Modèles prédictifs et adaptatifs pour la gestion énergétique du bâtiment résidentiel individuel : réseaux de neurones artificiels basés sur les données usuellement disponibles

Alexandra Delorme-Costil

▶ **To cite this version:**

Alexandra Delorme-Costil. Modèles prédictifs et adaptatifs pour la gestion énergétique du bâtiment résidentiel individuel : réseaux de neurones artificiels basés sur les données usuellement disponibles. Génie des procédés. Ecole des Mines d'Albi-Carmaux, 2018. Français. NNT : 2018EMAC0020 . tel-02475997

HAL Id: tel-02475997
https://theses.hal.science/tel-02475997

Submitted on 12 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Fédérale
Toulouse Midi-Pyrénées

THÈSE

en vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

délivré par
IMT – École Nationale Supérieure des Mines d'Albi-Carmaux

---

présentée et soutenue par

Alexandra DELORME-COSTIL

le 14 Juin 2018

Modèles prédictifs et adaptatifs pour la gestion énergétique du bâtiment résidentiel individuel : réseaux de neurones artificiels basés sur les données usuellement disponibles

---

École doctorale et discipline ou spécialité :
MEGEP : Énergétique et transferts

Unité de recherche :
Centre RAPSODEE, CNRS - UMR 5302, IMT Mines Albi

Directeur(s) de Thèse :
Jean-Jacques BÉZIAN, Ingénieur de recherche, IMT Mines Albi

Autres membres du jury :
Belkacem ZEGHMATI, Professeur, Université de Perpignan Via Domitia *(Rapporteur)*,
Christophe MENEZO, Professeur, Université de Savoie Mont Blanc *(Rapporteur)*,
Bernard BAUDOIN, Professeur, IMT Lille Douai *(Président)*,
Laurent ZALEWSKI, Professeur, Université d'Artois *(Examinateur)*,
Stéphanie GIROUX-JULIEN, Maître de Conférences, CETHYL, INSA de Lyon*(Examinateur)*,
Thomas PEDOT, Ingénieur de Recherche, Société BoostHeat *(Examinateur)*.

Modèles prédictifs et adaptatifs pour la gestion énergétique du bâtiment résidentiel individuel : réseaux de neurones artificiels basés sur les données usuellement disponibles

Alexandra DELORME-COSTIL

Sous la direction de

Jean Jacques BEZIAN

Thèse de doctorat soutenue le 14 juin 2018

IMT - École Nationale Supérieure des Mines d'Albi-Carmaux

Remerciements

J'adresse mes premiers remerciements aux membres de mon jury Mme Stéphanie Giroux Julien, M. Bernard Baudoin, M. Laurent Zalewski, M. Thomas PEDOT, M. Jean-Jacques Bézian et particulièrement à M. Belkacem Zeghmati et M. Christophe Menezo qui ont accepté d'être rapporteur. Je vous remercie d'avoir évaluer mon travail, et pour les passionnantes discutions que nous avons pu avoir. Merci à Jean-Jacques Bezian de m'avoir fait confiance pendant ces années de thèse et d'avoir soutenu les orientations que je souhaitais donner à ce travail. Merci à Olivier d'avoir entre autres été le premier à lire mon travail, et le premier à me conseiller python.

Je souhaite également remercier l'ensemble de l'équipe de BoostHeat, Jean-Marc Joffroy d'abord, initiateur des premières idées de mon sujet de thèse, Thomas Pedot ensuite, pour l'accompagnement, en Fortran, en python et surtout pour toutes les parties de ton travail que tu as partagé avec moi, et tout l'intérêt que tu as porté au mien, Guillaume Anies parce que c'est toi qui m'as "présentée" cette chaudière et essuyé mes questions naïves et pour finir tout les autres Flore, Guillaume D, Damien, Gwendoline, Matthieu, Martin, Thomas B, Julien, Yujie, Yassin, Mariano, Jerome, Loïc pour avoir fait d'un lieu de travail un espace si agréable.

La personne à qui je dois le plus grand des remerciements c'est l'homme qui partage ma vie, Geoffroy, parce que tu as été là, dans tous les instants de cette thèse, parce que sans toi je n'en serais pas là, parce que tu m'as apporté un soutien sans faille, et parce que tu m'as transmis un peu de la rigueur scientifique et de l'amour de la physique qui te caractérise, merci.

Merci aux membres de ma famille, au sens le plus large, d'avoir cru en moi, d'avoir fait ce long voyage à mes cotés (voyage spirituel et géographique) particulièrement à Corinne qui a donné de son temps pour que chacun puisse lire ce manuscrit sans avoir les yeux qui saignent devant mes fautes de français. Merci papa d'avoir traversé la France pour me voir devenir docteur. Merci à Cassiopée d'être un bébé si adorable et d'avoir fait ses nuits si tôt. Merci aux papis, mamies d'avoir pris soin d'elle lorsque mon attention se portait sur ce manuscrit.

Et pour finir, merci aux copains, Anastasia, Alexandra, Flore, François, Xavier,

Amélie, Julien, Christophe, Nelly, Victor, Olivier, Augustina, copains de toujours, du labo et d'ailleurs pour avoir partagé ces moments de vie avec un thè, un café, des rhum...

Abstract / Résumé

Modèles prédictifs et adaptatifs pour la gestion énergétique du bâtiment résidentiel individuel : réseaux de neurones artificiels basés sur les données usuellement disponibles

Résumé :

L'utilisation de régulation prédictive permet de diminuer la consommation d'énergie des bâtiments résidentiels sans diminuer le confort de l'habitant. C'est dans ce but que la société BoostHeat développe une chaudière thermodynamique à grande efficacité énergétique. La production simultanée de l'eau chaude sanitaire (ECS) et du chauffage permet de nombreuses stratégies de régulation pour en optimiser les performances. L'utilisation de régulations prédictives permet d'anticiper les besoins énergétiques, de prendre en compte l'impact de l'inertie du bâtiment sur la température intérieure et ainsi de faire des choix de gestion de productions minimisant la consommation énergétique. Les modèles utilisés aujourd'hui dans les régulations prédictives sont contraignants. En effet, ces modèles nécessitent de grandes quantités de données, soit sur un échantillon représentatif de bâtiments, soit sur chaque bâtiment modélisé. Ils peuvent également avoir besoin d'études détaillées sur le bâtiment, les occupants et leurs habitudes de consommation. Afin de permettre à la société BoostHeat d'utiliser une régulation prédictive sur sa chaudière sans passer par une étape complexe de modélisation à chaque installation de chaudière, nous proposons des modèles adaptatifs utilisant les informations communément disponibles sur une installation classique. Nous choisissons de développer des réseaux de neurones artificiels pour la prédiction d'une part des consommations d'ECS et d'autre part de la température intérieure du bâtiment. Les réseaux de neurones artificiels sont déjà utilisés pour modéliser les consommations énergétiques d'un bâtiment spécifique, cependant nos modèles sont génériques et s'adaptent automatiquement au bâtiment dans lequel la chaudière est installée. Plusieurs modèles sont développés afin d'étudier l'impact du choix des entrées, des quantités de données d'apprentissage et de l'architecture du réseau de neurones artificiels sur la qualité de la prédiction. Les modèles de prédiction de consommation d'ECS ainsi réalisés sont testés sur trois cas expérimentaux alors que les modèles de prédiction de température intérieure sont testés sur deux cas expérimentaux et cent-vingt cas simulés. Cela permet de tester leur adaptation à l'ensemble du parc immobilier français. Nous montrons, pour la prédiction de consommation d'ECS comme pour la prédiction de température

intérieure, que deux semaines de données collectées suffisent à une bonne adaptation des modèles à un cas spécifique. Le modèle le plus performant pour la prédiction de consommation d'eau chaude sanitaire a uniquement besoin des consommations des instants précédents. Le modèle de prédiction de température intérieure a de meilleures performances sur les bâtiments les moins isolés. Les résultats obtenus sont prometteurs pour l'application des régulations prédictives à grande échelle.

Mots-clés : Réseau de neurones artificiels, Apprentissage par rétropropagation de l'erreur, Prédiction de consommation d'ECS, Prédiction de température intérieure du bâtiment, Pompe à chaleur, Bâtiment résidentiel individuel.

Predictive and adaptive models for individual residential building energy management : Artificial neural networks based on commonly available data

Abstract :

The use of predictive control permits to reduce the energy consumption of residential buildings without reducing the comfort of the inhabitant. The company BoostHeat develops a thermodynamic furnace with high energy efficiency for this purpose. Simultaneous production of domestic hot water (DHW) and heating allows many control strategies to optimize performance. The use of predictive controls makes it possible to anticipate energy needs, to take into account the impact of building inertia on indoor temperature and thus to make production management choices that minimize energy consumption. The models used today in predictive controls are binding. Indeed, these models require large amounts of data, either on a representative sample of buildings or on each modeled building. They may also need detailed studies of the building, the occupants and their consumption practices. In order to allow BoostHeat to use predictive control without going through a complex modeling step at every furnace installation, we propose adaptive models using information commonly available on a typical installation. We choose to develop artificial neural networks for the prediction on the one hand of the consumptions of DHW and on the other hand of the ambiant temperature of the building. Artificial neural networks are already used to model the energy consumption of a specific building, however our models are generic and automatically adapt to the building in which the furnace is installed. Many models are developed to study the impact of the choice of inputs, amounts of learning data and artificial neural network architecture on the accuracy of prediction. The DHW consumption prediction models are tested on three experimental cases while the indoor temperature prediction models are tested on two experimental cases and one hundred and twenty simulated cases. This makes it possible to test their adaptation to the entire French housing stock. We show, for the prediction of DHW consumption as for the indoor temperature prediction, that two weeks of collected data are sufficient for a good adaptation of the models to a specific case. The most efficient model for the prediction of domestic hot water consumption only needs the consumptions of the previous instants. The indoor temperature prediction model performs better on less isolated buildings. The results obtained are promising for the application of predictive controls on a large scale.

Keywords : Artificial neural network, Back-propagation learning, Forecasting of DHW demand, Forecasting of indoor temperature, Heat pump, Individual residential building.

Table des matières

Abstract / Résumé **v**

Nomenclature **xi**

Introduction **1**

1 Etat de l'art **3**
1.1 Régulation intelligente : application à la chaudière thermodynamique BoostHeat . . . 3
1.1.1 La chaudière thermodynamique BoostHeat . . . 3
1.1.2 Régulation prédictive : intérêt pour la chaudière thermodynamique BoostHeat . . . 6
1.2 Les modèles prédictifs pour la gestion de chauffage : état de l'art . . . 10
1.3 Les modèles prédictifs de consommation d'eau chaude sanitaire dans l'habitat individuel : état de l'art . . . 15
1.4 Réseau de neurones artificiels : fondamentaux . . . 24
1.4.1 Perceptron multicouches . . . 24
1.4.2 Apprentissage supervisé par rétro-propagation de l'erreur . . . 27
1.4.3 Formatage des données d'entrée . . . 27
1.4.4 Développement de réseaux de neurones artificiels . . . 28
1.5 Conclusion . . . 29

2 Développement et évaluation des modèles : cas d'études utilisés **31**
2.1 Introduction . . . 31
2.2 Cas expérimentaux . . . 31
2.3 Cas simulés . . . 37
2.4 Conclusion . . . 41

3 Prédiction de la température intérieure du bâtiment **43**
3.1 Introduction . . . 43
3.2 Développement de l'architecture et de la méthodologie d'apprentissage 44
3.3 Évaluation des capacités d'un ANN à prédire la température intérieure d'un bâtiment . . . 50
3.3.1 Indicateurs de performance . . . 50

3.3.2 Cas expérimentaux : évaluation des architectures et méthodes d'apprentissage. 51
3.4 Évaluation du caractère adaptatif du réseau de neurones artificiels : Évaluation des prédictions sur les cas simulés 57
3.5 Conclusion 69

4 Prédiction des consommations d'eau chaude sanitaire 73
4.1 Introduction 73
4.2 Développement de l'architecture et de la méthodologie d'apprentissage. 74
4.3 Évaluation des capacités d'un ANN à prédire les consommations d'ECS. 79
4.3.1 Indicateur de performance 79
4.3.2 Cas 1 : évaluation des architectures et méthodes d'apprentissage. 80
4.4 Évaluation du caractère adaptatif du réseau de neurones A8. 89
4.5 Conclusion 96

Conclusion 99

Bibliographie 105

Liste des figures 120

Liste des tableaux 122

Nomenclature

Symboles latins	
C_p	capacité thermique spécifique
P	puissance
T	température
$\dot{V}$	débit volumique
V	volume
E	énergie
n	nombre d'exemple
min	minimum
max	maximum
t	instant

Indices	
int	intérieur
ext	extérieur
ECS	eau chaude sanitaire
$chauf$	chauffage
j	journalier
eau	eau
i	intermédiaire
e	entrée
S	sortie
app	apprentissage

Unité	
°C	degrée celsius
Wh	Watt heure
W	W
s	seconde
min	minute
h	heure

Symbole grecs	
Δ	symbole de la différence
ρ	masse volumique
ε	erreur signée
σ	écart-type
δ_{24h}	erreur sur 24h

Exposants	
dep	départ
ret	retour
$\hat{}$	prédiction
$\bar{}$	moyenne

Autres symboles	
$pers$	personne
$jour$	entier représentant le jour de la semaine

Sigles et acronymes	
ANN	réseau de neurones artificiels
ECS	Eau Chaude Sanitaire
SSC	Système Solaire Combiné
ADEME	Agence de l'environnement et de la maîtrise de l'énergie
RMSE	racine carré de l'erreur moyenne au carré

Introduction

Introduction et contexte

Le paquet énergie climat, adopté en 2008, fixe comme objectif la diminution de 20% des consommations d'énergie finale de l'Europe d'ici 2020. Le paquet énergie climat fixe également un objectif d'amélioration de l'efficacité énergétique de 27% à l'horizon 2030. À l'échelle française, la loi transition énergétique pour la croissance verte, fixe comme objectif la diminution de 20% des consommations d'énergie finale d'ici 2020, et de 50% d'ici 2050. Ces objectifs sont définis par rapport aux projections réalisées en 2012. Dans le secteur du bâtiment résidentiel, le Grenelle 1 fixe un objectif de diminution de 38% des consommations d'énergie en 2020. En France, les bâtiments résidentiels représentent 20% des consommations finales d'énergie, c'est plus que l'industrie [1]. Dans ce secteur le chauffage et l'eau chaude sanitaire sont les usages les plus consommateurs d'énergie. C'est dans ce contexte qu'est né le projet "APACHE" : Pompe À Chaleur gaz à Haute Efficacité pour les maisons individuelles neuves et existantes. Le porteur de projet, la société BoostHeat a conçu un compresseur thermique pour fluide frigorigène basé sur la compression thermique. BoostHeat a su réunir un consortium composé de la société ENGIE, et du laboratoire de recherche RAPSODEE, intégré à l'École des Mines d'Albi. Ce consortium a obtenu un financement du Fond Unique Interministériel afin de concevoir et de commercialiser une chaudière thermodynamique basée sur le compresseur développé par BoostHeat. En plus de concevoir une chaudière à haute efficacité la société BoostHeat a voulu voir plus loin et intégrer une régulation prédictive à leur produit. C'est avec cet objectif que les travaux de thèse présentés dans ce manuscrit ont débuté. Après de nombreuses recherches, il est apparu que le principal frein à l'intégration d'une régulation prédictive sur la chaudière BoostHeat soit la nécessité de développer des modèles de consommation d'énergie du bâtiment dans lequel la chaudière est installée. Nous nous intéressons donc au moyen de produire des prédictions fiables de consommation énergétique du bâtiment, avec une méthode automatique adaptée à tout type de bâtiment et d'habitant. Le travail se déroule en quatre phases : une étude de l'état de l'art afin d'identifier la méthode la plus adaptée à notre problème, le développement des modèles, l'évaluation de la précision des prédictions des modèles et l'évaluation de leur caractère adaptatif.

Structure du manuscrit

Le chapitre 1 a pour but d'identifier la méthode la plus appropriée à notre problème ainsi que les verrous technologiques empêchant sa mise en œuvre. Les performances et les besoins de chaque méthode sont mis en avant.

Le chapitre 2 présente les cas d'études qui servent à la fois à développer les réseaux de neurones artificiels et à tester la fiabilité de leurs prédictions, ainsi que leur adaptabilité.

Le chapitre 3 présente la méthodologie utilisée pour développer une architecture unique de réseaux de neurones artificiels capables de prédire la température intérieure du bâtiment. Les capacités de prédiction de ce réseau de neurones artificiels sont testées sur deux cas expérimentaux. Le caractère adaptatif de ce réseau de neurones artificiel est évalué sur cent-vingt cas simulés.

Le chapitre 4 présente la méthodologie utilisée pour développer une architecture unique de réseaux de neurones artificiels capables de prédire l'énergie consommée pour l'eau chaude sanitaire. Les capacités de prédiction de ce réseau de neurones artificiels et son caractère adaptatif sont testés sur trois cas expérimentaux.

CHAPITRE 1

Etat de l'art

1.1 Régulation intelligente : application à la chaudière thermodynamique BoostHeat

1.1.1 La chaudière thermodynamique BoostHeat

La chaudière thermodynamique BoostHeat est un système thermique destiné à fournir de l'eau chaude sanitaire (ECS) et le chauffage à un logement individuel. Elle fonctionne sur le principe d'une pompe à chaleur (PAC).

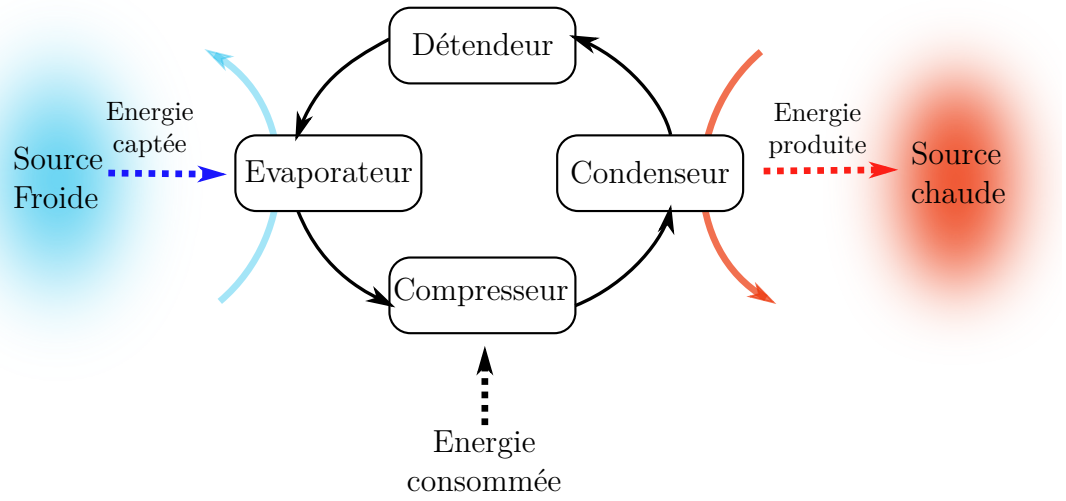


FIGURE 1.1 – Schéma d'une pompe à chaleur.

Fonctionnement d'une pompe à chaleur. Une pompe à chaleur peut être utilisée pour le chauffage ou la climatisation, ici nous nous intéressons uniquement au chauffage. Le principe de fonctionnement d'une PAC est de transférer de l'énergie

d'un milieu froid vers un milieu chaud. Elle est composée de quatre éléments fondamentaux : un compresseur, un condenseur, un détendeur et un évaporateur. Une représentation schématique est proposée en figure 1.1. Un fluide frigorigène circule à travers ces éléments. Le condenseur et l'évaporateur sont des échangeurs thermiques dont la fonction est de favoriser l'échange thermique entre le fluide frigorigène et la source froide ou la source chaude. Le rôle du compresseur est d'augmenter la pression du fluide frigorigène alors que le détendeur la diminue. Le fluide frigorigène subit un cycle thermodynamique, représenté en bleu sur le diagramme enthalpique en figure 1.2. La phase d'évaporation, 1 à 2 sur le diagramme, se fait à pression et température constante, l'énergie est transférée de la source froide vers le fluide, augmentant son enthalpie. Une surchauffe, 2 à 3 sur le diagramme, a lieu à la fin de l'évaporation afin de garantir un gaz sec en entrée du compresseur. La phase de compression, 3 à 4 sur le diagramme, augmente la pression et la température du fluide. C'est la phase consommatrice d'énergie. Pendant la phase de compression, 4 à 5 sur le diagramme, le fluide est refroidi par contact avec la source chaude, qu'il réchauffe, il passe alors à l'état liquide. C'est lors de cette étape qu'est produite l'énergie utile. Il peut être davantage refroidi, jusqu'au point 6 du diagramme. Le fluide frigorigène subit alors une détente isenthalpique(sans échange d'énergie), 6 à 1 sur le diagramme, sa pression diminue provoquant une chute de température et son évaporation partielle.

Les performances d'une PAC sont évaluées à l'aide du coefficient de perfor-

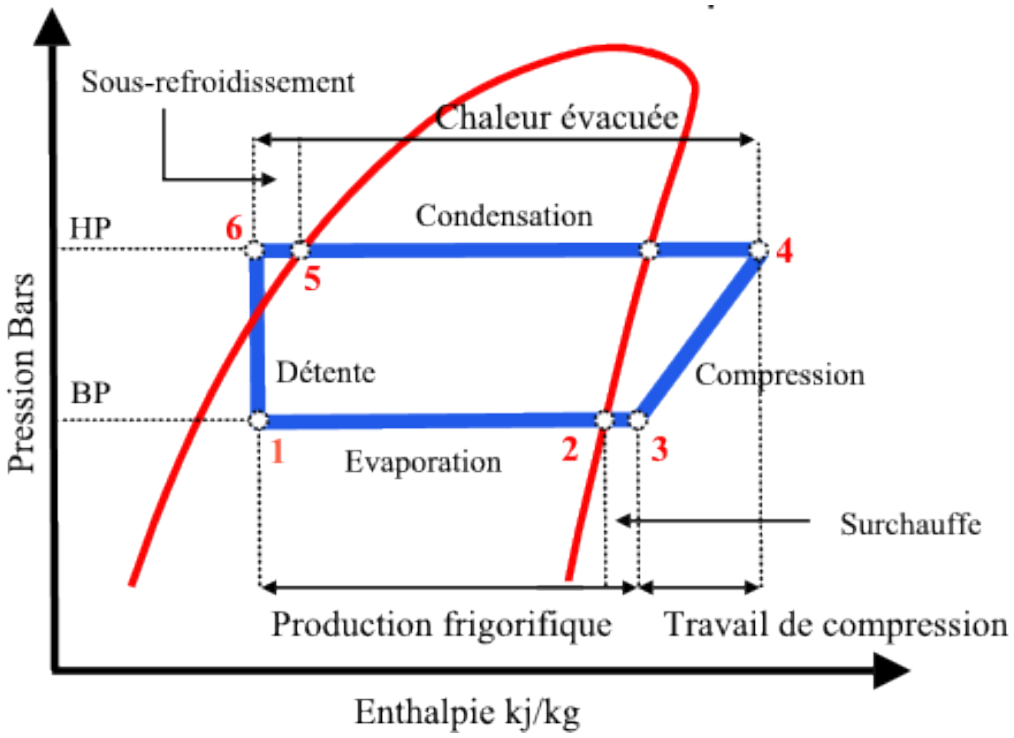


FIGURE 1.2 – Cycle thermodynamique d'un PAC sur le diagramme de Molier, extrait de [2].

mance(COP). Il est défini par le quotient de la puissance utile P_{utile} sur la puissance

consommée $P_{consommee}$, voir équation (1.1).

$$COP = \frac{P_{utile}}{P_{consommée}} \tag{1.1}$$

Les performances d'une pompe à chaleur sont liées à l'écart de température entre la source chaude et la source froide. Une plus grande différence de température viendra augmenter la phase de compression et diminuera le COP, voir figure 1.3.

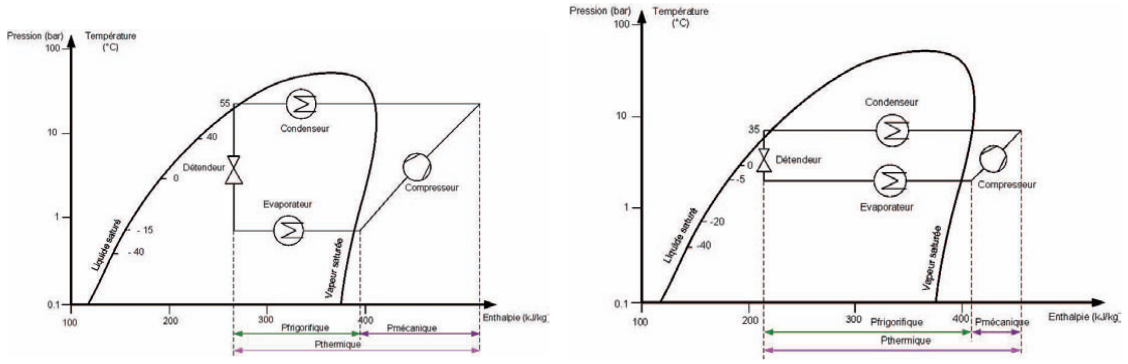


FIGURE 1.3 – Impact des températures sur le cycle thermodynamique, extrait de [2].

Le compresseur BoostHeat BoostHeat réalise une innovation sur la phase de compression. Le compresseur BoosthHeat est un compresseur thermique qui utilise la chaleur du brûleur gaz pour une mise en œuvre directe et à très haute efficacité de la compression du fluide frigorigène CO2 (R744). Inspiré de la technologie moteur Stirling, le compresseur joue à la fois le rôle d'un moteur et d'un compresseur dans une pompe à chaleur. Á la différence d'un compresseur volumique où la puissance est transmise par un travail mécanique, le compresseur BoostHEAT n'a pas un piston de travail mais un piston déplaceur, qui vient mettre le fluide frigorigène en contact successif avec la source chaude (brûleur) ou froide (air ambiant). En ce qui concerne l'intégration du compresseur dans le cycle d'une PAC, la compression est prévue pour être tri-étagée et pour travailler au-dessus du point critique du réfrigérant R744 (CO2). Une vue CAO en coupe du compresseur est proposée en figure 1.4.

- Phase de compression : sous l'apport de chaleur au niveau du régénérateur, le piston situé à gauche du cylindre se déplace vers la droite et la pression augmente. Les clapets de refoulement et d'aspiration sont fermés,
- Phase de refoulement : lorsque la pression est égale à P2 dans le cylindre (le piston se situe à la moitié du cylindre), le clapet de refoulement s'ouvre et le CO2 est refoulé à la pression P2 vers l'étage de compression suivant (ou vers le condenseur en sortie du 3ème étage de compression) jusqu'à ce que le piston soit à droite du cylindre,
- Phase de détente : l'impulsion est alors donnée au piston par le moteur électrique pour repartir vers la gauche du cylindre. La pression diminue et le clapet de refoulement se ferme,

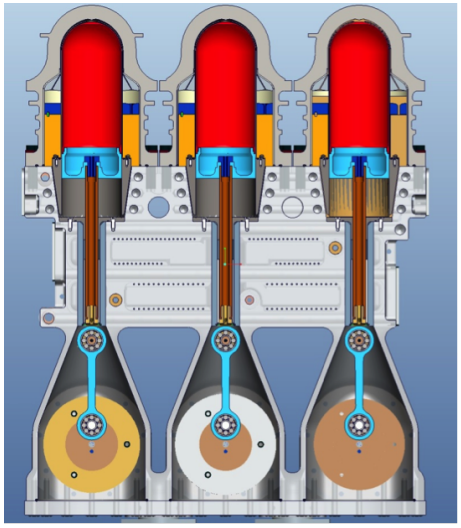


FIGURE 1.4 – Vue CAO du compresseur thermique BoostHeat.

- Phase d'aspiration : lorsque la pression a chuté à P1 dans le cylindre (le piston se situe à la moitié du cylindre), le clapet d'aspiration s'ouvre et le CO2 est aspiré à la pression P1 depuis l'évaporateur (ou depuis l'étage de compression supérieur) jusqu'à ce que le piston soit à gauche du cylindre.

Le circuit hydraulique Au circuit pompe à chaleur vient s'ajouter un circuit hydraulique permettant, via des échangeurs, de faire circuler de l'eau chaude dans une installation de chauffage classique et, via un ballon de stockage, de fournir l'eau chaude sanitaire (ECS). Un brûleur d'appoint est prévu dans le cas où la PAC ne pourrait pas produire la puissance demandée. Il permet aussi de produire l'ECS en instantané.

1.1.2 Régulation prédictive : intérêt pour la chaudière thermodynamique BoostHeat

Généralité. Le principe de la régulation prédictive est d'anticiper la variation future des éléments qui influence le système à contrôler ainsi que la variation de la commande. Dans notre cas, le système à contrôler est la chaudière, la commande est la température souhaitée dans l'habitat et la température et le débit d'ECS. Les éléments à prendre en compte sont les conditions climatiques, les habitants, la réaction thermique du bâti. Plusieurs méthodes de régulation prédictive sont utilisées pour les systèmes de production de chauffage et d'ECS : les réseaux de neurones artificiels, les régulations basées sur des modèles (communément appelées MPC de l'anglais Models Predictive Control), les systèmes multi-agents (MAST : multi-agent system technology). La plus répandue dans la littérature est la méthode MPC[3], que nous choisissons de détailler ici.

Régulation basée sur des modèles (MPC). La popularité des MPC (Model Predictive Control) a particulièrement augmenté ces dix dernières années [3]. Cette

méthode présente des propriétés intéressantes pour la régulation de systèmes dynamiques, particulièrement lorsqu'ils présentent des retards. Ce type de régulation consiste à prédire l'état du système sur un horizon de prédiction afin de déterminer un vecteur de contrôle sur un horizon de contrôle, comme illustré sur la figure 1.5. Les

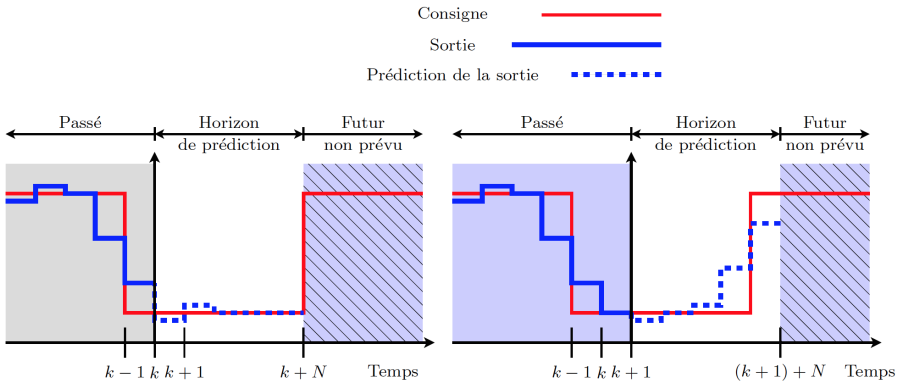


FIGURE 1.5 – Représentation temporelle du fonctionnement des MPC, extrait de [4].

trois temps caractéristiques mis en avant ici, le pas de temps, l'horizon de prédiction et l'horizon de contrôle ont un impact sur les performances du régulateur. Le pas de temps est la durée durant laquelle le vecteur de contrôle reste inchangé, l'horizon de prédiction est le temps pendant lequel les modèles vont prédire l'état du système, et l'horizon de contrôle est le temps sur lequel est calculé le vecteur de contrôle. Un schéma de principe est proposé en figure 1.6, il représente le calcul au temps t du vecteur de contrôle au temps t+k. Des modèles sont utilisés afin de prédire les états futurs du système, il faut distinguer les modèles du système, qui représentent son comportement propre, et les modèles des perturbations qui représentent les éléments qui viennent altérer l'équilibre du système. Les entrées de ces modèles sont les états réels du système, la consigne et le vecteur de contrôle aux pas de temps précédents. Ces prédictions alimentent une fonction coût qui sera minimisée sous contrainte à l'aide d'une méthode d'optimisation. Cette optimisation détermine le vecteur de contrôle. Appliqué au bâtiment, le vecteur de contrôle est composé des différents éléments permettant de piloter le système de production de chauffage ; le modèle du système représente les comportements thermiques du bâtiment et du système production de chauffage. Le modèle des perturbations peut être divisé en deux catégories : les perturbations internes et les perturbations externes. Le rendement d'une PAC est dépendant de la différence de température entre ses sources. Lorsque la source froide est l'air extérieur, comme c'est le cas dans la chaudière BoostHeat, la consommation d'énergie varie selon le moment de production dans la journée [5][6]. L'utilisation d'une régulation prédictive est donc pertinente afin d'optimiser les moments de production.

Chauffage : MPC appliqué à une PAC dans la littérature De nombreuses études montrent le potentiel d'économie énergétique et financière qu'offre l'application de MPC à des PAC. Afram et al. [7] développent une MPC pour contrôler

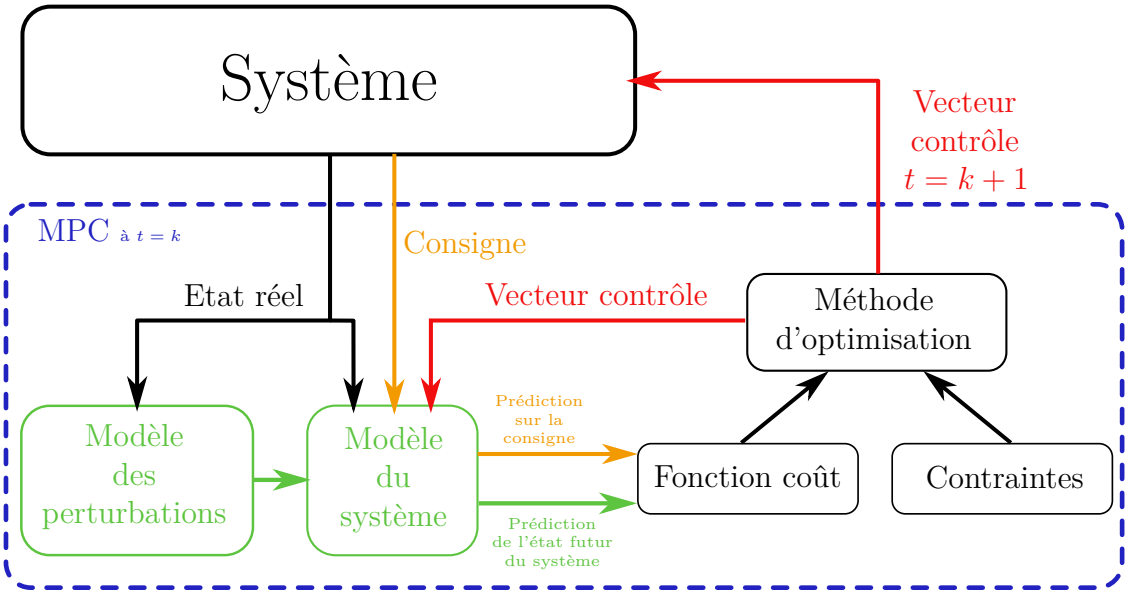


FIGURE 1.6 – Principe de fonctionnement des MPC.

une PAC géothermique. L'objectif est une économie financière. Cette MPC définit la température de consigne optimale en se basant sur les prédictions météorologiques et les informations de tarification de l'électricité. Ce système est testé par simulation et expérimentalement, il montre des résultats similaires dans les deux cas. Ils présentent une économie d'au moins 16% par rapport à une température de consigne fixe.
Kandler et al. [8] développent une MPC pour une PAC non-modulable avec stockage thermique. Ils montrent une diminution de 40% du temps de fonctionnement de la PAC et une diminution des pertes du système de stockage.
Tahersima et al. [9] travaillent sur un système de PAC géothermique. Cette fois l'objectif de la MPC est de maximiser le COP de la PAC et de minimiser la température de départ du circuit de chauffage. Ils s'aperçoivent que les deux objectifs sont opposés, la diminution de la température de départ engendre une baisse du COP.
Verhelst et al.[10] économisent de 5% sur la facture d'électricité en testant plusieurs formulations de la fonction coût d'une MPC sur une PAC. Hazyuk et al.[11] réduisent les consommations d'énergie de 18% sur une PAC par rapport à la régulation PID.

ECS : régulation prédictive. Bien qu'un très grand nombre de travaux soit réalisé sur l'intérêt des régulations prédictives dans la réduction des consommations énergétiques du bâtiment d'habitation individuelle, peu d'études s'intéressent exclusivement aux consommations d'ECS. Pourtant Fuentes et al. [12] montrent une variation de 13% du COP d'une pompe à chaleur selon le profil de consommation ECS. Les travaux présentés ici justifient l'intérêt de mettre en place des régulations prédictives propres à la production d'ECS. Ces travaux sont effectués pour différents systèmes de production et différents types de consommateurs.
Jin et al.[13] développent une régulation basée sur des modèles pour une pompe à chaleur (PAC) air/eau. Une représentation schématique du fonctionnement de cette régulation est proposée en figure 1.7. Les consommations d'ECS sont utilisées par un modèle afin de prédire les consommations futures, ces prédictions alimentent un

modèle simplifié de la PAC, qui couplé à un algorithme d'optimisation détermine le point de contrôle optimal.Ce type de régulation permet de pré-chauffer le ballon, ce qui diminue l'utilisation de l'appoint. Les auteurs s'appuient sur la coordination des sources de chaleur de la PAC. La régulation est testée par simulation sur deux maisons, sur des données recueillies pendant plus d'un an. L'économie d'énergie pour la maison consommant le moins est de 20% par rapport à une régulation classique.
Halvgaard et al. [14] mettent en place une régulation basée sur des modèles portant

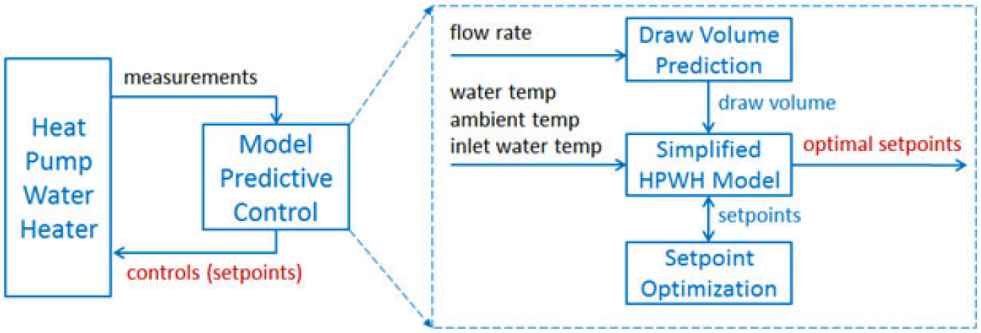


FIGURE 1.7 – Schématisation d'une MPC pour le contrôle de l'ECS, extrait de [13]

sur un système de production d'ECS solaire muni d'un appoint électrique. L'objectif ici est de minimiser le coût de production. Les auteurs montrent une économie de 25% à 30%. Cette étude est réalisée par simulation sur une durée d'un an. La MPC mise en place nécessite d'excellentes prédictions d'ensoleillement et de consommation d'ECS. Les prédictions d'ECS sont réalisées à partir de profil type créé à partir de données recueillies sur un groupe de maisons individuelles au Danemark.
Iglesias et al. [15] proposent une régulation intelligente d'un système de production d'ECS centralisé à l'échelle d'un quartier. Cette régulation est basée sur des profils de consommation. Les auteurs déplorent le manque d'intérêt apporté à la régulation de l'ECS par rapport au système de chauffage ou de climatisation, alors que cela représente un potentiel d'économie énergétique important.
Knudsen et al. [16] montrent une économie financière de 5% avec l'utilisation d'une régulation prédiction par rapport à une régulation classique. Les prédictions de consommation d'ECS sont faites sur 24h à partir d'un modèle très simple : la moyenne glissante des 60 dernières heures. Ce choix est fait dans le but d'utiliser un modèle de prédiction qui ne demande pas d'étude préalable.

Conclusion. Pour le chauffage de bâtiments résidentiels, l'utilisation de régulation prédictive se justifie par la forte inertie du bâti. Selon la qualité thermique du bâtiment, une variation de température extérieure ou de commande de chauffage mettra plusieurs heures à impacter la température intérieure de l'habitat. Il faut donc anticiper la modification de la consigne pour garantir un confort thermique optimal. Dans le cas de l'ECS, ce sont les pertes thermiques dues au stockage d'eau chaude qui justifient ce type de régulation. Dans le cas d'une PAC, c'est la variabilité du rendement en fonction du moment de production qui rend intéressante la mise en place de régulation prédictive. La chaudière BoostHeat a en plus l'avantage d'être bivalente, production de l'ECS et du chauffage et de posséder un système d'appoint.

Ce système d'appoint a un rendement inférieur à la partie PAC. Cette configuration technique complexifie la régulation du système et augmente le nombre de stratégies de contrôle possible. Un enjeu majeur des régulations prédictives est la connaissance des besoins futurs, en chauffage et en ECS. La suite de ce travail porte sur les modèles utilisés pour faire ces prédictions.

1.2 Les modèles prédictifs pour la gestion de chauffage : état de l'art

Dans la partie précédente, nous avons montré que l'utilisation d'une régulation prédictive pour la gestion du chauffage dans un logement résidentiel permet d'économiser de l'énergie. La mise en place de ce type de régulation nécessite des prédictions des consommations. Dans cette partie, nous allons décrire les quatre principales méthodes permettant de développer des modèles de prédiction de consommation de chauffage : les modèles basés sur le bilan énergétique, les modèles basés sur l'analogie électrique Résistance, Condensateur : RC, les modèles statistiques et les réseaux de neurones artificiels.

Modèle du bâtiment basé sur le bilan énergétique Faire le bilan énergétique d'un bâtiment consiste à répertorier l'ensemble des apports et pertes énergétiques du bâtiment. En considérant le bâtiment comme un système isolé, nous pouvons dire que la somme des apports et des pertes est nulle. Selon la précision souhaitée du modèle les échanges énergétiques pris en compte seront plus ou moins représentatifs de la réalité. Castilla et al. dans [17] appliquent cette méthode pour développer un modèle intégré à une MPC. Les auteurs considèrent le bâtiment comme une unique pièce. La température de l'air à l'intérieur de cette pièce est calculée à l'aide des équations (1.2). Les échanges énergétiques pris en compte sont le gain thermique par convection naturelle par les murs, Q_{conv}, le gain thermique par les fenêtres,Q_{glass},le gain thermique par ventilation forcée (chauffage) Q_{HVAC}, le gain thermique dû aux apports internes (cuisson, habitant, éclairage,etc.), Q_{iGain}, le gain thermique dû aux infiltrations, Q_{inf}, le gain thermique par ventilation naturelle, $Q_{natvent}$. L'ensemble de ces gains étant considéré comme des apports s'ils sont positifs et comme des pertes s'ils sont négatifs. Les gains sont calculés à partir d'un grand nombre d'informations : capacité spécifique de l'air, humidité relative, ensoleillement diffus et direct, masse d'air dans la pièce, nombre de personnes, températures de l'air intérieur, extérieur et pulsé, température des murs. Le premier membre de l'équation $m_a.C_{pa}.\frac{dT_{air}}{dt}$ représente l'apport thermique reçu par l'air de la pièce pendant la variation de temps dt.

$$m_a.C_{pa}.\frac{dT_{air}}{dt} = Q_{conv} + Q_{glass} + Q_{HVAC} + Q_{natvent} + Q_{inf} + Q_{igain} \tag{1.2}$$

Le même type de modèle est utilisé dans une MPC appliquée au stockage thermique pour le refroidissement [18]. Ces modèles bien que très précis et descriptifs du phénomène physique sont très peu appliqués car il nécessite un grand nombre d'informations et donc de capteurs. De plus, ces informations ne sont pas toujours

simples à mesurer, la température d'un mur ou de l'air d'une pièce est rarement homogène, le point de mesure a donc énormément d'importance. Les performances de ces modèles physiques sont fortement impactées par des mesures erronées ou biaisées.

Modèle basé sur l'analogie électrique Résistance, Condensateur : RC Une autre méthode répandue pour modéliser le bâtiment est l'analogie avec un circuit électrique composé de résistances et de condensateurs. Cette méthode introduite par L.Laret [19] propose de considérer les résistances thermiques des échanges conductifs, convectifs, et radiatifs comme des résistances électriques, les températures comme des potentiels, les inerties comme des capacités et les flux comme des puissances injectées à des nœuds. Une représentation de bâtiment en circuit électrique est proposée sur la figure 1.8. Ce modèle développé par Oldewurtel [20] est intégré à une MPC. Les

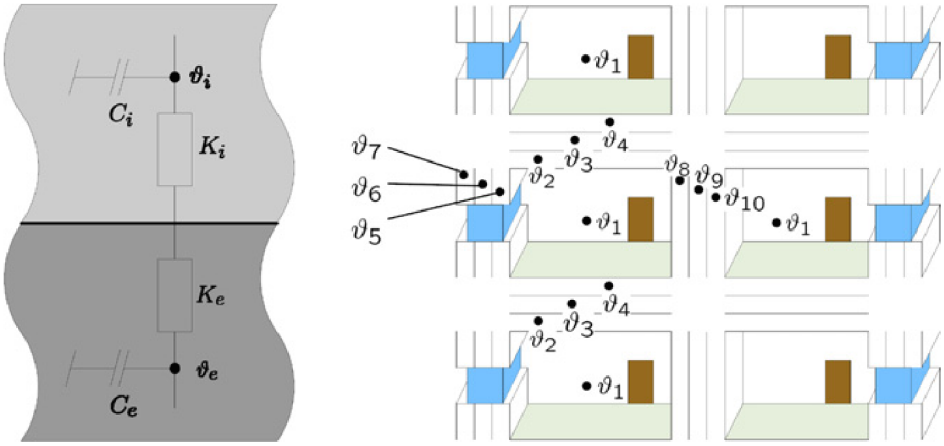


FIGURE 1.8 – Modèle du bâtiment par analogie électrique, extrait de [20]

nœuds représentent la température de l'air dans la pièce, du mur, du plafond ou du sol. Le transfert thermique est donné par l'équation (1.3), avec t :le temps, ϑ_i : la température de la couche i, ϑ_e : la température de la couche e, Q : l'énergie thermique, A : l'aire de la section d'échange, U_{ie} : le coefficient de transfert.

$$\frac{dQ}{d\vartheta_i}.\frac{d\vartheta_i}{dt} = A.U_{ie}(\vartheta_e - \vartheta_i) \ . \tag{1.3}$$

Ce calcul est reproduit pour chaque nœud. D'autres exemples d'utilisation de ce modèle sont disponibles en référence [21], [22] et [23]. Ce type de modèle est aussi appelé modèle déterministe semi-physique(DSPM de l'anglais Deterministic semi-physical Modeling) lorsque les paramètres sont ajustés à partir de simulation ou d'expérimentation [24]. Cette méthode est très répandue car elle reste descriptive et suffisamment précise bien que plus simple à mettre en œuvre que la méthode basée sur le bilan énergétique.

Modèle statistique. Concernant les modèles statistiques, deux méthodes se détachent, la méthode de l'erreur prédite (PEM de l'anglais Prediction Error Methods)

et la modélisation semi-physique probabiliste. La méthode de l'erreur prédite (PEM de l'anglais Prediction Error Methods) a pour but de minimiser la prédiction de l'erreur sur un pas de temps par optimisation de paramètres pré-spécifiés du modèle. Typiquement, le modèle ARMAX (de l'anglais AutoRegressive Moving Average with eXternal input) est décrit dans [25] comme suit :

$$y(t)+a_1.y(t-1)+...+a_{n_a}.y(t-n_a) = b_1.u(t-1)+...+b_{n_b}.u(t-n_b)+e(t)+c_1.e(t-1)+...+c_{n_c}.e(t-n_c)\ , \tag{1.4}$$

Les paramètres ajustables du modèle sont :

$$\theta = [a_1\ a_2...a_{n_a}\ \ b_1\ b_2...b_{n_b}\ \ c_1\ c_2...c_{n_c}\ \ ]^T\ , \tag{1.5}$$

avec

$$A(q^{-1}) = 1 + a_1.q^{-1} + ... + a_{n_a}.q^{-n_a}\ , \tag{1.6}$$

$$B(q^{-1}) = 1 + b_1.q^{-1} + ... + b_{n_b}.q^{-n_b}\ , \tag{1.7}$$

$$C(q^{-1}) = 1 + c_1.q^{-1} + ... + c_{n_c}.q^{-n_c}\ , \tag{1.8}$$

L'équation (1.5) devient

$$A(q^{-1}).y(t) = B(q^{-1}).u(t) + C(q^{-1}).e(t)\ , \tag{1.9}$$

équivalent à :

$$y(t) = \frac{B(q^{-1})}{A(q^{-1})}.u(t) + \frac{C(q^{-1})}{A(q^{-1})}.e(t)\ , \tag{1.10}$$

Où y est la sortie du système et u l'entrée du système.

Cette méthode nécessite une grande quantité de données afin de définir les paramètres du modèle par des méthodes statistiques. Ce modèle est utilisé dans [26],[27]et [28].

La modélisation semi-physique probabiliste est basée sur des équations différentielles stochastiques, une hiérarchie de modèles de complexité croissante est formulée à base de connaissances physiques, les paramètres sont adaptés à l'aide de la méthode du maximum de vraisemblance. Cette technique est détaillée dans [29]. Les méthodes statistiques sont plus simples à mettre en œuvre que les méthodes physiques présentées précédemment. Cependant leur dépendance à des données historiques peut biaiser le modèle, de plus ces données historiques ne sont pas facilement disponible.

ANN La méthode des réseaux de neurones artificiels (ANN, de l'anglais Artificial Neural Networks) a été développée par biomimétisme. Un ANN est composé d'entités mathématiques, mimant les neurones, reliées entre elles par des liaisons, mimant les synapses. La façon d'organiser les neurones et de les relier les uns aux autres est appelée architecture. Á chacune des liaisons est attribué un poids, considéré comme paramètre du ANN. Ces paramètres sont déterminés, par apprentissage, à

partir de données. Les ANN sont depuis longtemps reconnus pour la modélisation et le contrôle de systèmes physiques dynamiques, en particulier quand ce dernier n'est pas entièrement connu analytiquement[30]. Lindelöf et al.[31] utilisent des ANN pour modéliser le bâtiment et le climat. Ils prédisent la température extérieure, l'ensoleillement et la température intérieure du bâtiment. La figure 1.9 montre les architectures choisies pour ces ANN. Dans cette étude, les modèles prédictifs sont utilisés au sein d'une MPC. C'est l'énergie économisée par le système de régulation et non la qualité des prédictions qui est évaluée. Kusiak et al. [32] développent deux

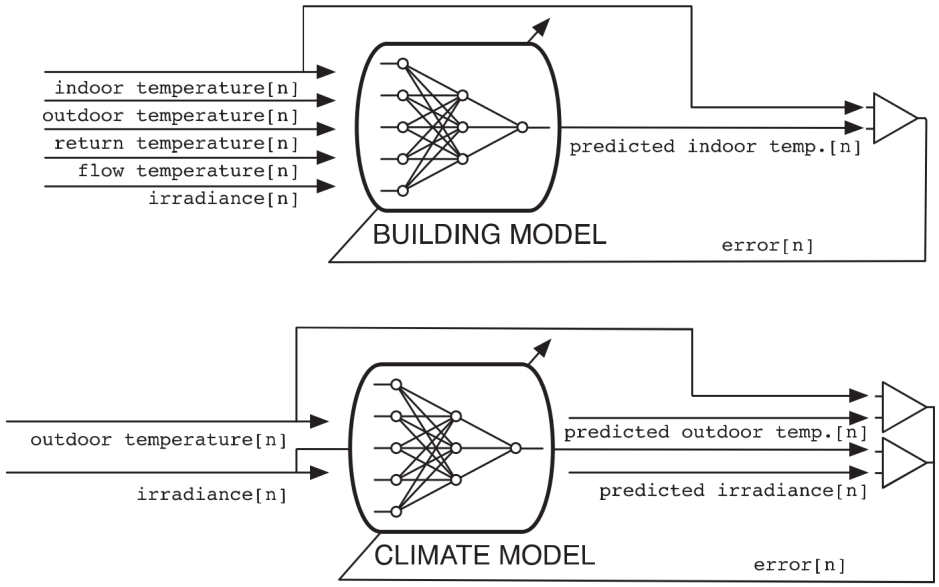


FIGURE 1.9 – Architecture de réseaux de neurones pour la modélisation du bâtiment et du climat, extrait de [31].

ANN. L'un prédit la consommation de chauffage, l'autre la température intérieure. Une prédiction est faite toutes les 30 minutes. Les résultats obtenus sur ces deux modèles sont présentés sur la figure 1.10. Le modèle de prédiction de consommation d'énergie montre une précision de 91,7% et celui de prédiction de température, 99,6%.

Aydinalp et al. [33] utilisent les réseaux de neurones pour prédire la consomma-

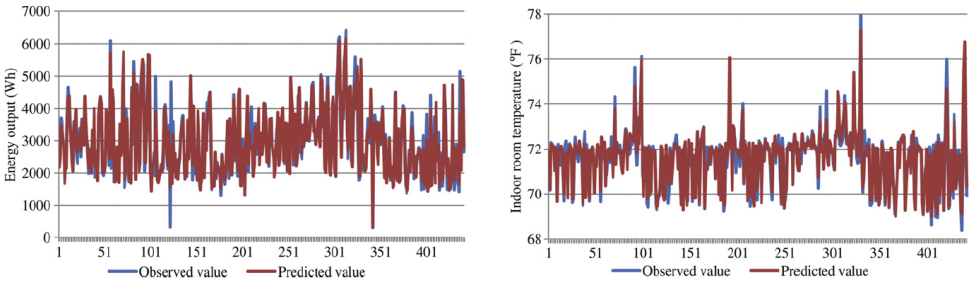


FIGURE 1.10 – Comparaison des grandeurs prédites aux grandeurs observées, à gauche l'énergie consommée pour le chauffage, à droite la température intérieure, extrait de [32].

tion annuelle de chauffage d'une habitation. Ils utilisent beaucoup d'entrées, voir

figure 1.11, regroupées en quatre catégories : caractéristiques du bâtiment, propriétés du système de chauffage, effets météorologiques et critères socio-économiques. Plusieurs architectures sont développées. Les modèles se montrent incapables de prédire les consommations pour les maisons consommant moins de 30GJ/an. Les auteurs expliquent ce phénomène par des influences non prises en compte dans les données d'entrée ou d'apprentissage, telles que de longues vacances, qui font chuter les consommations.

Table 2
SH NN model input units

	Input unit	Range
Dwelling characteristics	Dwelling type: 1 if single-detached; 0 if single-attached	0–1
	Number of doors	1–11
	Number of triple-glazed windows	0–30
	Number of double-glazed windows	0–48
	Number of single-glazed windows	0–24
	Wall area [m^2]	71–733
	Floor area [m^2]	17–265
	Basement wall-area [m^2]	0–163
	Basement floor-area [m^2]	0–265
	Roof area [m^2]	17–265
	Dwelling age category	1–6
	Wall-age category	1–6
	Roof-age category	1–6
	Basement wall-age category	0–6
	Basement floor-age category	0–6
	Percentage of the basement heated [%]	0–100
	Heated garage: 1 if heated; 0 if not heated	0–1
SH system and equipment properties	End-use efficiency of the SH equipment [%]	65–100
	Presence of heat-recovery ventilation system	0–1
	Presence of programmable thermostats	0–1
Indoor and outdoor temperatures	Average indoor temperature [°C]	16–24
	Heating degree-days [°C-day]	2930–6541
Socio-economic characteristics of the households	Income [$1,000/yr]	10–85
	Dwelling ownership: 1 if owner; 0 if renter	0–1
	Number of children	0–6
	Number of adults	1–6
	Daytime occupancy	0–1
	Size of area of residence: 1 If population is less than 15,000 2 If population is between 15,000 and 100,000 3 If population is 100,000 or over	1–3

FIGURE 1.11 – Présentation des entrées utilisées pour la prédiction de consommation annuelle de chauffage, extrait de [33].

Conclusion. Dans cette partie, nous avons présenté les avantages et inconvénients des méthodes permettant de développer des modèles prédictifs de bâtiment. La méthode des ANN, par son adaptabilité via des données mesurées sur le bâtiment est la plus adaptée aux travaux que nous développons.

1.3 Les modèles prédictifs de consommation d'eau chaude sanitaire dans l'habitat individuel : état de l'art

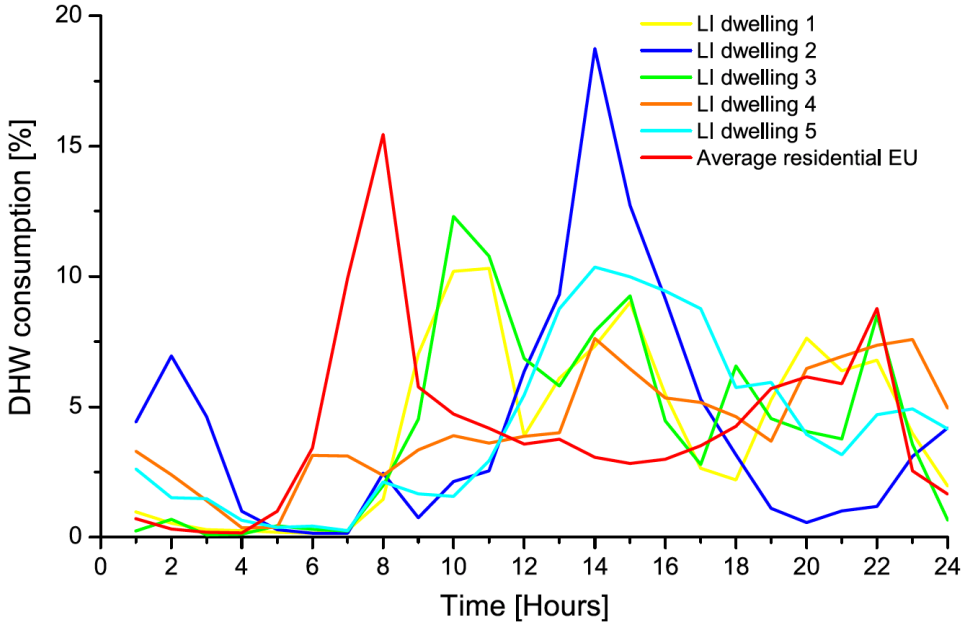


FIGURE 1.12 – Profils de consommations journalières de différents foyer à Manresa[34], comparaison avec le profil européen moyen[35], en rouge, extrait de [36].

Particularité des consommations d'ECS. Les profils de consommation d'ECS sont très variables d'un bâtiment résidentiel à l'autre. Á titre d'exemple, des profils de consommation sont représentés sur la figure 1.12. Ils représentent les consommations d'habitation situées à Manresa en Espagne, ils sont comparés au profil européen moyen. Plusieurs influences sont mises en avant par la communauté scientifique afin de définir des profils des consommations d'ECS : les variabilités journalières et saisonnières, l'impact socio-économique, la situation géographique [36]. Une baisse des consommations est observable l'été par rapport à l'hiver [36]. Plusieurs facteurs expliquent cette variation : la température de l'eau du réseau est plus élevée, il faut donc moins de temps à l'utilisateur pour obtenir la température souhaitée, une température d'air extérieur plus élevée minimise les pertes thermiques du stockage, et influence le comportement des usagers. Le jour de la semaine est également un facteur important. Ahmed et al. [37] observent une variation de 52L par jour et par habitant entre la semaine et le week-end. Cette étude a été réalisée en Finlande. George et al. [38] ont mené une étude sur 119 maisons au Canada et observent une augmentation de 12% des consommations d'ECS pendant le week-end en comparaison de la consommation moyenne sur l'ensemble de la semaine. Krippelovà et al. [39] observent un pic de consommation à 9h en semaine, alors qu'il a lieu à midi le week-end. Le comportement des habitants est le facteur principal influençant les profils

des consommations d'ECS [36]. Il dépend des habitudes culturelles, de l'éducation et du contexte socio-économique du cas considéré. Les principaux facteurs évoqués dans les études concernant les consommations d'ECS sont : l'occupation, la présence de femmes, d'adolescents ou d'enfants, le nombre de générations, l'âge, le revenu du ménage, le niveau d'éducation, la taille de la maison et le fait que la maison soit occupée par les propriétaires ou en location. Le salaire des ménages est le facteur le plus cité [40][41][42][43]. Le lien entre le revenu et la consommation d'ECS reste ambigu toutefois, en moyenne les ménages à haut revenu consomment plus que ceux à faible revenu[44][45].La présence de femmes, d'enfants ou d'adolescents peut être associée à une augmentation des consommations [40][46]. Les personnes âgées consommeraient davantage que les plus jeunes [45]. L'ensemble de ces influences complexes rendent difficile la prédiction des consommations d'ECS. C'est pourquoi dans la majorité des cas des profils utilisateurs sont utilisés. Il existe différentes méthodes afin de réaliser ces profils.

Modèles stochastiques Les méthodes stochastiques nécessitent une grande quantité de données expérimentales. Ces données peuvent être composées de mesures de consommation, de temps d'utilisation de l'ECS, de débit de puisage, mais aussi d'informations socio-économiques sur les usagers ou de caractéristiques techniques sur l'installation produisant l'ECS. Ces données sont fondamentales et définissent le domaine de validité des profils générés. Plusieurs méthodes stochastiques pour générer des profils de consommation d'ECS sont décrites dans ce paragraphe.
Une première méthode se décompose en deux étapes. Dans un premier temps, les données sont analysées afin de dégager des événements, des moments de consommation. Ces événements sont caractérisés par différents éléments selon les études. Á chaque événement est attribuée une densité de probabilité. Dans un deuxième temps, une méthode stochastique, par tirage de nombre aléatoire compose un profil de consommation, composé des événements précédemment définis .
Widen et al.[47] se basent sur des données expérimentales de temps d'utilisation sur des maisons suédoises. Ils répertorient 25 activités consommant de l'ECS. Les distributions de probabilités obtenues sont étalonnées avec des mesures de consommation et des données de temps d'utilisation, tandis que les transitions d'états de probabilité sont modélisées avec des approches tel un processus de chaînes de Markov.
Hendon et al.[48] se basent sur des données d'événements de puisages : durée, débit ainsi que la durée entre deux événements. Ils appliquent une méthode de regroupement, c'est-à-dire qu'ils posent l'hypothèse que les évènements se produisent en groupe. Ils définissent cinq catégories de puisage, en fonction de leur débit, leur durée et leur distribution de probabilité. Le tableau 1.13 synthétise ces informations. Une méthode de générateur de nombres aléatoires uniformes est utilisée pour affecter l'heure spécifique au cours de laquelle chaque groupe d'événements commence et un deuxième nombre aléatoire est utilisé pour modéliser l'heure de début des événements. Les densités de probabilité obtenues sont ajustées pour prendre en compte l'effet du jour de la semaine (semaine/week-end) et l'influence saisonnière. Le modèle se limite à des maisons équipées de deux douches, deux baignoires, quatre éviers, un lave-vaiselle et ne prend pas en compte le nombre d'habitants.

Yao et al.[49] définissent quatre catégories d'usages finaux. Ce modèle ne prend

Characteristics	Sink	Shwr	Bath	CW	DW
Average Daily Volume (gal/day, hot and cold)	25	28	7	15	5
Average Daily Events (events/day)	32.9	1.7	0.3	2.2	2.4
Annual Events (events/year)	12007	611	109	788	858
Maximum Time Between Events in Cluster (min)	15	60	60		60
Average Time Between Events in Cluster (min)	1.93	30.5			9.78
Average Events per Cluster	1.90	1.24	1.00	1.96	4.89
Number of Clusters per Year	6319	493	109	402	176
Maximum Time Between Events in Load (min)				30	
Maximum Time Between Loads in Cluster (min)				240	
Number of Loads per Cluster				1.40	
Average Number of Events per Load				1.40	
Average Time Between Events in Load (min)				4.97	
Average Time Between Loads in Cluster (min)				74.27	

Derived from AWWA 1200 house total water study
Derived from AWWA 20 house hot water study
Derived from Benchmark
Engineering Judgment

FIGURE 1.13 – Résultats des analyses de données de l'étude de [48].

pas en compte des durées d'utilisation mais des données sur les heures des premiers et derniers puisages. Les événements sont créés à l'aide d'un générateur aléatoire, chaque heure de la période d'utilisation ayant la même probabilité.
Jordan et al.[50] se basent sur les données d'une étude menée sur des maisons allemandes et suisses. Les événements sont classés selon quatre catégories : petit, moyen, douche et bain. Ces catégories sont définies en fonction de leur débit moyen, durée et nombre d'événements par heure. Le débit d'un événement dans le profil est donné par une distribution de probabilité gaussienne. L'influence de la saison

est prise en compte en appliquant une fonction sinusoïdale d'une amplitude de 10% entre février et septembre. Les profils sont comparés avec les profils moyens de la norme IEA Annexe 42, et montrent de bonnes corrélations.
Sandels et al.[51] développent un modèle de simulation stochastique qui prévoit les profils de consommation d'une habitation. Ils se basent sur l'approche des chaînes de Markov non-homogènes, où l'état des occupants varie dans une certaine limite de probabilité au cours du temps. L'état des occupants correspond à leur utilisation de l'ECS. Le modèle simule uniquement les douches et bains, mais il tient compte des pertes thermiques du ballon.

Une autre méthode stochastique consiste à considérer les événements de puisage comme des courbes rectangles, ou pulsation unique. Elle est basée sur l'hypothèse que l'événement se fait à puissance constante sur toute la durée du puisage. Chaque événement est alors caractérisé par sa durée, son intensité et sa fréquence de pulsation.
Buchberger et al.[52] appliquent cette méthode sur un an de données sur deux maisons individuelles. Cette étude met en évidence que les consommations d'ECS peuvent facilement être exprimées sous la forme d'impulsion unique. Elle montre une bonne fiabilité de la méthode utilisant la distribution cumulative, c'est-à-dire que plusieurs pulsations peuvent avoir lieu en même temps. Blokker et al.[53] développent le modèle SIMDEUM basé sur principe de pulsation unique. Les événements dépendent de l'utilisation finale et de l'utilisateur. Chaque utilisation finale a sa propre distribution de probabilité.
Creaco et al.[54] améliorent le modèle SIMDEUM en prenant en compte le lien entre l'intensité de l'événement et sa durée. Les performances de ces deux types de modèles (PRP et SIMDEUM) sont bonnes sur des prédictions à une heure. Le modèle PRP a besoin de mesures de débit pour être calibré, ce qui en fait un modèle descriptif. SIMDEUM a besoin de données sociologiques, ce qui permet de l'adapter aux cultures régionales.

Une autre méthode utilisée pour les profils de consommation d'ECS est la méthode de Monte Carlo. Negnevitsky et al.[55] développent un outil de gestion de la demande des systèmes de production d'ECS basé sur la méthode de Monte Carlo. Cet outil est adapté aux maisons individuelles, il nécessite un grand nombre d'informations : type de famille, moment de prise des douches (matin, soir, ou les deux), nombre de douches, nombre d'usages à faible volume, durée de chaque douche, durée entre deux douches consécutives, heure de début de chaque douche et chaque usage à faible volume.

La dernière méthode décrite ici est la méthode bottom-up. Fischer et al.[56] développent un modèle stochastique bottom-up de profil de consommation d'ECS basé sur un agenda des durées d'utilisation. Ce modèle prend en compte des facteurs socio-économiques.
Aki et al.[57] développent un modèle stochastique bottom-up qui estime les consommations pour chaque utilisation finale : remplissage de la baignoire, réchauffage de la baignoire, douche et cuisine. Ils se basent sur des données historiques pour extraire

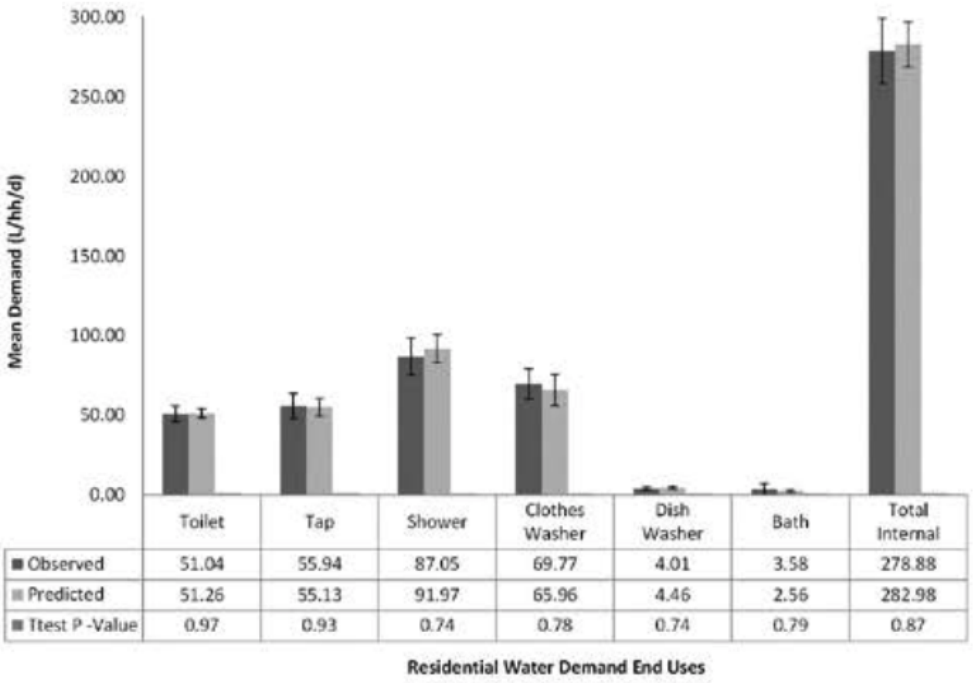
des profils de probabilité. Ferrantelli et al.[58] développent un modèle stochastique bottom-up qui prend en compte les variations entre la semaine et le week-end et l'occupation. Les données sont issues d'une étude menée à Osaka, Japon.

Modèles de prévision de séries temporelles L'étude des séries temporelles, ou séries chronologiques, correspond à l'analyse statistique d'observations régulièrement espacées dans le temps[59]. Le principe est d'utiliser un grand nombre de données représentatives pour déterminer statistiquement les paramètres d'un modèle. Les plus répandus sont les modèles de Box-Jenkins[60] qui permettent de modéliser des phénomènes temporels soumis à des perturbations aléatoires. Les phénomènes temporels peuvent être de plusieurs types, les plus communs sont : autorégressifs (AR, AutoRegressive), intégrés (I, Integrated), et moyennes mobiles (MA, Moving Average). Ces types de processus peuvent être associés dans un même modèle, formant ainsi les modèles ARMA ou ARIMA. Le modèle de Box-Jenkins est décrit par l'équation (1.11) [61] où $y(t)$ est le signal de sortie, $x(t)$ le signal d'entrée, $e(t)$ la perturbation extérieure, q l'opérateur de décalage, $A(q)$, $B(q)$,$C(q)$,$D(q)$, $F(q)$ sont des modèles polynomiaux dépendant des paramètres du modèle.

$$A(q)y(t) = \frac{B(q)}{F(q)}u(t - n_k) + \frac{C(q)}{D(q)}e(t) \tag{1.11}$$

Gelazanskas et al.[62] appliquent ces méthodes et comparent différentes techniques de prévisions : exponential smoothing (ETS), ARIMA saisonière, décomposition saisonnière de séries temporelles par le modèle de Loess (model seasonal decomposition of time series by Loess model, STL) et une combinaison de ces dernières. Ils montrent que les analyses de données temporelles pourraient être utilisées dans la gestion de système de production ECS. Les méthodes ayant donné les meilleurs résultats sont les couplages STL/ETS et STL/ARIMA.
Popescu et al. [61] appliquent les techniques d'analyse de données temporelles pour prédire la consommation thermique globale d'un quartier, ECS inclus.

Modèles d'apprentissage automatique L'apprentissage automatique est, suivant la définition de Tom Mitchell[63], l'étude des algorithmes qui permettent aux programmes de s'améliorer automatiquement par expérience. Les méthodes de réseaux de neurones artificiels, de machines à vecteur de support (Support Vector Machine, SVM), de réseaux bayésiens, les k plus proches voisins et les arbres de décisions sont des méthodes d'apprentissage automatique. Bennett et al.[40] montrent que les ANN peuvent être utilisés pour prédire la consommation d'ECS moyenne d'une habitation individuelle sur une journée avec une précision modérée, en utilisant en entrée le nombre d'adultes, le nombre d'adolescents, le nombre d'enfants, et l'efficacité du système de stockage. Les prédictions produites par le modèle de Bennett et al. sont comparées aux consommations observées sur la figure 1.14.
Bakker et al.[64] développent plusieurs modèles de type ANN qui prédisent l'énergie thermique nécessaire pour l'ECS sur une période de 24h. Les entrées du ANN sont la consommation énergétique de la semaine précédente et des informations météorologiques. Un exemple d'architecture testée est représenté en figure 1.15. Les auteurs



	Toilet	Tap	Shower	Clothes Washer	Dish Washer	Bath	Total Internal
Observed	51.04	55.94	87.05	69.77	4.01	3.58	278.88
Predicted	51.26	55.13	91.97	65.96	4.46	2.56	282.98
Ttest P -Value	0.97	0.93	0.74	0.78	0.74	0.79	0.87

FIGURE 1.14 – Comparaison des prédictions fait par un ANN par rapport aux observations, extrait de [40].

montrent que ce modèle fournit de bonnes estimations des profils ECS quotidiens, cependant des améliorations sont à apporter pour mieux prendre en compte le comportement de l'habitant, leurs résultats sont présentés en figure 1.16.

Aydinalp et al. [33] utilisent les réseaux de neurones pour prédire la consommation annuelle d'ECS d'une habitation. Ils utilisent beaucoup d'entrées, voir figure 1.17, regroupées en quatre catégories : propriétés du système de production d'ECS, type de consommation d'ECS, effets météorologiques et critères socio-économiques. Plusieurs architectures sont développées. L'évaluation des modèles sur 141 maisons présente de bonnes précisions avec un R^2 supérieur à 0,77.

Kepplinger et al.[65] développent un modèle à l'aide de la méthode des k plus proches voisins afin de prédire les consommations d'ECS pour un système d'optimisation de ballon électrique.

Autres modèles statistiques et modèle comportemental D'autres méthodes moins répandues sont utilisées pour générer des profils de consommation d'ECS. Pflugradt[66] développe un modèle de profils de puisage d'ECS basé sur un modèle psychologique qui stipule que le comportement des utilisateurs est déterminé par les désirs, ce qui conduit à mener des activités qui permettent de satisfaire ces désirs. L'évolution de l'état d'accomplissement de chaque désir est modélisée pour chaque individu occupant la maison. Lorsqu'une activité est exécutée, l'état d'exécution associé est 1 et il se désintègre après que l'activité soit terminée jusqu'à atteindre une valeur de seuil pour laquelle le désir devient pertinent. Cela conduit alors à la réalisation d'une autre activité qui répond au désir. Une limitation de ce modèle

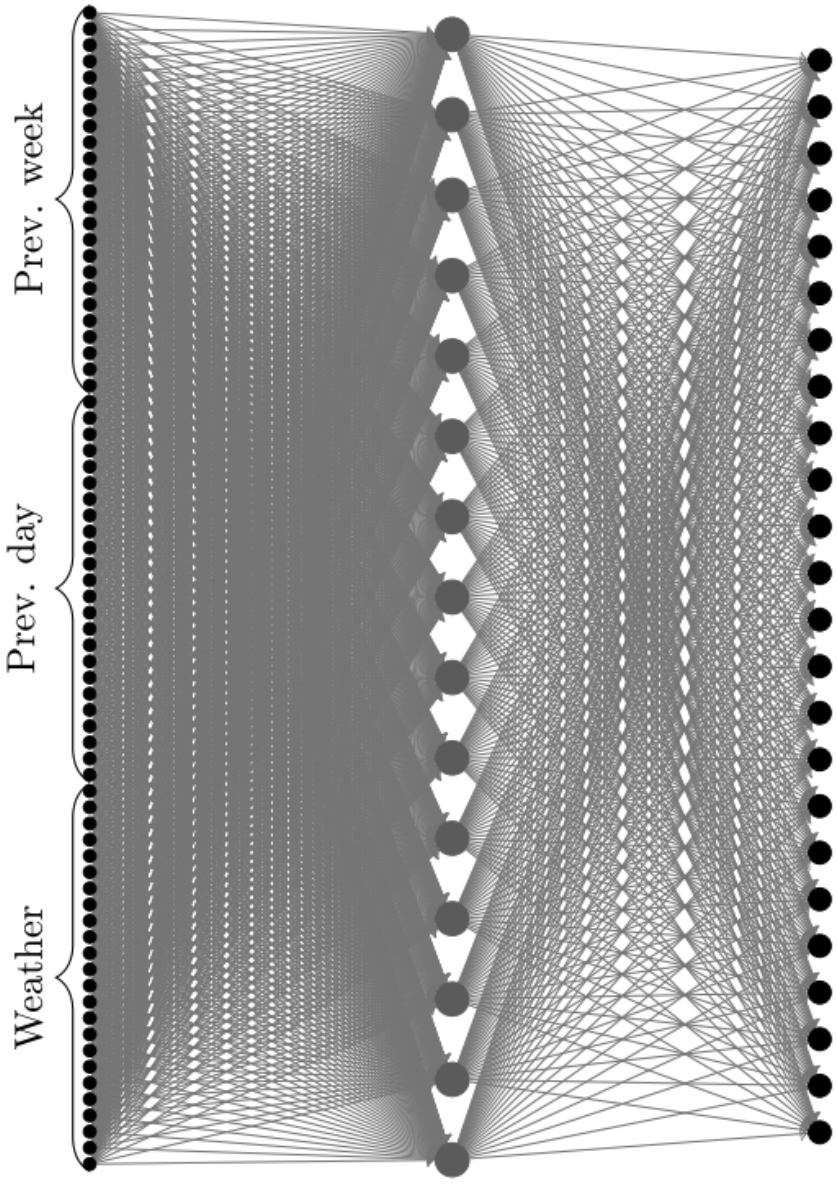


FIGURE 1.15 – Exemple de réseau de neurones artificiels pour la prédiction de consommation d'ECS, extrait de [64].

est qu'il est basé sur les désirs isolés des membres du ménage. Bien que le modèle considère la performance de certaines activités spécifiques se produisant lorsque les utilisateurs ont des désirs partagés coïncidants, il est difficile de rendre compte de toutes les interactions potentielles des utilisateurs domestiques qui déterminent leur comportement réel. Aux fins de validation, l'auteur de ce modèle a comparé l'eau chaude totale consommée par jour et par personne avec les données de consommation moyenne d'une base de données statistiques en Allemagne. Ces résultats bien qu'encourageants ne permettent pas la validation du profil de puisage.
Ladd et al.[67] développent le modèle EPRI (Electric Power Research Institute), c'est

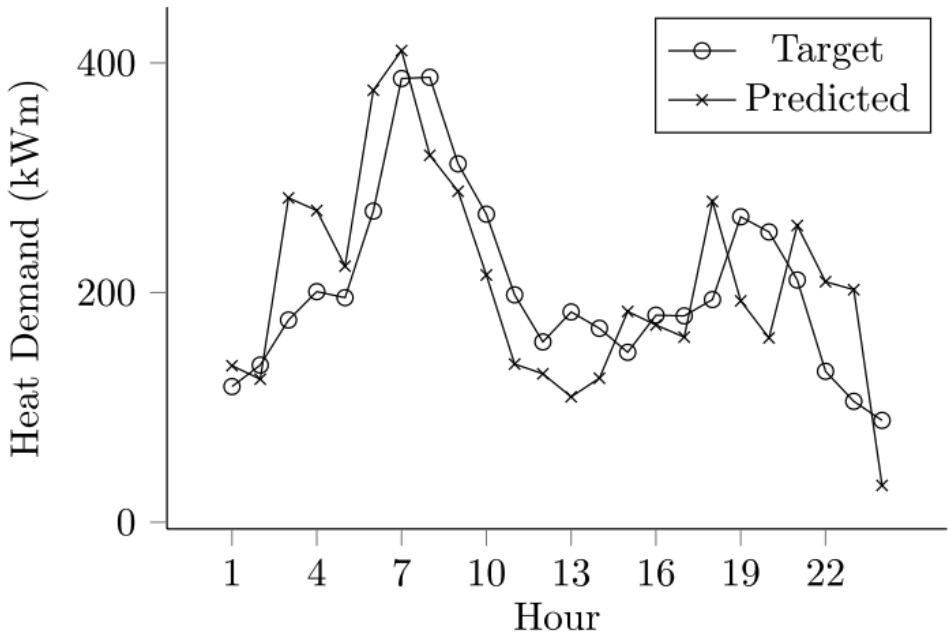


FIGURE 1.16 – Comparaison des prédictions de consommation d'ECS faite par un modèle développé par Bakker et al. et des observations, extrait de [64].

un modèle basé sur des corrélations pour les profils quotidiens de consommation d'ECS dérivés de données collectées auprès de 110 foyers américains, utilisé pour estimer la quantité horaire d'eau chaude selon le moment de la journée. Ils utilisent des méthodes statistiques, c'est-à-dire une multitude de régressions. Plutôt que de classifier les données démographiques en foyer type, ils déterminent des groupes d'âge. Les prédictions dépendent de trois types de variable : la démographie, le climat et le type de système de production d'ECS. Les variables importantes ,après l'analyse des données collectées, sont le nombre d'habitants (divisé en trois catégories : adulte, +14, enfant, entre 6 et 14 ans, et jeune enfant, moins de 6 ans), la température extérieure, la température de l'eau en entrée du système, la consigne du thermostat de l'élément le plus bas du système de chauffage, la taille du ballon, une variable indiquant s'il y a un habitant sans emploi, et quatre variables de saison indiquant l'absence ou la présence de certaines conditions. Chaque jour est divisé en huit périodes, de durées différentes, ayant une consommation ECS similaire en fonction de l'heure, du type de jour et de la saison. Les données ont été collectées avant 1980, les habitudes d'utilisation de l'ECS ayant changé, la validité de ce modèle est à remettre en question.

L'inconvénient des profils quelque soit la façon dont ils sont générés est le peu d'adaptabilité à l'usager, les informations recueillies sur l'usager servent uniquement à lui attribuer un profil. La majorité de ces méthodes demandent une grande quantité de données. Les consommations d'ECS étant très dépendante de l'usagé, la collecte de ces données peut être biaisée, par un facteur régional ou social par exemple. De plus les habitudes de consommation ont changé ces dix dernières années[36], la définition

Table 1
DHW NN model input units

	Input unit	Range
DHW heating system and equipment properties	End-use efficiency of the system	0.554–0.824
	Age of the system [years]	0.5–18
	Size of the water tanks [l]	130–280
	Number water heaters	1–2
	Shared with other dwellings	0–1
	Insulation around the water tank	0–1
	Insulation around hot water pipes	0–1
DHW consumption patterns	Number of children	0–5
	Number of adults	1–8
	Clothes washer [loads/week]	0–15
	Dishwasher [loads/week]	0–15
	Number of low-flow shower heads	0–3
	Number of aerators	0–4
Weather effects	Ground temperature [°C]	5–12
Socio-economic characteristics of the households	Income [$1,000/yr]	10–85
	Dwelling type: 1 if single-detached; 0 if single-attached	0–1
	Dwelling ownership: 1 if owner; 0 if renter	0–1
	Size of area of residence: 1 If population is less than 15,000 2 If population is between 15,000 and 100,000 3 If population is 100,000 or over	1–3

FIGURE 1.17 – Présentation des entrées utilisées pour la prédiction de consommation annuelle d'ECS, extrait de [33].

de profils à partir de données sans actualisation ou adaptation précise à l'usager pose un probléme de vieillissement des modèles.

Conclusion. Les réseaux de neurones artificiels ont montré leur capacité à prédire des événements temporels dans d'autres applications que les consommations d'ECS : une énergie solaire [68], une énergie éolienne [69], des paramètres économiques [70] ou encore des dispersions de poudre [71]. Ils commencent à être utilisés pour l'ECS et semblent pertinents. Leur capacité à apprendre à partir uniquement des données collectées auprès de l'utilisateur dont nous voulons prédire les consommations est très intéressante. Cependant, les cas de validation expérimentale ne sont pas assez nombreux dans la littérature pour montrer leur efficacité. Plusieurs points clés restent à démontrer : quelles architectures sont les plus efficaces pour la prédiction de consommation d'ECS, quels types de données faut-il utiliser et quelle quantité permet les meilleurs résultats ? Afin de répondre à la problématique de prédiction de consommation d'ECS dans le cadre d'une régulation prédictive, nous développons des ANN. Nous testerons plusieurs architectures, types de données, et quantités de données.

1.4 Réseau de neurones artificiels : fondamentaux

Les deux parties précédentes 1.1 et 1.3 ont montré l'intérêt d'utiliser des réseaux de neurones artificiels (ANN) pour la prédiction de température intérieure du bâtiment et de consommation d'ECS. Cette partie présente les fondamentaux nécessaires à la compréhension des travaux développés dans les chapitres suivants. Les ANN sont développés par biomimétisme, c'est-à-dire qu'ils cherchent à copier les réseaux de neurones biologiques et principalement leurs capacités de mémoire et d'apprentissage. La méthode des ANN aussi appelée réseaux de neurones formels est décrite pour la première fois en 1943 par W.McCulloch et W.Pitts, composés d'entités, représentant les neurones, reliées entre elles par des liaisons copiant les synapses. W.McCulloch et W.Pitts montrent qu'un ANN composé de trois couches est universel, c'est-à-dire capable de modéliser toutes les fonctions logiques. Mais la question reste comment adapter les paramètres de l'ANN pour chacune des fonctions logiques. Sont alors imaginés trois types d'apprentissages : supervisé, par renforcement et non supervisé. L'idée de l'apprentissage supervisé est que les observations d'apprentissages sont associées à une étiquette, une réponse, les poids sont alors adaptés pour qu'une observable soit associée à la bonne étiquette. Dans l'apprentissage par renforcement, l'idée est de fournir une "récompense" lorsque la réponse à une observable est correcte. L'apprentissage non supervisé, le plus complexe, se fait à partir uniquement des observables, l'apprentissage consiste alors à trouver des points communs aux différentes observables afin de créer des catégories. F.Rosenblatt développe en 1958 le perceptron, ANN composé de deux couches dont les poids sont déterminés par apprentissage supervisé. Cette partie décrit les principes fondamentaux du perceptron. Dans un premier temps, nous décrivons l'architecture, puis l'apprentissage, suivi du formatage des données et enfin quelques considérations concernant le développement de ce type de modèles.

1.4.1 Perceptron multicouches

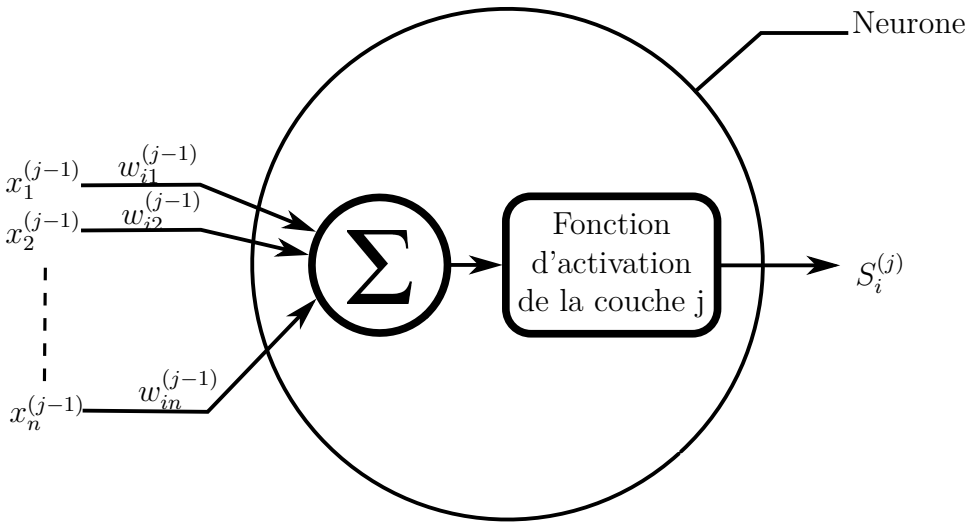


FIGURE 1.18 – Représentation du i-éme neurone artificiel de la couche j.

Architecture des ANN Le perceptron multicouche est un type de réseaux de neurones artificiels(ANN). Un ANN est composé de neurones organisés en couches. Les neurones sont reliés entre eux par des liaisons. Un neurone est représenté par un cercle, une liaison par un segment entre deux cercles et une couche de neurones par un alignement vertical de cercle. Chaque neurone i d'une couche j accepte n entrée $x_k|k = 1, 2...n$. À chaque liaison est associé un poids, le poids de la liaison entre le neurone k de la couche $j-1$ et le neurone i de la couche j est nommé $w_{ik}^{(j-1)}$. Nous notons $w^{(j)}$ la matrice composée de l'ensemble des poids des liaisons entre la couche j et la couche $j+1$, cette matrice compte autant de lignes qu'il y a de neurones sur la couche $j+1$ et autant de colonnes qu'il y a de neurones sur la couche j. La sortie S_i^j du neurone i de la couche j est calculée selon l'équation (1.12) c'est-à-dire en appliquant une fonction d'activation f à la somme pondérée des entrées. Le neurone artificiel i de la couche j est schématisé sur la figure 1.18.

$$S_i^{(j)} = f\left(\sum_{k=1}^{n} x_k^{(j-1)}.w_{ik}^{(j-1)}\right) \ . \tag{1.12}$$

La particularité des perceptrons multicouches est que l'information se propage uniquement vers l'avant. Chaque neurone d'une couche reçoit en entrée toutes les sorties des neurones de la couche précédente. Il faut distinguer deux types de couches neuronales, la couche de sortie dont chaque neurone fourni une sortie du ANN, et les couches intermédiaires situées entre les entrées et la couche de sortie, appelées couches cachées. Nous ne parlerons pas de couche d'entrées mais simplement d'entrées puisque ni somme, ni fonction d'activation ne sont calculées à cette étape. Chaque couche est composée d'un nombre de neurones qui lui est propre. Tous les neurones d'une couche ont la même fonction d'activation. Un ANN est défini par son architecture c'est-à-dire le nombre d'entrées, sorties, couches intermédiaires, ainsi que le nombre de neurones par couche.

Expression mathématique. A titre d'exemple, la figure 1.19 représente un perceptron multicouche à trois entrées x_1, x_2x_3 et une sortie $\widehat{S}$, la couche de sortie est donc constituée d'un neurone. L'ANN compte deux couches cachées, la première est constituée de trois neurones, la seconde de cinq. La matrice des poids entre la couche 1 et les entrées est nommée $w^{(0)}$, entre les couches 1 et 2, $w^{(1)}$, de même entre la couche 2 et la couche de sortie, $w^{(2)}$. Les couches 1,2 et de sortie ont respectivement pour fonction d'activation les fonctions quelconques f, g et h. En pratique, les fonctions d'activation utilisées sont variées, fonction identité, linéaire, sigmoïde, marche,etc. Les équations (1.13) permettent de calculer les sorties de tous les neurones de l'exemple présenté en figure 1.19. Ces équations sont présentées sous forme matricielle dans les équations (1.14).

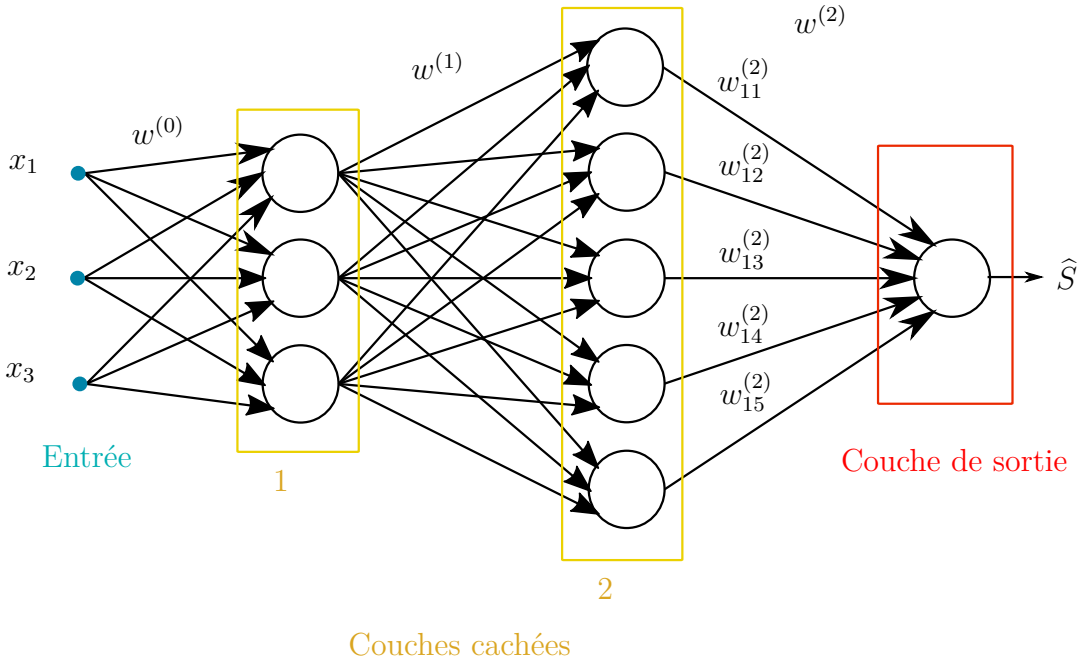


FIGURE 1.19 – Représentation d'un perceptron multicouche

Sorties des neurones de la couche 1 :

$$S_1^{(1)} = f(w_{11}^{(0)}.x_1 + w_{12}^{(0)}.x_2 + w_{13}^{(0)}.x_3) \tag{1.13a}$$
$$S_2^{(1)} = f(w_{21}^{(0)}.x_1 + w_{22}^{(0)}.x_2 + w_{23}^{(0)}.x_3) \tag{1.13b}$$
$$S_3^{(1)} = f(w_{31}^{(0)}.x_1 + w_{32}^{(0)}.x_2 + w_{33}^{(0)}.x_3) \tag{1.13c}$$
$$\tag{1.13d}$$

Sorties des neurones de la couche 2 :

$$S_1^{(2)} = g(w_{11}^{(1)}.S_1^{(1)} + w_{12}^{(1)}.S_2^{(1)} + w_{13}^{(1)}.S_3^{(1)}) \tag{1.13e}$$
$$S_2^{(2)} = g(w_{21}^{(1)}.S_1^{(1)} + w_{22}^{(1)}.S_2^{(1)} + w_{23}^{(1)}.S_3^{(1)}) \tag{1.13f}$$
$$S_3^{(2)} = g(w_{31}^{(1)}.S_1^{(1)} + w_{32}^{(1)}.S_2^{(1)} + w_{33}^{(1)}.S_3^{(1)}) \tag{1.13g}$$
$$S_4^{(2)} = g(w_{41}^{(1)}.S_1^{(1)} + w_{42}^{(1)}.S_2^{(1)} + w_{43}^{(1)}.S_3^{(1)}) \tag{1.13h}$$
$$S_5^{(2)} = g(w_{51}^{(1)}.S_1^{(1)} + w_{52}^{(1)}.S_2^{(1)} + w_{53}^{(1)}.S_3^{(1)}) \tag{1.13i}$$
$$\tag{1.13j}$$

Sortie de la couche de sortie :

$$\widehat{S} = h(w_{11}^{(2)}.S_1^{(2)} + w_{12}^{(2)}.S_2^{(2)} + w_{13}^{(2)}.S_3^{(2)} + w_{14}^{(2)}.S_4^{(2)} + w_{15}^{(2)}.S_5^{(2)}) \tag{1.13k}$$

Sortie des neurones de la couche 1 :

$$S^{(1)} = f(w^{(1)} \cdot x) \tag{1.14a}$$
$$\tag{1.14b}$$

Sortie des neurones de la couche 2 :

$$S^{(2)} = g(w^{(2)} \cdot S^{(1)}) \tag{1.14c}$$
$$\tag{1.14d}$$

Sortie de la couche de sortie :

$$\widehat{S} = h(w^{(2)} \cdot S^{(2)}) \tag{1.14e}$$
$$Soit \widehat{S} = h(w^{(2)} \cdot g(w^{(2)} \cdot f(w^{(1)} \cdot x))) \tag{1.14f}$$
$$\tag{1.14g}$$

1.4.2 Apprentissage supervisé par rétro-propagation de l'erreur

L'apprentissage L'apprentissage détermine les valeurs des poids $w^{(j)}$ affectées à l'ensemble des liaisons du ANN pour un jeu d'apprentissage donné. Pour cela, nous utilisons un algorithme d'optimisation. Dans nos travaux, nous utilisons l'algorithme de Levenberg-Marquardt utilisé pour la première fois dans [72]. Le problème à optimiser est mis sous la forme d'un problème des moindres carrés. Le jeu d'apprentissage est composé de m couples entrée,sortie $[x^{(i)}, S^{(i)}]$ appelés exemples d'apprentissage. La fonction coût est la différence des moindres carrés entre la sortie calculé par le ANN et la sortie souhaitée, moyennée sur l'ensemble du jeu d'apprentissage. La fonction coût, $J(w)$, est donnée par l'équation 1.15.

$$J(w) = \frac{1}{m}\left(\sum_{i=1}^{m}(\widehat{S}(x^{(i)}) - S^{(i)})^2\right) \tag{1.15}$$

1.4.3 Formatage des données d'entrée

Normalisation. Lorsque les grandeurs en entrée du réseau de neurones ne sont pas du même ordre de grandeur, il est nécessaire de les normaliser dans le but de faciliter la convergence de l'algorithme d'apprentissage. Cette normalisation est faite à l'aide de l'équation (1.16). L'équation (1.16) est donnée pour une normalisation entre 0 et 1 de la grandeur x_1, où $x_{1,min}$ et $x_{1,max}$ sont respectivement le minimum et le maximum que peut prendre la valeur x_1.

$$x_{1,norm} = \frac{x_1 - x_{1,min}}{x_{1,max} - x_{1,min}} \tag{1.16}$$

1.4.4 Développement de réseaux de neurones artificiels

Jeux de données. Afin de développer et de tester un réseau de neurones les données disponibles sont séparées en trois jeux de données, le jeu d'apprentissage, le jeu de validation et le jeu de test. Le jeu d'apprentissage permet de réaliser l'apprentissage, c'est-à-dire de déterminer les poids optimum pour une architecture de ANN. Le jeu de validation permet de déterminer l'architecture du ANN et le jeu de test d'en mesurer les performances.

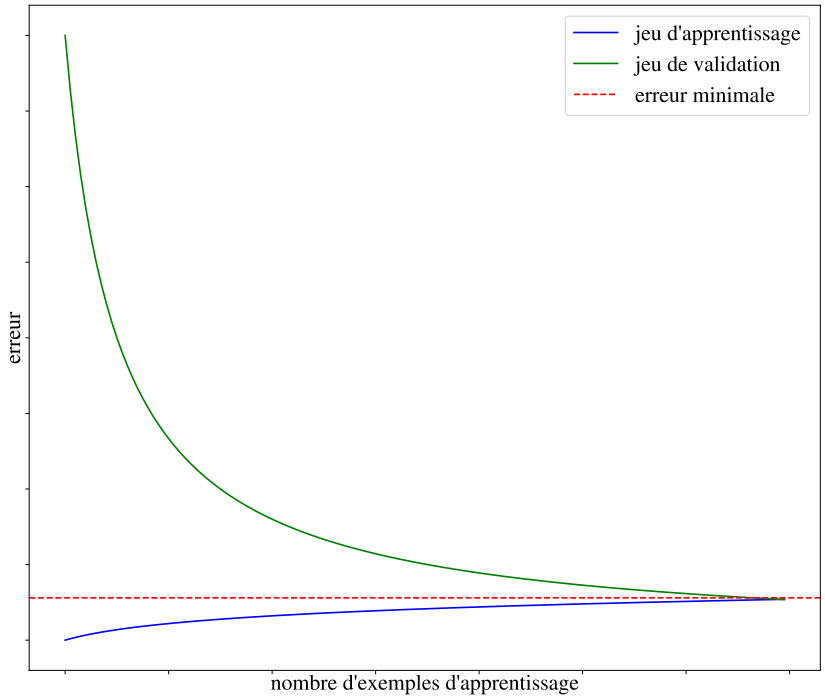


FIGURE 1.20 – Allure théorique des courbes d'apprentissage.

Courbes d'apprentissage. Les courbes d'apprentissage permettent d'évaluer la capacité d'un modèle de réseau de neurones à prédire un phénomène. La figure 1.20 représente l'allure théorique de ces courbes. Il s'agit de représenter l'erreur de prédiction du réseau de neurones en fonction du nombre d'exemples d'apprentissage proposés. Nous distinguons l'erreur faite sur le jeu d'apprentissage, en bleu, et sur le jeu de validation, en vert. Ces deux erreurs $RMSE_{app}$ et $RMSE_{val}$ sont calculées à l'aide de l'équation (1.17). Il est à noter que les deux courbes convergent vers une même valeur, l'erreur minimale que nous pourrons obtenir avec cet architecture de ANN. Si cette valeur est trop élevée pour le problème modélisé, il faudra alors modifier l'architecture du ANN. Nous observons également que les deux courbes évoluent de manière opposée : lorsque la quantité de données d'apprentissage augmente, l'erreur faite sur le jeu d'apprentissage augmente alors que celle sur le jeu de validation diminue.

Avec peu de données d'apprentissage, l'algorithme d'apprentissage va facilement trouver une solution très proche des valeurs proposées. Cependant lorsque le ANN rencontrera des valeurs qui n'étaient pas présentes dans le jeu d'apprentissage, il sera incapable de généraliser. Nous parlons alors de sur-apprentissage. Le sur-apprentissage peut également être expliqué par une architecture trop complexe. Il est possible de rencontrer le problème opposé, un sous-apprentissage, c'est-à-dire que l'erreur sur le jeu d'apprentissage est très élevée. Il faut alors questionner la convergence de l'algorithme d'apprentissage, la quantité de données dans le jeu d'apprentissage, une architecture trop simple ou la corrélation entre les grandeurs d'entrée et de sortie. Les courbes d'apprentissage permettent également de déterminer un nombre minimum de données d'apprentissage pour une erreur minimale déterminée. Il est à noter que la taille du jeu d'apprentissage nécessaire est dépendant de l'architecture du ANN.

$$RMSE = \frac{1}{m}\sum_{i=1}^{m}\sqrt{(\widehat{S}(x_i) - S(x_i))^2} \tag{1.17}$$

1.5 Conclusion

Ce chapitre présente l'état de l'art en terme de régulation prédictive des systèmes de production de chauffage et d'ECS pour l'habitat individuel. Dans un premier temps, nous montrons que ces méthodes permettraient de réduire les consommations d'énergie de la chaudière thermodynamique BoostHeat, donc d'en améliorer l'efficacité. Nous mettons en avant le besoin de modèle prédictifs fiables. Nous montrons pour la prédiction de température intérieure et de besoin d'ECS que la méthode des réseaux de neurones artificiels est la plus pertinente dans notre cas. En effet, son caractère intrinsèquement adaptatif permet de développer des modèles spécifiques à un bâtiment à l'aide uniquement de données mesurées sur site. Nous présentons les principes thèorique liés à la méthode des réseaux de neurones artificiels nécessaire à la compréhension de nos travaux. Nous nous attacherons dans la suite de ce manuscrit à réduire au maximum les données nécessaires au développement et à l'utilisation des ANN, dans le but de minimiser l'instrumentation du bâtiment. Afin de réduire la complexité de la prédiction, induit par cette limitation des choix d'entrée nous réalisons une prédiction à court terme. Nous posons l'hypothèse que l'architecture du ANN représente les phénomènes physiques liant les variables d'entrée aux variables de sortie. Les données d'apprentissage, donc les poids du ANN représentent les spécificités de chaque bâtiment. Dans le cas de la prédiction de température intérieure du bâtiment, nous avons présenté des études montrant les performances de cette méthode. Les études citées proposent une architecture de ANN pour un bâtiment. Les objectifs de ces travaux sont de développer une architecture unique pour tous les bâtiments et de tester son caractère adaptatif. Dans le cas de la prédiction de besoin d'ECS, les grandeurs à prendre en compte en entrées des modèles ANN sont encore mal connues, ainsi que la quantité de données nécessaire à une prédiction fiable. Les objectifs de ces travaux sont de développer une architecture unique, en mettant en avant l'influence du choix des entrées des exemples d'apprentissages. Nous testons également le caractère adaptatif du modèle.

CHAPITRE 2

Développement et évaluation des modèles : cas d'études utilisés

2.1 Introduction

Nous utilisons des réseaux de neurones artificiels comme modèles de prédictions de consommation d'eau chaude sanitaire(ECS) et de prédictions de température intérieure du bâtiment. Comme cela est présenté dans la partie 1.4, ces modèles nécessitent des données. Nous utilisons deux types de données, des données expérimentales dont l'origine et les techniques utilisées pour les recueillir sont présentées dans la première partie de ce chapitre (2.2), et des données obtenues par simulation dont le détail des modèles et des paramètres de simulation est présenté dans la deuxième partie de ce chapitre (2.3).

2.2 Cas expérimentaux

Généralité. Les données expérimentales que nous utilisons proviennent d'un projet dont l'objectif est la mesure des performances de Systèmes Solaires Combinés(SSC). Un SSC est une installation qui fournit de l'énergie thermique pour le chauffage et l'eau chaude sanitaire couplée à une autre source d'énergie. L'autre source d'énergie peut être de toute nature. Les SSC permettent de fournir de l'énergie quelle que soit la ressource solaire. Des habitations individuelles situés en France ont été équipées de SSC et ont été entièrement instrumentées. Ces instruments permettent de mesurer l'ensemble de leur production et de leur consommation d'énergie. La campagne de mesures s'est déroulée au cours des années 2004 et 2005 sur 20 habitations. Parmi les mesures récoltées, nous nous intéressons à celles permettant de calculer les consommations d'ECS et de chauffage, ainsi qu'aux mesures météorologiques et de confort thermique. Les mesures sont réalisées toutes les huit minutes. Les températures sont mesurées à l'aide de thermomètre à résistance de platine (PT100). Ceux-ci sont placés dans des doigts de gant lorsqu'il s'agit de mesurer des températures d'eau. Le débit d'ECS est mesuré à l'aide d'un débitmètre à ultrason, le débit du circuit de

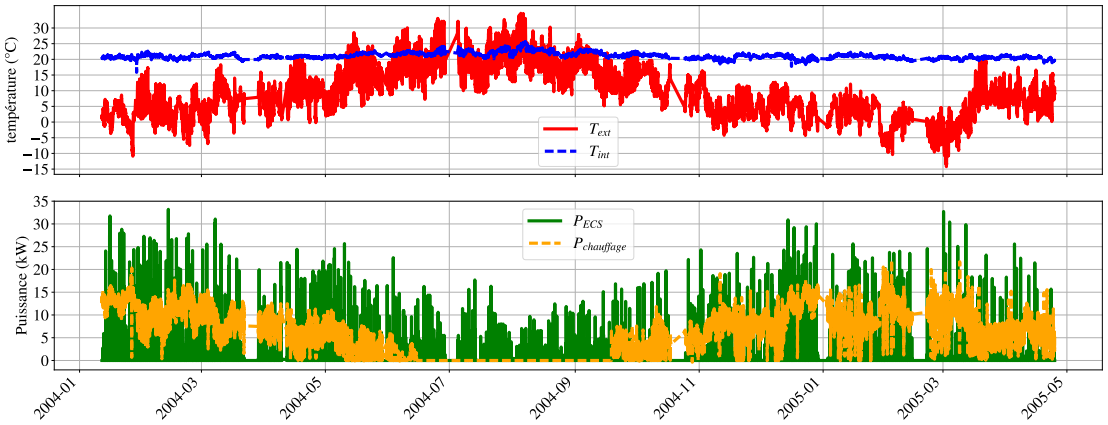


FIGURE 2.1 – Exemple de données expérimentales traitées. En vert est représenté la puissance d'ECS, en jaune la puissance de chauffage, en bleu la température intérieure,en rouge la température extérieure.

TABLE 2.1 – Informations sur les mesures expérimentales

Grandeur physique	Symbole	Type de capteur	Intervalle entre deux mesures
Température intérieure	T_{int}	thermomètre à résistance de platine	8 minutes
Température extérieure	T_{ext}	thermomètre à résistance de platine	8 minutes
Température de départ de l'eau dans le circuit ECS	T_{ECS}^{dep}	thermomètre à résistance de platine	8 minutes
Température de retour de l'eau dans le circuit ECS	T_{ECS}^{ret}	thermomètre à résistance de platine	8 minutes
Débit volumique de puissage de l'ECS	$\dot{V}_{ECS}$	débitmètre à ultrason	8 minutes
Température de départ de l'eau dans le circuit de chauffage	T_{chauf}^{dep}	thermomètre à résistance de platine	8 minutes
Température de retour de l'eau dans le circuit de chauffage	T_{chauf}^{ret}	thermomètre à résistance de platine	8 minutes
Débit volumique au départ du circuit de chauffage	$\dot{V}_{chauf}$	débitmètre	8 minutes

chauffage est mesuré avec un débitmètre classique, car il nécessite une mesure moins précise. Le tableau 2.1 synthétise les données utilisées. Les données ont été collectées automatiquement chaque jour sous la forme de fichier binaire. Un exécutable traite ces fichiers binaires pour en faire des fichiers texte, comme montré sur la figure 2.2. Dans ce fichier, chaque colonne correspond à un capteur, chaque ligne à un instant de mesure. Sur la figure 2.1 sont représentées des données expérimentales après traitement.

exemple_donnee_brute.txt

```
36399 40240 38481 44771 32568 28693 32976 33039 25545 41126 31197 4607 0 127 84 0 0 256 16040
36408 40037 38321 44110 32579 28727 32929 33043 25566 41088 31095 4608 0 129 88 0 0 256 16048
36435 39898 38632 44306 32578 28754 32922 33043 25526 41061 31029 4606 0 129 78 0 0 256 16056
36389 40087 38641 44763 32572 28778 32948 33045 25455 41047 30993 4605 0 130 84 0 0 256 16064
36456 39916 38504 43932 32558 28804 32952 33046 25440 41037 30969 4608 0 129 90 0 0 256 16072
36474 39950 38554 44818 32535 28828 32950 33048 25397 41020 30943 4609 0 130 80 0 0 256 16080
36498 40007 38632 44349 32509 28854 32957 33051 25469 41006 30905 4608 0 130 87 0 0 256 16088
36511 39893 38644 44040 32480 28876 32960 33054 25508 40991 30871 4609 0 130 92 0 0 256 16096
36521 40042 38575 44955 32454 28892 32972 33054 25519 40977 30859 4606 0 130 77 0 0 256 16104
36532 39874 38727 44326 32444 28907 32985 33056 25488 40974 30878 4613 0 130 85 0 0 256 16112
36537 40262 38437 44241 32445 28923 32993 33056 25442 40980 30909 4611 0 130 90 0 0 256 16120
36559 39897 38635 44605 32442 28933 33005 33056 25413 40978 30936 4609 0 131 85 0 0 256 16128
36567 40203 38608 43949 32429 28934 33012 33055 25414 40976 30969 4609 0 130 93 0 0 256 16136
36571 40212 38640 44907 32422 28949 32956 33055 25333 40964 30974 4606 0 130 77 0 0 256 16144
36588 40077 38722 44228 32421 28964 33026 33056 25315 40950 30975 4608 0 129 91 0 0 256 16152
36594 40157 38532 44134 32427 28973 33031 33057 25214 40937 30987 4608 0 130 86 0 0 256 16160
36606 40319 38706 44840 32431 28978 33034 33057 25261 40915 30964 4609 0 129 85 0 0 256 16168
36621 40245 38660 44229 32436 28974 33041 33059 25318 40886 30929 4609 0 129 88 0 0 256 16176
36622 40176 38574 44535 32437 28963 33036 33060 25334 40857 30884 4649 0 128 80 0 0 256 16184
36636 40118 38740 44382 32440 28950 33046 33061 25246 40844 30879 4609 0 129 91 0 0 256 16192
36646 39953 38591 44167 32447 28937 33057 33062 25264 40845 30891 4608 0 129 85 0 0 256 16200
36644 40299 38822 44979 32446 28940 33055 33064 25219 40823 30866 4605 0 128 81 0 0 256 16208
36661 40330 38796 44154 32436 28935 33034 33067 25280 40792 30825 4608 0 128 93 0 0 256 16216
36668 39809 38746 44575 32422 28929 33048 33069 25210 40775 30813 4610 0 128 76 0 0 256 16224
36664 40139 38816 44907 32407 28926 33047 33070 25242 40762 30796 4608 0 128 81 0 0 256 16232
36681 40119 38776 43967 32402 28926 33065 33071 25213 40748 30791 4610 0 129 95 0 0 256 16240
36700 40391 38727 44704 32406 28927 33068 33071 25165 40735 30619 4608 0 129 85 0 0 256 16248
36701 40293 38596 44367 32412 28928 33076 33070 25231 40727 30789 4610 0 129 87 0 0 256 16256
36710 40001 38763 44357 32409 28930 33079 33069 25268 40719 30789 4607 0 129 83 0 0 256 16264
36706 40311 38689 44962 32337 28927 33069 33068 25326 40688 30754 4610 0 129 80 0 0 256 16272
36712 40311 38842 44051 32404 28923 33069 33069 25251 40678 30761 4610 0 128 94 0 0 256 16280
```

FIGURE 2.2 – Exemple de données expérimentales brutes.

Pour des raisons de continuité dans les données nous avons séléctionné trois cas de maisons individuelles décrites ci-dessous :

FIGURE 2.3 – Cas 1 : Photographie panoramique de la maison et des panneaux solaires.

Cas 1. Le premier cas est une maison de $200m^2$ située à La Blaquerie, en France. Elle est habitée par 4 personnes, dont les besoins annuels en ECS sont de $1201kWh/an$. Le bâti est ancien avec des murs très épais et équipé d'un plancher chauffant. La maison est estimée dans la classe énergétique D avec une consommation annuelle pour le chauffage de $187kWh/m^2/an$. Les mesures sont réalisées de janvier 2004 jusqu'à avril 2005. Pour des raisons techniques, nous ne disposons pas de l'intégralité des mesures sur la période de la campagne, mais de 74 669 mesures, soit plus de 59 semaines de données.

FIGURE 2.4 – Cas 2 : Photographie de la maison et des panneaux solaires.

Cas 2. Le deuxième cas est une maison de $126m^2$ située à Saint-Pierre-de-Rivière, en France. Elle est habitée par 4 personnes, un couple et deux enfants. Leurs besoins annuels en ECS est de $2172kWh/an$. Le bâti est récent avec une inertie moyenne et équipé d'un plancher chauffant. La maison est estimée dans la classe énergétique B avec une consommation annuelle de $32kWh/m^2/an$. Les mesures sont faites de janvier 2004 jusqu'à mai 2005. Pour des raisons techniques, nous ne disposons pas de l'intégralité des mesures sur la période de la campagne, mais de 75 052 mesures, soit plus de 59 semaines de données.

FIGURE 2.5 – Cas 3 : Photographie panoramique de la maison.

Cas 3. Le troisième cas est une maison de $120m^2$ située dans un hameau de Millau, sur le plateau du Larzac, en France. Le gîte attenant à la maison n'est pas fourni en énergie par le SSC, il n'est donc pas instrumenté et ses consommations ne sont pas prises en compte. La maison est habitée par un couple de retraités, dont les besoins annuels en ECS sont de $1985kWh/an$. Cette consommation est élevée pour une

habitation de cette taille occupée par deux personnes. Cette consommation anormale est expliquée par la particularité de l'installation ECS. Dans le but d'économiser de l'eau, les propriétaires ont mis en place une boucle de recirculation. L'eau est maintenue chaude dans l'ensemble des circuits d'eau de la maison en permanence. Cela évite de jeter l'eau qui a refroidi dans les tuyaux entre deux utilisations. Le cas 3 sera utilisé uniquement pour la prédiction de consommation d'ECS, car les mesures de températures intérieures recueillies sont inutilisables. Les mesures sont réalisées de février 2004 à avril 2005. Pour des raisons techniques, nous ne disposons pas de l'intégralité des mesures sur la période de la campagne, mais de 71 939 mesures, soit plus de 57 semaines de données.

Ces informations sont synthétisées dans le tableau 2.2.

TABLE 2.2 – Description des cas d'études

	Cas 1	Cas 2	Cas 3
Surface habitable	$200m^2$	$126m^2$	$120m^2$
Nombre d'habitants	4	4	2
Besoin en ECS	$1021kWh$	$2172kWh$	$1985kWh$
Besoin en chauffage	$187kWh/m^2/an$	$32kWh/m^2/an$	
Caractéristique de l'enveloppe	ancienne, mur épais	récente	
Type d'émetteur	plancher chauffant	plancher chauffant	
Période de mesure	janvier 2004 à avril 2005	janvier 2004 à mai 2005	février 2004 à avril 2005
Quantité de mesures	74669	75052	71939

ECS : représentativité des cas d'étude. Une étude menée par l'ADEME [73] a estimé les besoins d'ECS à l'échelle d'une maison ou d'un appartement à partir de plus de 400 suivis instrumentés et 14 250 relevés annuels de compteurs en France. Cette étude indique les besoins d'ECS en volume d'eau consommée à 40°C. Ce choix est fait car la température de 40°C est proche de la température d'usage et permet d'obtenir des valeurs de besoins pratiquement indépendantes de la température d'eau froide du site. Afin de comparer nos cas d'études aux statistiques françaises, nous calculons le volume V_j d'eau consommée à 40°C par jour et par personne équivalent à la consommation annuelle des trois cas. Nous calculons le volume d'eau froide que l'énergie consommée en une année par chaque cas permet de chauffer à 40°C. Nous considérons que la température moyenne annuelle de l'eau froide est de 16°C. Enfin, nous divisons ce volume annuel par 365 (nombre de jours par an) et par le nombre de personnes consommant cette eau. L'ensemble de cette opération est indiquée dans la formule (2.1).

$$V_j = \frac{E}{\rho_{eau}.\Delta T.Cp.365.pers} \tag{2.1}$$

E désigne la consommation annuelle,ρ_{eau} désigne la masse volumique de l'eau, Cp la capacité thermique de l'eau et $pers$ le nombre de personnes habitant le cas étudié.

La figure 2.6 représente la répartition des besoins journaliers moyens par personne. Il est à noter que deux tiers des logements étudiés ont une consommation comprise

entre 35L et 80L par jour et par personne, la valeur moyenne étant de 56L par jour et par personne. Les consommations des trois cas d'études sont indiquées sur la figure 2.6. Nous remarquons que le cas 2 est représentatif d'une consommation moyenne, alors que le cas 1 représente une faible consommation et le cas 3 une forte consommation.

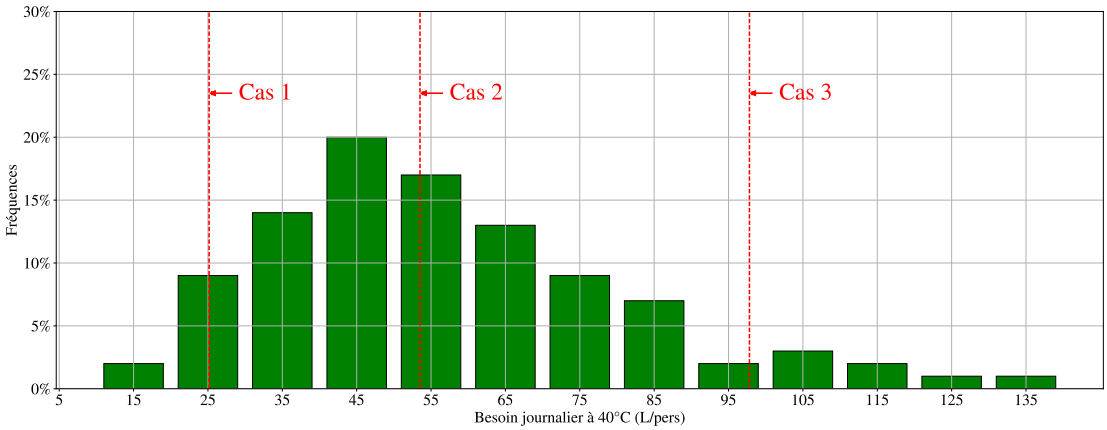


FIGURE 2.6 – ECS : Positionnement des cas d'études par rapport aux statitiques de consommation française.

Chauffage : représentativité des cas d'étude. La figure 2.7 représente la répartition des logements individuels construits avant 2013 selon leur étiquette énergétique. Ces résultats sont issus d'une étude menée par l'ADEME[74] sur 1 747 727 diagnostiques de performance énergétique. Le cas 1, d'étiquette énergétique C, est à la fois moyen et représentatif de la classe énergétique majoritaire. Le cas 2, d'étiquette énergétique A, est lui représentatif d'habitation peu énergivore.

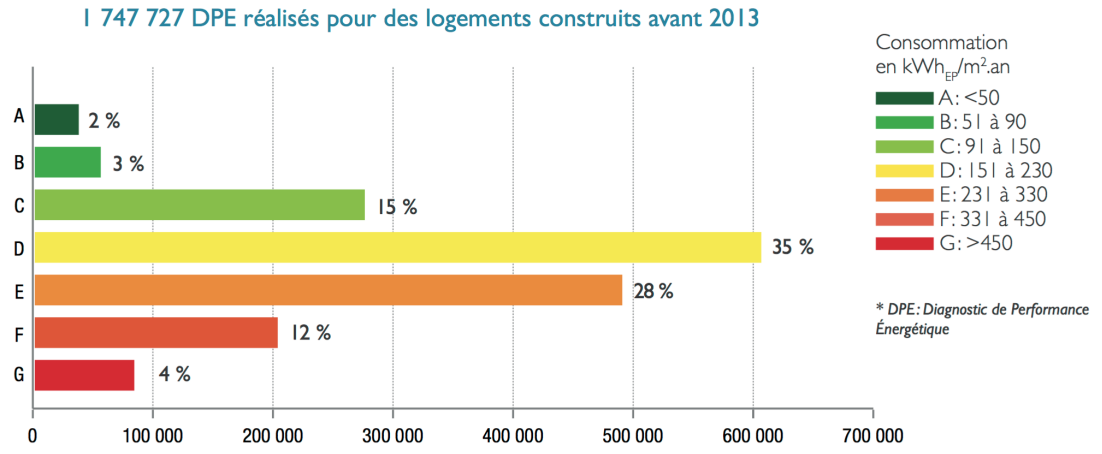


FIGURE 2.7 – Répartition des logements individuels français par classe énergétique, extrait de [74].

L'ensemble des cas d'étude présentés ici comporte la même particularité : la volonté des habitants de tester de nouveaux systèmes solaires combinés. Cela dénote

une sensibilité à leur impact environnemental et à leur consommation d'énergie et donc potentiellement une utilisation raisonnée ou économe des ressources énergétiques.

2.3 Cas simulés

Objectif de la modélisation de cas. Nous cherchons à montrer que nos modèles peuvent s'adapter à l'ensemble des habitations individuelles et à l'ensemble des climats français. Puisque nous ne disposons pas de mesures expérimentales représentatives de l'ensemble de ces cas, nous modélisons des cas complémentaires. Les consommations d'ECS étant fortement liées à l'usager et variable dans le temps, la création de données de consommation d'ECS par simulation n'a pas été retenue. Les modèles décrits dans cette partie sont utilisés afin de générer des données de consommations énergétiques pour le chauffage et de variation de température intérieure de l'habitat.

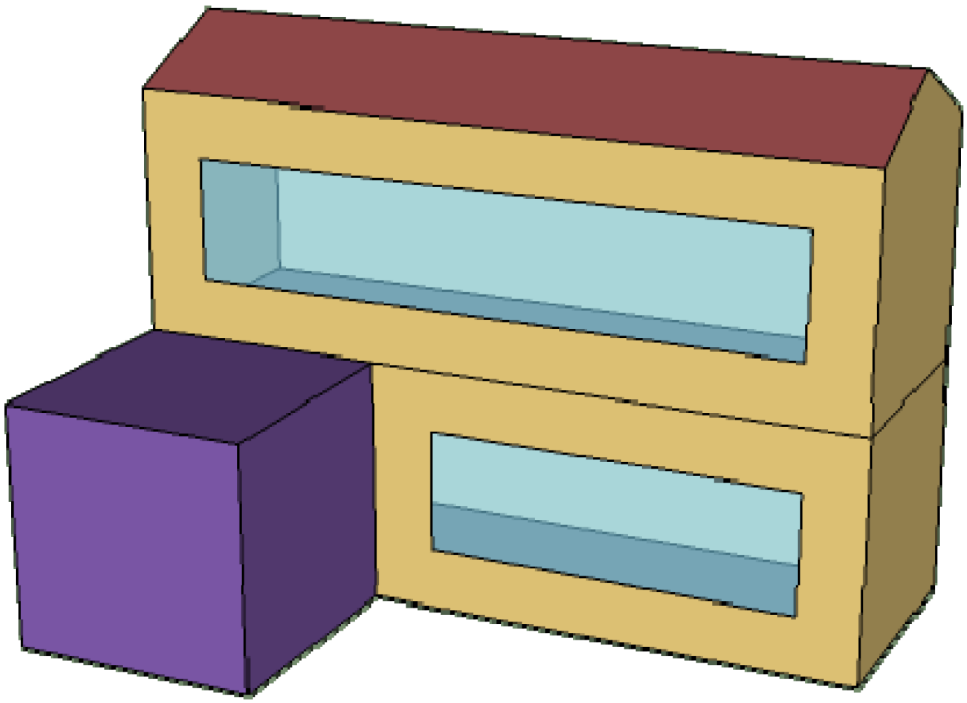


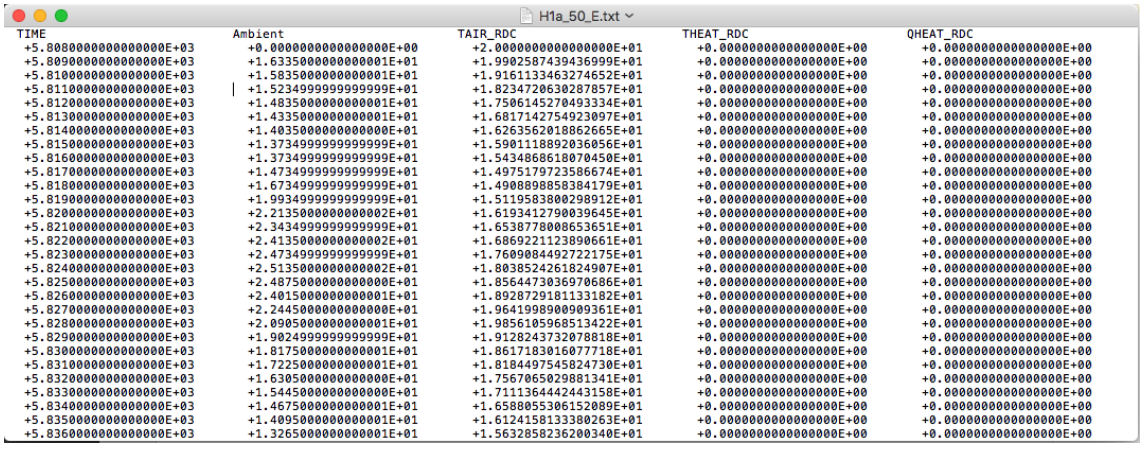
FIGURE 2.8 – Exemple de géométrie de bâtiment à l'aide du logiciel SketchUp.

Choix de l'outil : TRNSYS. L'un des objectifs du projet FUI Apache est la modélisation de la chaudière thermodynamique couplée à un bâtiment afin d'en évaluer les performances annuelles via une simulation thermique dynamique. Le logiciel TRNSYS, déjà utilisé par Engie, partenaire du projet, permet à la fois de modéliser les systèmes de chauffages et les bâtiments.

Présentation du logiciel TRNSYS. *Le logiciel permet d'intégrer toutes les caractéristiques d'un bâtiment et de son équipement (les systèmes de chauffage et la*

climatisation), pour l'étude détaillée du comportement thermique de ce bâtiment, en fonction de son emplacement, des matériaux de construction utilisés, de l'architecture globale, du concept énergétique choisi, etc. [75]. Des modules (appelés "types") permettent de modéliser séparément les phénomènes physiques et les systèmes. Un module calcule les valeurs de ses sorties à partir des valeurs intrinsèques (paramètres ou données) et des valeurs dépendantes (entrées). Tous les modules sont assemblés afin de respecter les couplages physiques et un processus de calcul itératif est mis en place afin de faire converger le calcul.

La modélisation thermique dynamique avec TRNSYS. La modélisation de bâtiment avec TRNSYS permet de décrire de façon détaillée un grand nombre de caractéristiques du bâtiment. Un plug-in permet de dessiner la géométrie du bâtiment et les zones thermiques à l'aide du logiciel SketchUp. L'image 2.8 montre à titre d'exemple un bâtiment de deux étages, avec des vitres en façade et sans mur intérieur. Ce bâtiment est modélisé avec deux grandes zones thermiques. SketchUp et TRNSYS permettent également de modéliser les ombrages subis par le bâtiment, représentés sur la figure 2.8 par le cube violet sur le devant du bâtiment. Une fois la géométrie et l'orientation définies, il reste à expliciter l'ensemble des paramètres thermiques du bâtiment : composition des parois (mur, sol, plafond, toit), fenêtre, infiltration d'air, ventilation, pont thermique, chauffage, climatisation, apport interne (personne, ordinateur, éclairage, etc). Pour les parois et les fenêtres, les caractéristiques des matériaux les plus couramment utilisés sont déjà disponibles. Les entrées et sorties du modèle sont personnalisables en fonction des besoins (différents bilans énergétiques, puissance ou énergie reçue par zone). Le chauffage et la climatisation peuvent être modélisés à l'intérieur du module bâtiment ou à l'extérieur via d'autres modules proposés dans TRNSYS. L'ensemble des modèles mathématiques utilisés pour simuler le comportement du bâtiment a été validés et est disponibles dans la documentation fournie avec le logiciel [76].


H1a_50_E.txt

TIME	Ambient	TAIR_RDC	THEAT_RDC	QHEAT_RDC
+5.8080000000000000E+03	+0.0000000000000000E+00	+2.0000000000000000E+01	+0.0000000000000000E+00	+0.0000000000000000E+00
+5.8090000000000000E+03	+1.6335000000000001E+01	+1.9902587439436999E+01	+0.0000000000000000E+00	+0.0000000000000000E+00
+5.8100000000000000E+03	+1.5835000000000001E+01	+1.9161133463274652E+01	+0.0000000000000000E+00	+0.0000000000000000E+00
+5.8110000000000000E+03	+1.5234999999999999E+01	+1.8234720630287857E+01	+0.0000000000000000E+00	+0.0000000000000000E+00
+5.8120000000000000E+03	+1.4835000000000001E+01	+1.7506145270493334E+01	+0.0000000000000000E+00	+0.0000000000000000E+00
+5.8130000000000000E+03	+1.4335000000000001E+01	+1.6817142754923097E+01	+0.0000000000000000E+00	+0.0000000000000000E+00
+5.8140000000000000E+03	+1.4035000000000000E+01	+1.6263562018862665E+01	+0.0000000000000000E+00	+0.0000000000000000E+00
+5.8150000000000000E+03	+1.3734999999999999E+01	+1.5901118892036056E+01	+0.0000000000000000E+00	+0.0000000000000000E+00
+5.8160000000000000E+03	+1.3734999999999999E+01	+1.5434868618070450E+01	+0.0000000000000000E+00	+0.0000000000000000E+00
+5.8170000000000000E+03	+1.4734999999999999E+01	+1.4975179723586674E+01	+0.0000000000000000E+00	+0.0000000000000000E+00
+5.8180000000000000E+03	+1.6734999999999999E+01	+1.4900898858384179E+01	+0.0000000000000000E+00	+0.0000000000000000E+00
+5.8190000000000000E+03	+1.9934999999999999E+01	+1.5119583800298912E+01	+0.0000000000000000E+00	+0.0000000000000000E+00
+5.8200000000000000E+03	+2.2135000000000002E+01	+1.6193412790039645E+01	+0.0000000000000000E+00	+0.0000000000000000E+00
+5.8210000000000000E+03	+2.3434999999999999E+01	+1.6538778008653651E+01	+0.0000000000000000E+00	+0.0000000000000000E+00
+5.8220000000000000E+03	+2.4135000000000002E+01	+1.6869221123890661E+01	+0.0000000000000000E+00	+0.0000000000000000E+00
+5.8230000000000000E+03	+2.4734999999999999E+01	+1.7609084492722175E+01	+0.0000000000000000E+00	+0.0000000000000000E+00
+5.8240000000000000E+03	+2.5135000000000002E+01	+1.8038524261824907E+01	+0.0000000000000000E+00	+0.0000000000000000E+00
+5.8250000000000000E+03	+2.4875000000000000E+01	+1.8564473036970686E+01	+0.0000000000000000E+00	+0.0000000000000000E+00
+5.8260000000000000E+03	+2.4015000000000001E+01	+1.8928729181133182E+01	+0.0000000000000000E+00	+0.0000000000000000E+00
+5.8270000000000000E+03	+2.2445000000000000E+01	+1.9641998900909361E+01	+0.0000000000000000E+00	+0.0000000000000000E+00
+5.8280000000000000E+03	+2.0905000000000001E+01	+1.9856105968513422E+01	+0.0000000000000000E+00	+0.0000000000000000E+00
+5.8290000000000000E+03	+1.9024999999999999E+01	+1.9128243732078818E+01	+0.0000000000000000E+00	+0.0000000000000000E+00
+5.8300000000000000E+03	+1.8175000000000001E+01	+1.8617183016077718E+01	+0.0000000000000000E+00	+0.0000000000000000E+00
+5.8310000000000000E+03	+1.7225000000000001E+01	+1.8184497545824730E+01	+0.0000000000000000E+00	+0.0000000000000000E+00
+5.8320000000000000E+03	+1.6305000000000000E+01	+1.7567065029881341E+01	+0.0000000000000000E+00	+0.0000000000000000E+00
+5.8330000000000000E+03	+1.5445000000000000E+01	+1.7111364442443158E+01	+0.0000000000000000E+00	+0.0000000000000000E+00
+5.8340000000000000E+03	+1.4675000000000001E+01	+1.6588055306152089E+01	+0.0000000000000000E+00	+0.0000000000000000E+00
+5.8350000000000000E+03	+1.4095000000000001E+01	+1.6124158133380263E+01	+0.0000000000000000E+00	+0.0000000000000000E+00
+5.8360000000000000E+03	+1.3265000000000001E+01	+1.5632858236200340E+01	+0.0000000000000000E+00	+0.0000000000000000E+00

FIGURE 2.9 – Exemple de fichier généré par TRNSYS.

Utilisation de TRNSYS dans la recherche. Depuis les années 1990 de plus en plus de recherches scientifiques utilisent TRNSYS pour simuler le comportement

thermique du bâtiment. Les applications de ces simulations sont variées. TRNSYS est utilisé :
Comme référence pour évaluer d'autres méthodes de modélisation du bâtiment[77][78] ;
Pour mesurer l'impact sur le bilan énergétique final d'une modification du bâtiment, typologie générale[79], ou de l'ajoût d'un équipement innovant au sein du bâtiment, façade biomimétique [80], façade photovoltaïque[81], isolation et fenêtre[82], orientation de stores vénitiens[83], ventilation croisée[84] ;
Pour estimer le comportement thermique d'un bâtiment[85][86][87][88] ;
Pour évaluer le potentiel d'un système de chauffage ou de climatisation[89][90] ;
Pour évaluer des stratégie de contrôle de systèmes de chauffage, de climatisation, et de ventilation[91][92][93] , au sein de salle de classe dans une université[94], dans un bâtiment commercial[95], pour déterminer le moment optimum pour remettre en marche le chauffage dans un bâtiment occupé par intermittence [96] ;
Pour étudier l'apport d'un équipement à énergie renouvelable aux consommations énergétiques du bâtiment, par exemple des échangeurs géothermiques[97],un système solaire combiné [98],un système de stockage thermique solaire [99][100], un chauffage et une climatisation solaire[101], une pompe à chaleur[102].

Classe énergétique. La réglementation thermique 2012(RT2012), en vigueur actuellement définit des étiquettes énergétiques, aussi appelées classes énergétiques. Ces classes permettent de catégoriser les bâtiments selon leurs consommations globales d'énergie primaire. La classe énergétique A, la plus économe en énergie, regroupe les bâtiments consommant moins de $50kWh/m^2/an$, alors que la classe G, la moins économe en énergie, regroupe les bâtiments consommant plus de $450kWh/m^2/an$. Les valeurs seuils pour les autre classes énergétiques sont indiquées sur la figure 2.10. Ces classes sont utilisées pour évaluer nos cas d'étude car elles sont un bon

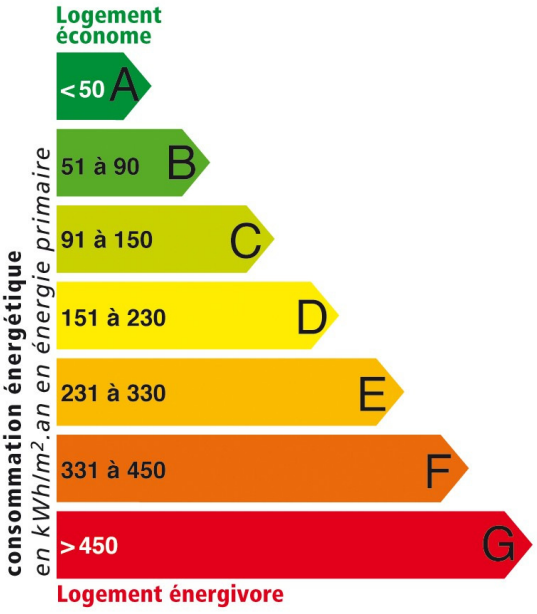


FIGURE 2.10 – Etiquettes énergétiques

indicateur des consommations de chauffages puisque dans l'habitat individuel le chauffage représente 67% des consommations totales.

Climats normés. Le climat a une forte influence sur les consommation d'énergie pour le chauffage. Il parrait evidant qu'il faudra plus d'énergie pour maintenir la température intérieure d'une maison si il fait 0°C à l'extérieur que s'il fait 10°C. Afin de déterminer les besoins énergétiques d'un bâtiment, la RT2012 définit huit climats types. Pour chacun de ces climats des données météorologiques sont créés afin de représenter une année type. *Ces années type de référence ont été définies à partir de la norme Européenne NF EN ISO 15927-4 « Performance hygrothermique des bâtiments. Calcul et présentation des données climatiques ; Données horaires pour l'évaluation du besoin énergétique annuel de chauffage et de refroidissement » en suivant une procédure basée sur le choix de mois représentatifs, qui sont ensuite raccordés aux limites afin d'éviter des sauts brusques de variables. La constitution de ces années type de référence a été effectuée sur la base de fichiers annuels des données mesurées par Météo-France (www.meteo.fr) sur la période de janvier 1994 à décembre 2008. Ces données climatiques ont ensuite été post-traitées de façon à ramener les données à une altitude nulle et à effectuer la séparation des rayonnements directs et diffus* (RT2012). Sur la figure 2.11 est représenté une carte de france découpée selon les climats réglementaires. Ces données climatiques sont utilisées lors de nos simulation afin de tester la pertinence de nos modèles sur l'ensemble du territoire français.

FIGURE 2.11 – Carte des climats réglementaires

Paramètres des modèles. Nous nous basons sur nos connaissances en thermique du bâtiment ainsi que sur la littérature et les données statistiques présentées dans

le chapitre 1 afin de déterminer les modèles représentatifs du parc immobilier français. Nous choisissons de décrire les bâtiments selon trois critères : la superficie, la classe énergétique et le climat réglementaire auquel le bâtiment est soumis. En ce qui concerne la classe énergétique, nous investiguons les classes A,B,C,D et E, ce qui représentent 71% des logements individuels[103]. Les superficies testées sont $50m^2$,$100m^2$ et $150m^2$ ce qui représente un cas minimal, moyen et maximal[104]. Dans le but d'être le plus généraliste possible, nous optons pour une géométrie simple et un système de chauffage modélisé par un simple apport thermique. La géométrie des bâtiments est semblable pour tous les cas, un étage unique, comportant des surfaces vitrées en façade sud, ouest et est. Le tout est surmonté d'un toit pentu. Les surfaces des parois, et des surfaces vitrées sont indiquées dans le tableau 2.3. Nous considérons les bâtiments comme mono-zones, c'est-à-dire une seule zone thermiquement homogène. Cette hypothèse nous place dans la situation d'un thermostat central considéré comme représentatif de la température de l'ensemble de l'habitat. Le type de matériaux d'isolation et leurs épaisseurs, ainsi que le type de fenêtres sont modifiés pour créer cinq modèles par géométrie, correspondant aux cinq étiquettes énergétiques.

TABLE 2.3 – Informations techniques sur les mesures expérimentales

Modèle	Surface de façade (m^2)				Surface vitrée (m^2)		
	sud	est	ouest	nord	sud	est	ouest
$50m^2$	25	12	12	25	4,2	0,8	0,8
$100m^2$	25	25	25	25	6	4,6	4,6
$150m^2$	25	38	38	25	3,77	9,2	9,2

Paramètres de simulation. Dans le but d'investiguer l'ensemble du territoire français, les simulations sont faites sous les 8 climats réglementaires, H1a, H1b, H1c, H2a, H2b, H2c, H2d et H3. Chacun des 120 cas, modèles et conditions de simulation, est nommé selon son triplé représentatif : superficie, classe énergétique, climat. Les scénarios d'occupation sont définis d'après les recommandations de la réglementation thermique (RT2012)[105]. La température de consigne est fixée à 18°C entre 6h et 22h, et à 16°C le reste de la journée. Nous simulons les consommations de chauffage et la température d'air de ces bâtiments sur une saison de chauffe, avec un pas de temps d'une heure. Le premier mois de simulation est supprimé afin d'éliminer l'influence des conditions initiales. Le logiciel TRNSYS génère en sortie un fichier texte, voir figure 2.9, comportant sur la première colonne le temps de simulation, et sur chaque colonne suivante une donnée de sortie, ici la température extérieure, la température intérieure, la température de consigne et la puissance de chauffage.

2.4 Conclusion

Dans ce chapitre, nous présentons les cas d'études qui servent à développer et à tester nos ANN. Nous distinguons deux types de cas : les cas expérimentaux et les cas simulés. Dans une première partie, nous décrivons les trois cas expérimentaux :

trois servent pour la prédiction d'énergie pour l'ECS, deux servent pour la prédiction de température intérieure du bâtiment. En terme de consommation d'ECS, le cas 2 représente une consommation moyenne, le cas 1 représente une faible consommation et le cas 3 représente une forte consommation. En terme de consommation énergétique globale, le cas 1, d'étiquette énergétique C, est à la fois moyen et représentatif de la classe énergétique majoritaire. Le cas 2, d'étiquette énergétique A, est lui représentatif d'habitation peu énergivore. Les trois cas comportent la même particularité : une volonté des habitants de tester de nouveaux systèmes solaires combinés. Cela dénote une sensibilité à leur impact environnemental et à leur consommation d'énergie et donc potentiellement une utilisation raisonnée ou économe des ressources énergétiques. Les cas simulés sont utilisés uniquement pour la prédiction de températures intérieures du bâtiment. Ils ont pour objectif de compléter les cas expérimentaux afin d'étudier des cas représentatifs du parc immobilier français.

CHAPITRE 3

Prédiction de la température intérieure du bâtiment

3.1 Introduction

Les réseaux de neurones artificiels (ANN) sont utilisés au sein de régulation prédictive appliquée à des systèmes de chauffage. Des modèles ont été développés pour prédire la puissance de chauffage ou la température intérieure d'un bâtiment. Bien que la méthode des ANN présente une grande adaptabilité, les études présentes dans la littérature s'intéressent à un bâtiment à la fois. Nous posons l'hypothèse qu'une architecture unique de ANN et une méthode d'apprentissage unique peuvent modéliser l'évolution de la température intérieure, et que le jeu d'apprentissage permet d'adapter le ANN à plusieurs habitations. Nous souhaitons développer un ANN capable de s'adapter à n'importe quel bâtiment, pour cela nous limitons les entrées possibles aux grandeurs mesurées sur une installation de chauffage classique. Afin de réduire la complexité de la prédiction, induit par cette limitation des choix d'entrée nous réalisons une prédiction à court terme, soixante-quatre minutes pour les modèles des cas expérimentaux et une heure pour les modèles des cas simulés. Afin de simplifier le discours et les notations, nous parlerons dans tous les cas de prédiction à une heure. Nous parlons de "ANN" pour désigner l'architecture et la méthode d'apprentissage que nous développons et nous parlons de "modèle" lorsque le ANN est adapté à un cas, c'est-à-dire après la phase d'apprentissage.
Dans la partie 3.2 de ce chapitre, nous développons un ANN qui prédit la température intérieure d'un bâtiment résidentiel. Pour cela, nous développons une architecture et sa méthodologie d'apprentissage à l'aide des données mesurées sur deux cas expérimentaux. Les performances de ce ANN sont mesurées afin d'évaluer le choix des grandeurs d'entrée. Dans la partie 3.4, nous évaluons la capacité du ANN à obtenir des performances équivalentes pour tous types de maisons individuelles sous les climats français, pour cela, nous adaptons le ANN à cent-vingt cas simulés.

3.2 Développement de l'architecture et de la méthodologie d'apprentissage

Démarche Comme présenté dans le chapitre 2, nous disposons d'une grande quantité de données à propos des consommations énergétiques de deux cas expérimentaux. Afin de développer des modèles de ces deux cas, nous mettons en place plusieurs étapes de traitement des données, de développement de l'architecture et de développement de la méthode d'apprentissage du ANN. Nous représentons en figure 3.1 un schéma de ces étapes. Nous y faisons référence tout au long de cette partie.

Étape 1 : définition du ANN. La première étape de notre démarche est de définir notre ANN. Comme nous le présentons dans la partie 1.4 un réseau de neurones est défini par son architecture, son nombre d'entrées, de sorties, de couches intermédiaires et de neurones sur chaque couche ainsi que par le jeu de données d'apprentissage choisi. Dans notre cas le choix des entrées est dépendant du phénomène physique à prédire : la température intérieure du bâtiment. Le jeu d'apprentissage permet de s'adapter au cas d'étude, c'est-à-dire à l'utilisateur et aux caractéristiques thermiques du bâti.

Définition des entrées et sorties Nous cherchons à prédire une grandeur : la température intérieure 1 heure après l'instant de prédiction : $\widehat{T}_{int}(t+1h)$. En se basant sur la littérature et nos connaissances du problème physique, nous déterminons quatre types d'influence dont dépend la température intérieure du bâtiment : les conditions météorologiques, le chauffage, les caractéristiques thermiques du bâti et les occupants, leur comportement ainsi que leurs notions de confort thermique. Prendre en compte le comportement des occupants nécessite l'ajout de capteurs, tels que capteur de présence par exemple, d'ouverture des portes et fenêtres, ou autres. Or, nous souhaitons un système non-invasif, ne nécessitant pas de capteur supplémentaire par rapport à une installation de chauffage classique. La sensibilité de l'occupant au confort thermique pourrait être pris en compte par l'utilisation de la température de consigne fournie au système de chauffage. Malheureusement, cette information n'est pas disponible dans les mesures expérimentales à notre disposition. Ces influences ne sont donc pas directement prises en compte par notre modèle. Cependant la variation de température intérieure dans les instants précédents la prédiction comporte indirectement des informations sur la sensibilité au confort thermique de l'occupant et sur son comportement. Beaucoup d'installations de chauffage récentes comportent une régulation basée sur les lois d'eau, ce type de régulation nécessite une mesure de température extérieure sur site, ainsi qu'une mesure de température intérieure. La puissance de chauffage consommée est usuellement disponible. Bien que les caractéristiques thermiques du bâti soient complexes, la variation de température intérieure par rapport à la variation de température extérieure peut permettre de déterminer l'inertie thermique du bâtiment, c'est-à-dire la capacité des matériaux à stocker de l'énergie, générant un décalage entre une variation de température extérieure et son impact sur la température intérieure. De multiples choix d'entrées sont disponibles dans la littérature. À titre d'exemple Huang et al. [106] développent

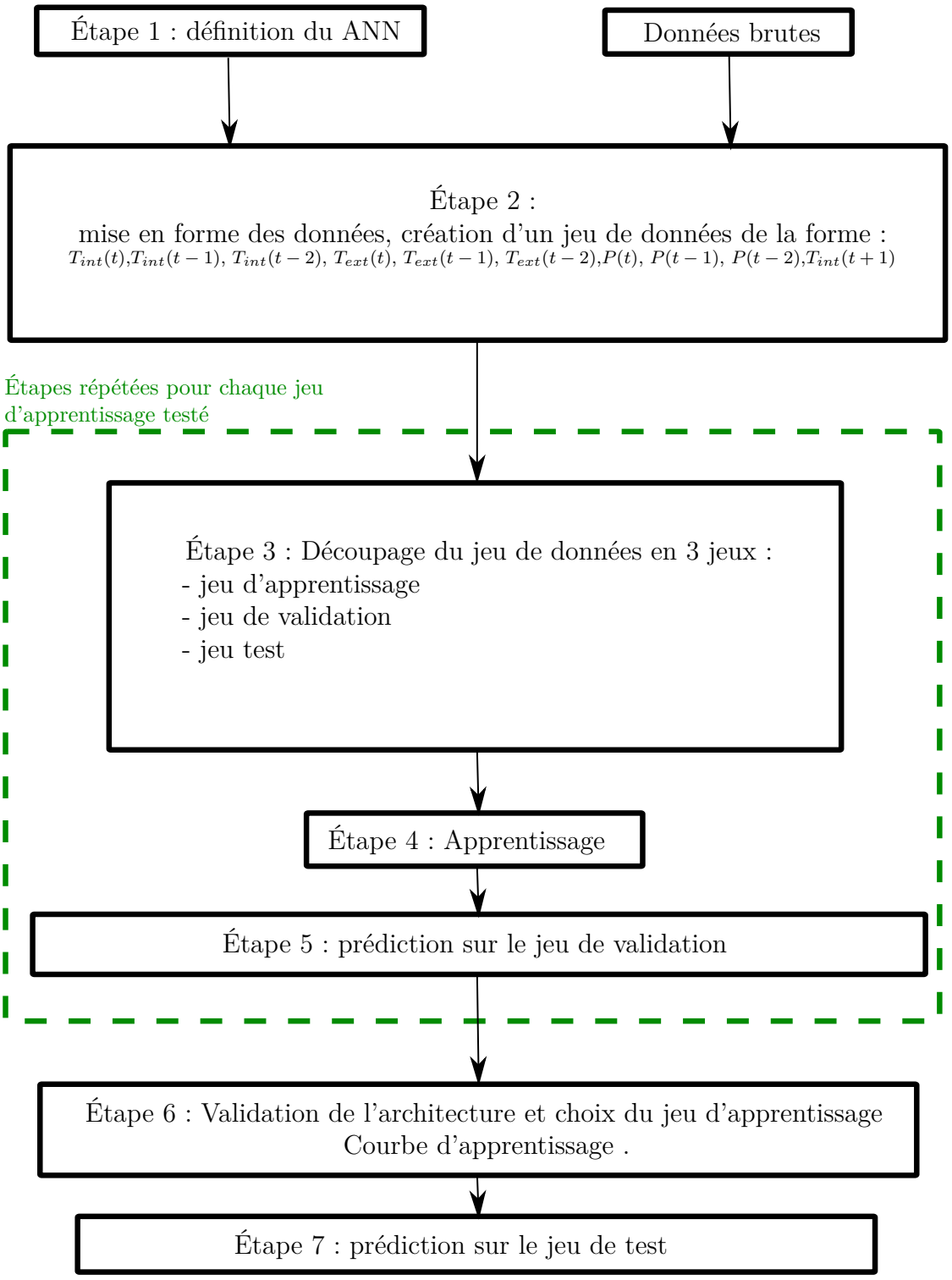


FIGURE 3.1 – Étape de développement des modèles.

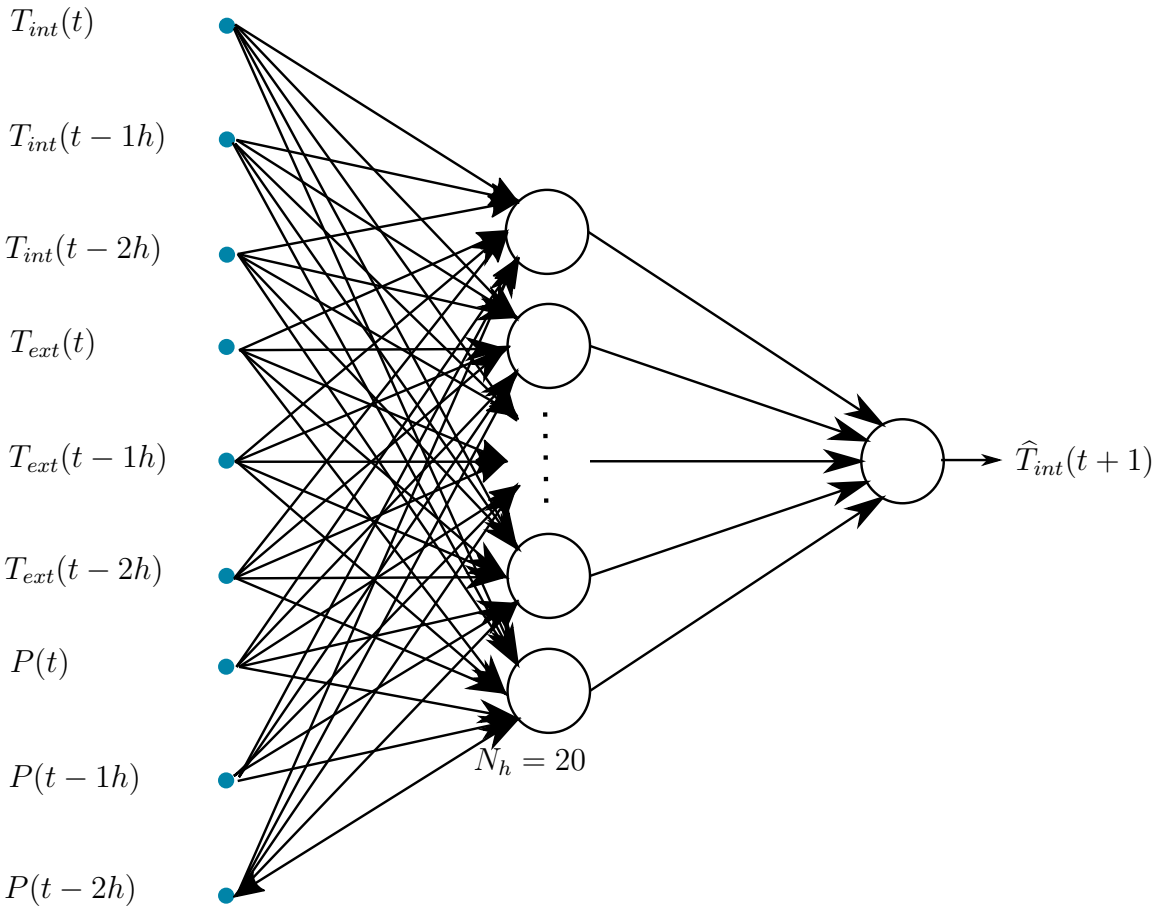


FIGURE 3.2 – Architecture du réseau de neurones.

un ANN qui prend en entrées les températures intérieures et extérieures aux pas de temps précédents, ainsi que des variables contrôlables, liées au système de chauffage. Leur modèle prédit la température intérieure d'un centre commercial avec une erreur moyenne inférieur à 0,7°C. Moon et al.[107] considèrent les températures intérieures et extérieures, ainsi que l'ouverture des systèmes de ventilation. Au vu de ce qui a été fait dans la littérature et des données dont nous disposons, nous choisissons de prendre comme entrées de notre ANN :

- la température intérieure à l'instant où la prédiction est faite : $T_{int}(t)$, une heure avant la prédiction : $T_{int}(t-1)$, et deux heures avant la prédiction : $T_{int}(t-2)$.
- la température extérieure à ces trois instants : $T_{ext}(t)$, $T_{ext}(t-1)$, $T_{ext}(t-2)$.
- la puissance de chauffage à ces trois instants : $P(t)$, $P(t-1)$, $P(t-2)$.

	Température(°C)	Puissance(kWh)
minimum	-15	0
maximum	40	65

TABLE 3.1 – Valeurs utilisées pour la normalisation

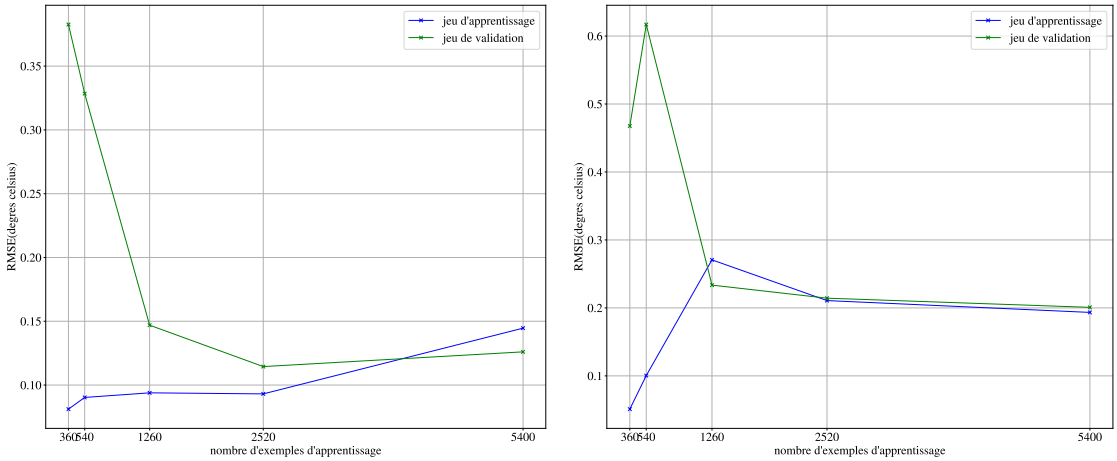
Architecture du ANN Nous montrons dans le chapitre 1 que les ANN ont été utilisés de nombreuses fois avec succès afin de prédire la température intérieure d'un bâtiment. En se basant sur ces travaux nous développons un ANN avec une couche intermédiaire dont le nombre de neurones est calculé d'après l'équation 3.1 extrait des travaux de Moon et al.[107]. Nous utilisons la fonction sigmoïde comme fonction d'activation sur la couche intermédiaire et la fonction identité sur la couche de sortie. L'utilisation de la fonction sigmoïde nécessite une normalisation des données d'entrées 1.4, le tableau 3.1 indique les valeurs utilisées pour la normalisation. L'architecture choisie est représentée sur la figure 3.2.

$$N_i = 2(N_e + 1) \tag{3.1}$$

Étape 2 : Mise en forme des données Une fois l'architecture du ANN défini, nous mettons en forme nos données. À cette étape les données sont sous la forme d'un tableau avec en colonne les grandeurs mesurées et en ligne l'instant de mesure. Pour utiliser les données, il faut les mettre sous la forme d'exemple, c'est-à-dire que chaque ligne comporte les valeurs d'entrées du ANN et la sortie correspondante. Dans notre cas, la ligne au temps t est composée de $T_{int}(t)$,$T_{int}(t-1)$, $T_{int}(t-2)$, $T_{ext}(t)$, $T_{ext}(t-1)$, $T_{ext}(t-2)$,$P(t)$, $P(t-1)$, $P(t-2)$,$T_{int}(t+1)$. Dans la suite, $T_{int}(t+1)$ est appelée valeur vraie de la prédiction faite au temps t. L'ensemble des données disponibles sont mises sous cette forme.

Étape 3 : création des trois jeux de données Nous venons de créer un jeu d'exemples correspondant à l'architecture entrée/sortie de notre ANN. À présent nous séparons ces exemples en trois catégories, le jeu d'apprentissage, le jeu de validation et le jeu de test. Le jeu d'apprentissage sert à l'apprentissage du ANN, c'est-à-dire à déterminer les poids du ANN en fonction du cas à modéliser. Le jeu de validation sert à valider l'architecture du ANN. Le jeu de test sert à évaluer les capacités de prédiction du modèle.
Habituellement la majorité des données disponibles sont utilisées pour l'apprentissage et le reste pour la validation et le test du modèle, typiquement deux tiers pour l'apprentissage, un sixième pour le jeu de validation et un sixième pour le jeu de test. Or, nous souhaitons développer un ANN utilisable rapidement sur un grand nombre de bâtiments, autrement dit l'apprentissage doit être composé de données collectées le plus rapidement possible après l'installation du système. La littérature ne définit pas de quantité minimale de données nécessaire à un bon apprentissage dans le cas de prédictions de températures intérieures du bâtiment, ni de critère de qualité des données d'apprentissage. Néanmoins, l'apprentissage doit être fait sur une période où le chauffage fonctionne, un apprentissage où les valeurs de puissance de chauffage sont nulle ne permet pas de prédire la température intérieure une fois que le chauffage fonctionne. Nous choisissons de créer plusieurs jeux d'apprentissage composés de toutes les données mesurées sur une période prédéterminée. Nous utilisons les premières données dont nous disposons pour créer cinq jeux d'apprentissage composés de 360, 540, 1260, 2520 et 5 400 exemples. Soit respectivement les deux, trois, sept, quatorze, et trente premiers jours des données disponibles. Le jeu de validation est composé de 10 800 exemples, soit 2 mois de

données. Les exemples restants composent le jeu de test, la taille de ce jeu est donc variable selon les cas d'étude.



(A) cas 1 (B) cas 2

FIGURE 3.3 – Courbes d'apprentissage, à gauche du cas 1, à droite du cas 2 : $RMSE$ sur le jeu d'apprentissage (en bleu) et sur le jeu de validation (en vert) en fonction du nombre d'exemples d'apprentissage.

Étape 4 : apprentissage Les jeux de données ainsi créés, nous réalisons la phase d'apprentissage. Nous nous basons sur la littérature afin de déterminer l'algorithme d'optimisation le plus adapté à l'apprentissage. Nous utilisons l'algorithme de Levenberg-Marquardt comme algorithme d'apprentissage. Lors de cette étape, nous avons défini pour chacun des cas expérimentaux les poids du ANN. Nous disposons alors de deux modèles adaptés automatiquement à nos cas.

Étape 5 : prédiction sur le jeu de validation. Lors de cette étape, nous utilisons les modèles précédemment développés pour prédire la température intérieure du bâtiment. Cette prédiction est comparée à la sortie attendue en calculant l'erreur moyenne $\bar{\varepsilon}$ sur le jeu d'apprentissage et sur le jeu de validation.

Étape 6 : Validation de l'architecture et choix du jeu d'apprentissage. Les étapes 3, 4 et 5 sont répétées pour chaque durée d'apprentissage testée. Afin d'évaluer la qualité de l'apprentissage et de déterminer la quantité minimale de données d'apprentissage, nous traçons les courbes d'apprentissage pour les deux cas expérimentaux sur la figure 3.3. Comme expliqué dans le chapitre 1, ces courbes représentent l'erreur faite sur la sortie d'un ANN en fonction du nombre d'exemples d'apprentissage. Nous observons sur la figure 3.3 que sur les deux cas, l'erreur sur le jeu de validation est suffisamment faible pour valider le choix de l'architecture. Nous observons également, toujours sur les deux cas, que l'augmentation de la durée d'apprentissage au-delà de deux semaines (2520 exemples) permet une diminution négligeable de l'erreur sur le jeu de validation. Nous choisissons donc d'évaluer les

performances de nos modèles avec une durée d'apprentissage de deux semaines (2520 exemples).

Étape 7 : prédiction sur le jeu de test. Afin d'évaluer les performances des modèles nous réalisons les prédictions de température sur l'ensemble des exemples du jeu de test. Cette évaluation est traitée dans la partie 3.3 de ce chapitre.

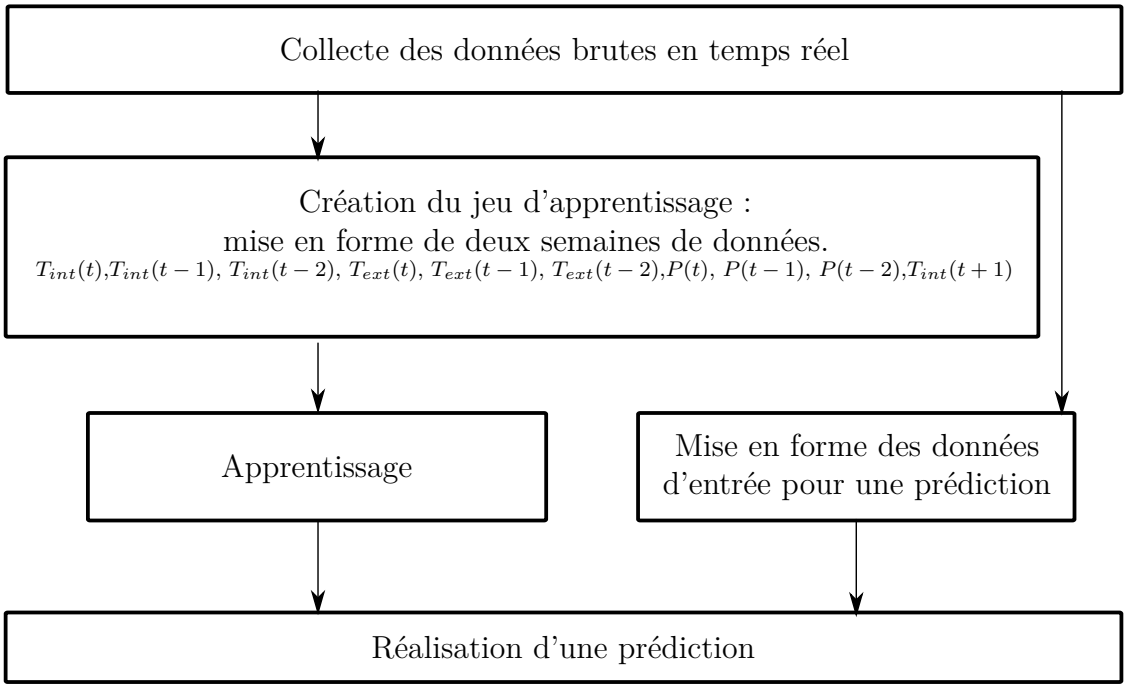


FIGURE 3.4 – Schéma des étapes nécessaire à l'adaptation et l'utilisation du ANN.

Temps de calcul. Le réseau de neurones artificiels que nous venons de développer a pour but d'être utilisé au sein d'une régulation de système de chauffage. De ce fait, nous devons nous assurer que les temps de calcul nécessaires à une prédiction sont cohérents avec cette utilisation. Les étapes précédant la réalisation d'une prédiction par nos modèles sont légèrement différentes pendant la phase de développement et la phase d'utilisation. Une représentation schématique des étapes réalisées lors d'une utilisation classique est proposée sur la figure 3.4. La première étape consiste à collecter deux semaines de données, ensuite l'apprentissage est réalisé, le bon fonctionnement de l'apprentissage est testé via le calcul de l'erreur sur le jeu d'apprentissage. Si l'apprentissage est validé, l'exemple de données nécessaire à la prédiction est mis en forme puis la prédiction est réalisée, sinon l'apprentissage est répété, sur les même données, jusqu'à sa validation. Le tableau 3.2 regroupe les temps de calcul des principales étapes mises en œuvre pendant la phase de développement du ANN sur les données des cas 1 et cas 2. Ces temps sont indiqués pour une machine équipée d'un processeur 3.1 GHz Intel Core i7 et de 16Go de mémoire vive. Il est évident qu'un régulateur aura des temps de calcul plus longs, les renseignements fournis ici

ont vocation à donner un ordre de grandeur des temps de calcul mis en œuvre. Le temps de calcul de la phase d'apprentissage est dépendant du nombre d'exemples d'apprentissage, le temps de réalisation de cette étape est fourni pour tous les jeux d'apprentissages testés. La phase la plus longue est la mise en forme des données, cette étape est réalisée pour plus d'un an de données, ce qui ne sera jamais le cas en réalité puisque les données sont mise en forme au fil de l'eau pour chaque instant de prédiction. L'étape la plus longue sera donc en réalité l'apprentissage dont la durée,pour un jeu d'apprentisssage de deux semaines (2520 exemple), est de cinq minutes, mise en forme des données comprise. Après deux semaines de collecte de données, quelques minutes pour réaliser l'apprentissage est un temps plus que raisonnable. Une fois cette étape terminée, chaque prédiction, mise en forme des données comprise ne prendra que quelques secondes. Ces temps de calculs sont parfaitement acceptables pour une utilisation en temps réel.

	nombre d'exemples d'apprentissage	Mise en forme des données	Apprentissage	Prédiction sur le jeu de validation	Prédiction sur le jeu de test
cas 1	180	345 s	1,2 s	$< 0,1s$	$< 0,1s$
	360		2,6 s		
	540		4,6 s		
	1260		20,9 s		
	2520		88,0 s		
	5400		460,0 s		
cas 2	180	325 s	0,9 s	$< 0,1s$	$< 0,1s$
	360		2,4 s		
	540		4,4 s		
	1260		19,1 s		
	2520		90,0 s		
	5400		477,0 s		

TABLE 3.2 – Temps de calcul, en seconde, des principales étapes de développement du ANN pour une machine équipée d'un processeur 3.1 GHz Intel Core i7 et de 16Go de mémoire vive

3.3 Évaluation des capacités d'un ANN à prédire la température intérieure d'un bâtiment

3.3.1 Indicateurs de performance

Les performances des modèles sont évaluées par plusieurs indicateurs :

- l'erreur signée ε faite sur une prédiction, définie par l'équation (4.1)

$$\varepsilon = T(t_i) - \hat{T}(t_i) \tag{3.2}$$

- La moyenne sur l'ensemble du jeu de test de la racine carré du carré de l'erreur entre la température prédite et la température réellement mesurée,

notée $RMSE$, calculée comme indiqué dans l'équation (3.3) où n représente le nombre d'exemples du jeu de test,

$$RMSE = \frac{1}{n}\sum_{i=1}^{n}\sqrt{\left(\hat{T}_{int}(t_i) - T_{int}(t_i)\right)^2} \tag{3.3}$$

- La moyenne de l'erreur absolue sur la prédiction de température intérieure sur le jeu de test, notée $\bar{\varepsilon}$, calculée comme indiqué dans l'équation (3.4),

$$\bar{\varepsilon} = \frac{1}{n}\sum_{i=1}^{n} T(t_i) - \hat{T}(t_i) \tag{3.4}$$

- l'écart-type associé, noté σ.

$$\sigma = \frac{1}{n}\sum_{i=1}^{n} \hat{T}(t_i) - \bar{T}(t_i) \tag{3.5}$$

Sur certains graphiques, le nombre important de données brouille la visibilité d'une tendance. C'est pourquoi nous avons alors recours à une moyenne glissante, définie ci-dessous sur l'équation (3.6), pour une grandeur x quelconque, avec N pair.

$$\bar{x}_n^N = \frac{1}{N}\sum_{k=n-\frac{N}{2}}^{n+\frac{N}{2}-1} x_k \tag{3.6}$$

3.3.2 Cas expérimentaux : évaluation des architectures et méthodes d'apprentissage.

Observations. Nous représentons sur la partie supérieure de la figure 3.5 les températures intérieures prédites et réelles pour le cas 1. Les mêmes représentations sont faites sur la figure 3.6 pour le cas 2. Nous observons de bonnes corrélations entre la température prédite et la température mesurée sur les premiers et derniers tiers de la période évaluer. C'est-à-dire entre janvier et avril 2004 puis entre octobre 2004 et avril 2005 pour le cas 1 et entre janvier et mars 2004 puis entre octobre 2004 et mars 2005 pour le cas 2. Cependant, entre ces périodes, les prédictions s'écartent visiblement des températures réelles. Dans les deux cas, cette période correspond à la période chaude de l'année, une explication plausible est que le chauffage ne fonctionnait pas pendant cette période. Afin d'établir le lien entre l'erreur faite par nos modèles et la puissance de chauffage, nous représentons sur le bas des figures 3.5 et 3.6 la puissance de chauffage consommée sur la même période, ainsi que l'erreur ε faite par le modèle. Les dates de fin de période de chauffe, 15 juin 2004 pour le cas 1, 11 mai 2004 pour le cas 2 et de début de période de chauffe suivante, 18 octobre 2004 pour le cas 1, 17 novembre 2004 pour le cas 2 sont représentées par des droites en pointillés sur les deux graphiques. Nous constatons que pour des puissances de chauffage nulles, l'erreur sur la prédiction augmente.

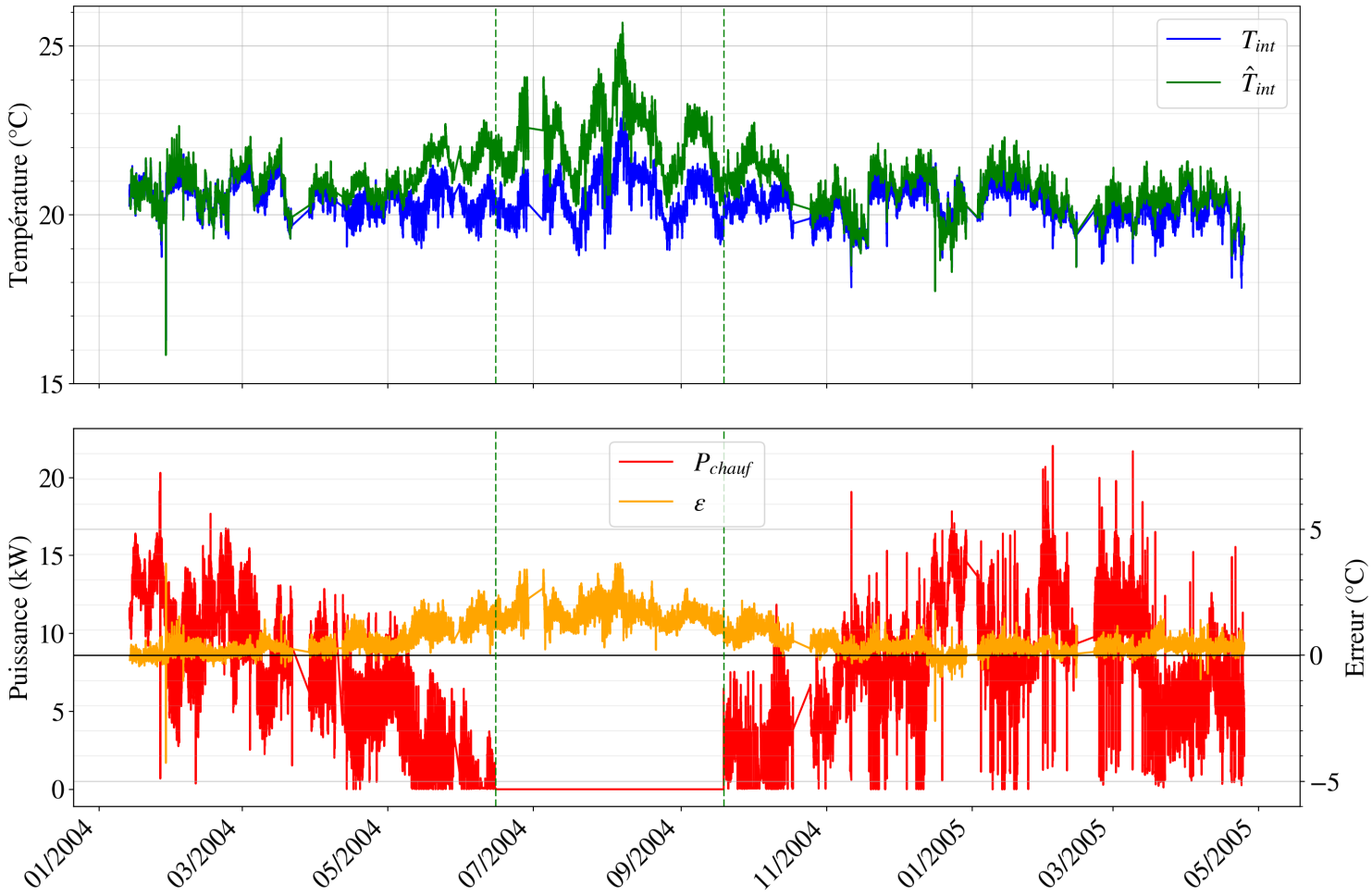


FIGURE 3.5 – Cas 1 : comparaison entre la température réelle et celle prédite sur tout le jeu de test. En haut : courbe de la température intérieure mesurée T_{int} (en bleu) et de la température intérieure prédite $\hat{T}_{int}$ (en vert), en °C, en fonction de la date. En bas : courbe de la puissance de chauffage mesurée P_{chauf} (en rouge), en kW, et de l'erreur sur la prédiction de la température $\varepsilon = T_{int} - \hat{T}_{int}$ (en orange), en °C, en fonction de la date.

FIGURE 3.6 – Cas 2 : comparaison entre la température réelle et celle prédite sur tout le jeu de test. En haut : courbe de la température intérieure mesurée T_{int} (en bleu) et de la température intérieure prédite $\hat{T}_{int}$ (en vert), en °C, en fonction de la date. En bas : courbe de la puissance de chauffage mesurée P_{chauf} (en rouge), en kW, et de l'erreur sur la prédiction de la température $\varepsilon = T_{int} - \hat{T}_{int}$ (en orange), en °C, en fonction de la date.

Influence de la puissance de chauffage consommée sur l'erreur de prédiction. Les observations faites précédemment semblent indiquer un lien entre la précision du modèle et la puissance de chauffage consommée. Nous représentons sur le haut de la figure 3.7 l'erreur de prédiction ε faite sur le cas 1 en fonction de la puissance de chauffage. Le grand nombre de valeurs sur l'ensemble du jeu de test rend difficile la lecture d'une tendance. Afin de pouvoir correctement visualiser les informations, nous représentons sur le bas de la figure 3.7 la même erreur ε, mais en ayant appliqué une moyenne glissante sur l'erreur et sur la puissance de chauffage sur trente jours, soit 5 400 mesures. Les mêmes représentations sont faites pour le cas 2 sur la figure 3.8. Nous observons une erreur plus importante lorsque la puissance de chauffage est nulle, puis l'erreur diminue, en valeur absolue, avec l'augmentation de la puissance de chauffage. La température intérieure est très fortement liée à la puissance de chauffage, de plus les données utilisées pendant la phase d'apprentissage sont collectées uniquement pendant une période de chauffe. Cela peut expliquer la forte augmentation de l'erreur lorsque la puissance de chauffage est nulle. Compte tenu de ces observations et du fait que les prédictions ont pour but d'être utilisées dans une régulation de chauffage, nous négligeons la période de non chauffe dans l'évaluation des performances des modèles sur les cas expérimentaux.

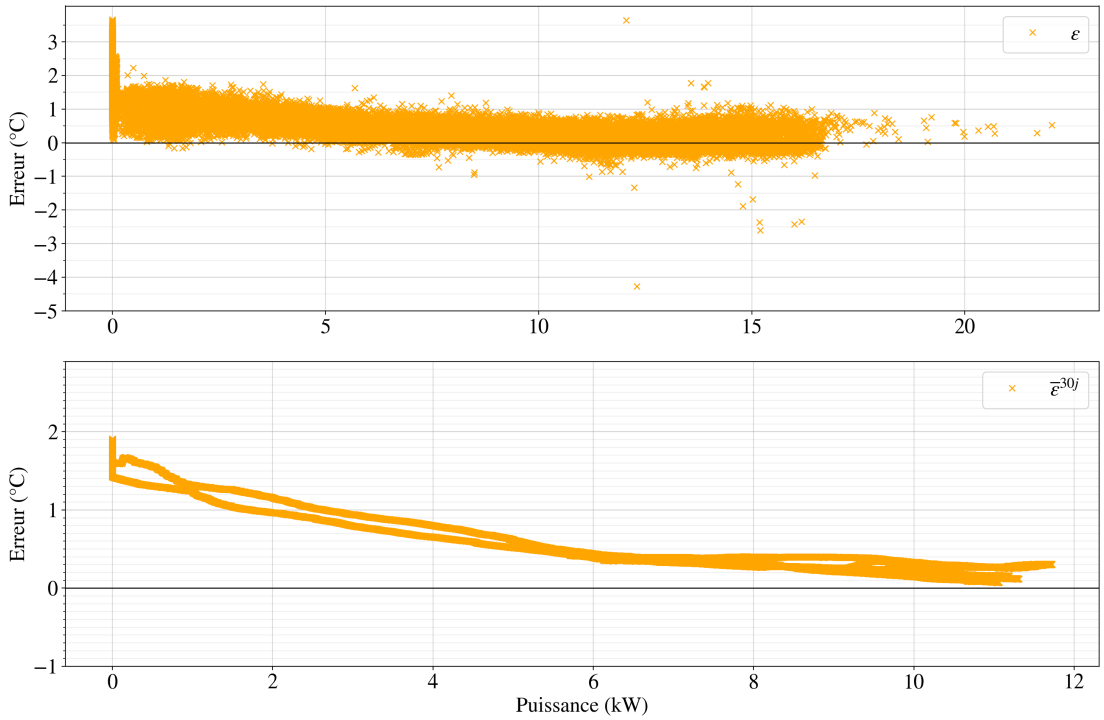


FIGURE 3.7 – Cas 1 : influence de la puissance de chauffage sur l'erreur de prédiction. En haut : erreur $\varepsilon = T_{int} - \hat{T}_{int}$, en °C, en fonction de la puissance de chauffage mesurée P_{chauf},en kW. En bas : erreur $\varepsilon = T_{int} - \hat{T}_{int}$, en °C, auquel nous avons appliqué une moyenne glissante, en fonction de la puissance de chauffage mesurée P_{chauf},en kW.

Observation sur une courte période. La représentation des prédictions du modèle sur l'ensemble du jeu de test ne permet pas une bonne visualisation des

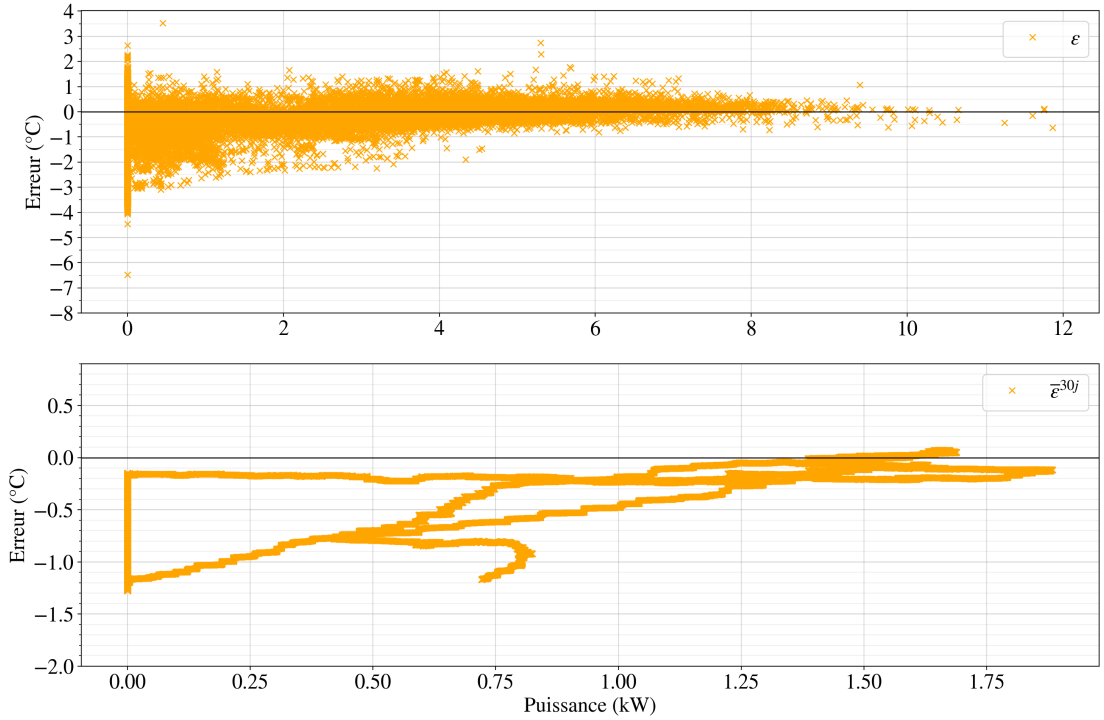


FIGURE 3.8 – Cas 2 : influence de la puissance de chauffage sur l'erreur de prédiction. En haut : erreur $\varepsilon = T_{int} - \hat{T}_{int}$, en °C, en fonction de la puissance de chauffage mesurée P_{chauf},en kW. En bas : erreur $\varepsilon = T_{int} - \hat{T}_{int}$, en °C, auquel nous avons appliqué une moyenne glissante, en fonction de la puissance de chauffage mesurée P_{chauf},en kW.

résultats. Dans ce but, nous représentons sur les figures 3.9 et 3.10, respectivement pour le cas 1 et le cas 2, la température prédite et réelle durant deux semaines du mois de février. Nous observons pour les deux cas étudiés que les modèles suivent parfaitement la tendance de variation de la température. Sur la période représentée, nous observons un écart maximal à la valeur vraie inférieur à 0,5°C. Toutefois le modèle du cas 1 semble majoritairement sous-estimer la température intérieure alors que celui de cas 2 la surestime.

Quantification de la précision des modèles. Les premières observations faites, bien qu'encourageantes sur la précision des modèles ne permet pas de quantifier la fiabilité des modèles. Nous calculons pour les deux modèles l'erreur signée et non signée pour chaque prédiction du jeu de test. Les valeurs moyennes $RMSE$ et $\bar{\varepsilon}$, l'écart-type $\sigma(RMSE)$ et $\sigma(\varepsilon)$ ainsi que les valeurs extrêmes $min(RMSE), max(RMSE)$ et $min(\varepsilon), max(\varepsilon)$ de ces deux erreurs sont regroupées dans le tableau 3.3. Ces valeurs permettent de confirmer la fiabilité des prédictions sur l'ensemble du jeu de test pour les deux cas expérimentaux. Pour le cas 1, l'éloignement moyen à la valeur vraie est de 0,45°C, pour le cas 2, il est de 0,53°C. Les valeurs extrêmes, à la fois de l'erreur signée et non signée, nous montrent que les modèles ne sont pas toujours justes, toutefois ces valeurs extrêmes, bien qu'inutilisables dans une régulation prédictive, reste du bon ordre de grandeur. L'erreur signée ε confirme les bonnes performances des modèles, mais aussi l'observation faite sur les figures 3.9 et 3.10. Dans le cas 1, la

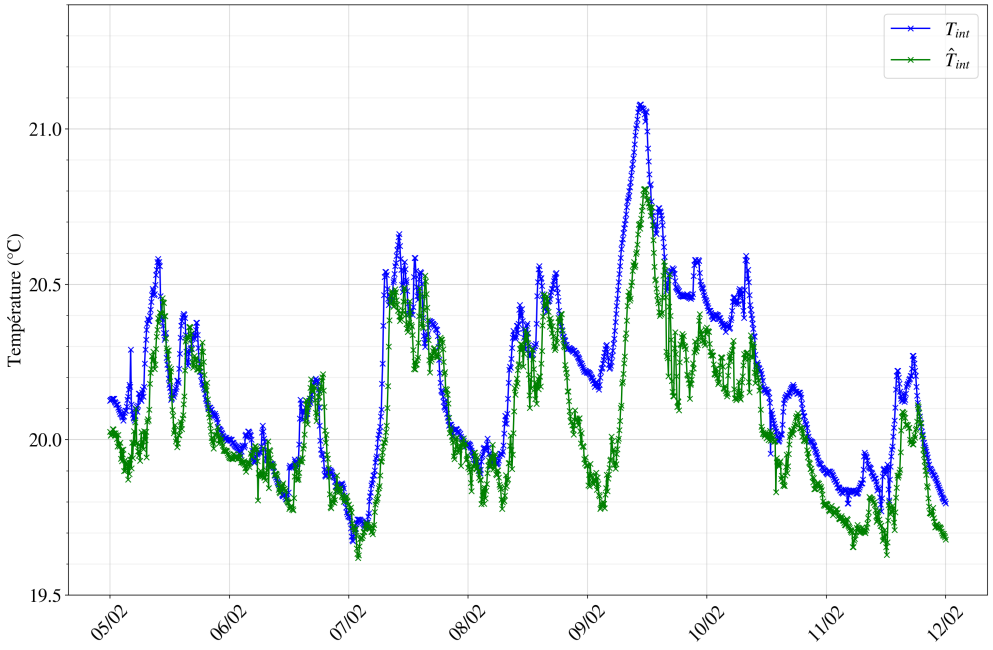


FIGURE 3.9 – Cas 1 : comparaison entre la température réelle et celle prédite du 05/02/05 au 12/02/05. Courbe de la température intérieure mesurée T_{int} (en bleu) et de la température intérieure prédite $\hat{T}_{int}$ (en vert), en °C, en fonction de la date.

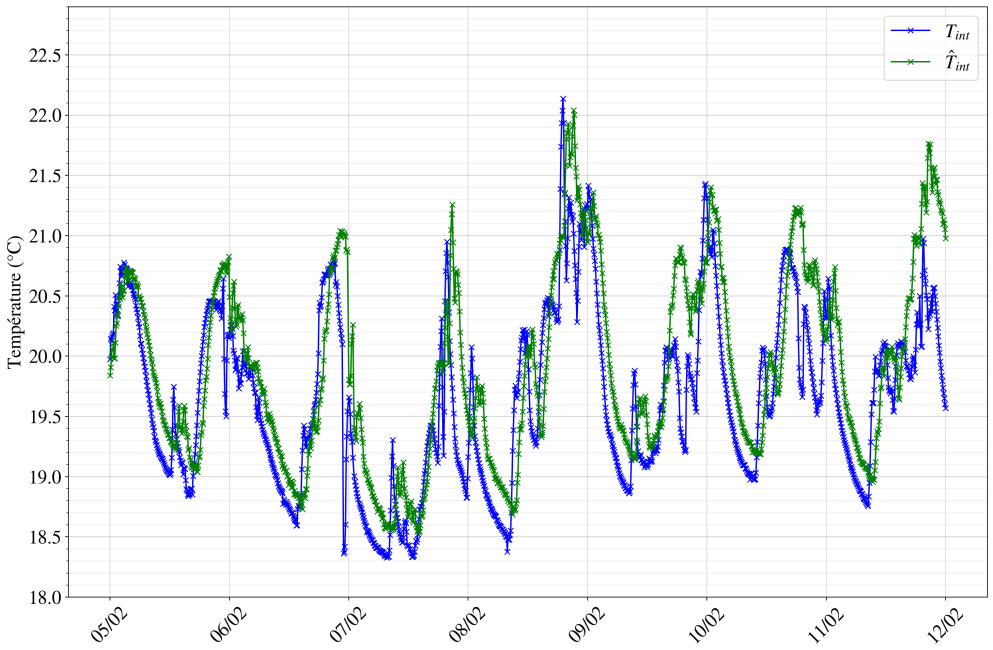


FIGURE 3.10 – Cas 2 : comparaison entre la température réelle et celle prédite du 05/02/05 au 12/02/05. Courbe de la température intérieure mesurée T_{int} (en bleu) et de la température intérieure prédite $\hat{T}_{int}$ (en vert), en °C, en fonction de la date.

température intérieure est sous-estimée par le modèle alors qu'elle est surestimée dans le cas 2. Il est difficile d'expliquer physiquement cette observation. Les valeurs de

l'écart-type $\sigma(\varepsilon)$ et $\sigma(RMSE)$ nous indiquent une concentration des valeurs autour de la moyenne.

	cas 1	cas 2
$RMSE$	0,45	0,53
$\sigma(RMSE)$	0,41	0,52
$min(RMSE)$	<0,01	<0,01
$max(RMSE)$	4,27	3,72
$\bar{\varepsilon}$	0,43	-0,34
$\sigma(\varepsilon)$	0,44	0,66
$min(\varepsilon)$	-4,27	-3,72
$max(\varepsilon)$	3,65	3,52

TABLE 3.3 – Valeur de l'erreur moyenne signée $\bar{\varepsilon}$ et non signée $RMSE$, ainsi que de l'écart-type σ et des extremum min et max associés pour les deux cas expérimentaux

Validation du ANN Nous observons et quantifions l'erreur faite par les deux modèles adaptés aux cas expérimentaux. Nous confirmons la capacité d'un ANN à prédire la température intérieure d'un bâtiment résidentiel et validons le choix des entrées, l'architecture et la méthode d'apprentissage de notre ANN.

3.4 Évaluation du caractère adaptatif du réseau de neurones artificiels : Évaluation des prédictions sur les cas simulés

Méthode Nous définissons les cent-vingt cas simulés selon trois caractéristiques : la superficie du bâtiment, la classe énergétique du bâtiment et le climat auquel il est soumis. Le chapitre 2 présente en détail les 120 cas simulés. La méthode utilisée pour adapter le ANN développé précédemment aux cas simulés est schématisée sur la figure 3.11. Pour chaque cas nous détectons le début de la période de chauffe, c'est-à-dire l'instant à partir duquel la puissance de chauffage est supérieure à 2kW. Cette étape a pour but de garantir que les deux semaines de données utilisées pour l'apprentissage sont bien prélevées sur une période où le chauffage est allumé. À partir du début de la saison de chauffe, nous créons un jeu de données correspondant aux entrées et sorties de notre ANN. Ce jeu de données est séparé en un jeu d'apprentissage, les deux premières semaines de données et en un jeu de test, le reste des données. Ensuite, nous réalisons l'apprentissage. Afin de garantir la qualité de l'apprentissage nous vérifions que la RMSE sur le jeu d'apprentissage est inférieure à 0,5°C, si c'est le cas nous réalisons les prédictions sur l'ensemble du jeu de test. Ces prédictions sont comparées aux valeurs fournies par la simulation.

Influence de la puissance de chauffage consommée sur l'erreur de prédiction. Nous représentons, comme pour les cas expérimentaux, l'erreur ε en fonction

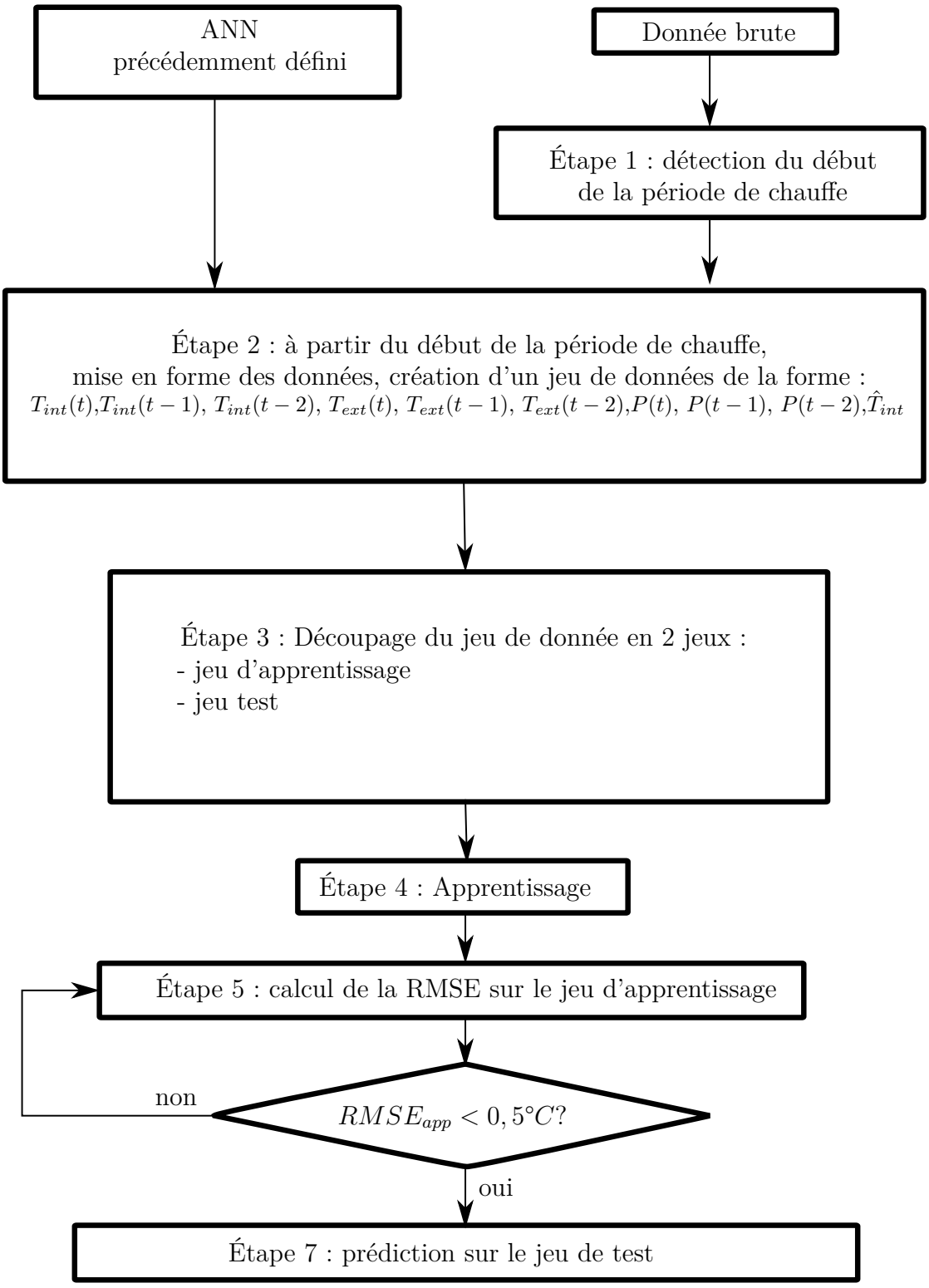


FIGURE 3.11 – Méthode utilisée pour adapter le ANN précédemment développé aux cas simulés.

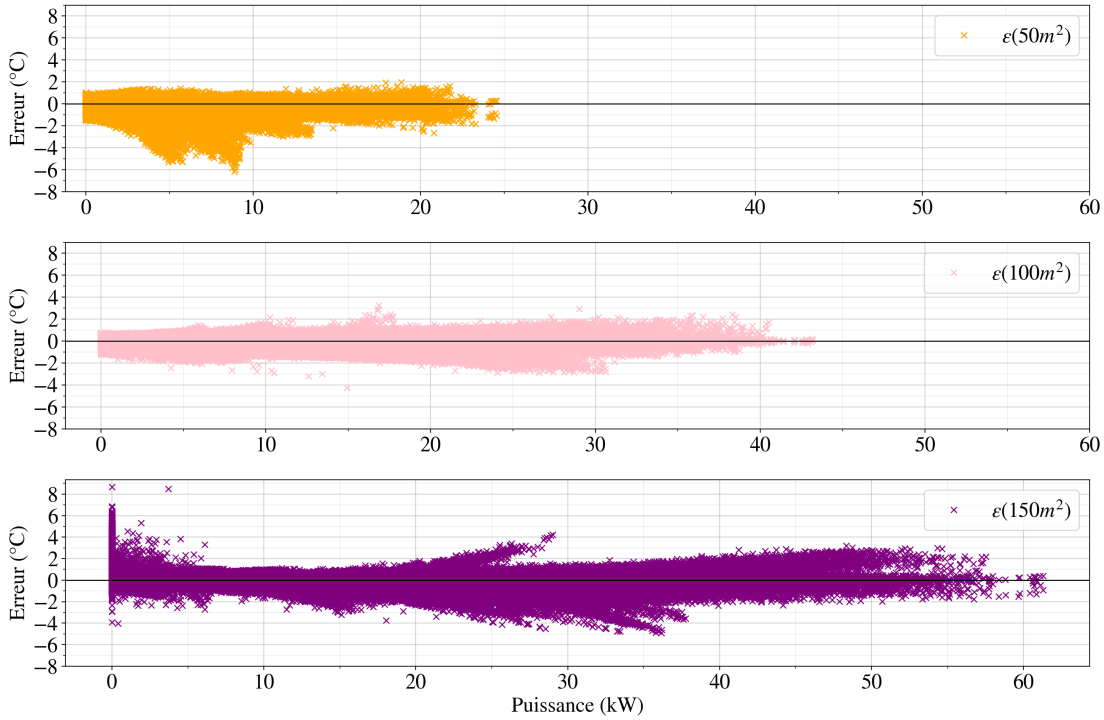


FIGURE 3.12 – Influence de la puissance de chauffage consommée, données regroupées selon la superficie du bâtiment. Sur les 3 graphiques, l'abscisse représente la puissance de chauffage consommée en kW, l'ordonnée représente l'erreur faite sur T_{int} en °C. Les graphiques représentent l'erreur ε faite sur toutes les prédictions des 120 cas simulés, soit 561 174 prédictions. Les données sont regroupées selon la superficie du cas : $50m^2$ sur la $1_{ére}$ ligne, $100m^2$ sur la $2^{éme}$ ligne et $150m^2$ sur la $3^{éme}$ ligne.

de la puissance de chauffage consommée. Plutôt que de représenter séparément les 120 cas, ce qui serait difficile à interpréter, nous regroupons les cas selon les critères dont nous cherchons à montrer l'influence sur nos modèles : la superficie 3.12, le climat 3.14 et la classe énergétique 3.13.

Influence d'une puissance de chauffage nulle. Contrairement aux cas expérimentaux, pour les cas simulés, les prédictions ne sont faites que sur la période de chauffe. Ceci explique que l'on ne voit pas de forte augmentation de l'erreur lorsque la puissance de chauffage est nulle, il y a très peu de prédiction réalisée dans cette situation. Nous observons tout de même cette augmentation sur le climat H3 et la superficie de $150m^2$. Le climat H3 est le climat le plus chaud, il y a donc davantage de période où le chauffage n'est pas nécessaire pour maintenir le confort thermique au cours de la saison de chauffe. De plus la saison de chauffe s'achève plus tôt, les simulations s'arrêtant toutes à la même date, les simulations sous le climat H3 comportent davantage de données où la puissance de chauffage P est nulle. De la même façon une habitation avec une plus grande superficie au sol, est constituée de plus de murs qu'une habitation plus petite. L'augmentation de la masse de murs, implique une augmentation de l'inertie thermique de l'habitat pouvant expliquer

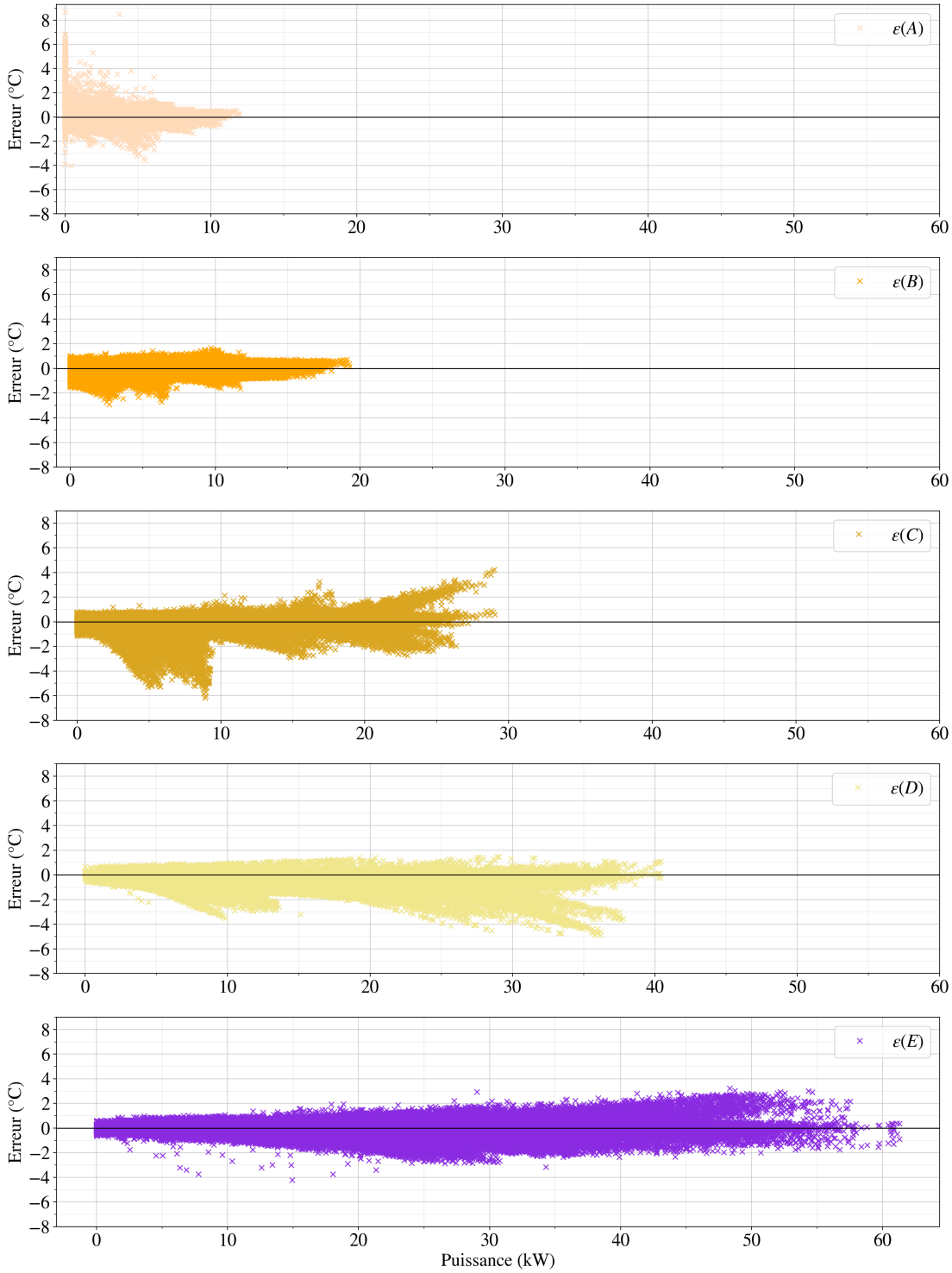


FIGURE 3.13 – Influence de la puissance de chauffage consommée, données regroupées selon la classe énergétique du bâtiment. Sur les 5 graphiques, l'abscisse représente la puissance de chauffage consommée en kW, l'ordonnée représente l'erreur en °C. Les graphiques représentent l'erreur ε faite sur toutes les prédictions des 120 cas simulés, soit 561 174 prédictions. Les données sont regroupées selon la classe énergétique du cas : A sur la $1_{ére}$ ligne, B sur la $2^{éme}$ ligne, C sur la $3^{éme}$ ligne, D sur la $4^{éme}$ ligne, E sur la $5^{éme}$ ligne.

qu'il y ait pour les cas de $150m^2$ plus de prédictions faites pour une puissance de chauffage nulle, que pour les cas de $50m^2$ et $100m^2$.

Influence de la puissance de chauffage consommée sur l'erreur de prédiction : augmentation de la puissance de chauffage. Dans aucune des situations représentées, huit climats sur la figure 3.14, trois superficies sur la figure 3.12, cinq classes énergétiques sur la figure 3.13, nous n'observons une diminution de l'écart à la valeur réelle lorsque la puissance de chauffage augmente, comme c'était le cas sur les cas expérimentaux. Au contraire, nous observons des tendances locales où l'erreur augmente, en valeur absolue, avec l'augmentation de la puissance de chauffage consommée P, notament sur les courbes représentant les cas de $50m^2$ et $150m^2$, de classe énergétique C, D et E et soumis au climat H2c et H2d. Pour le reste des cas, les courbes sont globalement constantes.

Influence de la puissance de chauffage consommée sur l'erreur de prédiction : lien avec le jeu d'apprentissage La première raison qui pourrait expliquer la variation entre les observations faite sur les cas expérimentaux et sur les cas simulés est la composition du jeu d'apprentissage. Pour les cas expérimentaux, nous réalisons les jeux d'apprentissages à partir des premières données dont nous disposons, les deux premières semaines de janvier. Par contre, pour les cas simulés, nous avons fait le choix de commencer les simulation en début de période de chauffe. Les jeux d'apprentissage sont donc, selon les cas, composés de données recueillies en octobre, novembre ou décembre. Indépendamment du bâtiment, de ses habitants ou du climat, les besoins de chauffage sont plus importants en janvier qu'en début de saison de chauffe. Nous représentons sur les figures 3.15 et 3.16 sous la forme d'histogramme la répartition des exemples d'apprentissage et de test pour les deux cas expérimentaux. Pour le cas 1, figure 3.15, nous observons que les exemples d'apprentissage ont une puissance de chauffage comprise entre 10kW et 16kW, alors que les exemples de test se situent entre 0kW et 18kW. Cela explique que l'erreur soit inférieure pour la gamme de puissance supérieure. Pour le cas 2, figure 3.16, nous observons un très grand nombre d'exemples d'apprentissage et de test dont la puissance de chauffage est nulle, et ce malgré la suppression des exemples hors période de chauffe. Cela s'explique par le fonctionnement du système de chauffage, ce dernier alterne période de chauffe et période de non chauffe afin de profiter de l'inertie du bâtiment. En dehors des puissances nulles de chauffage les jeux d'apprentissage se répartissent entre 3kW et 6kW, nous observons que très peu d'exemples d'apprentissage ont une puissance de chauffage comprise entre 1kW et 3kW. Les exemples de test ont une puissance comprise entre 0kW et 8kW. Le manque d'exemples d'apprentissage avec des valeurs faible de puissance de chauffage, mais non-nulle, explique la diminution de l'erreur de prédiction avec l'augmentation de la puissance de chauffage. De la même manière, nous représentons sur la figure 3.17, sous la forme d'histogramme, la répartition des exemples d'apprentissage et de test pour les 120 cas simulés. Nous observons que les exemples d'apprentissage ont une puissance de chauffage comprise entre 0kW et 30kW alors que les exemples de test sont répartis entre 0kW et 60kW. Toutefois, les répartitions des exemples sur ces plages de puissance sont très similaires, et les fortes puissances représentent pour le jeu d'apprentissage, comme pour le jeu

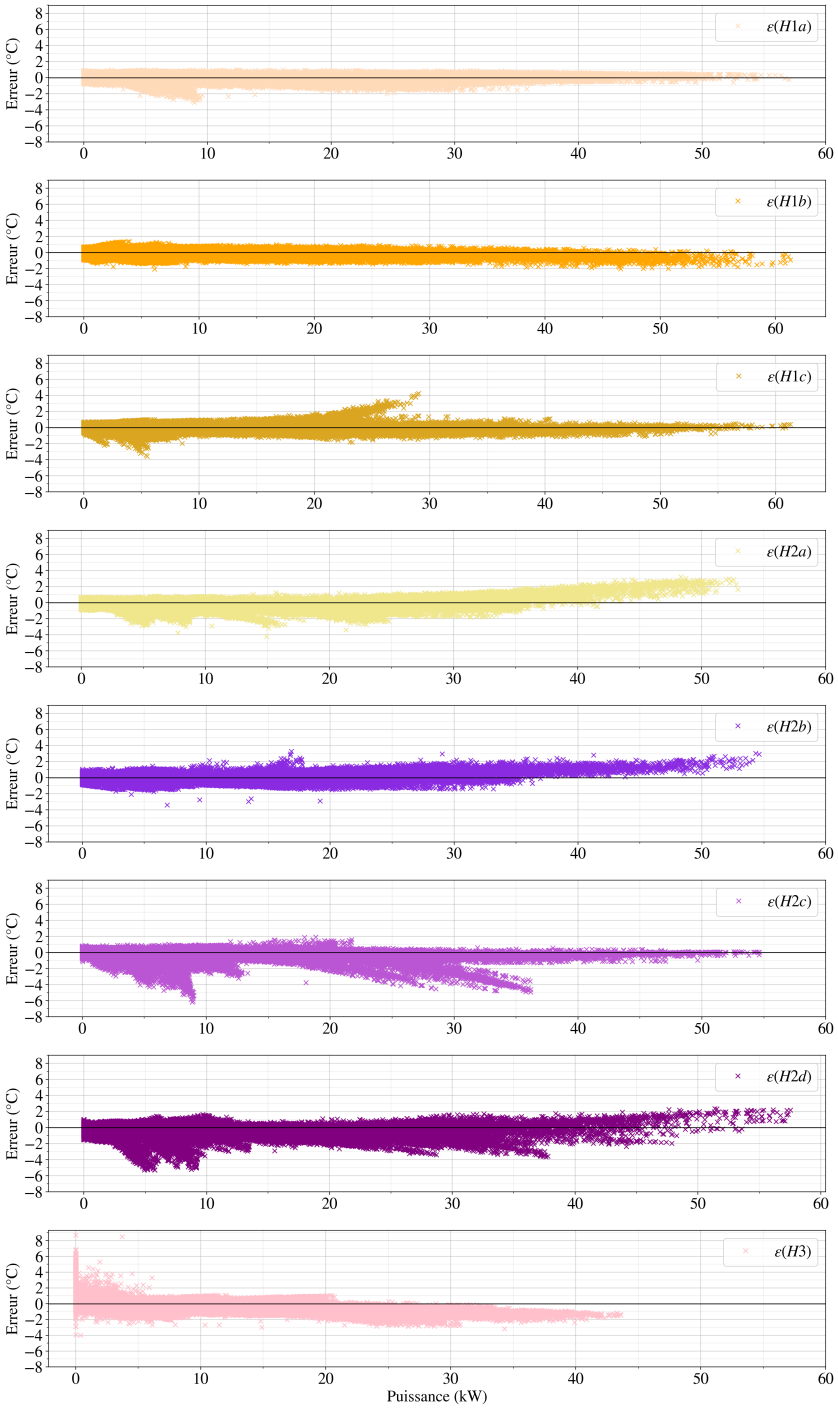


FIGURE 3.14 – Influence de la puissance de chauffage consommée, données regroupées selon le climat auquel est soumis le bâtiment. Sur les 8 graphiques, l'abscisse représente la puissance de chauffage consommée en kW, l'ordonnée représente l'erreur en °C. Les graphiques représentent l'erreur ε faite sur toutes les prédictions des 120 cas simulés, soit 561 174 prédictions. Les données sont regroupées selon le climat du cas : H1a sur la $1_{ére}$ ligne, H1b sur la $2^{éme}$ ligne, H1c sur la $3^{éme}$ ligne, H2a sur la $4^{éme}$ ligne, H2b sur la $5^{éme}$ ligne, H2c sur la $6^{éme}$ ligne, H2d sur la $7^{éme}$ ligne, H3 sur la $8^{éme}$ ligne.

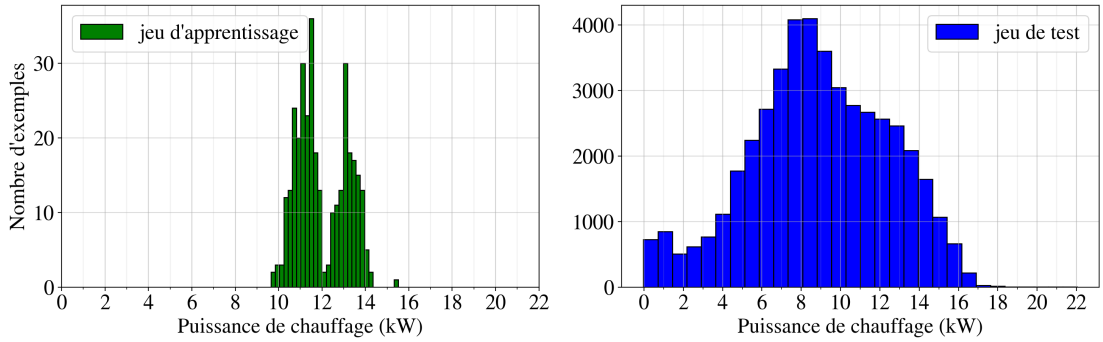


FIGURE 3.15 – Cas 1 : répartition des exemples d'apprentissage et de test en fonction de la puissance de chauffage. En ordonnée le nombre d'exemples, en abscisse la puissance de chauffage, en kW. À droite le jeu d'apprentissage, à gauche le jeu de test.

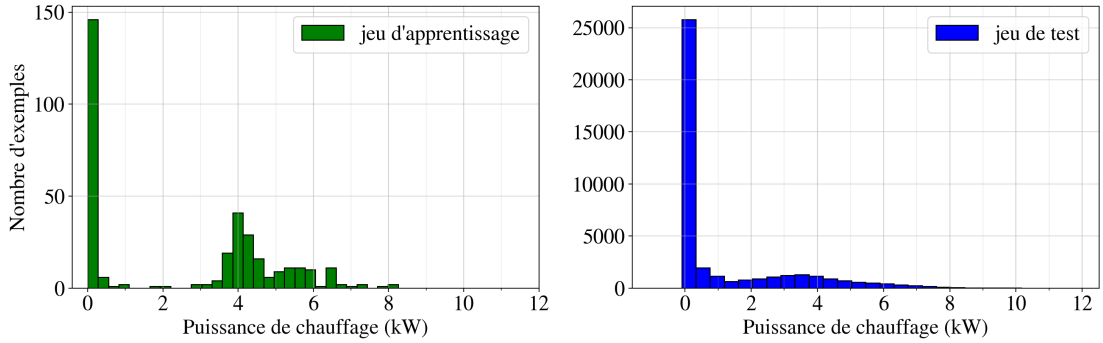


FIGURE 3.16 – Cas 2 : répartition des exemples d'apprentissage et de test en fonction de la puissance de chauffage. En ordonnée le nombre d'exemples, en abscisse la puissance de chauffage, en kW. À droite le jeu d'apprentissage, à gauche le jeu de test.

de test une faible partie des jeux de données. Ces observations sont cohérentes avec celles faites sur les figures 3.13, 3.14 et 3.12, où nous observons, dans certains cas, une augmentation de l'erreur de prédiction lorsque la puissance de chauffage augmente, expliquée par la faibles représentation des fortes puissances de chauffage dans le jeu d'apprentissage, par rapport au jeu de test. Cependant, ces fortes puissances restent minoritaires dans le jeu d'apprentissage comme dans le jeu de test, ce qui explique les tendances globalement constantes observées sur ces figures. Pour les cas expérimentaux, comme pour les cas simulés, il faut remarquer que l'erreur de prédiction faite sur des valeurs de puissance de chauffage non représentées dans le jeu d'apprentissage, bien que supérieure aux valeurs de puissance représentées, reste faible.

Influence des occupants. La dépendance de l'erreur de prédiction au jeu d'apprentissage explique très bien les observations faites à la fois sur les cas expérimentaux et sur les cas simulés. Toutefois une autre cause peut expliquer ces variations : le comportement des occupants. Les données obtenues par simulation ne permettent pas de simuler le comportement de l'habitant. Le fait que la diminution de l'erreur avec l'augmentation de la puissance de chauffage soit clairement visible sur les cas

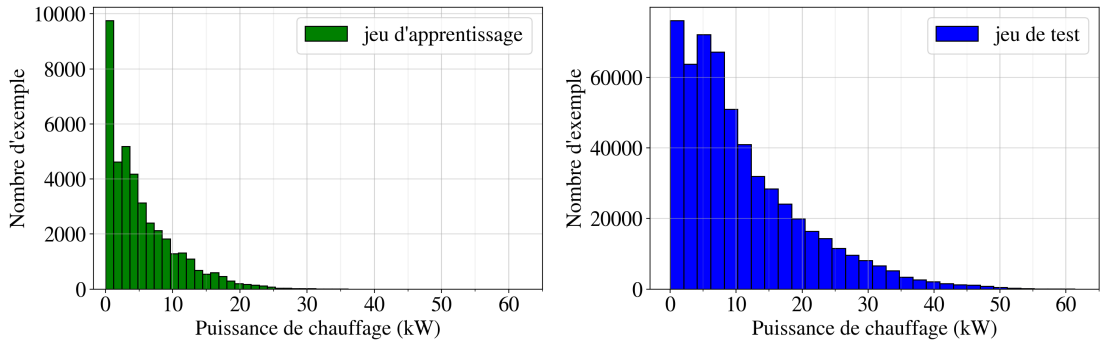


FIGURE 3.17 – 120 cas simulés : répartition des exemples d'apprentissage et de test en fonction de la puissance de chauffage. En ordonnée le nombre d'exemples, en abscisse la puissance de chauffage, en kW. À droite le jeu d'apprentissage, à gauche le jeu de test.

expérimentaux, mais pas sur les cas simulés, laisse penser que le comportement de l'habitant joue un rôle dans ce lien. En effet, l'augmentation de la puissance de chauffage consommée peut être due à une demande de température de consigne plus élevée de la part des occupants. Elle peut aussi être due à un rafraîchissement des températures extérieures, ce qui pourrait entraîner un changement de comportement des occupants. Ces propositions ne peuvent être vérifiées à l'aide des données que nous possédons, il faudrait mener une campagne de mesures in-situ où le comportement des habitants serait pris en compte afin de pouvoir affirmer ou infirmer ces propositions.

Diagramme en boite. Afin de représenter la distribution de l'erreur de prédiction sur les cas simulés nous représentons les diagrammes boite pour chaque critère superficie, climat, classe énergétique. Un diagramme boite est constitué d'un rectangle, dont la limite basse est la premier quartile $Q1$ et la limite haute est le troisième quartile $Q3$. La ligne à l'intérieur de la boite représente la médiane. Les traits au-dessus et en dessous du rectangle représentent 1,5 fois l'écart interquartile, $1,5(Q3 - Q1)$. Les valeurs au-delà de ces limites sont considérées comme atypiques car peu représentées.

Performance globale Le tableau 3.4 regroupe les informations statistiques de l'erreur signée et de l'erreur non signée faites par les 120 modèles sur les 561 174 prédictions réalisées, indépendamment des caractéristiques de superficie, de climat et de classe énergétique. Les modèles développés ont une valeur de RMSE de 0,44°C, et une valeur $\bar{\varepsilon}$ de -0,11°C. Ces valeurs moyennes très faibles indiquent de bonnes performances des modèles. Nous représentons sur la figure 3.18 le diagramme boite pour l'ensemble des valeurs de ε des prédictions faites sur les jeux de test des 120 cas simulés. Nous observons que la médiane et la moyenne sont très proches, indiquant une répartition équilibrée autour de la moyenne. Pour la majorité des prédictions testées l'erreur est comprise entre -1,5°C et +1,3°C, de plus la moitié des prédictions ont une erreur comprise entre -0,5°C et +0,2°C. Les valeurs extrêmes de -6,17°C et 8,67°C d'erreur ne sont évidement pas fiables car trop éloignées de la valeur vraie. Toutefois, nos modèles fournissent des prédictions fiables dans la grande majorité des

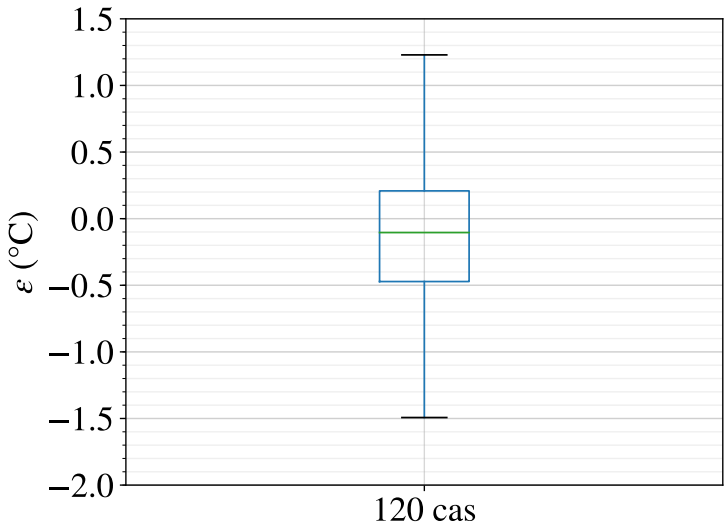


FIGURE 3.18 – ε : Diagramme en boite de l'erreur signée ε en ordonnée, en °C, pour l'ensemble des cas, en abscisse.

cas testés. Nous observons également que les modèles surestiment la température, ce qui favorise le confort thermique au détriment des économies d'énergie.

Superficie	120 cas
Nombre de prédictions	561 174
$RMSE$	0,44
$\sigma(RMSE)$	0,45
$min(RMSE)$	$< 0,01$
$max(RMSE)$	8,67
$\bar{\varepsilon}$	-0,17
$\sigma(\varepsilon)$	0,60
$min(\varepsilon)$	-6,17
$max(\varepsilon)$	8,67

TABLE 3.4 – Valeur de l'erreur moyenne signée $\bar{\varepsilon}$ et non signée $RMSE$, ainsi que de l'écart-type σ et des extremum min et max associés pour les 120 cas.

Influence de la superficie Nous regroupons les erreurs, $RMSE$ et ε faites sur chaque prédiction en trois groupes, selon la superficie du cas. Nous représentons le diagramme en boite de l'erreur signée ε pour chacun de ces trois groupes de données sur la figure 3.19. En complément le tableau 3.5 indique, pour les trois catégories, l'erreur moyenne signée $\bar{\varepsilon}$ et non signée $RMSE$, ainsi que l'écart-type σ et les extremum min et max associés. Pour les trois catégories, la majorité des valeurs de l'erreur ε sont comprises entre -1,5°C et 1,3°C. Les bornes de cet intervalle varient au maximum de 0,3°C selon la superficie des cas considérés. Les médianes sont comprises entre -0,2°C et 0°C. Nous n'observons pas d'influence de la superficie

du cas étudié sur la distribution de l'erreur de prédiction faite par les modèles. Le tableau 3.5 nous permet de confirmer cette observation. L'erreur moyenne non signée varie au maximum de 0,08°C entre les catégories étudiées. Les valeurs moyennes de l'erreur signée $\bar{\varepsilon}$ sont très proches des médianes lues sur la figure 3.19, ce qui indique une répartition équilibrée autour de la moyenne. Nous observons également, pour les trois catégories, que la température intérieure est en moyenne sur-estimée, car $\bar{\varepsilon}$ est négative. Nous concluons que la superficie du cas étudié n'influence pas la précision des modèles. Cela signifie que les entrées choisies pour le ANN, à savoir les températures intérieures et extérieures aux instants précédents et la puissance de chauffage consommée aux instants précédents suffisent à mesurer l'impact de la superficie du bâtiment sur la variation de températures intérieures.

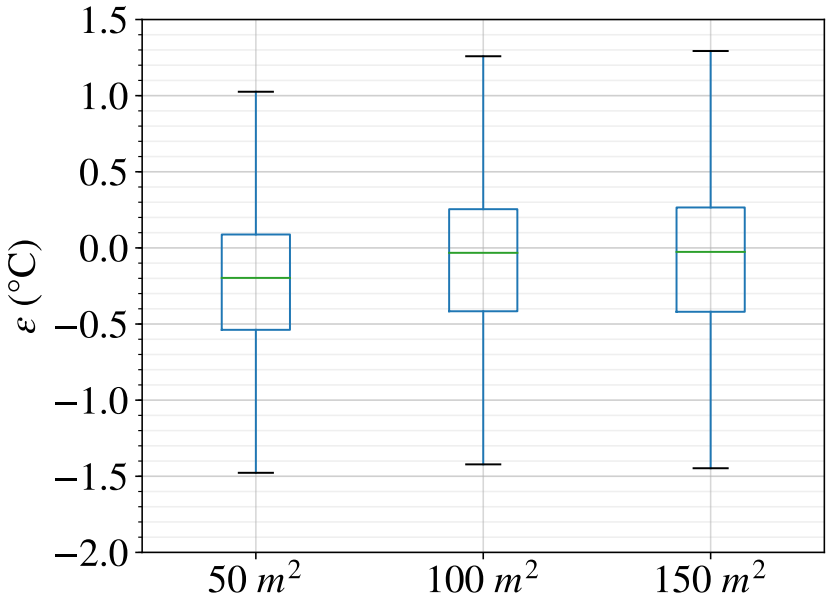


FIGURE 3.19 – ε : influence de la superficie. Diagramme en boite de l'erreur signée ε en ordonnée, en °C, en fonction de la superficie du cas, en abscisse.

Influence du climat Nous regroupons les erreurs, $RMSE$ et ε faites sur chaque prédiction en huit groupes, selon le climat auquel le cas est soumis. Nous représentons le diagramme en boite de l'erreur signée ε pour chacun de ces huit groupes de données sur la figure 3.20. En complément le tableau 3.6 indique, pour les huit catégories, l'erreur moyenne signée $\bar{\varepsilon}$ et non signée $RMSE$, ainsi que l'écart-type σ et les extremum min et max associés. Pour les trois catégories, la majorité des valeurs de l'erreur ε sont comprises entre -1,5°C et 1,3°C. Les bornes de cet intervalle varient au maximum de 0,3°C selon la superficie des cas considérés. Les médianes sont comprises entre -0,3°C et 0,1°C. Nous n'observons pas d'influence claire du climat du cas étudié sur la distribution de l'erreur de prédiction faite par les modèles. Le climat H2d semble donner de moins bons résultats avec la RMSE la plus élevé à 0,59°C. La RMSE augmente légèrement lorsque que le climat étudié s'adoucit,

Superficie	50	100	150
Nombre de prédictions	183 843	189 000	188 331
$RMSE$	0,45	0,40	0,48
$\sigma(RMSE)$	0,50	0,31	0,50
$min(RMSE)$	$< 0,01$	$< 0,01$	$< 0,01$
$max(RMSE)$	6,17	4,22	8,67
$\bar{\varepsilon}$	-0,30	-0,11	-0,10
$\sigma(\varepsilon)$	0,61	0,50	0,69
$min(\varepsilon)$	-6,17	-4,22	-4,90
$max(\varepsilon)$	1,96	3,26	8,67

TABLE 3.5 – Valeur de l'erreur moyenne signée $\bar{\varepsilon}$ et non signée $RMSE$, ainsi que de l'écart-type σ et des extremum min et max associés pour les trois superficies étudiées $50m^2, 100m^2, 150m^2$

cependant cette augmentation est faible avec une différence avec maximum de 0,24°C entre le climat H2h et H1a. L'erreur moyenne non signée varie au maximum de 0,42°C. Les valeurs moyennes de l'erreur signée $\bar{\varepsilon}$ sont très proches des médianes lues sur la figure 3.20, ce qui indique une répartition équilibrée autour de la moyenne. Nous observons également que pour les huit catégories la température intérieure en moyenne est surestimée, car $\bar{\varepsilon}$ est négative. Nous concluons que le climat du cas étudié n'influence pas la précision des modèles. Cela signifie que les entrées choisies pour le ANN, à savoir les températures intérieures et extérieures aux instants précédent et la puissance de chauffage consommée aux instants précédents suffisent à mesurer l'impact du climat sur la variation de températures intérieures.

	H1a	H1b	H1c	H2a	H2b	H2c	H2d	H3
Nombre de prédictions	71 150	71 258	71 258	70 722	70 305	69 845	70 226	66 410
$RMSE$	0,35	0,34	0,35	0,45	0,41	0,49	0,59	0,56
$\sigma(RMSE)$	0,28	0,24	0,29	0,40	0,31	0,59	0,62	0,59
$min(RMSE)$	$< 0,01$	$< 0,01$	$< 0,01$	$< 0,01$	$< 0,01$	$< 0,01$	$< 0,01$	$< 0,01$
$max(RMSE)$	3,10	2,06	4,24	4,22	3,40	6,17	5,30	8,57
$\bar{\varepsilon}$	-0,09	-0,08	-0,03	-0,22	0,02	-0,31	-0,40	-0,23
$\sigma(\varepsilon)$	0,44	0,40	0,46	0,56	0,52	0,70	0,76	0,78
$min(\varepsilon)$	-3,10	-2,06	-3,58	-4,22	-3,40	-6,17	-5,30	-4,01
$max(\varepsilon)$	0,99	1,38	4,24	3,22	3,26	1,96	2,36	8,67

TABLE 3.6 – Valeur de l'erreur moyenne signée $\bar{\varepsilon}$ et non signée $RMSE$, ainsi que de l'écart-type σ et des extrémum min et max associés pour les cinq climats étudiées A, B, C, D et E

Influence de la classe énergétique Nous regroupons les erreurs, $RMSE$ et ε faites sur chaque prédiction en cinq groupes, selon la classe énergétique du cas. Nous représentons le diagramme en boite de l'erreur signée ε pour chacun de ces trois groupes de données sur la figure 3.21. En complément le tableau 3.7 indique,

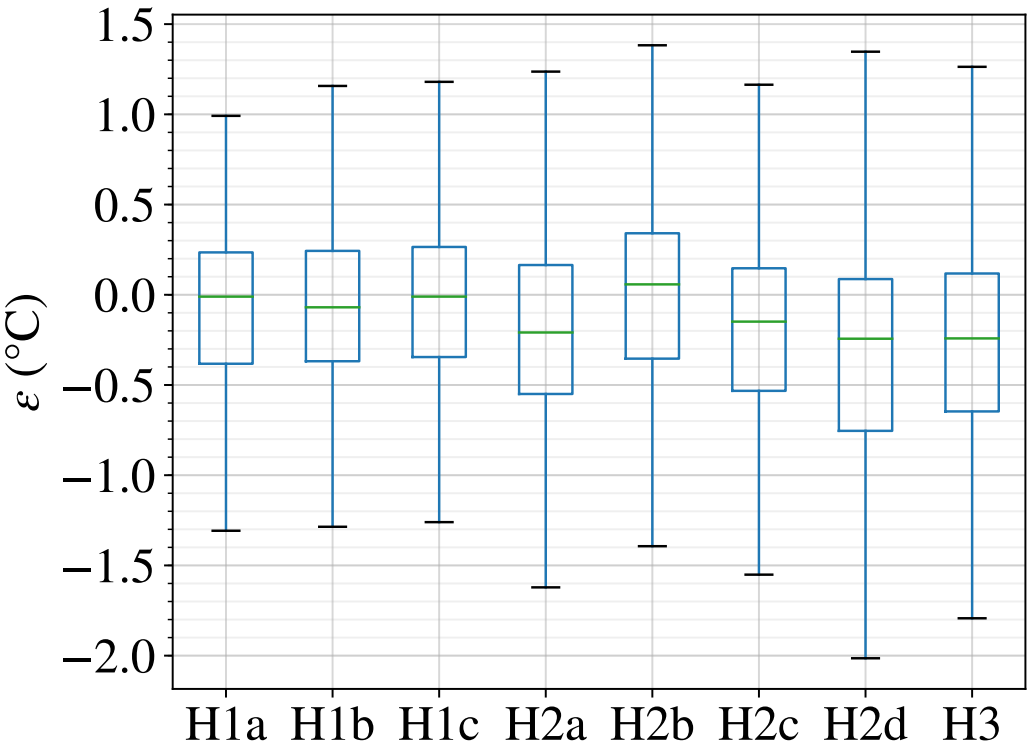


FIGURE 3.20 – ε : influence du climat. Diagramme en boite de l'erreur signée ε en ordonnée, en °C, en fonction du climat du cas, en abscisse.

pour les trois catégories, l'erreur moyenne signée $\bar{\varepsilon}$ et non signée $RMSE$, ainsi que l'écart-type σ et les extremum min et max associés. Pour les cinq catégories, la majorité des valeurs de l'erreur ε sont comprises entre -1,8°C et 1,3°C. Les bornes de cet intervalle varient au maximum de 1,3°C selon la classe énergétique des cas considérés. Les médianes sont comprises entre -0,4°C et 0,1°C. L'erreur moyenne non signée varie au maximum de 0,49°C entre les catégories étudiées. Nous observons davantage de variations selon la classe énergétique que selon le climat ou la superficie. Les classes énergétiques C et D présentent des résultats légèrement inférieurs aux autres classes en terme d'erreur moyenne et sont plus dispersés autour de cette moyenne. Toutefois, cette variation reste faible. De plus les prédictions faites sur des cas de classe énergétique E présentent des performances comparables à celles faites sur des cas de classe énergétique A et B, ce qui ne permet pas de donner une interprétation physique à ces variations. Nous concluons que la classe énergétique du cas étudié n'influence pas la précision des modèles. Cela signifie que les entrées choisies pour le ANN, à savoir les températures intérieures et extérieures aux instants précédents et la puissance de chauffage consommée aux instants précédents suffisent à mesurer l'impact de la classe énergétique du bâtiment sur la variation de températures intérieures.

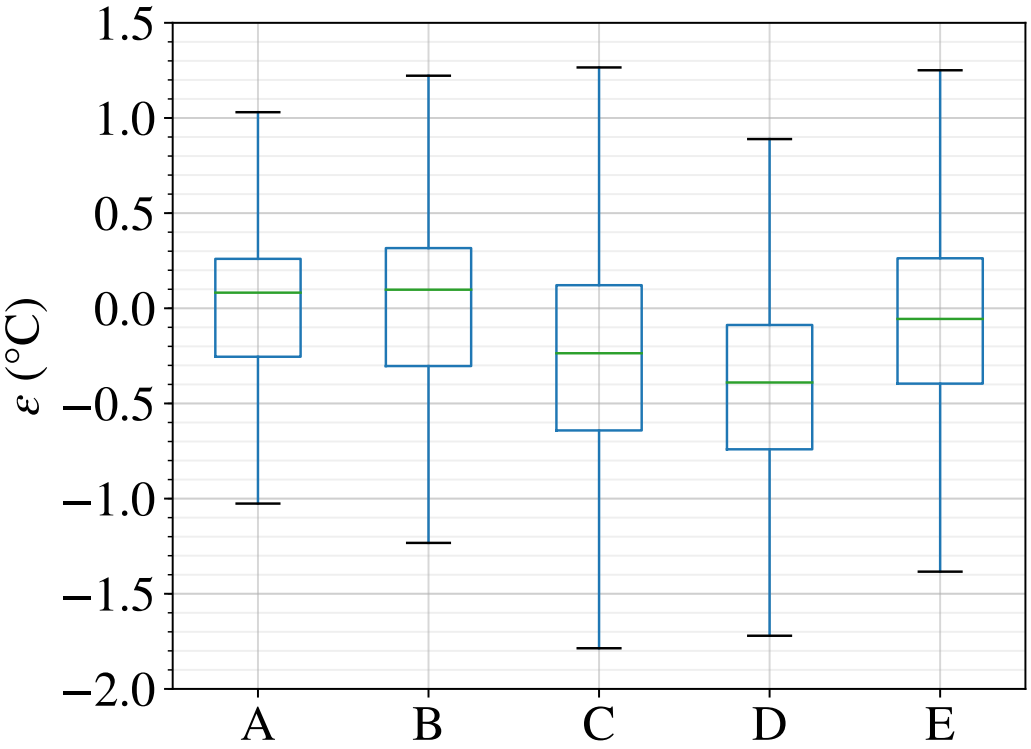


FIGURE 3.21 – ε : influence de la classe énergétique. Diagramme en boite de l'erreur signée ε en ordonnée, en °C, en fonction de la superficie du cas, en abscisse.

Classe	A	B	C	D	E
Nombre de prédictions	107 076	112 250	113 982	113 933	113 933
$RMSE$	0,34	0,33	0,55	0,54	0,44
$\sigma(RMSE)$	0,42	0,23	0,58	0,50	0,40
$min(RMSE)$	$< 0,01$	$< 0,01$	$< 0,01$	$< 0,01$	$< 0,01$
$max(RMSE)$	8,67	2,90	6,17	4,90	4,22
$\bar{\varepsilon}$	0,04	0,01	-0,34	-0,45	-0,08
$\sigma(\varepsilon)$	0,54	0,41	0,72	0,57	0,59
$\min(\varepsilon)$	-4,01	-2,90	-6,17	-4,90	-4,22
$max(\varepsilon)$	8,67	1,66	4,24	1,45	3,22

TABLE 3.7 – Valeur de l'erreur moyenne signée $\bar{\varepsilon}$ et non signée $RMSE$, ainsi que de l'écart-type σ et des extremum min et max associé pour les huit climats étudiés H1a, H1b, H1c, H2a, H2b, H2c, H2d, H3

3.5 Conclusion

Dans ce chapitre, nous définissons l'architecture et la méthode d'apprentissage choisies pour notre réseau de neurones artificiels. Notre ANN comporte neuf entrées, une couche intermédiaire de vingt neurones dont la fonction d'activation est la fonction

sigmoïde et une couche de sortie d'un unique neurone dont la fonction d'activation est la fonction identité. Nous validons l'architecture et les choix d'apprentissage sur les cas expérimentaux à l'aide des courbes d'apprentissage. Nous montrons qu'un jeu d'apprentissage composé de 2 520 exemples, soit deux semaines de données est suffisant pour permettre une adaptation de notre ANN à un cas particulier de bâtiment. Nous nous assurons que les temps de calcul nécessaires à la mise en place et à l'utilisation de ce ANN sont cohérents avec son utilisation au sein d'une régulation prédictive.

Nous évaluons les performances de notre ANN sur les deux cas expérimentaux ayant servi à son développement. Cela nous permet de confirmer la capacité d'un ANN à prédire la température intérieure d'un bâtiment résidentiel, en effet sur le premier cas expérimental le modèle commet en moyenne une erreur de 0,45°C, sur le deuxième cas expérimentale de 0,53 °C. Nous mettons en évidence que les modèles font une erreur de prédiction plus importante sur la période d'été ou le chauffage ne fonctionne pas. Ce n'est pas gênant puisque les modèles ne sont pas voués à être utilisés lorsqu'il n'y a pas besoin de chauffage. Nous observons sur les cas expérimentaux que l'erreur diminue avec l'augmentation de la puissance de chauffage. Nous expliquons ce phénomène par la dépendance des modèles à leurs jeux d'apprentissage.

Dans une deuxième partie, nous adaptons notre ANN à cent-vingt cas simulés représentatifs du parc immobilier français. Nous évaluons les performances de ces modèles en fonction de trois caractéristiques : la superficie du bâtiment, la classe énergétique du bâtiment et le climat auquel il est soumis. Nous calculons une valeur moyenne de la RMSE sur l'ensemble des cas de 0,45°C. Cette erreur en plus d'être très satisfaisant, est du même ordre de grandeur que celle calculée sur les cas expérimentaux. Les modèles des cas simulés présentent une dépendance moins marquée à la puissance de chauffage. Nous l'expliquons par la composition des jeux d'apprentissages. Les jeux d'apprentissage des modèles des cas simulé sont composés d'exemples mieux répartis sur la gamme de puissance de chauffage. Une autre explication serait que l'influence de la puissance de chauffage sur la précision des prédictions est corrélée au comportement des occupants. Etant donné que le comportement des occupants n'est pas pris en compte dans les simulations, nous ne pouvons pas observer cette influence sur les cas simulés. Nous montrons que notre ANN n'est pas sensible au type de bâtiment auquel il est adapté, c'est-à-dire que ses performances ne varient pas selon la superficie, la classe énergétique ou le climat du cas auquel nous appliquons notre ANN. L'étude ayant été menée par simulation, nous ne pouvons pas conclure que le ANN est adaptable à tout type de bâtiment. Nous proposons qu'une campagne de mesure soit menée sur des foyers type afin de valider l'adaptation du ANN à tout type d'occupants. Cependant, nous observons que l'erreur de prédiction est plus importante sur les prédictions dont la puissance de chauffage n'est pas représentée dans l'exemple d'apprentissage. Afin d'augmenter les performances déjà bonnes de notre ANN, nous proposons une amélioration de la méthode d'apprentissage : une fois plus de données disponibles, un apprentissage pourra être fait sur une gamme plus large de puissances de chauffage. Toutefois, une étude devra être menée afin de déterminer comment composer automatiquement le jeu d'apprentissage optimal pour n'importe quel bâtiment, afin de ne pas perdre le caractère généraliste de notre

ANN. Pour terminer, il sera nécessaire d'augmenter l'horizon de prédiction du ANN pour une utilisation au sein d'une régulation prédictive.

Nous montrons qu'un réseau de neurones artificiels permet de prédire la température intérieure d'un bâtiment sur un horizon d'une heure en utilisant uniquement les données usuellement disponibles sur une installation de production de chauffage. De plus, nous montrons le caractère adaptatif du réseau de neurones artificiels, puisqu'il n'est pas dépendant de la superficie, de la classe énergétique et du climat du bâtiment auquel il est adapté.

CHAPITRE 4

Prédiction des consommations d'eau chaude sanitaire

4.1 Introduction

Nous montrons dans le chapitre 1.3 l'intérêt de modèles prédictifs de l'énergie consommée pour l'eau chaude sanitaire (ECS). Encore très peu utilisés, les réseaux de neurones artificiels(ANN) nous semblent être une méthode adaptée à ce type de prédiction, car elle permet de s'adapter plus finement aux usagers que les méthodes basées sur des profils de consommation. Cependant, les entrées, et les données d'apprentissage nécessaires à ce type de prédiction sont encore mal connues. De plus, nous cherchons à développer un modèle non-invasif, capable de s'adapter à tous types de consommateurs et d'installations sans instrumentation supplémentaire. Afin de réduire la complexité de la prédiction, induite par cette limitation des choix d'entrée nous réalisons une prédiction à court terme : huit minutes. Nous posons l'hypothèse qu'une architecture unique de ANN et une méthode d'apprentissage unique peuvent modéliser l'évolution des consommations d'ECS, et que le jeu d'apprentissage permet d'adapter le ANN à plusieurs habitations. Nous parlons de "ANN" pour désigner l'architecture et la méthode d'apprentissage que nous développons et nous parlons de "modèle" lorsque un ANN est adapté à un cas, c'est-à-dire après la phase d'apprentissage. Dans la première partie de ce chapitre 4.2, nous développons trois ANN permettant de prédire l'énergie consommée pour l'ECS huit minutes après l'instant de prédiction : nous développons l'architecture et la méthode d'apprentissage associée en nous appuyant sur les données du cas 1. Ces différentes architectures permettent d'évaluer le jeu d'entrée le plus pertinent pour prédire l'énergie consommée pour l'ECS. Nous étudions également la taille du jeu d'apprentissage. Dans la deuxième partie 4.3, nous évaluons les performances des trois ANN sur le cas 1 afin d'identifier le ANN le plus performant. Dans la dernière partie de ce chapitre 4.4 nous évaluons la capacité d'adaptation à d'autres cas du ANN le plus performant, en le testant sur les cas 2 et cas 3.

Table 4.1 – Déscription des modéles

Nom du modéle	Entrées					Sortie	Architecture		
	$E(t)$	$E(t-8)$	$E(t-64)$	"$jour$"	T_{ext}	$E(t+1)$	N_i	f_i	f_S
Modèle A8	✓	✓	✓			✓	5	sigmoïde	identité
Modèle B8	✓	✓	✓	✓		✓	5	sigmoïde	identité
Modèle C8	✓	✓	✓		✓	✓	5	sigmoïde	identité

4.2 Développement de l'architecture et de la méthodologie d'apprentissage.

Le schéma proposé en figure 4.1 représente les différentes étapes mises en œuvre pour développer les trois architectures. Nous y fesons référence tout au long de cette partie.

Étape 1 : définition des ANN. La première étape de notre démarche est de définir nos ANN. Comme nous le présentons dans la partie 1.4un réseau de neurones est défini par son architecture, son nombre d'entrées, de sorties, de couches intermédiaires et de neurones sur chaque couche ainsi que par le jeu de données d'apprentissage choisi. Dans notre cas, le choix des entrées est dépendant du phénomène physique à prédire, la demande en énergie pour l'ECS. Le jeu d'apprentissage permet de s'adapter au cas d'étude.

Définition des entrées et sorties Nous développons un ANN, il prédit le besoin d'énergie pour l'ECS durant les huit minutes qui suivent l'instant de prédiction $\widehat{E}(t+8)$. Nous nous basons sur la littérature et sur nos connaissances du problème physique pour déterminer trois types d'influence dont dépend le besoin d'énergie pour l'ECS : la consommation pendant les instants précédents, la saison et le jour de la semaine. Nous choisissons de prendre en compte la consommation d'énergie sur trois instants précédents, les huit, seize et soixante-quatre minutes précédant la prédiction, respectivement $E(t-8)$,$E(t-16)$,$E(t-64)$. Nous représentons la saison par la température extérieure à l'instant de prédiction $T_{ext}(t)$. Le jour de la semaine est représenté par un entier "$jour$" dont la valeur est 1 pour lundi et 7 pour dimanche. Ce choix est fait en tenant compte des données disponibles sur une installation produisant à la fois le chauffage et l'ECS, la valeur de "$jour$" peut être facilement obtenue par un système connecté. Nous souhaitons étudier l'influence de ces trois types d'entrée sur la qualité de prédiction de notre modèle, c'est pourquoi nous développons trois modèles, chacun ayant des entrées différentes, ces entrées sont synthétisées dans le tableau 4.1.

Architecture du ANN Nous posons l'hypothèse que l'architecture permet de capter le lien entre les grandeurs physiques. C'est pourquoi nous développons l'architecture des modèles uniquement sur les données du cas 1. Il n'existe pas de méthodologie universelle pour déterminer l'architecture idéale d'un ANN. Nous avons

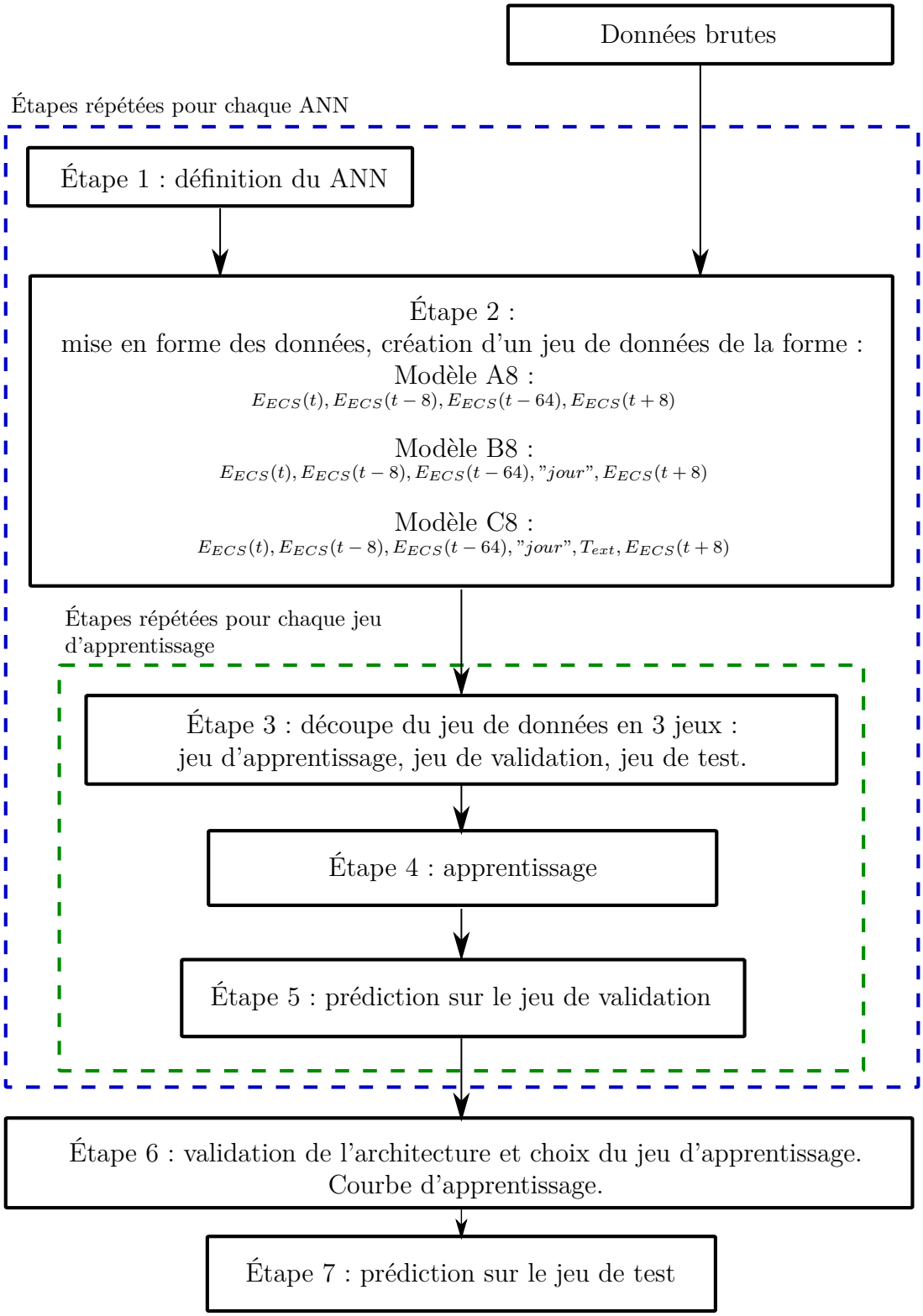


FIGURE 4.1 – Étapes de développement des modèles.

	Énergie (kWh)	"jour"	Température extérieure (°C)
minimum	0	1	-20
maximum	10	7	40

TABLE 4.2 – Valeurs utilisées pour la normalisation

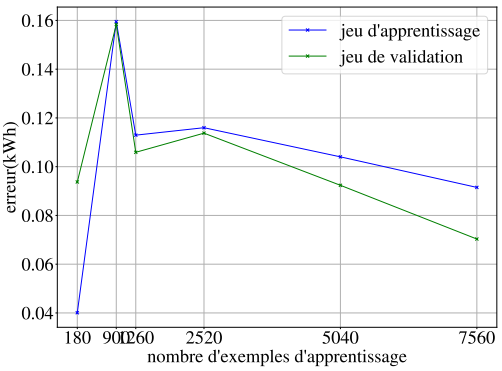
testé pour chacun de nos modèles plusieurs architectures. Nous avons simplifié ou complexifié l'architecture en fonction des observations faites sur les erreurs du jeu d'apprentissage et du jeu de test. Les étapes 1, 2, 3, 4 et 5 sont répétées pour chaque architecture testée.

Étape 2 : mise en forme des données Une fois l'architecture du ANN défini, nous mettons en forme nos données. À cette étape les données sont sous la forme d'un tableau avec en colonne les grandeurs mesurées et en ligne l'instant de mesure. Pour utiliser les données, il faut les mettre sous la forme d'exemple, c'est-à-dire que chaque ligne comporte les valeurs d'entrées du ANN et la sortie correspondante. Pour le modèle "A8", la ligne au temps t est composée de $E(t)$, $E(t-8)$, $E(t-64)$, $E(t+8)$. La forme d'un exemple est donnée pour chaque modèle sur la figure 4.1. Dans la suite, $E(t+8)$ est appelée valeur vraie de la prédiction à huit minutes, faite au temps t. L'ensemble des données disponibles sont mises sous cette forme.

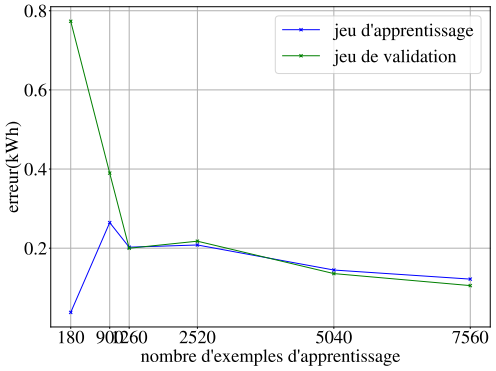
Étape 3 : création des trois jeux de données Nous venons de créer, pour chaque ANN, un jeu d'exemples correspondant à l'architecture entrée/sortie du ANN. À présent nous séparons ces exemples en trois catégories, le jeu d'apprentissage, le jeu de validation et le jeu de test. Le jeu d'apprentissage sert à l'apprentissage du ANN, c'est-à-dire à déterminer les poids du ANN en fonction du cas à modéliser. Le jeu de validation sert à valider l'architecture du ANN. Le jeu de test sert à évaluer les capacités de prédiction du modèle.
Habituellement, la majorité des données disponibles sont utilisées pour l'apprentissage et le reste pour la validation et le test du modèle, typiquement deux tiers pour l'apprentissage, un sixième pour le jeu de validation et un sixième pour le jeu de test. Or, nous souhaitons développer un ANN utilisable rapidement sur un grand nombre d'utilisateurs, autrement dit l'apprentissage doit être composé de données collectées le plus rapidement possible après l'installation du système. La littérature ne définit pas de quantité minimale de données nécessaire à un bon apprentissage dans le cas de prédictions de consommation d'énergie pour l'ECS, ni de critère de qualité des données d'apprentissage. Nous choisissons de créer plusieurs jeux d'apprentissage composés de toutes les données mesurées sur une période prédéterminée. Nous utilisons les premières données dont nous disposons pour créer cinq jeux d'apprentissage composés de 180, 900, 1 260, 2 520, 5 400 et 7 560 exemples. Soit respectivement le premier, les cinq, sept, quatorze, trente et quarante-deux premiers jours des données disponibles. Le jeu de validation est composé de 7 560 exemples, soit 1,5 mois de données. Les exemples restants composent le jeu de test, la taille de ce jeu est donc variable selon les cas d'étude.

Étape 4 : apprentissage Les jeux de données ainsi créés, nous réalisons la phase d'apprentissage. Nous nous basons sur la littérature afin de déterminer l'algorithme d'optimisation le plus adapté à l'apprentissage. Nous utilisons l'algorithme de Levenberg-Marquardt comme algorithme d'apprentissage. Lors de cette étape, nous avons défini pour chacun des cas expérimentaux les poids du ANN. Nous disposons alors de trois modèles adaptés automatiquement à notre cas.

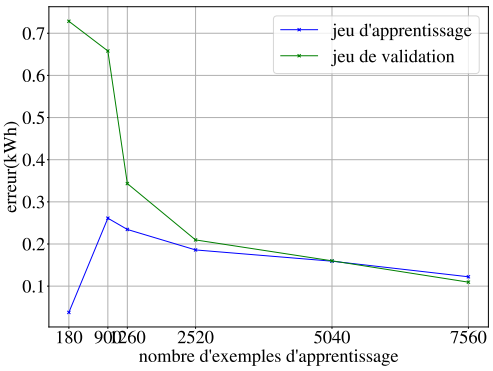
Étape 5 : prédiction sur le jeu de validation. Lors de cette étape, nous utilisons les modèles précédemment développés pour prédire l'énergie nécessaire à l'ECS sur l'ensemble des exemples composant le jeu de validation. Cette prédiction est comparée à la sortie attendue en calculant l'erreur moyenne $\bar{\varepsilon}$ sur le jeu d'apprentissage et sur le jeu de validation.



(A) Modéle A8



(B) Modéle B8



(C) Modéle C8

FIGURE 4.2 – Courbes d'apprentissage des trois modèles du cas 1.

Étape 6 : validation de l'architecture et choix du jeu d'apprentissage. Une architecture complexe (grand nombre de couches ou de neurones sur chaque couche) pourrait être trop spécifique au cas 1, sur lequel nous développons les modèles et empêcherait l'adaptation à d'autres cas d'étude. De plus une architecture complexe a de grandes chances de demander un plus grand nombre d'exemples

d'apprentissage pour ne pas être en sous-apprentissage, ou pour atteindre son erreur minimale. C'est pourquoi nous choisissons pour chaque modèle l'architecture la plus simple possible. Certaines caractéristiques de l'architecture sont communes à tous les modèles : les ANN comptent une couche intermédiaire, la fonction d'activation de cette couche intermédiaire est la fonction sigmoïde, la fonction d'activation de la couche de sortie est la fonction identité. L'utilisation de la fonction sigmoïde nécessite une normalisation des données d'entrées, voir partie 1.4, le tableau 4.2 indique les valeurs utilisées pour la normalisation. Le nombre de neurones de la couche intermédiaire est indiqué dans le tableau 4.1. Les étapes 3, 4 et 5 sont répétées pour chaque durée d'apprentissage testée. Afin d'évaluer la qualité de l'apprentissage et de déterminer la quantité minimale de données d'apprentissage, nous traçons les courbes d'apprentissage pour les trois modèles développés, représentées sur la figure 4.2. Comme expliqué dans la partie 1.4, ces courbes représentent l'erreur faite sur la sortie d'un ANN en fonction du nombre d'exemples d'apprentissage. Nous observons sur la figure 4.2 que l'erreur sur le jeu de validation est suffisamment faible pour valider le choix de l'architecture. Dans le but de comparer les modèles entre eux, une seule durée d'apprentissage est retenue pour l'ensemble des modèles. Nous observons sur la figure 4.2 que, pour la majorité des modèles, l'augmentation de la durée d'apprentissage au-delà de deux semaines (2 520 exemples) permet une diminution négligeable de l'erreur sur le jeu de validation. Nous choisissons donc d'évaluer les performances de nos modèles avec une durée d'apprentissage de deux semaines (2 520 exemples).

Étape 7 : prédiction sur le jeu de test. Afin d'évaluer les performances des modèles nous réalisons les prédictions de température sur l'ensemble des exemples du jeu de test. Cette évaluation est traitée dans la partie 4.4 de ce chapitre.

Étape	Cas 1
Mise en forme des données	210
Apprentissage	52
Prédiction sur les jeu de validation et de test	0,02

TABLE 4.3 – Temps de calcul, en seconde, des principales étapes de développement du ANN pour une machine équipée d'un processeur 3.1 GHz Intel Core i7 et de 16Go de mémoire vive.

Temps de calcul. Le réseau de neurones artificiels que nous venons de développer a pour but d'être utilisé au sein d'une régulation de système de production d'ECS. De ce fait, nous devons nous assurer que les temps de calcul nécessaires à une prédiction sont cohérents avec cette utilisation. Les étapes précédant la réalisation d'une prédiction par nos modèles sont légèrement différentes pendant la phase de développement et la phase d'utilisation. La première étape consiste à collecter deux semaines de données, ensuite les données sont mises en forme puis l'apprentissage est réalisé, le bon fonctionnement de l'apprentissage est testé via le calcul de l'erreur sur le jeu d'apprentissage. Si l'apprentissage est validé, l'exemple de données nécessaire à la prédiction est mis en forme puis la prédiction est réalisée, sinon l'apprentissage est

refait, sur les même données, jusqu'à sa validation. Le tableau 4.3 regroupe les temps de calcul des principales étapes mises en œuvre pendant la phase de développement du ANN sur les données du cas 1 . Ces temps sont indiqués pour une machine équipée d'un processeur 3.1 GHz Intel Core i7 et de 16Go de mémoire vive. Il est évident qu'un régulateur aura des temps de calcul plus longs, les renseignements fournis ici ont vocation à donner un ordre de grandeur des temps de calcul mis en œuvre. La phase la plus longue est la mise en forme des données, cette étape est réalisée pour plus d'un an de données, ce qui ne sera jamais le cas en réalité puisque les données sont mise en forme au fil de l'eau pour chaque instant de prédiction. L'étape la plus longue sera donc en réalité l'apprentissage dont la durée, pour un jeu d'apprentisssage de deux semaines (2 520 exemples), est de moins d'une minute. Après deux semaines de collecte de données, quelques minutes pour réaliser l'apprentissage est un temps plus que raisonnable. Une fois cette étape terminée, chaque prédiction, mise en forme des données comprise ne prendra que quelques secondes. Ces temps de calculs sont parfaitement acceptables pour une utilisation en temps réel.

4.3 Évaluation des capacités d'un ANN à prédire les consommations d'ECS.

4.3.1 Indicateur de performance

Nous allons maintenant nous intéresser à la manière d'évaluer les prédictions faites par les modèles. Dans cette partie, nous définissons les variables utiles qui nous permettent de comparer les performances des modèles, entre eux et vis-à-vis de l'objectif. Pour cela, nous utilisons plusieurs indicateurs :

- la moyenne de l'erreur signée ε faite sur les n exemples d'un jeu de données, défini par l'équation (4.1)

$$\bar{\varepsilon} = \frac{1}{n}\sum_{i=1}^{n} E(t_i) - \hat{E}(t_i) \tag{4.1}$$

- l'écart-type associé, défini par l'équation (4.2)

$$\sigma = \frac{1}{n}\sum_{t=1}^{n} \hat{T}(t) - \bar{T} \tag{4.2}$$

- l'erreur moyenne non signée faite sur un jeu de données composé de n exemples, c'est-à-dire la racine carré de l'erreur moyenne au carré (RMSE, de l'anglais Root Mean Square Error), défini par l'équation (4.3).

$$RMSE = \frac{1}{n}\sum_{i=1}^{n} \sqrt{\left(\hat{E}(t_i) - E(t_i)\right)^2} \tag{4.3}$$

- l'erreur relative faite sur une journée de prédiction, défini par l'équation (4.4)

$$\delta_{24h} = \frac{|\sum_{t=0}^{24h} E(t) - \sum_{t=0}^{24h} \hat{E}(t)|}{\sum_{t=0}^{24h} E(t)} \tag{4.4}$$

4.3.2 Cas 1 : évaluation des architectures et méthodes d'apprentissage.

4.3.2.1 Observations préliminaires

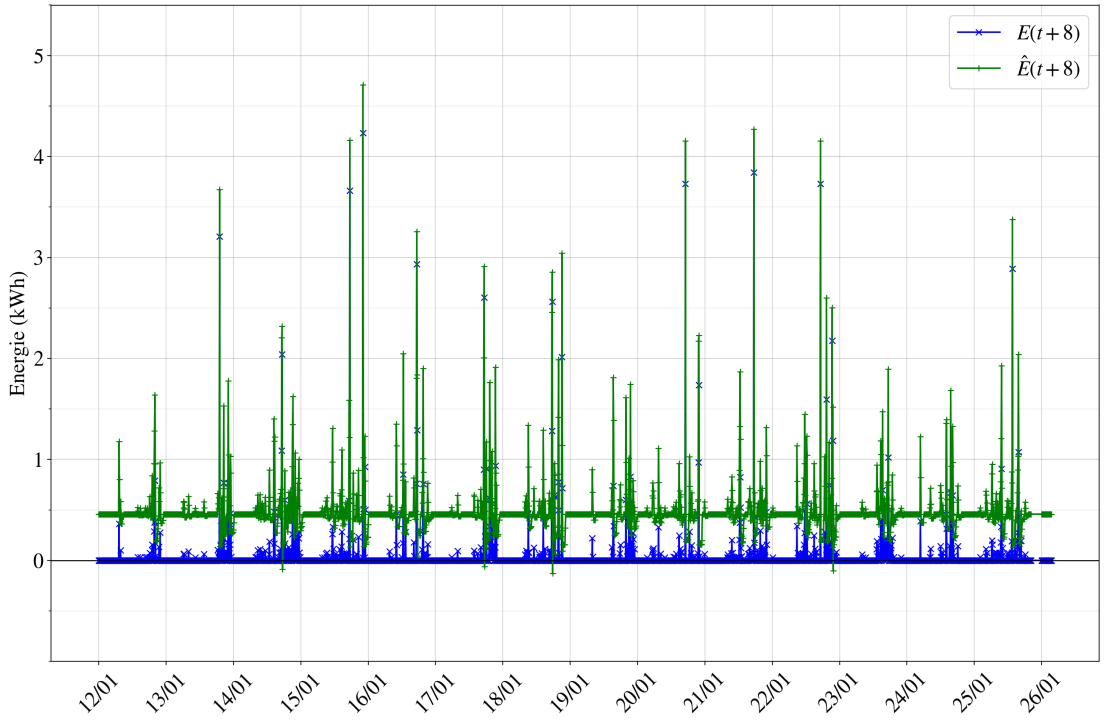
Observations Sur la figure 4.3, nous représentons l'énergie prédite et l'énergie réelle correspondante pour les prédictions du modèle A8, sur le jeu d'apprentissage et le jeu de validation. Nous faisons deux observations. D'une part, nous observons un décalage constant entre les valeurs prédites et les valeurs réelles. D'autre part le modèle prédit des valeurs d'énergie négatives, ce qui est physiquement impossible. Ces deux observations ont la même cause : lors de l'apprentissage, l'algorithme minimise l'erreur moyenne entre les prédictions et la valeur vraie sur l'ensemble des exemples composant le jeu d'apprentissage. Les données de consommation d'ECS ont la particularité d'être "en dents de scie", l'énergie consommée alterne de longues périodes nulles et des périodes plus ou moins élevées. Cette caractéristique des données fait que les deux solutions permettant d'obtenir une erreur nulle sur le jeu d'apprentissage sont soit de prédire exactement les données, soit de prédire une seule valeur pour tout le jeu correspondant à la valeur moyenne du jeu d'apprentissage. Ces observations sont moins visibles sur les prédictions faites par les modèles B8 et C8 représentées respectivement sur les figures 4.4 et 4.5. Pour ces deux modèles, nous observons également des prédictions négatives, toutefois le décalage n'est pas constant.

$$\begin{gathered} \text{Pour les } n \ t_i \text{ où } E_{app}(t_i) = 0 : \\ offset = \frac{1}{n} \sum \left(\hat{E}_{app}(t_i) - E_{app}(t_i) \right) \end{gathered} \tag{4.5}$$

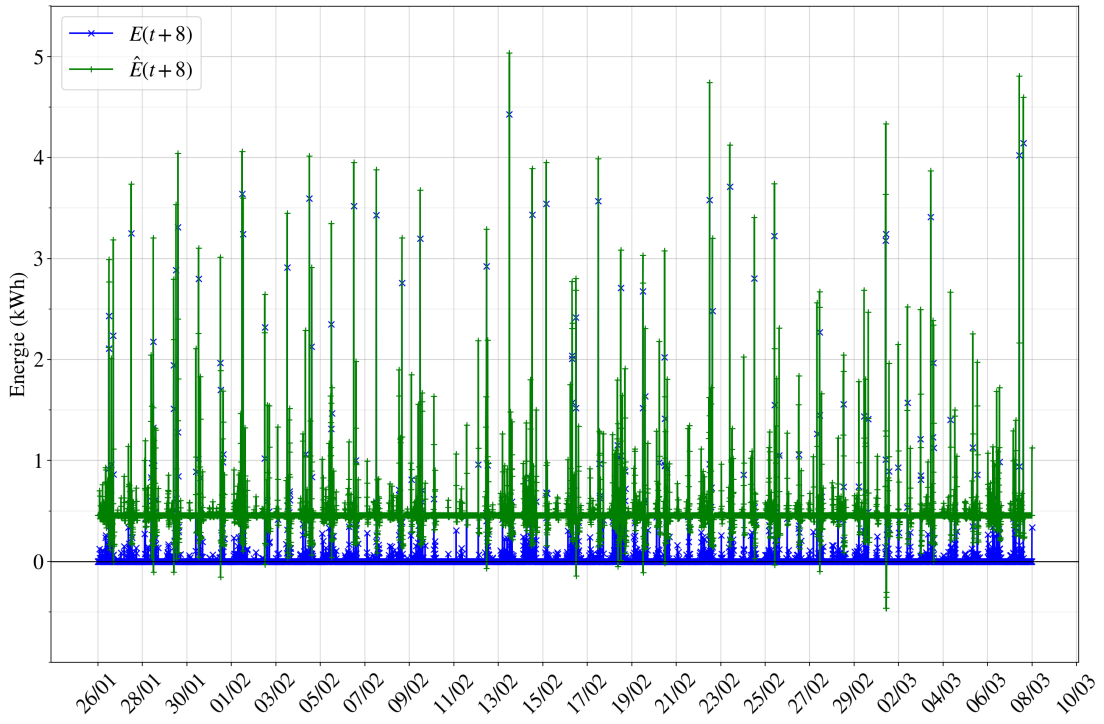
Post-traitement des prédictions. Afin de corriger les erreurs du modèle décrites dans le paragraphe précédent, nous proposons deux opérations à appliquer aux prédictions. Dans un premier temps, nous déterminons la valeur du décalage sur le jeu d'apprentissage, à l'aide de l'équation (4.5), les valeurs sont indiquées dans le tableau 4.4. Nous soustrayons cette valeur à l'ensemble des prédictions faites par la suite. Dans un deuxième temps, afin de supprimer les valeurs d'énergie négatives, nous appliquons la fonction valeur absolue aux prédictions. Cette solution est applicable en situation réelle.

TABLE 4.4 – Offset calculé sur les modèles du cas 1 pour chacun des ANN , en kWh.

Cas d'étude	A8	B8	C8
Cas 1	-0,430	-0,160	-0,399

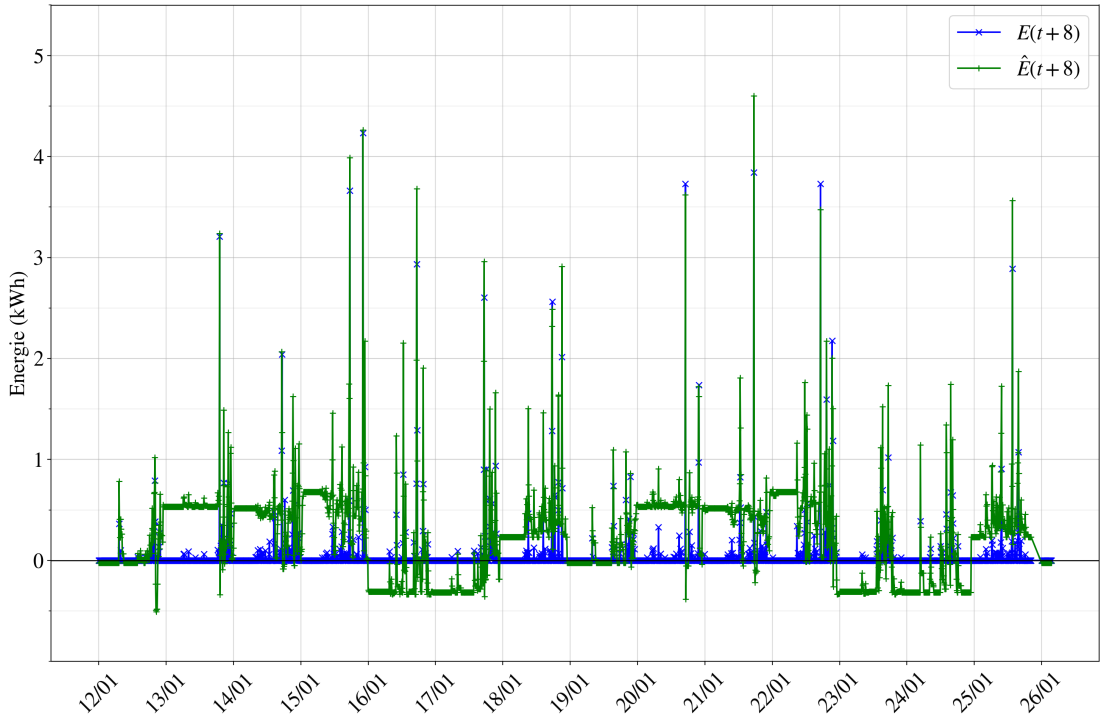


(A) jeu d'apprentissage

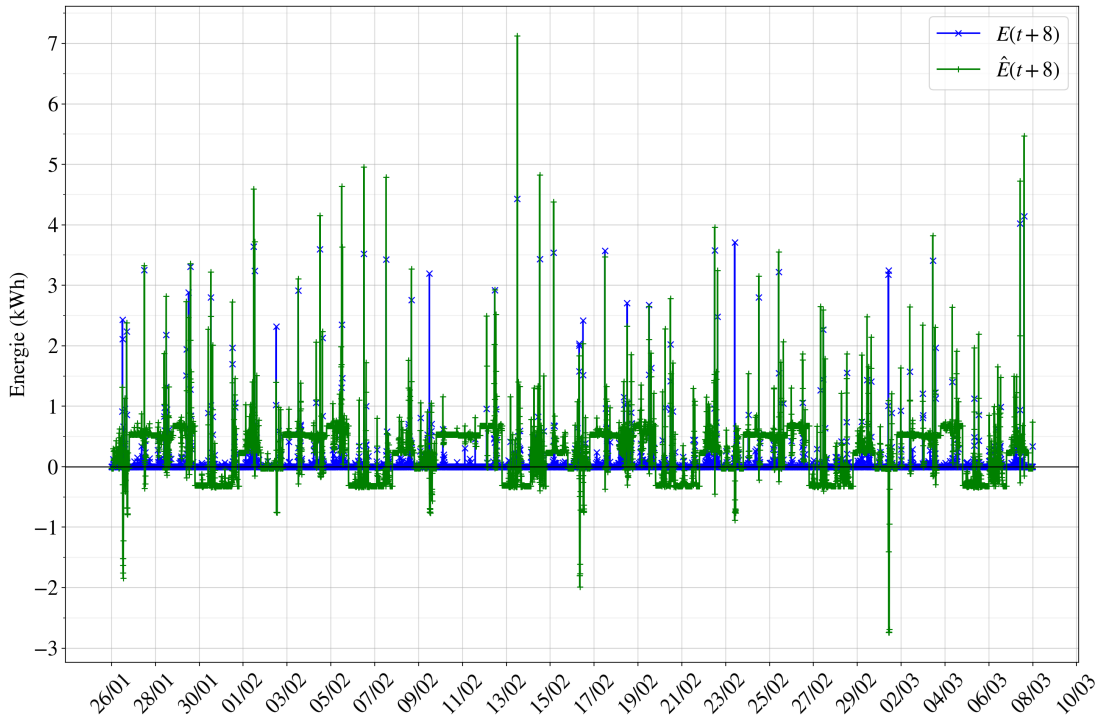


(B) jeu de validation

FIGURE 4.3 – Modèle A8 : visualisation de l'offset. Sur les 2 graphiques l'abscisse représente la date, au format jour/mois, l'ordonnée représente l'énergie en kWh. Le graphique du haut (A) représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et l'énergie prédite $\hat{E}(t+8)$ par le modèle A8 en vert, avec des marqueurs x, sur le jeu d'apprentissage. Le graphique du bas (B) représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et celle prédite $\hat{E}(t+8)$ par le modèle A8 en vert, avec des marqueurs x, sur le jeu de validation.

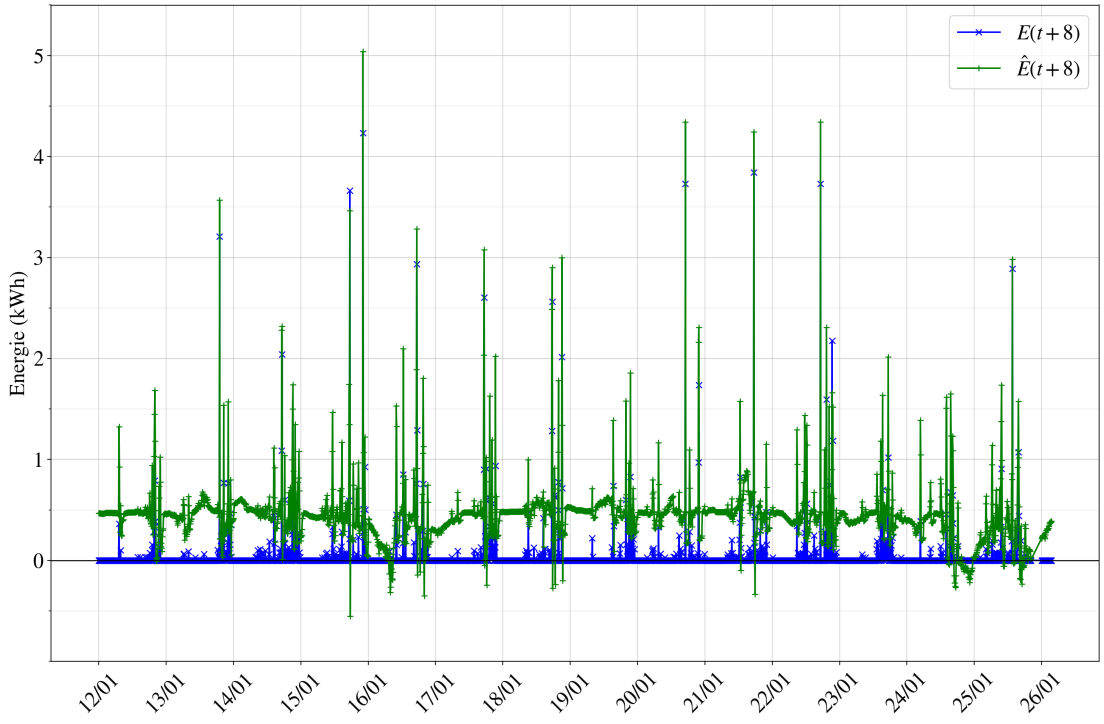


(A) jeu d'apprentissage

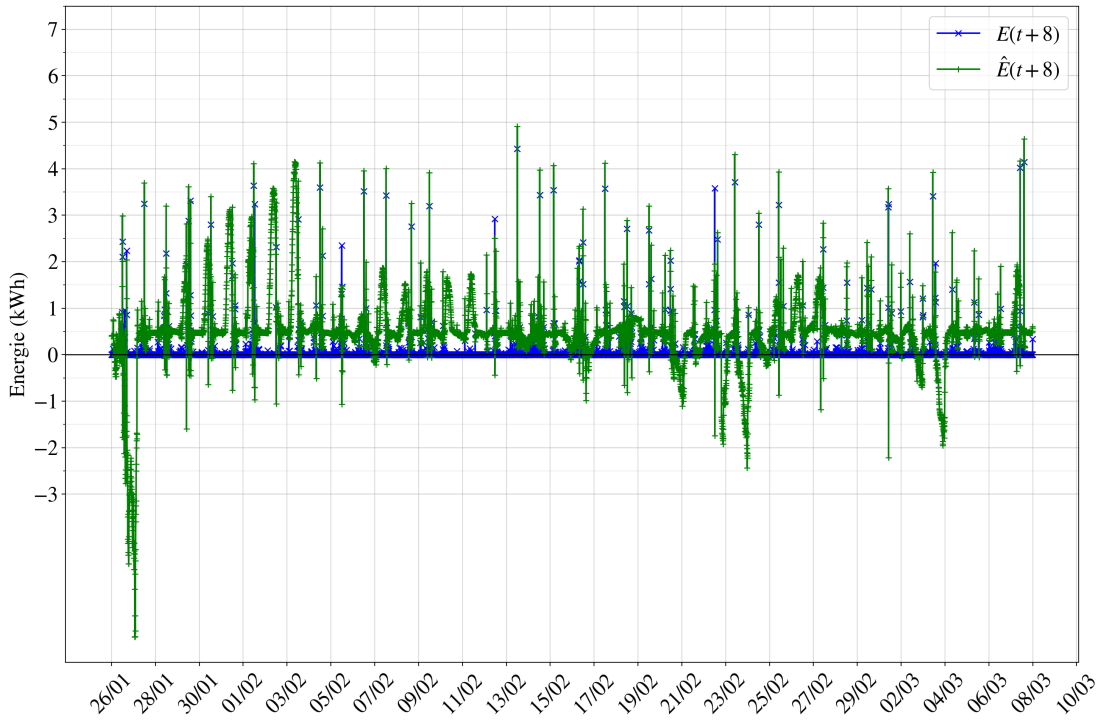


(B) jeu de validation

FIGURE 4.4 – Modèle B8 : visualisation de l'offset. Sur les 2 graphiques l'abscisse représente la date, au format jour/mois, l'ordonnée représente l'énergie en kWh. Le graphique du haut (A) représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et l'énergie prédite $\hat{E}(t+8)$ par le modèle B8 en vert, avec des marqueurs x, sur le jeu d'apprentissage. Le graphique du bas (B) représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et celle prédite $\hat{E}(t+8)$ par le modèle B8 en vert, avec des marqueurs x, sur le jeu de validation.



(A) jeu d'apprentissage



(B) jeu de validation

FIGURE 4.5 – Modèle C8 : visualisation de l'offset. Sur les 2 graphiques l'abscisse représente la date, au format jour/mois, l'ordonnée représente l'énergie en kWh. Le graphique du haut (A) représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et l'énergie prédite $\hat{E}(t+8)$ par le modèle C8 en vert, avec des marqueurs x, sur le jeu d'apprentissage. Le graphique du bas (B) représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et celle prédite $\hat{E}(t+8)$ par le modèle C8 en vert, avec des marqueurs x, sur le jeu de validation.

4.3.2.2 Comparaison des ANN développés.

Observation des prédictions. Nous représentons sur les figures 4.6, 4.7 et 4.8, les consommations réelles et les consommations prédites par les différents modèles, respectivement A8, B8, C8, ainsi que l'erreur ε sur chaque prédiction représentée. Sur chacune de ces figures, sur le graphique de gauche, est représentée la journée du mercredi 2 févier 2005, et sur le graphique de droite la journée du dimanche 8 août 2004. Nous choisissons l'échelle d'une journée, car elle permet de visualiser correctement l'évolution des consommations d'ECS. Ces dates sont choisies afin de visualiser le comportement des modèles sur une journée d'hiver et sur une journée d'été, ainsi que sur un jour de semaine et un jour de week-end. L'influence de la saison et du jour de la semaine sur les consommations d'ECS ayant été mis en avant dans la littérature, voir chapitre 1. Il est à noter que ces journées sont représentatives de ce que nous observons sur l'ensemble du test, les prédictions n'y sont ni meilleures, ni moins bonnes que sur d'autres journées du jeu de test.

Modèle	A8	B8	C8
$RMSE$	0,051	0,330	1,896
$\sigma(RMSE)$	0,081	0,170	2,346
$min(RMSE)$	$< 0,001$	$< 0,001$	$< 0,001$
$max(RMSE)$	0,868	3,234	22,759
$\bar{\varepsilon}$	-0,050	-0,326	-1,891
$\sigma(\varepsilon)$	0,082	0,177	2,351
$min(\varepsilon)$	-0,868	1,881	-22,758
$max(\varepsilon)$	0,385	3,234	1,716

TABLE 4.5 – Prédiction à 8 minutes sur le jeu de test : valeur de l'erreur moyenne signée $\bar{\varepsilon}$ et non signée $RMSE$, ainsi que de l'écart-type σ et des extremum min et max associés, pour les trois modèles du cas 1. Toutes les valeurs sont indiquées en kWh

Quantification de l'erreur de prédiction. Afin de comparer les performances des modèles, nous calculons, pour chaque modèle, les erreurs moyennes signées ε et non signées $RMSE$ faites sur l'ensemble de jeu de test. Les valeurs de ces erreurs ainsi que les informations statistiques permettant leurs analyses sont présentées dans le tableau 4.5.

Prise en compte des consommations précédentes. Le modèle A8 a comme entrées uniquement les consommations d'ECS précédant le moment de prédiction. Nous observons sur la figure 4.6 que les prédictions ont les mêmes tendances que les valeurs réelles de consommation. Nous observons des prédictions presque nulles lorsque la consommation réelle est nulle. Lors des phases de consommation, les prédictions augmentent et diminuent de façon similaire aux valeurs réelles. Cependant, nous observons sur la représentation de l'erreur ε que sur ces phases de consommation l'erreur est du même ordre de grandeur que la consommation. La valeur de l'erreur signée $RMSE$ de $0{,}051kWh$ est une valeur suffisamment basse pour considérer les prédictions comme fiables, en moyenne, voir 4.5. Toutefois, l'écart-type

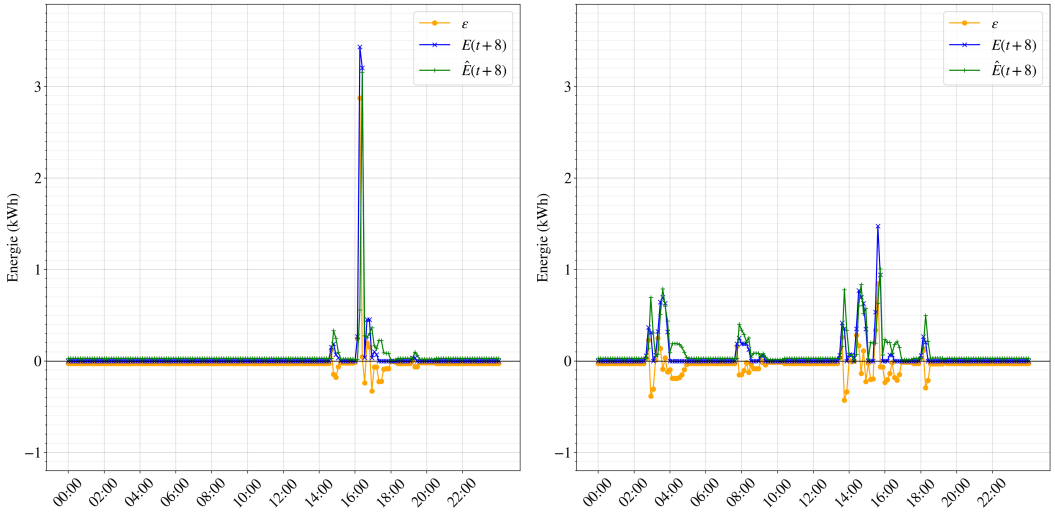


FIGURE 4.6 – Modèle A8 : visualisation des prédictions sur une journée extraite du jeu de test. Sur les deux graphiques, l'abscisse représente la date, au format heure/minute, l'ordonnée représente l'énergie en kWh. Le graphique de gauche représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et l'énergie prédite $\hat{E}(t+8)$ par le modèle A8 en vert, avec des marqueurs x, ainsi que l'erreur ε, en jaune avec des marqueurs ronds, sur la journée du 2 février 2005, incluse dans le jeu de test. Le graphique de droite représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et celle prédite $\hat{E}(t+8)$ par le modèle A8 en vert, avec des marqueurs x,ainsi que l'erreur ε, en jaune avec des marqueurs ronds, sur la journée du 8 août 2004, incluse dans le jeu de test.

de 0,081kWh, nous indique une dispersion importantes des erreurs autour de la moyenne. Ces informations statistiques couplées aux visualisations graphiques et au fait que les moments de consommation représentent seulement 15% du jeu de test, nous permettent de dire que les modèles font une erreur très faible sur les périodes de non-consommation, et une erreur beaucoup plus importante sur les périodes de consommation. Toutefois, nous observons sur les figures 4.6 que les prédictions sur les moments de consommation sont malgré tout très proches des consommations réelles. L'erreur très importante mesurée est dûe au fait que les prédictions sont légèrement décalées dans le temps des consommations réelles.

Modèle	A8
Nombre de prédictions totales	71 287
Nombre de prédictions où $E(t+8)=0$	60 170
Pourcentage de prédictions où $E(t+8)=0$	85%
$\bar{\varepsilon}$ pour $E(t+8)=0$	-0,038kWh
Nombre de prédiction où $\lvert\varepsilon\rvert > \lvert\bar{\varepsilon}\rvert$	6 617
Pourcentage de prédictions où $\lvert\varepsilon\rvert > \lvert\bar{\varepsilon}\rvert$	11%

TABLE 4.6 – Modèle du cas 1, A8 : Caractéristique des instants sans consommation et performance du modèle.

Période sans consommation : analyse des performances du modèle du cas 1, A8. Nous indiquons dans le tableau 4.6 les différents indicateurs qui nous permettent d'analyser la qualité des prédictions de notre modèle sur les périodes sans consommation, $E(t) = 0$. Nous indiquons que les moments sans consommation représentent 85% des moments étudiés. L'erreur moyenne $\bar{\varepsilon}$ est de -0,038kWh. Nous n'indiquons pas l'erreur non signée, car dans les cas étudiés, la valeur réelle est toujours zéro, l'erreur signée est donc toujours négative. L'erreur non signée n'apporte pas d'information supplémentaire. Les valeurs de l'erreur moyenne sont, comme nous nous y attendions, inférieures aux erreurs moyennes calculées sur l'ensemble de jeu de test. Cependant les prédictions représentées sur la figure 4.6 présentent une erreur proche de 0,010kWh lorsque les consommations réelles sont nulles. Nous expliquons cette différence par les moments où les modèles prédisent des consommations alors qu'il n'y en a pas réellement. Afin de quantifier l'impact de ces moments, nous calculons le nombre de prédictions où l'erreur est supérieure, en valeur absolue, à l'erreur moyenne faite sur les périodes sans consommation. Les moments où le modèle prédit des consommations alors qu'il n'y en a pas, représentent 11% du jeu de test. Compte tenu de l'erreur ε suffisamment faible et du peu d'exemples de prédiction où les modèles prédisent à tort une consommation, nous considérons que le modèle du cas 1, A8 est fiable pour des prédictions sur un horizon de huit, lorsque la consommation d'ECS réelle est nulle.

	2004									
Mois	Mars	Av.	Mai	Juin	Juil.	Août	Sept.	Oct.	Nov.	Déc.
δ_{24h}	0,21	0,23	0,19	0,10	0,16	0,20	0,13	0,24	0,27	0,26

	2005			
Mois	Janvier	Février	Mars	Avril
δ_{24h}	0,23	0,24	0,26	0,24

TABLE 4.7 – Modèle du cas 1, A8 : Valeurs de l'erreur relative sur une journée.

Période de consommation : analyse des performances du modèle du cas 1, A8. Afin de mesurer les performances du modèle sur les moments de consommation, en prenant en compte le décalage temporel observé sur la figure 4.6, nous utilisons l'erreur relative sur une journée, voir équation (4.4). Nous choisissons une erreur relative afin de pouvoir comparer les journées entre elles, sachant que l'énergie consommée pour l'ECS peut être très variable d'un jour à l'autre. Dans un objectif de synthèse, nous indiquons dans le tableau 4.7, les valeurs moyennes de δ_{24h}. Afin d'éviter toute confusion, nous précisons que c'est bien l'erreur relative faite sur la prédiction de l'énergie totale consommée sur une journée qui est moyennée sur un mois. Les résultats présentés ne sont pas l'erreur relative faite sur la prédiction de l'énergie totale consommée sur un mois. Nous calculons des valeurs moyennes de δ_{24h} sur un mois allant de 0,10 à 0,27. Une erreur de presque 30% peut paraître énorme, c'est pourquoi nous tenons à comparer nos résultats à d'autres modèles de la

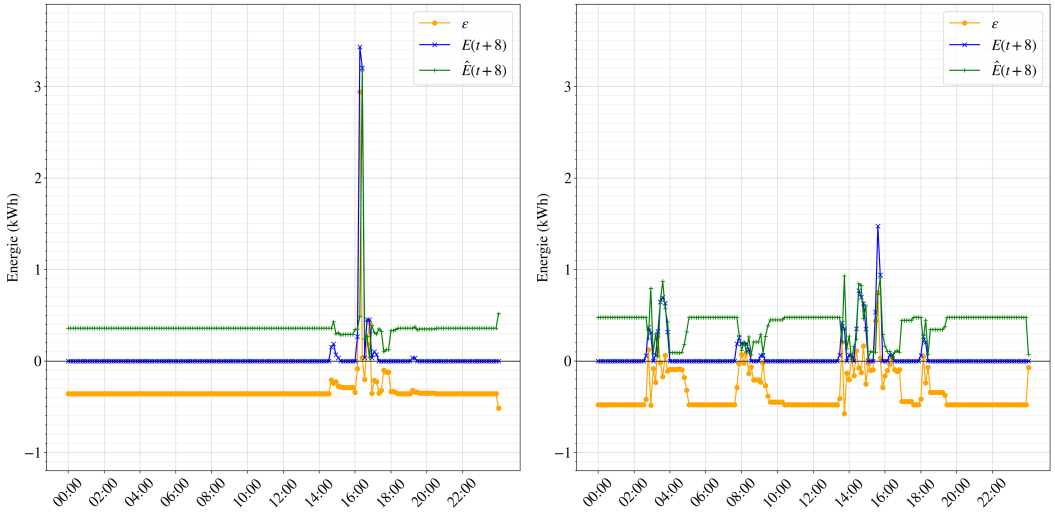


FIGURE 4.7 – Modèle B8 : visualisation des prédictions sur une journée extraite du jeu de test. Sur les deux graphiques, l'abscisse représente la date, au format heure/minute, l'ordonnée représente l'énergie en kWh. Le graphique de gauche représente l'énergie réelle consommée pour l'ECS pendant 64 minutes $E(t+8)$, en bleu, avec un marqueur + et l'énergie prédite $\hat{E}(t+8)$ par le modèle B8 en vert, avec des marqueurs x, ainsi que l'erreur ε, en jaune avec des marqueurs ronds, sur la journée du 2 février 2005, incluse dans le jeu de test. Le graphique de droite représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et celle prédite $\hat{E}(t+8)$ par le modèle B8 en vert, avec des marqueurs x,ainsi que l'erreur ε, en jaune avec des marqueurs ronds, sur la journée du 8 août 2004, incluse dans le jeu de test.

littérature. Malheureusement, ce type de modèle est rarement évalué seul, en général, c'est l'économie réalisée par l'utilisation d'une régulation prédictive basée sur des modèles qui est évaluée. Bennett et al. [40] obtiennent des erreurs relatives allant de 0,03 à 0,26 pour une prédiction de besoin moyen d'ECS par jour et par usage. Aydenalp et al. [33] prédisent la consommation annuelle d'une habitation avec un R^2 de 0,77, sachant qu'une prédiction est parfaite pour un R^2 égale à un. Bakker et al.[64] fournissent une représentation graphique des prédictions de leur modèle comparé aux consommations réelles, sur laquelle nous observons un décalage temporel d'un ordre de grandeur similaire à celui observé sur la représentation graphique de nos résultats. Ces études sont détaillées dans la partie 1.3. L'erreur mesurée sur les prédictions de nos modèles est du même ordre de grandeur que les travaux réalisé à partir d'un plus grand nombre de données d'entrée. Les valeurs moyennes de δ_{24h} présentées dans le tableau 4.7 sont plus en adéquation avec les observations graphiques réalisées sur la figure 4.6 que les valeurs de l'erreur ε. Cela corrobore l'hypothèse d'un décalage temporel de la prédiction. Bien que ce décalage soit difficilement mesurable, les erreurs bien plus basses faites sur la consommation d'une journée entière laissent penser que ce décalage n'empêchera pas l'utilisation des modèles dans le cadre d'une régulation prédictive, puisque dans la plupart des cas, les prédictions sont faites à vingt-quatre heures. Toutefois, il est nécessaire de valider l'impact de l'erreur de prédiction et du décalage temporel par une étude mesurant l'économie d'énergie réalisée en utilisant ces modèles au sein d'une régulation prédictive.

Prise en compte du jour de la semaine. Le modèle B8 a comme entrées les consommations d'ECS précédant le moment de prédiction ainsi que le paramétre 'jour', indiquant le jour de la semaine. Nous observons sur la figure 4.7 les mêmes différences entre les moments de consommation et de non-consommation que celles observées sur le modèle A8, à savoir que les modèles sont beaucoup plus précis sur les moments sans consommation. Toutefois l'erreur ε sur les périodes de consommation est plus importante que pour le modèle A8. Nous confirmons ces observations par les calculs d'erreurs présentés dans le tableau 4.5. L'erreur moyenne non signée $RMSE$ sur le jeu de test d'une valeur de $0{,}330kWh$ est nettement plus élevée pour le modèle B8 que pour le modèle A8. Nous observons sur la figure 4.7 que cette augmentation de l'erreur est imputable presque en totalité aux périodes de consommation. Nous observons sur la figures 4.6 que le modèle n'ayant pas le paramètre "jour" en entrée se comporte de la même façon sur un mercredi, graphique de gauche et sur un dimanche, graphique de droite. Les observations décrites précédemment laissent penser que l'information relative à l'influence du jour de la semaine est déjà présente dans les autres données sélectionnées comme entrées des modèles. Dans ce cas, l'ajout du paramètre "jour" en entrée créerait une sur-dépendance au jour de la semaine, puisqu'il est exprimé deux fois en entrée des modèles. Il n'est pas évident sur la figures 4.7 que le modèle ait les mêmes performances sur le mercredi 2 février et sur le dimanche 8 août. Toutefois, au vu de la dégradation globale des performances du modèle avec l'ajout du jour de la semaine en entrée, nous choisissons de ne pas pousser plus loin l'analyse afin de nous concentrer sur le jeu de données d'entrées le plus performant. Nous concluons au vu de ces observations que le jour de la semaine n'est pas une entrée pertinente pour des modèles prédisant la consommation d'énergie pour l'ECS à huit minutes.

Prise en compte de la température extérieure. Nous observons sur la figure 4.8 que le modèle prenant en compte la température extérieure en entrée ne permet pas de prédire la variation de consommation d'ECS sur un horizon de huit minutes. L'erreur moyenne sur le jeu de test permet de confirmer cette observation. En effet avec une $RMSE$ de $1{,}896kWh$ le modèle C8 ne peut pas être considéré fiable. La valeur plus faible de $\bar{\varepsilon}$ pour le modèle C8, indique que les erreurs sont autant négatives que positives. L'écart-type élevé indique que les valeurs de l'erreur ne sont pas concentrées autour de la moyenne, cette information est complétée par des valeurs extrêmes très élevées, dépassant les $20kWh$, voir 4.5. L'ensemble de ces indicateurs montrent des prédictions aberrantes comparées aux valeurs attendues. La première explication serait un mauvais apprentissage, nous écartons cette hypothèse dans la partie précédente 4.2. Des observations similaires sont faites sur la journée d'hiver, 2 février 2005, et sur la journée d'été, 8 août 2004. De ce fait, nous ne pouvons pas expliquer l'échec de prédiction des modèles par la variation des profils de consommation en hiver et en été. Nous observons sur les figures 4.6 et 4.7, que le modèle n'ayant pas la température extérieur en entrée se comporte de la même façon en hiver et en été. Les observations décrites précédemment laissent penser que l'information relative à l'influence de la saison est déjà présente dans les autres données sélectionnées comme entrées des modèles. Dans ce cas, l'ajout de la température extérieure en entrée créerait une sur-dépendance à la température

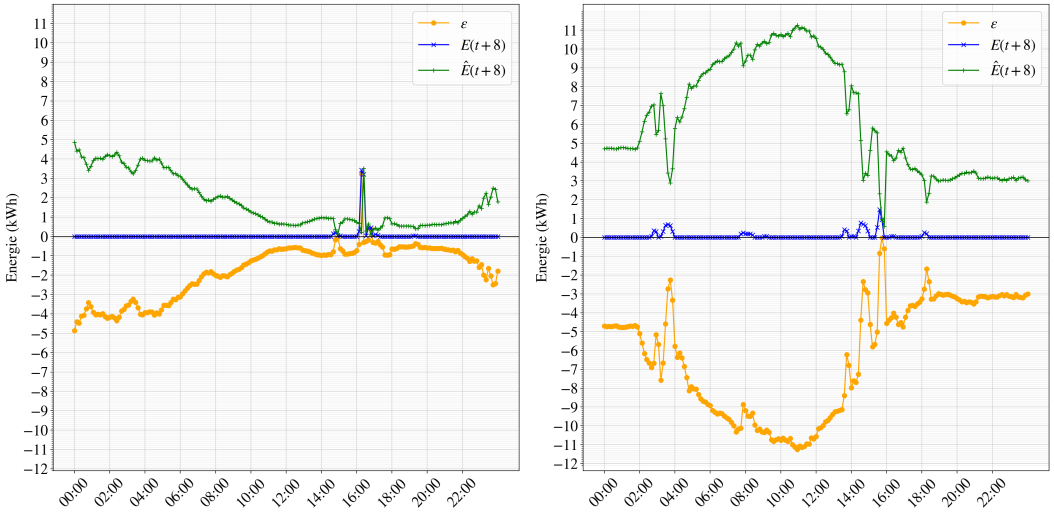


FIGURE 4.8 – Modèle C8 : visualisation des prédictions sur une journée extraite du jeu de test. Sur les deux graphiques, l'abscisse représente la date, au format heure/minute, l'ordonnée représente l'énergie en kWh. Le graphique de gauche représente l'énergie réelle consommée pour l'ECS pendant 64 minutes $E(t+8)$, en bleu, avec un marqueur + et l'énergie prédite $\hat{E}(t+8)$ par le modèle C8 en vert, avec des marqueurs x, ainsi que l'erreur ε, en jaune avec des marqueurs ronds, sur la journée du 2 février 2005, incluse dans le jeu de test. Le graphique de droite représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et celle prédite $\hat{E}(t+8)$ par le modèle C8 en vert, avec des marqueurs x,ainsi que l'erreur ε, en jaune avec des marqueurs ronds, sur la journée du 8 août 2004, incluse dans le jeu de test.

extérieure, puisqu'elle est exprimée deux fois en entrée des modèles. Nous concluons de ces observations que la température extérieure n'est pas une entrée pertinente pour des modèles prédisant la consommation d'énergie pour l'ECS à huit minutes.

ANN selectionné Les valeurs des différents indicateurs calculées lors de cette étude montrent que le ANN le plus performante est le ANN A8. Nous évaluons dans la partie suivante les capacités d'adaptations de ce ANN.

4.4 Évaluation du caractère adaptatif du réseau de neurones A8.

Méthode La méthode utilisée pour adapter les ANN développés précédemment au cas 2 et au cas 3 est schématisée sur la figure 4.9. Ces deux cas expérimentaux sont présentés dans le chapitre 2. Pour chaque cas, nous créons un jeu de données correspondant aux entrées et sorties de nos ANN. Ce jeu de données est séparé en un jeu d'apprentissage, les deux premières semaines de données et en un jeu de test, le reste des données. Ensuite, nous réalisons l'apprentissage. Afin de garantir la qualité de l'apprentissage nous vérifions que la RMSE sur le jeu d'apprentissage ne dépasse pas une valeur minimum $RMSE_{min}$ variant pour chaque cas. Si c'est le cas, nous réalisons les prédictions sur l'ensemble du jeu de test. Nous réalisons ensuite les deux étapes de traitement des prédictions définies dans la partie précédente, c'est-à-dire le

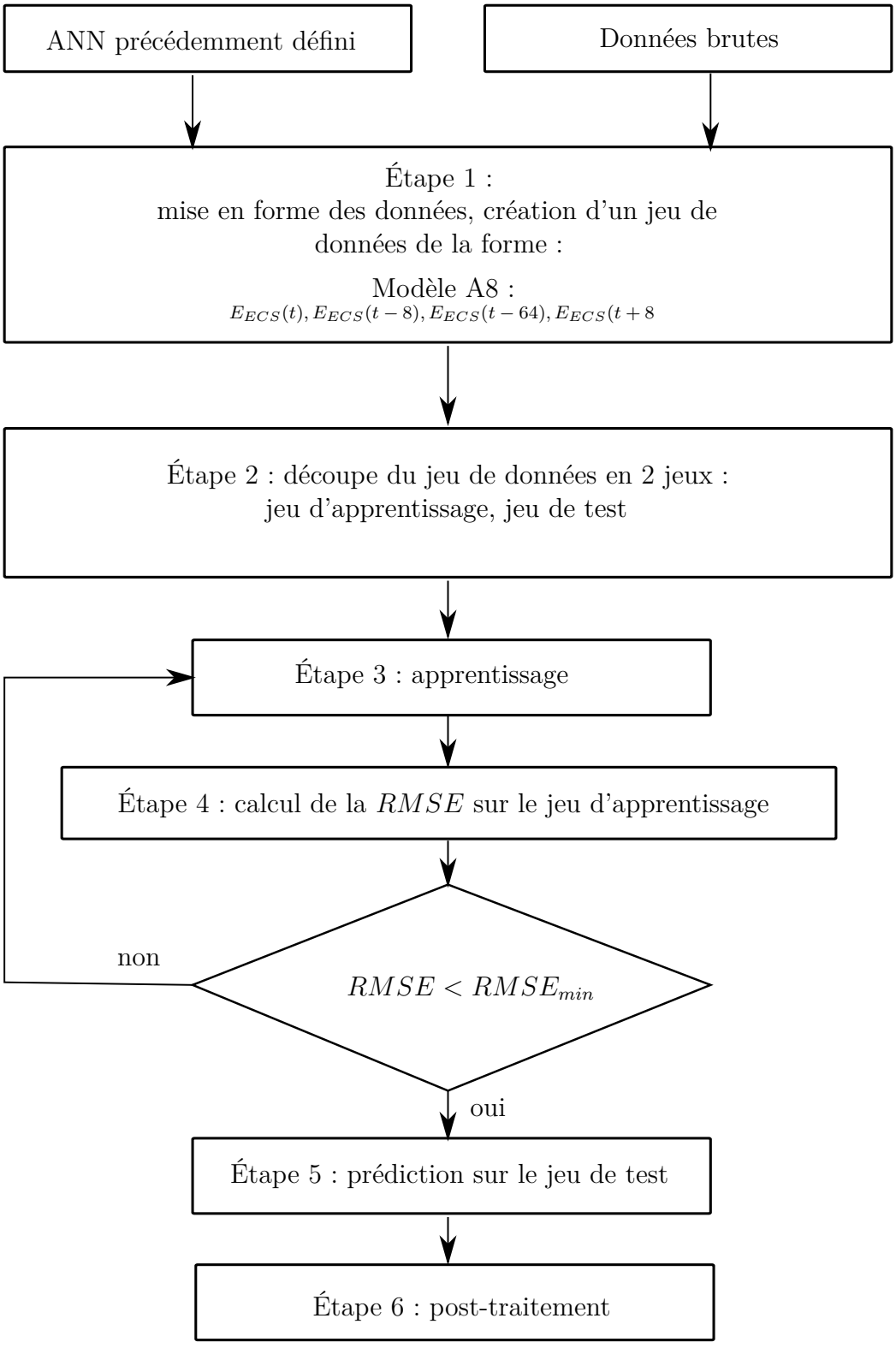


FIGURE 4.9 – Étapes d'adaptation du ANN au cas 2 et au cas 3.

calculé et l'addition de l'offset, puis l'application de la fonction valeur absolue. Ces prédictions sont comparées aux valeurs réelles.

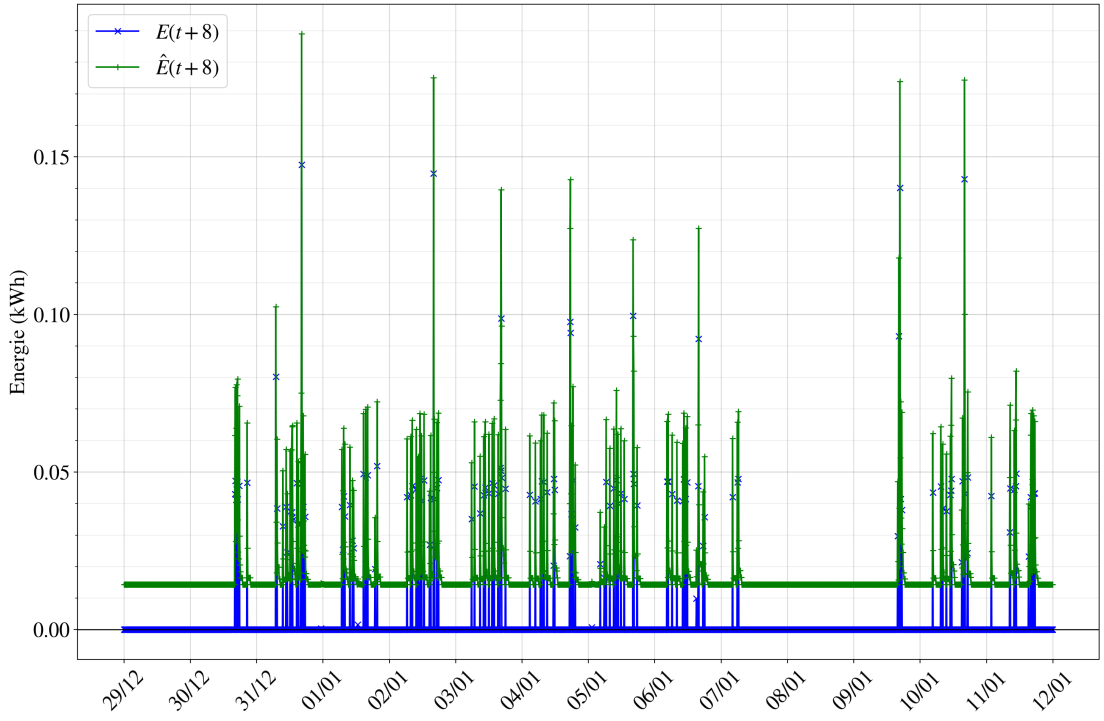


FIGURE 4.10 – Cas 2 : visualisation de l'offset.L'abscisse représente la date, au format jour/mois, l'ordonnée représente l'énergie en kWh. Le graphique représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et l'énergie prédite $\hat{E}(t+8)$ par le modèle, avec des marqueurs x, sur le jeu d'apprentissage.

Validation de l'apprentissage Le tableau 4.8 regroupe les valeurs de l'erreur moyenne non signée $RMSE_{app}$ faite sur le jeu d'apprentissage par les modèles des cas 2 et cas 3. Nous observons une $RMSE_{app}$ de 0,004kWh pour le modèle du cas 1 et de 0,026kWh pour le modèle du cas 2 ce qui nous permet de considérer l'apprentissage comme réussi.

TABLE 4.8 – RMSE faite sur le jeu d'apprentissage par les modèles du cas 2 et du cas 3, en kWh.

Cas d'étude	cas 2	cas 3
$RMSE_{app}$	0,004	0,026

Étapes de traitement automatique des prédictions Nous représentons sur les figures 4.10 et 4.11 les consommations prédites $\hat{E}(t+8)$ et réelles $E(t+8)$ sur les exemples du jeu d'apprentissage. Comme pour le modèle du cas 1 présenté précédemment, nous observons un décalage entre les valeurs réelles et les valeurs prédites. Des prédictions d'énergie négative sont observées uniquement sur le cas 3. Cependant, sur le cas 2, certaines prédictions sont inférieures au décalage moyen et peuvent devenir

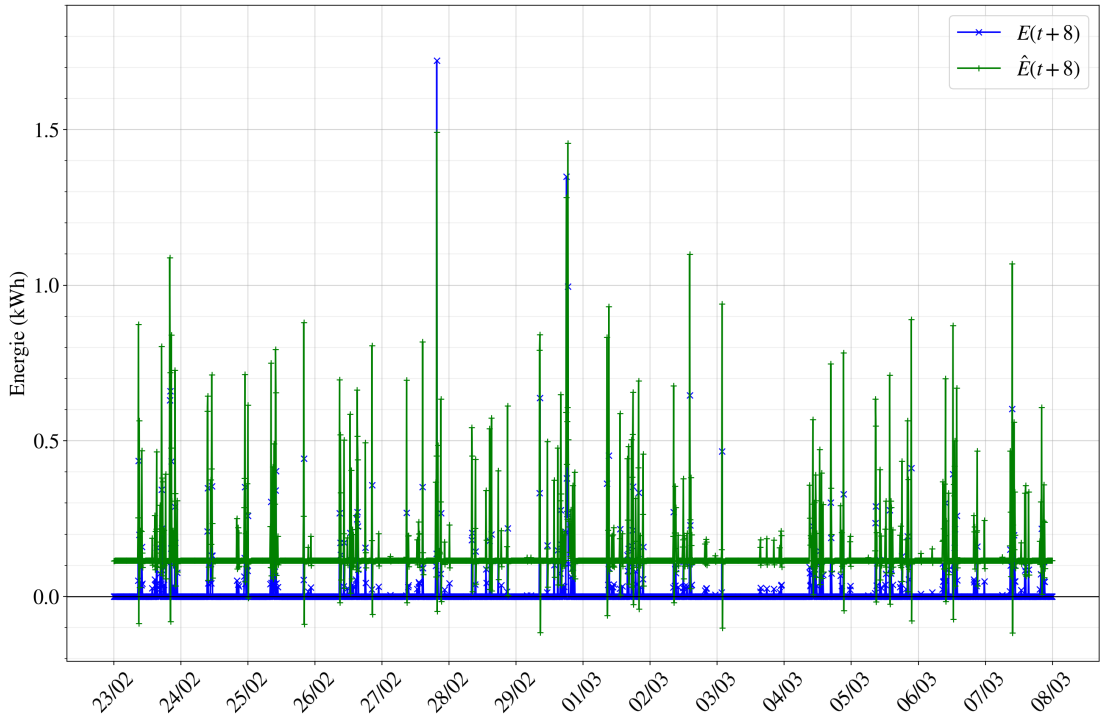


FIGURE 4.11 – Cas 3 : visualisation de l'offset. L'abscisse représente la date, au format jour/mois, l'ordonnée représente l'énergie en kWh. Le graphique du haut (A) représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et l'énergie prédite $\hat{E}(t+8)$ par le modèle, avec des marqueurs x, sur le jeu d'apprentissage.

négative après la suppression de l'offset. Comme proposé dans la partie précédente, nous calculons l'offset de façon automatique sur le jeu d'apprentissage, les valeurs sont indiquées pour les deux modèles dans le tableau 4.9. Une fois le décalage corrigé, nous appliquons la fonction valeur absolue aux prédictions, afin qu'elles soient toutes positives. Ce traitement des prédictions est réalisé sur l'ensemble du jeu de test. Nous notons un comportement similaire, pendant la phase d'apprentissage, du ANN A8 sur les trois cas évalués.

TABLE 4.9 – Offset calculé sur les modèles des cas 2 et cas 3, en kWh.

Cas d'étude	cas 2	cas 3
offset	-0,016	-0,111

Influence du jour de la semaine et de la saison. Nous représentons sur les figures 4.12 et 4.13 l'énergie nécessaire à l'ECS prédite $\hat{E}(t+8)$ et celle réellement consommée $E(t+8)$, ainsi que l'erreur ϵ entre ces deux grandeurs, respectivement pour les modèles du cas 2 et du cas 3. Sur le graphique de gauche, est représentées la journée du mercredi 2 février 2005, sur le graphique de droite, la journée du dimanche 8 août 2004. La première observation que nous faisons, sur le modèle du cas 2, est que la valeur moyenne de l'énergie consommée $E(t+8)$ est bien plus basse que pour

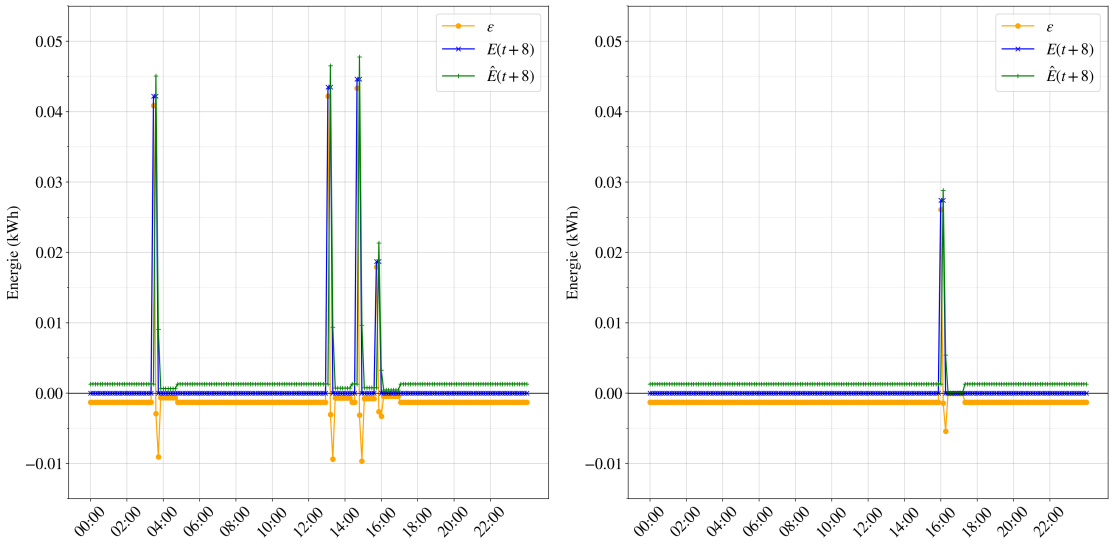


FIGURE 4.12 – Cas 2 : visualisation des prédictions sur une journée extraite du jeu de test. Sur les deux graphiques, l'abscisse représente la date, au format heure/minute, l'ordonnée représente l'énergie en kWh. Le graphique de gauche représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et l'énergie prédite $\hat{E}(t+8)$ par le modèle A8 en vert, avec des marqueurs x, ainsi que l'erreur ε, en jaune avec des marqueurs rond, sur la journée du 2 fevrier 2005, incluse dans le jeu de test. Le graphique de droite représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et celle prédite $\hat{E}(t+8)$ par le modèle A8 en vert, avec des marqueurs x,ainsi que l'erreur ε, en jaune avec des marqueurs rond, sur la journée du 8 août 2004, incluse dans le jeu de test.

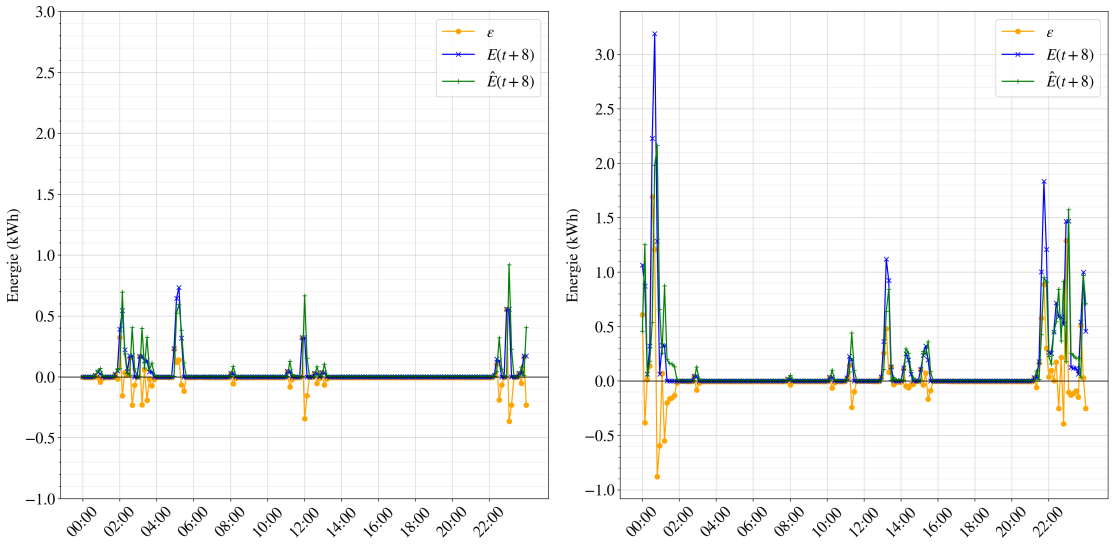


FIGURE 4.13 – Cas 3 : visualisation des prédictions sur une journée extraite du jeu de test. Sur les deux graphiques, l'abscisse représente la date, au format heure/minute, l'ordonnée représente l'énergie en kWh. Le graphique de gauche représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et l'énergie prédite $\hat{E}(t+8)$ par le modèle A8 en vert, avec des marqueurs x, ainsi que l'erreur ε, en jaune avec des marqueurs rond, sur la journée du 2 fevrier 2005, incluse dans le jeu de test. Le graphique de droite représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et celle prédite $\hat{E}(t+8)$ par le modèle A8 en vert, avec des marqueurs x,ainsi que l'erreur ε, en jaune avec des marqueurs rond, sur la journée du 8 août 2004, incluse dans le jeu de test.

le cas 1. Nous observons, comme pour le cas 1, une bonne précision des modèles sur les instants sans consommation, bien que le modèle du cas 2 semble d'avantage s'éloigner de zéro que ceux des cas 1 et cas 3. L'erreur ϵ est de l'ordre de grandeur de la consommation sur les moments de consommations. Nous n'observons pas de différence notable sur la performance des modèles entre la journée d'hiver, en semaine et la journée d'été en week-end. Cela laisse penser que notre choix d'architecture permet de capter l'influence de la saison et du jour de la semaine. Le tableau 4.10 regroupe les informations statistiques sur l'erreur signée et l'erreur non signée sur les cas 2 et 3. Nous observons des valeurs basses des erreur non-signée $RMSE$ et signée ε, indiquant une nouvelle fois une majorité de moment sans consommation, et un écart-type proportionnellement important, indiquant une distribution très étalé des erreurs sur le jeu de test. Cette distribution étalée est confirmée par des valeurs extrêmes très éloignées de la moyenne. La valeur de l'erreur signée ε est inférieur à celle de la $RMSE$ ce qui indique que l'erreur n'est pas majoritairement négative ou positive, mais équitablement répartie. Ces premières observations sont semblables à celle réalisées sur le modèle du cas 1. Nous caractérisons le comportement de modèle du cas 2 en distinguant les moments de consommations des moments de non-consommation, dans la suite de cette partie.

Modèle	cas 2	cas 3
Nombre de prédictions	64 203	61 666
$RMSE$	0,003	0,039
$\sigma(RMSE)$	0,008	0,118
$min(RMSE)$	$< 0,001$	$< 0,001$
$max(RMSE)$	0,200	3,322
$\bar{\varepsilon}$	-0,001	0,001
$\sigma(\varepsilon)$	0,009	0,125
$min(\varepsilon)$	-0,057	-2,302
$max(\varepsilon)$	0,200	3,322

Table 4.10 – Cas 2 et cas 3 : Prédiction à 8 minutes sur le jeu de test : valeur de l'erreur moyenne signée $\bar{\varepsilon}$ et non signée $RMSE$, ainsi que de l'écart-type σ et des extremum min et max associés, pour les modèles des cas 2et cas 3. Toute les valeurs sont indiquées en kWh

Période sans consommation Afin de caractériser les performances des modèles des cas 2 et cas 3, nous calculons le nombre d'exemples du jeu de test où l'énergie réellement consommée pour l'ECS est nulle, $E(t+8) = 0$. Nous calculons également l'erreur signée ε sur ces instants. Comme pour le cas 1, le calcul de l'erreur non signée n'apporterait aucune information supplémentaire, toutes les erreurs étant du même signe. Nous observons une forte majorité des moments sans consommation, qui représente 94% du jeu de test pour le cas 2 et 79% du jeu de test pour le cas 3. Pour le modèle du cas 3, comme pour le cas 1, l'erreur est nettement plus basse sur les moments sans consommation que sur l'ensemble du jeu de test et comme pour le cas 1, nous lisons sur le graphique 4.13 une erreur plus basse que l'erreur moyenne calculée, ce qui indique que le modèle prédit des moments de consommation alors que la

consommation réelle est nulle. Nous calculons que ces fausses prédictions représentent 8% des moments sans consommation. Pour le modèle du cas 2, l'erreur ne diminue pas de façon aussi significative que sur les deux autres cas. Nous l'expliquons par une erreur plus importante faite sur les moments de non-consommation, comme nous l'avons observé sur la figure 4.12. Les moments de fausse prédiction représentent pour le cas 2, 6% des moments sans consommation.

Modèle	cas 2	cas 3
Nombre de prédictions totales	64 203	61 666
Nombre de prédictions où $E(t+8)=0$	60 514	48 571
Pourcentage de prédictions où $E(t+8)=0$	94%	79%
$\bar{\varepsilon}$ pour $E(t+8)=0$	-0,001kWh	-0,008kWh
Nombre de prédictions où $\lvert\varepsilon\rvert > \lvert\bar{\varepsilon}\rvert$	3 614	3 808
Pourcentage de prédictions où $\lvert\varepsilon\rvert > \lvert\bar{\varepsilon}\rvert$	6%	8%

TABLE 4.11 – Modèles des cas 2 et cas 3 : Caractéristique des instants sans consommation et performance du modèle.

Période avec consommation Tout comme pour le cas 1, nous calculons les valeurs moyennes sur chaque mois de l'indicateur $\delta_2 4h$ mesurant l'erreur relative faite par les modèles sur la consommation d'énergie d'une journée. Ces valeurs sont indiquées dans le tableau 4.12 pour le modèle du cas 2 et dans le tableau 4.13 pour le modèle du cas 3. Le modèle du cas 3 a des performances similaires à celles du modèle du cas 1, que nous interprétons comme une très bonne adaptabilité du ANN A8 au cas 3. Cependant le modèle du cas 2 commet une erreur deux fois plus importante. Nous avons remarqué que les consommations du cas 2 sont très inférieures à celle des cas 1 et 3. Il est possible que les performances des modèles dépendent de la valeur de l'énergie, les modèles seraient alors plus précis sur les fortes consommations que sur les faibles. Cette tendance existe peut-être sur les modèles des cas 1 et cas 3, cependant nos indicateurs ne permettent pas de la mettre en avant. Une étude plus approfondie devra être menée afin d'affirmer ou d'infirmer cette hypothèse.

	2004								
Mois	Av.	Mai	Juin	Juil.	Août	Sept.	Oct.	Nov.	Déc.
δ_{24h}	0,41	0,43	0,43	0,42	0,43	0,43	0,43	0,41	0,41

	2005			
Mois	Janvier	Février	Mars	Avril
δ_{24h}	0,41	0,41	0,40	0,42

TABLE 4.12 – Modèle du cas 2 : Valeurs de l'erreur relative sur une journée

	2004								
Mois	Av.	Mai	Juin	Juil.	Août	Sept.	Oct.	Nov.	Déc.
δ_{24h}	0,12	0,17	0,13	0,20	0,16	0,16	0,13	0,10	0,13

	2005			
Mois	Janvier	Fevrier	Mars	Avril
δ_{24h}	0,11	0,12	0,11	0,15

TABLE 4.13 – Modèle du cas 3 : Valeurs de l'erreur relative sur une journée

4.5 Conclusion

Dans ce chapitre, nous développons trois architectures de réseaux de neurones artificiels afin de prédire l'énergie consommée pour l'ECS huit minutes après l'instant de prédiction. Nous déterminons trois jeux d'entrée afin d'évaluer la capacité des ANN à prendre en compte l'influence du jour de la semaine et l'influence de la saison sur les consommations d'ECS. Pour chacun de ces jeux d'entrée, nous développons l'architecture optimale en utilisant les données du cas 1. Nous montrons pour chacun de ces ANN qu'un jeu d'apprentissage comptant tous les exemples mesurés pendant deux semaines (2 520 exemples) est suffisant. Nous montrons qu'il est nécessaire d'appliquer un post-traitement aux prédictions afin de corriger l'offset et les prédictions d'énergie négative. L'évaluation de ces trois architectures montre que l'influence à la fois du jour de la semaine et de la saison est déjà comprise dans les valeurs de l'énergie consommée pendant les instants précédents la prédiction. Nous observons que le modèle du cas expérimental 1 fournit, en moyenne, des prédictions fiables, puisque la valeur de la $RMSE$ sur l'ensemble du jeu de test est de $51Wh$. Nous mettons en évidence que le modèle est plus précis sur les périodes sans consommation que sur les périodes de consommation. De plus, les prédictions sont décalées dans le temps par rapport aux valeurs vraies. Malgré cela, en la comparant avec les modèles présents dans la littérature, nous montrons que la précision des prédictions est suffisante pour que le modèle soit utilisé au sein d'une régulation prédictive. Nous nous assurons que les temps de calcul nécessaires aux prédictions permettent une utilisation en temps réel. Nous montrons qu'un ANN composé d'une couche intermédiaire de cinq neurones, et dont les entrées sont la consommation d'énergie pour l'ECS à l'instant de prédiction, huit minutes avant et soixante-quatre minutes avant l'instant de prédiction, en utilisant l'algorithme de Levenberg-Marcquardt sur deux semaines de données pour l'apprentissage, permet de prédire l'énergie consommée pour l'ECS huit minutes après l'instant de prédiction. Toutefois, une étude utilisant ces modèles au sein d'une régulation prédictive devra venir confirmer ces bons résultats.

Nous évaluons les capacités d'adaptation du ANN sur deux cas expérimentaux. Il apparaît sur les prédictions faites sur les exemples des jeux d'apprentissage que les deux modèles ont besoin du même post-traitement que le modèle du cas 1, utilisé pour le développement du ANN. Tout comme le modèle du cas 1, les modèles des cas 2 et cas 3 présentent de meilleurs performances sur les instants sans consommation que sur les instants avec consommation. Les deux modèles présentent également

un décalage temporel. Les performances du modèle du cas 3 sur les instants de consommation sont comparables à celles du modèle du cas 1. Ce qui nous permet de conclure qu'une architecture de ANN peut s'adapter au moins à un autre bâtiment que celui sur lequel elle a été développée. Cependant le modèle du cas 2, même si ses performances présentent des similitudes avec celles du modèle du cas 1, commet une erreur moyenne sur une journée deux fois plus importante que celle commise pas le modèle du cas 1. Nous supposons que cette erreur est due aux consommations beaucoup plus faibles du cas 2 par rapport au cas 1. Par conséquent, l'erreur faite par le ANN serait dépendante de l'énergie consommée, le ANN serait plus précis sur les fortes consommations que sur les faibles. Une étude complémentaire est nécessaire pour valider cette hypothèse. D'autres cas d'études sont nécessaires pour évaluer le caractère adaptatif global du ANN, cependant les résultats sont très encourageant puisque le ANN s'est parfaitement adapté au cas 3, en présentant des performances équivalentes à celle du modèle du cas 1. Des travaux devront être menés afin de tester ce ANN sur davantage de bâtiment. De plus l'horizon de prédiction de notre ANN est trop court pour une régulation prédictive. Nous proposons dans un premier temps d'utiliser ce modèle en chaîne, c'est-à-dire d'utiliser la prédiction d'un modèle à huit minutes comme entrée d'un même modèle, ce qui permettrait d'obtenir une prédiction à seize minutes. Il faudrait quatre-vingt-dix modèles pour obtenir une prédiction toutes les huit minutes sur un horizon de douze heures. Il est nécessaire d'étudier l'augmentation de l'erreur de prédiction due à l'utilisation de prédiction en entrée du modèle.

Nous montrons qu'un réseau de neurones artificiels permet de prédire l'énergie consommée pour l'ECS sur un horizon de huit minutes en utilisant uniquement les données usuellement disponibles sur une installation de production d'ECS. De plus, nous montrons que ce réseau de neurones artificiels peut s'adapter au moins à un bâtiment sur lequel il n'a pas été développé. Ces conclusions sont encourageantes pour l'utilisation à grande échelle de régulation prédictive pour la gestion de l'eau chaude sanitaire.

Conclusion

Conclusion

Ces travaux de thèse proposent une première solution non-intrusive afin de réaliser des prédictions pour la gestion énergétique de l'habitat. Ces travaux étant réalisés en partenariat avec le fabricant de chaudière thermodynamique BoostHeat, nous nous intéressons à la production de chauffage et d'eau chaude sanitaire. Nous nous plaçons du point de vue fabricant de chaudière qui souhaite proposer à l'installation de sa machine une régulation prédictive sans avoir à réaliser une phase longue et coûteuse de modélisation des usages.

Dans le chapitre1, nous présentons l'état de l'art en terme de régulation prédictive des systèmes de production de chauffage et d'eau chaude sanitaire pour l'habitat individuel. Dans un premier temps, nous montrons que ces méthodes permettraient de réduire les consommations d'énergie de la chaudière thermodynamique BoostHeat, donc d'en améliorer l'efficacité. Nous mettons en avant le besoin de modèles prédictifs fiables. Nous montrons pour la prédiction de température intérieure et de besoin d'eau chaude sanitaire que la méthode des réseaux de neurones artificiels est la plus pertinente dans notre cas. En effet, son caractère intrinsèquement adaptatif permet de développer des modèles spécifiques à un bâtiment à l'aide uniquement de données mesurées sur site. Nous présentons les principes théoriques liés à la méthode des réseaux de neurones artificiels nécessaires à la compréhension de nos travaux. Nous nous attachons à réduire au maximum les données nécessaires au développement et à l'utilisation des réseaux de neurones artificiels, dans le but de minimiser l'instrumentation du bâtiment et de facilité l'utilisation des modèles sur un très grand nombre d'habitations. Afin de réduire la complexité de la prédiction, induite par cette limitation des choix d'entrée nous réalisons une prédiction à court terme. Nous posons l'hypothèse que l'architecture du réseau de neurones artificiels représente les phénomènes physiques liant les variables d'entrée aux variables de sortie. Les données d'apprentissage, donc les poids du réseau de neurones artificiels représentent les spécificités de chaque habitation. Dans le cas de la prédiction de température intérieure du bâtiment, nous avons présenté des études montrant les performances de cette méthode. Les études citées proposent une architecture de réseaux de neurones artificiels pour un bâtiment. Les objectifs de ces travaux sont de développer une architecture unique pour tous les bâtiments et de tester son caractère adaptatif. Dans le cas de la prédiction de besoin d'eau chaude sanitaire, les grandeurs

à prendre en compte en entrées des réseaux de neurones artificiels sont encore mal connues, ainsi que la quantité de données nécessaires à une prédiction fiable. Les objectifs de ces travaux sont de développer une architecture unique, en mettant en avant l'influence du choix des entrées des exemples d'apprentissages. Nous testons également le caractère adaptatif du réseau de neurones artificiels. Pour la prédiction de consommation d'eau chaude sanitaire, comme pour la prédiction de température intérieure, le modèle doit s'adapter automatiquement à tous types de bâtiment et d'usager.

Dans le chapitre 2 nous présentons les cas d'études qui servent à développer et à tester nos réseaux de neurones artificiels. Nous distinguons deux types de cas : les cas expérimentaux et les cas simulés. Dans une première partie, nous décrivons les trois cas expérimentaux : trois servent pour la prédiction d'énergie pour l'eau chaude sanitaire, deux servent pour la prédiction de température intérieure du bâtiment. En terme de consommation d'eau chaude sanitaire, le cas 2 représente une consommation moyenne, le cas 1 représente une faible consommation et le cas 3 représente une forte consommation. En terme de consommation énergétique globale, le cas 1, d'étiquette énergétique C, est à la fois moyen et représentatif de la classe énergétique majoritaire. Le cas 2, d'étiquette énergétique A, est lui représentatif d'habitation peu énergivore. Les trois cas comportent la même particularité : une volonté des habitants de tester de nouveaux systèmes solaires combinés. Cela dénote une sensibilité à leur impact environnemental et à leur consommation d'énergie et donc potentiellement une utilisation raisonnée ou économe des ressources énergétiques. Les cas simulés sont utilisés uniquement pour la prédiction de températures intérieures du bâtiment. Ils ont pour objectif de compléter les cas expérimentaux afin d'étudier des cas représentatifs du parc immobilier français.

Dans le chapitre 3 nous nous intéressons à la prédiction de température intérieure du bâtiment. Nous définissons l'architecture et la méthode d'apprentissage choisies pour notre réseau de neurones artificiels. Notre réseau de neurones artificiels comporte neuf entrées, une couche intermédiaire de vingt neurones dont la fonction d'activation est la fonction sigmoïde et une couche de sortie d'un unique neurone dont la fonction d'activation est la fonction identité. Nous validons l'architecture et les choix d'apprentissage sur les cas expérimentaux à l'aide des courbes d'apprentissage. Nous montrons qu'un jeu d'apprentissage composé de 2 520 exemples, soit deux semaines de données est suffisant pour permettre une adaptation de notre réseau de neurones artificiels à un cas particulier de bâtiment. Nous nous assurons que les temps de calcul nécessaires à la mise en place et à l'utilisation de ce réseau de neurones artificiels sont cohérent avec son utilisation au sein d'une régulation prédictive. Nous évaluons les performances de notre réseau de neurones artificiels sur les deux cas expérimentaux ayant servi à son développement. Cela nous permet de confirmer la capacité d'un réseau de neurones artificiels à prédire la température intérieure d'un bâtiment résidentiel, en effet sur le premier cas expérimental le modèle commet en moyenne une erreur de 0,45°C, sur le deuxième cas expérimentale de 0,53 °C. Nous mettons en évidence que les modèles font une erreur de prédiction plus importante sur la période d'été où le chauffage ne fonctionne pas. Ce n'est pas gênant puisque les modèles ne sont pas voués à être utilisés lorsqu'il n'y a pas besoin de chauffage. Nous observons sur les cas expérimentaux que l'erreur diminue avec l'augmentation

de la puissance de chauffage. Nous expliquons ce phénomène par la dépendance des modèles à leurs jeux d'apprentissage. Dans une deuxième partie, nous adaptons notre réseau de neurones artificiels à cent-vingt cas simulés représentatifs du parc immobilier français. Nous évaluons les performances de ces modèles en fonction de trois caractéristiques : la superficie du bâtiment, la classe énergétique du bâtiment et le climat auquel il est soumis. Nous calculons une valeur moyenne de la RMSE sur l'ensemble des cas de 0,45°C. Cette erreur en plus d'être très satisfaisant, est du même ordre de grandeur que celle calculée sur les cas expérimentaux. Les modèles des cas simulés présentent une dépendance moins marquée à la puissance de chauffage. Nous l'expliquons par la composition des jeux d'apprentissages. Les jeux d'apprentissage des modèles des cas simulé sont composés d'exemples mieux répartis sur la gamme de puissance de chauffage. Une autre explication serait que l'influence de la puissance de chauffage sur la précision des prédictions est corrélée au comportement des occupants. Étant donné que le comportement des occupants n'est pas pris en compte dans les simulations, nous ne pouvons pas observer cette influence sur les cas simulés. Nous montrons que notre réseau de neurones artificiels n'est pas sensible au type de bâtiment auquel il est adapté, c'est-à-dire que ses performances ne varient pas selon la superficie, la classe énergétique ou le climat du cas auquel nous appliquons notre réseau de neurones artificiels. L'étude ayant été menée par simulation, nous ne pouvons pas conclure que le réseau de neurones artificiels est adaptable à tout type de bâtiment. Nous montrons qu'un réseau de neurones artificiels permet de prédire la température intérieure d'un bâtiment sur un horizon d'une heure en utilisant uniquement les données usuellement disponibles sur une installation de production de chauffage. De plus, nous montrons le caractère adaptatif du réseau de neurones artificiels, puisqu'il n'est pas dépendant de la superficie, de la classe énergétique et du climat du bâtiment auquel il est adapté.

Dans le chapitre 4 nous nous intéressons à la prédiction des besoin d'eau chaude sanitaire. Nous développons trois architectures de réseaux de neurones artificiels afin de prédire l'énergie consommée pour l'eau chaude sanitaire huit minutes après l'instant de prédiction. Nous déterminons trois jeux d'entrée afin d'évaluer la capacité des réseaux de neurones artificiels à prendre en compte l'influence du jour de la semaine et l'influence de la saison sur les consommations d'eau chaude sanitaire. Pour chacun de ces jeux d'entrée, nous développons l'architecture optimale en utilisant les données du cas 1. Nous montrons pour chacun de ces réseaux de neurones artificiels qu'un jeu d'apprentissage comptant tous les exemples mesurés pendant deux semaines (2 520 exemples) est suffisant. Nous montrons qu'il est nécessaire d'appliquer un post-traitement aux prédictions afin de corriger l'offset et les prédictions d'énergie négative. L'évaluation de ces trois architectures montre que l'influence à la fois du jour de la semaine et de la saison est déjà comprise dans les valeurs de l'énergie consommée pendant les instants précédents la prédiction. Nous observons que le modèle du cas expérimental 1 fournit, en moyenne, des prédictions fiables, puisque la valeur de la $RMSE$ sur l'ensemble du jeu de test est de $51Wh$. Nous mettons en évidence que le modèle est plus précis sur les périodes sans consommation que sur les périodes de consommation. De plus, les prédictions sont décalées dans le temps par rapport aux valeurs vraies. Malgré cela, en la comparant avec les modèles présents dans la littérature, nous montrons que la précision des prédictions est suffisante pour

que le modèle soit utilisé au sein d'une régulation prédictive. Nous nous assurons que les temps de calcul nécessaires aux prédictions permettent une utilisation en temps réel. Nous montrons qu'un réseau de neurones artificiels composé d'une couche intermédiaire de cinq neurones, et dont les entrées sont la consommation d'énergie pour l'ECS à l'instant de prédiction, huit minutes avant et soixante-quatre minutes avant l'instant de prédiction, en utilisant l'algorithme de Levenberg-Marcquardt sur deux semaines de données pour l'apprentissage, permet de prédire l'énergie consommée pour l'eau chaude sanitaire huit minutes après l'instant de prédiction. Toutefois, une étude utilisant ces modèles au sein d'une régulation prédictive devra venir confirmer ces bons résultats. Nous évaluons les capacités d'adaptation du réseau de neurones artificiels sur deux cas expérimentaux. Il apparaît sur les prédictions faites sur les exemples des jeux d'apprentissage que les deux modèles ont besoin du même post-traitement que le modèle du cas 1, utilisé pour le développement du réseau de neurones artificiels. Tout comme le modèle du cas 1, les modèles des cas 2 et cas 3 présentent de meilleurs performances sur les instants sans consommation que sur les instants avec consommation. Les deux modèles présentent également un décalage temporel. Les performances du modèle du cas 3 sur les instants de consommation sont comparables à celles du modèle du cas 1. Ce qui nous permet de conclure qu'une architecture de réseau de neurones artificiels peut s'adapter au moins à un autre bâtiment que celui sur lequel elle a été développée. Cependant le modèle du cas 2, même si ses performances présentent des similitudes avec celles du modèle du cas 1, commet une erreur moyenne sur une journée deux fois plus importante que celle commise pas le modèle du cas 1. Nous supposons que cette erreur est due aux consommations beaucoup plus faibles du cas 2 par rapport au cas 1. Par conséquent, l'erreur faite par le réseau de neurones artificiels serait dépendante de l'énergie consommée, le réseau de neurones artificiels serait plus précis sur les fortes consommations que sur les faibles. Une étude complémentaire est nécessaire pour valider cette hypothèse. Nous montrons qu'un réseau de neurones artificiels permet de prédire l'énergie consommée pour l'eau chaude sanitaire sur un horizon de huit minutes en utilisant uniquement les données usuellement disponibles sur une installation de production d'eau chaude sanitaire. De plus, nous montrons que ce réseau de neurones artificiels peut s'adapter au moins à un bâtiment sur lequel il n'a pas été développé. Ces conclusions sont encourageantes pour l'utilisation à grande échelle de régulation prédictive pour la gestion de l'eau chaude sanitaire.

Perspective.

Á court terme, nous proposons qu'une campagne de mesure soit menée sur des foyers type afin de valider l'adaptation des réseau de neurones artificiels à tout type d'occupants, pour la prédiction de température intérieure, comme pour la prédiction des besoins d'eau chaude sanitaire. La méthode d'apprentissage proposée permet une utilisation rapide des modèles après l'installation d'une chaudière, cependant nous avons montré que le choix des exemples d'apprentissage a un fort impact la fiabilité des prédictions. Nous proposons qu'un deuxième apprentissage soit réalisé une fois que davantage de données auront été enregistrées. Toutefois, une étude devra être menée afin de déterminer comment composer automatiquement le jeu d'apprentissage optimal pour n'importe quel bâtiment, afin de ne pas perdre le caractère généraliste de nos réseaux de neurones artificiels.

Pour terminer, il est nécessaire d'augmenter l'horizon de prédiction des réseaux de neurones artificiels pour une utilisation au sein d'une régulation prédictive. Nous proposons dans un premier temps d'utiliser ce modèle en chaîne, c'est-à-dire d'utiliser la prédiction d'un modèle comme entrée d'un même modèle. Il sera alors nécessaire d'étudier l'augmentation de l'erreur de prédiction due à l'utilisation de prédiction en entrée du modèle. Áprès quoi ces modèles pourront être intégrés à une régulation prédictive pour la chaudière BoosHeat.

Bibliographie

[1] ADEME. Climat air et énergie, chiffres-clés. Technical report, 2016.

[2] Bruno Béranger. Les pompes à chaleur. 2009.

[3] Pervez Hameed Shaikh, NursSyarizal Bin Mohd Nor, Perumal Nallagownden, Irraivan Elamvazuthi, and Taib Ibrahim. A review on optimized control systems for building energy and comfort management of smart sustainable buildings. *Renewable and Sustainable Energy Reviews*, 34 :409–429, 2014.

[4] Petru-daniel Morosan. *Commande prédictive distribuée pour la régulation thermique des bâtiments*. PhD thesis, 2011.

[5] U Jordan, K Vajen, and W Streicher. *Solar heating System for houses : A design handbook for solar combisystem*. 2003.

[6] J Burch and C Christensen. Toward Development of an Algorithm for Main Water Temperature. *Intersolar 2007 Conference*, pages 7–10, 2007.

[7] Abdul Afram and Farrokh Janabi-Sharifi. Supervisory model predictive controller (MPC) for residential HVAC systems : Implementation and experimentation on archetype sustainable house in Toronto. *Energy & Buildings*, 154 :268–282, 2017.

[8] Christian Kandler, Patrick Wimmer, and Johannes Honold. Predictive control and regulation strategies of air-to-water heat pumps. *Energy Procedia*, 78 :2088–2093, 2015.

[9] Fatemeh Tahersima, Jakob Stoustrup, Henrik Rasmussen, and Soroush a. Meybodi. Economic COP optimization of a heat pump with hierarchical model predictive control. *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, pages 7583–7588, 2012.

[10] Clara Verhelst, Filip Logist, Jan Van Impe, and Lieve Helsen. Study of the optimal control problem formulation for modulating air-to-water heat pumps connected to a residential floor heating system. *Energy and Buildings*, 45 :43–53, 2011.

[11] Ion Hazyuk, Christian Ghiaus, and David Penhouet. Model Predictive Control of thermal comfort as a benchmark for controller performance. *Automation in Construction*, 43 :98–109, jul 2014.

[12] E Fuentes, D Schaefer, and J Salom. On the impact of realistic domestic hot water demand profiles on thermal storage stratification and energy efficiency. *Proceedings of the 99 Eurotherm Seminar. Lleida, Spain, 28–30 June 2014*, 2014.

[13] Xin Jin, Jeff Maguire, Dane Christensen, and National Renewable. Model predictive control of heat pump water heaters for energy efficiency. In *Proceedings from the 18th ACEEE Summer Study on Energy Efficiency in Buildings*, pages 133–145, 2014.

[14] Rasmus Halvgaard, Peder Bacher, Bengt Perers, Elsa Andersen, Simon Furbo, John B Jørgensen, Niels K Poulsen, and Henrik Madsen. Model predictive control for a smart solar tank based on weather and consumption forecasts. In *Energy Procedia*, volume 30, pages 270–278, 2012.

[15] Felix Iglesias and Peter Palensky. Profile-Based Control for Central Domestic Hot Water Distribution. *IEEE Transactions on Industrial Informatics*, 10(1) :697–705, feb 2014.

[16] Michael Dahl Knudsen and Steffen Petersen. Model predictive control for demand response of domestic hot water preparation in ultra-low temperature district heating systems. *Energy and Buildings*, 146 :55–64, 2017.

[17] M. Castilla, J.D. Álvarez, M. Berenguel, F. Rodríguez, J.L. Guzmán, and M. Pérez. A comparison of thermal comfort predictive control strategies. *Energy and Buildings*, 43(10) :2737–2746, oct 2011.

[18] Yudong Ma Yudong Ma, F. Borrelli, B. Hencey, a. Packard, and S. Bortoff. Model Predictive Control of thermal energy storage in building cooling systems. *Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*, 2009.

[19] L.Laret. *Contribution au développement de modèles mathématiques du comportement thermique transitoire de structures d'habitation.* PhD thesis, 1980.

[20] Frauke Oldewurtel, Alessandra Parisio, Colin N. Jones, Dimitrios Gyalistras, Markus Gwerder, Vanessa Stauch, Beat Lehmann, and Manfred Morari. Use of model predictive control and weather forecasts for energy efficient building climate control. *Energy and Buildings*, 45 :15–27, feb 2012.

[21] Antoine Lefort, Romain Bourdais, Guillaume Ansanay-Alex, and Hervé Guéguen. Hierarchical control method applied to energy management of a residential house. *Energy and Buildings*, 64 :53–61, 2013.

[22] Jan Široký, Frauke Oldewurtel, Jiří Cigler, and Samuel Prívara. Experimental analysis of model predictive control for an energy efficient building heating system. *Applied Energy*, 88(9) :3079–3087, 2011.

[23] J.D. Álvarez, J.L. Redondo, E. Camponogara, J. Normey-Rico, M. Berenguel, and P.M. Ortigosa. Optimizing building comfort temperature regulation via model predictive control. *Energy and Buildings*, 57 :361–372, feb 2013.

[24] Samuel Prívara, Jiří Cigler, Zdeněk Váňa, Frauke Oldewurtel, Carina Sagerschnig, and Eva Žáčeková. Building modeling as a crucial part for building predictive control. *Energy and Buildings*, 56 :8–22, jan 2013.

[25] Cj Yiu. Statistical modelling and forecasting schemes for air-conditioning system. 2008.

[26] Roberto Z. Freire, Gustavo H C Oliveira, and Nathan Mendes. Predictive controllers for thermal comfort optimization and energy savings. *Energy and Buildings*, 40(7) :1353–1365, 2008.

[27] Samuel Prívara, Jan Široký, Lukáš Ferkl, and Jiří Cigler. Model predictive control of a building heating system : The first experience. *Energy and Buildings*, 43(2-3) :564–572, feb 2011.

[28] Peter Lute and Dolf van Paassen. Optimal indoor temperature control using a predictor. *IEEE Control Systems Magazine*, 15(4) :4–10, 1995.

[29] Klaus Kaae Andersen, Henrik Madsen, and Lars H. Hansen. Modelling the heat dynamics of a building using stochastic differential equations. *Energy and Buildings*, 31(1) :13–24, 2000.

[30] Hunt. Neural networks for control systems—a survey. 1992.

[31] David Lindelöf, Hossein Afshari, Mohammad Alisafaee, Jayant Biswas, Miroslav Caban, Xavier Mocellin, and Jean Viaene. Field tests of an adaptive, model-predictive heating controller for residential buildings. *Energy and Buildings*, 99 :292–302, 2015.

[32] Andrew Kusiak and Guanglin Xu. Modeling and optimization of HVAC systems using a dynamic neural network. *Energy*, 42(1) :241–250, 2012.

[33] Merih Aydinalp, V Ismet Ugursal, and Alan S Fung. Modeling of the space and domestic hot-water heating energy-consumption in the residential sector using neural networks. *Applied Energy*, 79 :159–178, 2004.

[34] Project beca : Balanced european conservation approach ict services for resource saving in social housing. Technical report.

[35] Swinton M Hajo MR. Knight IK, Manning N. European and canadian non-hvac electric anddhwload profiles for use in simulating the performance of residential cogeneration systems, energy conservation in buildings and community systems programme. 2007.

[36] E Fuentes, L Arce, and J Salom. A review of domestic hot water consumption profiles for application in systems and buildings energy performance analysis. *Renewable and Sustainable Energy Reviews*, 81(February 2017) :1530–1547, 2017.

[37] Kaiser Ahmed, Petri Pylsy, and Jarek Kurnitski. Monthly domestic hot water profiles for energy calculation in Finnish apartment buildings. *Energy & Buildings*, 97 :77–85, 2015.

[38] Dane George. Mech 6340 Project High time-step resolution measured domestic hot water consumption profiles of single detached Canadian homes. 2015.

[39] Zuzana Krippelova and Jana Peràckova. MEASUREMENT OF HOT WATER CONSUMPTION. 1(13) :49–54, 2014.

[40] Christopher Bennett, Rodney A Stewart, and Cara D Beal. ANN-Based Residential Water End-use Demand Forecasting Model. *Expert Systems with Application*, 40(4) :1014–1023, 2012.

[41] R Willis, R Stewart, M Talebpour, A Mousavinejad, S Jones, and D Giurco. Revealing the impact of socio-demographic factors and efficient devices on end use water consumption : case of Gold Coast, Australia. *International Water Association (IWA) Efficient 2009.*, 2009.

[42] Edward Vine, Rick Diamond, and Rich Szydlowski. DOMESTIC HOT WATER CONSUMPTION IN FOUR LOW-INCOME APARTMENT BUILDINGS DOMESTIC HOT WATER CONSUMPTION IN FOUR LOW-INCOME APARTMENT BUILDINGS.

[43] Douglas S. Kenney, Christopher Goemans, Roberta Klein, Jessica Lowrey, and Kevin Reidy. Residential Water Demand Management : Lessons from Aurora, Colorado. *JAWRA Journal of the American Water Resources Association*, 44(1) :192–207, feb 2008.

[44] Yixing Shan, Lili Yang, Kim Perren, and Yanmin Zhang. ScienceDirect Household Water Consumption : Insight from a Survey in Greece and Poland. *Procedia Engineering*, 119 :1409–1418, 2015.

[45] Cara Beal, Rodney Stewart, and Tsu-Te Huang. South East Queensland Residential End Use Study : Baseline Results - Winter 2010. pages 1836–5558, 2010.

[46] Anas A. Makki, Rodney A. Stewart, Kriengsak Panuwatwanich, and Cara Beal. Revealing the determinants of shower water end use consumption : enabling better targeted urban water conservation strategies. *Journal of Cleaner Production*, 60 :129–146, dec 2013.

[47] Joakim Widén, Magdalena Lundh, Iana Vassileva, Erik Dahlquist, Kajsa Ellegård, and Ewa Wäckelgård. Constructing load profiles for household electricity and hot water from time-use data-Modelling approach and validation. *Energy and Buildings*, 41(7) :753–768, 2009.

[48] Bob Hendron, Jay Burch, and Greg Barker. Tool for Generating Realistic Residential Hot Water Event Schedules. *SimBuild 2010*, (August) :328–335, 328.

[49] Runming Yao and Koen Steemers. A method of formulating energy load profile for domestic buildings in the UK. *Energy and Buildings*, 37(6) :663–671, 2005.

[50] Ulrike Jordan and K Vajen. Realistic domestic hot-water profiles in different time scales. *Report for IEA-SHC Task*, pages 1–18, 2001.

[51] C. Sandels, J. Widén, and L. Nordström. Forecasting household consumer electricity load profiles with a combined physical and behavioral approach. *Applied Energy*, 131 :267–278, oct 2014.

[52] Steven G. Buchberger and Greg J. Wells. Intensity, Duration, and Frequency of Residential Water Demands. *Journal of Water Resources Planning and Management*, 122(1) :11–19, jan 1996.

[53] E. J. M. Blokker, J. H. G. Vreeburg, and J. C. van Dijk. Simulating Residential Water Demand with a Stochastic End-Use Model. *Journal of Water Resources Planning and Management*, 136(1) :19–26, jan 2010.

[54] Enrico Creaco, Raziyeh Farmani, Lydia Vamvakeridou-Lyroudia, Steven G. Buchberger, Zoran Kapelan, and Dragan A. Savić. Correlation or not Correlation ? This is the Question in Modelling Residential Water Demand Pulses. *Procedia Engineering*, 119 :1455–1462, 2015.

[55] Michael Negnevitsky and Koon Wong. Demand-Side Management Evaluation Tool. *IEEE Transactions on Power Systems*, 30(1) :212–222, jan 2015.

[56] David Fischer, Tobias Wolf, Johannes Scherer, and Bernhard Wille-Haussmann. A stochastic bottom-up model for space heating and domestic hot water load profiles for German households. *Energy and Buildings*, 124 :120–128, jul 2016.

[57] Hirohisa Aki, Tetsuya Wakui, and Ryohei Yokoyama. Development of a domestic hot water demand prediction model based on a bottom-up approach for residential energy management systems. *Applied Thermal Engineering*, 108 :697–708, 2016.

[58] Andrea Ferrantelli, Kaiser Ahmed, Petri Pylsy, and Jarek Kurnitski. Analytical modelling and prediction formulas for domestic hot water consumption in residential Finnish apartments. *Energy and Buildings*, 143 :53–60, 2017.

[59] Arthur Charpentier. Modèles de prévision Séries temporelles. 2014.

[60] G.E.P. Box and G.M. Jenkins. Time Series Analysis : Forecasting and Control. *Holden-Day, San Francisco*, 1976.

[61] Daniela Popescu and Elena Serban. Simulation of Domestic Hot - Water Consumption Using Time - Series Models. *6th IASME/WSEAS International Conference on HEAT TRANSFER, THERMAL ENGINEERING and ENVIRONMENT (HTE'08) Rhodes, Greece, August 20-22, 2008 Simulation*, pages 277–282, 2008.

[62] Linas Gelažanskas and Kelum Gamage. Forecasting Hot Water Consumption in Residential Houses. *Energies*, 8(11) :12702–12717, nov 2015.

[63] T Mitchell. *Machine Learning*. 1997.

[64] Vincent Bakker, Albert Molderink, Johann L. Hurink, and Gerard J. M. Smit. Domestic Heat Demand Prediction Using Neural Networks. In *2008 19th International Conference on Systems Engineering*, pages 189–194. IEEE, aug 2008.

[65] Peter Kepplinger, Gerhard Huber, and Jörg Petrasch. Field testing of demand side management via autonomous optimal control of a domestic hot water heater. *Energy and Buildings*, 127 :730–735, 2016.

[66] Dr-Ing habil Bernd Platzer Dr-Ing habil Volker Quaschning. Modellierung von Wasser- und Energieverbräuchen in Haushalten. (April), 2016.

[67] G Ladd and J Harrison. Electric Water Heating for Single-Family Residences. *Tech.Rep., Gilbert Associates, Inc.*, 1985.

[68] Guido Cervone, Laura Clemente-Harding, Stefano Alessandrini, and Luca Delle Monache. Short-term photovoltaic power forecasting using artificial neural networks and an analog ensemble. *Renewable Energy*, 108 :274 – 286, 2017.

[69] Cristian-Dragos Dumitru and Adrian Gligor. Daily average wind energy forecasting using artificial neural networks. *Procedia Engineering*, 181 :829 – 836, 2017. 10th International Conference Interdisciplinarity in Engineering, INTER-ENG 2016, 6-7 October 2016, Tirgu Mures, Romania.

[70] Bojana Kordanuli, Lidija Barjaktarović, Ljiljana Jeremić, and Meysam Alizamir. Appraisal of artificial neural network for forecasting of economic parameters. *Physica A : Statistical Mechanics and its Applications*, 465 :515 – 519, 2017.

[71] Pierre Lauret, Frederic Heymes, Serge Forestier, Laurent Aprin, Alexis Pey, and Marcia Perrin. Forecasting powder dispersion in a complex environment using artificial neural networks. *Process Safety and Environmental Protection*, 110 :71 – 76, 2017. Loss prevention and safety promotion in the process industries : issues and challenges.

[72] Donald W. Marquardt. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2) :431–441, jun 1963.

[73] ADEME. Les besoins d'eau chaude sanitaire en habitat individuel et collectif. Technical report, 2015.

[74] ADEME. Les besoins d'eau chaude sanitaire en habitat individuel et collectif. 2015.

[75] CSTB description de trnsys. `http://boutique.cstb.fr/thermique/272-trnsys.html`. Consulté : 20/03/2018.

[76] S.A. Klein, W.A. Beckman, J.W. Mitchell, J.A. Duffie, N.A. Duffie, T.L. Freeman, J.C. Mitchell, J.E. Braun, B.L. Evans, J.P. Krummer, R.E. Urban, A. Fiksel, J.W. Thornton, N.J. Blair, P.M. Williams, D.E. Bradley, T.P. McDowell, M. Krummert, and D.A. Arias. *Trnsys 17*, volume 5. 2014.

[77] N K Bansal and M S Bhandari. COMPARISON OF THE PERIODIC SOLUTION METHOD WITH TRNSYS AND SUNCODE FOR THERMAL BUILDING SIMULATION. *Solar Energy*, 57(1) :9–18, 1996.

[78] Roberto Bruno, Gianluca Pizzuti, and Natale Arcuri. The prediction of thermal loads in building by means of the EN ISO 13790 dynamic model : a comparison with TRNSYS. *Energy Procedia*, 101(101) :192–199, 2016.

[79] Islam Tibermacine and Noureddine Zemmouri. Effects of building typology on energy consumption in hot and arid region. *Energy Procedia*, 139(16) :664–669, 2017.

[80] Matthew Webb, Lu Aye, and Ray Green. Simulation of a biomimetic façade using TRNSYS. *Applied Energy*, 213 :670–694, 2018.

[81] Li Mei, David Infield, Ursula Eicker, and Volker Fux. Thermal modelling of a building with an integrated ventilated PV façade. *Energy & Buildings*, 35 :605–617, 2003.

[82] Paolo Valdiserri, Cesare Biserni, Giacomo Tosi, and Massimo Garai. Retrofit strategies applied to a tertiary building assisted by Trnsys energy simulation tool. *Energy Procedia*, 78 :765–770, 2015.

[83] Sung-Hwan Cho, Kee-Shik Shin, and M Zaheer-Uddin. THE EFFECT OF SLAT ANGLE OF WINDOWS WITH VENETIAN BLINDS ON HEATING AND COOLING LOADS OF BUILDINGS IN SOUTH KOREA. *Energy*, 20(12) :1225–1236, 1995.

[84] Akashi Mochida, Hiroshi Yoshino, Satoshi Miyauchi, and Teruaki Mitamura. Total analysis of cooling effects of cross-ventilation affected by microclimate around a building. *Solar Energy*, 80 :371–382, 2006.

[85] M N Assimakopoulos, G Mihalakakou, and H A Flocas. Simulating the thermal behaviour of a building during summer period in the urban environment. *Renewable Energy*, 32 :1805–1816, 2007.

[86] Farraj F Al-Ajmi and V I Hanby. Simulation of energy consumption for Kuwaiti domestic buildings. *Energy and Buildings*, 40 :1101–1109, 2008.

[87] Liping Wang, Julie Gwilliam, and Phil Jones. Case study of zero energy house design in UK. *Energy and Buildings*, 41 :1215–1222, 2009.

[88] Zhiyong Tian, Shicong Zhang, Huai Li, Yiqiang Jiang, Jiankai Dong, Boxiang Zhang, and Rongbo Yi. Investigations of Nearly (net) Zero Energy Residential Buildings in Beijing. *Procedia Engineering*, 121(121) :1051–1057, 2015.

[89] U Eicker, M Huber, P Seeberger, and C Vorschulze. Limits and potentials of office building climatisation with ambient air. *Energy & Buildings*, 38 :574–581, 2006.

[90] H Breesch, A Bossaer, and A Janssens. Passive cooling in a low-energy office building. *Solar Energy*, 79 :682–696, 2005.

[91] M Santamouris, G Mihalakakou, D N Asimakopoulos, and J Owen Lewis. ON THE COUPLING OF THERMOSTATICALLY CONTROLLED BUILDINGS WITH GROUND AND NIGHT VENTILATION PASSIVE DISSIPATION TECHNIQUES. *Solar Energy*, 60(314) :191–197, 1997.

[92] P Blondeau, M Spérandio, and F Allard. NIGHT VENTILATION FOR BUILDING COOLING IN SUMMER. *Solar Energy*, 61(5) :327–335, 1997.

[93] Viktor Dorer and Dieter Breer. Residential mechanical ventilation systems : performance criteria and evaluations. *Energy and Buildings*, 27(247), 1998.

[94] Kezhi Yu, Zhaoyang Cao, and Yanling Liu. Research on the optimization control of the central air-conditioning system in university classroom buildings based on TRNSYS software. *Procedia Engineering*, 205 :1564–1569, 2017.

[95] Michael Kummert, Philippe Andre, and Jacques Nicolas. OPTIMAL HEATING CONTROL IN A PASSIVE SOLAR COMMERCIAL BUILDING. *Solar Energy*, 69(1-6) :103–116, 2000.

[96] Gilles Fraisse, Joseph Virgone, and Reneâ Yezou. A numerical comparison of different methods for optimizing heating-restart time in intermittently occupied buildings. *Applied Energy*, 62 :125–140, 1999.

[97] M Santamouris, G Mihalakakou, A Argiriou, and D N Asimakopoulos. ON THE PERFORMANCE OF BUILDINGS COUPLED WITH EARTH TO AIR HEAT EXCHANGERS. *Solar Energy*, 54(6) :375–380, 1995.

[98] A Argiriou, N Klitsikas, C A Balaras, and D N Asimakopoulos. Active solar space heating of residential buildings in northern Hellas- a case study. *Energy and Buildings*, 26 :215–221, 1997.

[99] L T Terziotti, M L Sweet, and J T Mcleskey. Modeling seasonal solar thermal energy storage in a large urban residential building using TRNSYS 16. *Energy & Buildings*, 45 :28–31, 2011.

[100] Marshall L Sweet and James T Mcleskey. Numerical simulation of underground Seasonal Solar Thermal Energy Storage (SSTES) for a single family dwelling using TRNSYS. *Solar Energy*, 86 :289–300, 2012.

[101] F Calise. Thermoeconomic analysis and optimization of high efficiency solar heating and cooling systems for different Italian school buildings and climates. *Energy and Buildings*, 42 :992–1003, 2010.

[102] R. Chargui and H. Sammouda. Modeling of a residential house coupled with a dual source heat pump using TRNSYS software. *Energy Conversion and Management*, 81 :384–399, 2014.

[103] Commissariat général au développement Durable. Bilan énergétique de la France en 2016 Données définitives. Technical report, 2017.

[104] CEREN. Données statistiques du CEREN. pages 1–32, 2015.

[105] Réglementation thermique 2012. `http://www.rt-batiment.fr`. Consulté : 20/03/2018.

[106] Hao Huang, Lei Chen, and Eric Hu. A neural network-based multi-zone modelling approach for predictive control system design in commercial buildings. *Energy and Buildings*, 97 :86–97, 2015.

[107] Jin Woo Moon, Ji-Hyun Lee, Jae. D Chang, and Sooyoung Kim. Preliminary performance tests on artificial neural network models for opening strategies of double skin envelopes in winter. *Energy and Buildings*, 75 :301–311, 2014.

Table des figures

1.1 Schéma d'une pompe à chaleur. 3
1.2 Cycle thermodynamique d'un PAC sur le diagramme de Molier, extrait de [2]. 4
1.3 Impact des températures sur le cycle thermodynamique, extrait de [2]. 5
1.4 Vue CAO du compresseur thermique BoostHeat. 6
1.5 Représentation temporelle du fonctionnement des MPC, extrait de [4]. 7
1.6 Principe de fonctionnement des MPC. 8
1.7 Schématisation d'une MPC pour le contrôle de l'ECS, extrait de [13] 9
1.8 Modèle du bâtiment par analogie électrique, extrait de [20] 11
1.9 Architecture de réseaux de neurones pour la modélisation du bâtiment et du climat, extrait de [31]. 13
1.10 Comparaison des grandeurs prédites aux grandeurs observées, à gauche l'énergie consommée pour le chauffage, à droite la température intérieure, extrait de [32]. 13
1.11 Présentation des entrées utilisées pour la prédiction de consommation annuelle de chauffage, extrait de [33]. 14
1.12 Profils de consommations journalières de différents foyer à Manresa[34], comparaison avec le profil européen moyen[35], en rouge, extrait de [36]. 15
1.13 Résultats des analyses de données de l'étude de [48]. 17
1.14 Comparaison des prédictions fait par un ANN par rapport aux observations, extrait de [40]. 20
1.15 Exemple de réseau de neurones artificiels pour la prédiction de consommation d'ECS, extrait de [64]. 21
1.16 Comparaison des prédictions de consommation d'ECS faite par un modèle développé par Bakker et al. et des observations, extrait de [64]. 22
1.17 Présentation des entrées utilisées pour la prédiction de consommation annuelle d'ECS, extrait de [33]. 23
1.18 Représentation du i-éme neurone artificiel de la couche j. 24
1.19 Représentation d'un perceptron multicouche 26
1.20 Allure théorique des courbes d'apprentissage. 28

2.1 Exemple de données expérimentales traitées. En vert est représenté la puissance d'ECS, en jaune la puissance de chauffage, en bleu la température intérieure,en rouge la température extérieure. 32

2.2 Exemple de données expérimentales brutes. 33
2.3 Cas 1 : Photographie panoramique de la maison et des panneaux solaires. 33
2.4 Cas 2 : Photographie de la maison et des panneaux solaires. 34
2.5 Cas 3 : Photographie panoramique de la maison. 34
2.6 ECS : Positionnement des cas d'études par rapport aux statitiques de consommation française. 36
2.7 Répartition des logements individuels français par classe énergétique, extrait de [74]. 36
2.8 Exemple de géométrie de bâtiment à l'aide du logiciel SketchUp. . . . 37
2.9 Exemple de fichier généré par TRNSYS. 38
2.10 Etiquettes énergétiques 39
2.11 Carte des climats réglementaires 40

3.1 Étape de développement des modèles. 45
3.2 Architecture du réseau de neurones. 46
3.3 Courbes d'apprentissage, à gauche du cas 1, à droite du cas 2 : $RMSE$ sur le jeu d'apprentissage (en bleu) et sur le jeu de validation (en vert) en fonction du nombre d'exemples d'apprentissage. 48
3.4 Schéma des étapes nécessaire à l'adaptation et l'utilisation du ANN. . 49
3.5 Cas 1 : comparaison entre la température réelle et celle prédite sur tout le jeu de test. En haut : courbe de la température intérieure mesurée T_{int} (en bleu) et de la température intérieure prédite $\hat{T}_{int}$ (en vert), en °C, en fonction de la date. En bas : courbe de la puissance de chauffage mesurée P_{chauf} (en rouge), en kW, et de l'erreur sur la prédiction de la température $\varepsilon = T_{int} - \hat{T}_{int}$ (en orange), en °C, en fonction de la date. 52
3.6 Cas 2 : comparaison entre la température réelle et celle prédite sur tout le jeu de test. En haut : courbe de la température intérieure mesurée T_{int} (en bleu) et de la température intérieure prédite $\hat{T}_{int}$ (en vert), en °C, en fonction de la date. En bas : courbe de la puissance de chauffage mesurée P_{chauf} (en rouge), en kW, et de l'erreur sur la prédiction de la température $\varepsilon = T_{int} - \hat{T}_{int}$ (en orange), en °C, en fonction de la date. 53
3.7 Cas 1 : influence de la puissance de chauffage sur l'erreur de prédiction. En haut : erreur $\varepsilon = T_{int} - \hat{T}_{int}$, en °C, en fonction de la puissance de chauffage mesurée P_{chauf},en kW. En bas : erreur $\varepsilon = T_{int} - \hat{T}_{int}$, en °C, auquel nous avons appliqué une moyenne glissante, en fonction de la puissance de chauffage mesurée P_{chauf},en kW. 54
3.8 Cas 2 : influence de la puissance de chauffage sur l'erreur de prédiction. En haut : erreur $\varepsilon = T_{int} - \hat{T}_{int}$, en °C, en fonction de la puissance de chauffage mesurée P_{chauf},en kW. En bas : erreur $\varepsilon = T_{int} - \hat{T}_{int}$, en °C, auquel nous avons appliqué une moyenne glissante, en fonction de la puissance de chauffage mesurée P_{chauf},en kW. 55

3.9 Cas 1 : comparaison entre la température réelle et celle prédite du 05/02/05 au 12/02/05. Courbe de la température intérieure mesurée T_{int} (en bleu) et de la température intérieure prédite $\hat{T}_{int}$ (en vert), en °C, en fonction de la date. 56
3.10 Cas 2 : comparaison entre la température réelle et celle prédite du 05/02/05 au 12/02/05. Courbe de la température intérieure mesurée T_{int} (en bleu) et de la température intérieure prédite $\hat{T}_{int}$ (en vert), en °C, en fonction de la date. 56
3.11 Méthode utilisée pour adapter le ANN précédemment développé aux cas simulés. 58
3.12 Influence de la puissance de chauffage consommée, données regroupées selon la superficie du bâtiment. Sur les 3 graphiques, l'abscisse représente la puissance de chauffage consommée en kW, l'ordonnée représente l'erreur faite sur T_{int} en °C. Les graphiques représentent l'erreur ε faite sur toutes les prédictions des 120 cas simulés, soit 561 174 prédictions. Les données sont regroupées selon la superficie du cas : $50m^2$ sur la $1_{ére}$ ligne, $100m^2$ sur la $2^{éme}$ ligne et $150m^2$ sur la $3^{éme}$ ligne. 59
3.13 Influence de la puissance de chauffage consommée, données regroupées selon la classe énergétique du bâtiment. Sur les 5 graphiques, l'abscisse représente la puissance de chauffage consommée en kW, l'ordonnée représente l'erreur en °C. Les graphiques représentent l'erreur ε faite sur toutes les prédictions des 120 cas simulés, soit 561 174 prédictions. Les données sont regroupées selon la classe énergétique du cas : A sur la $1_{ére}$ ligne, B sur la $2^{éme}$ ligne, C sur la $3^{éme}$ ligne, D sur la $4^{éme}$ ligne, E sur la $5^{éme}$ ligne. 60
3.14 Influence de la puissance de chauffage consommée, données regroupées selon le climat auquel est soumis le bâtiment. Sur les 8 graphiques, l'abscisse représente la puissance de chauffage consommée en kW, l'ordonnée représente l'erreur en °C. Les graphiques représentent l'erreur ε faite sur toutes les prédictions des 120 cas simulés, soit 561 174 prédictions. Les données sont regroupées selon le climat du cas : H1a sur la $1_{ére}$ ligne, H1b sur la $2^{éme}$ ligne, H1c sur la $3^{éme}$ ligne, H2a sur la $4^{éme}$ ligne, H2b sur la $5^{éme}$ ligne, H2c sur la $6^{éme}$ ligne, H2d sur la $7^{éme}$ ligne, H3 sur la $8^{éme}$ ligne. 62
3.15 Cas 1 : répartition des exemples d'apprentissage et de test en fonction de la puissance de chauffage. En ordonnée le nombre d'exemples, en abscisse la puissance de chauffage, en kW. À droite le jeu d'apprentissage, à gauche le jeu de test. 63
3.16 Cas 2 : répartition des exemples d'apprentissage et de test en fonction de la puissance de chauffage. En ordonnée le nombre d'exemples, en abscisse la puissance de chauffage, en kW. À droite le jeu d'apprentissage, à gauche le jeu de test. 63

3.17 120 cas simulés : répartition des exemples d'apprentissage et de test en fonction de la puissance de chauffage. En ordonnée le nombre d'exemples, en abscisse la puissance de chauffage, en kW. À droite le jeu d'apprentissage, à gauche le jeu de test. 64
3.18 ε : Diagramme en boite de l'erreur signée ε en ordonnée, en °C, pour l'ensemble des cas, en abscisse. 65
3.19 ε : influence de la superficie. Diagramme en boite de l'erreur signée ε en ordonnée, en °C, en fonction de la superficie du cas, en abscisse. . 66
3.20 ε : influence du climat. Diagramme en boite de l'erreur signée ε en ordonnée, en °C, en fonction du climat du cas, en abscisse. 68
3.21 ε : influence de la classe énergétique. Diagramme en boite de l'erreur signée ε en ordonnée, en °C, en fonction de la superficie du cas, en abscisse. 69

4.1 Étapes de développement des modèles. 75
4.2 Courbes d'apprentissage des trois modèles du cas 1. 77
4.3 Modèle A8 : visualisation de l'offset. Sur les 2 graphiques l'abscisse représente la date, au format jour/mois, l'ordonnée représente l'énergie en kWh. Le graphique du haut (A) représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et l'énergie prédite $\hat{E}(t+8)$ par le modèle A8 en vert, avec des marqueurs x, sur le jeu d'apprentissage. Le graphique du bas (B) représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et celle prédite $\hat{E}(t+8)$ par le modèle A8 en vert, avec des marqueurs x, sur le jeu de validation. 81
4.4 Modèle B8 : visualisation de l'offset. Sur les 2 graphiques l'abscisse représente la date, au format jour/mois, l'ordonnée représente l'énergie en kWh. Le graphique du haut (A) représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et l'énergie prédite $\hat{E}(t+8)$ par le modèle B8 en vert, avec des marqueurs x, sur le jeu d'apprentissage. Le graphique du bas (B) représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et celle prédite $\hat{E}(t+8)$ par le modèle B8 en vert, avec des marqueurs x, sur le jeu de validation. 82
4.5 Modèle C8 : visualisation de l'offset. Sur les 2 graphiques l'abscisse représente la date, au format jour/mois, l'ordonnée représente l'énergie en kWh. Le graphique du haut (A) représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et l'énergie prédite $\hat{E}(t+8)$ par le modèle C8 en vert, avec des marqueurs x, sur le jeu d'apprentissage. Le graphique du bas (B) représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et celle prédite $\hat{E}(t+8)$ par le modèle C8 en vert, avec des marqueurs x, sur le jeu de validation. 83

4.6 Modèle A8 : visualisation des prédictions sur une journée extraite du jeu de test. Sur les deux graphiques, l'abscisse représente la date, au format heure/minute, l'ordonnée représente l'énergie en kWh. Le graphique de gauche représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et l'énergie prédite $\hat{E}(t+8)$ par le modèle A8 en vert, avec des marqueurs x, ainsi que l'erreur ε, en jaune avec des marqueurs ronds, sur la journée du 2 février 2005, incluse dans le jeu de test. Le graphique de droite représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et celle prédite $\hat{E}(t+8)$ par le modèle A8 en vert, avec des marqueurs x,ainsi que l'erreur ε, en jaune avec des marqueurs ronds, sur la journée du 8 août 2004, incluse dans le jeu de test. 85
4.7 Modèle B8 : visualisation des prédictions sur une journée extraite du jeu de test. Sur les deux graphiques, l'abscisse représente la date, au format heure/minute, l'ordonnée représente l'énergie en kWh. Le graphique de gauche représente l'énergie réelle consommée pour l'ECS pendant 64 minutes $E(t+8)$, en bleu, avec un marqueur + et l'énergie prédite $\hat{E}(t+8)$ par le modèle B8 en vert, avec des marqueurs x, ainsi que l'erreur ε, en jaune avec des marqueurs ronds, sur la journée du 2 février 2005, incluse dans le jeu de test. Le graphique de droite représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et celle prédite $\hat{E}(t+8)$ par le modèle B8 en vert, avec des marqueurs x,ainsi que l'erreur ε, en jaune avec des marqueurs ronds, sur la journée du 8 août 2004, incluse dans le jeu de test. 87
4.8 Modèle C8 : visualisation des prédictions sur une journée extraite du jeu de test. Sur les deux graphiques, l'abscisse représente la date, au format heure/minute, l'ordonnée représente l'énergie en kWh. Le graphique de gauche représente l'énergie réelle consommée pour l'ECS pendant 64 minutes $E(t+8)$, en bleu, avec un marqueur + et l'énergie prédite $\hat{E}(t+8)$ par le modèle C8 en vert, avec des marqueurs x, ainsi que l'erreur ε, en jaune avec des marqueurs ronds, sur la journée du 2 février 2005, incluse dans le jeu de test. Le graphique de droite représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et celle prédite $\hat{E}(t+8)$ par le modèle C8 en vert, avec des marqueurs x,ainsi que l'erreur ε, en jaune avec des marqueurs ronds, sur la journée du 8 août 2004, incluse dans le jeu de test. 89
4.9 Étapes d'adaptation du ANN au cas 2 et au cas 3. 90
4.10 Cas 2 : visualisation de l'offset.L'abscisse représente la date, au format jour/mois, l'ordonnée représente l'énergie en kWh. Le graphique représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et l'énergie prédite $\hat{E}(t+8)$ par le modèle, avec des marqueurs x, sur le jeu d'apprentissage. . . . 91

4.11 Cas 3 : visualisation de l'offset. L'abscisse représente la date, au format jour/mois, l'ordonnée représente l'énergie en kWh. Le graphique du haut (A) représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et l'énergie prédite $\hat{E}(t+8)$ par le modèle, avec des marqueurs x, sur le jeu d'apprentissage. 92
4.12 Cas 2 : visualisation des prédictions sur une journée extraite du jeu de test. Sur les deux graphiques, l'abscisse représente la date, au format heure/minute, l'ordonnée représente l'énergie en kWh. Le graphique de gauche représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et l'énergie prédite $\hat{E}(t+8)$ par le modèle A8 en vert, avec des marqueurs x, ainsi que l'erreur ε, en jaune avec des marqueurs rond, sur la journée du 2 fevrier 2005, incluse dans le jeu de test. Le graphique de droite représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et celle prédite $\hat{E}(t+8)$ par le modèle A8 en vert, avec des marqueurs x,ainsi que l'erreur ε, en jaune avec des marqueurs rond, sur la journée du 8 août 2004, incluse dans le jeu de test. 93
4.13 Cas 3 : visualisation des prédictions sur une journée extraite du jeu de test. Sur les deux graphiques, l'abscisse représente la date, au format heure/minute, l'ordonnée représente l'énergie en kWh. Le graphique de gauche représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et l'énergie prédite $\hat{E}(t+8)$ par le modèle A8 en vert, avec des marqueurs x, ainsi que l'erreur ε, en jaune avec des marqueurs rond, sur la journée du 2 fevrier 2005, incluse dans le jeu de test. Le graphique de droite représente l'énergie réelle consommée pour l'ECS pendant 8 minutes $E(t+8)$, en bleu, avec un marqueur + et celle prédite $\hat{E}(t+8)$ par le modèle A8 en vert, avec des marqueurs x,ainsi que l'erreur ε, en jaune avec des marqueurs rond, sur la journée du 8 août 2004, incluse dans le jeu de test. 93

Liste des tableaux

2.1 Informations sur les mesures expérimentales 32
2.2 Description des cas d'études 35
2.3 Informations techniques sur les mesures expérimentales 41

3.1 Valeurs utilisées pour la normalisation 46
3.2 Temps de calcul, en seconde, des principales étapes de développement du ANN pour une machine équipée d'un processeur 3.1 GHz Intel Core i7 et de 16Go de mémoire vive 50
3.3 Valeur de l'erreur moyenne signée $\bar{\varepsilon}$ et non signée $RMSE$, ainsi que de l'écart-type σ et des extremum min et max associés pour les deux cas expérimentaux 57
3.4 Valeur de l'erreur moyenne signée $\bar{\varepsilon}$ et non signée $RMSE$, ainsi que de l'écart-type σ et des extremum min et max associés pour les 120 cas. 65
3.5 Valeur de l'erreur moyenne signée $\bar{\varepsilon}$ et non signée $RMSE$, ainsi que de l'écart-type σ et des extremum min et max associés pour les trois superficies étudiées $50m^2, 100m^2, 150m^2$ 67
3.6 Valeur de l'erreur moyenne signée $\bar{\varepsilon}$ et non signée $RMSE$, ainsi que de l'écart-type σ et des extrémum min et max associés pour les cinq climats étudiées A, B, C, D et E 67
3.7 Valeur de l'erreur moyenne signée $\bar{\varepsilon}$ et non signée $RMSE$, ainsi que de l'écart-type σ et des extremum min et max associé pour les huit climats étudiés H1a, H1b, H1c, H2a, H2b, H2c, H2d, H3 69

4.1 Déscription des modéles 74
4.2 Valeurs utilisées pour la normalisation 76
4.3 Temps de calcul, en seconde, des principales étapes de développement du ANN pour une machine équipée d'un processeur 3.1 GHz Intel Core i7 et de 16Go de mémoire vive. 78
4.4 Offset calculé sur les modèles du cas 1 pour chacun des ANN , en kWh. 80
4.5 Prédiction à 8 minutes sur le jeu de test : valeur de l'erreur moyenne signée $\bar{\varepsilon}$ et non signée $RMSE$, ainsi que de l'écart-type σ et des extremum min et max associés, pour les trois modèles du cas 1. Toutes les valeurs sont indiquées en kWh 84

4.6 Modèle du cas 1, A8 : Caractéristique des instants sans consommation et performance du modèle. 85
4.7 Modèle du cas 1, A8 : Valeurs de l'erreur relative sur une journée. . . 86
4.8 RMSE faite sur le jeu d'apprentissage par les modèles du cas 2 et du cas 3, en kWh. 91
4.9 Offset calculé sur les modèles des cas 2 et cas 3, en kWh. 92
4.10 Cas 2 et cas 3 : Prédiction à 8 minutes sur le jeu de test : valeur de l'erreur moyenne signée $\bar{\varepsilon}$ et non signée $RMSE$, ainsi que de l'écart-type σ et des extremum min et max associés, pour les modèles des cas 2et cas 3. Toute les valeurs sont indiquées en kWh 94
4.11 Modèles des cas 2 et cas 3 : Caractéristique des instants sans consommation et performance du modèle. 95
4.12 Modèle du cas 2 : Valeurs de l'erreur relative sur une journée 95
4.13 Modèle du cas 3 : Valeurs de l'erreur relative sur une journée 96

Modèles prédictifs et adaptatifs pour la gestion énergétique du bâtiment résidentiel individuel : réseaux de neurones artificiels basés sur les données usuellement disponibles

Résumé : L'utilisation de régulation prédictive permet de diminuer la consommation d'énergie des bâtiments résidentiels sans diminuer le confort de l'habitant. C'est dans ce but que la société BoostHeat développe une chaudière thermodynamique à grande efficacité énergétique. La production simultanée de l'eau chaude sanitaire (ECS) et du chauffage permet de nombreuses stratégies de régulation pour en optimiser les performances. L'utilisation de régulations prédictives permet d'anticiper les besoins énergétiques, de prendre en compte l'impact de l'inertie du bâtiment sur la température intérieure et ainsi de faire des choix de gestion de productions minimisant la consommation énergétique. Les modèles utilisés aujourd'hui dans les régulations prédictives sont contraignants. En effet, ces modèles nécessitent de grandes quantités de données, soit sur un échantillon représentatif de bâtiments, soit sur chaque bâtiment modélisé. Ils peuvent également avoir besoin d'études détaillées sur le bâtiment, les occupants et leurs habitudes de consommation. Afin de permettre à la société BoostHeat d'utiliser une régulation prédictive sur sa chaudière sans passer par une étape complexe de modélisation à chaque installation de chaudière, nous proposons des modèles adaptatifs utilisant les informations communément disponibles sur une installation classique. Nous choisissons de développer des réseaux de neurones artificiels pour la prédiction d'une part des consommations d'ECS et d'autre part de la température intérieure du bâtiment. Les réseaux de neurones artificiels sont déjà utilisés pour modéliser les consommations énergétiques d'un bâtiment spécifique, cependant nos modèles sont génériques et s'adaptent automatiquement au bâtiment dans lequel la chaudière est installée. Plusieurs modèles sont développés afin d'étudier l'impact du choix des entrées, des quantités de données d'apprentissage et de l'architecture du réseau de neurones artificiels sur la qualité de la prédiction. Les modèles de prédiction de consommation d'ECS ainsi réalisés sont testés sur trois cas expérimentaux alors que les modèles de prédiction de température intérieure sont testés sur deux cas expérimentaux et cent-vingt cas simulés. Cela permet de tester leur adaptation à l'ensemble du parc immobilier français. Nous montrons, pour la prédiction de consommation d'ECS comme pour la prédiction de température intérieure, que deux semaines de données collectées suffisent à une bonne adaptation des modèles à un cas spécifique. Le modèle le plus performant pour la prédiction de consommation d'eau chaude sanitaire a uniquement besoin des consommations des instants précédents. Le modèle de prédiction de température intérieure a de meilleures performances sur les bâtiments les moins isolés. Les résultats obtenus sont prometteurs pour l'application des régulations prédictives à grande échelle.
Mots-clés : Réseau de neurones artificiels, Apprentissage par rétropropagation de l'erreur, Prédiction de consommation d'ECS, Prédiction de température intérieure du bâtiment, Pompe à chaleur, Bâtiment résidentiel individuel

---

Predictive and adaptive models for individual residential building energy management : Artificial neural networks based on commonly available data.

Abstract : The use of predictive control permits to reduce the energy consumption of residential buildings without reducing the comfort of the inhabitant. The company BoostHeat develops a thermodynamic furnace with high energy efficiency for this purpose. Simultaneous production of domestic hot water (DHW) and heating allows many control strategies to optimize performance. The use of predictive controls makes it possible to anticipate energy needs, to take into account the impact of building inertia on indoor temperature and thus to make production management choices that minimize energy consumption. The models used today in predictive controls are binding. Indeed, these models require large amounts of data, either on a representative sample of buildings or on each modeled building. They may also need detailed studies of the building, the occupants and their consumption practices. In order to allow BoostHeat to use predictive control without going through a complex modeling step at every furnace installation, we propose adaptive models using information commonly available on a typical installation. We choose to develop artificial neural networks for the prediction on the one hand of the consumptions of DHW and on the other hand of the ambiant temperature of the building. Artificial neural networks are already used to model the energy consumption of a specific building, however our models are generic and automatically adapt to the building in which the furnace is installed. Many models are developed to study the impact of the choice of inputs, amounts of learning data and artificial neural network architecture on the accuracy of prediction. The DHW consumption prediction models are tested on three experimental cases while the indoor temperature prediction models are tested on two experimental cases and one hundred and twenty simulated cases. This makes it possible to test their adaptation to the entire French housing stock. We show, for the prediction of DHW consumption as for the indoor temperature prediction, that two weeks of collected data are sufficient for a good adaptation of the models to a specific case. The most efficient model for the prediction of domestic hot water consumption only needs the consumptions of the previous instants. The indoor temperature prediction model performs better on less isolated buildings. The results obtained are promising for the application of predictive controls on a large scale.
Keywords : Artificial neural network, Back-propagation learning, Forecasting of DHW demand, Forecasting of indoor temperature, Heat pump, Individual residential building.