

Chapitre 2

Modèles nécessaires à la régulation

Pour anticiper l'état futur du bâtiment et s'adapter aux usages et conditions environnementales changeantes, les stratégies de régulation doivent reposer sur l'utilisation de différents modèles. L'objectif de ce chapitre est de les présenter. Tout d'abord, nous décrivons le modèle thermique dynamique du bâtiment utilisé. Ensuite, un rapide état de l'art sera réalisé sur la modélisation de l'occupation et en particulier sur la modélisation du confort de l'occupant. Enfin, nous nous concentrerons sur le développement d'un modèle de prévision météorologique locale.

PLAN DU CHAPITRE 2

1.	<u>INTRODUCTION.....</u>	<u>51</u>
2.	<u>MODÈLE THERMIQUE DYNAMIQUE DU BÂTIMENT (COMFIE)</u>	<u>51</u>
3.	<u>MODÉLISATION DE L'OCCUPATION</u>	<u>55</u>
3.1	INTRODUCTION	55
3.2	ÉTAT DE L'ART SUR LE CONFORT	55
3.2.1	LE CONFORT THERMIQUE.....	56
3.2.2	MODÉLISATION	60
4.	<u>MODELISATION DU CLIMAT</u>	<u>61</u>
4.1	INTRODUCTION	61
4.2	ÉTAT DE L'ART DE LA MODÉLISATION DES CONDITIONS CLIMATIQUES	62
4.2.1	VARIABLES CLIMATIQUES À MODÉLISER	62
4.2.2	LES MODÈLES DE PRÉVISIONS MÉTÉOROLOGIQUES	62
4.3	CARACTÉRISATION DES MODÈLES ARMA.....	67
4.3.1	ÉLIMINATION DE LA TENDANCE ET DE LA SAISONNALITÉ	67
4.3.2	LES PROCESSUS ARMA	69
4.3.3	ESTIMATION ET CHOIX D'UN MODÈLE	72
4.4	MODÉLISATION DES CONDITIONS CLIMATIQUES.....	74
4.4.1	INDICATEURS DE LA QUALITÉ DE LA PRÉVISION.....	74
4.4.2	MÉTHODES DE MODÉLISATION DÉVELOPPÉES.....	76
4.4.3	DONNÉES UTILISÉES.....	77
4.4.4	MODÉLISATION DE LA TEMPÉRATURE	79
4.4.5	MODÉLISATION DU RAYONNEMENT SOLAIRE.....	88
4.5	CONCLUSION	92
5.	<u>CONCLUSION DU CHAPITRE.....</u>	<u>93</u>

1. INTRODUCTION

L'objectif de cette thèse est de développer des stratégies de régulation adaptée au temps réel. Ces stratégies doivent pouvoir s'adapter en continu aux usages et conditions environnementales changeantes tout en anticipant l'état futur du bâtiment. Ainsi, elles reposent sur l'utilisation de différents modèles :

- un modèle thermique du bâtiment,
- un modèle d'occupation du bâtiment,
- un modèle de prévision météorologique.

La validité du modèle thermique dynamique est un critère indispensable à la mise en place d'une régulation efficace. En effet, seule la fiabilité du modèle assurera la robustesse de la commande optimale calculée. Le temps de simulation du modèle est également un critère important, l'optimisation faisant de multiples appels au modèle du bâtiment. Enfin, concernant le système de chauffage sur lequel la régulation va s'appliquer, celui-ci est directement intégré dans le modèle thermique dynamique du bâtiment.

Pour fonctionner, le modèle thermique requiert des données d'occupation, telles que les heures de présence, l'utilisation des équipements ou encore les contraintes de température de confort à respecter. Dans cette étude, une modélisation simplifiée de l'occupation a été considérée étant données les nombreuses incertitudes dans ce domaine.

Enfin, le modèle thermique du bâtiment nécessite l'utilisation de données météorologiques telles que la température extérieure et le rayonnement. Dans le cadre de la régulation, le développement des stratégies se fait tout d'abord dans un cas idéal où la météorologie est parfaitement connue à l'avance. Des données mesurées ou des données de type TRY (*Test Reference Year*) (Lund 1985) représentant la climatologie typique d'une région sont alors utilisées. Ensuite, la robustesse des stratégies de régulation développées est évaluée dans des cas plus réalistes où les prévisions météorologiques ne sont pas connues parfaitement à l'avance. Il est alors nécessaire de développer un modèle basé sur les prévisions météorologiques régionales pour anticiper l'évolution du climat local.

2. MODÈLE THERMIQUE DYNAMIQUE DU BÂTIMENT (COMFIE)

Le modèle thermique dynamique du bâtiment utilisé dans cette thèse repose sur le concept de « zone thermique », sous-ensemble du bâtiment considéré à température homogène (Peuportier et Blanc-Sommereux 1990). Pour chaque zone, les murs sont décomposés en mailles (méthodes des volumes finis) sur lesquelles un bilan thermique est appliqué en supposant que la température de la

Chapitre 2 : Modèles nécessaires à la régulation

maille est uniforme. Pour assurer cette hypothèse, le maillage est d'autant plus fin que la maille est proche de l'ambiance intérieure. L'air, le mobilier et les cloisons légères éventuelles contenus dans la zone sont regroupés dans une maille unique. Le bilan thermique sur chaque maille prend alors la forme suivante (Neveu 1984) :

$$C_m \frac{dT_m}{dt} = \dot{Q}_G - \dot{Q}_P \quad (2.1)$$

avec

- C_m la capacité thermique de la maille ;
- T_m la température de la maille ;
- \dot{Q}_G les gains comprenant les apports solaires et internes, ainsi que les puissances de chauffage ou de refroidissement des équipements ;
- \dot{Q}_P les pertes dues à la ventilation et aux transferts thermiques par conduction, convection, rayonnement.

Certains phénomènes non stationnaires (ouverture d'un volet) ou non linéaires (mouvement d'air) sont pris en compte dans un vecteur des sollicitations noté U , ce qui permet d'écrire les équations (2.1) sous la forme d'un système d'équations linéaires (2.2).

En répétant ce bilan pour chaque maille de la zone considérée et en ajoutant une équation de sortie, l'ensemble des équations peut ainsi être représenté sous la forme d'un système linéaire continu et invariant (Bacot 1984 ; Lefebvre 1987) :

$$\begin{cases} C\dot{T}(t) = AT(t) + EU(t) \\ Y(t) = JT(t) + GU(t) \end{cases} \quad (2.2)$$

avec

- T le champ discrétisé des températures ;
- C la matrice diagonale des capacités thermiques ;
- U le vecteur des sollicitations (comprenant les paramètres climatiques, le chauffage, etc.) ;
- A la matrice d'état ou matrice des termes d'échange entre mailles ;
- E la matrice de commande ou matrice des termes d'échange entre mailles et sollicitations ;
- J la matrice d'observation ou matrice reliant les sorties aux températures des mailles ;
- G la matrice d'action directe ou matrice reliant les sorties aux sollicitations ;
- Y le vecteur des sorties (ici la température intérieure de la zone prenant en compte l'air et la température des murs).

Le système étant linéaire, il est possible de décomposer le vecteur des températures $T(t)$ en un terme purement dynamique et un terme de régime permanent glissant (Palomo Del Barrio et al. 2000) :

Chapitre 2 : Modèles nécessaires à la régulation

$$T(t) = T_d(t) + T_p(t) \quad (2.3)$$

où $T_p(t)$ représente le comportement du système lorsque sa capacité thermique est supposée être égale à 0 (régime permanent). On obtient alors $T_p(t) = -A^{-1}EU(t)$.

En utilisant l'équation (2.3) dans l'équation (2.2), on obtient alors :

$$\begin{cases} \dot{T}_d(t) = C^{-1}AT_d(t) + A^{-1}E\dot{U}(t) \\ Y(t) = JT_d(t) + (G - JA^{-1}E)U(t) \end{cases} \quad (2.4)$$

Pour résoudre simplement ce système on diagonalise la matrice $C^{-1}A$. On réalise alors un changement de base en utilisant la matrice de passage P formée des vecteurs propres de $C^{-1}A$. Le vecteur des températures $T_d(t)$ est ainsi transformé en un vecteur d'état $X(t)$ par la relation suivante :

$$T_d(t) = PX(t) \quad (2.5)$$

Le système (2.4) devient alors :

$$\begin{cases} \dot{X}(t) = FX(t) + B\dot{U}(t) \\ Y(t) = HX(t) + SU(t) \end{cases} \quad (2.6)$$

avec

- $F = P^{-1}C^{-1}AP$
- $B = P^{-1}A^{-1}E$
- $H = JP$
- $S = G - JA^{-1}E$

F est une matrice diagonale dont le $i^{\text{ème}}$ terme diagonal est $-1/\tau_i$, τ_i étant la $i^{\text{ème}}$ constante de temps de la zone. Ces constantes de temps sont ordonnées dans F de la plus grande à la plus petite. L'avantage de ce changement de base est de pouvoir réduire l'ordre du système. En effet, certaines constantes de temps sont petites et les termes correspondants du vecteur $X(t)$ atteignent très rapidement leur régime permanent. Il est alors possible de les négliger dans le calcul de la partie dynamique (on considère qu'ils sont en régime permanent). On peut ainsi obtenir un modèle d'ordre réduit en ne conservant que les N plus grandes constantes de temps différentes (Marshall 1966). Suite à des analyses de sensibilité, la valeur maximale de N a été fixée à 10 dans le logiciel.

On obtient alors pour chaque zone thermique du bâtiment le modèle réduit suivant :

$$\begin{cases} \dot{X}_r(t) = F_r X_r(t) + B_r \dot{U}(t) \\ Y(t) = H_r X_r(t) + SU(t) \end{cases} \quad (2.7)$$

Il est alors possible de coupler les différents modèles de zones entre eux (Blanc Sommereux et Lefebvre 1989 ; Blanc Sommereux 1991) et d'aboutir au système couplé suivant (où $Y_g(t)$ est le vecteur des variables de couplage et où l'on sépare dans les matrices B_r et S les colonnes concernant U_g regroupées dans les matrices B_g^α et S_g^α , et celles concernant Y_g , regroupées dans B_g^β et S_g^β) :

$$\begin{cases} \dot{X}_g(t) = F_g X_g(t) + B_g^\alpha \dot{U}_g(t) + B_g^\beta \dot{Y}_g(t) \\ Y_g(t) = H_g X_g(t) + S_g^\alpha U_g(t) + S_g^\beta Y_g(t) \end{cases} \quad (2.8)$$

Chapitre 2 : Modèles nécessaires à la régulation

On peut alors intégrer le système ainsi obtenu sur un pas de temps Δt en notant X_g^{n+1} et X_g^n les vecteurs d'état aux temps $(n+1)\Delta t$ et $n\Delta t$ (on procède de même pour les vecteurs Y_g et U_g). Le système (2.8) devient alors après intégration :

$$\begin{cases} X_g^{n+1} = \exp(F_g \Delta t) X_g^n + W_g^\alpha (U_g^{n+1} - U_g^n) + W_g^\beta (Y_g^{n+1} - Y_g^n) \\ Y_g^{n+1} = MGIF \cdot X_g^n + MGIE \cdot U_g^{n+1} - MGID \cdot U_g^n - MGIC \cdot Y_g^n \end{cases} \quad (2.9)$$

avec (en notant I la matrice identité) :

- $MGIF = (I - H_g W_g^\beta - S_g^\beta)^{-1} H_g F_g$
- $MGIE = (I - H_g W_g^\beta - S_g^\beta)^{-1} (H_g W_g^\beta + S_g^\alpha)$
- $MGID = (I - H_g W_g^\beta - S_g^\beta)^{-1} H_g W_g^\alpha$
- $MGIC = (I - H_g W_g^\beta - S_g^\beta)^{-1} H_g W_g^\beta$
- $W_i^j = e^{-\Delta t / \tau_i} B_i^j$

Les phénomènes non linéaires ou impliquant des paramètres variables (ventilation externe, ventilation interne, résistance variable due à l'usage intermittent des occultations, etc.) sont pris en compte après le couplage des modèles zonaux réduits modaux. Par exemple pour la ventilation externe, une valeur moyenne est considérée pour les débits de ventilation externe dans les systèmes linéaires. C'est après le couplage que les écarts par rapport à la moyenne sont pris en compte et que le vecteur des sollicitations U_g est corrigé.

La simulation de ce modèle nécessite de connaître les sollicitations, en particulier les émissions de chaleur par les occupants et les équipements, mais également les données météorologiques locales concernant la température extérieure et le rayonnement solaire. Toutes ces données sont contenues dans le vecteur des sollicitations.

Dans le cadre de cette thèse, nous nous intéressons uniquement à un système de chauffage électrique pour la régulation, directement intégré dans le modèle thermique dynamique du bâtiment. On suppose que le système de chauffage dispose d'un rendement de 100 % et qu'il fournit l'ensemble de la puissance à la maille d'air. Ce type de système permet ainsi d'être très réactif (sa dynamique sera négligée) et de ne pas avoir de décalage entre le moment où la commande est envoyée au système et celui où la puissance de chauffage est injectée dans la maille d'air.

3. MODÉLISATION DE L'OCCUPATION

3.1 Introduction

La modélisation de l'occupation est une notion complexe mais néanmoins essentielle pour la mise en place d'une régulation. En effet, l'impact de l'occupation sur la consommation énergétique finale du bâtiment est non négligeable. Il devient alors indispensable de la modéliser. Selon les manières d'aborder le sujet, deux approches peuvent être distinguées (Vorger 2014).

La première approche considère le comportement des occupants comme des processus stochastiques (J. Fergus Nicol et Humphreys 2004). L'objectif de la modélisation est alors de reproduire, à partir de données d'observations et à travers des modèles statistiques, le comportement des occupants (durée et fréquence des actions) en fonction des conditions environnementales. Plus précisément, il s'agit d'observer la réalité et de la reproduire sans expliciter les relations de cause à effet entre les conditions environnementales et le choix d'une action. Ces modèles permettent ainsi de prendre en compte un ensemble de phénomènes cachés (liés au comportement de l'occupant) qu'il est difficile de modéliser de manière explicite. Ils sont également capables de reproduire la variété des comportements existants. L'utilisation de ces modèles est ainsi pertinente dans le cadre du développement de la Garantie de Performance Énergétique (GPE) qui vise à garantir (par contrat) la performance énergétique d'un bâtiment réhabilité. Ces modèles peuvent ainsi être utilisés par les logiciels de simulation thermique dynamique pour prévoir les consommations énergétiques.

La deuxième approche considère quant à elle que les actions des occupants sont réalisées dans le but d'assurer certaines conditions de confort (Fanger 1986). En effet, les bâtiments sont chauffés ou refroidis dans le but d'assurer aux occupants des conditions de confort acceptables. La définition de ce confort est alors un pré-requis pour la modélisation de l'occupation. On modélise ainsi de manière explicite la logique menant au choix de l'action de l'occupant.

Dans le cadre de la régulation énergétique des bâtiments, les différents systèmes du bâtiment (chauffage, climatisation, ventilation, protections solaires) sont pilotés par le système de régulation et non par les occupants. Le comportement des occupants est pris en compte sous la forme de contraintes à respecter. Dans ce cas, la modélisation de l'occupation se focalise sur le confort, et plus précisément sur le respect des conditions acceptables par l'occupant (deuxième approche).

3.2 État de l'art sur le confort

Comme nous l'avons vu dans le chapitre 1, le confort peut être défini par plusieurs facteurs, tels que le confort thermique, le confort visuel, le confort acoustique et la qualité de l'air intérieur. Dans le cadre de la régulation énergétique des bâtiments en période hivernale, c'est la notion de confort thermique qui va nous intéresser tout particulièrement.

3.2.1 Le confort thermique

L'amélioration du confort de l'occupant, en particulier le confort thermique, est un des objectifs principaux de la régulation dans le bâtiment. La définition de cette notion est alors indispensable pour la prendre en compte dans l'optimisation, soit sous forme de contrainte à respecter, soit en tant qu'objectif même.

Le phénomène de confort thermique est lié à l'interaction du corps humain avec son environnement à travers des échanges de chaleur (sensibles et latents) et à la régulation de sa température par « des processus de régulation thermo-physiologique involontaires qui modulent la production et le transfert de masse et de chaleur, internes et externes » (Thellier et al. 2009). Ainsi, en réaction à des sensations d'inconfort, le corps humain va réagir en s'adaptant à son environnement à travers un ensemble de réactions physiologiques dans le but de retrouver un état de confort. Dans cet état de l'art du confort thermique, nous allons dans un premier temps nous intéresser au bilan thermique du corps humain pour ensuite présenter deux méthodes d'évaluation du confort thermique, à savoir l'approche statique et l'approche adaptative.

3.2.1.1 Bilan thermique chez l'homme

Le corps humain est homéotherme, c'est-à-dire qu'il tend à maintenir une température corporelle constante autour de 37°C. Cette stabilité de la température impose un équilibre entre la production de chaleur à l'intérieur du corps (résultat du métabolisme de base ainsi que du travail physique) et les pertes de chaleur par échange vers le milieu extérieur. Le métabolisme correspond à l'ensemble des transformations de la nourriture consommée en constituants du corps (protéines, etc.) d'une part et en énergie d'autre part. L'énergie ainsi produite est alors transformée en travail et en chaleur. Ces échanges de chaleur entre l'individu et le milieu extérieur se font selon divers mécanismes et dépendent des activités et de l'habillement (Figure 5).

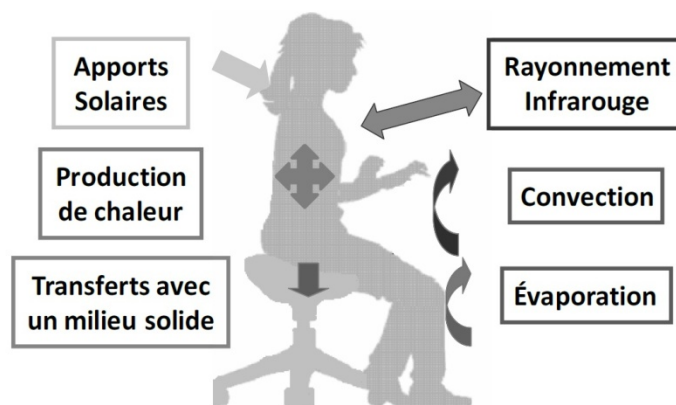


Figure 5 : Echanges entre le corps humain et le milieu extérieur (Thellier et al. 2012)

Ce bilan thermique (en régime transitoire) par unité de surface corporelle peut alors s'écrire sous la forme (Fanger 1970) :

$$M - W = C_{res} + E_{res} + K + C + R + E \quad (2.10)$$

Chapitre 2 : Modèles nécessaires à la régulation

avec

- $M - W$, production de chaleur interne du corps (en W/m^2), avec M le métabolisme et W le travail physique,
- C_{res} , échange de chaleur au niveau des voies respiratoires par convection (en W/m^2),
- E_{res} , échange de chaleur au niveau des voies respiratoires par évaporation (en W/m^2),
- K , échange de chaleur au niveau de la peau par conduction (en W/m^2),
- C , échange de chaleur au niveau de la peau par convection (en W/m^2),
- R , échange de chaleur au niveau de la peau par rayonnement (en W/m^2),
- E , échange de chaleur au niveau de la peau par évaporation (en W/m^2).

Lorsque le corps humain est à l'équilibre avec son environnement, on dit qu'il est à neutralité thermique. Cette situation idéale correspond à des réactions thermo-physiologiques minimales avec un bilan thermique équilibré. Cependant, si les réactions thermo-physiologiques ne parviennent pas à maintenir l'équilibre thermique du corps, le corps est alors dans une situation d'hyper ou d'hypothermie (température corporelle supérieure à $38^{\circ}C$ ou inférieure à $35^{\circ}C$ respectivement). Ainsi, si un individu a chaud, des mécanismes de vaso-dilatation et de transpiration seront utilisés alors que dans le cas contraire des mécanismes de frissons et de vaso-constriction seront enclenchés.

La notion de confort thermique correspond à une appréciation subjective. Par exemple, la norme ASHRAE (ASHRAE 2003) définit le confort thermique comme « la condition d'esprit qui exprime la satisfaction en présence d'un environnement thermique donné ». Ainsi, la sensation de confort ne sera pas la même pour différentes personnes au sein d'une même pièce, rendant la définition du confort thermique, au delà des aspects purement physiques, très difficile.

Il est important de noter que la température de zone calculée par COMFIE est proche d'une température opérative, correspondant à une synthèse des échanges convectifs (température de l'air) et des échanges radiatifs (température des parois). Cette température opérative correspond, de manière simplifiée, à la température de confort ressentie par les occupants.

3.2.1.2 L'approche statique

Une des approches le plus couramment utilisée pour décrire le confort thermique est l'approche développée par Fanger (1967) avec les indices PMV (*Predicted Mean Vote*) et PPD (*Predicted Percentage Dissatisfied*).

Dans cette approche, Fanger considère que la situation de confort thermique dépend de trois conditions :

- un bilan thermique équilibré ;
- une température moyenne de la peau située dans les limites du confort ;
- une sudation située dans les limites du confort.

Chapitre 2 : Modèles nécessaires à la régulation

Dans ces conditions, l'équilibre thermique est caractérisé par 6 paramètres :

- le métabolisme (en W/m^2) ;
- le clo, correspondant à l'unité d'isolement vestimentaire ($1 \text{ clo} = 0,155^\circ\text{C}\cdot\text{m}^2/\text{W}$) ;
- la température de l'air (en $^\circ\text{C}$) ;
- la température des parois (en $^\circ\text{C}$) ;
- l'humidité relative de l'air (en %) ;
- la vitesse relative de l'air (en m/s).

A partir de ces paramètres et en écrivant le bilan thermique, il est possible de calculer l'indice PMV qui décrit la sensation thermique en fonction de l'écart entre le flux de chaleur cédé par le corps à l'environnement et celui qui correspond aux conditions de confort. L'indice PMV permet de prédire l'avis moyen d'un groupe important de personnes qui exprimeraient un vote de sensation de confort thermique. Ces votes reposent sur une échelle (Tableau 1) allant de -3 à +3 où un indice PMV de 0 exprime une sensation de confort thermique idéale. Un indice PMV négatif signifie que la température est plus basse que la température idéale et inversement, un indice positif indique une température trop élevée.

Tableau 1 : Echelle des indices PMV

+3	Chaud
+2	Tiède
+1	Légèrement tiède
0	Neutre
-1	Légèrement frais
-2	Frais
-3	Froid

La sensation thermique n'étant pas suffisante pour exprimer le confort, Fanger a alors proposé un autre indice pour compléter le PMV : il s'agit du PPD (pourcentage prévu des insatisfaits). Cet indice permet de prévoir le pourcentage de personnes insatisfaites à une sensation de confort thermique donnée en fonction du PMV (Figure 6).

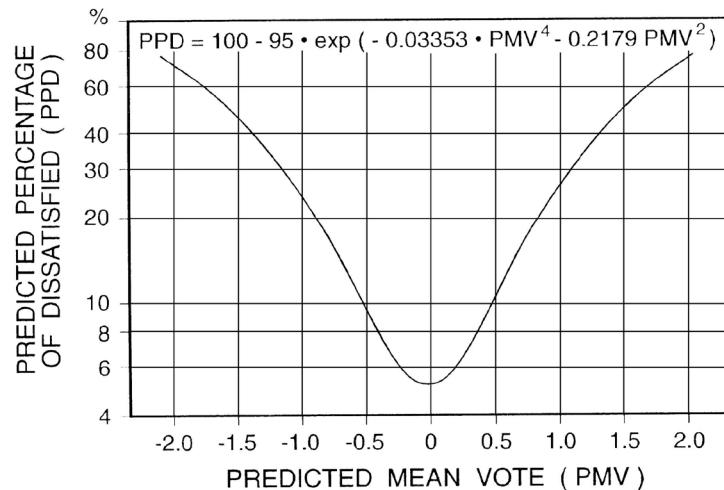


Figure 6 : Relation entre le PMV et le PPD (ASHRAE 2004)

On peut remarquer que le PPD n'atteint jamais 0 %, illustrant le fait qu'aucune condition n'est jugée optimale pour l'ensemble de la population.

Cependant, la mise en place de l'approche statique implique l'utilisation de nombreux capteurs pour mesurer les différents paramètres, entraînant une augmentation du coût ainsi que de la logistique nécessaire à la régulation. De plus, cette approche ne prend pas en compte les possibilités d'actions sur l'environnement dont dispose l'occupant. L'approche adaptative, reposant sur une hypothèse d'adaptation des occupants à leur environnement, permet de prendre en compte ce phénomène.

3.2.1.3 L'approche adaptative

L'approche adaptative se différencie de l'approche statique en intégrant dans sa réflexion la capacité d'adaptation climatique et la capacité d'interaction de l'occupant avec son environnement. Cette adaptation des occupants peut se décliner en trois domaines :

- **Comportemental** : correspond aux réactions de l'occupant à des situations ressenties, qui va d'une part s'adapter à son environnement (cela inclut par exemple la modification de l'habillement, le changement de posture, le déplacement ou encore l'absorption de boissons chaudes ou froides) et d'autre part modifier son environnement (on peut citer par exemple la modification de la consigne de chauffage ou de climatisation, l'ouverture ou la fermeture des fenêtres, ou encore l'action sur des occultations).
- **Physiologique** : sur une échelle de quelques jours ou semaines, les individus vont ajuster de manière inconsciente les réactions et les processus de régulation du corps.

Chapitre 2 : Modèles nécessaires à la régulation

- **Psychologique** : selon le degré de liberté dont dispose les occupants, les zones de confort thermique et les exigences de confort des occupants sont plus ou moins grandes. Ainsi, si les occupants disposent, ou pensent disposer, d'une capacité d'action sur la gestion du bâtiment, leur zone de confort thermique sera plus grande que s'ils n'avaient aucune possibilité d'action. L'anticipation des conditions climatiques (permettant d'adapter les vêtements) les rend également plus tolérants.

Ainsi, plusieurs études sur le terrain se sont intéressées à la relation existante entre la température de confort T_{conf} et la température extérieure. Cette relation, obtenue par régression linéaire, est de la forme suivante (de Dear et Brager 2002 ; J.F. Nicol et Humphreys 2002) :

$$T_{conf} = a T_{ext} + b \quad (2.11)$$

avec T_{ext} la moyenne mensuelle ou la moyenne glissante de la température extérieure.

Ces modèles adaptatifs ont été intégrés dans des standards tels que le standard ASHRAE 2004 ou encore le standard CEN avec la norme EN15251 (Nicol et Wilson 2010).

Les différentes approches du confort thermique (statique ou adaptative) présentent cependant quelques limites. En effet, aucun modèle ne s'impose par sa robustesse ou sa qualité de prévision (Vorger 2014). Par exemple, il existe des différences entre les standards ASHRAE et CEN. De plus, les modèles rationnels et adaptatifs reposent principalement sur des données de bâtiments de bureaux et non des logements, limitant leur validité. Enfin, la difficulté de mise en place de ces modèles (en particulier pour le modèle statique) peut être un frein à leur utilisation.

3.2.2 Modélisation

La prise en compte de la variabilité des comportements des occupants (avec par exemple la probabilité d'ouverture d'une fenêtre ou de l'activation de l'éclairage) peut être réalisée à partir de modèles statistiques (première approche). Cependant, cette approche impliquerait l'utilisation de méthodes de commande prédictive stochastique, méthodes qui sont hors du cadre de cette thèse.

Étant donné que le système de régulation a pour objectif de déterminer une séquence de commandes optimales de la puissance, le système de chauffage n'est plus contrôlé directement par l'occupant. La prise en compte de l'occupation est alors faite de manière indirecte à travers le respect des contraintes de confort, et en particulier la modélisation du confort thermique (deuxième approche). Cependant, un rapide état de l'art a permis de révéler les limites des différentes modélisations du confort thermique (statique ou adaptative). Les incertitudes associées nous ont conduits à adopter une approche déterministe (et plus précisément invariante) qui consiste à modéliser l'occupation à partir de scénarios définis par l'utilisateur. Il s'agit de l'approche la plus répandue dans les logiciels de simulation thermique dynamique.

Chapitre 2 : Modèles nécessaires à la régulation

Les scénarios permettent ainsi de renseigner les données suivantes :

- la présence (avec le nombre d'occupants et leur chaleur métabolique en W) ;
- les consignes de chauffage et de climatisation (en °C) ;
- les apports internes (liés à la puissance dissipée par les appareils électriques en W) ;
- les taux d'occultation des vitrages.

Ces scénarios peuvent être affinés à partir de données collectées sur le site, dans l'objectif d'être les plus réalistes possibles. Nous ne considérerons que le confort thermique. Notre cas d'application étant l'effacement de la consommation électrique de pointe en hiver, nous modéliserons le confort thermique par une zone de confort définie par une température minimale de 19°C et une température maximale de 24°C permettant une surchauffe du bâtiment. Les autres aspects du confort, tels que la qualité de l'air intérieur, ou la luminosité ne sont pas pris en compte dans le cadre de cette thèse car ils complexifieraient trop la contrainte de confort. Pour le lecteur intéressé, Vorger (2014) a réalisé un état de l'art approfondi sur la modélisation de l'occupant.

4. MODÉLISATION DU CLIMAT

4.1 Introduction

Dans le cadre du développement de stratégies de régulation, le climat est pris en compte, dans un premier temps, de manière simplifiée à partir de données horaires. On se place alors dans le cas idéal où la météorologie est parfaitement connue à l'avance. Ces données météorologiques peuvent ainsi être issues de données mesurées ou de données types (données TRY par exemple).

Dans un deuxième temps, la robustesse des stratégies de régulation est évaluée dans des cas plus réalistes où les prévisions météorologiques ne sont pas parfaites. Il est alors indispensable de développer des modèles permettant de prévoir l'évolution de la météorologie locale et d'avoir un ordre de grandeur sur les erreurs de prévisions.

L'objectif de la prévision météorologique est de concevoir un dispositif permettant de fournir en temps réel des données météorologiques prévisionnelles pour les heures ou les jours à venir pour un bâtiment donné. En effet, les données fournies par les instituts nationaux de prévisions météorologiques sont pour la plupart régionales, difficiles d'accès et payantes. Dans ce contexte, le développement d'un modèle de prévision météorologique locale est indispensable pour la gestion optimisée du bâtiment.

Dans cette partie, un état de l'art sur la modélisation des conditions climatiques est présenté (§4.2), suivi par une étude approfondie sur les modèles ARMA (§4.3). Enfin, les méthodes choisies pour modéliser les conditions climatiques locales ainsi que leurs résultats sont exposées dans le §4.4.

4.2 État de l'art de la modélisation des conditions climatiques

4.2.1 Variables climatiques à modéliser

Le pré-requis pour développer notre modèle est d'étudier les variables météorologiques nécessaires à la prévision des performances énergétiques d'un bâtiment et d'identifier les plus importantes et pertinentes. On peut citer par exemple le rayonnement solaire, la température extérieure, l'humidité, la vitesse du vent ainsi que la pression atmosphérique. En suivant l'approche de Crawley (1998), parmi les données météorologiques disponibles, seuls le **rayonnement solaire** et la **température de l'air** seront modélisés. En effet, la France, avec un climat tempéré, n'est pas sujette à de forts taux d'humidité et les zones soumises à des vents violents y sont limitées.

Le rayonnement solaire, ou plus précisément le rayonnement solaire global incident sur une surface donnée correspond à la somme des trois composantes : directe, réfléchi et diffus. Il participe au chauffage, à la surchauffe, ainsi qu'à l'éclairage du bâtiment. En France métropolitaine, le rayonnement solaire global horizontal moyen varie dans une plage 1100-1900 kWh/m² sur un an.

La température de l'air extérieur a, quant à elle, une importance non négligeable sur les besoins de chauffage. En effet, elle influe directement sur les déperditions thermiques du bâtiment.

4.2.2 Les modèles de prévisions météorologiques

Il existe deux catégories de modèles de prévisions météorologiques. La première concerne les modèles physiques, utilisés par les services de météorologie ou certains centres de recherche. Ils mettent en jeu les différentes grandeurs caractérisant le climat et leurs interactions, et permettent d'effectuer des prévisions météorologiques fines. La deuxième catégorie regroupe les modèles purement mathématiques reposant uniquement sur l'étude statistique des données passées dans le but de caractériser le comportement du système.

4.2.2.1 Les modèles physiques

Les modèles physiques, ou plus précisément les modèles de prévision numérique de la météorologie, servent à produire une représentation la plus exacte possible de l'état présent et futur de l'atmosphère. Leur principale utilisation est la prévision météorologique pour les prochains jours. Ils sont donc utilisés principalement par les services de météorologie nationale. Ces modèles sont basés sur une division de l'atmosphère en de nombreuses mailles et utilisent les équations primitives atmosphériques, qui sont une version simplifiée des équations de Navier-Stokes, qui permettent de simuler le comportement futur de l'atmosphère à partir de nombreux points de mesure (stations terrestres, navires, aéronefs, radiosondes, données satellitaires de mesures à distance, etc.).

Chapitre 2 : Modèles nécessaires à la régulation

De nombreux modèles physiques ont été développés. On peut tout d'abord citer les modèles de Météo-France (Météo-France 2015) avec le modèle ARPEGE (système de prévision numérique couvrant le globe), le modèle ALADIN (système couvrant des régions de quelques milliers de km de taille) et le modèle AROME (système utilisé opérationnellement sur la France métropolitaine). Les tailles de mailles pour ces modèles sont respectivement de l'ordre de 15 km, 10 km et 2,5 km. Diagne et al. (2013) présentent un état de l'art complet des modèles physiques existants avec l'estimation de leurs erreurs.

L'avantage de ces modèles physiques est la précision des prévisions météorologiques réalisées sur plusieurs jours par rapport aux modèles mathématiques. Cependant, ces modèles très détaillés nécessitent l'utilisation de supercalculateurs qui ne sont accessibles qu'à des professionnels de la météorologie. De plus, les prévisions sont généralement faites à l'échelle régionale et non locale, ce qui induit une perte de précision et la non-prise en compte des spécificités locales telles que les microclimats, les masques, les montagnes, etc. Enfin, l'accès aux données est généralement payant et nécessite une connexion fiable.

A l'échelle de la ville, des modèles physiques existent également permettant de reproduire les échanges énergétiques et hydriques entre le milieu urbain et l'atmosphère (Musy et Bozonnet 2013). De nombreux modèles ont ainsi été développés pour modéliser les microclimats à différents niveaux d'échelle (ville, quartier, bâtiment) et évaluer l'interaction entre microclimat urbain et demande énergétique des bâtiments. Cependant, la modélisation des microclimats est complexe du fait de l'interdépendance des différents flux existants : ensoleillement, vents dominants, convection naturelle, polluants et production anthropique (Bozonnet et al. 2011). De plus, des problèmes se posent quant aux temps de calcul, à la fiabilité des modèles et à la disponibilité des données d'entrée.

4.2.2.2 Les modèles mathématiques

Les modèles purement mathématiques reposent sur l'analyse des relevés des données météorologiques et donc sur l'utilisation de séries temporelles (ou chronologiques). Ces modèles sont moins fiables et moins précis mais ont l'avantage d'être plus simples et ne nécessitent pas des moyens informatiques lourds. On peut donc imaginer leur utilisation sur des prévisions locales.

Il se dégage de cette catégorie quatre grandes familles de modèles (Voyant 2011) : les modèles de type « naïf », les modèles à probabilités conditionnelles, les réseaux de neurones et les modèles ARMA. Avant une présentation succincte de ces modèles, une définition des séries temporelles est nécessaire.

4.2.2.2.1 Définition et caractérisation des séries temporelles

Une série temporelle (ou série chronologique) à temps discret est une suite finie d'observations d'une famille de variables aléatoires réelles $(X_t)_{1 \leq t \leq n}$, où t représente le temps (en minute, jour, année...). Dans le cas des séries temporelles météorologiques, ces dernières sont constituées de valeurs instantanées séparées par un intervalle de temps fixe. Un des objectifs principaux de l'étude des séries temporelles est de décrire, expliquer et prévoir l'évolution au cours du temps d'un phénomène.

Chapitre 2 : Modèles nécessaires à la régulation

Une série temporelle est la résultante de différentes composantes fondamentales :

- La tendance (Z_t), qui représente l'évolution à long terme de la série étudiée. La tendance peut être linéaire, polynomiale, logarithmique, exponentielle, etc.
- La composante saisonnière ou périodique (S_t), qui correspond à un phénomène qui se répète à intervalles de temps réguliers.
- La composante résiduelle (ϵ_t), qui correspond à des fluctuations irrégulières, en général de faible intensité mais de nature aléatoire.

Il existe différents modèles de représentation générale d'une série temporelle. Plus précisément, les tendances et les facteurs saisonniers rencontrés le plus fréquemment peuvent être présentés par trois types de modèles :

- Le modèle additif :

$$X_t = Z_t + S_t + \epsilon_t \quad (2.12)$$

- Le modèle multiplicatif :

$$X_t = Z_t(1 + S_t)(1 + \epsilon_t) \quad (2.13)$$

- Les modèles hybrides, correspondant à différentes combinaisons de modèles additifs et de modèles multiplicatifs :

$$X_t = (Z_t + S_t)(1 + \epsilon_t) \quad (2.14)$$

4.2.2.2 Les modèles de type « naïf »

On peut citer dans les modèles de type « naïf », le lissage exponentiel ou encore les k plus proches voisins.

➤ Le lissage exponentiel

Les méthodes de lissage constituent l'ensemble des techniques empiriques de prévision qui accordent plus ou moins d'importance aux valeurs du passé d'une série temporelle. Ces méthodes reposent sur le principe d'une mise à jour facile de la prédiction de la série lorsqu'une nouvelle observation est disponible.

Soit une série temporelle (X_t) dont on dispose des N premières observations X_1, \dots, X_N , l'objectif est de faire une prévision de la série à l'horizon h , c'est-à-dire de prévoir la valeur \hat{X}_{N+h} ou $\hat{X}_N(h)$.

Le lissage exponentiel simple

Le lissage exponentiel simple permet d'effectuer des prévisions pour des séries temporelles sans saisonnalité et dont la tendance est constante. Il s'agit d'une technique permettant de faire des prévisions à $\hat{X}_N(1)$.

Chapitre 2 : Modèles nécessaires à la régulation

Étant donnée la constante de lissage β tel que $0 < \beta < 1$, la prévision de la série est :

$$\hat{X}_N(1) = (1 - \beta) \sum_{j=0}^{N-1} \beta^j X_{N-j} \quad (2.15)$$

Si β est proche de 0, la prévision est fortement influencée par les observations récentes. Au contraire, si β est proche de 1, l'influence des observations éloignées est d'autant plus importante et remonte loin dans le passé.

Le lissage exponentiel double

Le lissage exponentiel double permet de généraliser le lissage exponentiel simple au cas où la série peut être ajustée à une droite au voisinage de l'instant N . La prévision de la série à l'horizon h , $\hat{X}_N(h)$, est donnée par :

$$\hat{X}_N(h) = \hat{a}_N h + \hat{b}_N \quad (2.16)$$

où le couple (\hat{a}_N, \hat{b}_N) minimise la fonction :

$$\sum_{j=0}^{N-1} \beta^j (X_{N-j} - (aj + b))^2 \quad (2.17)$$

Les méthodes de Holt-Winters

Il s'agit des méthodes à privilégier parmi les techniques de lissage exponentiel dans le cas de séries d'observations présentant à la fois une tendance et une saisonnalité. On peut citer la méthode non saisonnière, qui s'applique à des séries sans saisonnalité, la méthode additive adaptée aux séries admettant une décomposition additive, et la méthode multiplicative adaptée aux séries admettant une décomposition multiplicative.

La différence entre les méthodes de Holt-Winters et le lissage exponentiel double porte principalement sur les formules de mise à jour.

➤ K plus proches voisins

La méthode des K plus proches voisins peut également être utilisée pour la prédiction de séries temporelles. Cette méthodologie a pour but de trouver dans les données d'apprentissage, c'est-à-dire l'historique, un groupe de données (formé de K données) se rapprochant le plus du groupe de données (formé également de K données) dont on veut prédire les valeurs futures (Sharif et Burn 2006). De nombreux algorithmes existent pour réaliser cet apprentissage supervisé.

4.2.2.2.3 Les chaînes à probabilités conditionnelles : chaîne de Markov

Certains auteurs ont eu recours aux chaînes de Markov pour la prévision de séries temporelles (Bauer 1998 ; Muselli et al. 2001). Une chaîne de Markov est un processus stochastique qui possède la propriété Markovienne qui signifie que, étant donné l'état présent, les états futurs sont indépendants des états passés. Les états futurs sont donc modélisables grâce à un processus probabiliste dépendant

Chapitre 2 : Modèles nécessaires à la régulation

uniquement des états présents et d'une matrice de transition (ou matrice de Markov). Pour initialiser le processus, il est nécessaire d'identifier la matrice de transition du phénomène en utilisant l'historique.

Soit $p_{i,j}$ la probabilité de transition de l'état i à l'état j :

$$p_{i,j} = P(X_t = j | X_{t-1} = i) \quad (2.18)$$

La matrice de transition R_M est constituée de la famille de ces probabilités. La prédiction à l'horizon h est donc donnée par :

$$\hat{X}_N(h) = X_N * (R_M)^h \quad (2.19)$$

4.2.2.2.4 Les réseaux de neurones

Une alternative aux approches conventionnelles est l'utilisation des réseaux de neurones artificiels. Il s'agit de modèles à intelligence artificielle dont le fonctionnement est analogue à celui du cerveau humain. Ils permettent de résoudre des problèmes très complexes tels que la classification ou la prédiction de séries temporelles. Les réseaux de neurones sont une solution appropriée lorsque l'on dispose de nombreuses données et que les règles régissant les phénomènes que l'on souhaite modéliser sont inconnues. De nombreuses études basées sur différentes techniques d'intelligence artificielle ont été réalisées. On peut citer par exemple Sfetsos et Coonick (2000), Mellit et al (2006) ou encore Tasadduq et al (2002).

Les réseaux de neurones artificiels ont un fonctionnement assez simple. Ils mettent en œuvre le principe de l'induction, c'est-à-dire la capacité à apprendre par l'expérience. Ainsi, par confrontation avec des données connues lors de la phase d'apprentissage, un système de décision est généré (un modèle sous forme de boîte noire) dont l'efficacité et le caractère générique sont fonction du nombre de cas d'apprentissage et de leur complexité par rapport au problème à modéliser et à résoudre.

4.2.2.2.5 Les modèles ARMA

Les modèles ARMA sont des modèles couramment utilisés pour représenter les séries temporelles. Ils associent deux types de processus : un processus autorégressif (AR) et un processus en moyenne mobile (MA). Ces processus sont construits à partir de l'idée que l'observation au temps t s'explique linéairement par les observations précédentes et par les observations d'un bruit blanc. Par exemple, Matyasovszky (2001) illustre l'utilisation des modèles ARMA pour la prévision météorologique de séries temporelles alors que Guerrier et al (1980) appliquent ces modèles pour la prévision du rayonnement solaire.

4.2.2.2.6 Choix de la méthode de modélisation

Cette section a permis de présenter les principaux modèles mathématiques pour modéliser une série temporelle. D'après la littérature, les réseaux de neurones et les modèles ARMA sont les méthodes classiquement utilisées pour modéliser les conditions climatiques (Voyant 2011). Par exemple, Favre (2013) a utilisé les réseaux de neurones pour prévoir la température et le rayonnement solaire (en

considérant une période d'apprentissage de cinq ans de données climatiques). Les résultats obtenus ne furent pas d'une grande précision (par exemple pour la prévision de la température à un horizon de 24 h, l'erreur moyenne absolue était de 1,6°C). D'autre part, différentes études semblent indiquer que les modèles ARMA peuvent être plus efficaces (Reikard 2009 ; Diagne et al. 2013). Par conséquent, les modèles ARMA ont été retenus dans la suite de l'étude pour tester la mise en œuvre d'un modèle de prévision du climat local. L'approche est appliquée sur l'Île-de-France, où le phénomène d'îlot de chaleur est important.

4.3 Caractérisation des modèles ARMA

Les processus ARMA s'appliquent à des séries stationnaires, c'est-à-dire des séries dont l'espérance est constante au cours du temps, la variance est constante au cours du temps et non infinie, et dont la structure de covariance reste aussi constante.

Ainsi pour modéliser la série temporelle à l'aide d'un processus ARMA, la démarche suivante doit être adoptée :

- Représenter graphiquement la série afin de repérer les tendances et saisonnalité.
- Estimer et supprimer les tendances et saisonnalités (partie déterministe du processus stochastique).
- Choisir un modèle pour les résidus (partie aléatoire du processus stochastique) et l'estimer.
- Prédire les réalisations futures à l'aide de ce modèle.

4.3.1 *Élimination de la tendance et de la saisonnalité*

4.3.1.1 Série temporelle avec tendance et sans périodicité

Soit une série temporelle de la forme :

$$X_t = Z_t + \epsilon_t \quad (2.20)$$

Une première méthode est l'estimation paramétrique de la tendance. On suppose que la tendance de la série temporelle est de forme polynomiale. Le but est de présumer le degré du polynôme à partir de sa représentation graphique et d'estimer les paramètres du polynôme par méthode des moindres carrés. Cependant, cette méthode peut devenir rapidement fastidieuse. On opte ainsi pour une estimation non paramétrique : la moyenne mobile. Le principe est de supposer la tendance linéaire sur un petit intervalle et de l'estimer avec la moyenne sur cet intervalle. Cette moyenne est dite mobile car elle est recalculée à chaque pas de temps Δt .

4.3.1.2 Série temporelle avec tendance et périodicité

Une autre méthode pour éliminer la tendance et la saisonnalité sans les estimer est la méthode des différences. Pour cela, on utilise les opérateurs de Box-Jenkins (Box et Jenkins 1976) :

- l'opérateur de recul **B** :

$$BX_t = X_{t-1} \quad (2.21)$$

- l'opérateur de différenciation **∇** qui permet d'éliminer la tendance de la série :

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t \quad (2.22)$$

Ainsi, si la série X_t admet une tendance polynomiale d'ordre k , alors ∇X_t admet une tendance d'ordre $k-1$. En appliquant k fois ∇ on élimine donc la tendance.

- l'opérateur de désaisonnalisation **∇_s** qui permet d'éliminer la saisonnalité de période s :

$$\nabla_s X_t = X_t - X_{t-s} = (1 - B^s)X_t \quad (2.23)$$

Par exemple, si un processus admet une tendance polynomiale d'ordre k et une saisonnalité d'ordre s , alors $\nabla_s^k X_t = (1 - B)^k (1 - B^s) X_t$ est un processus désaisonné et sans tendance.

On peut évaluer la stationnarité d'une série temporelle, et donc déterminer si cette dernière nécessite une différenciation, à l'aide des tests de racine unitaire. Il existe par exemple le test ADF (*Augmented Dickey-Fuller*) (Dickey et Fuller 1981), réalisé sous l'hypothèse de la présence d'une racine unitaire dans la série temporelle, appelée hypothèse nulle. Si le résultat du test est au-dessus d'une certaine valeur critique (par exemple 0,05), alors l'hypothèse nulle est acceptée et la série est non stationnaire. Sinon, l'hypothèse nulle est rejetée et la série est considérée comme stationnaire.

On peut également citer le test de Kwiatkowski-Phillips-Schmidt-Shin (KPSS) (Kwiatkowski et al. 1992) qui est un des tests de stationnarité le plus couramment utilisé. Contrairement au test ADF, ce test est réalisé sous l'hypothèse nulle de stationnarité de la série étudiée.

4.3.1.3 Vérifications sur le résidu

L'objectif des méthodes présentées ci-dessus est d'obtenir une série résiduelle stationnaire (ou le plus stationnaire possible), et en particulier sans tendance et sans saisonnalité. Il est ensuite nécessaire de tester s'il y a dépendance entre les différents termes de la série obtenue. Si ce n'est pas le cas, on est en présence d'un bruit blanc et il est inutile d'aller plus loin si ce n'est d'estimer la moyenne et la variance du bruit blanc. Dans le cas contraire, on peut modéliser la série résiduelle par un processus ARMA. Il existe différents tests de bruits blancs, tels que le test du porte-manteau ou le test de Ljung et Box. Ces tests seront présentés plus précisément par la suite.

4.3.2 Les processus ARMA

4.3.2.1 Les fonctions d'autocovariance et d'autocorrélation

L'autocovariance d'ordre (ou de décalage) h est définie par :

$$\gamma(h) = Cov(X_t, X_{t-h}) \quad (2.24)$$

La fonction de covariance (ou d'autocovariance) pour un processus stationnaire X_t est donnée par :

$$h \rightarrow \gamma(h), \quad h = 0, 1, 2, \dots \quad (2.25)$$

La fonction d'autocovariance peut être « normalisée » et la nouvelle fonction obtenue est la fonction d'autocorrélation (ou ACF pour *AutoCorrelation Function*) :

$$h \rightarrow \rho(h) = \frac{\gamma(h)}{\gamma(0)}, \quad h = 0, 1, 2, \dots \quad (2.26)$$

Le graphe de cette fonction est appelé corrélogramme. Ces fonctions caractérisent la dépendance entre les variables de X_t .

Cependant, la fonction d'autocorrélation n'est pas suffisante pour expliquer comment présent et passé sont reliés car les variables du passé sont elles-mêmes reliées entre elles. Autrement dit, X_{t-h} dépend en général des variables $X_{t-h+1}, \dots, X_{t-1}$. On a donc recours à une nouvelle fonction, la **fonction d'autocorrélation partielle (PACF)** qui exprime le lien entre X_t et X_{t-h} lorsqu'on a retiré leur meilleure explication affine en termes de $X_{t-h+1}, \dots, X_{t-1}$. L'algorithme de Durbin-Watson permet d'estimer les autocorrélations partielles d'un processus stationnaire.

4.3.2.2 Les processus autorégressif AR(p)

Ces processus sont construits à partir de l'idée que l'observation au temps t s'explique linéairement par les observations précédentes. On appelle processus autorégressif d'ordre p , tout processus (X_t) stationnaire tel que :

$$X_t - \varphi_1 X_{t-1} - \dots - \varphi_p X_{t-p} = \varepsilon_t \quad (2.27)$$

où $\varphi_1, \varphi_2, \dots, \varphi_p$ sont des réels fixés et (ε_t) est un bruit blanc centré de variance σ^2 .

On pose le polynôme caractéristique $\varphi(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p$. On peut donc écrire :

$$\varphi(B)X_t = \varepsilon_t \quad (2.28)$$

➤ Propriétés d'un processus AR(p)

L'autocovariance et l'autocorrélation d'un processus AR(p) décroît exponentiellement avec h tandis que l'autocorrélation partielle est nulle à tout ordre strictement supérieur à p . Par exemple, dans la Figure 7 est illustré **le processus AR(2)** : $X_t + 0.5X_{t-1} + 0.9X_{t-2} = \varepsilon_t$ (la figure du haut correspond à la simulation du processus AR(2)). On peut constater que d'une part l'autocorrélation du processus modélisé (ACF) décroît exponentiellement, et que d'autre part l'autocorrélation partielle (PACF) du

Chapitre 2 : Modèles nécessaires à la régulation

processus est nulle (aux incertitudes près) pour tout ordre strictement supérieur à 2. Par conséquent, le processus modélisé est un processus autorégressif d'ordre 2 (AR(2)).

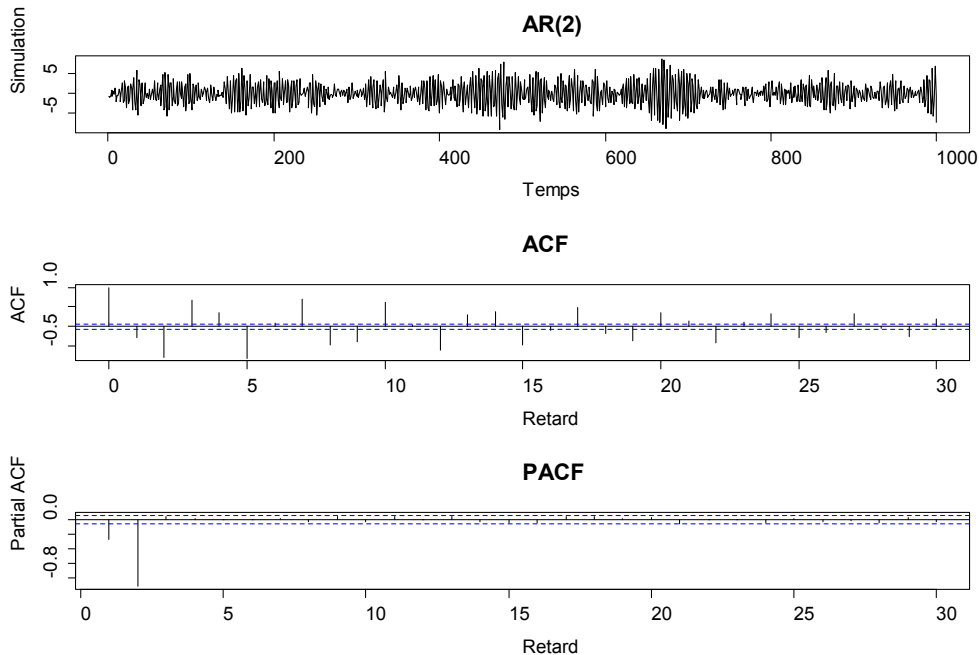


Figure 7: Graphe de trajectoire, corrélogramme et corrélogramme partiel du processus AR(2)

L'abscisse du corrélogramme et du corrélogramme partiel (retard) correspond à l'ordre (ou au décalage) des fonctions d'autocorrélation et d'autocorrélation partielle.

4.3.2.3 Les processus à moyenne mobile MA(q)

Ces processus sont construits à partir de l'idée que l'observation au temps t s'explique linéairement par les observations d'un bruit blanc.

On appelle processus à moyenne mobile d'ordre q tout processus (X_t) stationnaire tel que :

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (2.29)$$

où $\theta_1, \theta_2, \dots, \theta_q$ sont des réels fixés et ε_t est un bruit blanc de variance σ^2 .

On pose le polynôme caractéristique $\theta(B) = I - \theta_1 B - \dots - \theta_q B^q$. On peut donc écrire :

$$X_t = \theta(B)\varepsilon_t \quad (2.30)$$

➤ Propriétés d'un processus MA(q)

L'autocovariance et l'autocorrélation d'un processus MA(q) est nulle à tout ordre strictement supérieur à q tandis que l'autocorrélation partielle tend vers 0 à vitesse exponentielle lorsque h tend vers l'infini. Par exemple, dans la Figure 8, on peut remarquer que l'autocorrélation (ACF) du processus est nulle à tout ordre supérieur à 1 et que l'autocorrélation partielle (PACF) tend exponentiellement vers 0. Le processus modélisé est donc un processus à moyenne mobile d'ordre 1 (MA(1)).

Exemple d'un processus MA(1) : $X_t = \varepsilon_t - 0.8\varepsilon_{t-1}$

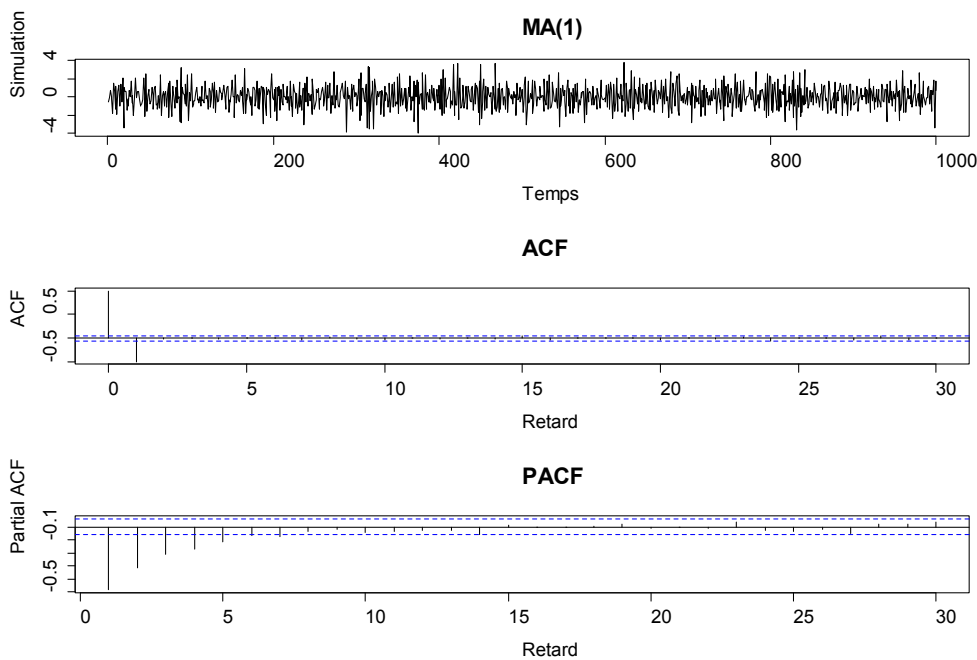


Figure 8: Graphe de trajectoire, corrélogramme et corrélogramme partiel du processus MA(1)

4.3.2.4 Les processus mixtes ARMA(p,q)

On appelle processus autorégressif à moyenne mobile d'ordre (p,q) tout processus (X_t) stationnaire tel que :

$$X_t - \varphi_1 X_{t-1} - \dots - \varphi_p X_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \tag{2.31}$$

Avec les polynômes caractéristiques, on a :

$$\varphi(B)X_t = \theta(B)\varepsilon_t \tag{2.32}$$

Il est immédiat qu'un ARMA(p,0) est un AR pur et qu'un ARMA(0,q) est un MA pur. Les seuls processus admettant simultanément une représentation MA pure et une représentation AR pure correspondent au cas ARMA(0,0), c'est-à-dire aux bruits blancs.

$\varphi(B)$ et $\theta(B)$ forment les deux polynômes caractéristiques du processus. On supposera que ces polynômes n'ont pas de racine commune, afin de s'assurer qu'il n'y a pas de représentation plus courte.

➤ Propriétés d'un processus ARMA(p,q)

L'autocovariance et l'autocorrélation d'un processus ARMA(p,q) vont tendre exponentiellement vers 0 lorsque h tend vers l'infini, à partir de l'ordre q+1. Cependant, à partir de cette seule propriété, il est difficile de déterminer graphiquement l'ordre d'un modèle ARMA.

4.3.2.5 Les processus ARMA saisonnier (modèle SARMA(p,q)(P,Q)_s)

Les modèles SARMA sont des cas particuliers des modèles ARMA. Soit une série temporelle X_t montrant une saisonnalité. Par une modélisation ARMA(p,q), nous obtenons le polynôme caractéristique :

$$\varphi(B)X_t = \theta(B)\varepsilon_t \quad (2.33)$$

Cependant, si la série présente une saisonnalité, le résidu de ε_t ne se comportera pas comme un bruit blanc, mais aura une structure de corrélation saisonnière. Pour traiter cela, on peut modéliser le résidu par un processus ARMA qui aura comme unité de temps la période de la saisonnalité (ses ordres seront alors notés P et Q). On notera le processus : SARMA(p,q)(P,Q). Ainsi, une série temporelle X_t suit un processus SARMA(p,q)(P,Q)_s avec période « s » si l'on peut écrire :

$$\varphi(B)\Gamma(B^s)X_t = \theta(B)\Psi(B^s)\varepsilon_t \quad (2.34)$$

où $\Gamma(B^s) = 1 - \gamma_1 B^s - \dots - \gamma_P B^{Ps}$, et $\Psi(B^s) = 1 - \psi_1 B^s - \dots - \psi_Q B^{Qs}$.

4.3.2.6 Les processus ARIMA(p,d,q) et SARIMA(p,d,q)(P,D,Q)_s

Les modèles ARMA ou SARMA sont utilisés pour modéliser des séries stationnaires. Cependant, en pratique, les séries temporelles étudiées (et en particulier les séries temporelles météorologiques) ne sont pas stationnaires et nécessitent donc un prétraitement pour éliminer les tendances et saisonnalités. Les processus ARIMA et SARIMA sont ainsi la généralisation des modèles ARMA et SARMA à des processus non stationnaires, admettant soit une tendance (ARIMA) soit une tendance et une saisonnalité (SARIMA).

Plus précisément (en prenant les notations définies au §4.3.1.2), une série temporelle X_t peut être modélisée par un processus ARIMA(p,d,q) si la série temporelle $Y_t = (1 - B)^d X_t$ peut être modélisée par un processus ARMA(p,q). Ainsi, les processus ARIMA(p,d,q) sont destinés à modéliser des séries temporelles présentant une tendance polynômiale d'ordre d. On modélise la série sous la forme :

$$\varphi(B)(1 - B)^d X_t = \theta(B)\varepsilon_t \quad (2.35)$$

De même, une série temporelle X_t peut être modélisée par un processus SARIMA(p,d,q)(P,D,Q)_s si la série temporelle $Y_t = (1 - B)^d (1 - B^s)^D X_t$ peut être modélisée par un processus SARMA(p,q)(P,Q)_s. Ainsi, les processus SARIMA(p,d,q)(P,D,Q)_s sont destinés à modéliser des séries temporelles présentant des variations saisonnières (ces variations pouvant elles-mêmes présenter un caractère aléatoire). On modélise ainsi la série sous la forme :

$$\varphi(B)\Gamma(B^s)(1 - B)^d (1 - B^s)^D X_t = \theta(B)\Psi(B^s)\varepsilon_t \quad (2.36)$$

4.3.3 Estimation et choix d'un modèle

A partir de l'observation d'une série temporelle, il est nécessaire de choisir un modèle (ou plusieurs), et estimer ses paramètres. Lorsque l'on hésite entre différents modèles possibles, des critères de choix de modèles sont utilisés pour sélectionner le meilleur d'entre eux.

4.3.3.1 Estimation du modèle

L'estimation des paramètres des modèles ARMA est faite par maximum de vraisemblance (ou MLE pour *Maximum Likelihood Estimation*). Cette technique cherche les valeurs des paramètres qui maximisent la probabilité d'obtenir les valeurs de la série temporelle observée. Cette maximisation est réalisée à l'aide d'algorithmes d'optimisation non linéaires (type Newton).

4.3.3.2 Choix du modèle

Une fois les paramètres des modèles estimés, l'utilisation de critères de choix de modèles permet de choisir le modèle qui effectue le meilleur compromis entre :

- l'ajustement à la série de données ;
- la complexité du modèle.

En effet, si l'on s'intéresse seulement à l'ajustement à la série de données, un modèle ARMA avec un très grand nombre de paramètres serait sélectionné. Cependant, plus il y a de paramètres à estimer et plus le nombre de données nécessaires à l'estimation est important. Or, pour une série temporelle donnée nous avons un nombre d'observations fixé, et donc plus le modèle sera complexe et moins bien les paramètres seront estimés.

Soit x_1, \dots, x_n une série d'observations indépendantes d'un processus aléatoire de loi g inconnue et soit $f(x_1, \dots, x_n | \theta)$ la densité de probabilité spécifiée par le paramètre θ . On appelle L la vraisemblance du paramètre θ :

$$L(\theta) = \prod_{j=1}^n f(x_j | \theta) \quad (2.37)$$

Les trois critères de choix de modèle les plus couramment utilisés sont :

- **le critère AIC** (*Akaike's Information Criterion*), qui sera généralement préféré si l'objectif de l'étude est de faire de la prévision :

$$AIC = -2 \log L(\theta) + 2\omega \quad (2.38)$$

où θ représente les paramètres du modèle et ω le nombre de ces paramètres ;

- **le critère AIC corrigé :**

$$AICc = AIC + \frac{2\omega(\omega + 1)}{n - \omega - 1} \quad (2.39)$$

où n représente le nombre d'observations ;

- **le critère BIC** (*Bayesian Information Criterion*), qui sera généralement préféré si l'objectif de l'étude est de s'ajuster à la série observée :

$$BIC = -2 \log L(\theta) + n\omega \quad (2.40)$$

Chapitre 2 : Modèles nécessaires à la régulation

Ces critères permettent ainsi de sélectionner les modèles ayant la plus grande vraisemblance tout en pénalisant la complexité du modèle. Le modèle sélectionné sera celui ayant la plus petite valeur du critère.

➤ Test du Bruit Blanc

Lorsque le processus est bien estimé, les résidus entre les valeurs observées et les valeurs estimées doivent se comporter comme un bruit blanc. Soit (ε_t) la série des résidus. S'ils obéissent à un bruit blanc, il ne doit pas exister d'autocorrélation dans la série. Différents tests permettent de le vérifier :

- Étude du corrélogramme et du corrélogramme partiel : on vérifie qu'il n'existe pas d'autocorrélation ou d'autocorrélation partielle significativement non nulle pour le processus étudié.
- Les tests du « porte-manteau » (deux tests sont utilisés en pratique) :

a) Test de Box et Pierce

On note $\rho(h)$ l'autocorrélation d'ordre h de la série des résidus (ε_t) . Pour un ordre K , le test de Box et Pierce est le test de l'hypothèse $H_0: \rho(1) = \dots = \rho(K) = 0$ contre l'hypothèse $\exists j \in [1, K]$ tel que $\rho(j) \neq 0$. Pour un processus ARMA(p,q), la statistique de ce test est :

$$Q_{BP} = n \sum_{h=1}^K \rho(h) \xrightarrow{n \rightarrow \infty} \chi^2(K - p - q) \quad (2.41)$$

L'hypothèse H_0 est rejetée au seuil de 5 % si Q_{BP} est supérieur au quantile 0,95 de la loi du χ^2 correspondant.

b) Test de Ljung-Box

Ces statistiques, définies pour un ordre K , correspondent à l'hypothèse nulle : $H_0: \rho(h) = 0 \forall h \leq K$ et sont reconstruites de la façon suivante :

$$Q_K = n(n+2) \sum_{h=1}^K \frac{\rho(h)^2}{n-h} \xrightarrow{n \rightarrow \infty} \chi^2(K - p - q) \quad (2.42)$$

4.4 Modélisation des conditions climatiques

4.4.1 Indicateurs de la qualité de la prévision

Les indicateurs de la qualité de la modélisation et de la prévision permettent d'évaluer la performance du modèle, c'est-à-dire de mesurer l'erreur de prévision. Différents indicateurs peuvent être utilisés pour cela.

Chapitre 2 : Modèles nécessaires à la régulation

Soit une série temporelle (X_t). Soit X_{t+h} la réalisation à $t + h$ de la grandeur à prévoir, et soit $\hat{X}_t(h)$ sa prédiction calculée au temps t pour le même instant $t + h$. Les indicateurs définis ci-dessous caractérisent l'erreur commise sur N mesures.

On définit l'erreur de prévision e à l'instant t de la grandeur à l'instant $t + h$ comme :

$$e(t + h|t) = X_{t+h} - \hat{X}_t(h) \quad (2.43)$$

Il existe différents indicateurs permettant de juger de la qualité de prévision. On peut citer par exemple Diagne et al. (2013) ou Siebert (2008) :

i) Le biais

Le biais ou MBE (*Mean Biased Error*) est défini comme la moyenne de l'erreur sur la période considérée :

$$MBE(h) = \frac{1}{N} \sum_{t=1}^N e(t + h|t) \quad (2.44)$$

Le biais est utilisé pour déterminer si un modèle est plus biaisé qu'un autre.

ii) L'erreur moyenne absolue

L'erreur moyenne absolue ou MAE (*Mean Absolute Error*) est un des indicateurs le plus utilisé et est défini par :

$$MAE(h) = \frac{1}{N} \sum_{t=1}^N |e(t + h|t)| \quad (2.45)$$

Il s'agit de la moyenne des déviations ou erreurs absolues sur toutes les données.

iii) MSE et RMSE

Un autre indicateur couramment utilisé est le RMSE (*Root Mean Square Error*) qui correspond à :

$$RMSE(h) = \sqrt{MSE(h)} \quad (2.46)$$

où on définit MSE (*Mean Squared Error*) comme :

$$MSE(h) = \frac{1}{N} \sum_{t=1}^N e(t + h|t)^2 \quad (2.47)$$

iv) MAPE

On peut également définir le MAPE (*Mean Absolute Percentage Error*) tel que :

$$MAPE(h) = Moyenne(|\frac{e(t + h|t)}{X_{t+h}} * 100|) \quad (2.48)$$

Dans cette étude, les indicateurs MBE et MAE seront utilisés. Le premier a été choisi car il permet de représenter l'erreur globale sur le cumul des prédictions. Ainsi, il permettra d'analyser si un modèle aura tendance à surestimer ou sous-estimer les prédictions par rapport aux mesures. Le deuxième a été retenu car il est moins sensible aux erreurs importantes que l'indicateur RMSE qui

pénalise lourdement les gros écarts. Ainsi, une diminution du MAE signifiera une amélioration en moyenne de la prédiction alors qu'une diminution du RMSE signifiera plutôt une limitation des grosses erreurs de prédiction.

4.4.2 Méthodes de modélisation développées

Dans le cadre de cette étude, deux méthodes de modélisation ont été développées.

4.4.2.1 Modélisation directe

La première méthode correspond à la modélisation « directe » de la température et du rayonnement solaire. Elle consiste à utiliser seulement les données d'une station météorologique (données séparées en données d'apprentissage et données de validation), de déterminer un modèle ARMA sur les données d'apprentissage et d'analyser la qualité de prévision du modèle sur les données de validation.

4.4.2.2 Modélisation des écarts (contextualisation)

La deuxième méthode correspond à la modélisation de l'écart (ou contextualisation) entre deux stations (une station dont on souhaite prévoir l'évolution des conditions météorologiques, et une autre dont on dispose des prévisions météorologiques). Cette méthode se base sur l'idée d'utiliser la prévision de Météo France (ou d'autres fournisseurs) pour prévoir la température ou le rayonnement solaire qu'il y aura localement. En effet, les prévisions fournies par les modèles physiques sont plus précises que celles fournies par les modèles purement mathématiques. Cependant, il s'agit de prévisions météorologiques à l'échelle régionale et non locale. L'idée est alors d'utiliser les modèles mathématiques pour prévoir l'écart entre les données météorologiques fournies par les modèles physiques (données plus précises mais régionales) et les données météorologiques locales (autour du bâtiment à réguler). Cette méthode permet de s'appuyer sur la précision des données issues des modèles physiques tout en gardant le caractère local des prévisions météorologiques.

Par exemple, on utilise un modèle ARMA pour modéliser l'écart de température (ou de rayonnement solaire) entre deux stations : une première station où nous disposons des prévisions de Météo France et une deuxième où l'on souhaite prévoir la température (ou le rayonnement solaire). Notons X_t la série temporelle de la station où nous souhaitons réaliser les prévisions et soit Y_t la série temporelle de la station dont nous disposons des prévisions de Météo France. Enfin, notons ε_t l'écart entre Y_t et X_t , soit : $\varepsilon_t = X_t - Y_t$. Ainsi, avec les notations du §4.4.1, la prévision de X_t calculée au temps $t + h$ vaut $\hat{X}_t(h) = \hat{\varepsilon}_t(h) + Y_{t+h}$, où $\hat{\varepsilon}_t(h)$ correspond à la prévision de l'écart au temps $t + h$ et où Y_{t+h} correspond à la prévision à $t + h$ donnée par Météo France. Le but de la deuxième méthode est de modéliser ε_t à l'aide d'un modèle ARMA.

Cette méthode de modélisation des écarts (ou contextualisation) peut permettre de prendre en compte le phénomène d'îlot de chaleur urbain. Ce dernier correspond à l'écart de température observé entre une agglomération et son environnement périphérique moins urbanisé. Essentiellement nocturne,

Chapitre 2 : Modèles nécessaires à la régulation

il est favorisé par trois facteurs prépondérants (Météo-France et Agence Parisienne du Climat, 2013) : le mode d'occupation des sols, les propriétés radiatives et thermiques des matériaux (par exemple leur albédo) et la morphologie de la ville. Par exemple, l'îlot de chaleur urbain centré sur Paris se traduit par des différences nocturnes de température de l'ordre de 2°C à 3°C en moyenne annuelle entre Paris et les zones rurales alentour.

4.4.3 Données utilisées

Cette étude a pour objectif d'évaluer les performances des modélisations directe et indirecte présentées précédemment. L'exemple d'application concerne la prévision météorologique d'une station de Paris. Dans le cadre de la modélisation des écarts (contextualisation), la station régionale considérée est celle d'Orly.

4.4.3.1 Origine des données

Les données météorologiques utilisées pour la modélisation de la température sont issues du site <http://www.meteociel.fr> fournissant les données horaires de la température des stations météorologiques de Paris (Saint-Germain-des-Prés) et d'Orly pour l'année 2013. Certaines données – moins de 0,3 % d'entre elles – étaient manquantes. Celles-ci ont été calculées comme étant la moyenne des températures des heures précédentes et suivantes. Les données recueillies correspondent aux températures horaires selon l'heure de Paris (*Central European Summer Time*, correspondant au temps universel UTC+1h en hiver, UTC+2h en été). Étant donné que les données recueillies pour le rayonnement solaire sont exprimées en heure universelle, celles de la température ont été également exprimées en heure universelle.

Les données météorologiques utilisées pour la modélisation du rayonnement solaire sont issues de la base de données HelioClim-3 de SoDa (Solar Radiation Data) pour les stations de Paris et d'Orly. Ces données sont issues d'observations satellite.

Perez et al. (1994) ont montré que les valeurs issues de satellites marquent une différence de l'ordre de 20 % par rapport aux mesures. De plus, les données manquantes concernant la température diminuent également la qualité des données et donc la fiabilité du modèle météorologique. Ainsi, l'utilisation de données (température de l'air extérieur et rayonnement solaire) issues directement de mesures sur le bâtiment permettrait d'améliorer la fiabilité des modèles développés.

4.4.3.2 Analyse statistique des données

Le phénomène d'îlot de chaleur urbain entre Paris et Orly peut être mis en évidence par une analyse statistique de la répartition sur l'année de l'écart de température entre ces deux stations. Pour cela, l'outil *box plot* est utilisé pour figurer le profil statistique de la série. Il permet de représenter une distribution de valeurs sous forme simplifiée avec :

- la médiane (trait épais) ;
- une boîte, s'étendant du premier quartile au troisième quartile ;
- des moustaches, qui s'étendent jusqu'à la valeur distante d'au maximum 1,5 fois la distance interquartile ;
- des ronds, qui correspondent à des valeurs supérieures ou inférieures aux moustaches.

La Figure 9 présente le *box plot* de la série des écarts de température nocturne (de 22h à 6h) entre Paris et Orly. On peut remarquer que sur l'ensemble de l'année, la médiane est en moyenne légèrement inférieure à 2°C. On peut également constater que la boîte est plus étendue sur le mois d'août. Ceci peut s'expliquer par le fait que les conditions anticycloniques de la période estivale sont plus favorables à la survenue d'îlots de chaleur urbain que les perturbations océaniques de l'hiver (Météo-France et Agence Parisienne du Climat, 2013).

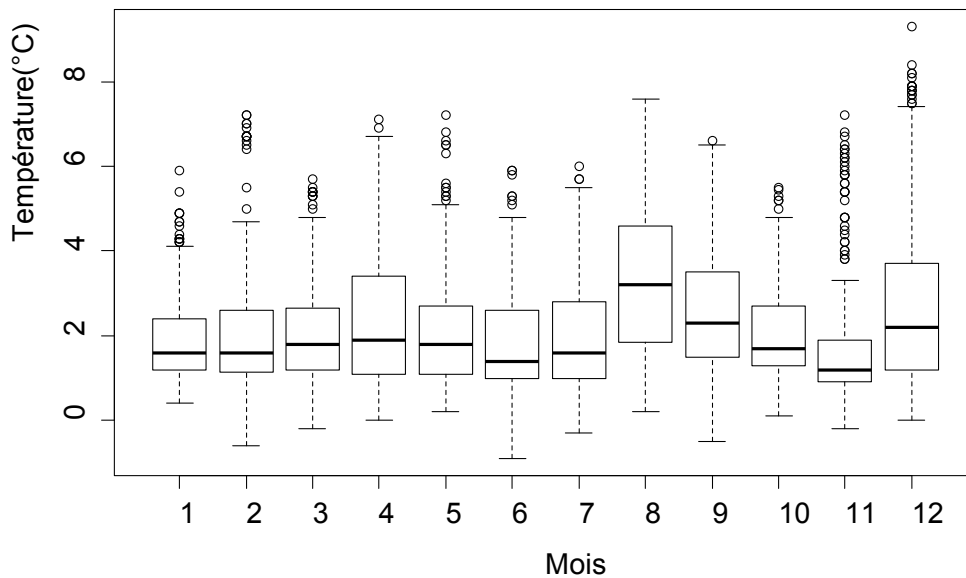


Figure 9: *Box plot* de la série des écarts de température nocturne entre Paris et Orly pour l'année 2013

4.4.4 Modélisation de la température

4.4.4.1 Première méthode : modélisation directe de la température

Pour la modélisation directe de la température, un modèle ARMA a été développé pour chaque mois de l'année. Seuls les résultats du mois d'août (correspondant au mois où l'îlot de chaleur est le plus marqué) seront détaillés.

4.4.4.1.1 Modélisation du mois d'août

➤ Etude de la tendance et de la saisonnalité

Les 14 premiers jours du mois d'août ont été utilisés pour déterminer l'ordre du modèle ARMA permettant de modéliser l'évolution de la température à Paris sur le mois d'août.

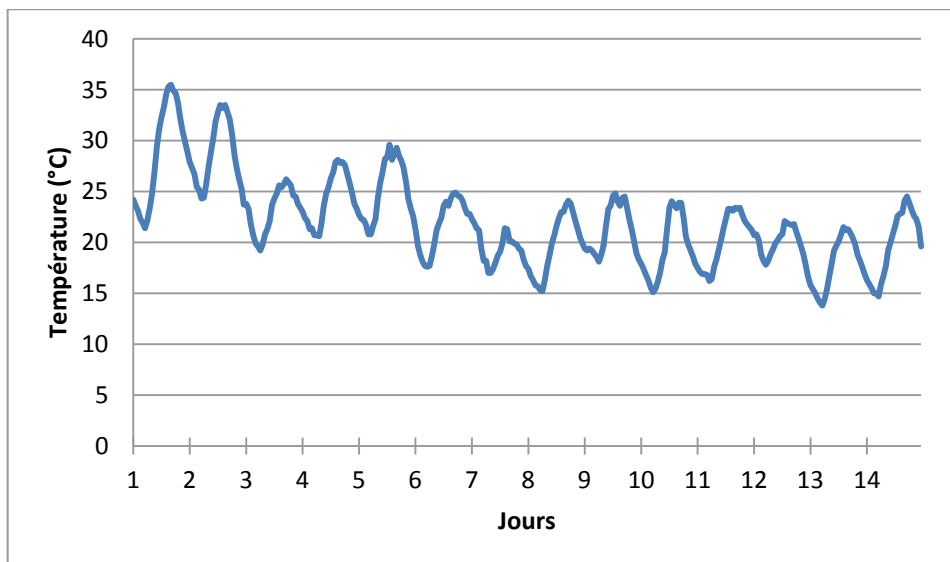


Figure 10: Température à Paris (mois d'août 2013)

Le graphe de l'évolution de la température à Paris lors du mois d'août (Figure 10) montre que cette série temporelle n'est pas stationnaire (son espérance et sa variance ne sont pas constantes au cours du temps). L'étude ACF de cette série est donc nécessaire.

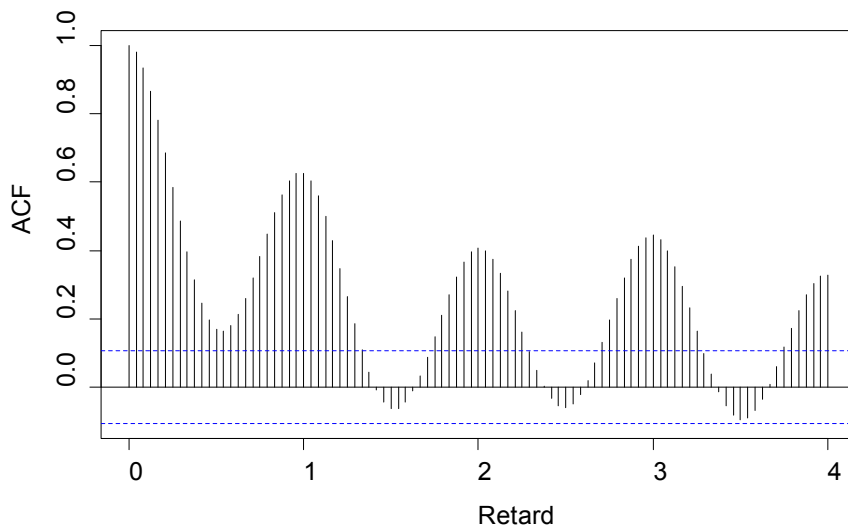


Figure 11: ACF de la température à Paris (mois d'août 2013)

Dans la Figure 11, les retards en abscisse sont indiqués en jours. Ainsi, un retard de 1 correspond à 24 h, un retard de 2 correspond à 48 h et ainsi de suite. On peut observer que l'ACF de la température présente des pics aux retards multiples de 24 qui ne s'atténuent que très lentement. On peut donc en déduire une non-stationnarité saisonnière. Cette non-stationnarité s'explique par la périodicité de 24 h de la série. De plus, on peut remarquer que l'ACF de la température présente un pic important au retard 1. Ces observations amènent à penser que les différenciations à l'ordre 1 ($d=1$) et à l'ordre 24 ($D=1, s=24$ h) sont nécessaires pour obtenir une série stationnaire.

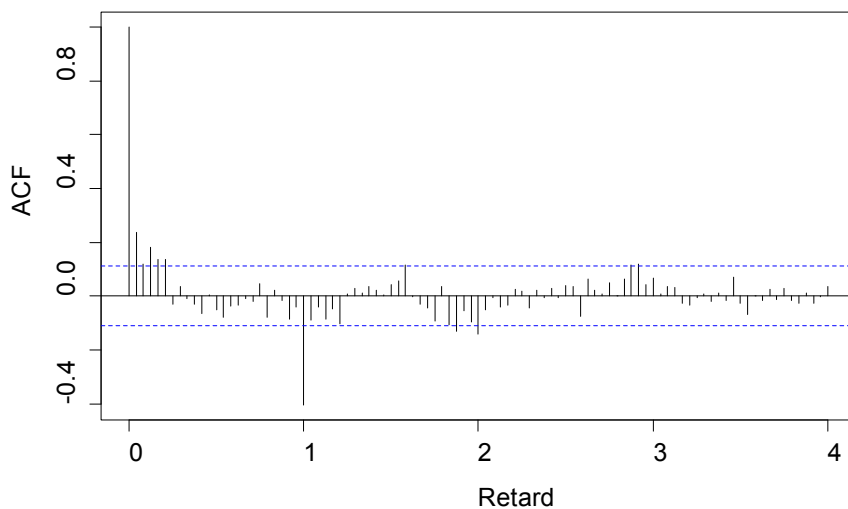


Figure 12: ACF de la température différenciée (Paris, mois d'août 2013)

On peut constater que l'ACF de la température différenciée (Figure 12) à l'ordre 24 et à l'ordre 1 $((1 - B)(1 - B^{24})X_t)$ montre un pic à 24 h (1 jour), caractéristique d'une série stationnaire avec saisonnalité. On peut donc conclure que la série temporelle de l'évolution de la température à Paris au mois d'août peut être modélisée par un modèle SARIMA($p,1,q$)($P,1,Q$) $_{24}$.

Chapitre 2 : Modèles nécessaires à la régulation

➤ Détermination du modèle

Pour déterminer l'ordre du modèle correspondant le mieux à la série temporelle, le logiciel R (R Core Team 2014), mettant en œuvre différents packages pour l'étude des séries temporelles et des modèles ARMA, a été choisi. Plus précisément, la fonction « *auto.arima* » du package « *forecast* » (Hyndman et Khandakar 2007) a été utilisée pour la sélection du modèle SARIMA se basant sur les critères automatiques (AIC, AICc, BIC). Le modèle déterminé par la fonction est un **SARIMA (1,1,0) (0,1,1)₂₄**.

➤ Prévisions

Après avoir déterminé l'ordre du modèle SARIMA à utiliser pour modéliser notre série de données, l'objectif est de prévoir la température sur un horizon de temps donné et de comparer la prévision avec les données de validation. L'horizon de temps choisi correspond à $h + 24$ à partir de la journée de prévision, c'est-à-dire que le modèle SARIMA prévoit les 24 prochaines heures de la station météorologique. Par exemple, si nous sommes à la 12^{ème} heure du jour J, alors $h + 1$ correspond à la 13^{ème} heure du jour J, $h + 3$ à la 15^{ème} heure du jour J et $h + 24$ correspond à la 12^{ème} heure du jour J+1.

Pour déterminer les coefficients du modèle SARIMA connaissant l'ordre du modèle (SARIMA (1, 1, 0) (0, 1, 1)₂₄), les données d'apprentissage correspondent aux deux semaines précédant le jour J de prévision (principe de la fenêtre glissante). Ces coefficients sont déterminés par la fonction « *auto.arima* » sur les données d'apprentissage. Ces coefficients sont à chaque fois mis à jour lors de la modification des données d'apprentissage. Les résultats de la prévision du 15 au 30 août sont présentés dans le Tableau 2. Ils correspondent à la moyenne sur 16 jours des indicateurs MAE et MBE.

Tableau 2 : Résultats de la prévision de la température (mois d'août 2013)

Indicateur	MAE	MAE	MAE	MAE	MAE	MBE	MBE	MBE	MBE	MBE
Horizon	1 h	3 h	6 h	12 h	24 h	1 h	3 h	6 h	12 h	24 h
Paris (mois d'août)/ Température (°C)	0,34	0,72	1,20	1,77	2,36	0,00	0,00	0,01	0,01	-0,08

On peut remarquer que les prévisions à court terme ($h + 1$, $h + 3$) réalisées par la modélisation directe sont performantes avec un MAE de 0,34°C et 0,72°C respectivement. De plus, la valeur des MBE indique que ces prévisions ne sont que très peu biaisées. Cependant, pour un horizon plus lointain, les performances se détériorent rapidement. Ainsi, pour une prévision à $h + 24$, l'erreur absolue est en moyenne de 2,36°C indiquant la difficulté pour la modélisation directe de faire une prévision fiable à des horizons de temps plus grands. Enfin, on peut remarquer que les prévisions pour tous les horizons sont peu biaisées.

➤ Analyse des résidus

Les différents tests de blancheur du résidu ont pu montrer que le résidu se comporte comme un bruit blanc.

4.4.4.2 Deuxième méthode : modélisation de l'écart (contextualisation)

Dans cette méthode, ce n'est pas la température de la station locale qui est prédite, mais son écart avec la station régionale.

4.4.4.2.1 Modélisation du mois d'août

➤ Etude de la tendance et de la saisonnalité

Les 14 premiers jours du mois d'août ont été également utilisés pour déterminer l'ordre du modèle ARMA permettant de modéliser l'évolution de l'écart de température entre Paris et Orly sur le mois d'août.

Le graphe de l'évolution de l'écart de température entre Paris et Orly (Figure 13) montre que cette série temporelle n'est pas stationnaire (son espérance et sa variance ne sont pas constantes au cours du temps). L'étude ACF de cette série est donc nécessaire.

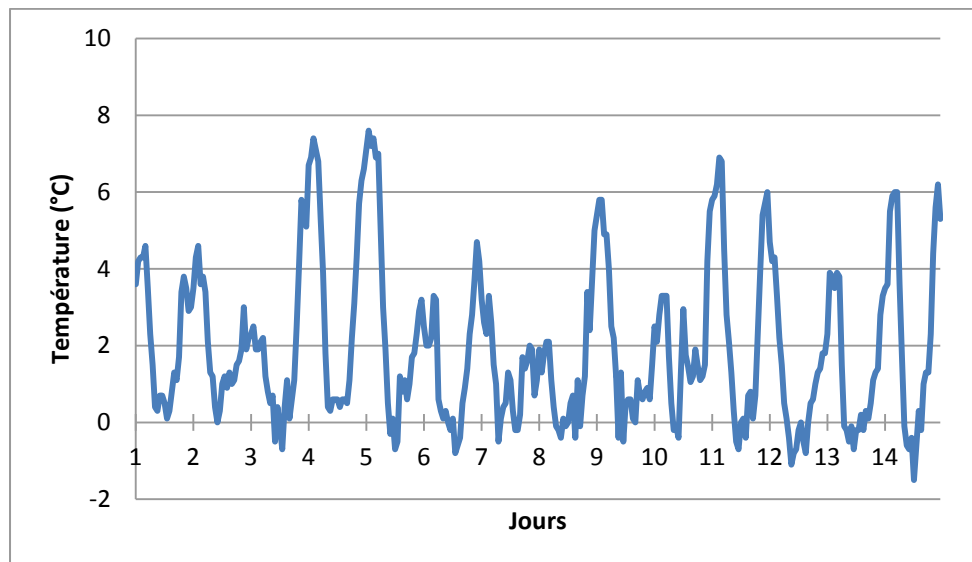


Figure 13: Écarts de température entre Paris et Orly (mois d'août 2013)

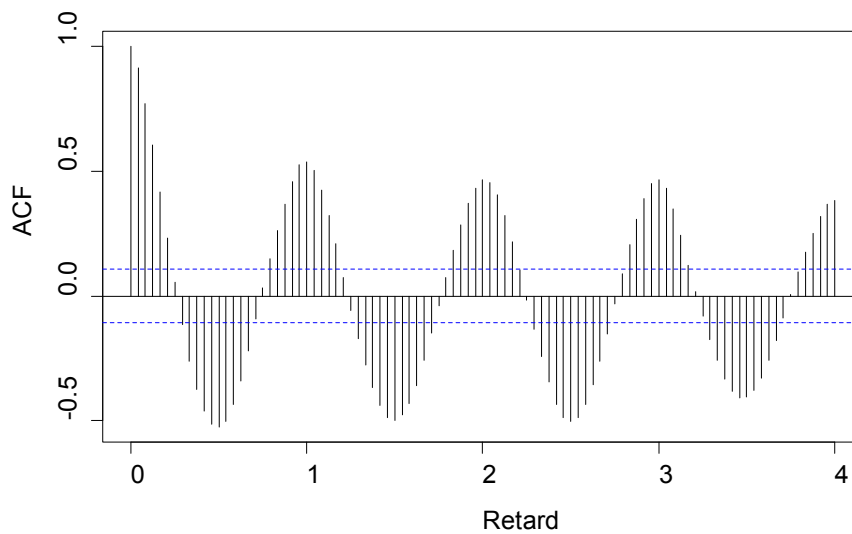


Figure 14: ACF des écarts de température entre Paris et Orly (mois d'août 2013)

Dans la Figure 14, on peut observer que l'ACF présente des pics (positifs) aux retards multiples de 24 h (1 jour) ainsi qu'un pic important au retard 1 h. Ces observations amènent à penser que les différenciations à l'ordre 1 et à l'ordre 24 sont nécessaires pour obtenir une série stationnaire.

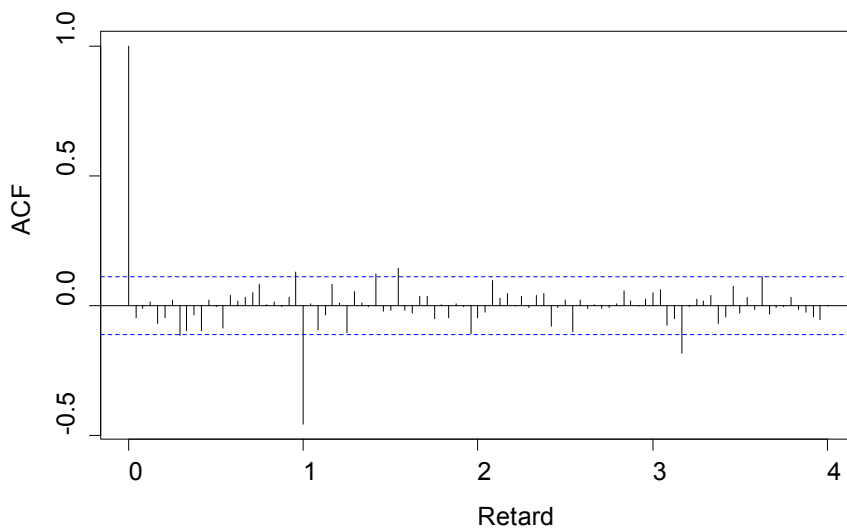


Figure 15: ACF des écarts différenciés de température entre Paris et Orly (mois d'août 2013)

On peut observer sur la Figure 15 la présence d'un pic à 24 h (1 jour), caractéristique d'une série stationnaire avec saisonnalité. On peut donc conclure que la série temporelle de l'évolution de l'écart de température entre Paris et Orly peut être modélisée par un modèle SARIMA(p,1,q)(P,1,Q)₂₄.

➤ Détermination du modèle

L'ordre du modèle déterminé par la fonction est un **SARIMA (0, 1, 0) (2, 1, 0)₂₄**.

Chapitre 2 : Modèles nécessaires à la régulation

➤ Prévisions

Les données d'apprentissage suivent le même principe que celles pour la modélisation directe de la température. Les résultats de la prévision sont présentés dans le Tableau 3. Ils correspondent à la moyenne sur 16 jours (du 15 au 30 août) des indicateurs MAE et MBE.

Tableau 3 : Résultats de la prévision des écarts de température (mois d'août 2013)

Indicateur	MAE	MAE	MAE	MAE	MAE	MBE	MBE	MBE	MBE	MBE
Horizon	1 h	3 h	6 h	12 h	24 h	1 h	3 h	6 h	12 h	24 h
Écarts (mois d'août)/ Température (°C)	0,64	1,03	1,39	1,80	1,40	0,00	0,00	-0,01	-0,04	-0,09

On peut remarquer que, comme pour la modélisation directe, la prévision à court terme ($h + 1$ par exemple) est performante (MAE de 0,64°C) et présente un biais négligeable. Ces performances sont cependant inférieures à celle de la modélisation directe. Au contraire, pour un horizon plus lointain, les performances de la modélisation des écarts sont supérieures à celles de la modélisation directe avec un MAE à $h + 24$ de 1,40°C (à comparer à un MAE de 2,36°C pour la modélisation directe). De plus, on peut noter que les prévisions sont peu biaisées dans l'ensemble.

➤ Analyse des résidus

Les différents tests de blancheur du résidu ont pu montrer que le résidu se comporte comme un bruit blanc.

4.4.4.3 Analyse des résultats sur l'ensemble de l'année

Le Tableau 28 de l'Annexe A récapitule les résultats obtenus concernant la modélisation de la température sur l'ensemble de l'année. Il permet de comparer les erreurs de la modélisation directe par rapport à celles de la modélisation des écarts. Le Tableau 4 présente un récapitulatif de ces résultats sur l'année (les résultats correspondent à la moyenne des indicateurs MAE et MBE mensuels).

Tableau 4 : Récapitulatif des résultats sur l'année 2013 de la modélisation de la température

Indicateur	MAE	MAE	MAE	MAE	MAE	MBE	MBE	MBE	MBE	MBE
Horizon	1 h	3 h	6 h	12 h	24 h	1 h	3 h	6 h	12 h	24 h
Modélisation directe	0,36	0,76	1,21	1,71	2,09	0	-0,01	-0,02	-0,03	-0,03
Modélisation écarts	0,51	0,75	0,97	1,15	1,09	0	-0,01	-0,02	-0,02	-0,03

On peut tout d'abord constater que pour la prévision à $h + 1$, la modélisation directe est plus performante que la modélisation des écarts. En effet, sur l'année, la modélisation directe permet d'obtenir une meilleure prévision de la température à $h + 1$ que la modélisation des écarts avec 0,36°C et 0,51°C respectivement. Pour la prévision à $h + 3$ les résultats sont quant à eux très proches avec en moyenne sur l'année 0,76°C pour la modélisation directe et 0,75°C pour la modélisation des écarts.

Concernant les prévisions à $h + 6$, $h + 12$ et $h + 24$, la modélisation des écarts est plus performante que la modélisation directe. Cette différence de performance est d'autant plus

Chapitre 2 : Modèles nécessaires à la régulation

remarquable pour la prévision à $h + 24$: l'erreur moyenne absolue de prévision obtenue par la modélisation des écarts est en moyenne de 1,09°C sur l'ensemble de l'année, alors qu'elle est de 2,09°C pour la modélisation directe. Ces résultats sont résumés dans la Figure 16, où l'on peut voir l'évolution sur l'année de l'écart moyen absolu des prévisions à $h + 1$ (MAE 1h), et $h + 24$ (MAE 24h) pour les deux types de modélisation.

Enfin, on peut constater que les erreurs moyennes (c'est-à-dire le biais, noté MBE) sont faibles pour la modélisation directe et la modélisation des écarts (Tableau 4). Par exemple, pour la modélisation directe les erreurs de prévision à $h + 1$ et $h + 24$ sont de 0°C (avec un écart type de 0,01°C) et de -0,03°C (avec un écart type de 0,35°C) respectivement. De même, pour la modélisation des écarts, les erreurs de prévision à $h + 1$ et $h + 24$ sont de 0°C (avec un écart type de 0,02°C) et de -0,03°C (avec un écart type de 0,11°C) respectivement (Tableau 28 de l'Annexe A).

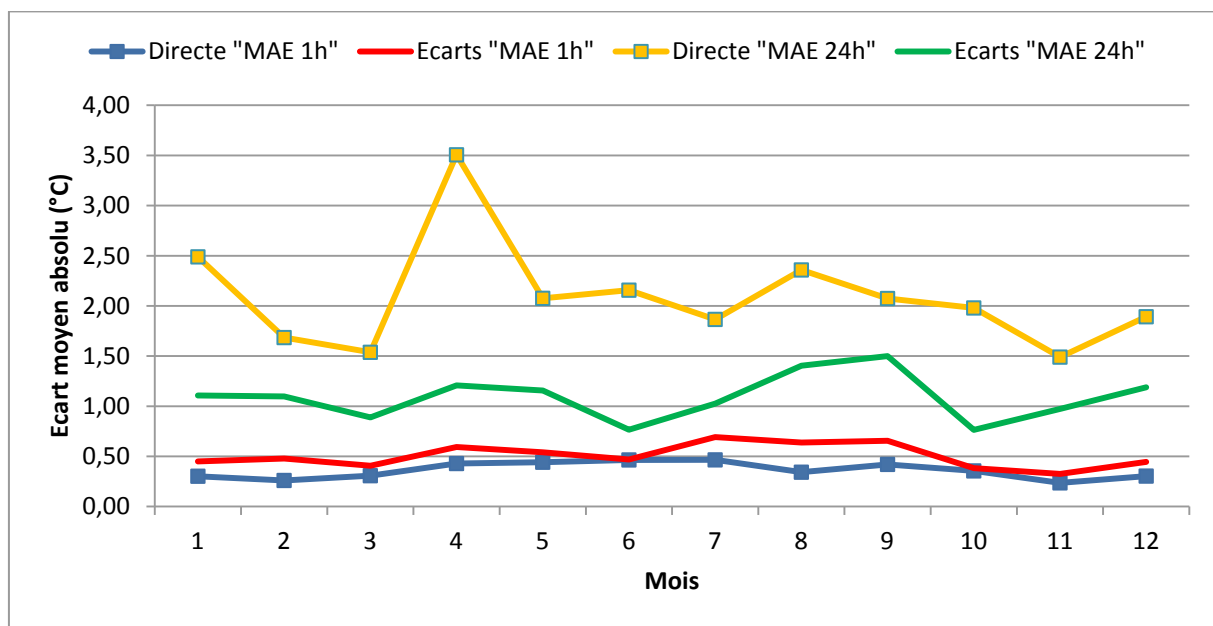


Figure 16 : Résultats des modélisations (directe et écarts) pour la prévision à 1 h et 24 h

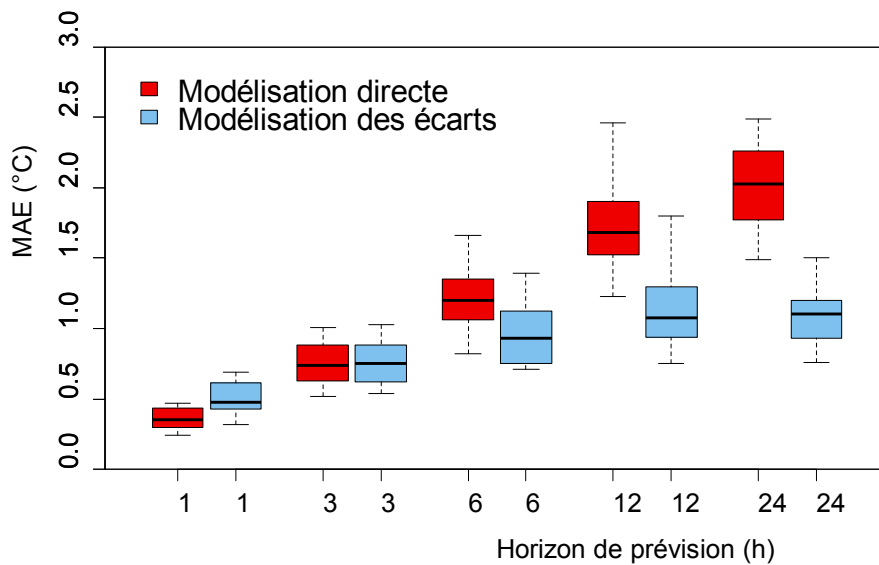


Figure 17 : Box plot des MAE obtenus par les modélisations de la température suivant l'horizon de prévision

L'utilisation de l'outil *box plot* (§4.4.3.2) dans la Figure 17 permet de présenter la dispersion statistique de l'écart moyen absolu (MAE) obtenu par les deux types de modélisation de la température suivant les différents horizons de prévision. Ce graphique illustre les remarques faites précédemment. On peut par exemple noter qu'à l'horizon $h + 1$ les performances de la modélisation directe sont nettement supérieures à celle de la modélisation des écarts avec une dispersion statistique plus faible. Le constat est inversé pour les horizons de temps plus grand ($h + 6$, $h + 12$ et $h + 24$) où la modélisation des écarts permet d'obtenir des résultats plus précis et moins dispersés statistiquement (en particulier pour l'horizon $h + 24$).

Il est également intéressant d'analyser la pertinence des résultats obtenus par la modélisation des conditions climatiques locales. En effet, il est utile d'étudier si les résultats obtenus par la modélisation de la météorologie locale (modélisation directe ou des écarts) sont plus précis que ceux obtenus par une prévision régionale. Ne disposant pas d'un modèle de prévision pour la station régionale, un modèle de prévision régional « parfait » a été considéré. Ce modèle de prévision parfait fait correspondre la prévision régionale à la mesure régionale (par exemple, la prévision de la température à $h + 1$ réalisée par le modèle régional correspond à la mesure de la température régionale à $h + 1$). Il est à noter que pour la modélisation des écarts, la prévision régionale utilisée est également considérée comme parfaite. L'utilisation de ce modèle parfait revient alors à comparer les résultats obtenus par la modélisation des conditions climatiques locales avec l'écart moyen entre les températures de Paris et Orly sur les données de validation. Les écarts moyens absolus entre les températures à Paris et Orly sont résumés dans le Tableau 6.

Chapitre 2 : Modèles nécessaires à la régulation

Tableau 5 : Écarts moyens absolus par mois entre les températures à Paris et Orly sur l'année 2013

Indicateur	MAE	MAE	MAE	MAE	MAE	MAE	MAE	MAE	MAE	MAE	MAE	MAE
Mois	01	02	03	04	05	06	07	08	09	10	11	12
Écarts entre Paris et Orly (°C)	1,75	1,9	1,4	1,84	1,3	1,15	1,31	1,6	1,64	1,63	1,34	1,7

On peut constater qu'en moyenne la modélisation des écarts est toujours plus performante que la prévision régionale parfaite (Figure 18 et Figure 19). Ainsi, même pour la prévision à $h + 24$, qui correspond à l'horizon où les performances sont les plus faibles pour la modélisation des écarts, les résultats obtenus par la modélisation de la météorologie locale sont meilleurs que ceux obtenus par la prévision régionale. La modélisation directe est quant elle plus performante que la prévision régionale « parfaite » pour les horizons de temps courts mais pas pour des horizons de temps longs (Figure 19).

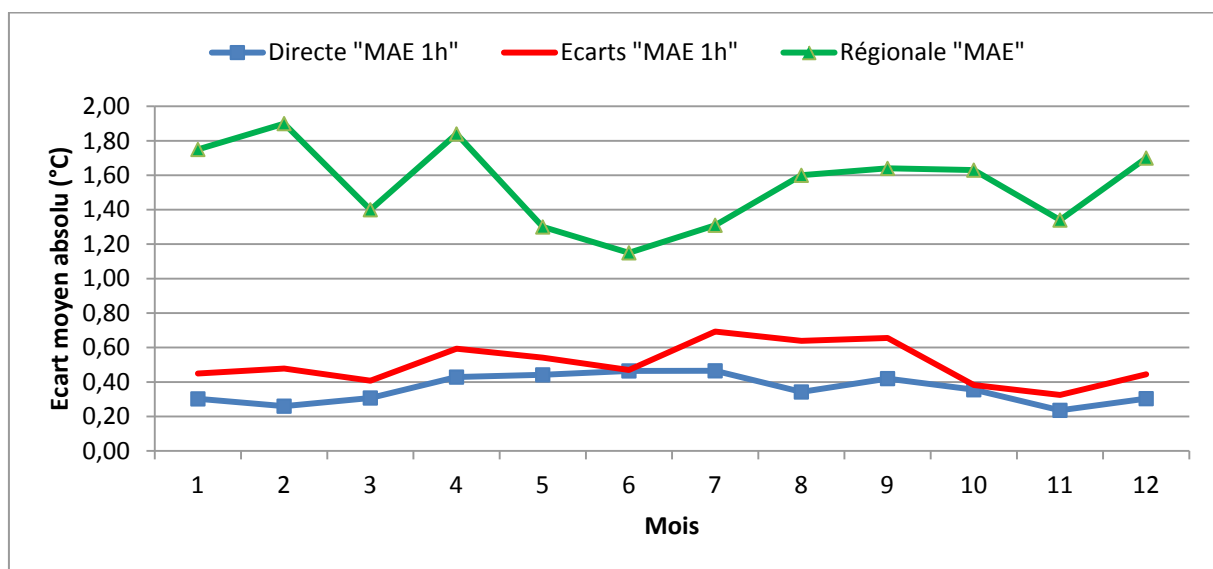


Figure 18 : Résultats des différents types modélisations de la température pour la prévision à 1 h

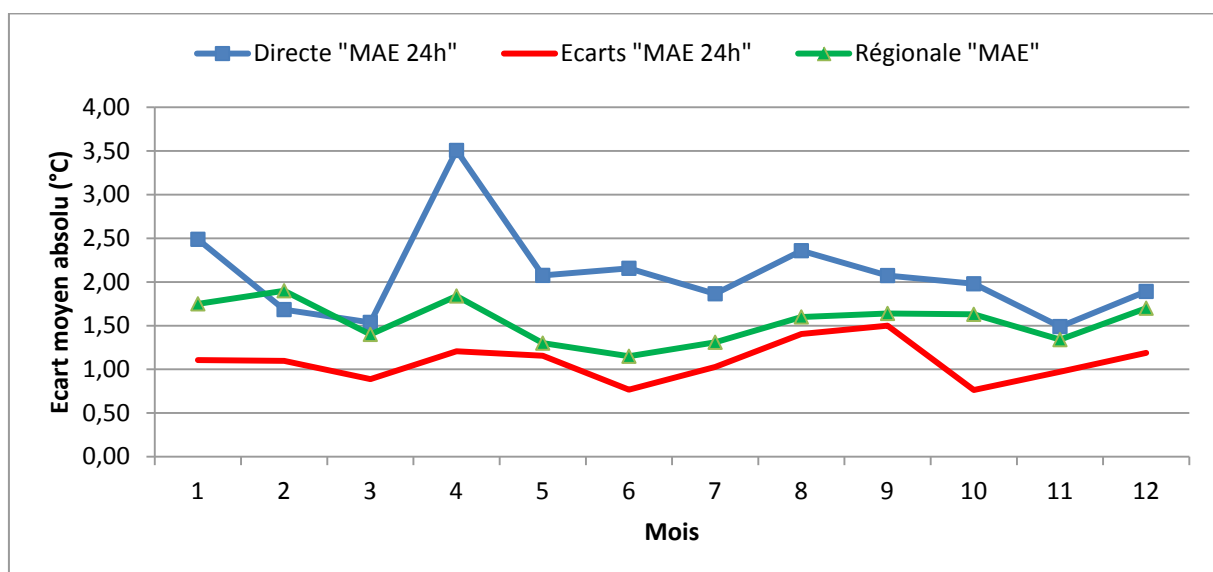


Figure 19: Résultats des différents types modélisations de la température pour la prévision à 24 h

Chapitre 2 : Modèles nécessaires à la régulation

Ces résultats confirment l'intérêt de la modélisation des écarts pour des horizons de prévision longs (supérieurs à trois heures). Cette méthode permet ainsi de profiter de la précision des prévisions issues des modèles physiques tout en gardant le caractère local des prévisions météorologiques. Cependant, pour des horizons de prévision plus courts (inférieurs à trois heures), la modélisation directe qui présente, dans cette étude, des résultats légèrement plus précis, est à privilégier pour obtenir des résultats plus précis.

4.4.5 Modélisation du rayonnement solaire

4.4.5.1 Méthodologie suivie

La méthodologie suivie pour modéliser le rayonnement solaire est là même que celle présentée dans le §4.4.4. Pour évaluer la performance des modèles identifiés, seules les prévisions concernant la journée ont été retenues dans le calcul des MAE et MBE.

4.4.5.2 Analyse des résultats sur l'ensemble de l'année

Le Tableau 29 de l'Annexe B rassemble les résultats obtenus pour la modélisation du rayonnement solaire. Le Tableau 6 présente un récapitulatif de ces résultats sur l'année.

Tableau 6 : Récapitulatif des résultats sur l'année 2013 de la modélisation du rayonnement (en W/m^2)

Indicateur	MAE	MAE	MAE	MAE	MAE	MBE	MBE	MBE	MBE	MBE
Horizon	1 h	3 h	6 h	12 h	24 h	1 h	3 h	6 h	12 h	24 h
Modélisation directe	45	76	81	81	88	0,4	0,3	-3	-3	-4
Modélisation écarts	25	27	26	26	27	0,2	0,2	-0,2	-0,8	0,1

Contrairement à la modélisation de la température, on peut constater que la modélisation des écarts est plus performante que la modélisation directe pour l'ensemble des horizons de prévision. Par exemple, pour la prévision à $h + 1$, l'erreur moyenne absolue sur les 12 mois est de $25 W/m^2$ pour la modélisation des écarts alors qu'elle est de $45 W/m^2$ pour la modélisation directe. De même, pour la prévision à $h + 24$, l'erreur moyenne absolue sur les 12 mois est de $27 W/m^2$ pour la modélisation des écarts alors qu'elle est de $88 W/m^2$ pour la modélisation directe. Ces résultats sont résumés dans la Figure 20 où l'on peut voir l'évolution sur l'année de l'écart moyen absolu des prévisions à $h + 1$ (MAE 1 h) et $h + 24$ (MAE 24 h) pour les deux types de modélisation.

On peut également constater que les erreurs moyennes (MBE) sont faibles pour la modélisation des écarts avec des erreurs moyennes de prévision à $h + 1$ et $h + 24$ de $0,2 W/m^2$ (avec un écart type de $1,5 W/m^2$) et $0,1 W/m^2$ (avec un écart type de $2,2 W/m^2$) respectivement (Tableau 29 de l'Annexe B).

La Figure 21 permet de juger de l'importance de l'erreur faite par la modélisation pour la prévision du rayonnement solaire à midi (avec comme horizon $h + 1$ et $h + 24$) en la comparant à la moyenne mesurée à midi. On peut constater que l'erreur est en proportion plus importante lors des mois d'hiver que des mois d'été. De plus, on peut remarquer que l'erreur obtenue par la modélisation directe à

Chapitre 2 : Modèles nécessaires à la régulation

l'horizon $h + 24$ est beaucoup plus importante que celle obtenue à l'horizon $h + 1$, contrairement à la modélisation des écarts pour laquelle les erreurs sont du même ordre de grandeur.

L'utilisation de l'outil *box plot* (§4.4.3.2) dans la Figure 22 permet de présenter la dispersion statistique de l'écart moyen absolu (MAE) obtenu par les deux types de modélisation du rayonnement suivant les différents horizons de prévision.

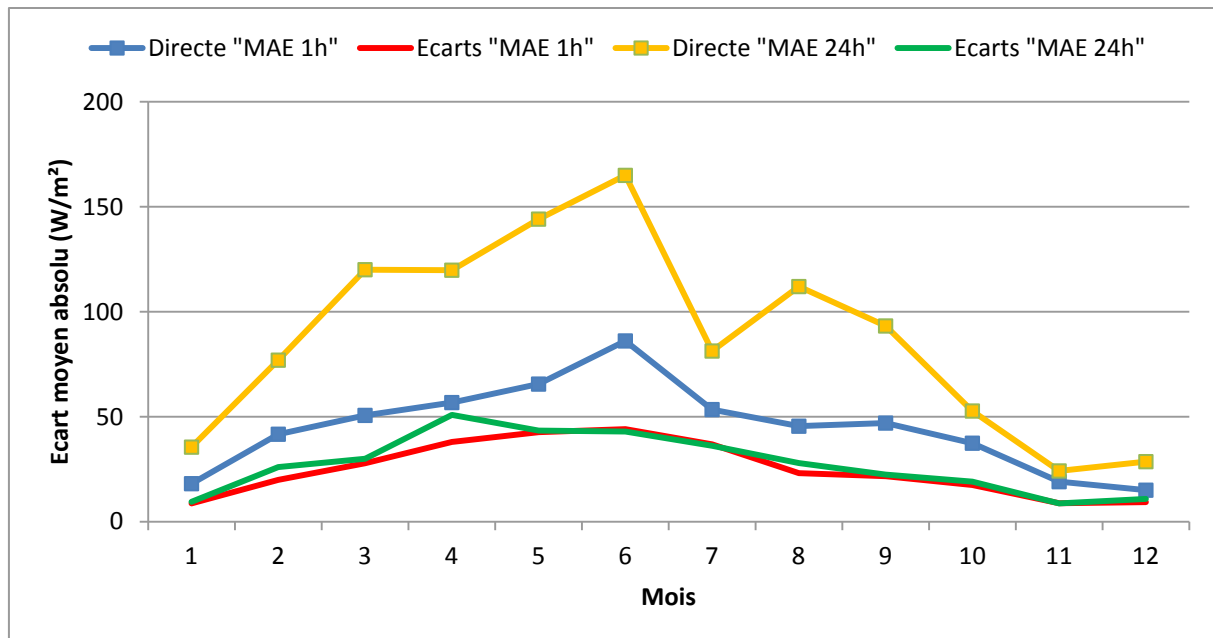


Figure 20: Résultats des modélisations (directe et écarts) pour la prévision à 1 h, et 24 h

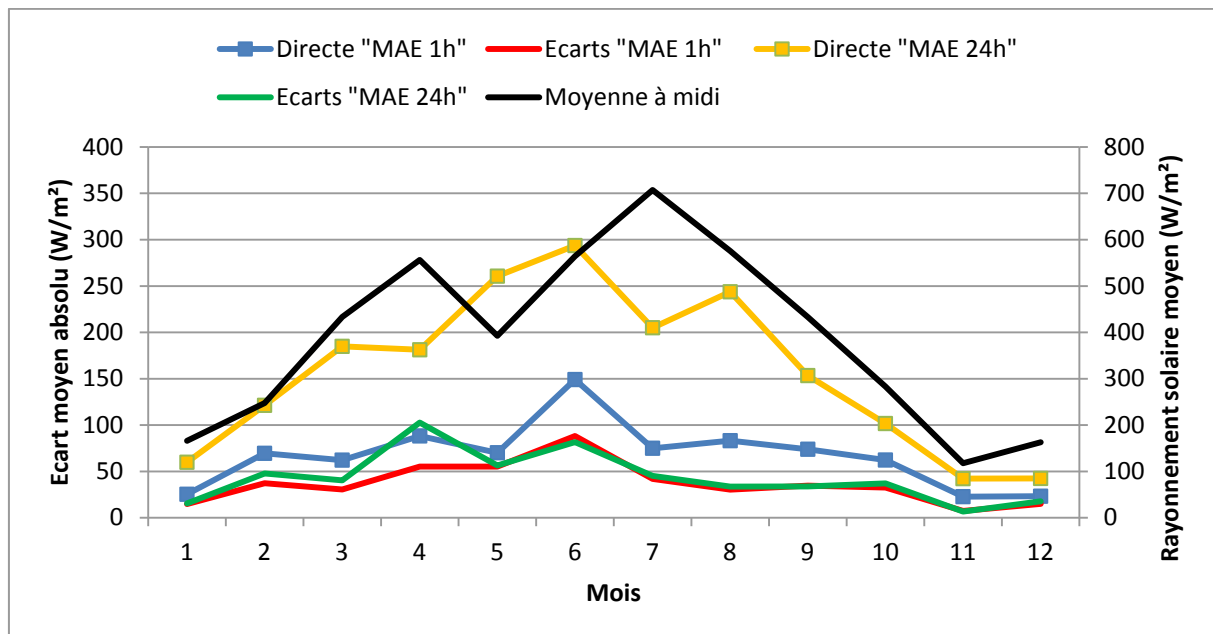


Figure 21: Comparaison entre les résultats de la prévision à midi et de la moyenne mesurée à midi

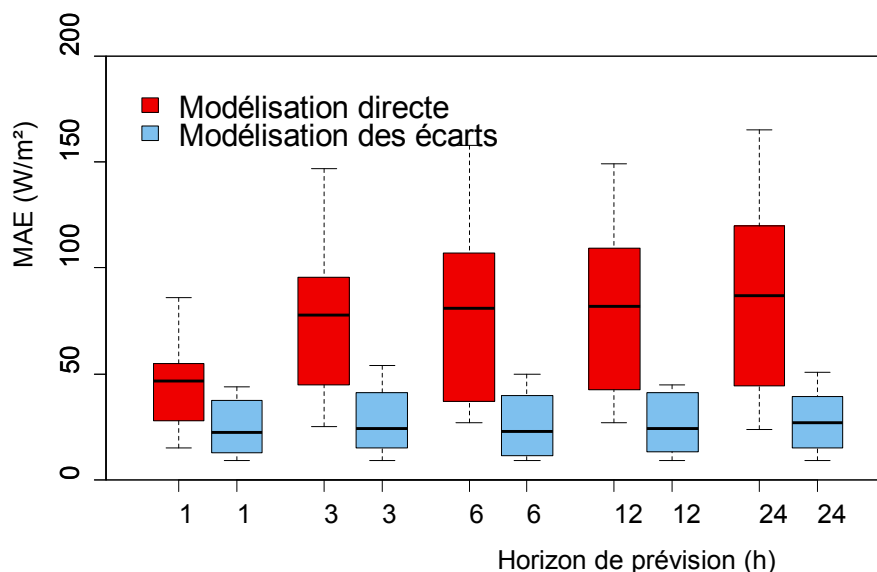


Figure 22: Box plot des MAE obtenus par les modélisations du rayonnement suivant l'horizon de prévision

Ce graphique illustre les remarques faites précédemment, à savoir que les performances de la modélisation des écarts sont nettement supérieures à celles de la modélisation directe, et cela pour tous les horizons de prévision. On peut également constater une plus faible dispersion statistique pour la modélisation des écarts par rapport à la modélisation directe.

Comme dans le cas de la modélisation de la température, il est intéressant d'analyser la pertinence des résultats obtenus par la modélisation des conditions climatiques locales en les comparant à ceux obtenus par une prévision régionale « parfaite ». Pour rappel, l'utilisation d'un tel modèle revient à comparer les résultats obtenus par la modélisation des conditions climatiques locales avec l'écart moyen entre le rayonnement à Paris et Orly sur les données de validation. Les écarts moyens absolus entre le rayonnement à Paris et Orly sont résumés dans le Tableau 7.

Tableau 7 : Écarts moyens absolus par mois entre le rayonnement à Paris et Orly sur l'année 2013

Indicateur	MAE	MAE	MAE	MAE	MAE	MAE	MAE	MAE	MAE	MAE	MAE	MAE
Mois	01	02	03	04	05	06	07	08	09	10	11	12
Écarts entre Paris et Orly (W/m ²)	10	20	27	31	34	32	26	21	21	16	7	10

On peut constater (Figure 23 et Figure 24) que la prévision régionale « parfaite » est plus performante que la modélisation directe et même que la modélisation des écarts (les écarts restent cependant proches), contrairement à la modélisation de la température.

Chapitre 2 : Modèles nécessaires à la régulation

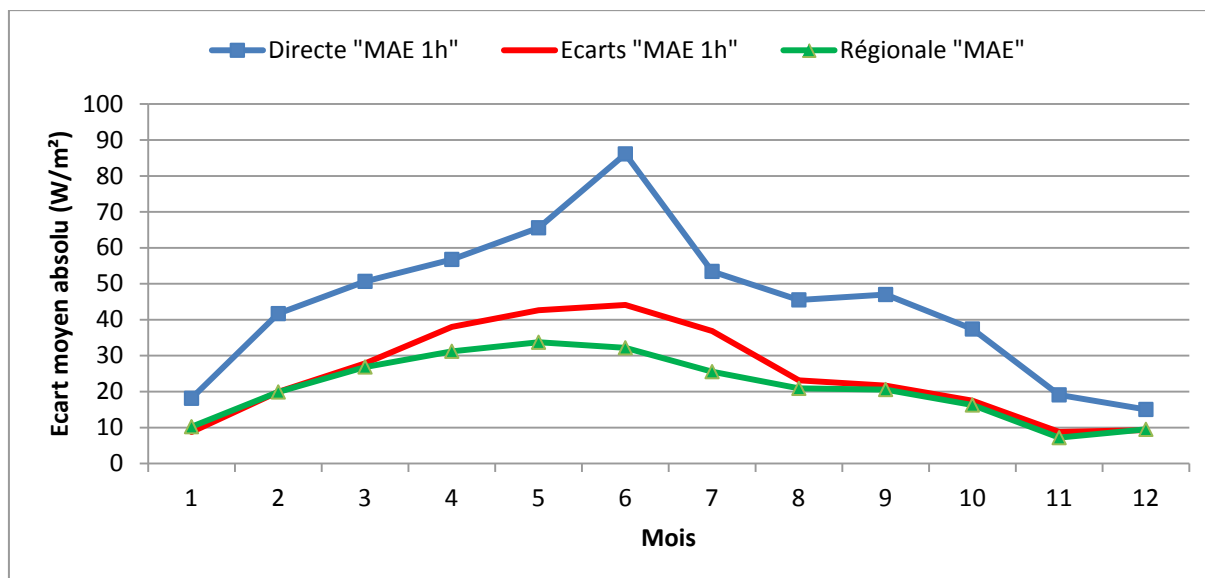


Figure 23 : Résultats des différents types modélisations du rayonnement pour la prévision à 1 h

Ces résultats confirment également l'intérêt de la modélisation des écarts par rapport à la modélisation directe. En effet, cette méthode permet de profiter de la précision des prévisions issues des modèles physiques tout en gardant le caractère local des prévisions météorologiques. Cependant, on constate qu'en ce qui concerne le rayonnement, la modélisation des écarts n'est pas plus performante que la prévision régionale parfaite (la modélisation des écarts n'apporte pas plus de précisions quant au rayonnement local que les prévisions régionales). Dans ce cas d'étude, la modélisation des écarts pour le rayonnement ne présente donc pas d'intérêt par rapport à la prévision régionale. Il faut cependant rappeler qu'il ne s'agit pas de mesures de rayonnement, mais d'estimations obtenues à partir d'observations par satellite.

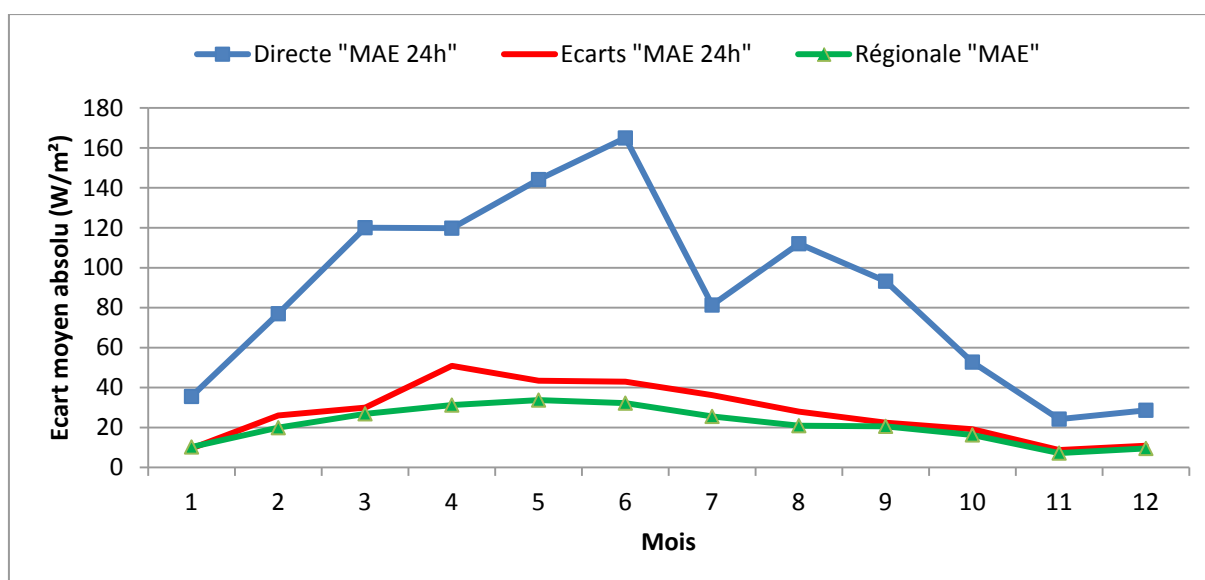


Figure 24 : Résultats des différents types modélisations du rayonnement pour la prévision à 24 h

4.5 Conclusion

L'objectif de cette étude était de développer une méthode de prévision à court terme de variables climatiques locales. Un état de l'art sur la modélisation des conditions climatiques a permis d'identifier les deux principales grandeurs à modéliser (la température extérieure et le rayonnement solaire) ainsi que les modèles, physiques ou mathématiques, les plus couramment utilisés. Suite à cette étude bibliographique, les modèles ARMA ont été choisis pour modéliser les conditions climatiques locales, et leur fonctionnement a été présenté.

Deux méthodes de modélisation ont été proposées : la modélisation directe des variables climatiques et la modélisation des écarts (ou contextualisation). La première méthode repose sur l'utilisation de l'historique des données météorologiques d'une station pour prévoir l'évolution des deux variables climatiques. La deuxième méthode repose quant à elle sur la modélisation de l'écart entre deux stations (une station dont on souhaite calculer les prévisions météorologiques, et une autre dont on dispose des prévisions météorologiques fournies par Météo France). Cette méthode a pour avantage de s'appuyer sur la précision des données issues des modèles physiques (Météo France) tout en gardant le caractère local des prévisions météorologiques.

Pour la prévision de la température, les résultats ont montré une plus grande précision concernant la modélisation des écarts pour des horizons de prévision supérieurs à trois heures. Par exemple, l'erreur moyenne absolue de prévision à $h + 24$ obtenue par la modélisation des écarts est en moyenne de $1,09^{\circ}\text{C}$ sur l'ensemble de l'année, alors qu'elle est de $2,09^{\circ}\text{C}$ pour la modélisation directe. Cependant, pour des horizons de prévision inférieurs à trois heures, la modélisation directe présente de meilleurs résultats. Ainsi, pour la prévision de la température à $h + 1$ l'erreur moyenne absolue sur l'année est de $0,36^{\circ}\text{C}$ pour la modélisation directe et $0,51^{\circ}\text{C}$ pour la modélisation des écarts. Concernant la prévision du rayonnement solaire, les performances de la modélisation des écarts sont supérieures à celles de la modélisation directe, et cela pour tous les horizons. Par exemple, l'erreur moyenne absolue de prévision à $h + 24$ sur les 12 mois est de 27 W/m^2 pour la modélisation des écarts alors qu'elle est de 88 W/m^2 pour la modélisation directe.

Les résultats ont ensuite été comparés à ceux obtenus en considérant une prévision locale identique à la prévision régionale (prévision régionale « parfaite »). Pour la modélisation de la température, la modélisation des écarts montre une plus grande précision que la prévision régionale. Concernant la modélisation du rayonnement solaire, la précision de la modélisation des écarts est légèrement inférieure à celle de la prévision régionale. Il faut cependant rappeler que les mesures du rayonnement sont issues d'observations par satellite.

Ces résultats confirment l'intérêt de modéliser les écarts entre deux stations (en particulier pour la température, dans une moindre mesure pour le rayonnement solaire) pour s'appuyer sur la précision des prévisions météorologiques fournies par les modèles physiques. De plus, cette méthode ne nécessite pas de supercalculateurs (contrairement aux modèles physiques) pour créer un modèle du climat local. Dans le cadre de la gestion énergétique des bâtiments, il est intéressant d'identifier les

conséquences de la précision des prévisions météorologiques sur la stratégie de régulation optimale. Par conséquent, les erreurs de prévisions de la température et du rayonnement obtenues respectivement par la modélisation des écarts et la prévision régionale seront utilisées pour évaluer la robustesse des stratégies de régulation (chapitre 4).

5. CONCLUSION DU CHAPITRE

Dans ce chapitre, les principaux modèles nécessaires à la mise en place d'une régulation ont été présentés. Le modèle thermique dynamique du bâtiment COMFIE sera utilisé. Ce dernier repose sur une modélisation multizone couplée à une réduction modale permettant une bonne précision ainsi que des temps de calculs réduits. La simulation de ce modèle de bâtiment nécessite de connaître les sollicitations auxquelles le bâtiment est soumis, telles que l'occupation et le climat. Ainsi, la prise en compte de l'occupation par des scénarios prédéfinis reposera sur une approche déterministe en considérant une zone de confort thermique définie par une température minimale et maximale. Le système de chauffage électrique sera quant à lui directement intégré dans le modèle du bâtiment. Concernant le climat, un cas idéal où les données météorologiques seront connues par avance avec exactitude permettra de développer des stratégies de régulation. Un modèle de prévision météorologique local sera ensuite utilisé pour évaluer la robustesse des stratégies de régulation développées ainsi que l'impact des erreurs de prévision sur la régulation, en prenant en compte les résultats obtenus par la modélisation des écarts pour la prévision de la température, et ceux obtenus par la prévision régionale (observations par satellite) pour la prévision du rayonnement.

Les modèles numériques et la simulation fournissent des outils pour appréhender le comportement du bâtiment. Mais des écarts subsistent entre les performances énergétiques prévues par la simulation et les consommations réelles. Il est alors nécessaire, dans le but de calculer les stratégies de gestion énergétique les plus performantes possibles, de calibrer le modèle de bâtiment à partir de relevés in situ. Ainsi, nous nous intéresserons dans le chapitre 3 aux méthodes de calibrage de modèle.

