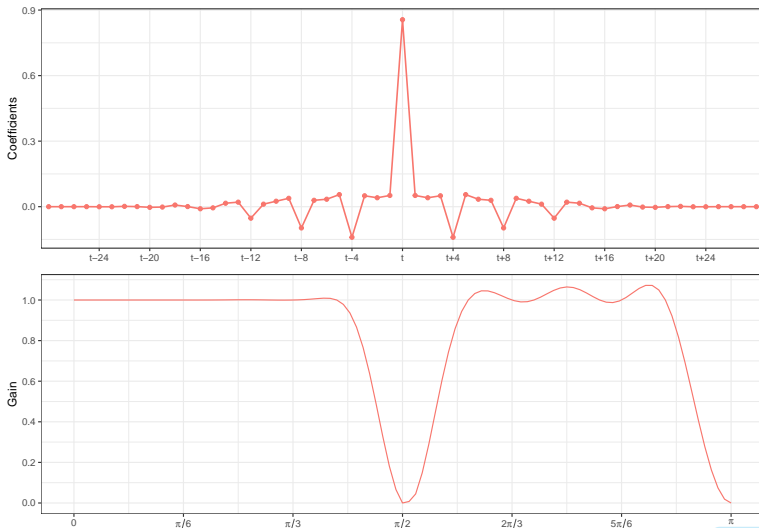
Filtre symétrique de X-11 (mens)

Pour séries mensuelles, Henderson 13 et Macurve 3×5 , 84 mois = 7 ans.



Filtre symétrique de X-11 (trim)

Pour séries trimestrielles, Henderson 5 et Macurve 3×5 , 84 mois = 7 ans.



Sommaire

1. Introduction aux moyennes mobiles

2. Choix de la longueur de la MM

2.1 Rappels sur le principe itératif de X11

2.2 Longueur du filtre de Henderson

2.3 Longueur du filtre de Macurves

3. Diagnostics sur la qualité de la décomposition

Principe itératif de X11 (1/2)

Une première estimation de la CVS :

1. Estimation de la **tendance-cycle** par moyenne mobile 2×12 (ou 2×4) :

$$TC_t^{(1)} = M_{2 \times 12}(X_t)$$

2. Estimation de la composante **saisonnier-irrégulier** :

$$(S_t + I_t)^{(1)} = X_t - TC_t^{(1)}$$

3. Estimation de la composante **saisonnrière** par moyenne mobile 3×3 sur **chaque mois** :

$$S_t^{(1)} = M_{3 \times 3} \left[(S_t + I_t)^{(1)} \right] \text{ et normalisation } Snorm_t^{(1)} = S_t^{(1)} - M_{2 \times 12} \left(S_t^{(1)} \right)$$

4. Estimation de la série corrigée des variations saisonnières :

$$Xsa_t^{(1)} = (TC_t + I_t)^{(1)} = X_t - Snorm_t^{(1)}$$

Principe itératif de X11 (2/2)

Une seconde estimation de la CVS :

1. Estimation de la **tendance-cycle** par moyenne de Henderson (généralement 13 termes ou 5 pour séries trim) :

$$TC_t^{(2)} = H_{13}(Xsa_t^{(1)})$$

2. Estimation de la composante **saisonnier-irrégulier** :

$$(S_t + I_t)^{(2)} = X_t - TC_t^{(2)}$$

3. Estimation de la composante **saisonnrière** par moyenne mobile 3×5 (généralement) sur **chaque mois** :

$$S_t^{(2)} = M_{3 \times 5} \left[(S_t + I_t)^{(2)} \right] \text{ et normalisation } Snorm_t^{(2)} = S_t^{(2)} - M_{2 \times 12} \left(S_t^{(2)} \right)$$

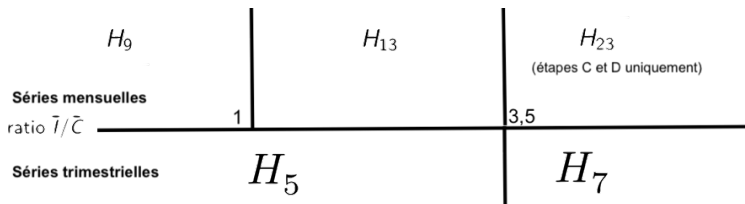
4. Estimation de la série corrigée des variations saisonnières :

$$Xsa_t^{(2)} = (TC_t + I_t)^{(2)} = X_t - Snorm_t^{(2)}$$

Ratio I/C

Choix du filtre de Henderson dépend du ratio \bar{I}/\bar{C} qui mesure l'importance de l'irrégulier dans la série. $C = H_{13}(SA)$ et $I = SA \text{ op } C$ avec $op = -$ ou $/$, et

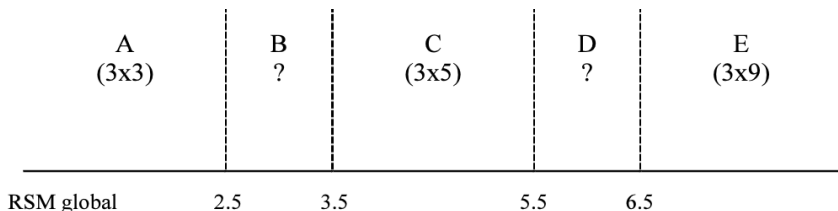
$$\bar{C} = \frac{1}{n-1} \sum_t |C_t \text{ op } C_{t-1} - xbar| \text{ avec } xbar = 0 \text{ ou } 1$$



Ratio de saisonnalité mouvante (RSM/MSR)

Choix du filtre saisonnier basé sur le calcul du ratio I/S, qui mesure la part du bruit dans la composante S-I (également calculable par mois j)

$$MSR = \frac{\sum_j n_j \bar{l}_j}{\sum_j n_j \bar{s}_j} \text{ avec } \bar{l}_j = \frac{1}{n_j - 1} \sum_t |l_{t,j} \text{ op } l_{t-1,j} - xbar|$$



Si zone B ou D, on enlève la dernière année et on recommence (au plus 5 fois, sinon M3X5)

i L'étude des MSR peut révéler des mois à saisonnalité \pm variables

Sommaire

1. Introduction aux moyennes mobiles

2. Choix de la longueur de la MM

3. Diagnostics sur la qualité de la décomposition

3.1 Les statistiques M

3.2 Sliding spans, révisions et stabilité du modèle

3.3 Tests de saisonnalité

Les statistiques M (1/3)

11 Statistiques, moyenne pondérée = statistique Q ou $Q - M2$

	Poids	Description	Si problème..
M1	10	Contribution de l'irrégulier à la variance totale (stationnarisation par différence d'ordre 3). Si trop élevé, difficile d'extraire la saisonnalité.	Points atypiques ou taille des filtres
M2	11	Contribution de l'irrégulier à la variance totale (stationnarisation par une droite).	Points atypiques ou taille des filtres
M3	10	Mesuré à partir du ratio I/C. Si trop grand on aura du mal à séparer les deux composantes.	Points atypiques ou taille des filtres
M4	8	Test autocorrélation sur l'irrégulier (réduire le filtre saisonnier).	Filtre saisonnier plus court
M5	11	Mesuré à partir du MCD (nombre de mois nécessaires pour que les variations absolues de la TC l'emporte sur I).	Points atypiques

Les statistiques M (2/3)

	Poids	Description	Si problème
M6	10	Vérifie si la moyenne mobile M3x5 est appropriée ($1.5 < I/S < 6.5$).	Prendre filtre plus long
M7	18	Permet de voir si la saisonnalité est identifiable (compare part relative de la saisonnalité stable et mobile).	Schéma multiplicatif ?
M8	7	Mesure l'évolution de la S de court terme.	Changer filtre saisonnier
M9	7	Mesure l'évolution de la S de long terme.	Changer filtre saisonnier
M10	4	M8 sur dernières années ($N - 2$ à $N - 5$).	Changer filtre saisonnier
M11	4	M9 sur dernières années ($N - 2$ à $N - 5$).	Changer filtre saisonnier

Les statistiques M (3/3)

M1 et M2 : calibrées pour des séries dont le cycle contribue de 5 à 10 % à la variance totale : si pas de cycle peut dépasser 1 et si contribution du cycle $> 10\%$ alors le seuil devrait être plus grand que 1

M3 et M5 : si tendance plate à ignorer, important si on analyse les cycles économiques

M4 échoue souvent (peut venir des CJO, de la collecte...) et indépendant des autres statistiques

M8 et M10 : ignorer si l'on s'intéresse qu'aux estimations courantes

M9 et M11 : ignorer si l'on s'intéresse qu'aux estimations "historiques"

MSR et Calendarsigma

Hypothèse générale : la saisonnalité évolue au même rythme dans tous les mois, l'irrégulier est stationnaire. . . Pas toujours le cas :

- MSR peuvent être différents entre les mois ➡ filtres saisonniers mal adapté aux mois plus volatils (paramètre `x11.seasonalma`)
- variance de l'irrégulier peut être différente en fonction des mois (test de cochrane) ➡ sur-correction des points atypiques sur certains mois (paramètre `x11.calendarsigma`)

💡 : l'étude des S-I ratio peuvent aider

Contribution à la variance

Sous Decomposition > Quality measures > Details tables F de X-11 avec des statistiques sur les contributions des différentes composantes :

- O = série initiale
- CI = série désaisonnalisée
- I = irrégulier, C = cycle (TC - tendance), S = saisonnalité
- P = Préajustement
- TD&H = effets de calendrier
- Mod.O, Mod.CI, Mod.I = séries corrigées des valeurs extrêmes

Comparaison à travers différentes spécifications

Fenêtre Diagnostics > Matrix permet de comparer les résultats sur différentes spécifications :

- Créer des nouvelles spécifications
- Pour une matrice personnalisée : Tools > Option > Custom matrix (diagnostics) > réduire la liste

Juger la fiabilité des estimations

Sliding spans

Objectif : juger de la fiabilité des estimations

Méthode : comparaison des estimations de la CVS-CJO sur différents intervalles (ex : [2007 - 2015], [2008 - 2016], [2009 - 2017], [2010 - 2018])

Révision

Objectif : juger de la fiabilité des dernières estimations

Méthode : comparaison de la première et de la dernière estimation de la CVS-CJO sur les 4 dernières années

Model stability

Objectif : juger de la fiabilité des estimations du modèle Reg-ARIMA

Méthode : comparaison des estimations des coefficients du modèle Reg-ARIMA

Tests de saisonnalité

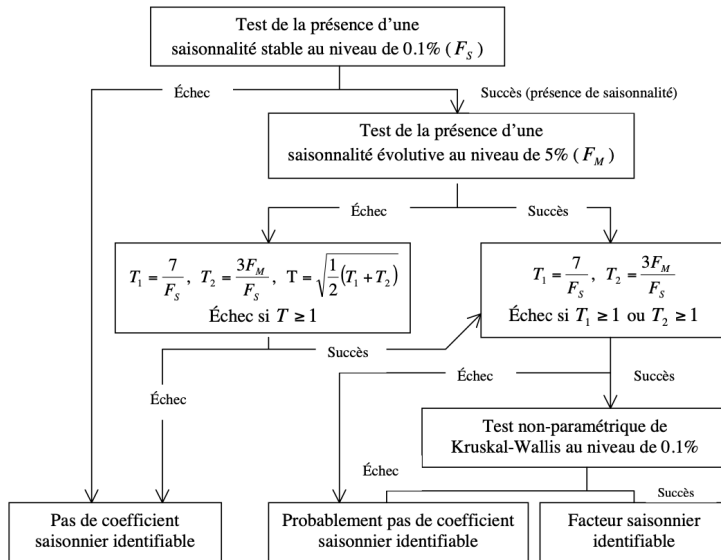
Test	Description
1. Auto-correlations at seasonal lags (QS)	Test "Ljung-Box" sur lags saisonniers
2. Friedman (non parametric)	Test de saisonnalité stable entre mois (test de rang, proche du W de Kendall)
3. Kruskal-Wallis (non parametric)	Test ANOVA de saisonnalité stable entre mois
4. Spectral peaks	Test spectral (Tukey et AR(30)) sur les pics saisonniers
5. Periodogram	Test spectral (periodogramme, transformée de Fourier)
6. Seasonal dummies (F-test)	F-test sur coefficients saisonniers avec modèle ARIMA (0,1,1) + mean
6bis. Seasonal dummies (AMI)	F-test sur coefficients saisonniers avec modèle ARIMA automatique + mean

i Dans "Main results" : qs test effectué sur l'ensemble de la série et f-test sur 8 dernières années.

"Residual seasonality" = test combiné

! Tests non effectués sur les séries finales !

Test combiné, source : Ladiray et Quenneville (1999)



Les essentiels

X11 est une application successive de différentes MM ayant des fonctions spécifiques







Les 3 principales propriétés des MM : conservation des droites ou polynômes, suppression des fonctions périodiques et pouvoir réducteur du bruit

La finalité de X11 étant d'estimer les coefficients saisonniers par une très longue MM

Les CVS sur la fin de série sont déphasées

Important de vérifier la qualité de la décomposition : statistiques M et saisonnalité résiduelle. . . mais dépend de l'objectif !

Bibliographie

-  Documentation de JDemetra+
<https://jdemetradocumentation.github.io/JDemetra-documentation/>
-  Kirchner R., Ladiray D. et Mazzi G.L, Quality Measures and Reporting for Seasonal Adjustment, in Handbook on Seasonal Adjustment, edited by G. L. Mazzi, co-edited by D. Ladiray, European Union, Luxembourg.
ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-18-001.
-  Ladiray D., et Quenneville B. (1999). Comprendre la méthode X11, *Institut National de la Statistique et des Études Économiques*,
<https://www.census.gov/pub/ts/papers/x11doc.pdf>
-  De Antonio A. et Palate J. (2019), 'JD+ Seasonality tests'.
<https://github.com/palatej/ESTP>.
-  Findley D. F., Lytras D. P. et McElroy, T. S. (2017). Detecting seasonality in seasonally adjusted monthly time series. *Statistics*, 3,
<https://www.census.gov/content/dam/Census/library/working-papers/2017/adrm/rrs2017-03.pdf>
-  Lytras D. P., Feldpausch R. M. et Bell W. R. (2007). Determining seasonality: a comparison of diagnostics from X-12-ARIMA. *US Census Bureau*,
<http://www.census.gov/ts/papers/ices2007dpl.pdf>.