

Untitled

2025-02-05

1 Introduction

La consommation énergétique et les variations de température sont deux phénomènes interdépendants qui jouent un rôle crucial dans la gestion des ressources et la planification des infrastructures. La température influence directement la demande énergétique, notamment à travers les besoins en climatisation lors des épisodes de chaleur. Comprendre cette relation et modéliser ces dynamiques est essentiel pour anticiper les pics de consommation, optimiser la production d'énergie et réduire les coûts associés.

L'objectif de cette étude est d'analyser la consommation énergétique en fonction des variations de température dans la période estivale (Juillet-Aout) et d'explorer notamment comment les modèles de lissage exponentiel (comme Holt-Winters) et les modèles SARIMA peuvent être utilisés pour prédire la demande énergétique et les fluctuations climatiques. Enfin, cette dernière sera l'occasion de s'intéresser à plusieurs moyens de calcul de corrélation entre la demande en énergie et les variations de température.

2 Données

Les données analysées dans le cadre de ce projet ont été relevées en 2017 dans la ville de Tetouan, au nord du Maroc (10375 km², estimation à 583374 habitants en 2017). Localisée le long de la mer méditerranée, la température est élevée et l'atmosphère est sèche durant la période d'été.

Ces données contiennent $n = 52416$ relevés qui, toutes les 10 minutes, fournissent les informations (variables) suivantes :

```
## [1] "Variables : 'Datetime' (date), 'Temperature' (°C), 'conso' (KW/h)"
```

Afin de faciliter l'analyse, les données de consommation d'énergie provenant des trois fournisseurs existants (Zone1, Zone2, Zone3) ont été rassemblées en une seule mesure (*conso*) décrivant la consommation totale, en KWh, pour la ville de Tetouan. Aussi, nous allons rassembler les données par heures, afin d'avoir 24 mesures pour une journée, puis trier les données pour ne garder que la période estivale (Debut : 01/07, fin : 31/08) donnant $T = 63$ [1 : 63] jours et $t = 1488$ mesures au total.



Sur les données de *data_ete* (Juillet / Aout) :

Pour la *Temperature*, celle-ci varie de 19°C à 40°C, avec une moyenne journalière de 26,5°C.

Pour la *conso*, celle-ci varie de 42642.2 KW/h à 133194.1 KW/h, avec une moyenne journalière de 87964.8 KW/h.

Aucune valeur manquante ni valeur abérante n'est à déplorer dans les données.

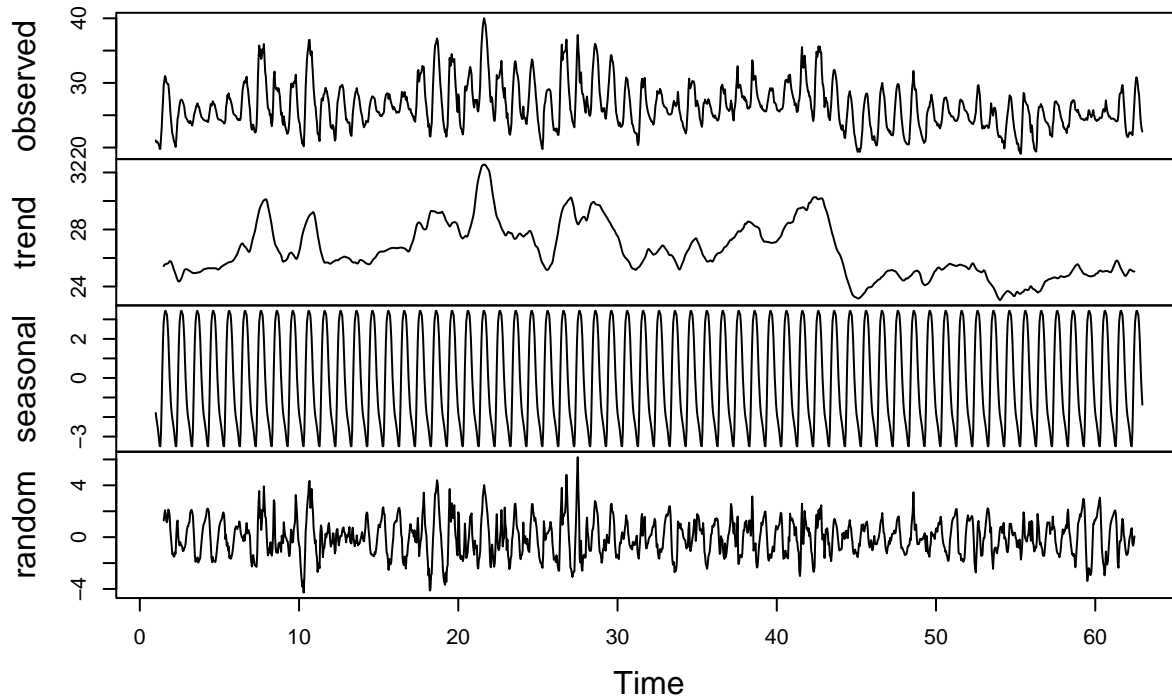
L'analyse sera séparée en 3 temps : l'analyse de la variable *Temperature*, l'analyse de la variable *conso* et l'analyse de la corrélation entre les deux mesures.

	Unique	Missing Pct.	Mean	SD	Min	Median	Max	Histogram
Temperature	876	0	26.5	3.5	19.0	26.0	40.0	
conso	1488	0	86884.2	18194.2	42642.2	87964.8	133194.1	

3 Analyse des séries temporelles

3.1 Temperature

Decomposition of additive time series



A partir de la décomposition de la série temporelle *st_temperature* :

- 1) La tendance semble linéaire : malgré les fluctuations visibles, la période estivale analysée donne l'intuition de températures relativement constantes dans les deux mois sélectionnés
- 2) Une saisonnalité journalière semble se dessiner : l'intuition derrière cette analyse est que malgré une tendance constante, les températures baissent la nuit avant de remonter en journée (pic haut à 14, pic bas à 6h **A VERIFIER**)
- 3) La partie résiduelle **A COMPLETER**

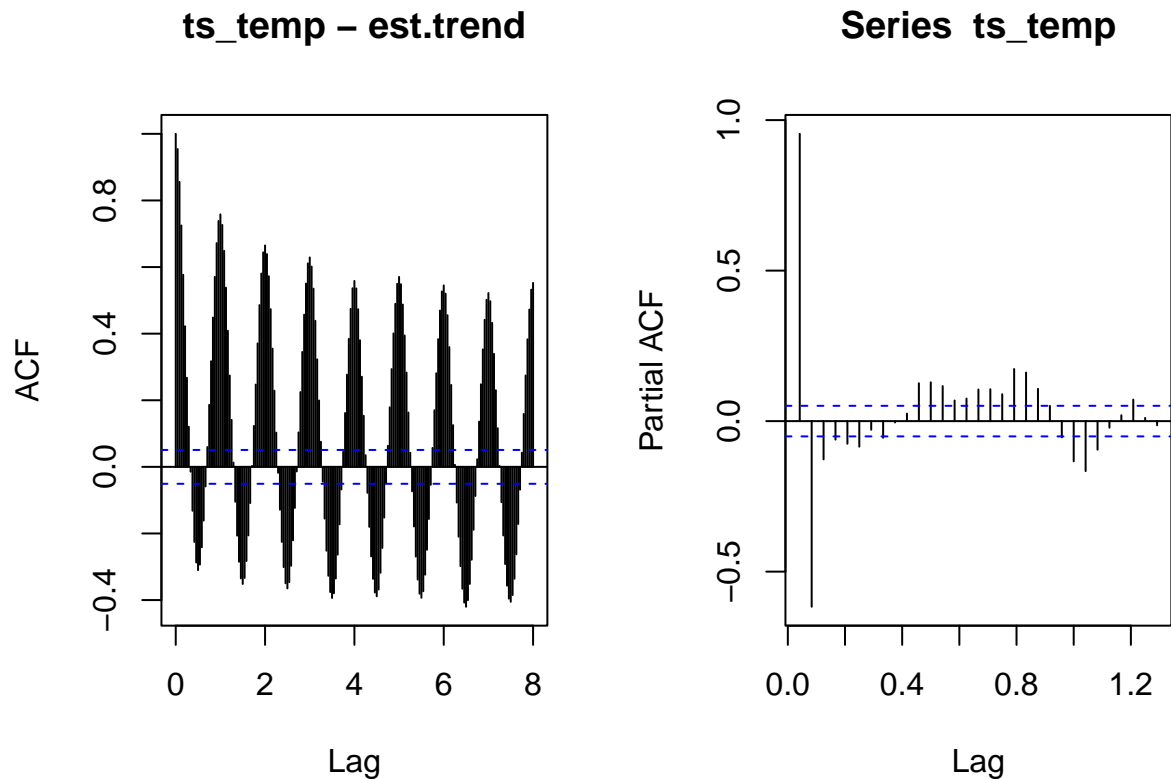
Comme la trend a une amplitude de moins de 10°C sur une période de plus de 60 jours, il ne semble pas nécessaire de modéliser la tendance autrement que par la température moyenne de cette dernière sous peine de complexifier le modèle. Néanmoins, on peut essayer d'ajuster un modèle linéaire. Sans grande surprise, le coefficient directeur de la droite de regression est de l'ordre de -10^{-3} . Toutefois, ce choix va influencer le **test kpss de non-stationnarité** car ce dernier, à la différence du **test de racine unitaire augmenté de Dickey-Fuller**, va détecter la trend non modélisée avec comme conséquence une non-stationnarité de la série. **ATTENTION ON DOIT CHOISIR SI OUI OU NON ON MODELISE LA TREND ET EXPLIQUER EXPLICITEMENT LE CHOIX**

```
## Warning in `-.default`(ts_temp, mod_trend_lin$fitted.values): la taille d'un
## objet plus long n'est pas multiple de la taille d'un objet plus court

##
## KPSS Test for Level Stationarity
##
## data: ts_temp - mod_trend_lin$fitted.values
## KPSS Level = 0.61084, Truncation lag parameter = 7, p-value = 0.02165
```

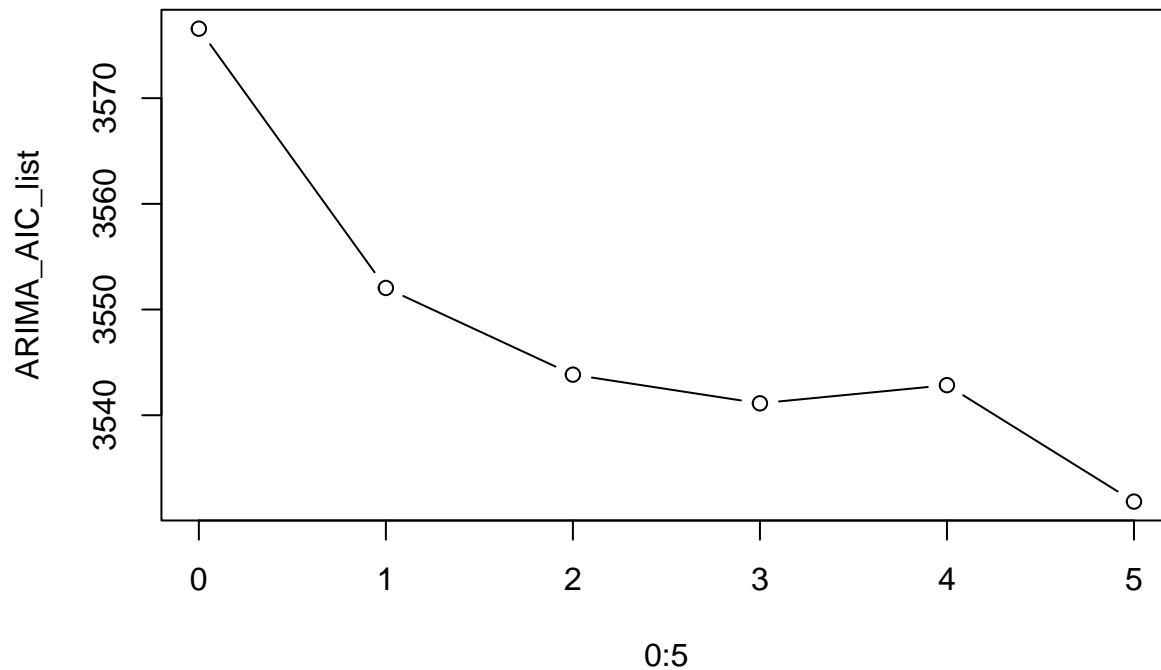
Maintenant, analysons l'ACF et le Partial ACF de la série temporelle afin de pouvoir sélectionner et paramétrer

au mieux le modèle final :



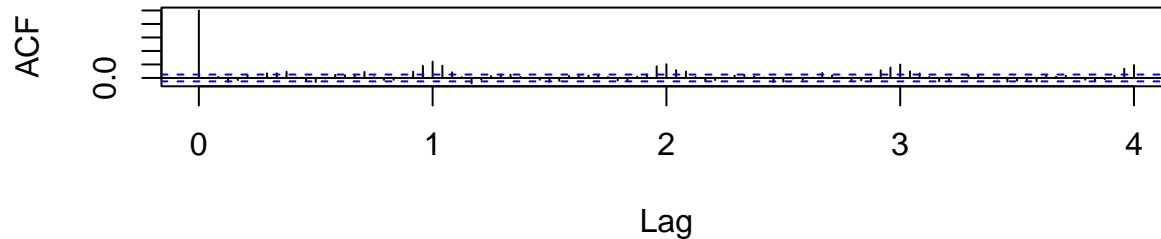
- Pour l'ACF, la forme de sinusoïde montre une saisonnalité claire, toutes les 24h (ce qui vient confirmer l'intuition de départ). Les “pics” aux lags 2 à 8 suggère une part de type moyenne-mobile dans le modèle.
- Par ailleurs, pour le partial ACF, les “pics” aux lags 1 et 2 suggère une part de type auto-regressive de paramètre 2.

On part donc sur un modèle ARIMA de paramètre $(p=2, i=0, q)$. Pour choisir le paramètre q , on choisit le modèle minimisant l'AIC (ce choix sera justifié plus tard)

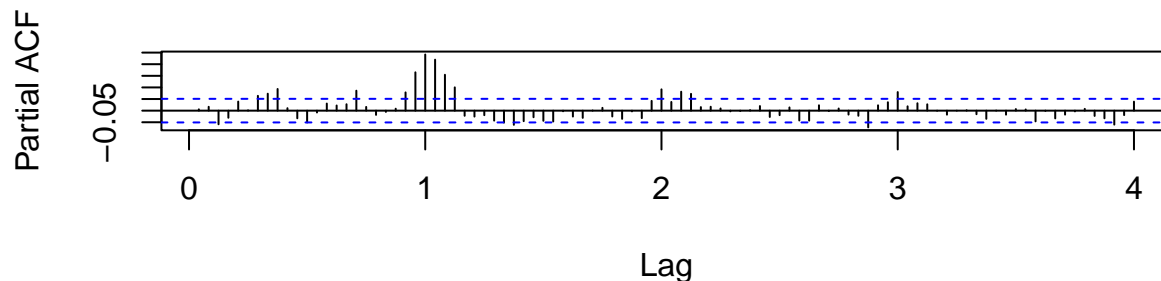


On voit que pour $q=2$, l'AIC est de 3544 alors qu'il est de 3541 pour $q=3$. Nous avons fait le choix de ne pas trop complexifier le modèle quitte à perdre en précisions : ainsi, nous choisissons $q=2$.

Series mod1\$residuals



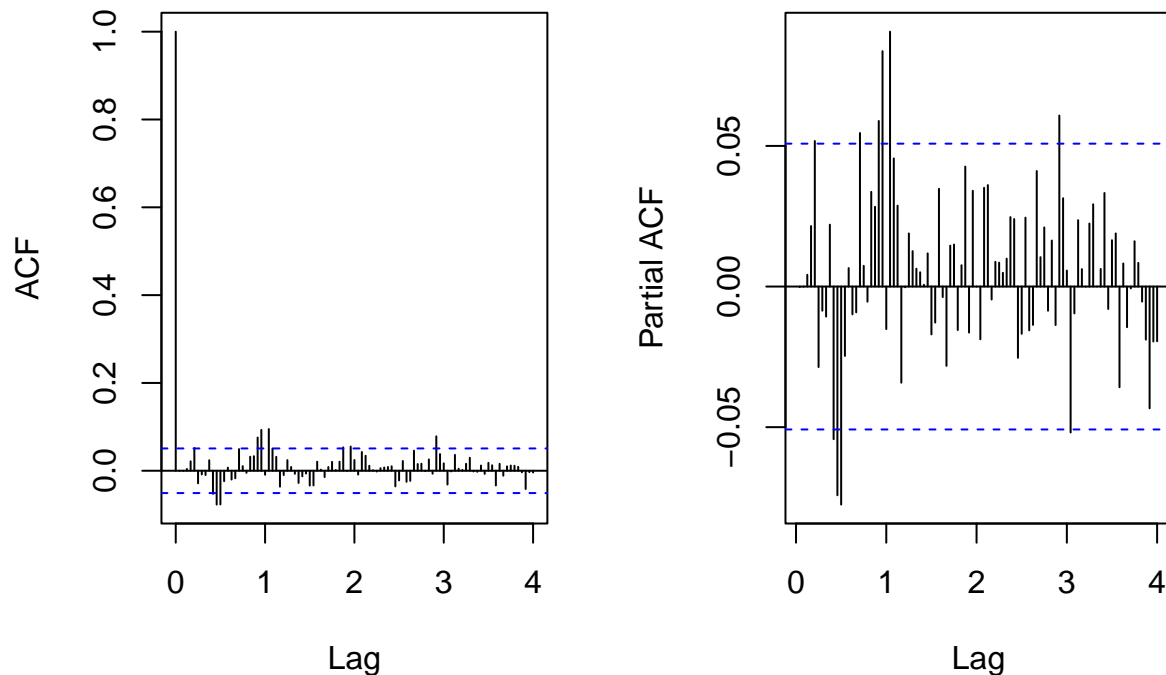
Series mod1\$residuals



[1] 0.8244877

Bien que le test de Box-Ljung admet une p-valeur de 0.82, l'acf et le pacf montre des “pics” à chaque période. Ceci est un symptôme de saisonnalité ; le modèle permettant de mieux modéliser la partie résiduelle est donc un modèle SARIMA. Les pics aux lags 24 sur l'acf et le pacf suggère un modèle SARIMA($p=2, i=0, q=2$)($P=1, D=1, Q=1$)[24]

Series manual_arima_temp\$residuals Series manual_arima_temp\$residuals



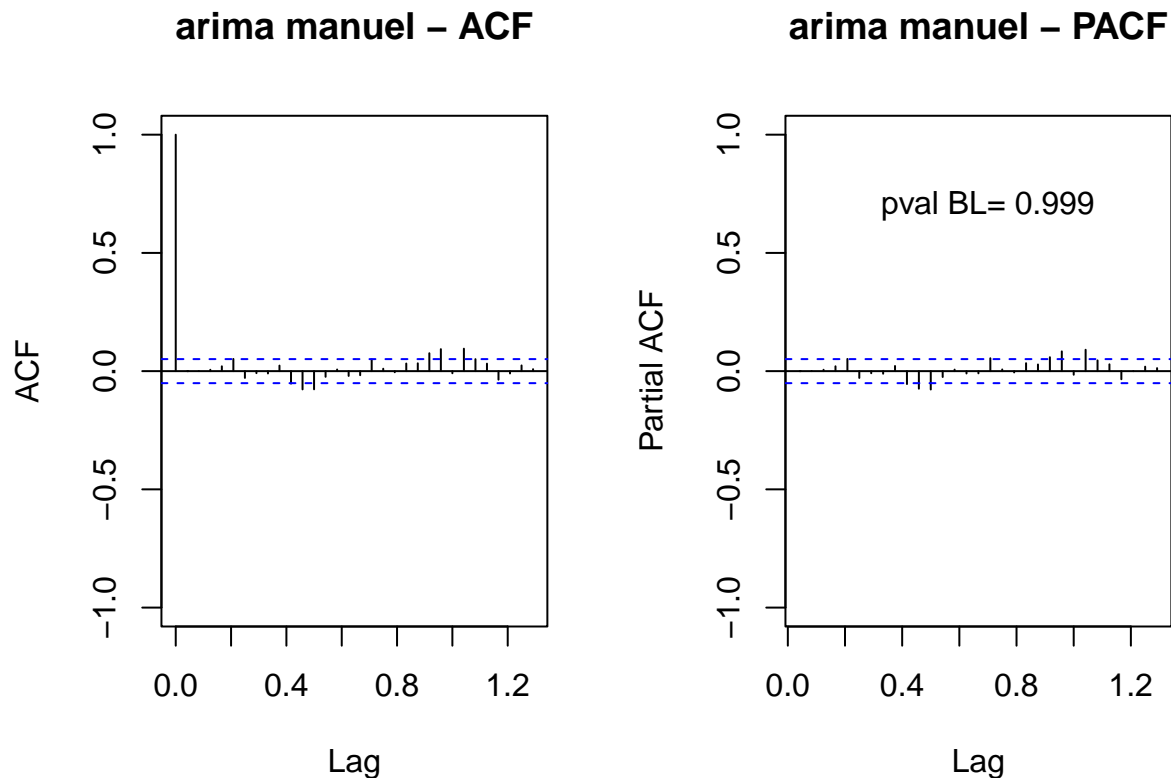
Il nous resterait à comparer les aic de tous les modèles SARIMA tels que $P + D + Q = 3$ mais on trouverait que le modèle précédent a l'aic le plus bas.

```
## [1] 0.9935056
```

On constate que le test de Box-Ljung admet une p-valeur encore plus proche de 1 et que les pics symptomatiques du modèle ARIMA sont moins prononcés et plus rares. De plus, l'AIC affiché est meilleur : 3161 pour le modèle SARIMA contre 3544 pour le modèle ARIMA. On retiendra donc le nouveau modèle.

Le **test de racine unitaire augmenté de Dickey Fuller** sur la partie résiduelle de nos données étant **significative**, nous avons fait le choix de constituer ****ATTENTION ON IGNORE LA TREND OU NON ?***

Nous avons donc fait le choix de constituer un modèle $ARIMA(p = 1, d = 0, q = 2)(P = 1, D = 1, Q = 1)$ [24] : ce modèle contient un terme auto-régressif d'ordre $p = 1$, aucun terme de différenciation car la série est supposée stationnaire **ATTENTION MODELISATION OU NON TREND**, un terme de moyenne mobile d'ordre $q = 1$. Pour la partie saisonnière, celle-ci a été modélisée



PEUT ETRE RAJOUTER DES COMPARAISONS AVEC D'AUTRES MODELES ?? AIC ??

Le test de **Ljung-Box** indique que les résidus de notre modèle ARIMA est un bruit blanc (au seuil $\alpha = 99\%$). De plus; l'interprétation graphique de l'ACF et du PACF nous indique que notre modèle capture la quasi-totalité des variations non-aléatoires de notre série temporelle.

Après analyse, les perturbations du modèle semblent homoscédastiques : l'intuition sur la non-variabilité des température moyennes durant ces deux mois viennent confirmer l'analyse, même si les résidus ne sont pas parfaitement heteroscédastiques (**test de Breuch-Pagan d'heteroscédasticité significatif**) **EXPLIQUER L'INTUITION AVEC LA TEMPERATURE** Les résidus du modèle ne suivent pas une loi normale : malgré l'absence de skew (**distribution centrée**), la distribution de nos rédidus reste **fortement leptokurtique**.

En réalité, nous n'effectuons pas de test paramétriques sur nos modèles. De plus, l'objectif principal étant la modélisation à très court-terme, nous vons choisi de négliger les hypothèses d'homoscédasticité (parfaite) et de normalité des résidus. Il ne faudra donc pas tenir compte des intervalles de confiance dont le calcul sur R n'est valable que pour des résidus suivant une loi normale.

Afin de s'assurer de la pertinence de notre modèle, nous allons le comparer aux modèles conçus par les algorithmes *auto.arima*, configurés en fonction du BIC, afin d'avoir une meilleure prévision au long terme (en effet, un BIC moins élevé suggère un modèle plus robuste) et au modèle de lissage exponentiel automatique de Holt-Winters *ets* **EXPLIQUER UN PEU POURQUOI C'EST BIEN POUR NOTRE MODELE ET QU'EST CE QUE ÇA PEUT AIDER A CAPTURER EN +**. Ces modèles seront utilisés pour prédire, à 3 jours, la température à chaque heure. Le graphique en rouge, utilisé comme référence, représente les valeurs réelles de température issue de nos données pour le septembre-01.

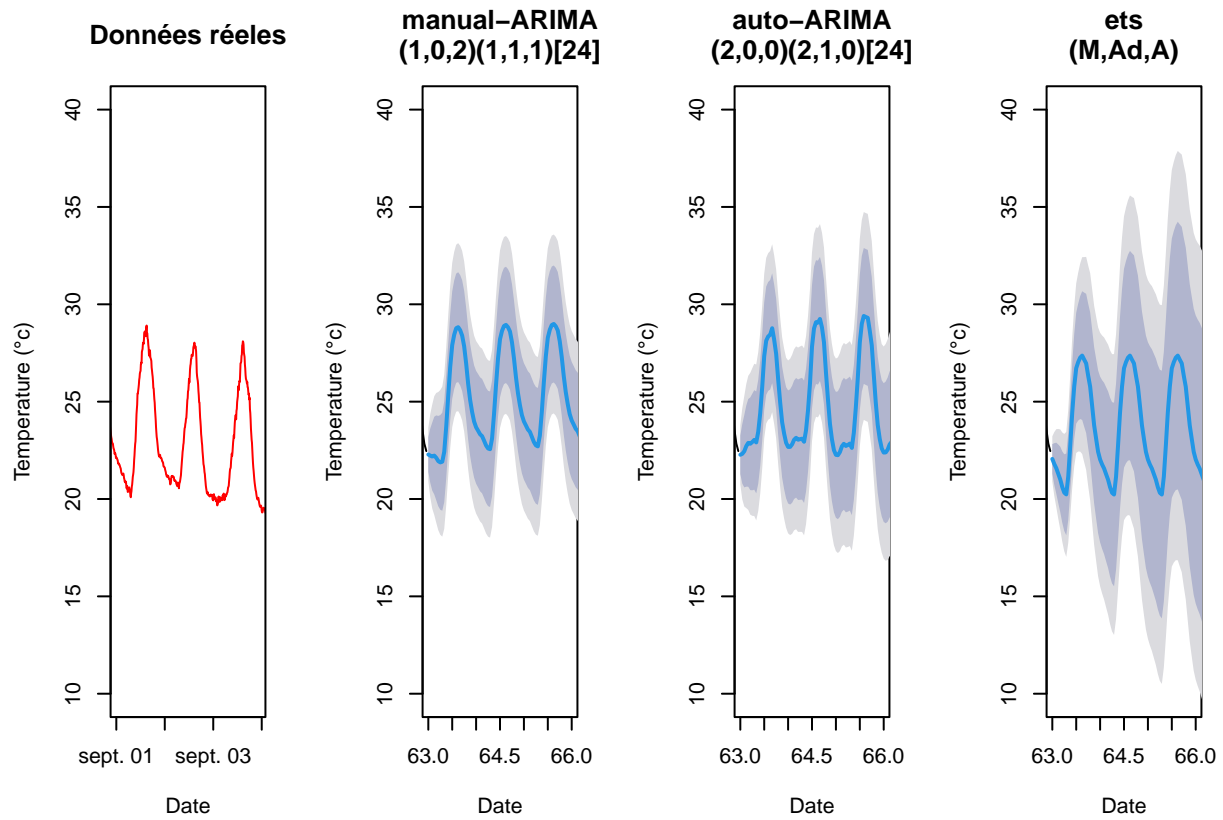
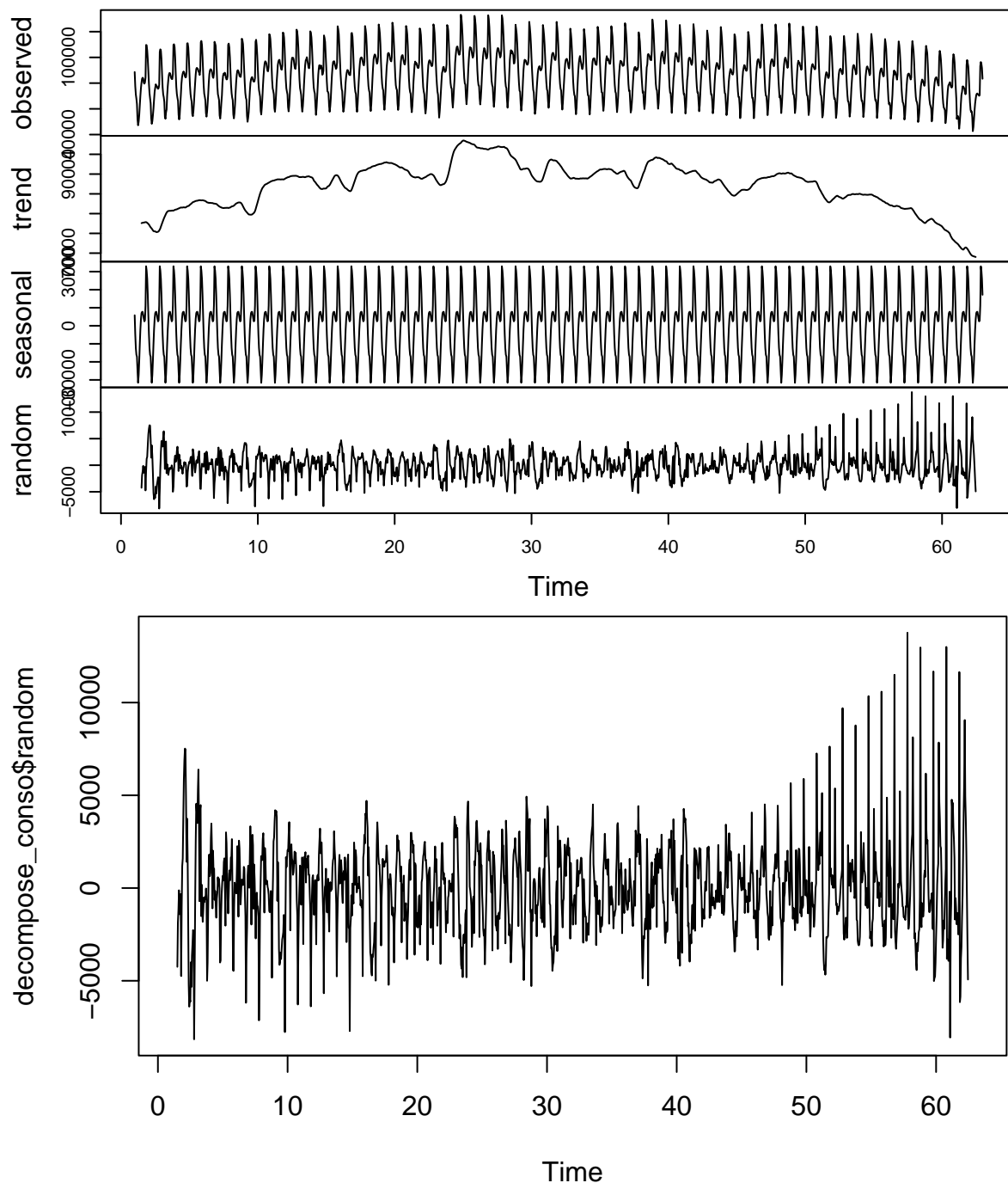


Table 1: Performances de prédiction des modèles

Modele	AIC	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Manual ARIMA	3162.542	-0.004	0.691	0.498	-0.072	1.848	0.656	0
Auto ARIMA	3425.987	0	0.768	0.557	-0.047	2.067	0.323	0.015
ETS	9857.847	-0.001	0.733	0.537	-0.038	1.992	0.311	0.26

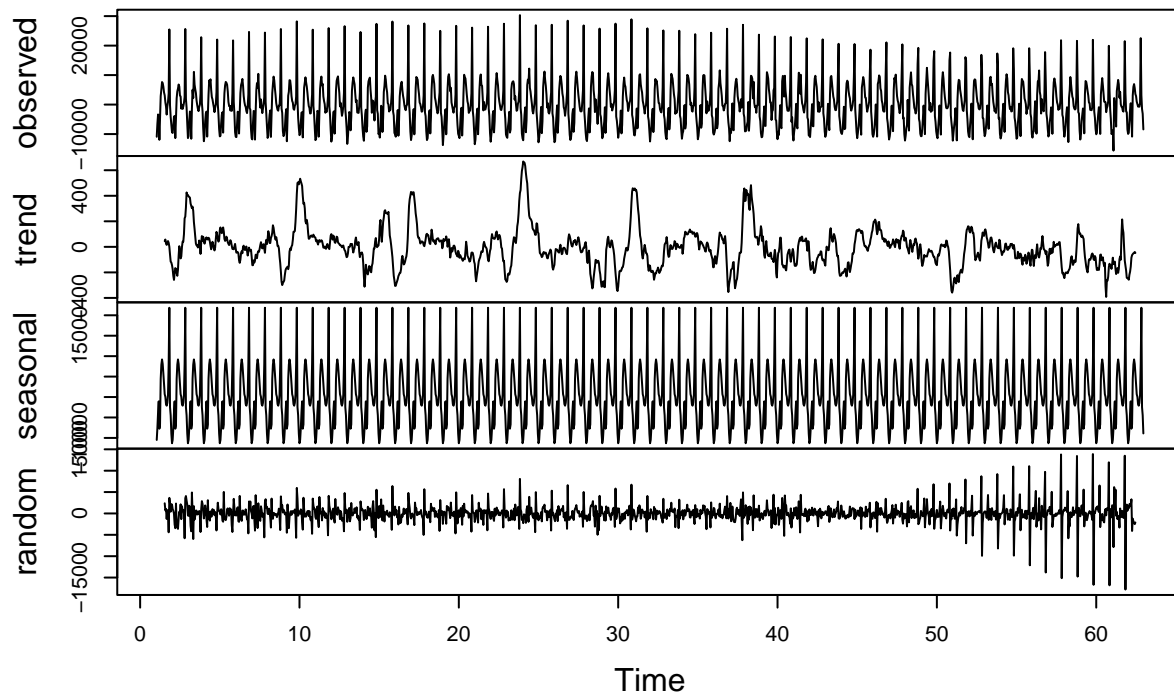
4 Consommation électrique

Decomposition of additive time series



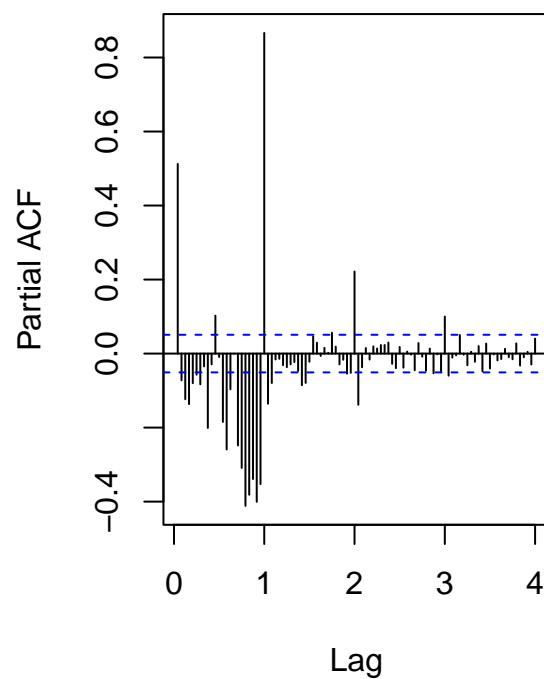
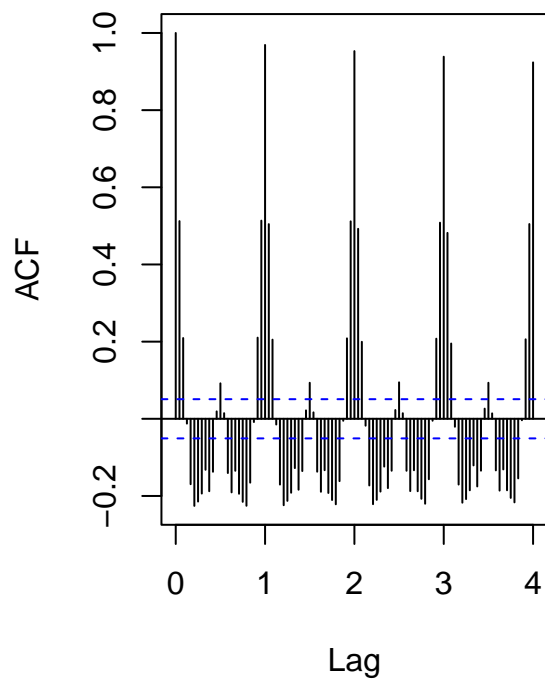
La trend est comprise entre environ 70000 et 100000 kWh (amplitude de 30000 kWh). Cette dernière est globalement croissante sur le mois de juillet puis décroissante sur le mois d'août. On constate une saisonnalité journalière avec une amplitude de 60000 kWh (consommation minimale à 6h et maximale à 20h) Enfin, la partie résiduelle **A COMPLETER**

Decomposition of additive time series



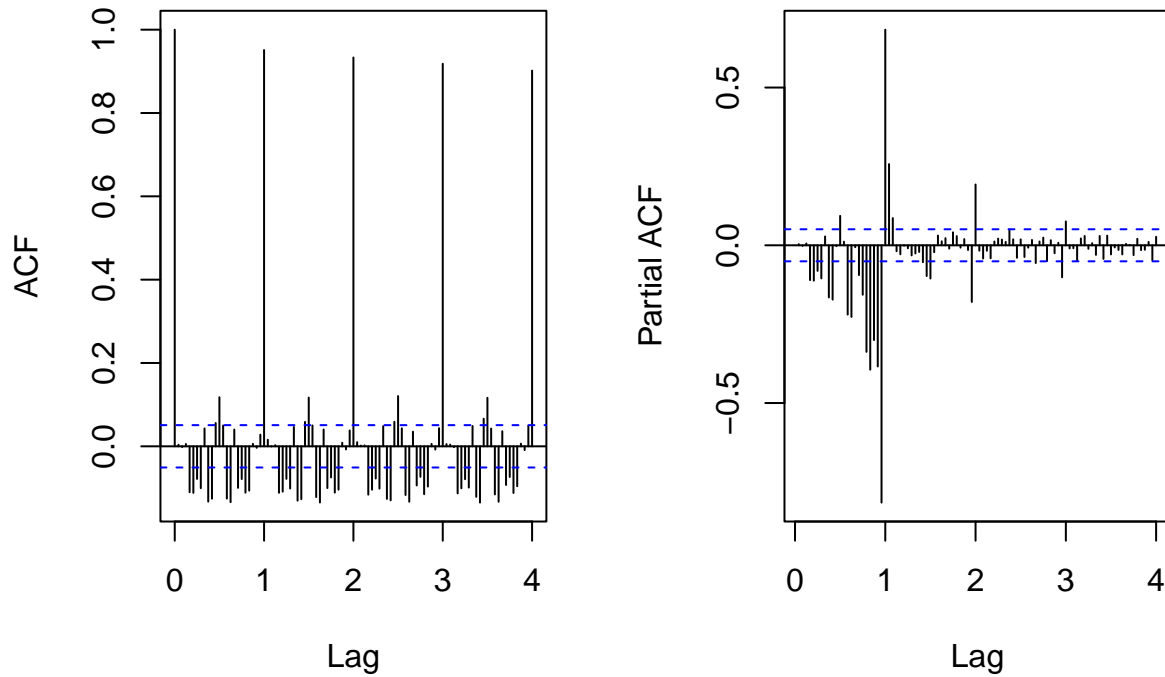
Series diff(ts_conso)

Series diff(ts_conso)



L'acf et le pacf montre une saisonnalité claire. Essayons d'abord de modéliser la consommation à l'aide d'un modèle ARIMA : si ce dernier ne s'avère pas assez précis, nous nous orienterons vers un modèle SARIMA. L'acf et le pacf suggère d'abord un modèle ARIMA(1,1,2)

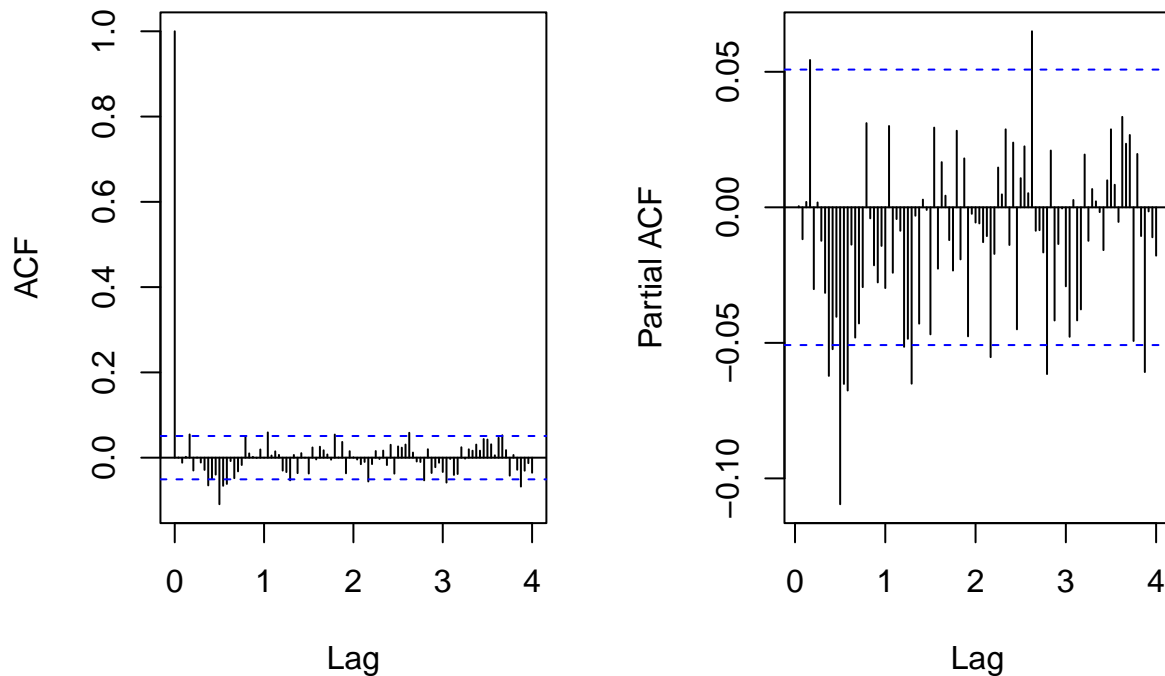
Series mod_conso_arima\$residu **Series mod_conso_arima\$residu**



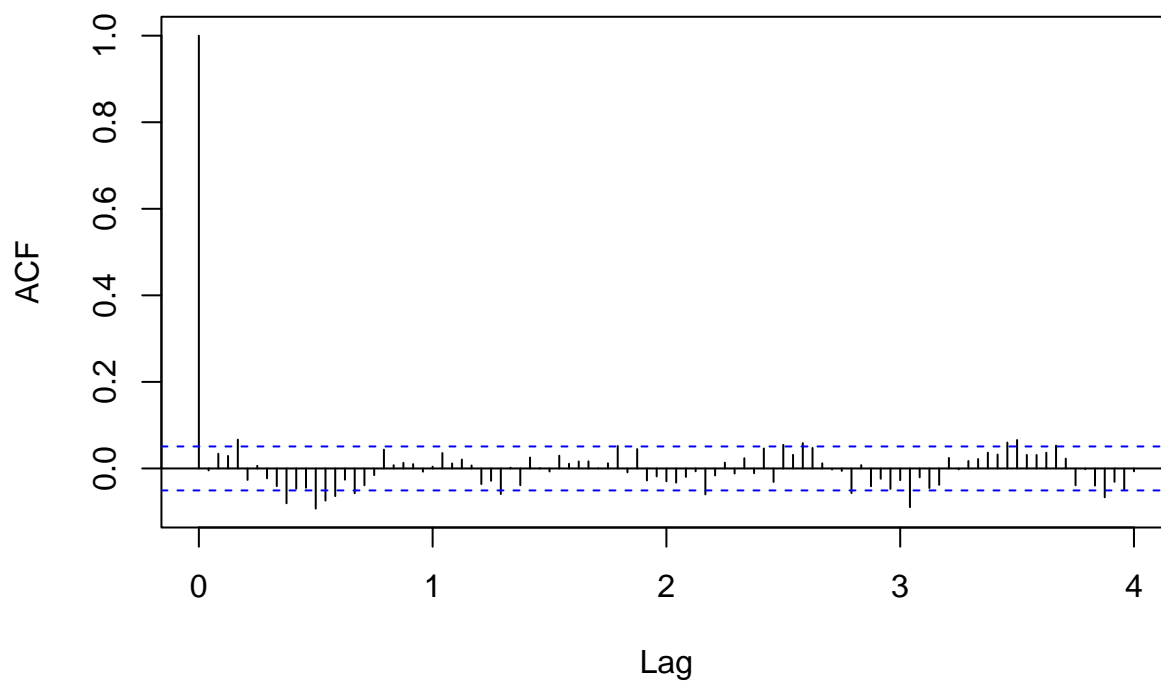
On constate mieux l'aspect saisonnier de la série temporelle. Les pics sur l'acf et le pacf suggère alors un modèle SARIMA(1,1,2)(1,1,1)[24].

```
## Warning in log(s2): Production de NaN
## Warning in log(s2): Production de NaN
```

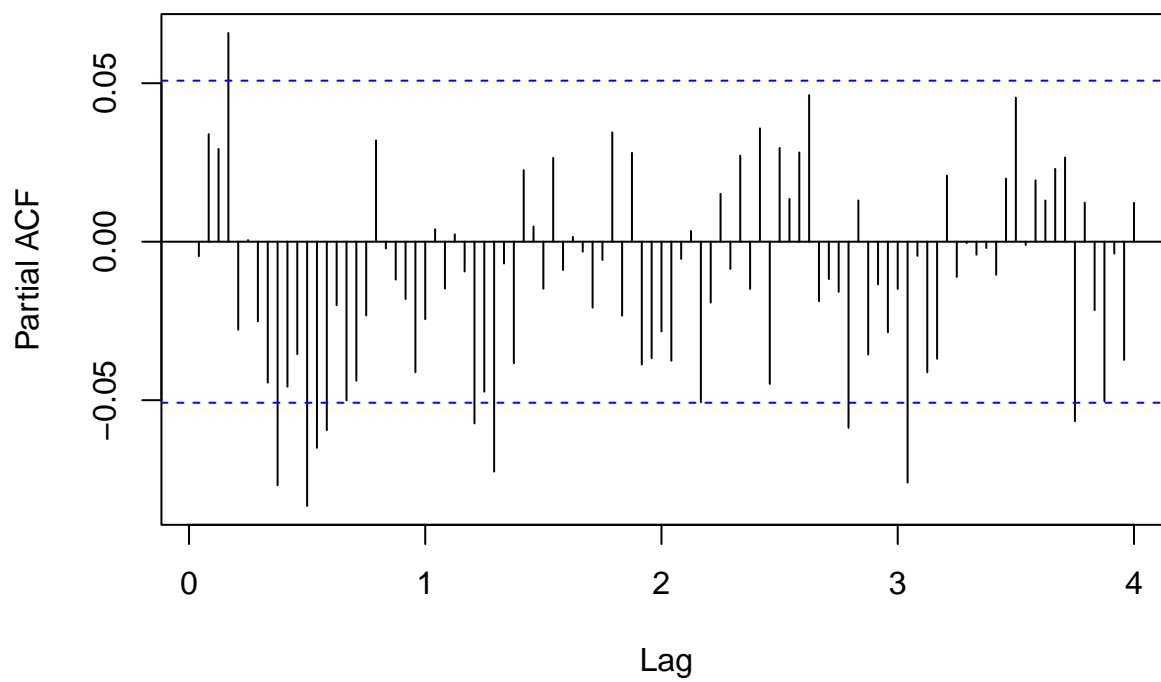
Series mod_conso_sarima\$residu **Series mod_conso_sarima\$residu**



Series auto_arima_conso\$residuals



Series auto_arima_conso\$residuals



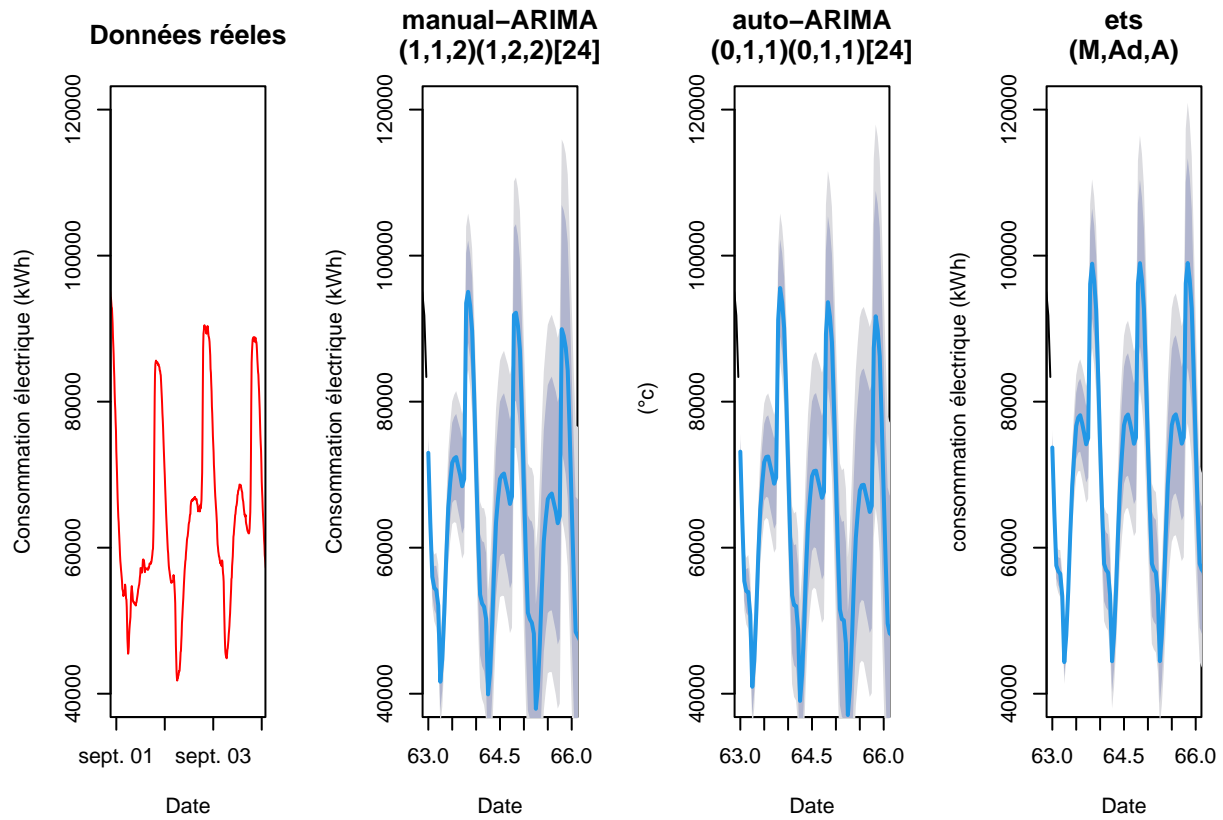


Table 2: Performances de prédiction des modèles

Modele	AIC	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Manual ARIMA	24675.046	-3.313	1198.751	873.273	-0.022	1.056	0.148	0
Auto ARIMA	25053.908	-3.893	1248.279	914.898	-0.003	1.105	0.343	-0.005
ETS	32443.877	15.954	1322.303	979.495	-0.025	1.18	0.367	0.171