Détection d'ouvertures de fenêtres au moyen de réseaux de neurones artificiels LSTM

Enguerrand de Rautlin de la Roy*[1], Thomas Recht[1], Akka Zemmari[2], Pierre Bourreau[3], Kamel Houchat[3], Laurent Mora[1]

[1] Univ. Bordeaux, CNRS, Bordeaux INP, I2M, UMR 5295, F-33400, Talence, France
[2] Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400, Talence, France
[3] Nobatek/INEF4, 9 rue Jean-Paul Alaux, F-33000 Bordeaux, France
***enguerrand.de-rautlin-de-la-roy@u-bordeaux.fr**

RESUME. L'utilisation croissante de systèmes de Gestion Technique de Bâtiment (GTB) et d'équipements connectés dans le secteur du bâtiment, offre la possibilité d'étudier et modéliser, à l'aide des données issues de mesures (température, concentration de CO_2, etc.), les sollicitations auxquelles le bâtiment est réellement soumis lors de son utilisation. Parmi ces sollicitations, l'ouverture des fenêtres peut influencer fortement la consommation énergétique des bâtiments, mais est rarement mesurée faute de moyen ou car jugée trop intrusive. Cependant il est possible, à l'aide de réseaux de neurones, d'observer les impacts des ouvertures de fenêtres sur d'autres grandeurs au travers de motifs récurrents afin de les reconstruire. Une comparaison de modèles de réseaux de neurones « Long Short-Term Memory » (LSTM), est dans ce but proposée pour évaluer la performance qui peut être attendue sur la détection d'ouvertures de fenêtres en saison de chauffe selon les données à disposition (température, concentration de CO_2, etc.).

MOTS-CLÉS : Intelligence artificielle, LSTM, détection d'ouvertures de fenêtres

ABSTRACT. The increasing use of Building Management System (BMS) and Internet of Things (IoT) devices in the building sector provide the opportunity to study and to model the users' impact on the building with the collected data (temperature, CO_2 concentration, etc.). Among those, windows' opening status may have a strong influence on the energy consumption of buildings but are rarely measured due to a lack of funds or because it is considered too intrusive. However, it is possible, with the help of neural networks, to observe the impacts of window openings on other measurements through specific patterns in order to reconstruct them. A comparison of « Long Short-Term Memory » (LSTM) neural network models is therefore proposed to evaluate the performance that can be expected on window openings detection during the heating season depending of the available data (temperature, CO_2 concentration, etc.).

KEYWORDS: Artificial intelligence, LSTM, window opening detection

1. INTRODUCTION

L'utilisation croissante de systèmes de Gestion Technique de Bâtiment (GTB) et d'équipements connectés dans les bâtiments offre la possibilité d'étudier et de modéliser, à l'aide des données issues de mesures (température, concentration de CO_2, ouverture des portes et fenêtres, *etc*.), les sollicitations auxquelles le bâtiment est réellement soumis lors de son utilisation. Parmi ces sollicitations, le comportement des occupants peut influencer fortement la consommation énergétique des bâtiments. Un geste courant est souvent cité en exemple : l'ouverture des fenêtres [1]. Cependant et bien qu'essentiels à la compréhension du fonctionnement et du comportement thermique d'un bâtiment, les moyens de mesures des ouvertures sont rarement mis en œuvre et exploités le cas échéant. Contrairement aux capteurs d'ambiance (tels que la température et l'humidité relative de l'air) bien plus répandus, les capteurs d'ouverture de fenêtres sont généralement considérés comme plus intrusifs, nombreux à installer et pouvant présenter de problèmes de fiabilité. Ainsi, les ouvertures sont souvent approchées *via* un modèle statistique ou l'application de règles expertes (ratios, *etc*.) pouvant conduire à des écarts importants entre les modélisations et les observations [2]. Comme détaillé par [3], il n'y a, à l'heure actuelle, pas de consensus sur la manière dont l'occupant interagit avec le bâtiment ainsi que sur les facteurs influençant son comportement. Il n'est pas aisé d'obtenir une modélisation fidèle des actions des occupants en l'absence de mesures dédiées ou en cas d'utilisation de données de mauvaise qualité (anomalies, *etc*.). Cependant, il est souvent possible d'observer les impacts de ces actions sur d'autres grandeurs au travers de motifs récurrents et ainsi remonter à l'origine de l'action pour la modéliser. De nos jours, la reconnaissance de motifs (ou classification) au moyen de techniques d'apprentissage automatique est utilisée dans de nombreux domaines afin de répondre à des besoins variés [4,5]. Cependant, cette approche reste encore peu abordée dans le domaine du bâtiment ou utilise des métriques d'évaluation peu adaptées, malgré certaines études récentes [6]. Dans ce contexte, le présent travail vise à apporter une contribution sur la détection (par modélisation) d'ouverture-fermeture de fenêtres. Une comparaison de modèles de réseaux de neurones « *Long Short-Term Memory* » (LSTM), est dans ce but proposée pour évaluer la performance qui peut être attendue sur la détection d'ouvertures de fenêtres en saison de chauffe selon les données à disposition (température et humidité relative de l'air, concentration de CO_2).

2. METHODOLOGIE

2.1. RESEAU DE NEURONES ET STRUCTURE : LSTM

Les réseaux de neurones récurrents (*Recurrent Neural Network RNN*) sont une sous-classe des réseaux de neurones artificiels. Ce sont des algorithmes possédant une mémoire interne, particulièrement adaptés aux données séquentielles. Ils sont régulièrement utilisés pour de la traduction automatique, de la reconnaissance vocale ou encore de la classification d'images. Cependant, en raison du problème de dissipation du gradient (lié à leur méthode d'apprentissage par descente du gradient), les informations anciennes sont facilement oubliées. On dit qu'ils ont une mémoire courte. Le réseau « *Long Short-Term Memory* » (LSTM) est un réseau de neurones récurrents à mémoire court-terme et long-terme qui permet, grâce à sa structure, de répondre partiellement au problème de dissipation du gradient. Cela fait des modèles LSTM de bons candidats pour effectuer de la reconnaissance de motifs sur des séries temporelles.

Le modèle LSTM utilisé pour cette étude est le résultat de plusieurs essais et est composé d'une couche d'entrée de 128 unités suivi d'une couche intermédiaire cachée de 64 unités avec décrochage et une couche de sortie dense de 2 unités, chacune renvoyant la probabilité que la fenêtre soit ouverte (valeur associée : 1) ou fermée (valeur associée : 0). La probabilité la plus importante des deux correspond à la prédiction de l'état de la fenêtre réalisée par le modèle. Enfin, les paramètres initiaux de chaque couche sont générés suivant une loi de distribution normale aléatoire.

2.2. ÉVALUATION DES PERFORMANCES

La modélisation des ouvertures de fenêtres est un problème de classification binaire relativement courant en apprentissage automatique. Ce problème est composé de deux classes traduisant l'état des fenêtres : état ouvert (minoritaire) et état fermé (majoritaire). Comme présenté par [6], de nombreuses métriques peuvent être utilisées pour évaluer les performances d'un tel modèle. Certaines évaluent les performances du modèle sur une classe en particulier telles que le *F1-score*, le taux de vrais positifs ou le taux de vrais négatifs. D'autres évaluent plutôt les performances globales du modèle telles que l'AUC-ROC (*Area Under The Curve - Receiver Operating Characteristics*) ou l'exactitude. Cependant, ces métriques d'évaluation ne permettent pas toujours de différencier les modèles entre eux. En effet, comme illustré par la Figure 1, une ouverture se traduit par un ou plusieurs états ouverts (cases pleines) encadrés par un ou plusieurs états fermés (cases vides). En utilisant l'une des métriques d'évaluation évoquées, le score de performance obtenu serait identique pour les deux modélisations (a) et (b) avec le même nombre d'état ouvert simulés sur la période. Ce score ne permettrait donc pas d'expliquer que, pour la modélisation (a), une ouverture n'a pas été détectée alors que pour la modélisation (b), les deux ouvertures ont été détectées mais sous-estimées (en durée). Aussi, pour cette étude, quatre métriques d'évaluation, orientées autour des besoins liés à la détection d'ouvertures de fenêtres pour le domaine du bâtiment, sont proposées : le nombre de vraies ouvertures, le temps de vraies ouvertures, le temps de fausses ouvertures et le temps d'ouvertures sur ou sous-estimées.

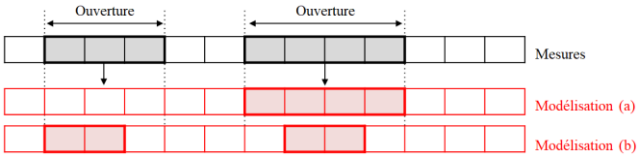


Figure 1: Exemple de simulations d'ouvertures de fenêtres avec deux modélisations

2.2.1. Estimation du nombre de vraies ouvertures

Une vraie ouverture est une ouverture simulée (par le modèle) correspondant à une ouverture mesurée dans un intervalle de temps donné. Une vraie ouverture est donc un vrai positif ou un ensemble successif de vrais positifs pouvant inclure un ou plusieurs faux positifs. Cette métrique évalue le nombre de vraies ouvertures notées « Nb_Ouv_{vrai} » retournées par le modèle et a pour objectif d'évaluer la capacité du modèle à estimer des ouvertures sans tenir compte de leurs durées. Ainsi, comme présenté dans la Figure 2, la modélisation (a) simule une vraie ouverture composée de trois vrais positifs et d'un faux positif (cases pleines) alors que la modélisation (b) en simule une composée de deux vrais positifs et d'un faux positif, décalée d'un pas de temps de la mesure.

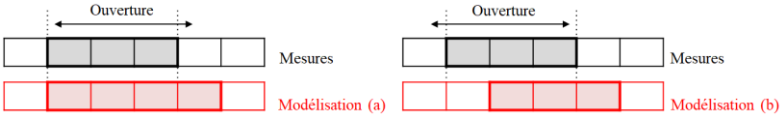


Figure 2: Exemple de deux vraies ouvertures pour une même période de mesures

2.2.2. Estimation du temps de vraies ouvertures et du temps de fausses ouvertures

Le temps de vraies ouvertures correspond à la durée totale des vraies ouvertures et est noté « $Temps_Ouv_{vrai}$ ». Le temps de fausses ouvertures correspond à la différence entre le temps total d'ouvertures prévu par le modèle et le temps total de vraies ouvertures, il est noté « $Temps_Ouv_{fausse}$ ».

2.2.3. Estimation du temps de vraies ouvertures sur ou sous-estimées

Cette métrique évalue la différence entre la durée d'une vraie ouverture et la durée mesurée de l'ouverture en question. Elle permet ainsi de déterminer si le modèle a tendance à surestimer ou sous-estimer les temps d'ouvertures qu'il détecte. Les temps totaux des sous-estimations et des surestimations sont respectivement notés « $Temps_Ouv_{vrai}$ sous-estimé » et « $Temps_Ouv_{vrai}$ surestimé ». À titre d'exemple, on constate avec la Figure 2 que l'ouverture vraie de la modélisation (a) surestime la durée de l'ouverture d'un pas de temps contrairement à la modélisation (b).

3. CAS D'ETUDE

3.1. DONNEES

Les données utilisées pour ce cas d'étude sont issues de mesures réalisées sur deux ans (juillet 2019 à février 2021) dans la chambre principale exposée nord-ouest d'un appartement de type T3 situé au centre-ville de Bordeaux. Les mesures de température d'air, humidité relative et concentration de CO_2 sont issues du même capteur d'ambiance et sont récoltées toutes les minutes. Les mesures d'ouverture des fenêtres sont réalisées au moyen d'un contacteur (état ouvert : 1, état fermé : 0) et sont récoltées toutes les heures sauf en cas de changement d'état. Sur les deux années de mesures disponibles, quelques mois, présentant peu d'anomalies, ont été retenus et séparés en trois périodes. La première, destinée à l'entrainement, s'étale sur une période de 4 mois de fin octobre 2019 à début mars 2020 (environ 12 900 points de mesure). La deuxième, destinée à la validation de l'entrainement, débute le 1er décembre 2020 et termine le 24 décembre de cette même année (environ 2 300 points de mesure). La dernière, utilisée pour tester et présenter les performances des modèles débute le 25 décembre 2020 et termine le 15 janvier 2021 (environ 2 000 points de mesure).

3.2. ANALYSE DES DONNEES

L'étude des données de la période de test permet de mettre en évidence deux périodes différentes d'utilisation de la chambre : *(i)* une période d'absence des usagers du 24 décembre 2020 au 03 janvier 2021, période pendant laquelle la température intérieure va dériver, allant jusqu'à perdre 2 °C comme présenté dans la Figure 3 ; *(ii)* ainsi qu'une période d'occupation du 04 au 15 janvier 2021 avec une remontée en température de la pièce à 22 °C.

La Figure 3 permet aussi d'illustrer que certaines ouvertures, probablement très courtes, comme l'ouverture 4 (O_4) ou probablement en soufflet comme les ouvertures 2, 13 et 14 (respectivement O_2, O_{13} et O_{14}), ont très peu d'impact sur l'ambiance intérieure avec une faible chute de température observée (0,2 °C ou moins). Une tendance similaire est observée sur les mesures d'humidité relative ou de concentration de CO_2. Ces ouvertures semblent donc, de prime abord, particulièrement délicates à traduire en période d'ouvertures mais ne correspondent qu'à 2h30 d'ouverture sur les 14h30 mesurées. Enfin, il est important de noter que la troisième ouverture (O_3) a été ajoutée manuellement car présentant un impact notable à la fois sur les mesures de température, de CO_2 et d'humidité. L'origine de son absence peut s'expliquer par un défaut des contacteurs, problème rencontré à plusieurs reprises.

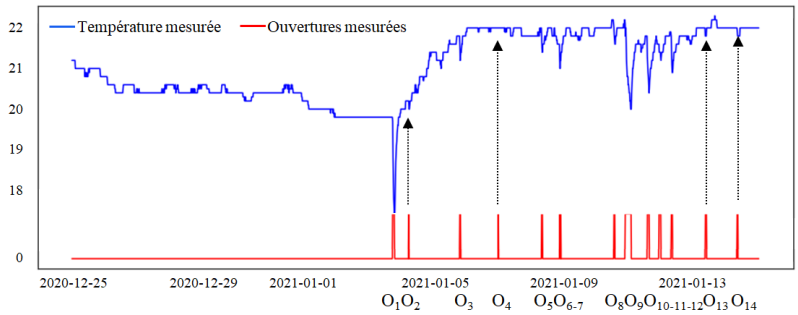


Figure 3: Données de température et d'ouverture de fenêtres sur la période de test

4. RESULTATS

4.1. CONSTRUCTION DES ENTREES DES MODELES

Afin d'optimiser la qualité de la phase d'entrainement, les anomalies ponctuelles sur les mesures de température, humidité et CO_2 ont été corrigées. Puis, les données mesurées ont été agrégées à une période de 15 minutes pour se rapprocher d'une période communément utilisée dans l'étude des consommations énergétiques d'un lieu d'habitation (entre 10 et 60 minutes). Enfin, ces données d'entrées (aussi appelées *features*) fournies au modèle LSTM sont associées à leur horodatage (jour de la semaine, heure et minute) puis découpées suivant l'approche d'une fenêtre glissante centrée de deux heures. Cette approche consiste à découper les données selon une fenêtre de deux heures qui se décale d'un pas de temps jusqu'à la fin du jeu de données.

4.2. DESCRIPTION DES DONNEES UTILISEES COMME ENTREES

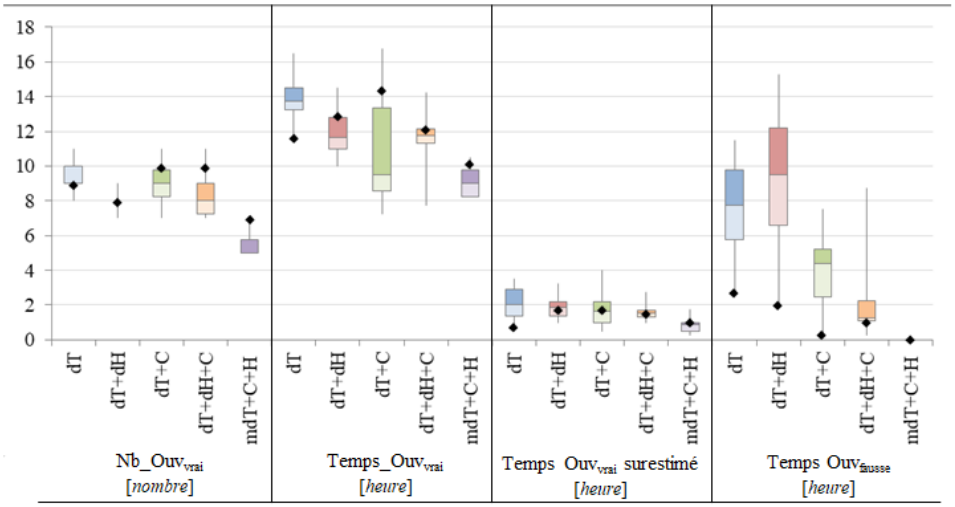
Différentes données d'entrées, ou associations de données d'entrées, ont été utilisées afin d'entrainer les modèles LSTM. Bien que toutes aient pour base les mesures de températures, humidités et de concentrations de CO_2 détaillées dans la partie 3.1, certaines transformations telles que la dérivée ou un lissage par moyenne mobile exponentielle ont été appliquées. Le but de ces transformations est d'extraire des informations différentes d'une même mesure qui, une fois traitées par les modèles, pourraient permettre d'améliorer les résultats. Ainsi, l'utilisation de la dérivée des mesures a pour objectif de permettre au modèle de mieux capter la dynamique régissant les ouvertures et fermetures de fenêtres bien souvent traduites par des baisses (ou hausses) localement importantes des valeurs mesurées. L'utilisation d'une moyenne mobile exponentielle avec une constante de lissage *alpha* relativement forte ($0{,}01 < \alpha < 0{,}25$) est une approche visant à réduire les bruits pouvant être observés sur les mesures ou leurs dérivées.

4.3. PRESENTATION DES RESULTATS

Plus d'une trentaine de combinaisons de *features* ont été testées en se basant sur quatre types de combinaisons : *(i)* les données issues des mesures de température (température, dérivée, moyenne mobile exponentielle à différents α), *(ii)* l'association des données issues des mesures de température et d'humidité, *(iii)* l'association des données issues des mesures de température et de concentration de CO_2 ainsi que l'association des trois mesures *(iv)*. La Figure 4 (a) présente sous la forme d'un diagramme en boite la répartition, par métrique d'évaluation, des meilleurs résultats obtenus pour chaque combinaison de *features* sur un ensemble de 50 modèles. Un même modèle entrainé deux fois avec les mêmes données pouvant produire des résultats différents, cette approche permet d'évaluer la

dispersion des résultats obtenus. La Figure 4 (b) répertorie quant à elle le nombre d'ouvertures prévues par les modèles pour chaque association de données présentées.

(a)



Avec: T→Température ; H→Humidité ; C→CO_2 ; d→dérivée ; m→moyenne mobile exponentielle ($\alpha = 0{,}05$)

◆ → Les résultats du meilleur modèle pour la ou les *features* renseignées. Le meilleur modèle est choisi parmi les modèles détectant le plus de vraies ouvertures tout en ayant un faible temps de fausses ouvertures.

Exemple: dT + C → dérivée de la Température associée à la concentration de CO_2 (sans transformations)

Information: nombre d'ouverture mesurées total → **14** et pas de temps des données → 15 minutes

(b)

	O_1	O_2	O_3	O_4	O_5	O_6	O_7	O_8	O_9	O_{10}	O_{11}	O_{12}	O_{13}	O_{14}
dT	100%	93%	100%	**0%**	100%	100%	**6%**	100%	100%	100%	53%	100%	**6%**	**0%**
dT+dH	100%	100%	60%	**20%**	**0%**	100%	**0%**	60%	100%	100%	50%	100%	**10%**	**0%**
dT+C	100%	70%	90%	**0%**	100%	**10%**	**0%**	100%	50%	100%	70%	100%	100%	**0%**
dT+dH+C	100%	100%	60%	**0%**	60%	**10%**	**10%**	100%	90%	100%	80%	100%	**30%**	**0%**
mdT + C + H	100%	**0%**	**30%**	**0%**	**10%**	100%	**0%**	**0%**	100%	100%	**0%**	100%	**0%**	**0%**

Figure 4: (a) Résultats des modèles LSTM pour différentes associations de données, (b) Nombre de correspondance entre ouvertures modélisées et simulées

4.4. ANALYSE DES RESULTATS

D'après les résultats présentés dans la Figure 4 (a), on constate, par son absence, que l'utilisation de la température sans transformations semble donner des résultats en deçà de sa dérivée (en bleu notée « *dT* »). En effet, les modèles utilisant la température comme donnée d'entrée présentent un taux très élevé de fausses ouvertures avec plus de 100 heures quelles que soient les combinaisons réalisées. Ces résultats peuvent s'expliquer par le fait que les modèles semblent associer une température intérieure basse comme élément principal traduisant une ouverture. Ce phénomène provoque un nombre important de fausses ouvertures pendant la période sans occupation de l'appartement où la température intérieure est plus basse de 2 °C. L'utilisation de la dérivée de la température, traduisant plutôt sa dynamique, semble permettre au modèle de mieux approcher les ouvertures mesurées en diminuant drastiquement le temps de fausses ouvertures des modélisations avec une médiane de 7h30. On constate cependant qu'il délimite mal les temps d'ouvertures, les surestimant d'environ 2 heures au total. Ainsi, le meilleur modèle, représenté par une pastille noire, n'utilisant que la dérivée de la température, permet d'obtenir des résultats plutôt prometteurs avec 9 ouvertures détectées sur 14, 11h45 de temps d'ouverture, une surestimation totale des ouvertures de 45 minutes pour 2h45 de fausses ouvertures.

L'association de la dérivée de la température et de la dérivée de l'humidité (en rouge notée « *dT* + *dH* »), est la meilleure combinaison pour les données associées aux mesures de température et d'humidité. On constate cependant une détérioration générale légère des résultats sur la détection des ouvertures et le temps des fausses ouvertures, tous deux inférieurs aux modèles n'utilisant que la dérivée de la température et ceci, malgré une diminution du temps de surestimation des ouvertures. Ces résultats peuvent s'expliquer par la redondance possible des informations apportées par l'humidité à la température. Le modèle pourtant déjà capable de simuler certaines ouvertures à l'aide de la température, pourrait se voir conforté dans ses bonnes décisions mais aussi induit en erreur par de nouvelles informations issues des mesures d'humidité. D'une manière générale, on constate que l'ajout de l'humidité en entrée a tendance à détériorer les résultats des modélisations.

Contrairement à l'humidité et bien que très fluctuante avec des variations qui ne sont pas uniquement liées à l'ouverture et à la fermeture des fenêtres (ex. : présence ou non d'occupants, positionnement dans la pièce, mouvements d'air naturels…), la combinaison des mesures de concentration de CO_2 aux mesures de température semble donner de meilleurs résultats que ceux observés précédemment. La meilleure combinaison est représentée par l'association de la dérivée de la température à la mesure de CO_2 sans modifications (en vert notée « *dT* + *C* »). On constate en effet une diminution notable du temps de fausses ouvertures avec une médiane située vers 4 heures ainsi qu'une très légère amélioration sur le temps d'ouverture surestimé. Bien que non présents sur la figure, il est à noter que l'utilisation de la dérivée de la concentration de CO_2 (avec ou sans lissage par moyenne mobile exponentielle) détériore sensiblement les résultats avec en moyenne 15 heures de fausse ouvertures. Ainsi, le meilleur modèle de cette combinaison permet d'obtenir de très bons résultats avec 10 ouvertures détectées pour 14h30 de temps d'ouvertures, une surestimation totale des ouvertures de 1h45 pour 15 minutes de fausses ouvertures. Contrairement à l'humidité, la concentration de CO_2 dans la pièce n'est pas une mesure redondante, mais complémentaire.

En combinant les trois mesures de température, humidité et concentration de CO_2, on constate que deux associations sortent du lot. La première est composée des meilleures associations observées sur les précédents essais à savoir la dérivée de la température, la dérivée de l'humidité et la concentration de CO_2 sans modifications (en orange notée « *dT* + *dH* + *C* »). Il est probable que l'ajout de la dérivée de l'humidité à la combinaison vue précédemment conforte le modèle dans ses décisions sans toutefois trop l'induire en erreur contrairement à ce qui a été observé pour l'association température et humidité. Le meilleur modèle de cette combinaison permet d'obtenir de bons résultats avec 10 ouvertures détectées pour 12h30 de temps d'ouvertures modélisés et une surestimation totale des ouvertures de 1h30 pour 1 heure de fausses ouvertures. La deuxième combinaison est constitué d'une dérivée de la température lissée par moyenne mobile exponentielle ($\alpha = 0{,}05$) associée aux mesures d'humidité et de concentration de CO_2 sans modifications (en violet notée « *mdT* + *C* + *H* »). Contrairement aux essais précédents, le modèle obtenu ne détecte que 5 à 6 ouvertures mais avec peu de dispersion tout en ne surestimant que peu les ouvertures. Ce résultat, qui n'a pas été observé sur d'autres associations reprenant la moyenne mobile de la dérivée, pourrait s'expliquer par la réduction du bruit de la dérivée induit par le lissage. Il donnerait donc plus d'importance à la dérivée lissée pour détecter les ouvertures et utiliserait les autres mesures afin d'affiner les détections. En utilisant cette association, il est possible de modéliser les ouvertures les plus importantes avec une relative précision, permettant d'estimer, pour le meilleur modèle, 10h15 d'ouvertures, avec 1 heure de surestimation, sans fausses ouvertures, sur les 14h30 totales.

Enfin, on remarque avec la Figure 4 (b) que les modèles ont, dans l'ensemble, du mal à détecter les ouvertures 4, 13 et 14 ayant un impact faible sur les conditions intérieures, comme détaillé dans la partie 3.2, bien que l'ouverture 2 semble régulièrement détectée. On constate cependant que les ouvertures ayant un impact important sur l'environnement ambiant intérieur sont souvent détectées.

5. CONCLUSION

Cette étude présente une comparaison des performances de modèles de réseau de neurones LSTM à prévoir des ouvertures de fenêtres selon plusieurs données issues de mesures (température et humidité relative de l'air, concentration de CO_2, *etc*.) dans le domaine de l'énergétique du bâtiment. Les résultats ont montré que le choix des données d'entrées était primordial pour obtenir des résultats satisfaisants mais aussi qu'il n'était pas toujours souhaitable d'ajouter de l'information en entrées des modèles et qu'un tri préalable était nécessaire au risque de détériorer leurs performances. Ces travaux ont aussi mis en évidence qu'une transformation des données en amont (dérivée, moyenne mobile exponentielle) pouvait avoir un impact très positif sur la qualité des prévisions en passant de résultats inexploitables (température seule) à des résultats satisfaisants (dérivée de la température). Bien que le nombre d'ouvertures modélisées vraies puisse sembler faible de prime abord (entre 7 et 10 prédites sur 14 mesurées au total), ce résultat doit être nuancé. En effet, plusieurs ouvertures n'impactent pas ou très peu l'environnement intérieur, n'offrant ainsi pas d'information claire aux modèles pour les détecter. Leur non-détection n'est alors pas forcément problématique dans des cadres d'application comme l'estimation des pertes thermiques d'un bâtiment liées aux ouvertures de fenêtres par exemple. Enfin, les résultats obtenus pourraient être améliorés en utilisant d'autres données mesurées telle que l'occupation de la pièce, la température, l'humidité et la concentration de CO_2 extérieure ou encore en couplant les réseaux LSTM à d'autres réseaux de neurones tels que les auto-encodeurs (*Autoencoder : AE*) ou des réseaux de neurones convolutifs (*Convolutional Neural Network : CNN*).

Remerciements

Ces travaux ont été menés dans le cadre de l'équipe de recherche commune GP2E (Garantie des Performances Énergétique et Environnementale des bâtiments) entre l'Université de Bordeaux et l'ITE Nobatek/INEF4, soutenue par la région Nouvelle-Aquitaine.

6. BIBLIOGRAPHIE

[1] Carlucci *et al*. 2020. « Modeling Occupant Behavior in Buildings ». Building and Environment 174.

[2] Enguerrand de Rautlin de la Roy *et al*. « Méthodologie pour la comparaison des performances simulées et mesurées de maisons à énergie positive ». Conférence Francophone de l'International Building Performance Simulation Association IBPSA, 2020, France.

[3] D'Oca, Simona, et Tianzhen Hong. 2014. « A Data-Mining Approach to Discover Patterns of Window Opening and Closing Behavior in Offices ». Building and Environment 82

[4] Raghavendra C., and Sanjay C. 2019. « Deep learning for anomaly detection: A survey ». arXiv preprint arXiv:1901.03407(2019).

[5] Chandola, V., Banerjee, A., and Kumar, V. 2009. « Anomaly detection: A survey ». ACM Comput. Surv. 41, 3, Article 15 (July 2009), 58 pages. DOI=10.1145/1541880.1541882

[6] Xilei, Dai & Liu, Junjie & Zhang, Xin. (2020). « A review of studies applying machine learning models to predict occupancy and window-opening behaviours in smart buildings. Energy and Buildings ». 223. 110159. 10.1016/j.enbuild.2020.110159.