

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.

Dans ce travail, nous nous intéressons à l'analyse de la consommation énergétique en fonction des variations de température dans la période d'été (Juillet-Aout). Nous explorerons 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, nous explorerons plusieurs moyens de calcul de corrélation entre demande énergétique et 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 forte et l'atmosphère est sec 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 3 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 d'été (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 à 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 aberrante n'est à déplorer dans nos données.

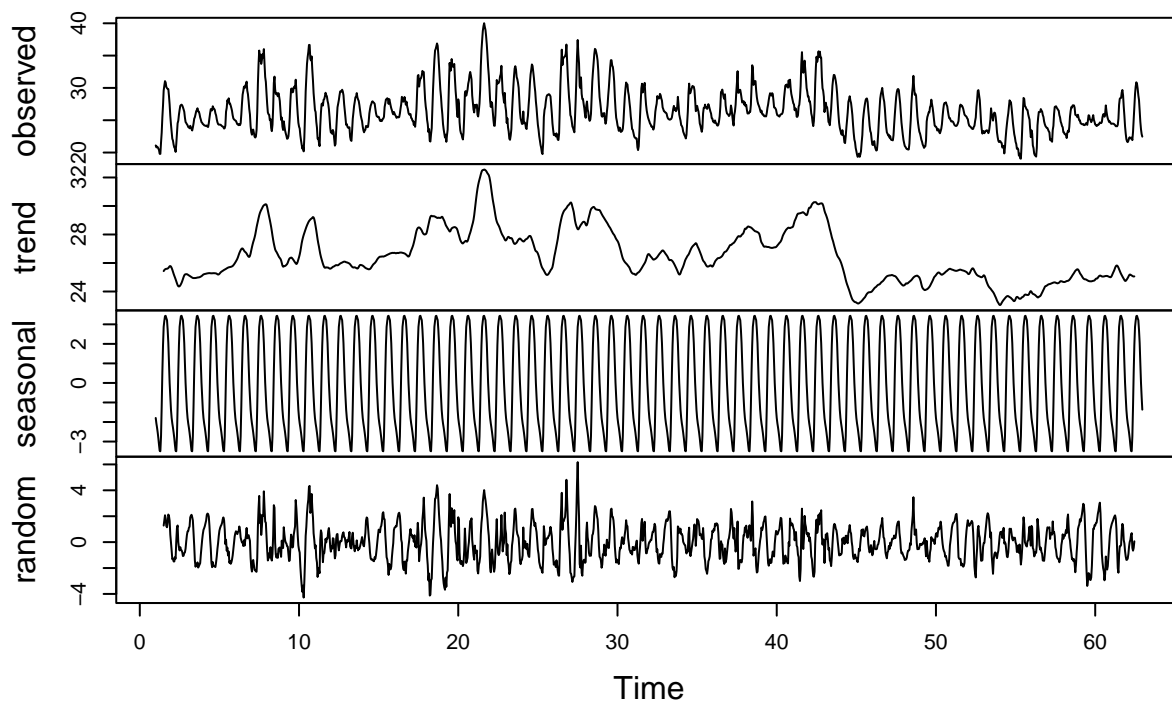
L'analyse sera séparée en 3 temps : l'analyse de la variable *Temperature*, l'analyse de la variable *conso* et l'analyse de la corrélation entre les deux mesures.

	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

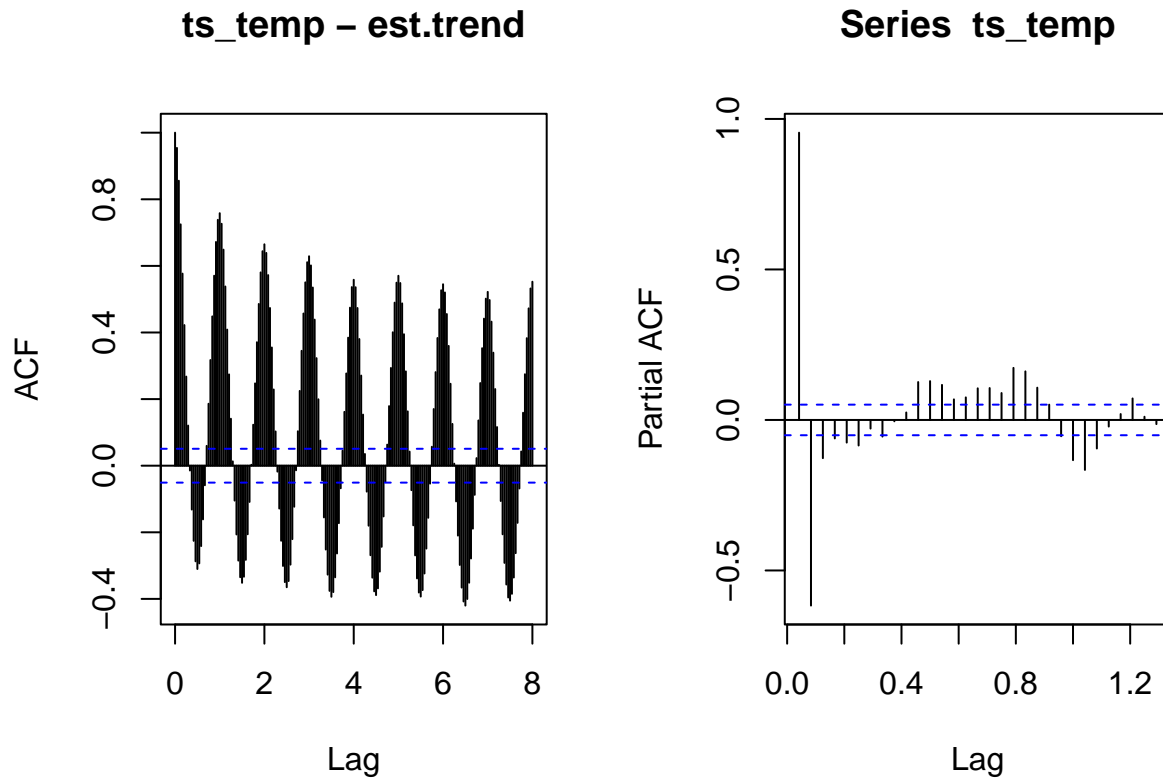


A partir de la décomposition de la série temporelle *st_temperature* :

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

Selon ces premières intuitions, nous avons fait le choix de ne pas tenter de modéliser la trend de manière linéaire car les tentatives de modélisations quadratiques et supérieures n'était pas significativement différent de la modélisation linéaire simple, rajoutant cependant une quantité non négligeable de paramètres au modèle final. Ce choix va cependant 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 notre série. **ATTENTION ON DOIT CHOISIR SI OUI OU NON ON MODELISE LA TREND ET EXPLIQUER EXPLICITEMENT LE CHOIX**

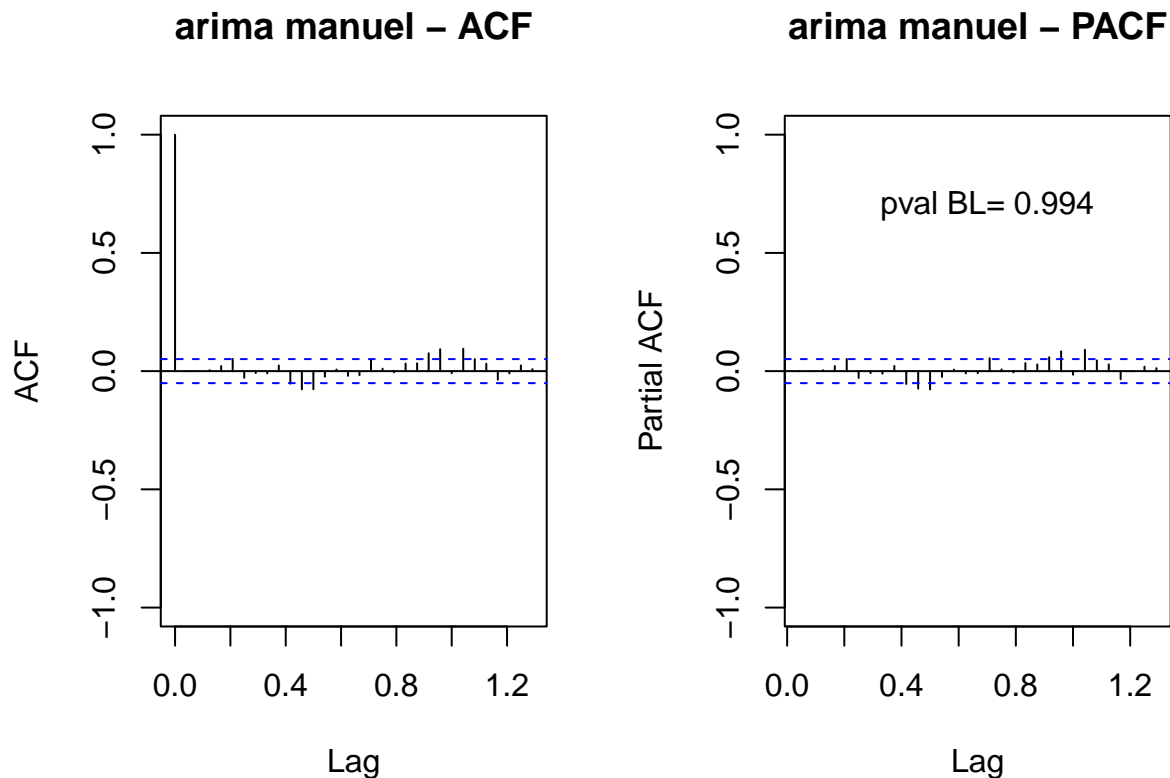
Maintenant, analysons l'ACF et le Partial ACF de notre série temporelle afin de pouvoir sélectionner et paramétrer au mieux notre modèle final :



- Pour l'ACF, la sinuïsalité montre une saisonnalité claire, toutes les 24h (ce qui vient confirmer notre intuition de départ). La décroissance lente nous pousse à inclure un terme de moyenne mobile saisonnier de 1.
- Pour le partial ACF,

Le **test de racine unitaire augmenté de Dickey Fuller** sur la partie résiduelle de nos données étant **significatif**, 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 de notre modèle semblent homoscedastiques : 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 heteroscedastiques (**test de Breuch-Pagan d'heteroscedasticité significatif**) **EXPLIQUER L'INTUITION AVEC LA TEMPERATURE** Les résidus de notre modèle ne suivent pas une loi normale : malgré l'absence de skew (**distribution centrée**), la distribution de nos résidus reste **fortement leptokurtique**.

En réalité, nous n'effectuons pas de test paramétriques sur nos modèles. De plus, notre objectif étant la modélisation à très court-terme, nous nous choisis de négliger les hypothèses d'homoscedasticité (parfaite) et de normalité des résidus. **ATTENTION BIEN EXPLIQUER POURQUOI ET PARLER DES INTERVALLES DE CONFIANCE**

Afin de s'assurer de la pertinence de notre modèle, nous allons le comparer aux modèles conçus par les algorithmes *auto.arima* (configuré en fonction du BIC **EXPLIQUER POURQUOI**) et modèles 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.

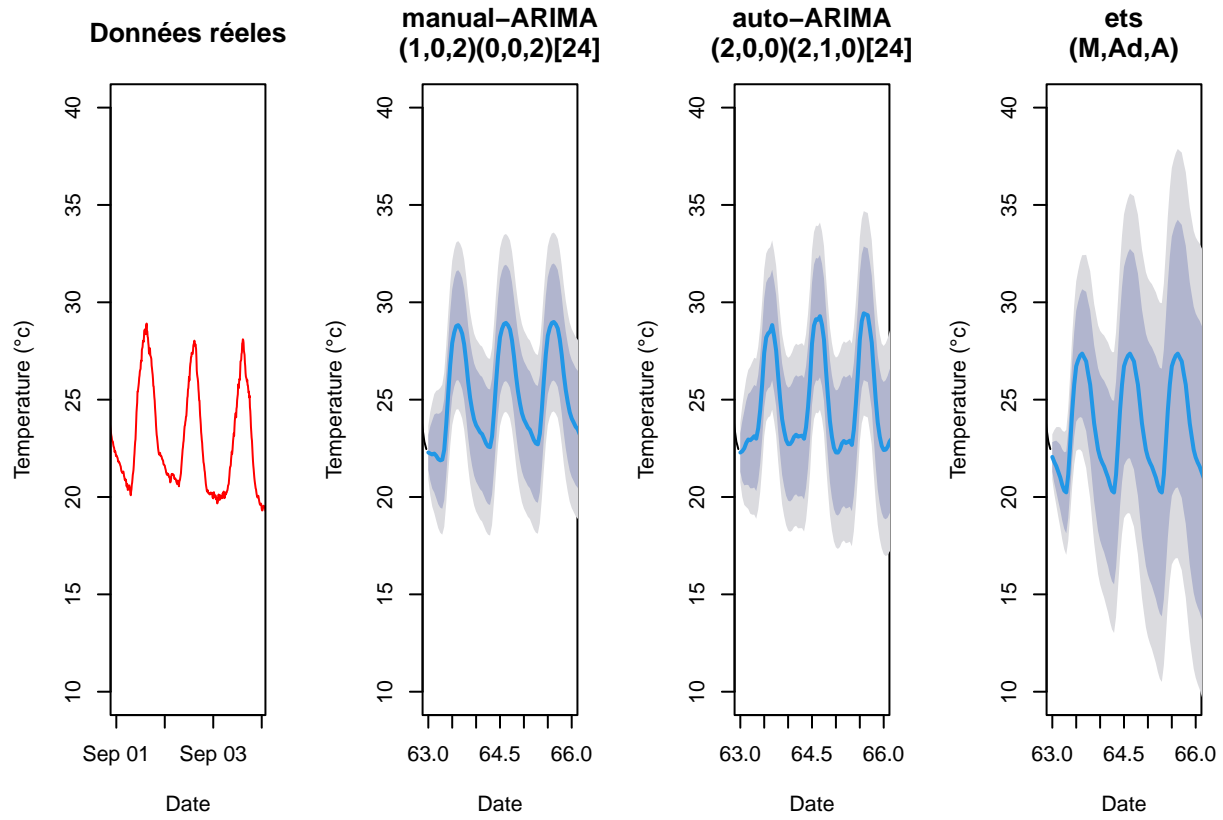


Table 1: Performances de prédiction des modèles

Modele	AIC	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Manual ARIMA	3160.548	-0.004	0.691	0.498	-0.072	1.848	0.656	0
Auto ARIMA	3421.592	0	0.766	0.554	-0.047	2.056	0.321	0.006
ETS	9857.847	-0.001	0.733	0.537	-0.038	1.992	0.311	0.26