

UNIVERSITÉ DU QUÉBEC

MÉMOIRE PRÉSENTÉ À
L'UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN MATHÉMATIQUES ET INFORMATIQUE APPLIQUÉES

PAR
MOUHAMADOU MOUNTAKHA NDAO

APPROCHES STATISTIQUES ET NEURONALES POUR LA MODÉLISATION
PRÉDICTIONNELLE D'UNE MACHINE À COMBUSTION À BIOMASSE.

Juin 2024

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

L'auteur de ce mémoire, de cette thèse ou de cet essai a autorisé l'Université du Québec à Trois-Rivières à diffuser, à des fins non lucratives, une copie de son mémoire, de sa thèse ou de son essai.

Cette diffusion n'entraîne pas une renonciation de la part de l'auteur à ses droits de propriété intellectuelle, incluant le droit d'auteur, sur ce mémoire, cette thèse ou cet essai. Notamment, la reproduction ou la publication de la totalité ou d'une partie importante de ce mémoire, de cette thèse et de son essai requiert son autorisation.

Résumé

En raison des nombreux phénomènes thermiques, physiques et chimiques qui sont liés, les machines à combustion à biomasse peuvent être extrêmement difficiles à utiliser. Dès lors, trouver un modèle mathématique précis pour représenter la majorité de ces machines reste encore un travail en cours. Cette étude porte sur la conception d'un modèle de prédiction des intrants d'une machine à biomasse. Le modèle proposé est un modèle dynamique inverse. Il est basé sur des données obtenues expérimentalement à partir d'une machine à biomasse réelle. Compte tenu de la complexité de la machine, il est essentiel de comprendre les signaux qui doivent être réglés à l'entrée de la machine pour qu'elle fonctionne de manière optimale. Quatre modèles de réseaux de neurones ont été évalués : le perceptron multicouche (PMC), le réseau neuronal récurrent (RNN), la mémoire à long et court terme (LSTM) et les unités récurrentes fermées (GRU) qui tiennent compte de la non-linéarité. Le modèle inverse basé sur l'approche GRU a surpassé les autres modèles testés, avec une perte de 11,97% et une précision de 79,89%. Ces deux indicateurs de performance ont été calculés sur la base de tests de nos données (20% de notre base initiale). Les vitesses primaire, secondaire, tertiaire et celle du ventilateur ont été prédites. La différence entre les prévisions et les valeurs réelles est inférieure à 6 %. Il s'agit d'un pas en avant pour mieux comprendre les machines à combustion à biomasse et pouvoir configurer les signaux d'entrée appropriés pour que la machine fonctionne dans le mode souhaité.

Abstract

Because of many thermal, physical, and chemical interrelated phenomena, combustion biomass machines can be extremely difficult to operate. Finding a precise mathematical model to represent the majority of these machines is still a work in progress. This study covers the establishment of a model for predicting the inputs of a biomass machine. The model proposed is an inverse model. It is based on data experimentally obtained from a real-world biomass machine. Given the machine's complexity, it is critical to understand the signals that must be set at the machine's input in order for it to perform optimally. Four ANN models were evaluated : the multilayer perceptron (MLP), the recurrent neural network (RNN), the long short term memory (LSTM), and the gated recurrent unit (GRU). These models take into account the data's nonlinearity. The inverse model based on the GRU approach outperformed the other ANNs tested, with a Loss of 11.97% and an Accuracy of 79.89%. These two performance indicators were calculated on the test basis of our data (20% of data). The primary, secondary, tertiary and fan speeds were predicted using the GRU model. The difference between the predictions and the actual values for each speed is less than 6%. This is a step forward in having a better understanding of biomass combustion machines and being able to configure appropriate input signals for the machine to operate in the desired mode.

Avant-propos

Ce mémoire de recherche se concentre principalement sur les réseaux de neurones artificiels et leur application en science des données, en particulier dans le domaine de la biomasse. Dans le cadre de nos expériences, nous avons utilisé quatre types de réseaux de neurones : les perceptrons multicouches (PMC), les réseaux à mémoire à long et court terme (LSTM), les réseaux de neurones récurrents (RNN) et les unités récurrentes fermées (GRU).

Les méthodes neuronales ont émergé dès les années 50, mais elles n'ont véritablement connu une expansion significative qu'à partir des années 80, notamment avec la découverte de l'algorithme de rétropropagation. Dans cette étude, nous commencerons par décrire l'approche méthodologique impliquant ces algorithmes neuronaux. Par la suite, nous procéderons à une revue de la littérature portant sur la biomasse, puis nous aborderons leur application dans le domaine de la biomasse, en particulier dans la prédiction des paramètres d'une machine destinée à ce secteur.

Remerciements

Je tiens tout d'abord à exprimer ma sincère gratitude envers ma directrice de recherche, Mme Nadia Ghazzali, pour le soutien inestimable et la confiance qu'elle m'a accordés tout au long de mon parcours à l'UQTR. Son esprit critique et sa rigueur ont grandement contribué à l'amélioration de ce document. Je lui serai éternellement reconnaissant pour le partage de ses connaissances et de son expérience tout au long de ma maîtrise, ainsi que pour son soutien financier. J'adresse également mes remerciements à ma codirectrice, Mme Nadjet ZIOUI, pour son soutien pédagogique, financier et son expertise en matière de recherche, dont j'ai pu bénéficier tout au long de ce parcours.

Je tiens à exprimer ma reconnaissance envers l'Institut des Sciences Mathématiques du Québec (ISM), le laboratoire d'Intelligence Artificielle Appliquée (LI2A) de l'UQTR, le conseil de recherches en sciences naturelles et en génie (CRSNG), le Mitacs, l'entreprise Combustion Expert Énergie pour leur soutien financier précieux. Je remercie également l'ensemble du personnel du département de mathématiques et d'informatique de l'UQTR, pour leurs services rendus, leur réactivité et leur orientation qui ont grandement facilité mon intégration.

Mes remerciements s'étendent également à mes parents, ma superbe femme Adjji, mes frères et mes sœurs, qui m'ont constamment encouragé, accompagné et motivé tout au long de mon parcours scolaire.

*Je dédie ce modeste travail à ma mère Ndèye Arame et
à mon défunt père Momath. À mes frères et à mes
sœurs. Vous êtes une source d'inspiration et de
motivation perpétuelle.*

*À ma Femme Adjì Fatou Diawara ! Au fil du temps, tu
es devenue la personne la plus essentielle dans ma vie.*

*Je veux que tu comprennes à quel point je t'aime
profondément, combien tu comptes pour moi, et que je
trouve mon bonheur à tes côtés. Je prie DIEU de
préservèr éternellement cette félicité d'être à tes côtés,
de demeurer parmi tes priorités. Ce travail est dédié à
toi et à toute notre famille. Merci du fond du cœur.*

*Sénèque "Tirons notre courage de
notre désespoir même."*

Acronymes

RNN Réseau de neurones récurrent

GRU Unités récurrentes fermées

LSTM Mémoire à long et court terme

PMC Perceptron multicouches

ACP Analyse en composantes principales

PID Régulateur Proportionnel Intégral Dérivé

Table des matières

Résumé	ii
Abstract	iii
Avant-propos	iv
Remerciements	v
Acronymes	vii
Table des matières	viii
Table des figures	xi
Liste des tableaux	xii
Introduction	1
1 Approche méthodologique	4
1.1 Les méthodes statistiques	4
1.1.1 Le coefficient de corrélation linéaire	4
1.1.2 Le test de Kolmogorov-Smirnov (K-S)	5
1.1.3 L'Analyse en Composantes Principales (ACP)	8
1.2 Les méthodes neuronales	11
1.2.1 Architecture d'une couche neuronale	11
1.2.2 Un réseau de neurones (plusieurs couches)	13
1.3 Les modèles de réseaux de neurones étudiés	14

1.3.1	Le perceptron multicouches (PMC)	14
1.3.2	Les réseaux de neurones récurrents (RNN)	17
1.3.3	Les réseaux à mémoire à long et court terme (LSTM)	18
1.3.4	Les unités récurrentes fermées (GRU)	21
2	Application au cas de la machine à combustion à biomasse	23
2.1	Revue de littérature en biomasse et problématique de la recherche . . .	23
2.2	Description de la machine à biomasse et des données provenant de cette machine	26
2.2.1	La machine à combustion à biomasse	26
2.2.2	Statistique descriptive sur la base de données	29
2.3	Présentation des analyses exploratoires	33
2.4	Présentation des résultats du modèle inverse	36
2.5	Discussion des résultats	43
	Conclusion	45
	Bibliographie	47

Table des figures

1.1	<i>Exemple de données normales : Courbe de densité empirique (F) vs Loi normale (G) [10]</i>	7
1.2	<i>Architecture d'un réseau neuronal</i>	12
1.3	<i>Fonction de transfert : (a) du neurone «seuil»; (b) du neurone «linéaire», (c) du neurone «tangente hyperbolique», (d) du neurone «RELU» et (e) du neurone «Sigmoidé»</i>	13
1.4	<i>Architecture d'un réseau de S neurones</i>	14
1.5	<i>Représentation matricielle d'un réseau PMC à 2 couches cachées avec 4 sorties.</i>	15
1.6	<i>Réseau neuronal récurrent RNN de 2 couches cachées avec 4 sorties [6].</i>	17
1.7	<i>Représentation de l'architecture d'un LSTM. Un seul bloc de mémoire est représenté pour plus de clarté [40].</i>	19
1.8	<i>Représentation de la structure du réseau GRU (gated recurrent unit) [52].</i>	21
2.1	<i>Schéma de la machine à combustion à biomasse.</i>	27
2.2	<i>Système d'entrée/sortie de la machine à biomasse.</i>	28
2.3	<i>Système d'entrée/sortie de la machine à combustion à biomasse avec modèle inverse.</i>	29
2.4	<i>Courbes de densité des neuf variables.</i>	32
2.5	<i>Diagrammes en boîte des neuf variables.</i>	33
2.6	<i>Graphique des corrélations deux par deux des variables.</i>	34
2.7	<i>Représentation du cercle de corrélation des variables.</i>	36
2.8	<i>Validation croisée [5]</i>	37

2.9	<i>Représentation de la perte et de la précision à l'aide de la méthode GRU.</i>	40
2.10	<i>Représentations a), b), c) et d) de la différence entre les prédictions et les valeurs réelles à l'aide de la méthode GRU.</i>	42
11	Partie 1 du workflow du traitement des données	54
12	Partie 2 du workflow du traitement des données	55
13	Workflow de la modélisation	58

Liste des tableaux

- 2.1 *Extrait de la base de données expérimentales* 30
- 2.2 *Statistique descriptive de chaque variable à l'étude* 31
- 2.3 *Perte et précision des quatre modèles NN.* 39

Introduction

La biomasse constitue la matière qui compose les organismes vivants ou qui en est issue [8]. Il s'agit de toute matière organique d'origine végétale ou provenant d'organismes vivants. C'est une source d'énergie propre et renouvelable. Au 21^e siècle, l'utilisation d'énergies renouvelables est devenue incontournable pour toutes les nations. Conformément à une étude citée dans [4], les résultats expérimentaux montrent qu'il existe une relation bidirectionnelle entre la consommation d'énergie renouvelable et la croissance économique à la fois à court et à long terme. Il est donc évident que la puissance énergétique joue un rôle central dans la quête de la richesse. En outre, [7] met en exergue cette urgence de se diriger vers les énergies renouvelables. Les États allouent des budgets considérables pour garantir leur approvisionnement énergétique et la biomasse constitue une part significative de cette richesse énergétique.

C'est dans ce contexte que l'industrie a déployé des équipements pour la production d'énergie renouvelable, comme en témoigne l'étude [28]. Cela s'applique, en particulier, aux systèmes de chauffage à biomasse qui utilisent la biomasse forestière résiduelle. Ces systèmes requièrent un entretien méticuleux et l'optimisation de leurs paramètres pour garantir une production efficace.

L'étude menée par l'auteur dans [37] a tenté de dresser un bilan de la valorisation énergétique de la biomasse par combustion. Cette recherche a permis d'identifier des caractéristiques influentes telles que la teneur en azote, en cendres et en potassium, qui peuvent varier en fonction du mode de fonctionnement de la machine. Les experts

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

L'auteur de ce mémoire, de cette thèse ou de cet essai a autorisé l'Université du Québec à Trois-Rivières à diffuser, à des fins non lucratives, une copie de son mémoire, de sa thèse ou de son essai.

Cette diffusion n'entraîne pas une renonciation de la part de l'auteur à ses droits de propriété intellectuelle, incluant le droit d'auteur, sur ce mémoire, cette thèse ou cet essai. Notamment, la reproduction ou la publication de la totalité ou d'une partie importante de ce mémoire, de cette thèse et de son essai requiert son autorisation.

travaillant sur ces machines sont souvent confrontés à des problèmes tels que des arrêts inattendus, des problèmes de surchauffe ou un déficit de production de vapeur, parmi d'autres.

Dans nombreuses études antérieures (par exemple [42]) sur l'optimisation des machines à biomasse, les chercheurs se sont principalement concentrés sur des approches directes, examinant les paramètres de conception et les configurations de la machine pour améliorer son efficacité. Cependant, ces méthodes peuvent être limitées par la nécessité d'une compréhension préalable des relations entre les variables et la complexité des modèles requis. Dans cette étude, nous proposons une approche innovante basée sur la méthode inverse, visant à prédire les entrées optimales nécessaires pour maximiser les performances des machines à biomasse. Cette approche offre une perspective alternative en identifiant les combinaisons d'entrées qui conduisent à des résultats optimaux. En utilisant les réseaux de neurones, notre objectif est de proposer des solutions plus efficaces et innovantes pour l'optimisation des machines à biomasse.

Ce mémoire, suggère d'utiliser quatre algorithmes alternatifs basés sur des réseaux de neurones artificiels pour représenter le comportement dynamique d'une unité à combustion à biomasse. Cette recherche se concentre sur une machine à combustion à biomasse, visant principalement à développer un modèle prédictif des paramètres d'entrée et de sortie en utilisant des réseaux de neurones. Les algorithmes des perceptrons multicouches, des réseaux à mémoire à long et court terme, des réseaux de neurones récurrents et des unités récurrentes fermées ont été expérimentés. Ces algorithmes utilisent des données expérimentales acquises sur le terrain. Les quatre algorithmes ont été évalués en termes de perte (Loss) et de précision (Accuracy) des données après une analyse descriptive et un pré-traitement complet des données. Pour prédire le comportement de la machine, l'algorithme avec les meilleures performances est choisi. En fait, l'application des réseaux de neurones dans certains travaux sur la biomasse, comme [17], a produit des résultats satisfaisants en terme de pourcentage d'erreur absolue moyenne et un coefficient de détermination. Ces travaux visent à four-

nir un outil sur mesure aux scientifiques et ingénieurs impliqués dans la modélisation et l'optimisation des machines à combustion à biomasse.

Le document est organisé en deux principaux chapitres. Le premier chapitre traite de l'approche méthodologique et explore les différentes méthodes utilisées pour l'analyse des données, en se concentrant sur les méthodes statistiques et les méthodes neuronales. Le second chapitre, intitulé application au cas de la machine à combustion à biomasse, met l'accent sur l'application pratique de ces méthodes dans le contexte spécifique de la machine à combustion à biomasse.

Chapitre 1

Approche méthodologique

Dans cette partie, nous présenterons les méthodes statistiques standards utilisées pour l'analyse des données provenant d'une machine à biomasse. Par la suite, nous ferons un survol des méthodes neuronales qui seront considérées pour la commande et l'optimisation d'une machine à combustion à biomasse. Finalement, nous passerons à la sélection des modèles candidats pour faire notre étude.

1.1 Les méthodes statistiques

1.1.1 Le coefficient de corrélation linéaire

Pour toute étude statistique, il est essentiel de comprendre la relation entre les variables. C'est dans cette optique que nous avons utilisé le coefficient de corrélation. Il consiste à diviser la covariance ($cov(X, Y)$) par le produit des écarts-types (σ_X et σ_Y) des variables [34].

Soit X et Y deux variables quantitatives. La corrélation entre X et Y est r_{XY} défini

par :

$$r_{XY} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \quad (1.1),$$

avec :

- $\text{cov}(X, Y)$: la covariance entre X et Y ;
- σ_X et σ_Y les écarts-types respectives de X et Y .

Les propriétés du coefficient de corrélation sont :

- $-1 \leq r_{XY} \leq +1$
- Il a le même signe que $\text{cov}(X, Y)$.
- X et Y sont indépendantes $\implies r = 0$. La réciproque n'est pas toujours vraie. En fait, si le couple de variables (X, Y) suit une loi normale bivariée, et uniquement dans ce cas, l'équivalence $r = 0 \iff X$ et Y sont indépendants est vérifiée. Dans cette situation, le coefficient de corrélation caractérise parfaitement la relation entre X et Y . Cependant, dans d'autres cas, le coefficient de corrélation représente l'une des mesures de l'intensité de la corrélation parmi d'autres [34].
- $r_{XX} = 1$ est la corrélation de X avec elle même.
- Le coefficient de corrélation (r_{XY}) quantifie l'intensité de la relation linéaire entre X et Y . Il est important de noter que r_{XY} peut être égal à zéro ($r_{XY} = 0$) tout en présence d'une relation fonctionnelle entre les variables. Cela se produit lorsque leur relation est non linéaire [34].
- $r_{XY} = +1$ signifie que la liaison entre X et Y est linéaire, positive et parfaite. Autrement dit, lorsque la valeur de X augmente, la valeur de Y augmente de manière proportionnelle.
- $r_{XY} = -1$ la liaison est linéaire et négative.

1.1.2 Le test de Kolmogorov-Smirnov (K-S)

Dans une telle étape, nous examinons la distribution de nos données. Nous avons travaillé avec un ensemble de données expérimentales de grande taille. Pour le cas

particulier de l'hypothèse de données suivant une distribution normale, le test de Kolmogorov-Smirnov est le choix le plus courant [25]. Ce test compare la distribution de nos données à celle d'une distribution normale. Nous traitons le problème comme étant non paramétrique avec deux échantillons indépendants. Plus précisément, nous supposons que nous sommes confrontés au scénario suivant [1] :

- Les variables X_1, \dots, X_m sont i.i.d avec comme fonction de répartition F_m .
- Les variables Y_1, \dots, Y_n sont i.i.d avec comme fonction de répartition G_n .
- Les variables X_1, \dots, X_m sont indépendantes des variables Y_1, \dots, Y_n .
- La loi de probabilité F est inconnue et celle de G connue (la loi normale dans notre cas). Les fonctions de densités sont f et g .

La statistique de test du test K-S, notée D , est définie comme la distance maximale entre les fonctions de distribution cumulative (FDC) des deux distributions à comparer. Pour deux distributions empiriques $F_m(x)$ et $G_n(x)$, la statistique de test est donnée par :

$$D_{n,m} = \sup_x |F_m(x) - G_n(x)|,$$

où \sup_x est le supremum sur toutes les valeurs de x .

Du point de vue statistique, la p -value associée au test K-S est la probabilité de trouver une valeur de $D_{n,m}$ au moins aussi extrême que celle observée, sous l'hypothèse nulle (H_0). L'hypothèse H_0 stipule que les deux distributions sont identiques.

Pour calculer la p -value, nous utilisons la distribution de $D_{n,m}$ sous l'hypothèse H_0 . En pratique, cette distribution est obtenue via des tables de valeurs critiques ou des méthodes de simulation.

Si la p -value est inférieure à un seuil de signification α (par exemple, 0.05), nous rejetons l'hypothèse H_0 et concluons que les deux distributions sont significativement différentes. Si la p -value est supérieure à α , nous ne pouvons pas rejeter l'hypothèse

H_0 .

Du point de vue visuel, on se base sur les courbes de distribution empirique des deux lois de probabilité. La problématique se décline comme suit :

$$H_0 : F = G$$

contre

$$H_1 : F \neq G$$

On rejette H_0 si les histogrammes de F et G sont trop différents l'un de l'autre. Les deux courbes de la figure 1.1 est utile pour comparer visuellement une distribution de données (courbe empirique F) à une distribution normale théorique G . On peut voir si les données empiriques suivent ou s'écartent de la distribution normale attendue en observant la différence entre les deux courbes.

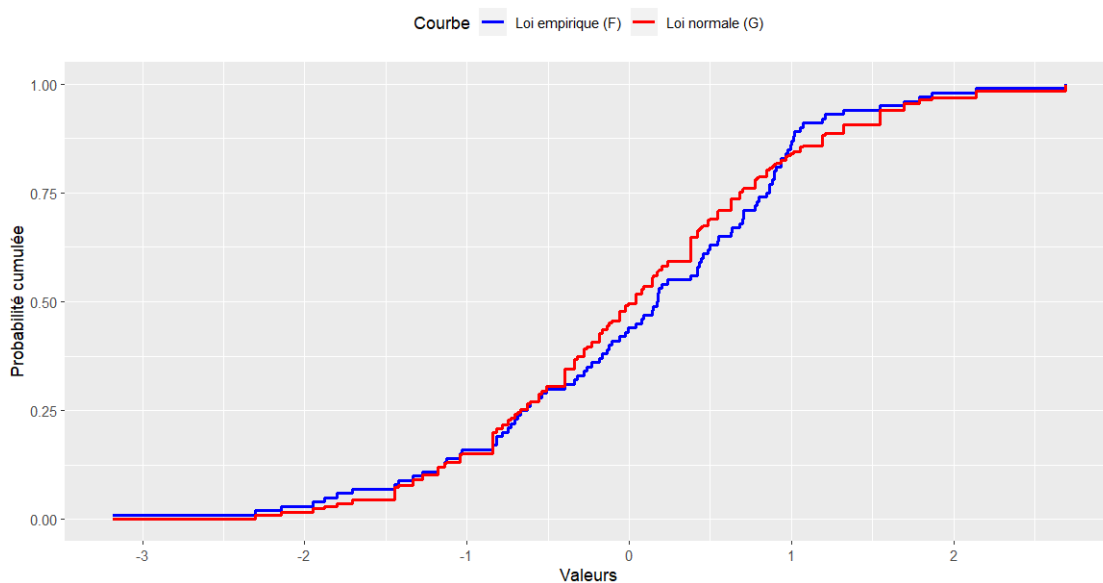


FIGURE 1.1 – Exemple de données normales : Courbe de densité empirique (F) vs Loi normale (G) [10]

1.1.3 L'Analyse en Composantes Principales (ACP)

Nous avons aussi utilisé l'analyse en composantes principales (ACP). Cette dernière fait partie des méthodes d'analyses factorielles. Ce sont des méthodes qui ont pour but d'analyser et de faire figurer géométriquement dans un espace euclidien de faible dimension, généralement deux axes factorielles nommées Dim 1 et Dim 2 par les logiciels utilisés. Les informations les plus diverses sont consignées dans un tableau numérique. Ces informations sont parfois difficilement synthétisées compte tenu de la taille et de la nature des données du tableau.

L'ACP permet d'obtenir une représentation des individus en fonction de leur proximité et une représentation des variables en fonction de leur corrélation. L'ACP s'applique généralement aux données quantitatives et cherche, par ses représentations, à :

- repérer les individus qui se ressemblent vis-à-vis des variables ;
- relever les différences entre individus et mettre en évidence ceux dont le comportement est atypique par rapport à l'ensemble des variables ;
- savoir si l'information brute contenue dans le tableau des données ne pourrait pas être obtenue à partir d'un nombre plus restreint de variables ; et
- décrire simultanément les liaisons multiples entre les variables.

Les données à analyser se présentent sous la forme d'un tableau Individus par variables. Les individus sont en ligne et sont repérés par l'indice i avec $i = 1, \dots, N$, sur chacun de ces individus nous observons les valeurs de K variables disposées en colonne et repérées par l'indice j avec $j = 1, \dots, K$.

On notera par x_{ij} la valeur de la variable X_j pour l'individu i . Le tableau se présente de la façon suivante :

$$X = \begin{matrix} & X_1 & \dots & X_j & \dots & X_K \\ \begin{matrix} 1 \\ \vdots \\ i \\ \vdots \\ N \end{matrix} & \begin{pmatrix} x_{11} & \dots & x_{1j} & \dots & x_{1K} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i1} & \dots & x_{ij} & \dots & x_{iK} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{N1} & \dots & x_{Nj} & \dots & x_{NK} \end{pmatrix} \end{matrix}$$

Ainsi nous pouvons remarquer que :

- pour chaque individu, le tableau donne une suite hétérogène de variables X_j ($j = 1, \dots, K$) de telle sorte que $\sum_{j=1}^K x_{ij}$ n'ait aucun sens. En effet, les unités de mesure des X_j sont souvent différentes (à titre d'exemple, on aura : poids+taille+..... qui n'a aucun sens) ;
- pour chaque variable j nous pouvons calculer la somme $\sum_{i=1}^N x_{ij}$ qui a un sens puisqu'elle représente à un facteur près la moyenne ;

Pour effectuer une ACP, il est courant de centrer et de réduire les données (ACP normée), ce qui permet de résoudre le problème des unités et de garantir que les variables sont comparables. De plus, l'ACP calcule les valeurs propres (λ_j avec $j = 1, \dots, K$) de la matrice de covariance du tableau des données X , qui représentent la quantité de variance expliquée par chaque composante principale. Les composantes principales dans une ACP sont de nouvelles variables synthétiques qui capturent l'essentiel de l'information contenue dans les données initiales. Elles sont obtenues en combinant linéairement les variables d'origine de telle sorte que la première composante principale capture le maximum de variance dans les données, la deuxième composante principale capture le maximum de variance restante tout en étant non corrélée avec la première, et ainsi de suite. En d'autres termes, les composantes principales permettent de réduire la dimensionnalité des données. Les individus (ou observations) sont projetés sur ces composantes principales pour obtenir leurs coordonnées dans l'espace des composantes \mathbb{R}^K (nuage des individus). Les variables sont également projetées sur les composantes principales pour comprendre leur contribution à chaque axe : c'est le nuage des variables dans l'espace \mathbb{R}^N . En résumé, l'ACP permet de réduire

la dimensionnalité des données tout en préservant au mieux l'information importante contenue dans celles-ci, ce qui facilite l'analyse et l'interprétation des relations entre les individus et les variables [22].

L'interprétation des résultats de l'ACP repose principalement sur les valeurs propres engendrées par les vecteurs propres. Bien que l'ACP soit une méthode d'analyse factorielle, elle utilise des représentations graphiques pour faire recours à l'examen graphique afin de diminuer le risque d'erreur. Ainsi des aides à l'interprétation comme **le cercle de corrélation, choix du nombre d'axes factoriels, la mise en élément supplémentaire**, appuient le diagnostic graphique pour les deux nuages.

Choix du nombre d'axes factoriels : Après avoir effectué une ACP, on ne saurait étudier tous les axes déterminés. On se demande combien d'axes factoriels faut-il analyser pour reconstituer les données. En pratique, il existe plusieurs critères qui guident le choix du nombre d'axes.

- **Critère de Joliffe** : Il s'agit de définir une proportion fixée de la variance expliquée et de s'arrêter dès que la proportion globale atteint cette valeur fixée (par exemple à 90%).
- **Critère du taux d'inertie** : Le taux d'inertie expliqué par les q premiers axes factoriels est défini par :

$$r_q = \frac{\sum_{j=1}^q \lambda_j}{\sum_{j=1}^K \lambda_j} \quad (1.2).$$

La valeur de r_q donne une indication sur la part d'information restituée par les q premiers axes. La valeur de q est choisie de sorte que r_q soit supérieure à une valeur seuil fixée à priori par les praticiens, en générale au moins 60%.

- **Critère de Kaiser** : Il consiste à ne retenir que les valeurs propres $\lambda_1, \dots, \lambda_K$

supérieures à la moyenne.

$$\frac{\lambda_1 + \lambda_2 + \cdots + \lambda_K}{K} = 1 \quad (1.3).$$

En outre, pour l'interprétation, nous adoptons donc la démarche suivante :

- On décide du nombre d'axes que l'on désire conserver pour le sous espace de projection.
- On interprète le nuage des individus en terme de ressemblance.
- On interprète en termes de corrélation les positions des variables sur le plan factoriel. On précise éventuellement, par exemple s'il y a des groupes de variables, l'origine de leurs corrélations en examinant le représentant factoriel correspondant des individus ;

1.2 Les méthodes neuronales

Les réseaux de neurones sont de plus en plus utilisés pour résoudre des problèmes liés aux séries temporelles telle que la prédiction. On parle de d'apprentissage profond quand la méthode consiste en un réseau de plusieurs couches successives de neurones [50]. Les méthodes neuronales sont des techniques de classification et de prédiction différentes des méthodes classiques car tenant compte de la non linéarité de certaines données[46].

1.2.1 Architecture d'une couche neuronale

L'architecture d'une méthode neuronale se présente sous forme d'un réseau neuronal qui est l'association, en un graphe plus ou moins complexe, d'objets élémentaires et des neurones artificiels. Les principaux réseaux se distinguent par l'organisation

du graphe, c'est-à-dire leur architecture, leur niveau de complexité par le type des neurones (leurs fonctions de transition ou d'activation), par l'objectif visé tel que **l'apprentissage supervisé ou non, ou l'optimisation.**

La figure 1.2 extrait de [30] montre l'architecture d'un réseau neuronal. On peut y voir que pour avoir la sortie a , il y'a un certain nombre de paramètres qui interviennent. Ces paramètres sont :

- les poids d'une couche de neurones W ;
- le biais d'une couche de neurones b ;
- le stimulus d'un reseau p ;
- le niveaux d'activation d'une couche de neurones n ; et
- la fonction d'activation f .

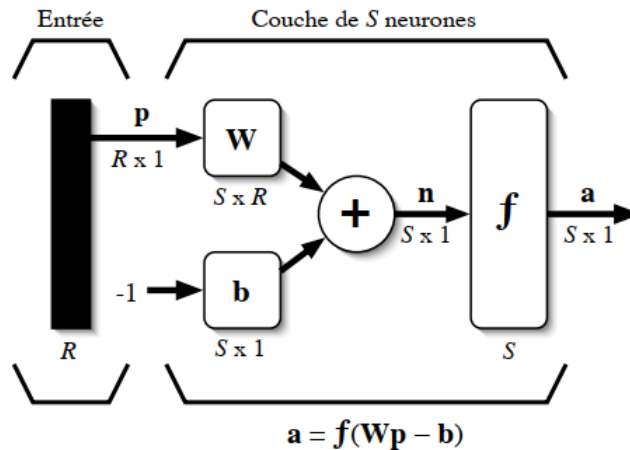


FIGURE 1.2 – Architecture d'un réseau neuronal

La figure 1.3 (basée sur [30]), montre les cinq fonctions de transfert les plus utilisées comme fonctions d'activation d'un réseau neuronal.

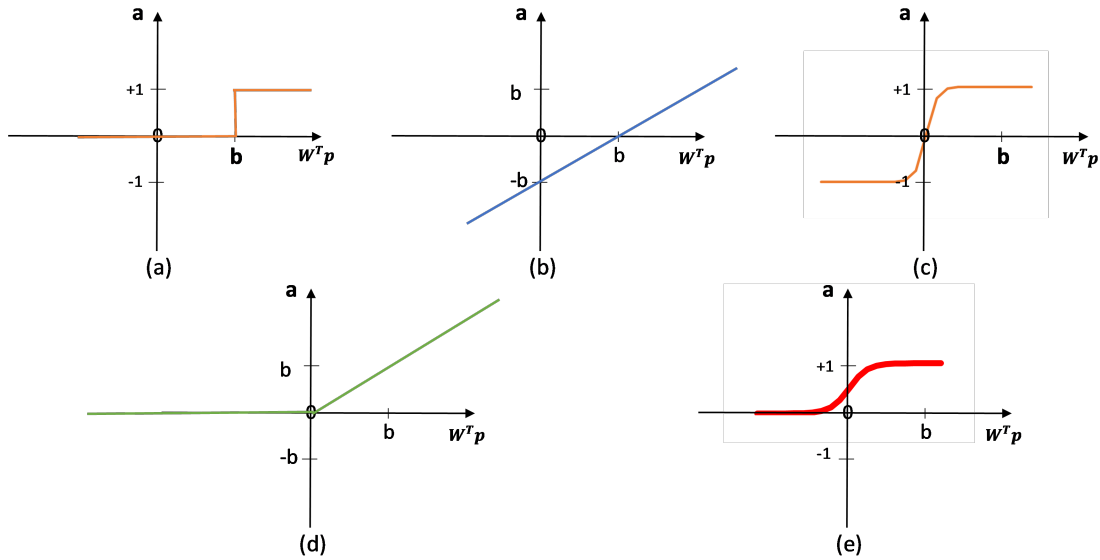


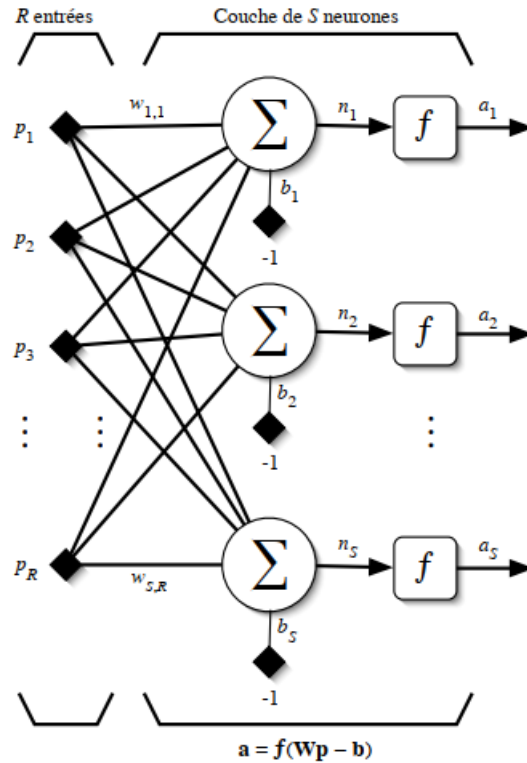
FIGURE 1.3 – Fonction de transfert : (a) du neurone «seuil»; (b) du neurone «linéaire», (c) du neurone «tangente hyperbolique», (d) du neurone «RELU» et (e) du neurone «Sigmoïde»

1.2.2 Un réseau de neurones (plusieurs couches)

Un réseau de neurones est un maillage de plusieurs neurones, généralement organisé en couches. Pour construire une couche de S neurones, il s'agit simplement de les assembler. En effet, la couche est totalement connectée parce que les S neurones d'une même couche sont tous branchés aux R entrées [30]. Ensuite, un poids $w_{i,j}$ (designe l'élément (i, j) de W) est associé à chacune des connexions. Ainsi on a la matrice des poids W définie par :

$$\begin{pmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,R} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,R} \\ \vdots & \vdots & \ddots & \vdots \\ w_{R,1} & w_{R,2} & \cdots & w_{R,R} \end{pmatrix}$$

Les couches précédentes sont les entrées des couches suivantes, ainsi de suite jusqu'à avoir la couche de sortie. Dans la figures 1.4 (extrait de [30]) on a un aperçu global des réseaux de neurones.

FIGURE 1.4 – Architecture d'un réseau de S neurones

1.3 Les modèles de réseaux de neurones étudiés

Nous allons relater les quatre réseaux de neurones qui seront testés avec les données expérimentales recueillies.

1.3.1 Le perceptron multicouches (PMC)

Le perceptron multicouches (PMC) est un réseaux de neurones basé sur un apprentissage supervisé où la structure des classes est connue *a priori*. Il est notamment utilisé à des fins de classifications et de prédictions.

Dans un PMC, un neurone de la couche cachée k est connecté en sortie à tous les neurones de la couche $k - 1$ et en entrée à tous les neurones de la couche $k + 1$. Étant

donné la complexité des données et des problèmes à résoudre, les neurones simples ne sont pas appropriés (par exemple, en raison de la non-linéarité des données) [36]. Les étapes suivantes sont réalisées lors de la création d'un PMC[45]. La figure 1.5 (extrait de [50]) montre un exemple de réseau neuronal PMC avec 5 entrées, 4 sorties et 2 couches cachées.

- chaque neurone a des entrées, qui sont des variables à valeur réelle, notées X_1, \dots, X_n , et des sorties, notées Y_1, \dots, Y_k ;
- l'entrée X_i est associée à un poids, noté w_i ;
- un calcul est effectué sur chaque neurone, consistant en un produit vectoriel entre le vecteur d'entrée $X^T = (X_1; \dots; X_n)$ et le vecteur de poids $w^T = (w_1, \dots, w_n)$ et le résultat est ajouté à une constante w_0 (appelée le biais);
- le résultat est lui-même l'entrée d'une fonction non linéaire $f(w_0 + \langle X, w \rangle)$, appelée fonction d'activation;

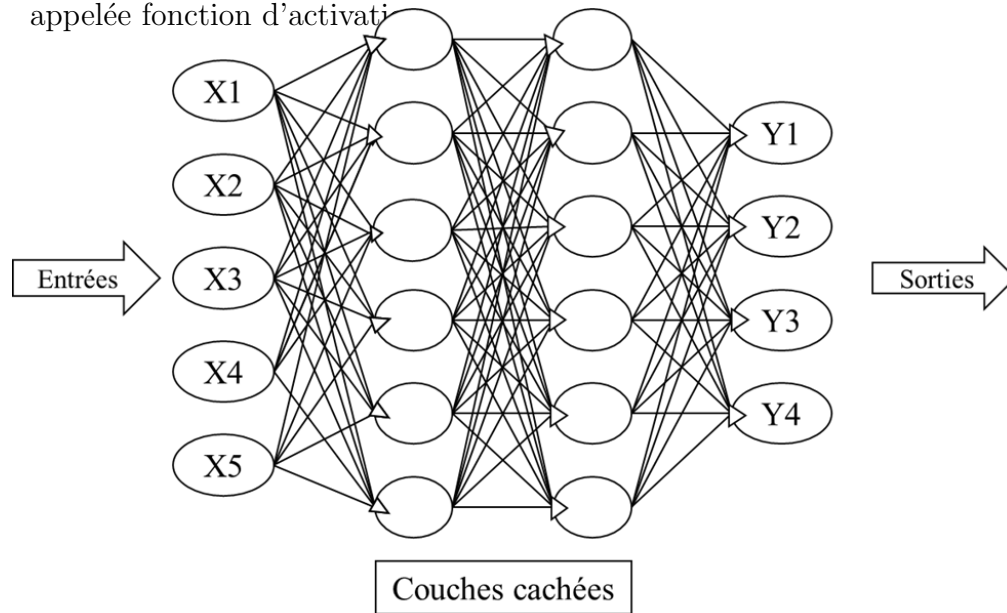


FIGURE 1.5 – Représentation matricielle d'un réseau PMC à 2 couches cachées avec 4 sorties.

L'objectif d'un réseau neuronal non linéaire composé de m couches cachées et k couches de sortie est d'atteindre les valeurs optimales des poids w et des biais w_0 . Cependant, l'obtention de ces valeurs est difficile en raison de la non-linéarité, du

grand nombre de paramètres et de la complexité du modèle. Pour résoudre ce défi, nous faisons appel à l'algorithme de rétropropagation de l'erreur [39]. Ce dernier est un mécanisme utilisé dans l'apprentissage des réseaux neuronaux, en particulier des PMC. Cet algorithme est utilisé pour ajuster les poids et les biais du réseau neuronal afin de minimiser l'erreur de prédiction pendant l'apprentissage. Le processus commence par une initialisation aléatoire des poids et des biais au début de la formation. Ensuite, une fois que les entrées ont été présentées au réseau neuronal, elles sont propagées à travers le réseau, de la couche d'entrée à la couche de sortie. Ainsi, chaque neurone calcule une valeur de sortie en appliquant une fonction d'activation f aux entrées pondérées par les poids w et en ajoutant le biais w_0 (propagation vers l'avant). La perte est calculée en comparant les sorties du réseau aux sorties attendues. La rétropropagation commence lorsque l'erreur est propagée vers l'arrière du réseau, de la couche de sortie à la couche d'entrée. L'objectif est de déterminer comment chaque poids et chaque biais contribuent à l'erreur totale du réseau. Pour ce faire, on utilise la dérivée de la fonction de perte par rapport aux poids et aux biais. Enfin, les poids et les biais sont mis à jour en minimisant l'erreur commise (exemple de l'optimiseur d'Adam) [39].

En règle générale, des critères sont choisis pour déterminer quand le processus d'itération doit cesser. Par exemple :

- convergence des paramètres (les poids et les biais cessent de changer) ;
- fixer un nombre maximum de boucles d'apprentissage ;
- choisir une limite à ne pas dépasser pour l'erreur quadratique moyenne ; et
- contrôler un indice de performance.

L'importance d'établir de tels critères découle de la nécessité d'éviter les phénomènes de sur-apprentissage et de saturation, qui correspondent à un indice de performance qui se détériore une fois que son optimum a été atteint et à une vitesse de convergence si lente que le réseau ne s'améliore plus.

1.3.2 Les réseaux de neurones récurrents (RNN)

Depuis 1979, les réseaux neuronaux récurrents (RNN) sont parmi les réseaux neuronaux les plus utilisés. Ils sont conçus pour apprendre des modèles séquentiels. Il s'agit de réseaux de neurones avec des connexions de rétroaction, souvent appelées boucles fermées [14]. Les techniques basées sur ces types de modèles ont été utilisées pour relever un large éventail de défis. À la fin des années 80, un certain nombre de chercheurs ont établi des réseaux neuronaux simples partiellement destinés à l'apprentissage de chaînes de lettres. Ils sont construits à partir de réseaux de type feedforward et fonctionnent de manière similaire au cerveau humain [33]. En d'autres termes, ces modèles d'apprentissage peuvent anticiper des données séquentielles d'une manière que d'autres algorithmes ne peuvent pas anticiper. La structure d'un RNN avec cinq entrées et quatre sorties est illustrée à la figure 1.6.

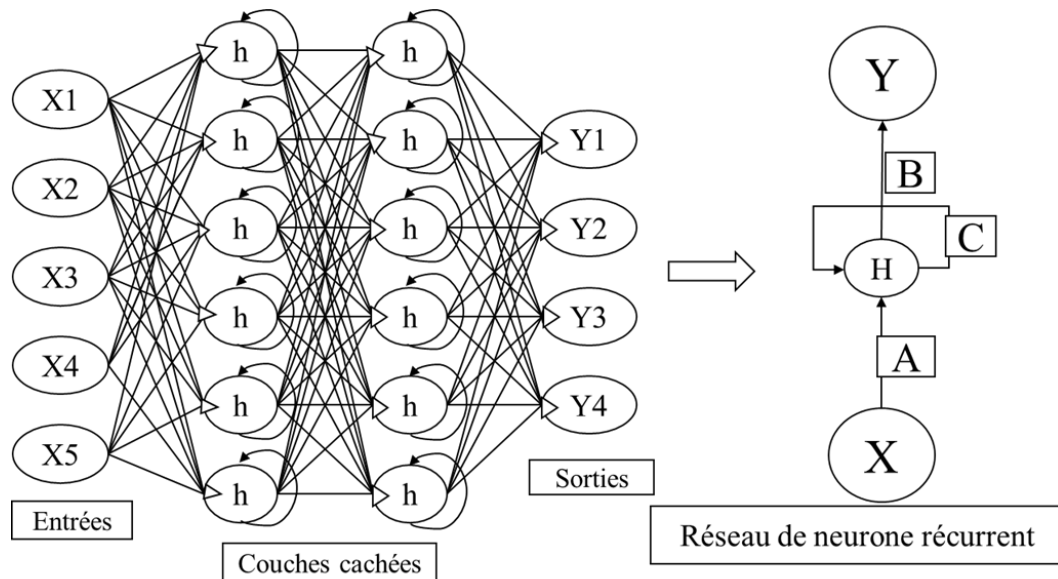


FIGURE 1.6 – Réseau neuronal récurrent RNN de 2 couches cachées avec 4 sorties [6].

Le côté droit de la figure 1.6 montre en détail les paramètres du modèle RNN avec X comme entrée et Y comme sortie. En effet, H est un état caché du temps et sert de *mémoire* au réseau. Son calcul est basé sur l'entrée actuelle X et l'état caché du pas de temps précédent. En outre, il s'agit d'une matrice C de poids paramétrés. Comme

le RNN est utilisé sur les données temporelles, la matrice C est mise à jour en utilisant la rétropropagation, connue sous le nom de rétropropagation dans le temps C .

La partie gauche de la figure 1.6 représente les entrées et les sorties des réseaux neuronaux qui sont totalement indépendantes les unes des autres. Toutefois, dans d'autres situations, par exemple lorsqu'il s'agit de prédire la valeur suivante d'une variable, les valeurs précédentes doivent être mémorisées. C'est ainsi que le RNN a été développé, avec les deux couches cachées pour résoudre le problème. Les états cachés, qui conservent des informations spécifiques sur une séquence, sont les composants les plus importants du RNN [23].

1.3.3 Les réseaux à mémoire à long et court terme (LSTM)

Il s'agit d'une approche de réseau neuronal récurrent. Elle est conçue pour mieux refléter les séquences temporelles et leurs dépendances au long terme comparativement aux RNNs standards. En ce qui concerne la conception du modèle LSTM, des unités spéciales sont incorporées dans chaque couche cachée récurrente, comme le montre la figure 1.7. L'état temporel du réseau est contenu dans des cellules qui sont conservées dans des blocs de mémoire par le biais d'auto connexions (par exemple r_{t-1} dans la figure 1.7). Parallèlement, le flux d'informations est contrôlé par des unités multiplicatives spéciales appelées portes (f_t , i_t et o_t). Dans certains modèles, les blocs de mémoire peuvent également comporter des portes d'entrée qui régissent le flux des activations d'entrée dans les cellules de mémoire, ainsi que des portes de sortie.

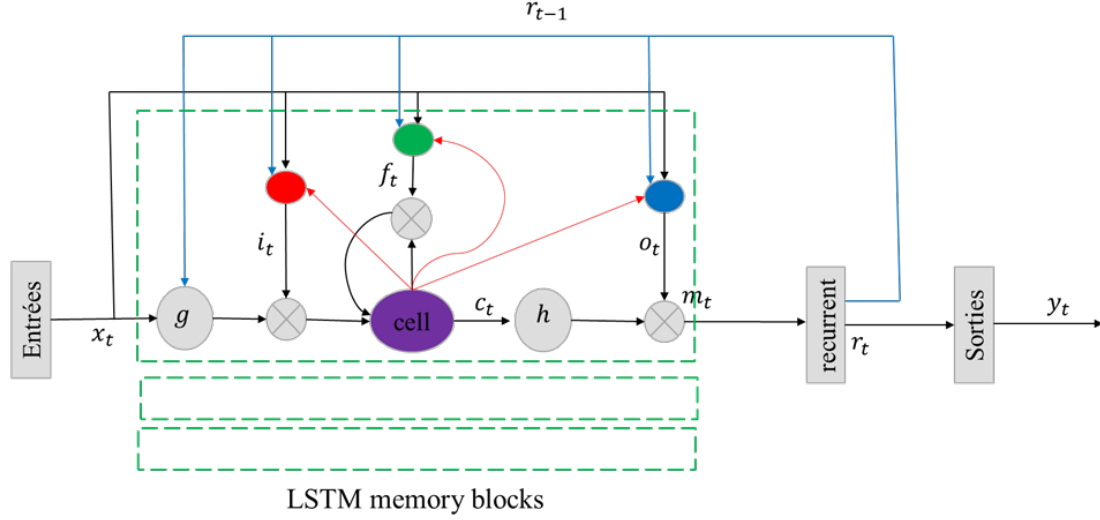


FIGURE 1.7 – Représentation de l'architecture d'un LSTM. Un seul bloc de mémoire est représenté pour plus de clarté [40].

La porte de sortie (o_t) contrôle le flux d'activation dans le réseau, comme le montre la figure 1.7. La porte d'oubli (f_t) est ensuite ajoutée au bloc de mémoire. La mise à l'échelle de l'état interne de la cellule (c_t) par la porte de sortie précède son inclusion en tant qu'entrée dans la cellule via la porte d'oubli. Le modèle de réseau neuronal LSTM prend en compte la composante temporelle. Par conséquent, un calcul des correspondances entre les séquences d'entrée $X = (X_1, \dots, X_T)$ et $Y = (Y_1; \dots; Y_k)$ est effectué. Ainsi, les activations des unités du réseau sont déterminées à l'aide des équations (1.4) à (1.9), en tenant compte de la plage de temps $t = 1$ à T [40].

$$i_t = \sigma(W_{iX}X_t + W_{im}m_{t-1} + W_{ic}c_{t-1} + b_i) \quad (1.4),$$

$$f_t = \sigma(W_{fX}X_t + W_{fm}m_{t-1} + W_{fc}c_{t-1} + b_f) \quad (1.5),$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g(W_{cX}X_t + W_{cm}m_{t-1} + b_c) \quad (1.6),$$

$$o_t = \sigma(W_{oX}X_t + W_{om}m_{t-1} + W_{oc}c_t + b_o) \quad (1.7),$$

$$m_t = o_t \odot h(c_t) \quad (1.8),$$

$$Y_t = \phi(W_{Ym}m_t + b_Y) \quad (1.9).$$

Avec :

- W La matrice de poids ;
- W_{ix} La matrice des poids de porte à l'entrée ;
- $W_{im}, W_{fm}, W_{cm}, W_{Ym}, W_{om}, W_{fX}, W_{cX}$ et W_{oX} les matrices de poids respectivement de la passerelle, de la porte d'oubli, de la cellule, des sorties, de la porte de sortie, de l'entrée de la porte d'oubli, de l'entrée dans la cellule et de l'entrée à la porte de sortie ;
- W_{ic}, W_{fc} et W_{oc} sont les matrices de poids diagonale pour les points de connexion ;
- b sont les vecteurs de biais ;
- σ sont les vecteurs de biais ;
- i la fonction sigmoïde logistique de la passerelle ;
- f la fonction sigmoïde logistique de la porte d'oubli ;
- o la fonction sigmoïde logistique de la porte de sortie ;
- c le vecteur d'activation cellulaire ;
- m le vecteur d'activation de la sortie de la cellule ;
- \odot le produit élément par élément des vecteurs ;
- g et h sont les fonctions d'activation pour l'entrée et la sortie de la cellule ; et
- ϕ La fonction d'activation de sortie réseau.

Les données non linéaires recueillies au fil du temps sont examinées dans ce travail. Il est essentiel de comprendre que les algorithmes LSTM sont des architectures profondes dans le sens où elles peuvent être considérées comme un réseau neuronal de feedforward se déroulant dans le temps avec les mêmes paramètres de modèle partagés par chaque couche. Les entrées du modèle sont gérées par de nombreuses couches non linéaires, tandis que les caractéristiques d'un instant donné sont traitées par une seule couche non linéaire avant de contribuer à la sortie [11, 13, 47].

1.3.4 Les unités récurrentes fermées (GRU)

Les réseaux d'unités récurrentes fermées (GRU) sont bien connus pour leur capacité à bien fonctionner dans les tâches séquentielles d'apprentissage des données. En fait, cette méthode a résolu certains problèmes telle que la disparition du gradient et l'explosion dans les réseaux de neurones récurrents classiques (RNN) lors de l'apprentissage des dépendances à long terme [16, 38, 44]. L'approche GRU est également destinée à traiter le problème du temps d'apprentissage LSTM en plus du problème du gradient. En fait, la réduction du temps de formation a entraîné la mise en place de l'unité récurrente fermée proposée en 2014 [18]. Comme il n'y a pas de porte de sortie, il s'agit d'une version améliorée de LSTM qui a un temps d'apprentissage plus court que les LSTM ordinaires avec moins de paramètres [27]. L'architecture du modèle GRU est illustrée à la figure 1.8.

La figure 1.8 montre la présence de deux types d'entrées à chaque itération t .

- $h(t-1)$ Vecteur de sortie du calque caché précédent ; et
- $x(t)$ Vecteur d'entrée actuel.

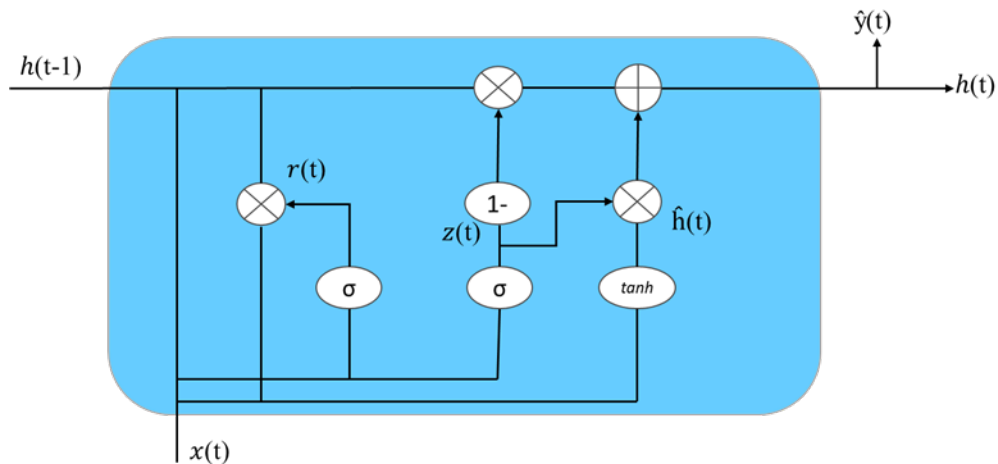


FIGURE 1.8 – Représentation de la structure du réseau GRU (gated recurrent unit) [52].

Il convient de noter que dans le processus d'apprentissage avec le modèle GRU, chaque porte peut être obtenue par une opération logique et une transformation non linéaire de l'entrée [51] comme le montrent les équations (1.10) à (1.13).

$$r(t) = \sigma_g(W_r x(t) + U_r h(t-1) + b_r) \quad (1.10),$$

$$z(t) = \sigma_g(W_z x(t) + U_z h(t-1) + b_z) \quad (1.11),$$

$$\hat{y}_t = h(t) = (1 - z(t)) \circ h(t-1) + z(t) \circ \hat{h}(t) \quad (1.12),$$

$$\check{h}(t) = \sigma_h(W_h x(t) + U_h(r(t) \circ h(t-1)) + b_h) \quad (1.13).$$

Avec :

- $z(t)$ Le vecteur de porte de mise à jour ;
- $r(t)$ Le vecteur de réinitialisation de la porte ;
- W et U les matrices et vecteurs de paramètres ;
- \circ représente le produit terme à terme des vecteurs ;
- σ_g Une fonction sigmoïde ; et
- σ_h Une tangente hyperbolique.

La technique d'apprentissage GRU utilise souvent la rétropropagation dépendante du temps (BPTT) et/ou l'apprentissage récurrent en temps réel (RTRL) [9]. L'optimiseur d'Adam surpasse également les algorithmes tels que la descente de gradient stochastique (SGD), RMSProp, Adadelta et Adagrad [53]. Toutes ces méthodes visent à ajuster les poids du modèle GRU de manière à minimiser l'erreur lors de la prédiction de séquences temporelles.

Chapitre 2

Application au cas de la machine à combustion à biomasse

Dans toute étude, il est important de connaître la structure des données. Dans cette section, il s'agit de décrire la machine et de lister les variables avec lesquelles on travaille. Par la suite, nous ferons une statistique descriptive et une étude exploratoire des données et terminer par l'application des réseaux de neurones les données de la machine à combustion à biomasse.

2.1 Revue de littérature en biomasse et problématique de la recherche

La combustion à biomasse est une alternative viable à la combustion de combustibles fossiles pour la production de chaleur industrielle et le chauffage des locaux [42]. En conséquence, l'incorporation de la biomasse serait un aspect important dans le développement de systèmes énergétiques durables. La technologie la plus fréquemment

utilisée à cette fin est la combustion à grille [43]. Ces dernières années, la biomasse a pris de l'importance en tant que combustible pour le chauffage, conformément à l'objectif à atteindre tel que la réduction des émissions de dioxyde de carbone [35]. Le contrôle des chaudières à biomasse est donc devenu plus important, car elles ont des objectifs de faibles émissions, un rendement élevé, une bonne capacité de modulation de charge et une grande flexibilité du combustible [43].

De nombreuses enquêtes sur le contrôle des réseaux de combustion ont été menées [42, 12]. On peut prendre l'exemple des contrôleurs PID (Régulateur Proportionnel Intégral Dérivé) qui sont utilisés à ces fins [3, 32, 19]. Ces types de contrôleurs (PID) sont généralement utilisés pour réguler la combustion en ajustant les paramètres du système en fonction de la différence entre la température désirée (consigne) et la température réelle. Les performances de la machine sont malheureusement limitées puisqu'il s'agit d'un système non linéaire avec différentes entrées et sorties. Mais la plupart des contributions ont étudié des modèles linéaires ou linéarisés produits empiriquement qui sont utilisés comme base pour les méthodes de contrôle prédictif des modèles [21, 29]. C'est dans ce contexte que de nombreux modèles et systèmes de contrôle basés sur des modèles ont été développés, promettant des avancées majeures dans ce domaine particulier.

Les chaudières modernes à grille à biomasse sont géométriquement optimisées et permettent d'avoir des stratégies efficaces de distribution de l'air. Cette conception se traduit par des émissions polluantes minimisées et une performance énergétique élevée [49]. Cependant, en raison des qualités variables de la biomasse (teneur en humidité, teneur en cendres, composition chimique) et des problèmes associés aux cendres (encrassement des échangeurs de chaleur), son contrôle est difficile et fait toujours l'objet d'études et de développements continus [42]. Cela rend difficile pour les techniciens de gérer les alertes inattendues qui se produisent lorsqu'il y a un problème. Cela a souvent un impact sur le fonctionnement optimal des machines à combustion à biomasse [35].

L'apprentissage automatique a été utilisé dans les recherches récentes sur la biomasse pour la prévision et la modélisation de l'optimisation des paramètres des machines à combustion à biomasse. Les auteurs de [26, 54] offrent des données spécifiques sur le traitement hydrothermal des biocarburants, des produits chimiques et des matériaux. Parallèlement, les auteurs de [48] ont mené les plus grandes investigations dans ce domaine dans le cadre des études sur le réchauffement climatique avec la fabrication d'hydrogène vert et durable issu du procédé de gazéification de la biomasse. Il y a également eu des études sur la gestion de la biomasse, dans le but de développer un mécanisme de contrôle des chaudières à grille à biomasse [16, 38, 43]. En outre, les auteurs de [24] ont examiné l'apprentissage automatique pour la conversion thermique de l'énergie de la biomasse. Les auteurs de [2] ont mené une étude des méthodologies de prédiction de la biomasse avec l'apprentissage automatique sur des données d'humidité du sol dérivées de la télédétection, où des réseaux neuronaux artificiels sont utilisés pour reproduire le comportement dynamique d'une machine à combustion à biomasse.

Le contrôle flou [20], les réseaux neuronaux [49] ou une combinaison des deux [31] sont des moyens de modéliser les machines à combustion à biomasse. Chacune de ces approches a une applicabilité généralement limitée. Bien que l'entraînement des réseaux neuronaux nécessite généralement une grande quantité de données [42, 49], l'expertise d'un spécialiste de la biomasse pourrait améliorer considérablement les résultats obtenus avec ces réseaux. Par conséquent, étant donné la nature des données provenant des machines à combustion à biomasse, les réseaux neuronaux pourraient constituer une option de contrôle prometteuse [44].

En raison des propriétés variables de la biomasse, telles que la teneur en humidité, la teneur en cendres, et la composition chimique, ainsi que des problèmes associés aux cendres, comme l'encrassement des échangeurs de chaleur, son contrôle s'avère complexe. Par conséquent, la gestion de la biomasse est un domaine de recherche et de développement en constante évolution. Les techniciens sont souvent confrontés à des

alarmes inattendues lors de dysfonctionnements, ce qui perturbe le fonctionnement optimal des machines à combustion à biomasse.

Dans ce contexte, se pose la question sur la manière d’obtenir les données d’entrée nécessaires pour garantir des performances optimales en sortie. Notre objectif principal est de développer un modèle prédictif pour les paramètres de ces machines, permettant ainsi d’anticiper et de répondre de manière plus efficace aux fluctuations et aux problèmes liés aux valeurs d’entrées et de sorties de la machine à biomasse.

2.2 Description de la machine à biomasse et des données provenant de cette machine

Tout d’abord, des données expérimentales ont été collectées à partir d’une machine à biomasse. Ces données comprennent des variables d’entrée et de sortie enregistrées pendant 24 heures de fonctionnement de la machine. La contribution présentée dans ce document repose sur l’utilisation de réseaux de neurones pour établir un modèle prédictif, appelé modèle inverse, qui permettra de déterminer les paramètres d’entrée nécessaires pour obtenir les sorties souhaitées.

2.2.1 La machine à combustion à biomasse

La figure 2.1 illustre une machine à combustion à biomasse. Elle est spécialement conçue pour produire de la vapeur en chauffant des écorces d’arbres. Dans le processus de production de bois, il est crucial de sécher le bois après l’avoir écorcé et coupé. Pour ce faire, une chaudière remplie d’eau est chauffée, et une partie de la vapeur produite est utilisée pour sécher le bois fraîchement coupé. Les paramètres d’entrée du

système de chauffage sont essentiels pour son fonctionnement. Une fois ces paramètres configurés, le système est activé, ce qui génère les paramètres de sortie nécessaires.

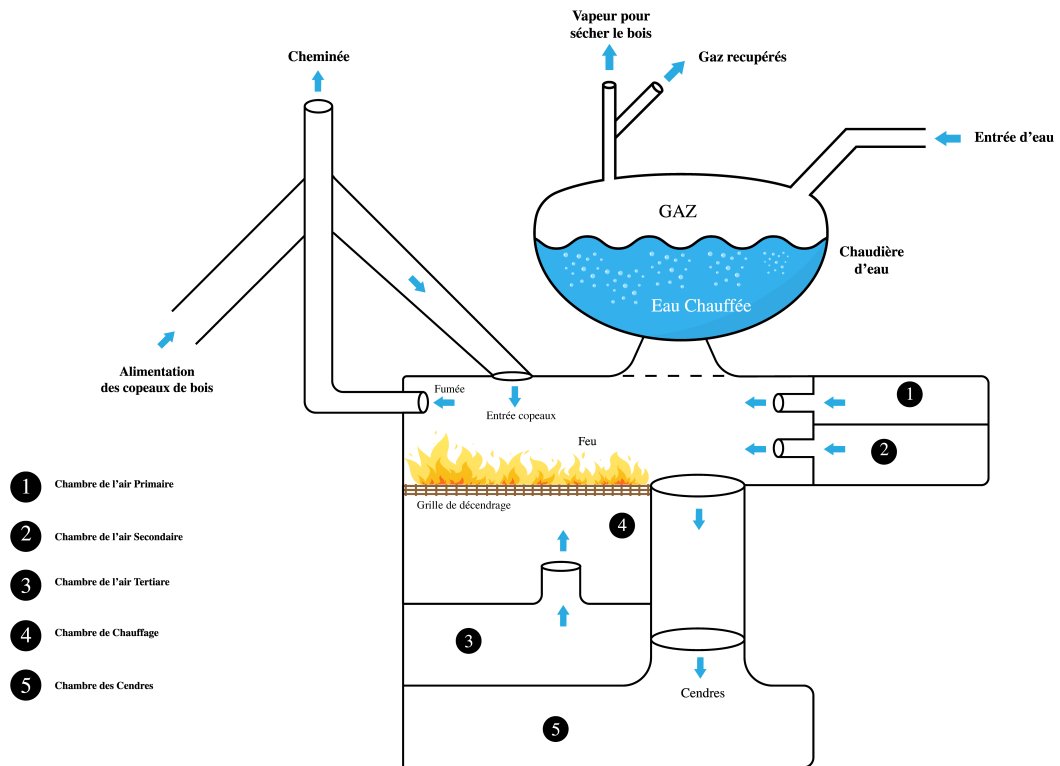


FIGURE 2.1 – Schéma de la machine à combustion à biomasse.

Ce processus de configuration peut s'avérer complexe pour les experts travaillant avec la machine. En effet, pour obtenir la quantité de vapeur nécessaire au séchage du bois, il est impératif de paramétrer les valeurs optimales de ces paramètres. Dans la figure 2.2, nous présentons la représentation systémique du système de la machine à combustion à biomasse que nous étudions.

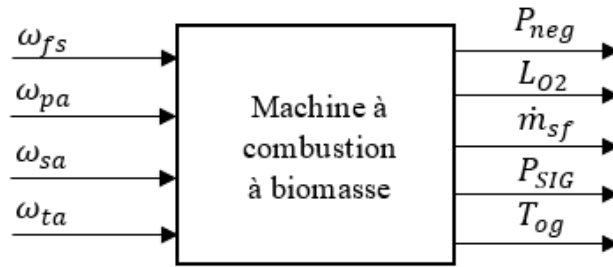


FIGURE 2.2 – Système d'entrée/sortie de la machine à biomasse.

L'ensemble des variables dans notre base de données tirées à partir de la machine est listé ci-dessous.

Les entrées de la machine :

- ω_{fs} vitesse du ventilateur de tirage de la chambre de combustion (en %) ;
- ω_{pa} vitesse du ventilateur primaire de la chaudière (en %) ;
- ω_{sa} vitesse du ventilateur secondaire de la chaudière (en %) ;
- ω_{ta} vitesse du ventilateur tertiaire de la chaudière (en %).

Les sorties de la machine :

- P_{neg} pression négative dans la chambre de combustion (en poH2O ou pouces d'eau) ;
- L_{O_2} taux d'oxygène des gaz de combustion (en %) ;
- \dot{m}_{sf} débit de vapeur de la chaudière (en Lb/h ou livres par heure) ;
- P_{SIG} pression de vapeur de la chaudière (en livres par pouce carré) ; et
- T_{og} Température des gaz de combustion à la sortie de convection de la chaudière en °F.

Le schéma du système présenté à la figure 2.2 est couramment utilisé pour simuler la machine à combustion à biomasse dans la plupart des travaux de modélisation disponibles dans la littérature de la biomasse. Cette approche est appropriée lorsque l'objectif principal est de prédire les sorties du système en fonction des entrées connues.

Cependant, dans le contexte de l'étude présentée ici, la principale inconnue est

constituée des signaux d'entrée. L'objectif est d'identifier les intrants nécessaires pour obtenir les sorties requises de la machine. En conséquence, nous examinons un modèle prédictif inverse, dans lequel les sorties du modèle deviennent ses nouvelles entrées et vice versa, comme illustré à la figure 2.3.

Par conséquent, les variables d'entrée du modèle inverse sont P_{neg} (par pouces d'eau), L_{O_2} (%), \dot{m}_{sf} (livres par heure), P_{SIG} (Livres par pouce carré) et T_{og} (°F). Les sorties du modèle seront alors l'ensemble des vitesses de la machine, c'est-à-dire ω_{fs} (%), ω_{pa} (%), ω_{sa} (%) et ω_{ta} (%). Cependant, avant la modélisation, les données doivent être traitées, recodées pour une meilleure visualisation des variables dans les graphiques et organisées pour améliorer les performances de l'algorithme de prédiction.

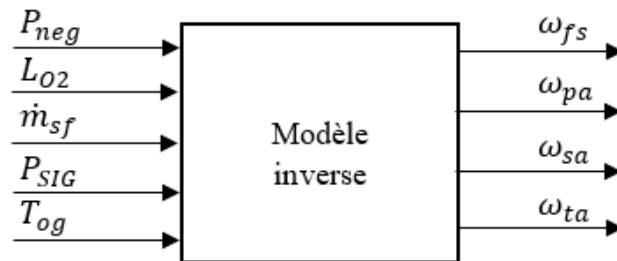


FIGURE 2.3 – Système d'entrée/sortie de la machine à combustion à biomasse avec modèle inverse

2.2.2 Statistique descriptive sur la base de données

La présente étude est basée sur un ensemble de données industrielles d'une machine à combustion à biomasse d'un client de **Combustion Expert Énergie**¹ situé à la Beauce à Québec. Les données expérimentales comprennent les valeurs des nombreux réglages et paramètres de la machine. Elles ont été saisies durant le mois d'Avril de l'année 2023. La période d'échantillonnage est fixée à 5 secondes. Les premières données ont été collectées pour créer une base de données de 17281 lignes (une

1. Cliquez ici pour plus d'informations.

ligne toutes les 5 secondes), caractérisées par les neuf (9) mesures effectuées entre 00h00min01sec et 23h59min56sec.

TABLE 2.1 – Extrait de la base de données expérimentales

Date	Time	Wfs	Wpa	Wsa	Wta	Pneg	Lo2	msf	Psig	Tog	
2023-04-16	00:00:01	75,24368286	47,5	41,25	41,25	41,25	0,71680707	6,34724998	14343,69434	88,91249847	390,7000122
2023-04-16	00:00:06	76,20526886	47,5	41,25	41,25	41,25	0,67414814	6,36956263	14153,25	88,79062653	390,5
2023-04-16	00:00:11	76,04463959	47,5	41,25	41,25	41,25	0,66614687	6,38137484	14024,13477	88,74375153	390,3999939
2023-04-16	00:00:16	74,97409821	47,5	41,25	41,25	41,25	0,73466372	6,37612486	13949,89355	88,803750153	390,1000061
2023-04-16	00:00:21	75,03568268	47,5	41,25	41,25	41,25	0,72992396	6,38662481	14072,55273	88,94062805	390
2023-04-16	00:00:26	74,84815979	47,5	41,25	41,25	41,25	0,7260555	6,39318752	14079,00879	88,72499847	390
2023-04-16	00:00:31	74,76290894	47,5	41,25	41,25	41,25	0,72811824	6,41025019	13812,70898	88,57499695	389,7999878
2023-04-16	00:00:36	74,99942017	47,5	41,25	41,25	41,25	0,70891094	6,42731237	13722,3291	88,60312653	389,8999939
2023-04-16	00:00:41	74,42790222	47,5	41,25	41,25	41,25	0,73189658	6,42731237	13865,96875	88,76249695	389,7999878
2023-04-16	00:00:46	75,39483643	47,5	41,25	41,25	41,25	0,67589456	6,41025019	13927,29883	88,875	389,7000122
2023-04-16	00:00:51	75,07598114	47,5	41,25	41,25	41,25	0,69517648	6,38137484	14056,41406	89,00624847	389,7000122
2023-04-16	00:00:56	75,18428802	47,5	41,25	41,25	41,25	0,68419009	6,34724998	14064,4834	89,02500153	389,7000122
2023-04-16	00:01:01	74,09275818	47,5	41,25	41,25	41,25	0,74157602	6,33018732	14022,52148	89,13749695	389,8999939
2023-04-16	00:01:06	74,94752502	47,5	41,25	41,25	41,25	0,68848294	6,31443739	14114,51465	89,23124695	389,7000122
2023-04-16	00:01:11	74,11203003	47,5	41,25	41,25	41,25	0,74230236	6,2973752	14282,36523	89,31562805	389,8999939
2023-04-16	00:01:16	74,54908752	47,5	41,25	41,25	41,25	0,70649129	6,26324987	14643,88672	89,31562805	389,8999939
2023-04-16	00:01:21	74,23049927	47,5	41,25	41,25	41,25	0,719688	6,23962498	14682,62109	89,09062195	389,8999939
2023-04-16	00:01:26	74,40790558	47,5	41,25	41,25	41,25	0,70403188	6,22387505	14627,74805	88,98750305	390
2023-04-16	00:01:31	74,64121246	47,5	41,25	41,25	41,25	0,68390518	6,22387505	14480,87891	88,875	390
2023-04-16	00:01:36	74,37343597	47,5	41,25	41,25	41,25	0,70835495	6,24093771	14303,3457	88,71562195	389,8999939
2023-04-16	00:01:41	74,49624634	47,5	41,25	41,25	41,25	0,69735271	6,28031254	14150,02246	88,57499695	389,7999878
2023-04-16	00:01:46	74,29429626	47,5	41,25	41,25	41,25	0,71511847	6,2973752	14035,43164	88,40625	389,8999939
2023-04-16	00:01:51	74,50153351	47,5	41,25	41,25	41,25	0,69932932	6,30262518	13875,65332	88,29374695	389,8999939
2023-04-16	00:01:56	74,19035339	47,5	41,25	41,25	41,25	0,70824808	6,30262518	13857,89941	88,34999847	389,8999939
2023-04-16	00:02:01	75,04807281	47,5	41,25	41,25	41,25	0,66971737	6,2973752	13920,84277	88,53749847	389,8999939

Le tableau 2.1 présente un échantillon des données collectées au niveau de la machine. Le tableau contient des données temporelles qui nécessitent un traitement supplémentaire tel que l'analyse des densités, les diagrammes en boîte et les calculs de corrélation. Ce traitement permettra une meilleure compréhension de la structure des données et contribuera à améliorer les performances du modèle inverse proposé.

Le tableau 2.2 représente la moyenne, l'écart type, les valeurs minimales et maximales, les quartiles et les écarts types des variables de la base de données. La moyenne du taux d'oxygène est de 5% avec un écart de 1,57%. L'étude de l'écart type est une étape importante de la modélisation, les variables avec un écart type nul ou presque nul n'améliore pas la performance d'un modèle et il est recommandé parfois de les supprimer. Les variables avec un écart type élevé doivent subir une analyse particulière pour vérifier l'existence de valeurs aberrantes.

TABLE 2.2 – Statistique descriptive de chaque variable à l'étude

Wfs	Wpa	Wsa	Wta	Pneg
Min : 10,00	Min :10,00	Min :13,00	Min :33,00	Min :0,0000
1st Qu : 67,00	1st Qu :30,00	1st Qu :40,00	1st Qu :40,00	1st Qu :0,0000
Median : 70,00	Median :30,00	Median :40,00	Median :40,00	Median :0,0000
Mean : 69,43	Mean :31,61	Mean :38,69	Mean :40,59	Mean :0,1257
3rd Qu : 77,00	3rd Qu :37,00	3rd Qu :40,00	3rd Qu :40,00	3rd Qu :0,0000
Max :100,00	Max :45,00	Max :50,00	Max :50,00	Max :2,0000
sd : 16,425330	sd : 7,574616	sd : 7,421926	sd : 3,500309	sd : 0,33264

Lo2	msf	Psig	Tog
Min :1,000	Min : 1596	Min : 76,00	Min :344,0
1st Qu :4,000	1st Qu : 4684	1st Qu : 83,00	1st Qu :380,0
Median :5,000	Median : 9387	Median : 90,00	Median :392,0
Mean :4,996	Mean : 9987	Mean : 89,68	Mean :391,6
3rd Qu :6,000	3rd Qu :13026	3rd Qu : 94,00	3rd Qu :402,0
Max :9,000	Max :23086	Max : 107,00	Max :429,0
sd : 1,574673	sd : 5840,087238	sd : 8,129100	sd : 16,997847

La variable P_{neg} présente une particularité dans ses statistiques descriptives, comme indiquée dans le tableau 2.2. En effet, après un examen plus attentif, nous avons remarqué que les valeurs du premier quartile et du troisième quartile sont toutes deux nulles. Cela peut indiquer que la distribution des données est très concentrée autour de zéro. Pour cette raison, nous avons choisi de la traiter séparément des huit autres variables dans la suite de l'analyse descriptive.

La figure 2.4 montre les densités des variables. On peut voir que les variables de sortie vitesse primaire (ω_{pa}), vitesse secondaire (ω_{sa}) et vitesse tertiaire (ω_{ta}) de l'air ont

une distribution proche de la distribution normale. Cependant, le test de Kolmogorov-Smirnov réfute l'hypothèse de normalité pour les neuf variables des données car les valeurs des p -value étant très faibles pour chaque variable. Les interruptions (alarmes) au cours du fonctionnement font que ces variables n'ont pas une distribution normale. Par conséquent, pour améliorer les performances du modèle, toutes ces variables seront transformées en variables gaussiennes.

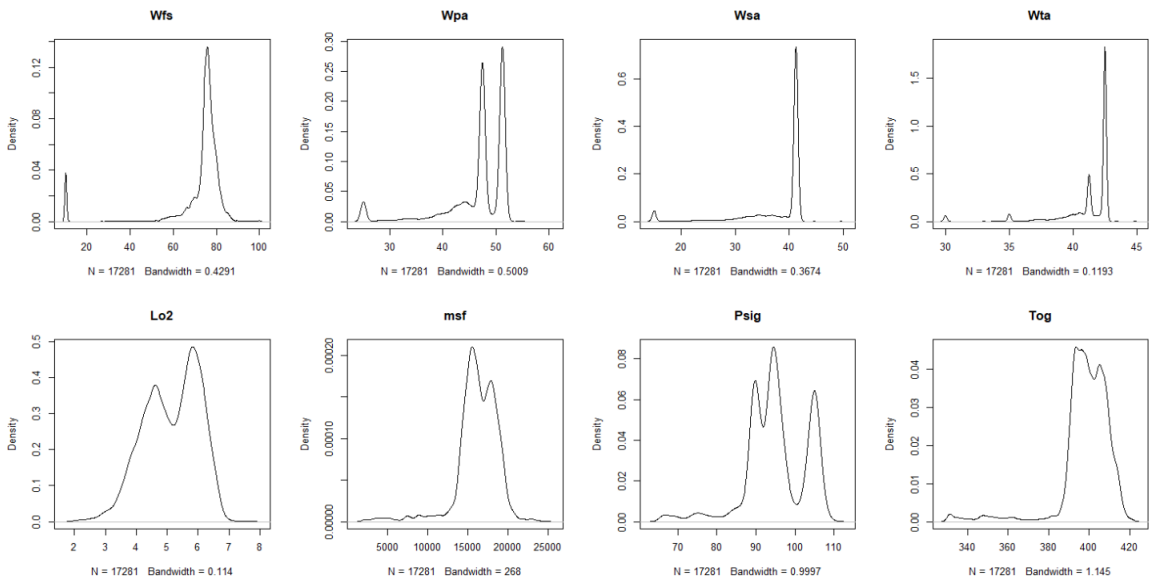


FIGURE 2.4 – Courbes de densité des neuf variables.

La figure 2.5 montre les diagrammes en boîte pour les neuf variables de la base de données. Les variables primaires (ω_{pa}), secondaires (ω_{sa}) et tertiaires (ω_{ta}) de vitesse de l'air présentent chacune des valeurs aberrantes. Cela pourrait être dû à des problèmes de paramétrage à la suite d'un défaut de l'équipement au cours du fonctionnement de la machine à combustion à biomasse. Pour L_{o2} , P_{sig} , et T_{og} les valeurs atypiques ne sont que d'un côté. Et enfin, comme la variable m_{sf} prend de grandes valeurs, elle a des valeurs aberrantes des deux côtés du diagramme.

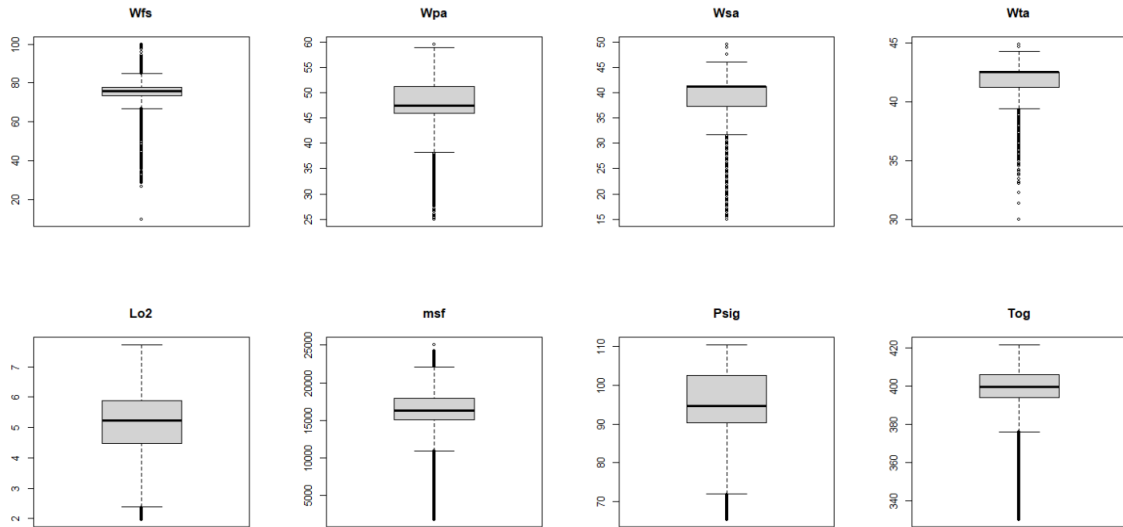


FIGURE 2.5 – Diagrammes en boîte des neuf variables.

2.3 Présentation des analyses exploratoires

Après avoir effectué le traitement des données, une analyse de la corrélation entre les variables a été entreprise. L'objectif de cette étude est de déterminer quelles variables sont effectivement associées et devraient être prises en compte dans le modèle proposé. Pour ce faire, une analyse de corrélation bivariée des variables a d'abord été réalisée, suivie d'une analyse en composantes principales (ACP) sur le jeu de données.

La Figure 2.6 illustre les corrélations entre les variables. Les couleurs varient de 1 (bleu foncé) à -1 (rouge foncé), passant par 0 (blanc). Les petits cercles symbolisent les liens entre deux variables. Plus le diamètre est important, plus la corrélation est forte. Les cercles bleu foncé indiquent une forte corrélation positive, ce qui signifie que les deux variables augmentent ou diminuent dans le même sens. Les cercles rouge foncé indiquent une forte corrélation négative. Les cercles blancs signalent l'absence de lien entre les variables. Les corrélations des variables suggèrent que les variables

de sortie (vitesses, ω_{fs} , ω_{pa} , ω_{sa} et ω_{ta}) sont toutes positivement corrélées. De plus, la température du gaz de sortie (T_{og}) est positivement corrélée avec les deux variables d'entrée (m_{sf}) ainsi qu'avec la pression de vapeur de la chaudière (P_{sig}).

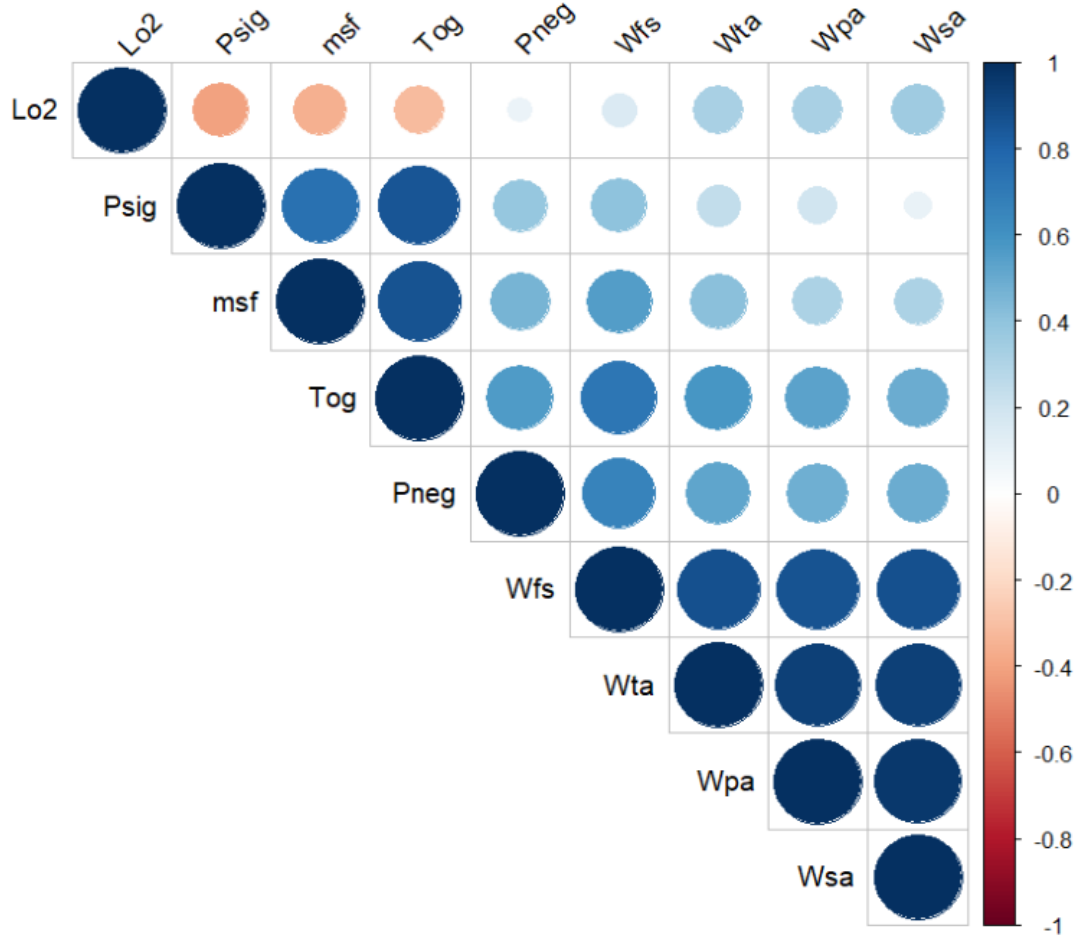


FIGURE 2.6 – Graphique des corrélations deux par deux des variables.

La méthode de l'ACP est utilisée pour réduire le nombre de variables dans un ensemble de données, en ne conservant que les informations les plus pertinentes selon un critère approprié [22]. Elle permet également de déterminer quels individus et quelles variables ont le plus d'influence sur les axes sélectionnés (Dim 1 et Dim 2 sur la figure 2.7). Il réduit donc une base de données de K variables en un ensemble de deux (ou plus) composantes qui sont des combinaisons linéaires des variables d'origine et sont choisies pour maximiser la variance des données. Dans notre étude, nous utilisons

le cercle de corrélation permettant d'identifier toutes les variables qui sont liées aux composantes principales et comment elles interagissent.

La Figure 2.7 montre une analyse plus détaillée des facteurs clés utilisés pour construire les axes de l'analyse en composantes principales (ACP). Les axes factoriels de l'ACP, notamment Dim 1 et Dim 2, sont des axes virtuels résultant de la synthèse des variables de l'analyse. L'angle formé par deux variables au départ de l'origine indique leur degré de corrélation. Si l'angle est droit (90°), les deux variables sont indépendantes, comme c'est le cas entre ω_{pa} et P_{sig} . En revanche, un angle plus petit suggère une corrélation forte et positive, par exemple entre ω_{sa} et ω_{pa} . Enfin les deux variables ont une corrélation forte négative si l'angle est obtus (180°).

La proximité d'une variable par rapport au premier axe factoriel, ou Dim 1 dans la Figure 2.7, indique que cette variable est fortement représentée par cet axe. À l'exception de la teneur en oxygène (L_{o2}), toutes les variables contribuent positivement à la formation du premier axe factoriel. Dans la Figure 2.7, l'axe factoriel Dim 1 représente 57,66% de l'inertie, tandis que le second axe factoriel Dim 2 représente 25,79% de l'inertie. Le taux d'oxygène (L_{o2}) contribue principalement à la création de Dim 2. Ces variables sont fortement corrélées, comme indiqué dans le diagramme de corrélation, qui capture 83,45% de l'information contenue dans les données.

En ce qui concerne l'analyse des variables à chaque composante (Dim 1 et Dim 2), les vitesses (ω_{fs} (0,95 ; 0,13), ω_{pa} (0,86 ; 0,41), ω_{sa} (0,85 ; 0,46), ω_{ta} (0,90 ; 0,35)) ainsi que L_{o2} (0,08 ; 0,82) ont des contributions positives sur les deux axes. En revanche, les variables P_{neg} , m_{sf} , P_{sig} et T_{og} affichent des contributions négatives sur Dim 2, respectivement de -0,05, -0,59, -0,72 et -0,50. Les valeurs attribuées à chaque paramètre sur une composante donnée indiquent comment chaque variable d'origine contribue à cette composante principale (Dim 1 et Dim 2).

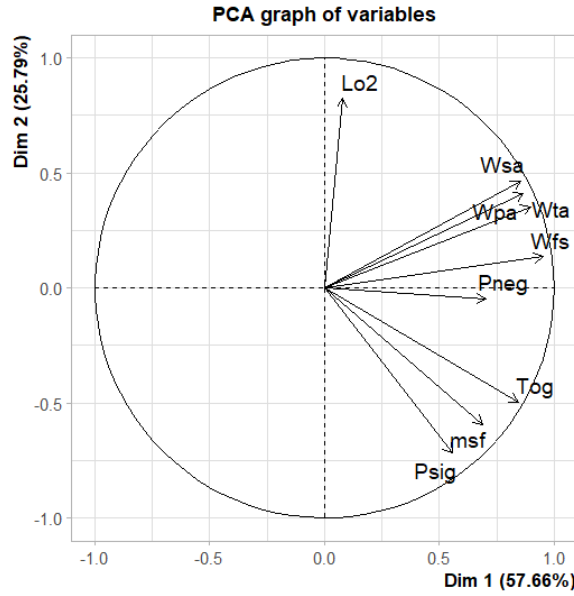


FIGURE 2.7 – Représentation du cercle de corrélation des variables.

2.4 Présentation des résultats du modèle inverse

Afin de valider le modèle avec de nouvelles données, pour déterminer si le modèle fonctionne efficacement à des fins de prédiction, les données sont divisées en données d'apprentissage (80%) et données de test (20%). L'échantillon d'apprentissage est nécessaire pour la prédiction et l'échantillon test est utilisé pour l'analyse des performances. La prédiction des paramètres d'entrée avec chaque modèle a été réalisée par validation croisée. Cette dernière consiste en la division de l'échantillon d'apprentissage en n sous-groupes ($n = 10$ dans notre étude) et l'utilisation de $(n - 1)$ sous-groupes pour prédire les valeurs du $n^{\text{ième}}$ sous-groupe [5]. La figure 2.8 montre une illustration de la validation croisée. Cette opération est répétée pour les autres sous-groupes afin de prédire les probabilités du sous-groupe restant jusqu'à ce que toutes les prédictions soient complètes. L'avantage de cette méthode réside dans son potentiel d'apprentissage élevé, permettant de mieux discriminer les cas pour lesquels le modèle n'a pas été formé.

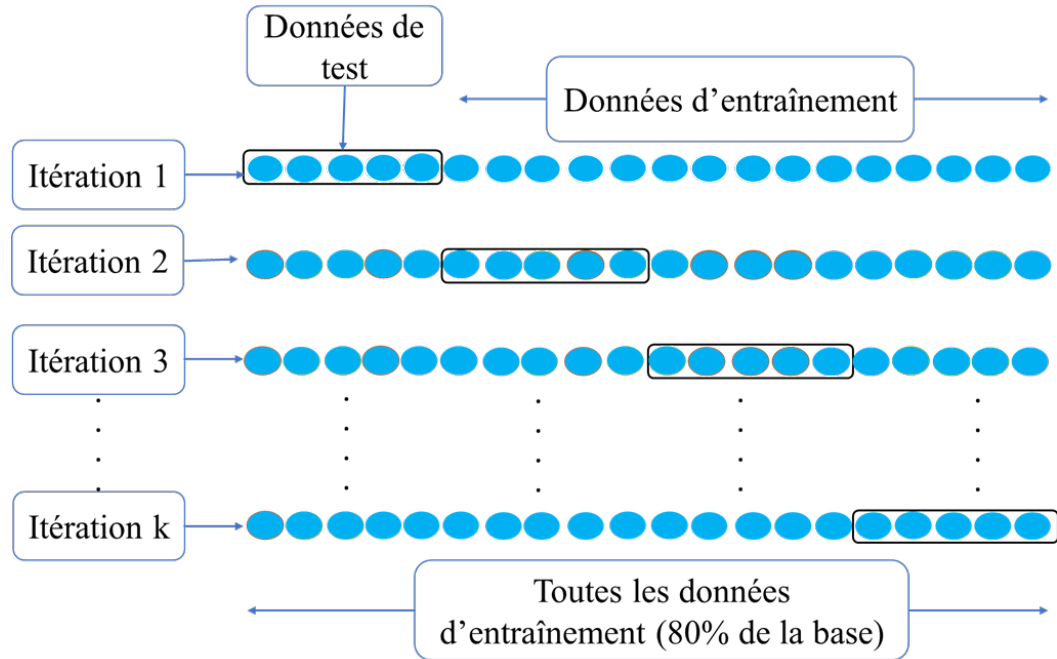


FIGURE 2.8 – Validation croisée [5]

Pour tenir compte de la composante temporelle, les données ont été structurées en séquences temporelles. Les quatre modèles de réseaux de neurones (PMC, RNN, LSTM et GRU) ont été programmés avec *R-Studio*, compilés, entraînés, testés, puis analysés de manière comparative. Les résultats de l'évaluation peuvent être examinés pour déterminer les performances du modèle sur les données de test. Les performances des modèles sont ensuite évaluées à l'aide de mesures de perte (Loss) et de précision (Accuracy).

La précision (Accuracy) est une mesure couramment utilisée pour évaluer les performances des modèles de classification.

$$\text{Précision} = \frac{\text{Nombre de prédictions correctes}}{\text{Nombre total de prédictions}}$$

C'est la proportion de vraies prédictions faites par le modèle par rapport à toutes les prédictions faites par le modèle. La précision du modèle inverse proposé, qui prédit

quatre (4) variables de sortie, est la proportion de bonnes prédictions pour les quatre variables de sortie combinées. Une précision de 1 indique que toutes les prédictions du modèle sont correctes. Avec une précision de 0,7, le modèle prédit avec précision 70% des sorties. Une précision faible peut signifier que le modèle ne parvient pas à capturer les relations complexes entre les entrées et les sorties, ou qu'il y a des failles dans les données ou les méthodes d'apprentissage [15].

La perte (Loss) est l'erreur absolue moyenne (MAE), est une mesure de la différence moyenne entre les données réelles et prévues.

$$Perte = \frac{\sum_{i=1}^N |\hat{y}_i - y_i|}{N}$$

Avec

- \hat{y}_i valeur prédite par le modèle ;
- y_i valeur réelle.

Une MAE plus faible suggère que les prédictions du modèle sont plus précises, avec une différence moyenne plus faible entre les prévisions et les valeurs réelles. Une MAE nulle indique que toutes les prévisions sont parfaites et exemptes d'erreurs [51]. Lors de nos expériences, nous n'avons utilisé que l'hyperparamètre : le nombre d'époques est pris en compte dans notre étude pour la comparaison des différents modèles. Après plusieurs expériences, nous avons retenu les quatre valeurs de ce paramètre 25, 75, 100 et 125, à des fins de comparaison. Alors que les performances (perte et précision) convergent vers la 125e époque, nous avons décidé de nous arrêter à 125 itérations.

Le tableau 2.3 résume la précision et la perte pour les quatre modèles utilisés dans cette étude. Les performances du modèle LSTM montrent une diminution de la perte avec l'augmentation du nombre d'époques, atteignant une précision maximale de 71,02% pour 100 époques. Le modèle GRU est le plus performant car démontrant une amélioration continue au fil des époques, avec une perte de prédiction de 11,97% et une précision de 79,89% pour 125 époques. Le modèle RNN présente des résultats

variables, avec une précision maximale de 71,92% pour 100 époques, bien que la perte augmente à 19,19% à cette époque. En revanche, le modèle PMC montre une performance inférieure avec une précision maximale de 49,47% pour 125 époques. En effet, les variables de notre base de données sont très stochastiques et bruitées, avec des variances élevées $\sigma_{\omega_{fs}}$, $\sigma_{\omega_{pa}}$, $\sigma_{\omega_{sa}}$, $\sigma_{\omega_{ta}}$, $\sigma_{P_{neg}}$, $\sigma_{L_{o2}}$, $\sigma_{m_{sf}}$, $\sigma_{P_{sig}}$ et $\sigma_{T_{og}}$. Cela peut rendre la modélisation PMC difficile.

TABLE 2.3 – Perte et précision des quatre modèles NN.

Model	Epoques	Loss	Accuracy
LSTM	25	0,1979	0,6818
	75	0,2113	0,6902
	100	0,2084	0,7102
	125	0,1948	0,7084
GRU	25	0,142	0,7062
	75	0,15	0,7616
	100	0,1314	0,7865
	125	0,1197	0,7989
RNN	25	0,1753	0,6874
	75	0,1415	0,7054
	100	0,1919	0,7192
	125	0,1812	0,674
PMC	25	0,5105	0,3254
	75	0,5103	0,3451
	100	0,5103	0,3187
	125	0,5098	0,4947

Sur la base des résultats du tableau 2.3, le modèle GRU a été sélectionné pour développer le modèle inverse de la machine. Ainsi, la figure 2.9 montre comment le modèle GRU est entraîné. La perte et la précision diminuent et augmentent respecti-

vement, jusqu'à ce que leurs valeurs convergent en fonction du nombre d'itérations.

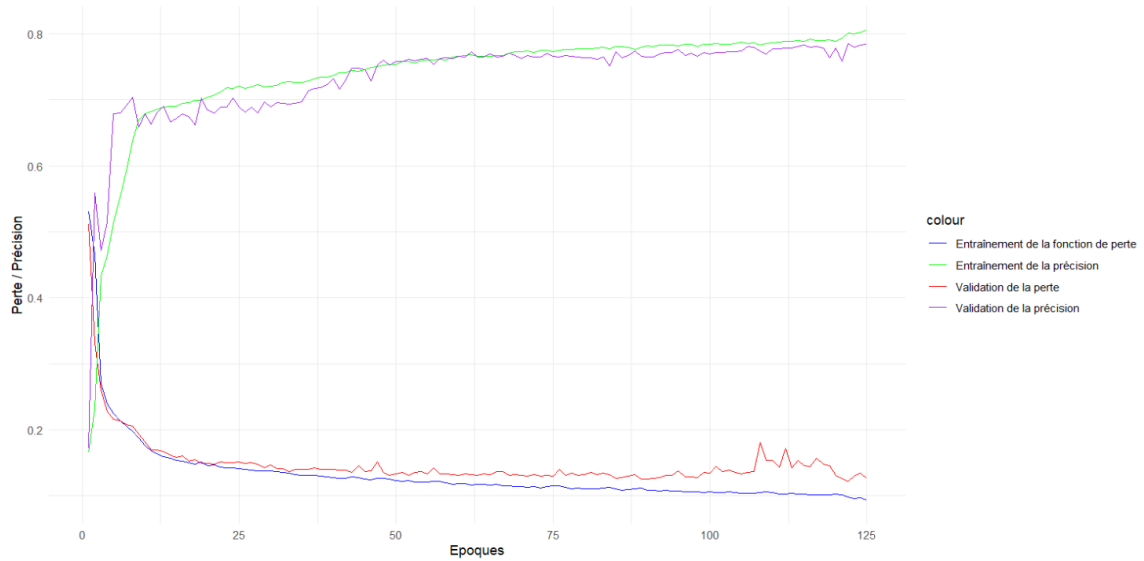
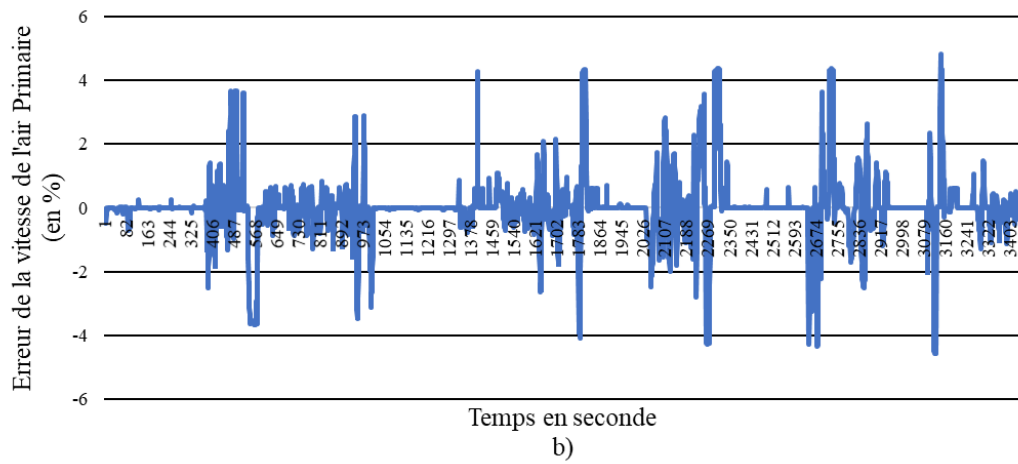
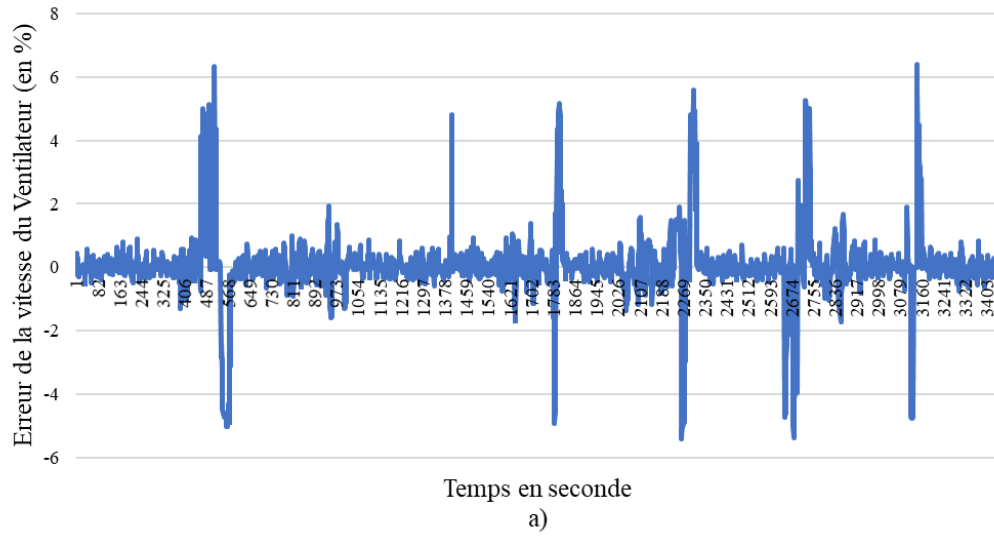


FIGURE 2.9 – Représentation de la perte et de la précision à l'aide de la méthode GRU.

Les courbes de la figure 2.10 montrent les pourcentages d'erreurs commises pour chaque sortie lors de la prédiction avec le modèle GRU. Elle se compose de quatre graphiques intitulés a), b), c) et d). Ceux-ci représentent la différence entre les prédictions et les valeurs réelles (expérimentales), pour chaque sortie du modèle GRU (ω_{fs} , ω_{pa} , ω_{sa} et ω_{ta}). Tout d'abord, dans la courbe a), les différences de vitesse du ventilateur vont de 6% à environ -5%. Ensuite, dans les courbes b) et c) respectivement, les vitesses de l'air primaire et secondaire présentent des différences plus faibles autour de 5% et -4,9%. Enfin, pour la courbe d), qui est la vitesse tertiaire, les différences sont comprises entre 4,8% et -4,9%. Dans le domaine de la commande des systèmes, une erreur de 5% à 10% correspond aux exigences standard. Nos résultats sont donc satisfaisants.



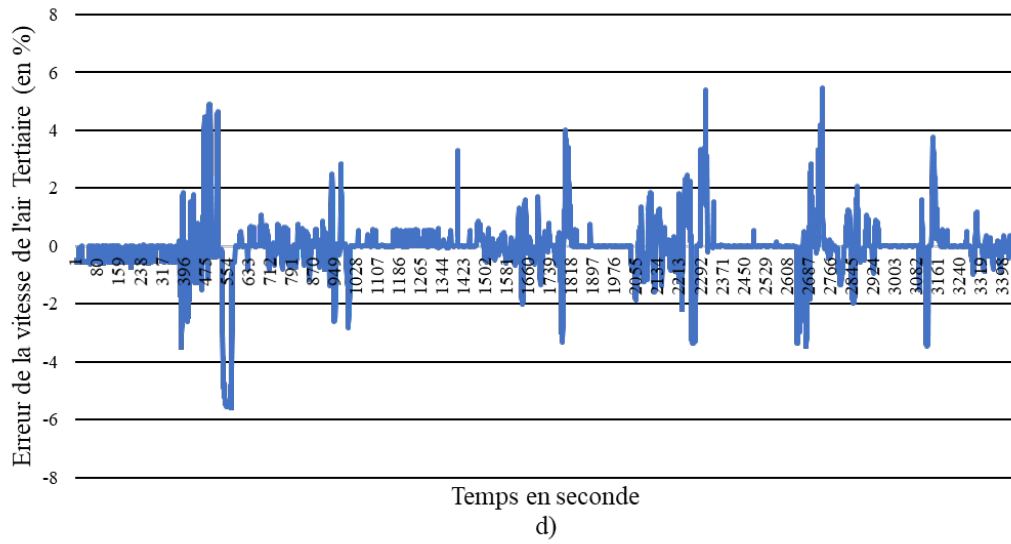
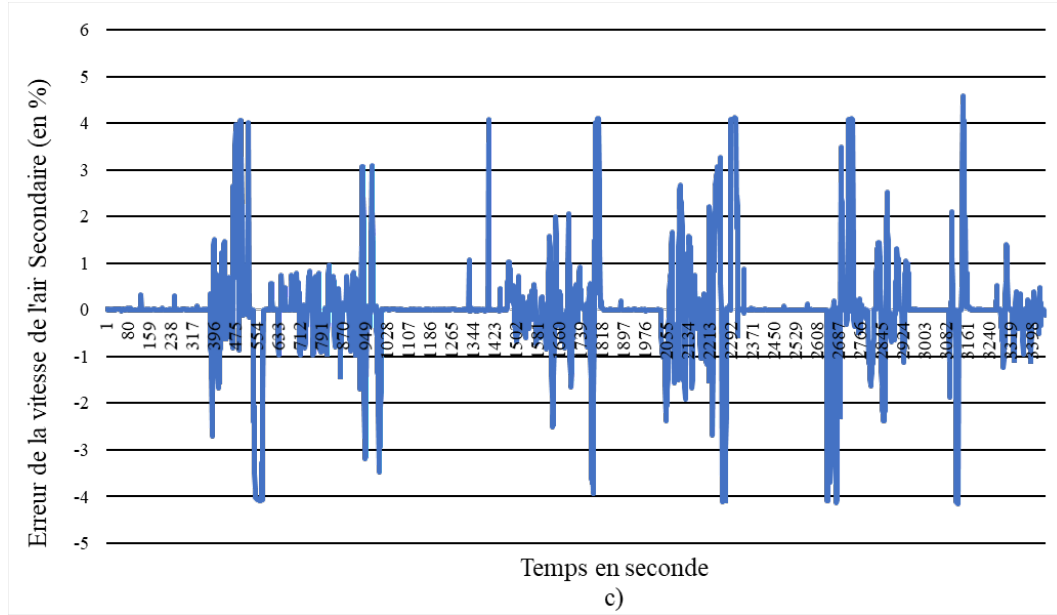


FIGURE 2.10 – Représentations a), b), c) et d) de la différence entre les prédictions et les valeurs réelles à l'aide de la méthode GRU.

2.5 Discussion des résultats

L'analyse des données révèle que les vitesses étaient toutes liées, comme l'illustre la figure 2.6. L'opérateur de l'usine régule les vitesses de la machine à biomasse pour produire plus de chaleur. Les associations sont fortes et positives (bleu foncé). Cela signifie que si la vitesse du ventilateur (ω_{fs}) est augmentée, les autres vitesses (ω_{pa} , ω_{sa} et ω_{ta}) doivent également être modifiées. Par conséquent, lors de la modification des réglages, l'opérateur doit tenir compte de ce lien entre les vitesses. Le fait que la température des gaz récupérés à la sortie (T_{og}), la pression de vapeur de la chaudière (P_{sig}) et le débit de vapeur de la chaudière (m_{sf}) soient tous positivement liés illustre qu'une valeur élevée de l'un d'entre eux est accompagnées par une valeur élevée des autres. Ces résultats sont essentiels au bon fonctionnement de la machine car ils sont directement liés aux résultats expérimentaux de la machine.

De même, une ACP a été réalisée sur les données recueillies. Cette méthode se concentre sur les variables les plus importantes qui ont un impact direct sur le système et le modèle [41]. Les premier et deuxième axes factoriels retiennent 83,45% de l'information globale contenue dans cette base de données (cf figure 2.7). Seule la teneur en oxygène (L_{o2}) a une faible contribution sur le premier axe selon le cercle de corrélation dans la figure 2.7. Cela signifie qu'elle n'est pas liée à la première composante principale qui est une combinaison linéaire des variables initiales qui capture la plus grande proportion de la variance totale dans les données. Autrement dit, si la contribution de L_{o2} à Dim 1 est faible, cela suggère que sa variation ne contribue pas significativement à la variance totale expliquée par le premier axe.

À des fins de modélisation, la modification des hyperparamètres a donné des résultats satisfaisants. En effet, les hyperparamètres du modèle doivent être pris en compte afin d'obtenir une convergence et une meilleure précision des valeurs réelles. Les facteurs architecturaux du modèle tels que le nombre de couches, les fonctions

d'activation pour traiter la non-linéarité des données, la fonction de perte appropriée pour le modèle, l'optimiseur et le taux d'apprentissage, et le nombre d'époques (itérations) sont des exemples d'hyperparamètres. Le tableau 2.3 montre les performances des modèles en termes d'époques, ainsi que la fonction perte, en l'occurrence l'erreur absolue moyenne (MAE). La perte pour le modèle LSTM augmente entre 25 et 100 itérations, passant à 19,48% à la 125e itération. Pour le GRU, la perte diminue au fur et à mesure que le nombre d'itérations augmente, atteignant 11,97% à la 125e itération. Le modèle RNN a montré une augmentation de la perte à mesure que le nombre d'itérations augmentait, mais le PMC a affiché une perte de 50,98% à la 125e itération. Les résultats du tableau 2.3 révèlent que le modèle GRU a obtenu les meilleurs résultats.

La figure 2.9 illustre l'entraînement du modèle GRU avec la fonction de perte et la superposition de graphiques de précision. Pour ce graphique, le nombre d'itérations est défini sur 125 afin d'obtenir une vue plus claire de la convergence des performances du modèle. Pour le graphique de la perte, la convergence commence autour de la 100e itération à environ 11%, tandis que pour le graphique de précision, la convergence commence vers 95 itérations, avec une précision d'environ 79%.

Les quatre graphiques (figure 2.10) de vitesse, qui sont les résultats du modèle inverse, montrent empiriquement que le GRU donne de bons résultats. Ce résultat est corroboré par les résultats de la différence entre les valeurs réelles et prédites en pourcentage affichés à la figure 2.10, car cette différence est inférieure à 6% pour les quatre vitesses, ce qui est un résultat assez satisfaisant.

Conclusion et perspectives

L'objectif de cette recherche est de développer un modèle inverse pour une machine à combustion à biomasse afin d'anticiper les intrants qui doivent être introduits pour atteindre les rendements requis. À l'aide d'une base de données expérimentale, des réseaux neuronaux tels que le perceptron multicouche (PMC), le réseau de neurones récurrent (RNN), les réseaux à mémoire à long et court terme (LSTM) et l'unité récurrente fermée (GRU) ont été utilisés, entraînés et évalués.

Après avoir analysé les données expérimentales, on a pu constater que les vitesses sont toutes positivement corrélées, tout comme la température du gaz de sortie, qui est positivement corrélée avec les vitesses d'entrée et la pression de vapeur de la chaudière. De plus, le modèle GRU a donné les meilleurs résultats parmi les trois approches restantes, avec la perte la plus faible et la précision la plus élevée, ce qui a conduit à sa sélection. Enfin, en utilisant le modèle GRU, l'écart entre les prévisions et les valeurs réelles était inférieur à 6 %. Pour toutes les variables prédites.

Ces résultats empiriques sont importants pour alimenter les techniques de stratégie de commande et d'optimisation des machines à biomasse, car le modèle proposé démontre la capacité à prévoir les valeurs d'entrée d'une machine à combustion à biomasse qui permettra une opération de sortie prédéfinie. La prochaine étape de cette recherche consisterait à déterminer les meilleures performances afin que le modèle proposé puisse offrir les entrées correctes pour que l'ensemble du système fonctionne

de manière optimale et de pouvoir justifier les résultats empiriques dans un cadre conceptuel formel robuste. Ce travail a ainsi donné naissance à un article de recherche qui est sous-évaluation (voir annexe).

Bibliographie

- [1] L. AL-LABADI et M. ZAREPOUR : Two-sample kolmogorov-smirnov test using a bayesian non parametric approach. *Mathematical Methods of Statistics*, 26:212–225, 2017.
- [2] I. ALI, F. GREIFENEDER, J. STAMENKOVIC, M. NEUMANN et C. NOTARNICOLA : Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data. *Remote Sensing*, 7(12):16398–16421, 2015.
- [3] G. BAUMBACH, H. HARTMANN, I. HÖFER, H. HOFBAUER, T. HÜLSMANN, M. KALTSCHMITT, V. LENZ, U. NEULING, T. NUSSBAUMER et I. OBERNBERGER : Grundlagen der thermo-chemischen umwandlung biogener festbrennstoffe. pages 579–814, 2016.
- [4] M. BEHNAME : La consommation d’énergie renouvelable et la croissance économique dans l’europe de l’ouest. *Romanian Journal of Economics*, 2:160–171, 2012.
- [5] D. BERRAR : Cross-validation. *Encyclopedia of Bioinformatics and Computational Biology*, 1(April):542–545, 2019.
- [6] M. BODEN : A guide to recurrent neural networks and backpropagation. *the Dallas project*, 2(2):1–10, 2002.
- [7] M. DAKHMOUCHE et F. B. BOURBIA : Utilisation de l’énergie renouvelable pour un musee scientifique autonome. 2020.
- [8] B. DAVIRON : *Biomasse*. éditions Quae, 2019.

- [9] O. DE JESUS et M. T. HAGAN : Backpropagation algorithms for a broad class of dynamic networks. *IEEE Transactions on Neural Networks*, 18(1):14–27, 2007.
- [10] L. DIÉMÉ : Tests d’adéquation des copules bivariées et applications. 2022.
- [11] F. EYBEN, M. WÖLLMER, B. SCHULLER et A. GRAVES : From speech to letters—using a novel neural network architecture for grapheme based asr. *In 2009 IEEE Workshop on Automatic Speech Recognition & Understanding*, pages 376–380. IEEE, 2009.
- [12] M. FORSYTH, T. CHIMDI, A. SEEBER, D. GUNZELMANN et P. C. HOWLETT : Structure and dynamics in an organic ionic plastic crystal, n-ethyl-n-methyl pyrrolidinium bis (trifluoromethanesulfonyl) amide, mixed with a sodium salt. *Journal of Materials Chemistry A*, 2(11):3993–4003, 2014.
- [13] A. GRAVES, A. MOHAMED et G. HINTON : Speech recognition with deep recurrent neural networks. pages 6645–6649, 2013.
- [14] G. G. S. H. U. V. HUV : View metadata, citation and similar papers at core.ac.uk. *Journal of vascular surgery*, 43(2), 2020.
- [15] R. J. HYNDMAN : Measuring forecast accuracy. *Business forecasting : Practical problems and solutions*, pages 177–183, 2014.
- [16] A. JAWOREK, A. SOBCZYK, A. MARCHEWICZ, A. KRUPA et T. CZECH : Particulate matter emission control from small residential boilers after biomass combustion. a review. *Renewable and Sustainable Energy Reviews*, 137:110446, 2021.
- [17] B. KHALIDA, Z. MOHAMED, S. BELAID, H. O. SAMIR, K. SOBHI et S. MIDANE : Prediction of higher heating value hhv of date palm biomass fuel using artificial intelligence method. *In 2019 8th International Conference on Renewable Energy Research and Applications (ICRERA)*, pages 59–62, 2019.
- [18] M. KIM, M. D. SINGH et M. LEE : Towards abstraction from extraction : Multiple timescale gated recurrent unit for summarization. *arXiv preprint arXiv :1607.00718*, 2016.
- [19] J. KOPPEJAN et S. VAN LOO : The handbook of biomass combustion and co-firing. 2012.

- [20] T. M. KORPELA, T. K. BJÖRKQVIST et P. A. LAUTALA : Control strategy for small-scale wood chip combustion. *IFAC Proceedings Volumes*, 42(9):119–124, 2009.
- [21] J. KORPELA et S. JÄMSÄ-JOUNELA : Model predictive control for biopower combined heat and power (chp) plant. 31:435–439, 2012.
- [22] T. KURITA : Principal component analysis (pca). *Computer Vision : A Reference Guide*, pages 1–4, 2019.
- [23] L. R. MEDSKER et L. JAIN : Recurrent neural networks. *Design and Applications*, 5:64–67, 2001.
- [24] S. R. NAQVI, Z. ULLAH, S. A. A. TAQVI, Muhammad N. A. KHAN, W. FAROOQ, M. T. MEHRAN, D. JUCHELKOVÁ et L. ŠTĚPANEC : Applications of machine learning in thermochemical conversion of biomass-a review. *Fuel*, 332:126055, 2023.
- [25] C. D. NGUYEN, J. B. CARLIN et K. J. LEE : Diagnosing problems with imputation models using the kolmogorov-smirnov test : a simulation study. *BMC medical research methodology*, 13:1–9, 2013.
- [26] H. K. NIELSEN, T. S. VEHUS, S. RUDRA et N. R. KRISTIANSEN : Biomass production, storage, and pretreatment for the production of solid biofuels. pages 339–380, 2022.
- [27] Z. NIU, G. ZHONG, G. YUE, L. WANG, H. YU, X. LING et J. DONG : Recurrent attention unit : A new gated recurrent unit for long-term memory of important parts in sequential data. *Neurocomputing*, 517:1–9, 2023.
- [28] M. NOURI, D. TRENTESAUX et A. BEKRAR : Easysched : une architecture multi-agent pour l’ordonnancement prédictif et réactif de systèmes de production de biens en fonction de l’énergie renouvelable disponible dans un contexte industrie 4.0. *Génie industriel et productique*, 2(1), 2019.
- [29] N. PACES, A. VOIGT, S. JAKUBEK, A. SCHIRRER et M. KOZEK : Combined control of combustion load and combustion position in a moving grate biomass furnace. pages 1447–1452, 2011.

- [30] M. PARIZEAU : Réseaux de neurones. *GIF-21140 et GIF-64326*, 124, 2006.
- [31] V. PLAČEK et J. HRDLIČKA : Influence of control on environmental and economical aspects of small-scale biomass boiler. pages 1–3, 2011.
- [32] V. PLACEK, B. ŠULC, S. VRÁNA, J. HRDLIČKA et J. PITELL : Investigation in control of small-scale biomass boilers. pages 308–311, 2011.
- [33] D. C. PLAUT *et al.* : Experiments on learning by back propagation. 1986.
- [34] R. RAKOTOMALALA : Analyse de corrélation. *Cours statistique à l'université de lumière Lyon*, 2:89, 2015.
- [35] D. S. RETSCHITZEGGER, D. F. J. KELZ, D. M. HITZL, V. FRANCESCATO et C. HOCHENAUER : Evaluation of the combustion characteristics of potentially new biomass fuels such as htc pellets and vineyard pruning pellets within the eu-project brisk. *In World Sustainable Energy Days 2016 : European Pellet Conference*, 2016.
- [36] M. RIEDMILLER et A. LERNEN : Multi layer perceptron. *Machine Learning Lab Special Lecture, University of Freiburg*, pages 7–24, 2014.
- [37] Y. ROGAUME : La combustion du bois et de la biomasse. *Pollution atmosphérique*, 65, 2009.
- [38] L. M. ROMEO et R. GARETA : Fouling control in biomass boilers. *Biomass and bioenergy*, 33(5):854–861, 2009.
- [39] D. E. RUMELHART, G. E. HINTON et R. J. WILLIAMS : Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.
- [40] H. SAK, A. W. SENIOR et F. BEAUFAYS : Long short-term memory recurrent neural network architectures for large scale acoustic modeling. pages 338–342, 2014.
- [41] B. SALEM : Principal component analysis (pca). *La Tunisie Médicale*, 99(4):383–389, 2021.
- [42] C. SCHÖRGHUBER, M. GÖLLES, M. REICHHARTINGER et M. HORN : Control of biomass grate boilers using internal model control. *Control Engineering Practice*, 96:104274, 2020.

- [43] R. SEEBER, M. GÖLLES, N. DOURDOUMAS et M. HORN : Reference shaping for model-based control of biomass grate boilers. *Control engineering practice*, 82:173–184, 2019.
- [44] A. SHAFIZADEH, H. SHAHBEIG, M. H. NADIAN, H. MOBILI, M. DOWLATI, V. K. GUPTA, W. PENG, S. S. LAM, M. TABATABAEI et M. AGHBASHLO : Machine learning predicts and optimizes hydrothermal liquefaction of biomass. *Chemical Engineering Journal*, 445:136579, 2022.
- [45] A. J. SHEPHERD : Second-order methods for neural networks : Fast and reliable training methods for multi-layer perceptrons. 2012.
- [46] T. STÉPHANE : Data mining et statistique décisionnelle : l’intelligence des données. 2012.
- [47] M. SUNDERMEYER, R. SCHLÜTER et H. NEY : Lstm neural networks for language modeling. 2012.
- [48] S. VALIZADEH, H. HAKIMIAN, A. FAROOQ, B. JEON, W. CHEN, S. H. LEE, S. JUNG, M. W. SEO et Y. PARK : Valorization of biomass through gasification for green hydrogen generation : A comprehensive review. *Bioresource technology*, page 128143, 2022.
- [49] S. VRÁNA, V. PLAČEK, C. OSWALD, B. ŠULC et P. NEUMAN : Neural network evaluation of combustion process for continuous control of small scale biomass fired boilers. *IFAC Proceedings Volumes*, 47(3):1440–1445, 2014.
- [50] N. WAGNER : Detection of changes in circadian organisation of animal activities related to pre-pathological states, stress, or reproduction events. 2020.
- [51] W. WANG et Y. LU : Analysis of the mean absolute error (mae) and the root mean square error (rmse) in assessing rounding model. 324(1):012049, 2018.
- [52] Y. WANG, W. LIAO et Y. CHANG : Gated recurrent unit network-based short-term photovoltaic forecasting. *Energies*, 11(8):2163, 2018.
- [53] E. YAZAN et M. F. TALU : Comparison of the stochastic gradient descent based optimization techniques. In *2017 International Artificial Intelligence and Data Processing Symposium (IDAP)*, pages 1–5. IEEE, 2017.

- [54] W. ZHANG, Q. CHEN, J. CHEN, D. XU, H. ZHAN, H. PENG, J. PAN, M. VLASKIN, L. LENG et H. LI : Machine learning for hydrothermal treatment of biomass : a review. *Bioresource Technology*, page 128547, 2022.

Article sous évaluation et le traitement des données

Article sous évaluation

Titre de l'article : Experimental data analysis and ANN implementation for inverse dynamics-based prediction of biomass combustion machine operation.

Auteurs : Mouhamadou Mountakha Ndao ; Nadia Ghazzali ; Nadjjet Zioui

Soumis aux revues : *Springer Journal et Neural Computing and Applications (NCAA)*

Traitement des données

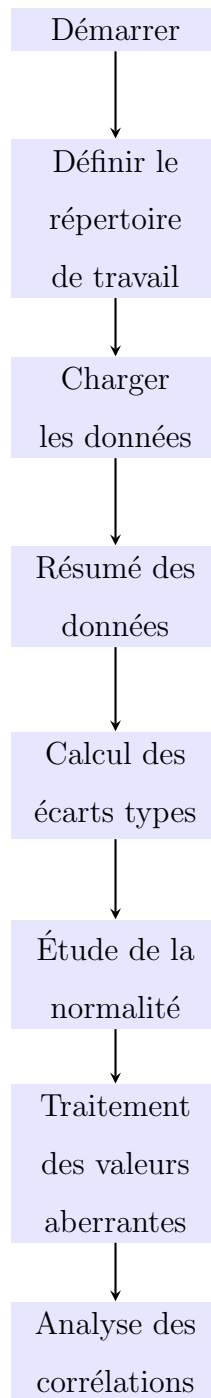


FIGURE 11 – Partie 1 du workflow du traitement des données

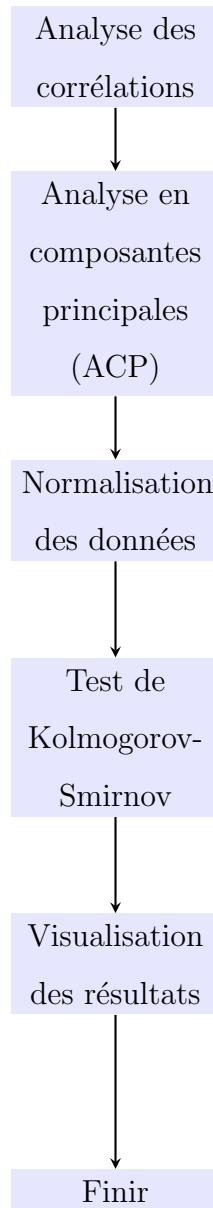


FIGURE 12 – Partie 2 du workflow du traitement des données

Explications des étapes

1. **Définir le répertoire de travail** : Utilisation de `'setwd("C :/memoire/Projet de recherche/codes R")'` pour définir le répertoire de travail où se trouvent les fichiers de données.

2. **Charger les données** : Chargement des données avec `'read_excel("Gdonnées_Projet_Clermond_Hamel.xlsx")'`.
3. **Résumé des données** : Utilisation de `'summary(data2)'` pour obtenir un résumé statistique des données.
4. **Calcul des écarts types** : Calcul des écarts types des variables avec `'apply(data2[,3 :11], 2, sd)'`.
5. **Étude de la normalité** : Utilisation d'histogrammes et de courbes de densité pour évaluer la normalité des données.
6. **Traitement des valeurs aberrantes** : Identification et traitement des valeurs aberrantes en utilisant des critères basés sur l'écart type.
7. **Analyse des corrélations** : Calcul et visualisation des corrélations entre les variables avec `'cor(data2t)'` et `'corrplot(mcor)'`.
8. **Analyse en composantes principales (ACP)** : Réalisation d'une ACP avec `'PCA(data2[,3 :11])'` et exploration des charges des variables.
9. **Normalisation des données** : Normalisation des données pour des analyses ultérieures avec `'scale(data2t)'`.
10. **Test de Kolmogorov-Smirnov** : Test de normalité des données avec `'ks.test(data, "pnorm")'`.
11. **Visualisation des résultats** : Utilisation de `'ggplot2'` pour créer des graphiques qui comparent les distributions empiriques et théoriques, facilitant l'interprétation visuelle des résultats.

Algorithme et Workflow du Codage de la modélisation des données

Algorithme du modèle GRU

Initialisation du Modèle :

- Créer un réseau de neurones séquentiel

Définition des Couches du Modèle :

- Ajouter une couche GRU :
 - Nombre d'unités = 64
 - Forme d'entrée = (9, 5)
- Ajouter une couche dense :
 - Nombre d'unités = 4

Compilation du Modèle :

- Compiler le modèle :
 - Fonction de perte = "mae" (Mean Absolute Error)
 - Optimiseur = Adam avec taux d'apprentissage de 0.001
 - Métriques = "accuracy" (précision)

Entraîner le modèle :

- Données d'entraînement = train_seq
- Époques = Exemple 125

— Données de validation = test_seq

Workflow du processus de la modélisation

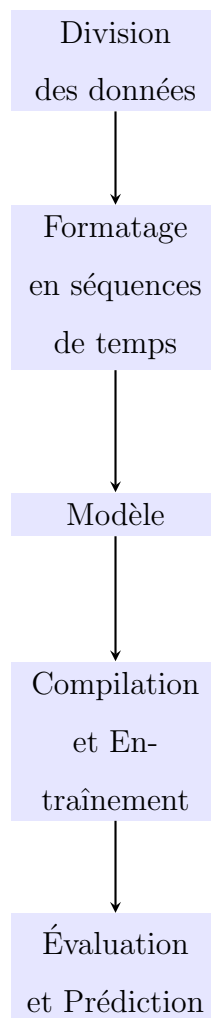


FIGURE 13 – Workflow de la modélisation

Explications des étapes

1. **Division des données en ensembles d'entraînement et de test** : Un échantillon aléatoire des données est sélectionné pour l'ensemble d'entraînement, tandis que le reste est attribué à l'ensemble de test.

2. **Formatage des données en séquences de temps** : Les données sont transformées en séquences de temps pour être utilisées dans les modèles de séries chronologiques.

3. **Création du modèle** : Chaque modèle est défini, comprenant son propre architecture avec une couche de dropout et une couche dense de sortie.

4. **Compilation et entraînement des modèles** : Chaque modèle est compilé avec une fonction de perte MAE et optimisé avec l'optimiseur Adam avec l'accuracy comme métrique.

Entraînement du modèle sur les données d'entraînement avec validation sur les données de test.

5. **Évaluation et prédiction avec chaque modèle** : Le modèle est évalué sur les données de test pour calculer la perte et l'accuracy.

Les valeurs prédites sont générées pour les données de test et enregistrées dans un fichier Excel.