

# 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 km<sup>2</sup>, 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.

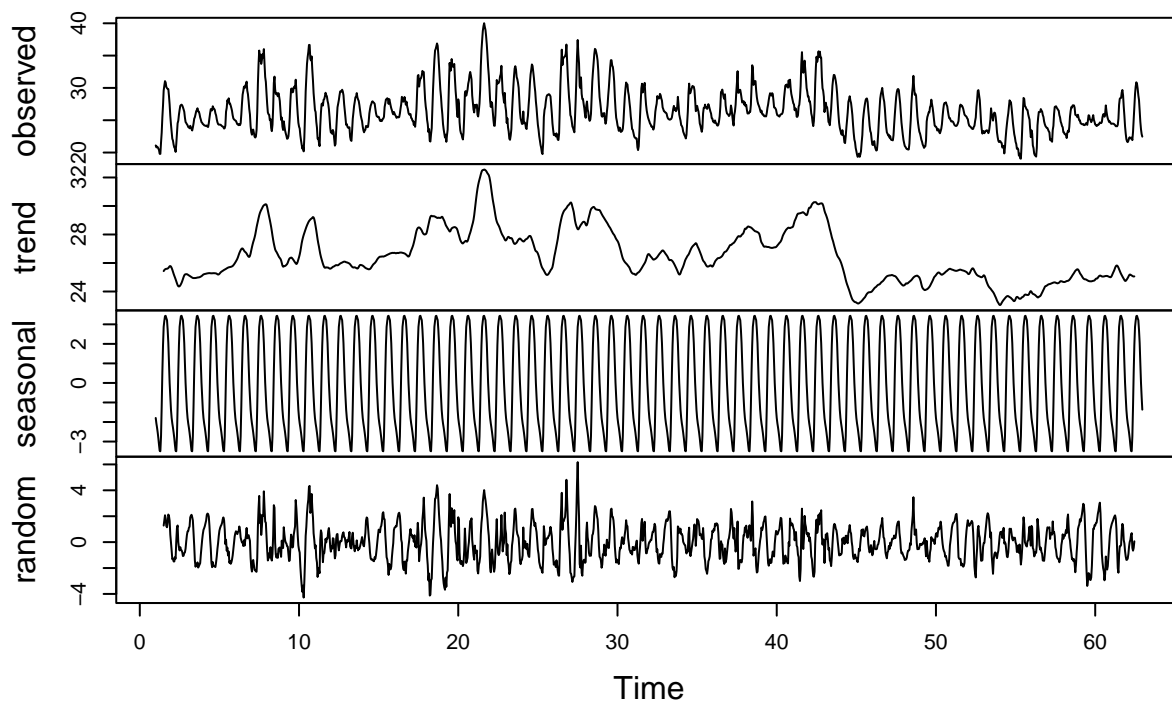
L'analyse sera séparée en 3 temps : l'analyse de la variable *Temperature*, l'analyse de la variable *conso* et l'analyse de la corrélation entre les deux mesures.

	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

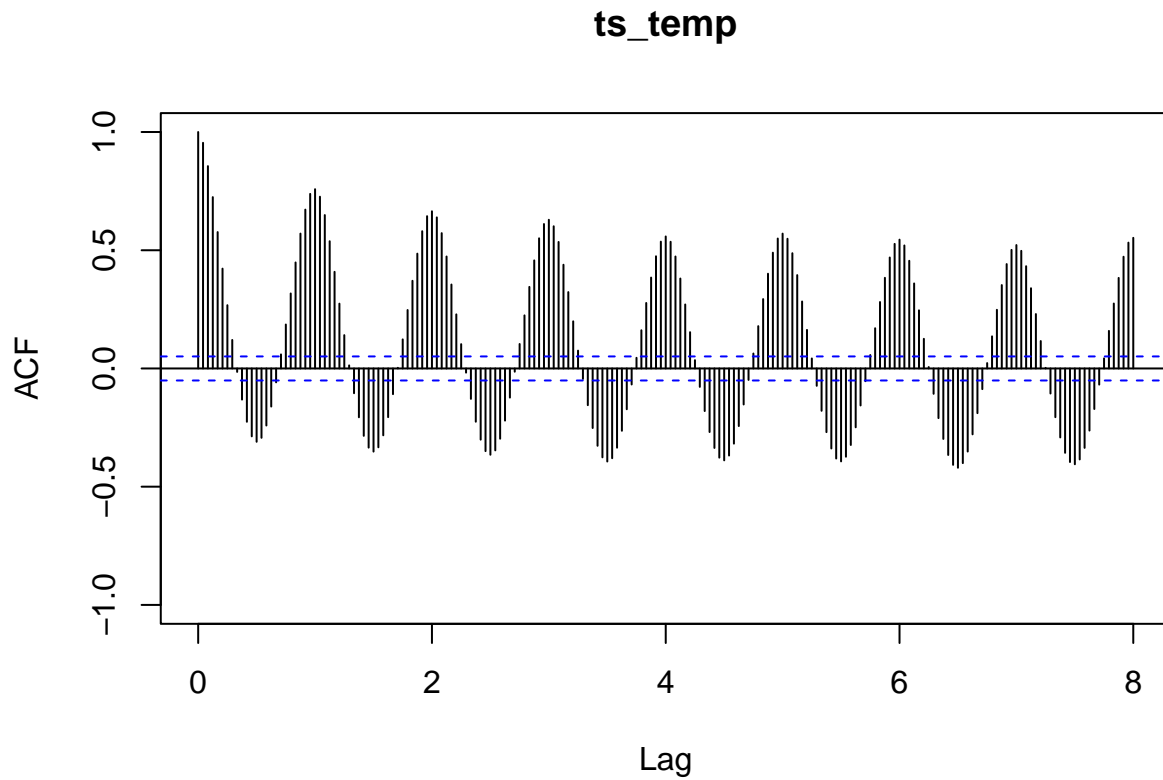


A partir de la décomposition de la série temporelle *st\_temperature* :

- 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 **\*\*CALCULER LES MOYENNES\***
- 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**)

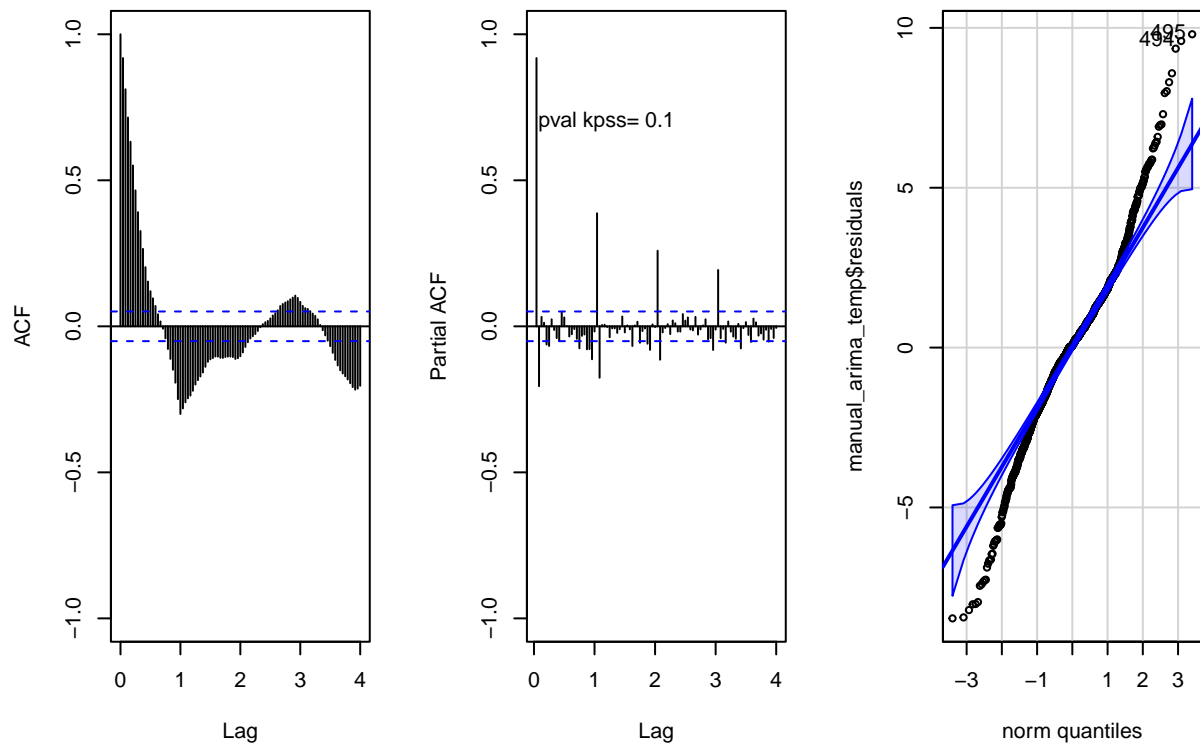
Comme la trend a une amplitude de moins de  $10^{\circ}\text{C}$  sur une période de plus de 60 jours, il ne semble pas nécessaire de modéliser la tendance autrement que par la température moyenne de cette dernière sous peine de complexifier le modèle. Néanmoins, on peut essayer d'ajuster un modèle linéaire. Sans grande surprise, le coefficient directeur de la droite de regression est de l'ordre de  $-10^{-3}$ . Nous ferons le choix de ne pas modéliser la trend pour le moment : lorsque nous effectuerons le test de kpss de stationnarité, celui-ci montrera s'il reste une part de tendance dans notre partie résiduelle.

Maintenant, analysons l'ACF et le Partial ACF de la série temporelle afin de pouvoir sélectionner et paramétrer au mieux le modèle final :



- La forme de sinusoïde de l'ACF montre une saisonnalité claire à chaque lag (toutes les 24h) (ce qui vient confirmer l'intuition de départ). Ainsi, nous allons commencer par modéliser notre série avec un modèle SARIMA en ajoutant une différenciation saisonnière ( $D=1$ ). Le test kpss donnera ensuite une indication sur la stationnarité ou non des résidus de notre modèle **ATTENTION A LA FORMULATION**.

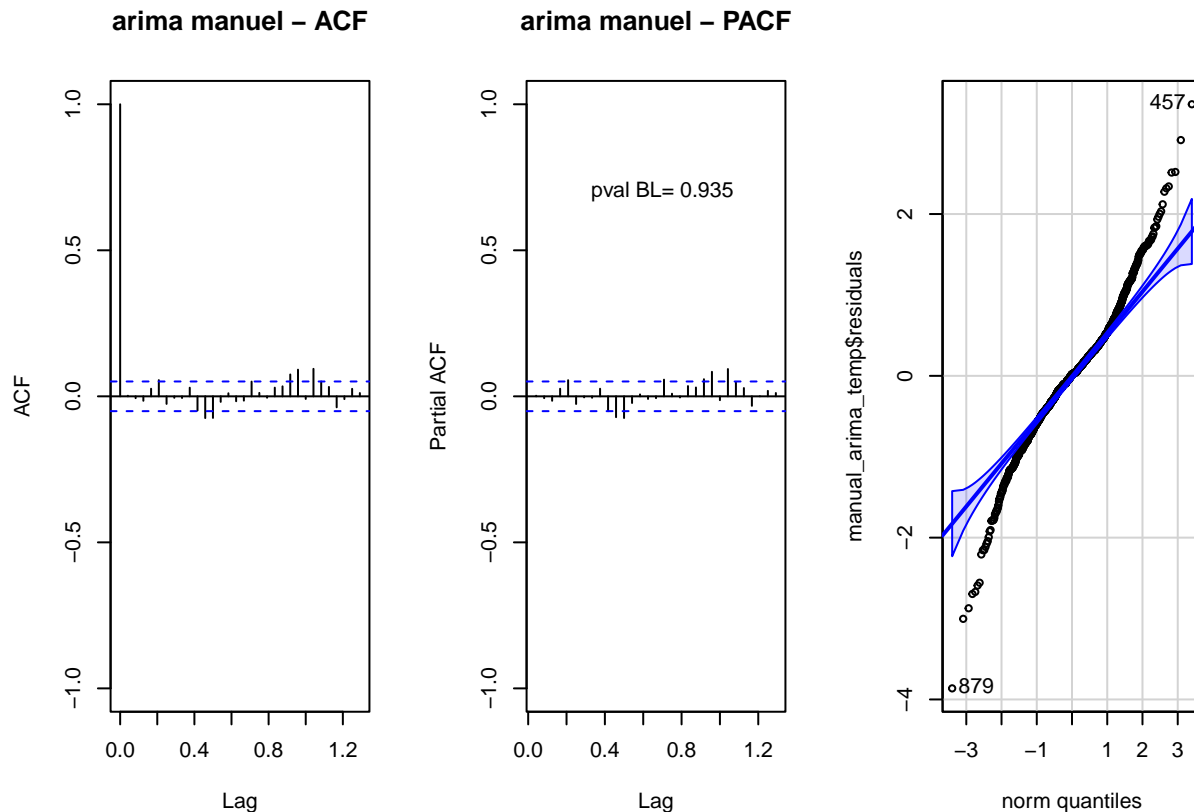
```
series manual_arima_temp$resid
```



```
## [1] 495 494
```

- l'ACF nous montre que la différenciation saisonnière appliquée a bien permis de supprimer la saisonnalité (le pattern sinusoidal a disparu).
- Le test kpss (non-significatif) indique que les résidus de notre modèle sont stationnaires : en plus de la suppression effective de la saisonnalité par différenciation, le test nous montre aussi qu'il n'est pas nécessaire de modéliser la trend par différenciation car celle-ci n'est pas présente dans nos résidus (elle n'influence pas la stationnarité dans notre cas).
- Le fait que l'ACF et le PACF sont significatifs à chaque lag nous pousse à appliquer un terme auto-régressif (PACF) et de moyenne mobile (ACF) tout deux saisonniers ( $P = 1$ ,  $Q = 1$ ).
- Sur le PACF, deux pics sont significatifs pour chaque période, ce qui nous pousse à prendre un terme auto-régressif d'ordre 2 ( $p = 2$ ).

Analysons maintenant les résidus de notre modèle final SARIMA(2,0,0)(1,1,1)[24]



```
## [1] 879 457
```

Selon les deux graphiques, la quasi-totalité des variations significatives ont été captées par notre modèle. Le test de Box vient confirmer cette intuition en montrant que les résidus de notre modèle sont bien un bruit blanc (test non significatif).

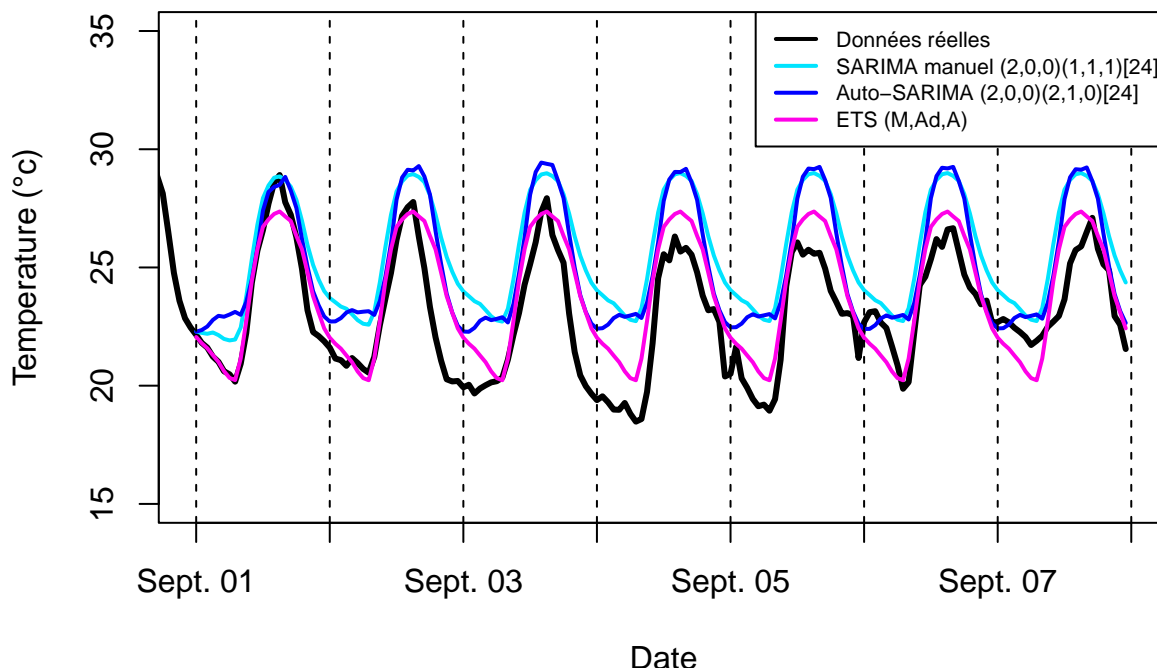
### PARLER EN TERME D'AIC

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

En réalité, nous n'effectuons pas de tests paramétriques sur nos modèles. De plus, l'objectif principal étant la modélisation à très court-terme, nous avons 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.

## Comparaison des modèles de prévision de température



Selon ce graphique, nous pouvons voir que les 3 modèles présentent des différences significatives. Le modèle ETS se démarquant davantage des modèles SARIMA. Ainsi, sur les 7 jours : - Le modèle manuel ARIMA(2,0,0)(1,1,1)[24] capture relativement bien le pic saisonnier du premier jour. Cependant, il surestime les valeurs dans la période creuse, réagit mal aux chutes abruptes de températures et sa précision lors des pics hauts décroît fortement lors des jours suivants (car la saisonnalité n'est pas la même en septembre ?). - Le modèle automatique ARIMA(2,0,0)(2,1,0)[24] capture le pic de manière similaire et possède une meilleure adéquation que le modèle manuel lors des chutes abruptes de température. Cependant, il surestime davantage les valeurs dans la période creuse par rapport au modèle manuel. Comme ce dernier, sa capacité d'estimation des pics et des variations abruptes décroît avec le temps. - Le modèle ETS (M,Ad,A) possède une précision élevée dans les creux : il lisse efficacement les variations brutales et ne les surestime pas. Cependant, il sous-estime le pic de température (probablement à cause d'une composante saisonnière trop lissée).

Ainsi, le modèle SARIMA manuel serait plus adapté pour les saisonnalités marquées au détriment des changements brusques. Le modèle SARIMA automatique est plus flexible pour les fluctuations rapides mais toujours sujet au surajustement. ETS excelle pour les tendances lissées et les creux, cependant au détriment des pics de température. Pour finir, aucun de nos modèles n'est performant lorsque l'on dépasse le deuxième jour.

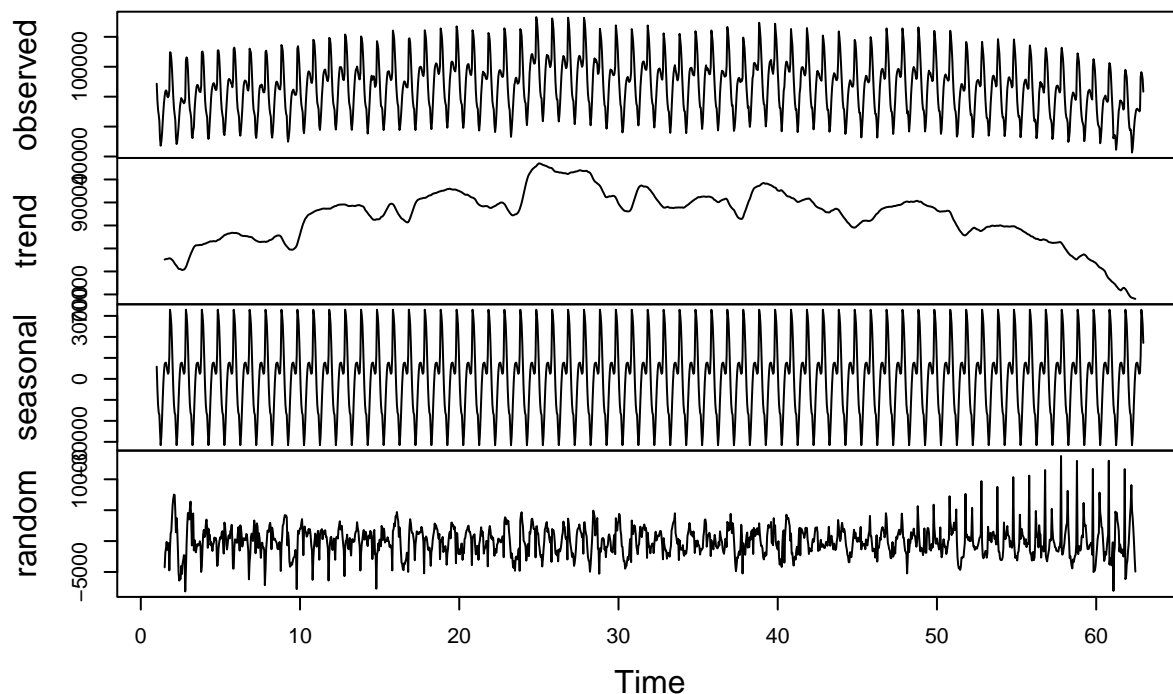
Table 1: RMSE des modèles pour la semaine du 01/09 au 07/09

Nom du modèle	AIC	jour_1	jour_2	jour_3	jour_4	jour_5	jour_6	jour_7
SARIMA	3158.98	1.81	3.34	3.61	3.74	3.32	2.29	2.31
Auto-SARIMA	3421.59	1.73	2.94	2.89	3.13	2.92	2.03	1.79
ETS	9857.85	1.18	2.15	2.13	1.91	1.66	1.10	1.23

Comme on peut le voir sur ce tableau, le modèle ETS reste, en moyenne, celui qui prédit le mieux avec une variation de l'erreur moyenne ne dépassant pas les 2%. Pour les modèles SARIMA, celui constitué manuellement est en moyenne moins précis que le modèle automatique sur les 7 jours.

## 4 Consommation électrique

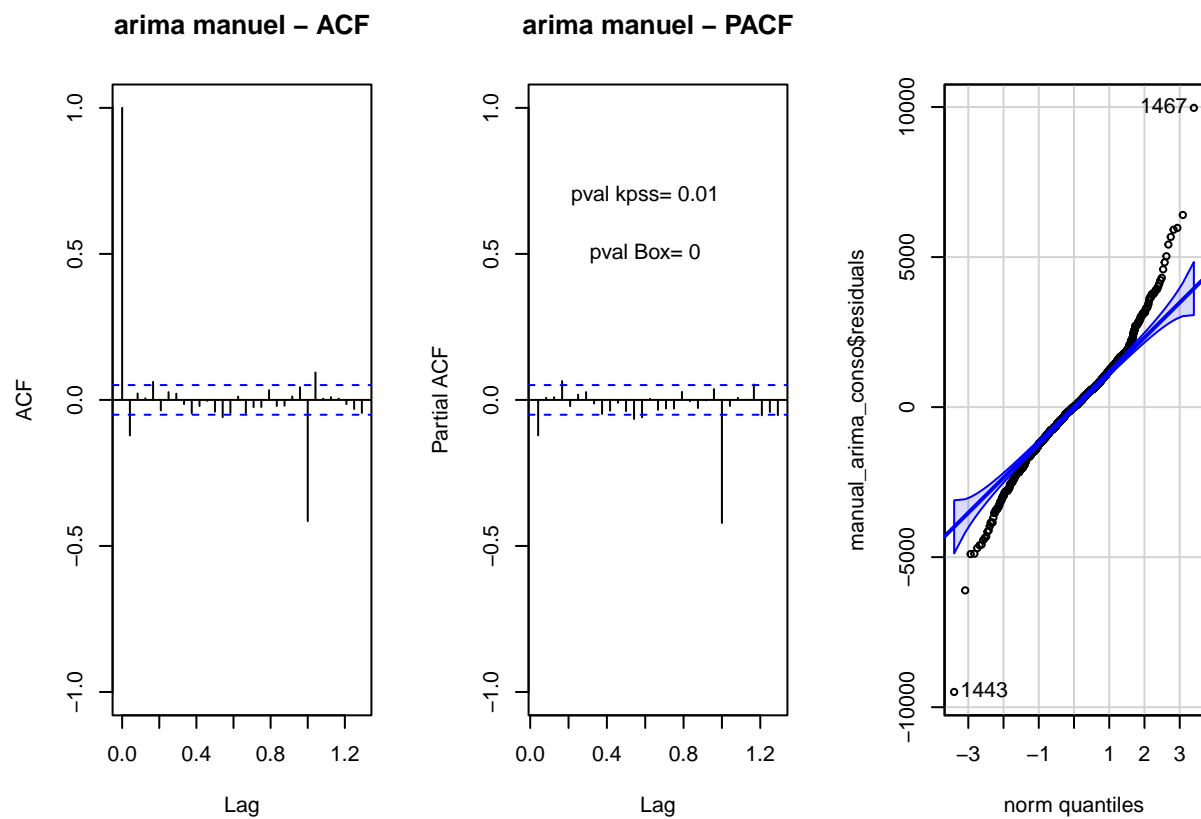
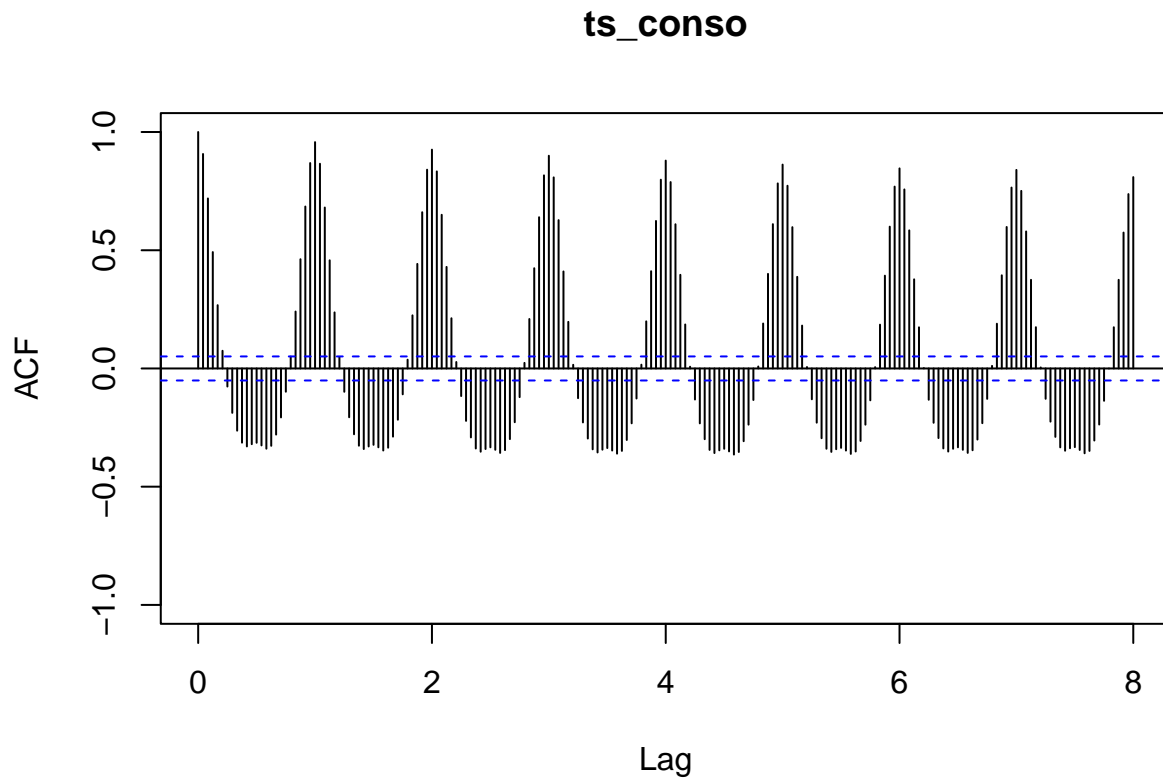
### Decomposition of additive time series



A partir de la décomposition de la série temporelle *ts\_consommation* : 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)

- La tendance est ici non négligeable, comprise entre 70000 et 100000 kWh (amplitude de 30000 kWh). Il sera donc important d'appliquer une différenciation dans notre modèle afin de capter cette tendance.
- Il y'a une forte saisonnalité journalière ici aussi, avec une amplitude de 60000 kWh (consommation minimale à 6h et maximale à 20h) : en effet, en partant des analyses de température, celles-ci chutent la nuit à une température acceptable sans chauffage (20°C) et remontent en journée vers des températures très hautes (entre 30 et 40 degrés) nécessitant, si disponible, l'utilisation de climatisation **PARLER AUSSI DE LA PRESENCE A LA MAISON ET DU FAIT QUE C'EST LA CONSO TOTALE ET NON PAS DES MENAGES**, qui est source de consommation énergétique. Ainsi, selon cette intuition, cette saisonnalité serait présente à chaque temps (toutes les 24h). Il sera aussi important d'appliquer une différenciation saisonnière afin de capter la saisonnalité de nos données.

Afin de confirmer nos intuitions, regardons l'ACF de la série temporelle :



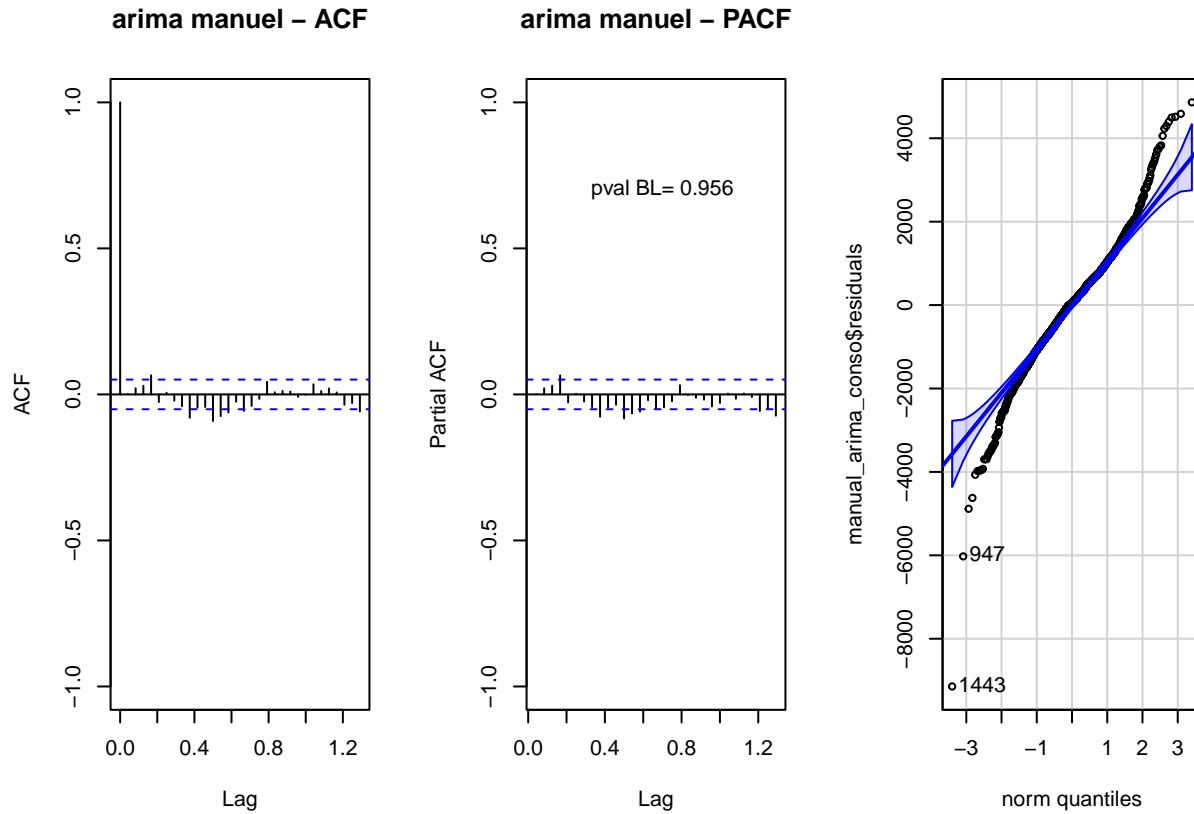
```
## [1] 1467 1443
```

- L'ACF et le PACF montrent des pics significatifs à chaque période (tout les 24 pics) montrant ainsi une saisonnalité dans les deux graphs **A REFORMULER**. Afin de corriger cela, nous appliquerons



un paramètre auto-régressif et de moyenne mobile saisonnier d'ordre 1 ( $P = 1, Q = 1$ ).

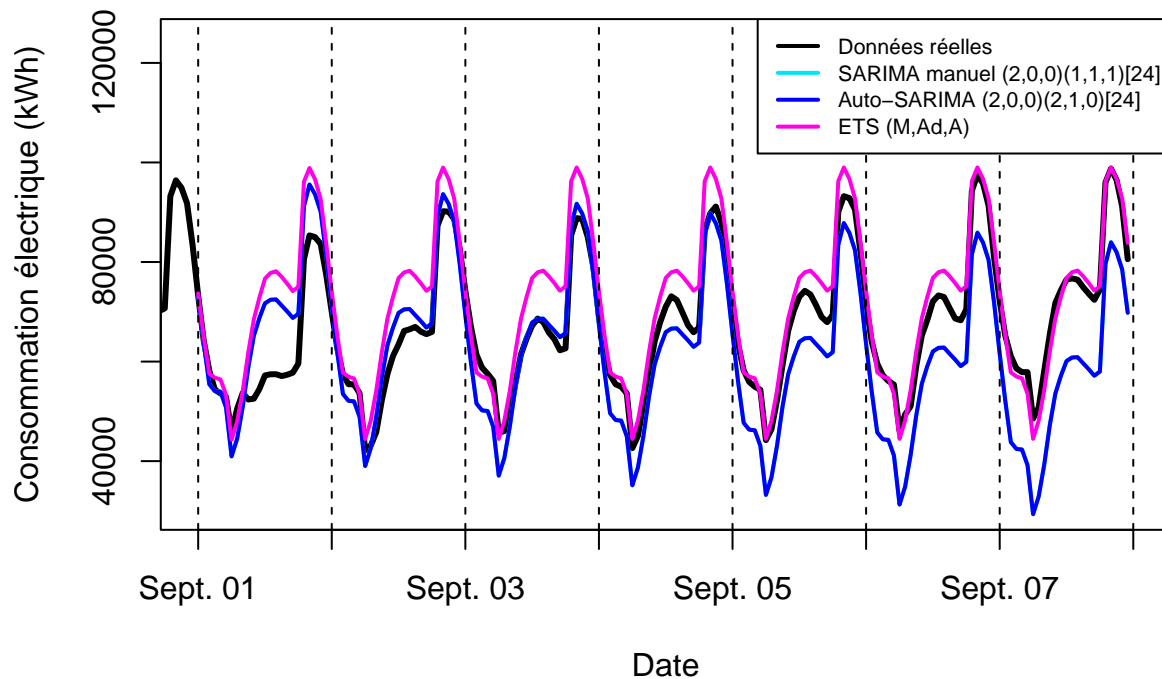
- On voit le premier pic du PACF significatif. Ainsi, nous allons appliquer un terme auto-régressif d'ordre  $p = 1$ .



```
## [1] 1443 947
```

Selon les deux graphiques, la quasi-totalité des variations significatives ont été captées par notre modèle. Le test de Box vient confirmer cette intuition en montrant que les résidus de notre modèle sont bien un bruit blanc (test non significatif).

## Comparaison des modèles de prévision de consommation électrique



Le graphique nous montre un événement spécial lors du premier jour, événement absolument pas capté par les modèles de prévisions.

Selon ce graphique : - On voit que les modèles SARIMA manuel et automatique sont très similaires dans leurs estimations. Les pics de basse consommation sont bien estimés lors des deux premiers jours mais sous-estimés de manière croissante ensuite. Ils estiment bien les pics hauts à partir du jour 2 (**POURQUOI**) et capturent davantage les variations intermédiaires (chute partielle de consommation avant atteinte du pic). Les variations abruptes sont aussi bien estimées. Cependant, ces deux modèles tendent à sous-estimer les valeurs réelles avec le temps, toute heure confondue. - Le modèle ETS est meilleur que les modèles SARIMA pour estimer les pics base de consommation, mais il a tendance à surestimer les pics hauts, surtout lors des premiers jours. Aussi, malgré une moins bonne estimation des variations intermédiaires et des pics hauts comparé aux deux autres modèles, ETS estime mieux et de manière plus stable la consommation, là où les modèles SARIMA perdent fortement en précision dans le temps.

Les pics journaliers montrent qu'ETS surestime mais suit la tendance globale, tandis que les modèles SARIMA divergent. Les creux nocturnes relèvent une sous-estimation critique des modèles SARIMA. Ainsi, pour la gestion des pics, ETS est plus fiable grâce à sa capacité à prévoir des amplitudes élevées, et est aussi optimal pour les prévisions nocturnes.

Table 2: RMSE des modèles pour la semaine du 01/09 au 07/09

Nom du modèle	AIC	jour_1	jour_2	jour_3	jour_4	jour_5	jour_6	jour_7
SARIMA	25054.82	9130.53	5902.42	6479.52	7904.91	10468.77	12750.60	17335.14
Auto-SARIMA	25053.91	9132.03	5895.50	6480.20	7917.95	10492.39	12780.70	17371.33
ETS	32443.88	12422.99	8709.75	8681.81	7981.13	7169.48	7294.17	7667.31

Comme on le voit sur ce tableau, le modèle ETS est en moyenne moins précis que les modèles SARIMA pour les 4 premiers jours (différence d'erreur de 3000 KWh), mais cette tendance s'inverse ensuite et ETS devient largement plus précis que les modèles SARIMA (différence d'erreur moyenne de 10000 KWh à j+7).

Les pics journaliers montrent qu'ETS surestime mais suit la tendance globale, tandis que les modèles SARIMA divergent. Les creux nocturnes relèvent une sous-estimation critique des modèles SARIMA. Ainsi, pour la gestion des pics, ETS sera plus fiable grâce à sa capacité à prévoir des amplitudes élevées, et est aussi optimal pour les prévisions nocturnes. Cependant, si l'on cherche à estimer la consommation au court terme dans les périodes intermédiaires, les modèles SARIMA seront plus adaptés. Enfin, aucun des modèles n'arrive à capter le changement brutal du premier jour (**QUEL JOUR C'EST**). Ces modèles ne sont donc pas robustes aux événements rares et critiques.