

## Travaux pratiques de Séries Temporelles

F. LAVANCIER, A. PHILIPPE

### *Étude préliminaire d'une série temporelle*

*Quelques fonctions utiles :*

- `ts` [objet  $\mathbb{R}$  pour représenter une série temporelle],
- `acf` [auto corrélations/covariances],
- `diff` [calcul des différences  $x_t - x_{t-h}$  où  $h$  est le `lag`],
- `filter` [par exemple pour la simulation des modèles AR étudiés dans l'exercice 1 ]

**Ex 1.** *Quelques simulations de séries*

Soit  $(\epsilon_n)$  une suite de variables i.i.d. gaussiennes centrées réduites.

- 1) Pour  $|a| = 0, .5, .9$ , simuler des trajectoires du processus défini par

$$X_n = \epsilon_n + a\epsilon_{n-1}$$

ainsi que leur suite des auto-corrélations.

- 2) Pour  $|a| = 0, .5, .9$ , simuler des trajectoires du processus défini par

$$X_n + aX_{n-1} = \epsilon_n$$

ainsi que leur suite des auto-corrélations.

**Ex 2.**

On considère une série chronologique  $(x_1, \dots, x_n)$  de longueur  $n$  et on note  $\hat{\rho}_n^x(h)$  la suite des auto-corrélations empiriques

1) Soit  $(\epsilon_n)$  un bruit blanc gaussien centré et de variance 1. Calculer et représenter les covariances empiriques  $\hat{\sigma}_n(h)$  pour les séries  $z_j = j + \epsilon_j$  pour  $j \in \{1, \dots, n\}$  (*modèle additif*) et  $w_j = j\epsilon_j$  pour  $j \in \{1, \dots, n\}$  (*modèle multiplicatif*).

- 2) Commentez les résultats obtenus dans cet exercice.

**Ex 3.**

- 1) Écrire une fonction qui retourne une série de la forme

$$y_j = a \cos(\omega j) + bj + \epsilon_j$$

- 2) Pour  $a = 0$  et  $b \in \mathbb{R}$ , simuler une trajectoire, puis représenter

2-1) la série et sa suite d'auto-corrélations empiriques

2-2) la série  $X_n - X_{n-1}$

2-3) la suite des auto-corrélations empiriques associée à la série  $X_n - X_{n-1}$

- 3) Pour  $b = 0$  et  $\omega = \pi/6$  simuler une trajectoire, puis représenter
  - 3-1) la série et sa suite des auto-corrélations empiriques
  - 3-2) la série  $X_n - X_{n-12}$
  - 3-3) la suite des auto-corrélations empiriques associée à la série  $X_n - X_{n-12}$
- 4) Pour  $a \neq 0$  et  $b \neq 0$ , simuler une trajectoire puis représenter
  - 4-1) la série et sa suite des auto-corrélations empiriques
  - 4-2) les séries  $X_n - X_{n-1}$  et  $X_n - X_{n-12}$
  - 4-3) la suite des auto-corrélations empiriques associée à la série  $X_n - X_{n-12}$  et  $X_n - X_{n-1}$

**Ex 4.** *Réchauffement climatique*

Télécharger les séries :

- “globtemp.dat” : températures annuelles depuis 1856 (plus exactement : différences avec la moyenne constatée entre 1961 et 1990)
- “varve.dat” : épaisseurs de 634 varves de la période glaciaire. Les varves sont des dépôts sédimentaires issus de la fonte des glaciers. Leur épaisseur est directement liée à la température moyenne de l’année.

On se pose la question d’un réchauffement climatique significatif au 20ème siècle.

- 1) Mettre ces séries au format ts sous R. Les représenter.
  - 2) Zoomer *globtemp* sur la période 1900- . Effectuer une régression linéaire : une tendance positive significative apparait-elle ?
  - 3) Représenter les résidus de la regression précédente. Effectuer un test d’auto-corrélation des résidus (par exemple le test de Durbin Watson avec `durbinWatsonTest` de la librairie ‘car’ ou le test de Breusch Godfrey avec `bgtest` de la librairie `lmtest`).
  - 4) Représenter le nuage de points des résidus à l’instant  $n$  par rapport aux résidus à l’instant  $n - 1$ . Représenter les auto-corrélations empirique des résidus. Faire le lien avec le résultat du test précédent.
  - 5) En quoi le résultat du test remet en cause la significativité de la pente constatée à la question 2 ?
  - 6) On admet que les résidus du modèle suivent un modèle autorégressif d’ordre 1 (AR1), comme en 2) de l’exercice 1. Mettre en oeuvre une estimation de la régression linéaire précédente à l’aide de la méthode des moindres carrés généralisés. On pourra utiliser la fonction `gls` de la librairie `nlme` en précisant la structure AR1 des résidus grâce à l’option `correlation=corAR1()`.
  - 7) Une tendance positive apparait-elle suite à cette nouvelle estimation ?
  - 8) Récupérer les résidus normalisés de la régression précédente grâce à la fonction `residuals` en précisant l’option `type="normalized"`. Le test de significativité précédent vous parait-il valide ?
- Note : le vecteur des résidus normalisés est le vecteur des résidus de la régression multiplié par l’inverse de sa matrice de covariance estimée.*
- 9) Peut-on extraire de la série *varve* une période de 100 ans ayant une tendance similaire à celle observée sur *globtemp* ?

**Ex 5.**    *Stationnarité de la série varve*

1) Calculer les moyennes et variances glissantes par tranches de 50 ans de la série *varve*, notée  $X_n$  par la suite (utiliser la fonction `rollapply` de la librairie `zoo`). La série semble-t-elle stationnaire?

2) On désire transformer  $X_n$  pour stabiliser la variance. Montrer que la transformation  $\ln$  est raisonnable. La série  $Y_n = \ln(X_n)$  semble-t-elle stationnaire?

3) Représenter la série  $Z_n = Y_n - Y_{n-1}$ . Vous paraît-elle stationnaire? (Procéder comme à la première question)

4) Calculer la corrélation entre  $Z_n$  et  $Z_{n-1}$ . La série  $Z_n$  est-elle un bruit blanc? Calculer la corrélation entre  $Z_n$  et  $Z_{n-2}$ . Commenter.

5) En vous inspirant des exemples traités exercice 1, Comment pourrait-on modéliser la série  $Z_n$ ? Comment estimer les paramètres inconnus de ce modèle, en s'appuyant notamment sur l'autocorrélogramme  $(\hat{\rho}(h))_h$ .

6) Comment pourrait-on prédire la valeur de  $X_{635}$ ?

Indication : Si  $X_n$  est un MA(1) i.e.  $X_n = \epsilon_n + a\epsilon_{n-1}$  avec  $|a| < 1$  et  $\epsilon$  un bruit blanc faible de variance  $\sigma^2$ , alors la meilleure prévision linéaire de  $X_{n+1}$  sachant le passé est

$$\hat{X}_{n+1} = a \sum_{j \geq 0} (-a)^j X_{n-j}.$$

## Travaux pratiques de Séries Temporelles

F. LAVANCIER, A. PHILIPPE

**Ex 6.** *Données des ventes de champagne*

Télécharger le fichier `champagne.asc`. La série des ventes de champagne est notée  $v_t$

- 1) Tracer la série et sa suite des auto-corrélations empiriques.
- 2) Peut-on détecter la présence d'une fonction périodique ou d'une tendance dans cette série ?
- 3) Quels filtres peut on appliquer pour éliminer la tendance et la composante saisonnière.
- 4) Tracer la série obtenues après élimination de la tendance et de la saisonnalité et sa suite des auto-corrélations empiriques.
- 5) La série obtenue peut elle être modélisée par un bruit blanc ?
- 6) Donner une estimation de la tendance ( $\hat{m}_t$ ) et de la composante saisonnière ( $\hat{s}_t$ ).
- 7) Tracer la série des résidus c'est à dire  $v_t - \hat{s}_t - \hat{m}_t$  et ses auto-corrélations empiriques.
- 8) La série des résidus peut elle être modélisée par un bruit blanc ?
- 9) Proposer une prévision à l'horizon 12 des ventes de champagne.

**Ex 7.** *Tendance et saisonnalité :étude de la série de vente de voitures*

Télécharger la série "voiture". On conserve uniquement le début de la série avant la rupture.

- 1) A l'aide de la suite des covariances empiriques, mettre en évidence la présence d'une tendance et d'une composante saisonnière dans cette série.

*Approche 1*

- 2) Estimer la tendance et la composante saisonnière de cette série.
- 3) Éliminer la tendance et la composante saisonnière à partir de vos estimations. On note  $(R_j)_j$  la série obtenue.
- 4) La série  $(R_j)_j$  peut elle être modélisée par un bruit blanc ? Discuter le résultat.

*Approche 2*

- 5) Éliminer la tendance et la composante saisonnière en appliquant un ou plusieurs filtres linéaires de la forme  $(I - L^s)^d$ . On note  $(R'_j)_j$  la série obtenue.
- 6) La série  $(R'_j)_j$  peut elle être modélisée par un bruit blanc ? Discuter le résultat.

*Conclusion*

## Travaux pratiques de Séries Temporelles

F. LAVANCIER, A. PHILIPPE

*Processus ARMA. Processus  $L^2$*

### Ex 8. *Simulations sous SAS*

1) Créer sous SAS les séries suivantes, pour  $n = 1 \dots 200$  :

- $T_n$  une tendance linéaire :  $T_n = 100 + 3n$
- $S_n^1$  une série périodique de période 3 :  $S_n^1 = 50 \cos\left(\frac{2\pi}{3}n\right)$
- $S_n^2$  une série périodique de période 4 :  $S_n^2 = 50 \sin\left(\frac{\pi}{4} + \frac{\pi}{2}n\right)$
- $S_n$  une série périodique de période 12 :  $S_n = S_n^1 + S_n^2$
- $\epsilon_n$ , un bruit blanc i.i.d. suivant une  $\mathcal{N}(0, 15^2)$
- Le processus  $X_n = T_n + S_n + \epsilon_n$

2) A l'aide de la procédure EXPAND, appliquer les moyennes mobiles suivantes centrées d'ordre 3, 4 et 12 sur les six séries précédentes.

- $M_3 = \frac{1}{3}F(I + B + B^2)$
- $M_4 = \frac{1}{4}F^2\left(\frac{1}{2}I + B + B^2 + B^3 + \frac{1}{2}B^4\right)$
- $M_{12} = \frac{1}{12}F^6\left(\frac{1}{2}I + B + B^2 + \dots + B^{10} + B^{11} + \frac{1}{2}B^{12}\right)$ .

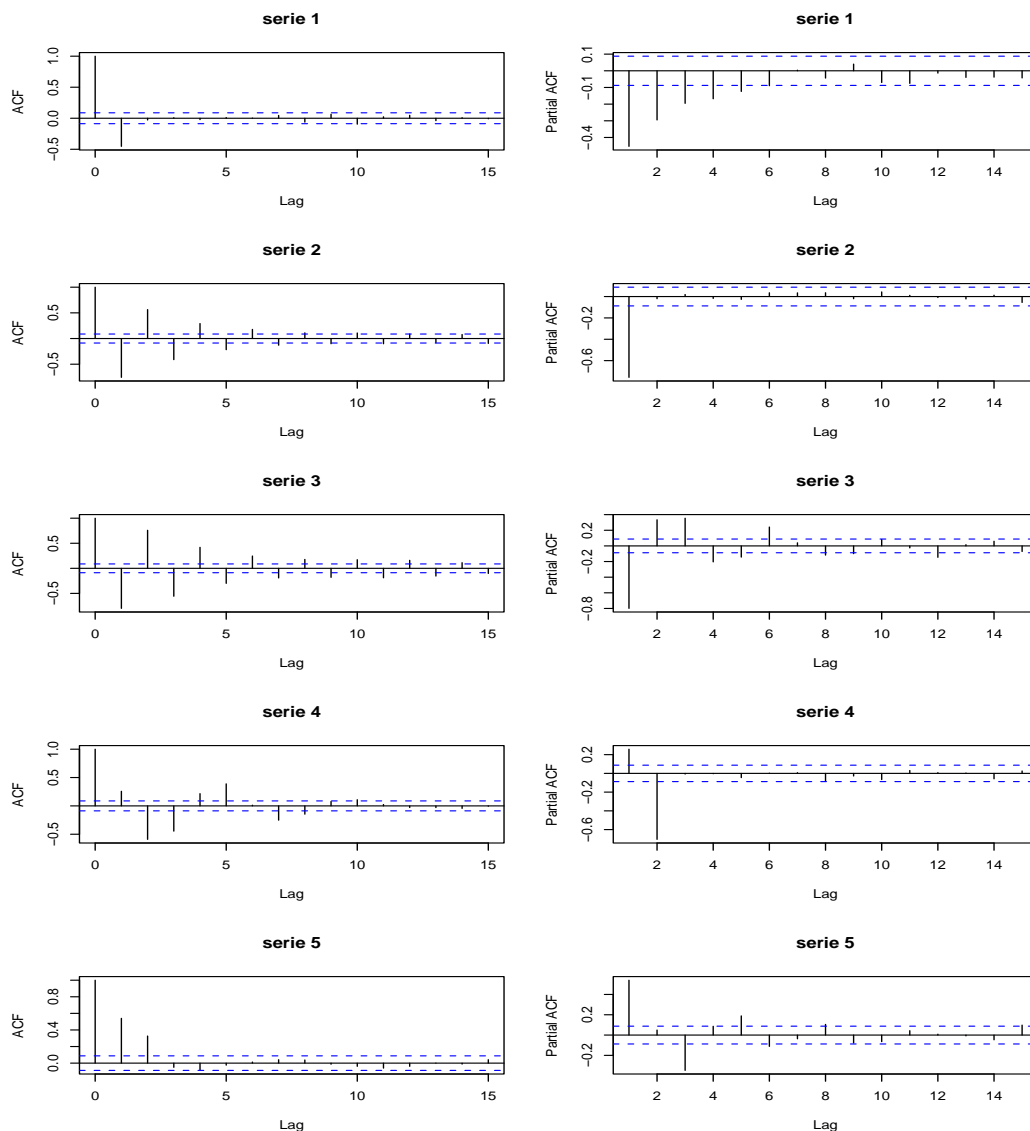
Quelles conclusions pouvez vous en tirer ?

### Ex 9. *Identification de modèles*

Le graphique suivant, représente les autocorrélations et autocorrélations partielles empiriques de processus ARMA simulés (les trajectoires sont de longueur 500). A partir de ces graphes et de simulations que vous effectuerez sous R, identifier les processus AR, MA et ARMA (préciser les ordres).

*Indication pour les simus sous R :*

- `ARMAacf` pour le calcul des ACF et des PACF théoriques d'un ARMA
- `arima.sim` pour simuler des trajectoires de processus ARMA
- `arima` ou `ar` pour estimer les paramètres d'un modèle ARMA ou AR



**Ex 10.** Modéliser à l'aide de modèles ARMA les séries *simul1*, *simul2*, *simul3* et *simul4* contenues dans les fichiers du même nom. On pourra éventuellement les stationariser au préalable en les différenciant.

**Ex 11.** Modéliser à l'aide d'un modèle SARIMA la série *elec\_austr* (contenue dans le fichier '*elec\_austr.txt*') qui correspond à la production mensuelle australienne d'électricité (en millions de kWh) de 1956 à avril 1990.

**Ex 12.** Importer le fichier de données "SNCF" dans SAS. Il s'agit du trafic voyageur mensuel.

- 1) Tracer la série. L'adéquation à un modèle pour la totalité de la série vous semble-t-elle possible?
- 2) On choisit de tronquer la série et de ne modéliser que sa partie la plus récente. Déterminer le point de rupture de la série (à la main).
- 3) Effectuer une modélisation SARIMA de cette sous-série sans la dernière année (utilisée ultérieurement comme année de validation) à l'aide d'une PROC ARIMA ou de l'outil "*Application de Prédiction de Séries Temporelles*".

4) Proposer une prévision avec un intervalle de confiance à 95% pour l'année future et comparer avec les valeurs réelles (i.e. l'année mise de côté à la question précédente).

5) Proposer une prévision de type Holt-Winters pour la même semaine. Comparer les erreurs quadratiques moyennes.

6) Refaire la même étude sous R. Effectuer une prévision non-paramétrique et comparer l'erreurs quadratique moyenne avec celles des méthodes précédentes (sous R, on pourra utiliser la fonction `HoltWinters()`, ainsi que les fonctions (fournies) `demo.np` et `np` pour le non-paramétrique.)

#### Utilisation de l'outil "Application de Prédiction de Séries Temporelles"

On le lance depuis `Solutions\Analyse` puis on choisit "Develop Models". On sélectionne les caractéristiques demandées (bibliothèque, table, variable à modéliser). Il convient notamment de choisir la variable `Time ID`. Cette variable correspond à l'indice temporel de la série étudiée. La nature de cette variable (mensuelle, quotidienne ou autres) est spécifiée dans `INTERVAL`, sous le choix `TIME ID`, et peut-être modifiée en cliquant sur `SPECIFY`. Elle peut même être créée en cliquant sur `CREATE`. Si `Time ID` est composée des mois de l'année, SAS prendra naturellement  $s = 12$  comme période, si `Time ID` est les jours ouvrés, SAS prendra  $s = 6$ , etc. Par contre, si `Time ID = _N_` (le numéro des observations), SAS ne pourra pas choisir de saisonnalité et leur modélisation sera impossible.

Une fois la série choisie, l'environnement est composé de boutons poussoirs. Au milieu de la fenêtre à droite, se trouve celui permettant de tracer la série (un dessin de série temporelle  $y$  est dessiné en bleu). La visualisation de la série nous permet d'analyser les autocorrélogrammes simple et partiel, de la différencier éventuellement une fois (par le bouton presseur "Delta" en haut) ou de la désaisonnaliser une fois (par le bouton presseur "Delta s"). Attention, la valeur de la période ne peut pas être choisie : elle est automatiquement proposée par le logiciel d'après la variable `Time ID` choisie au moment de la sélection de la table et des variables d'étude. Une fois la visualisation de la série faite et les paramètres a priori choisis grâce aux corrélogrammes, on peut lancer un modèle SARIMA à partir de la fenêtre "Develop Models" en cliquant sur le bouton poussoir en haut en forme de triangle bleu ("Fit Arima Models" en glissant dessus avec la souris). On remplit les champs :  $p$ ,  $d$ ,  $q$ ,  $P$ ,  $D$ ,  $Q$  et on a accès aux estimations et tests pour le modèle soit en cliquant droit sur son intitulé dans la fenêtre, soit en pressant les boutons poussoirs en haut à droite (corrélogrammes, estimations, résidus, prévisions, test de bruit blanc, etc).

Pour n'effectuer la modélisation que sur une partie de la série (c'est le cas ici car le début de la série SNCF a visiblement un comportement différent du reste), il faut sélectionner dans la fenêtre "Develop Models" l'intervalle correspondant en cliquant sur `Set Ranges` et en modifiant la `Period of Fit`.

La démarche d'identification des ordres de la partie ARIMA est la même que d'habitude (sauf qu'il n'y a pas de "méthode du coin" proposé par cet outil). Pour la partie saisonnière, la saison est automatiquement choisie par la variable `Time ID`. Les ordres de la partie saisonnière (ordres de la partie SMA et de la partie SAR) sont à choisir à partir du comportement des auto-corrélogrammes simple et partiel aux lags multiples de  $s$  en suivant les mêmes règles que dans le cas des ARIMA. Si par exemple l'autocorrélogramme simple présente une valeur non nulle en  $h = 12$ , on pourra proposer un ordre SMA égal à 1, etc.

Des ajustements dans les sorties du modèle sont possibles via les boutons de la barre du haut : nombre de lags pour les autocorrélogrammes, statistiques affichées (AIC, BIC,

etc.), niveau des ICs, etc.

**Ex 13.** Reprendre la série “champagne” et comparer sous R les prévisions obtenues sur la dernière année à l’aide 1) d’un filtrage simple ( $1 - B^{12}$ ), 2) d’un lissage de type Holt-Winters, 3) d’un modèle SARIMA(0,1,13)(0,1,0) censé inclure Holt-Winters, 4) d’une prévision non-paramétrique. Dans le cas de la prévision non-paramétrique, faire varier la taille de la fenêtre pour en analyser l’impact.

**Ex 14.** Faire le même type de comparaisons que ci-dessus sur la série *elec\_austr* en ajoutant la prévision issue du SARIMA ajusté lors d’un exercice précédent.

**Ex 15.** *Regréssion linéaire avec bruit AR*

Reprendre les données “globtemp” de températures au 20ème siècle. Le but est d’estimer la tendance linéaire en tenant compte de la dépendance des résidus (sous R) :

1) Analyser dans un premier temps la dépendance des résidus issus d’une régression simple et en proposer un modèle.

2) Effectuer une régression linéaire généralisée à l’aide de la fonction `gls()` de la librairie `nlme` (on peut spécifier un modèle de dépendance des résidus à l’aide de l’option `correlation`).

3) Effectuer la même estimation à l’aide de la fonction `arima()` (on peut préciser les variables sur lesquelles régresser à l’aide de l’option `xreg`).

**Ex 16.** *Une modélisation ARMAX*

Télécharger les données “LApollution.dat” qui regroupent les valeurs hebdomadaires, relevées à Los Angeles, de `mort` : mortalité due à un accident cardiovasculaire ; `temp` : températures ; `part` : pollution (poussières). L’objectif est de tenter d’expliquer au mieux la mortalité. On se place dans une optique de prévision : on choisit de n’expliquer la mortalité à un instant  $t$  qu’à l’aide des informations disponibles jusqu’à l’instant  $t - 1$ .

1) Représenter les trois séries et représenter leurs nuages de points croisés. Une dépendance apparaît-elle ?

2) Modéliser à l’aide d’une régression la mortalité à l’instant  $t$  en fonction de la température et de la pollution à l’instant  $t - 1$ . Que penser des résidus ?

3) Modéliser la dépendance de la série `mort` à l’aide d’un modèle ARIMA.

4) Analyser le lien temporel des résidus du modèle précédent avec la température et la pollution (fonction `ccf()` sous R ; option `crosscor` de la commande `identify` dans la `proc arima` sous SAS).

5) Introduire des régresseurs (éventuellement avec une dynamique propre de type ARMA) dans le modèle ARIMA initial, afin d’obtenir le meilleur modèle possible (=bruit blanc non corrélé avec les variables exogènes, variables significatives, etc.).

Pour cela sous R : option `xreg` de la fonction `arima()` (la dynamique des régresseurs ne peut être que de type AR en ajoutant les variables décalées dans le temps) ; sous SAS : option `crosscor` de la commande `identify` et option `input` de la commande `estimate` dans la `proc arima` (cf l’aide pour spécifier la dynamique des régresseurs).