

وزارة التعليم العالي والبحث العلمي



**BADJI MOKHTAR –ANNABA
UNIVERSITY
UNIVERSITE BADJI MOKHTAR–
ANNABA**

**جامعة باجي مختار
- عنابة**

Faculté des Sciences

Année : 2023/2024

**Département de Mathématiques
Laboratoire de Modélisation Mathématique et Simulation Numérique**



THÈSE

Présentée en vue de l'obtention du diplôme de doctorat

MODÉLISATION MATHÉMATIQUE ET RÉOLUTION NUMÉRIQUE DU SYSTÈME DYNAMIQUE : APPLICATION EN INSTRUMENTATION ET CONTRÔLE

Filière

Mathématiques Appliquées

Option

Équations Différentielles et Applications

Par

AZZEDINE Abdel Hakim

DIRECTEUR DE THÈSE : NOURI Fatma Zohra Prof. U.B.M. ANNABA

CO-DIRECTEUR DE THÈSE : BOUHOUCHE Salah DR. CRTI. ALGER

Devant le jury

PRESIDENT : MAZOUZI Said Prof. U.B.M. ANNABA

EXAMINATEUR : MERZOUGUI Mouna Prof. U.B.M. ANNABA

EXAMINATEUR : KIMOUCHE Karima MCA UNIV. SKIKDA

Remerciements

Grand merci à **Allah** le tout puissant de m'avoir donné le courage, la patience, et la santé, ainsi que l'audace pour dépasser toutes les difficultés et d'achever ce modeste travail.

Je voudrais dans un premier temps remercier, ma directrice de thèse **Mme Nouri Fatma Zohra**, professeur à l'université de Annaba. qui m'a fait l'honneur de me proposer un sujet de thèse d'actualité, aussi pour sa patience, sa disponibilité, ses suggestions, et surtout ses judicieux conseils qui ont contribué pour une bonne conduite de cette thèse de recherche, et de ma formation doctorale, j'ai énormément appris.

Mes remerciements vont également à **M. Bouhouche Salah**, directeur de recherche à l'unité de Recherche en Technologies Industrielles CRTI, d'avoir accepté de co-encadrer ce travail. Aussi pour le temps qu'il a consacré pour mettre à ma disposition les outils indispensables de domaine du data mining ; ainsi que pour son aide à obtenir des données réelles.

Je tiens à remercier **M. Mazouzi Said**, professeur à l'université de Annaba, pour l'honneur qu'il m'a fait en acceptant de présider le jury de cette thèse.

J'adresse également mes sincères remerciements à **Mme Merzougui Mouna**, professeur à l'université de Annaba, ainsi que **Mme Kimouche Karima**, Maitre de Conférences A à l'université de Skikda, qu'elles m'ont fait le grand honneur pour leurs participation à ce jury en qualité d'examinatrices, et pour le temps consacré à la lecture de mon travail.

Je tiens à remercier tous les enseignants du département de mathématiques, en particulier **Pr. Hadji Mohamed Lakhder** et **Pr. Mazouzi Said** qui m'ont fourni les outils nécessaire à la réussite de mes études universitaires, et pour m'avoir donné les bases de mon savoir mathématique et qui ont partagé avec moi leurs connaissances.

Tous les membres du laboratoire **Modélisation Mathématique et Simulation Numérique LAM²SIN**, toute fonction confondue, trouvent ici l'expression de mes vifs remerciements pour leur accueil et aide multiforme inoubliable.

Enfin ces remerciements ne seraient pas complets sans mentionner **mes parents, mon frère et ma soeur**. Je les remercie pour leur patience, leur soutien moral et matériel, leurs encouragements qu'ils ont pu m'offrir en leur promettant que le meilleur reste à venir.

المخلص

في السنوات الأخيرة، تم تطوير تقنيات استخراج البيانات في العديد من مجالات التطبيق بهدف استخراج المعلومات الإحصائية الرئيسية من البيانات الكبيرة والمعقدة. تُستخدم هذه المعلومات لوصف العلاقات التي قد توجد بين البيانات المختلفة (التنبؤ)؛ تتيح التقنيات الأخرى تجميع البيانات بطريقة تجعلها تبدو واضحة (التجميع والتصنيف). هذا العمل هو مساهمة في النمذجة الرياضية للأنظمة المعقدة بهذه الطرق المختلفة.

تتطلب نمذجة الأنظمة المعقدة تحليل التفاعلات المرتبطة بالبيانات التحليلية واختيار الهيكل، فنحن مهتمون بالبحث عن نماذج معقدة بين معلمات التشغيل لهذه الأنظمة، وبالتالي تم دراسة خوارزميات التعلم الخاضعة للإشراف وغير الخاضعة للإشراف وتنفيذها عدديًا باستخدام نهج تكيفي للخوارزميات الحالية لمراعاة الاختلافات المحلية في المعلمات البيئية. سيتم إجراء التحقق من صحة الاستخدام وحدوده من خلال البيانات الحقيقية التي تم جمعها من معمل معتمد لمشكلة عمليات صناعة الحديد والصلب وكذلك استخدام بيانات تتعلق بالصناعات الزراعية.

الكلمات المفتاحية:

النمذجة الرياضية، تحليل البيانات تحت إشراف (NN، SVR، RLS)، التعلم الاصطناعي غير الخاضع للإشراف (PCA، PLS)، الاحتمالية القصوى المتبقية (REML).

Résumé

Ces dernières années, les méthodes d'exploration de données ont été développées dans de nombreux domaines d'application dans le but de tirer les principales informations statistiques des données volumineuses et complexes. Ces informations permettent de décrire et caractériser les relations qui peuvent exister entre plusieurs données (Prédiction); certaines approches peuvent être utilisées pour regrouper les données de manière à faire apparaître claire (clustering et classification). Ce travail est une contribution à la modélisation mathématique des systèmes complexes par ces différentes méthodes. La modélisation des systèmes complexes nécessite une analyse des interactions associées aux données analytiques et la sélection de structure, nous nous sommes intéressés à la recherche de modèles complexes entre les paramètres de fonctionnement de ces systèmes. Les algorithmes d'apprentissage supervisés et non supervisés seront ainsi étudiés et implémentés numériquement avec une approche adaptative des algorithmes existants pour tenir compte des variations locales des paramètres du milieu. La validation et les limites d'usage seront effectuées par des données réelles collectées des processus sidérurgiques typiques (Haut fourneau). Ainsi qu'en utilisant les données de contrôle de qualité d'un laboratoire accrédité pour le problème agro-industriel.

Mots-clés :

Modélisation Mathématique, Analyse de données, Apprentissage supervisé (NN, SVR, RLS) et non supervisé (PCA, PLS), Vraisemblance maximum (REML).

Abstract

In recent years, data mining techniques have been developed in many application areas with the aim of extracting key statistical information from large and complex data. This information allows to describe the relationships that can exist between different data (prediction) ; other techniques allow to group data in order to make them clear (clustering and classification). This work is a contribution to the mathematical modeling of complex systems by these different methods. The modeling of complex systems requires an analysis of the interactions associated with the analytical data and the selection of structure, we are interested in the search for complex models between the parameters of operation of these systems. Supervised and unsupervised learning algorithms will be studied and implemented numerically with an adaptive approach to existing algorithms to take into account local variations in the parameters of the environment. The validation and the limits of use will be carried out by real data collected from typical iron and steel processes (blast furnace). As well as using quality control data from an accredited laboratory for the agro-industrial problem.

Keywords : Mathematical modeling, Data analysis, Artificial learning supervised (NN, SVR, RLS) and unsupervised (PCA, PLS), Residual Estimation Maximum Likelihood (REML).

Table des matières

Table des figures	vii
Liste des tableaux	viii
Introduction générale	ix
1 Préliminaires et notions fondamentales	1
1.1 Qu'est que l'exploration des données	1
1.2 Étapes d'un projet d'exploration des données	2
1.3 Domaines d'application	3
1.4 Tâches d'exploration des données	3
1.5 Différents types d'apprentissage automatique	4
1.5.1 Caractérisation des méthodes : apprentissage supervisé	4
1.5.2 Apprentissage non-supervisé	5
1.5.3 Caractérisation des méthodes : apprentissage semi-supervisé	6
1.5.4 Caractérisation des méthodes : apprentissage par renforcement	6
1.6 Définitions et concepts	6
1.6.1 Définitions ISO 5725	6
1.6.2 Répétabilité	7
1.6.3 Reproductibilité	7
1.6.4 Fidélité	7
1.7 Termes métrologiques généraux	8
1.8 Techniques d'évaluation d'incertitude	9
1.9 Facteurs de l'importance des incertitudes	10
1.10 Termes et concepts statistiques fondamentaux	10
2 Techniques d'exploration des données	12
2.1 Réseau de neurones artificiels	12
2.1.1 Historique	12
2.1.2 Définition et principe des neurones artificiels	12
2.1.3 Types de fonctions d'activation	14

2.1.4 Réseau de neurones artificiels de type fonction à base radiale (RBANN)	15
2.1.5 Lissage exponentiel combiné à RBANN(ES-RBANN)	16
2.1.6 Algorithmes	16
2.1.7 Apprentissage profond (Deep learning)	17
2.2 Méthode des moindres carrés récursive(RLS)	18
2.3 Méthode de Moindres carrés partiels (PLS)	21
2.4 Méthode de séparateur à vaste marge (SVR)	24
2.5 Techniques adaptatives	26
2.6 Méthode de vraisemblance résiduelle maximum	27
2.6.1 Vraisemblance maximum de la composante de variance dans le modèle mixte général	28
2.6.2 Estimation de vraisemblance résiduelle maximum (REML)	29
2.7 Distribution de l'évaluation de la fiabilité	30
2.7.1 Distribution normale	30
2.7.2 Probabilité de conformité et niveaux de risque associés	31
3 Résultats mathématiques et numériques pour le contrôle qualité du métal chaud dans le haut fourneau	33
3.1 Introduction	33
3.2 Modélisation et identification	34
3.3 Application au contrôle de la qualité du métal chaud	35
3.3.1 Expérimentation numérique	37
3.4 Application de l'analyse multivariée des risques	41
3.5 Expérimentation numérique	41
3.5.1 Résultats et discussion	42
3.6 Conclusion	45
4 Méthode d'analyse de variance pour les données de contrôle qualité	47
4.1 Introduction	47
4.1.1 Quelques approches mathématiques	47
4.2 Application au contrôle qualité en agro-industrie	49
4.2.1 Expérimentation numérique	49
4.3 Conclusion	54
5 Modèles multivariés et méthodes de mesure	55
5.1 Introduction	55
5.2 Mesure et analyse des matériaux par spectrométrie	56
5.2.1 Principe de mesure	56
5.3 Application au contrôle de qualité des matériaux	57
5.3.1 Résultats et discussion	60
5.4 Conclusion	66

6 Conclusion générale et Perspectives	68
--	-----------

Bibliographie	70
----------------------	-----------

Table des figures

1.1 Exploration des données comme confluent de plusieurs disciplines . . .	2
1.2 Schéma des différents notions de base	8
3.1 Entrées et sorties expérimentales des modèles	37
3.2 Prédiction de la température de la fonte et des erreurs	38
3.3 Résultats calculés et prédits pour des données stables	39
3.4 Modèle des entrées expérimentales	42
3.5 Risques des différents paramètres	43
3.6 Risques globaux additifs et multiplicatifs	43
3.7 Modélisation et prévision du risque et des erreurs de température de la fonte	44
4.1 Modèle des entrées et sorties expérimentales	50
4.2 Résultats de prédiction et erreur de modélisation de la méthode RLS	51
4.3 Résultats de prédiction et erreur de modélisation des méthodes REML et REML adaptative	51
4.4 Valeurs réelles et calculées du gonflement de la pâte	52
5.1 Principe de l'analyse d'échantillons métalliques par fluorescence X . .	57
5.2 E.S.U : structure du modèle sans interaction entre les concentrations	58
5.3 E.S.M : structure du modèle avec interaction entre les concentrations	59
5.4 Entrées et sorties expérimentales	59
5.5 Prédiction de la concentration de carbone et erreurs	60
5.6 Prédiction de la concentration en silicium et erreurs	61
5.7 Prédiction de la concentration en manganèse et erreurs	61
5.8 Prédiction de la concentration de soufre et erreurs	62
5.9 Prédiction de la concentration de carbone et erreurs (test)	63
5.10 Prédiction de la concentration de silicium et des erreurs (Test)	64
5.11 Prédiction de la concentration de manganèse et des erreurs (Test) . .	64
5.12 Prédiction de la concentration en soufre et des erreurs (Test)	65

Liste des tableaux

2.1 Fonctions de transfert	15
3.1 Nomenclature des données d'entrée et de sortie	36
3.2 Description statistique des données collectées	37
3.3 Matrice de corrélation	39
3.4 Indicateurs de performances des différentes méthodes	40
3.5 Statistiques descriptives des échantillons collectés	42
3.6 Matrice de corrélation	43
3.7 Décision de conformité et de non-conformité	44
3.8 Indicateurs de performance des méthodes RBANN et ES-RBFANN	44
4.1 Nomenclature des données d'entrée et de sortie	49
4.2 Descriptions statistiques des échantillons collectés	50
4.3 Indicateurs de performance statistique des différentes méthodes	52
5.1 Nomenclature des données d'entrée et de sortie	59
5.2 Statistiques descriptives des échantillons collectés	60
5.3 Indicateurs statistiques de performance des différentes méthodes dans l'étape de modélisation	63
5.4 Indicateurs statistiques de performance des différentes méthodes dans l'étape de prédiction	65

Introduction

Au cours des deux dernières décennies la quantité de données augmente de manière exponentielle connu sous l'appellation « Big Data ». Elles sont généralement d'une nature particulière et différente, caractérisant les actions multi-échelles du système considéré. Bien que la quantité d'informations stockées soit importante, il existe de nombreuses caractéristiques associées qui ont aussi la même importance, comme l'exactitude, la variété (format, nature et type) et la valeur (perspectives et impacts) de ces données. De plus, malgré l'absence de bruit, les valeurs aberrantes, manquantes et incorrectes avec la rareté des données pourraient introduire du bruit. L'analyse du Big Data est confrontée à de nombreux défis ainsi qu'aux problèmes importants, tels que la qualité et la vérification des données, les dimensions élevées, la diffusion et la représentation de leurs ressources, leurs visualisations et leurs tests, et la capacité de développer des algorithmes appropriés. Afin de déduire des modèles à partir de données à grande échelle et de comprendre la valeur de l'analyse des big data, nous devons appliquer et adapter certaines méthodes, telles que les paradigmes et algorithmes d'apprentissage automatique et l'apprentissage profond. L'apprentissage automatique est un domaine important et une frontière de recherche dans l'intelligence artificielle, qui ont une partie significative dans l'analyse des grandes données longtemps avec la puissance de calcul. Sa tâche est d'analyser la plus grande quantité de données à tous les niveaux, qu'ils soient simples ou complexes, et de veiller à ce que des résultats plus précis et des décisions soient obtenus plus rapidement. On peut ce domaine classer en deux catégories de modèles : La première est appelée apprentissage supervisé, où pour ce type de données d'entrée, il est généralement nécessaire de former des représentations de données fiables qui peuvent être étendues à de nouvelles données dans le même domaine d'applications big data. Il peut être classé en deux groupes principaux, la régression et la régression logistique. Le réseau de neurone artificiel (NN) et les machines à vecteurs de support (SVM) sont considérés comme des méthodes de classification permettant d'obtenir la meilleure prédiction ou de découvrir certains modèles linéaires ou non linéaires. Pour cela, il suffit de modifier les paramètres de structure, tels que le noyau pour le SVM, la fonction d'activation et le nombre de couches dans le NN. Différents algorithmes basés sur des algorithmes robustes, allant de la famille des gradients à Levenberg Marquardt, sont utilisés pour déterminer le nombre de couches d'un réseau. Ils sont

utilisés pour minimiser les erreurs de modélisation pour ces types de modèles. Le second type d'apprentissage est appelé apprentissage non supervisé, dans lequel il n'y a pas d'erreur de modélisation à superviser ni d'algorithme d'apprentissage direct pour suivre la sortie d'un modèle. Les moindres carrés partiels (PLS), l'analyse en composantes principales (ACP), la réduction de la dimension et le regroupement, appartiennent à l'apprentissage non supervisé. En général, nous utilisons ces méthodes pour extraire les caractéristiques des données bruitées, comme une étape de pré-traitement. Les données pré-traitées sont appliquées comme entrées pour l'étape d'apprentissage supervisé. Les techniques d'apprentissage automatique ont été de plus en plus appliquées au contrôle et à l'identification des systèmes complexes. Ces derniers se caractérisent par leurs interactions non linéaires, leurs grande dimension, et l'imprécision des modèles mathématiques. Les industries sidérurgiques sont caractérisées par des phénomènes complexes en particulier là où il y a des changements de phase considérables, comme la transformation liquide-solide. Au cours de ce processus, des réactions complexes qui dépendent des matières premières et des paramètres de production. Ces réactions définissent la qualité finale du lingot et des brames. En fait, de nombreux travaux ont été effectués pour obtenir une production sans défauts avec des coûts de production minimales (voir [4] ; [77]). Pour réaliser cela, un système d'assurance qualité basé sur la modélisation avancée a été développé et appliqué dans différentes industries sidérurgiques (voir par exemple [65] ; [90]). L'optimisation d'un grand système, tel qu'une aciérie, repose sur la division du système global en différents sous-systèmes. Ce travail fait notamment l'objet d'une étude approfondie de plusieurs approches de l'apprentissage automatique et propose une contribution à l'amélioration de certains aspects de ces techniques, en utilisant les algorithmes des fenêtres glissantes. Le travail de cette thèse est réparti comme suit :

Chapitre 1 : est consacré à la définition de domaine d'exploration des données, et toutes les notions fondamentales liées à ce domaine, précisant une bonne motivation et l'importance de ce travail dans la recherche.

Chapitre 2 : est consacré à l'état d'art des nombreuses techniques d'optimisation, et leur utilisation principale pour la modélisation des systèmes complexes ; ainsi que les notions mathématiques de bases utiles dans notre étude.

Chapitre 3 : est consacré à décrire le problème de l'industrie où on présente la structure générale des hauts fourneaux et leur principe de fonctionnement en appliquant certaines méthodes d'intelligence artificielle pour réaliser cette étude. Afin d'identifier les modèles et prédire les paramètres cruciaux pour améliorer la qualité de la fonte avec un coût raisonnable ; l'optimisation et le développement des aspects de contrôle des principaux procédés sidérurgiques ; estimation des risques multivariées ; pour aboutir à la validation de modèles à partir de données industrielles brutes acquises auprès de la sidérurgie Sider El Hadjar.

Chapitre 4 : est consacré au problème de l'agro-industrie où on présente la structure générale du processus, et on applique certaines techniques, en utilisant les données de contrôle de qualité d'un laboratoire accrédité pour l'agro-industrie

(CDAAI-Constantine-Algérie). L'objectif de ce travail est de développer et de mettre en oeuvre un modèle mixte basé sur l'algorithme des moindres carrés récurrents (RLS), ainsi que sur l'estimation du maximum de vraisemblance résiduelle (REML) et une nouvelle forme adaptative de la technique REML. Cela sera réalisé en utilisant des données de contrôle de qualité provenant d'un laboratoire accrédité. La nouveauté consiste à utiliser une stratégie de fenêtre glissante, basée sur une technique de calcul prenant en compte un sous-ensemble d'éléments de la série temporelle au lieu de la totalité des données.

Chapitre 5 : est consacré au développement et la mise en oeuvre d'une évaluation réelle de la méthode analytique basée sur la technologie des rayons X, en utilisant des modèles et des méthodes multivariés. Habituellement, un modèle linéaire simple est utilisé pour prédire la relation complexe entre la réponse instrumentale (Y_t) et la valeur réelle de l'étalon analysé (X_t). La précision du modèle basée sur l'incertitude et d'autres propriétés telles que la linéarité, la sensibilité, etc., seront prises en compte. Ce modèle est étendu à des systèmes multivariés, par exemple la matrice de l'échantillon analysé en établissant la composition chimique clé de l'acier telle que la teneur en carbone (C(wt%)), en manganèse (Mn(wt%)), en silicium (Si(wt%)) et en soufre (S(wt%)). Une étude comparative entre la procédure habituelle basée sur les méthodes simples et multivariées est développée. Le modèle simple est linéaire et utilise des entrées-sorties uniques, où chaque atome est considéré sans interaction avec les autres, ignorant ainsi l'influence de l'un sur l'autre. La nouveauté ici est l'utilisation d'un modèle multivarié imposant le fait réel des interactions entre les concentrations d'atomes, ce qui permet un meilleur étalonnage et un meilleur réglage de l'analyse chimique multi-éléments de l'acier à l'aide de la technologie des rayons X.

Préliminaires et notions fondamentales

Dans ce chapitre, nous rappelons brièvement des définitions et théorèmes connus dans le domaine d'analyse des données. On commence tout d'abord par présenter ce domaine intéressant, la procédure pour étudier n'importe quel projet d'exploration des données et les domaines d'application. Ensuite, nous nous intéressons aux types d'apprentissages. Enfin, nous donnons un aperçu sur les termes et les concepts mathématiques et métrologiques fondamentaux.

1.1 Qu'est que l'exploration des données

En français, les termes « forage de données », « explorations de données » ou « fouille de données » désignent l'ensemble des algorithmes et techniques destinés d'exploiter d'immenses bases de données informatiques, sans idée préconçue. L'objectif est de trouver ou de découvrir des règles, des relations, des corrélations ou des modèles entre de nombreuses bases de données numériques. L'exploration des données est aussi essentiellement un ensemble de méthodes incorporées dans un logiciel. Ces techniques proviennent de différents domaines basés sur les statistiques avancées et l'intelligence artificielle. Parmi eux, on trouve des outils statistiques fondamentaux et multidimensionnels d'exploration et de classification, ainsi que des méthodes de modélisation numérique utilisant diverses généralisations des modèles linéaires traditionnels (régression multiple, régression logistique et analyse discriminante). Elles sont ensuite confrontées à des techniques complémentaires, plus algorithmiques, qui produisent des arbres de décision, des règles d'induction ou des réseaux neuronaux [\[3\]](#).

On peut dire que l'exploration des données est l'assemblage de diverses disciplines.

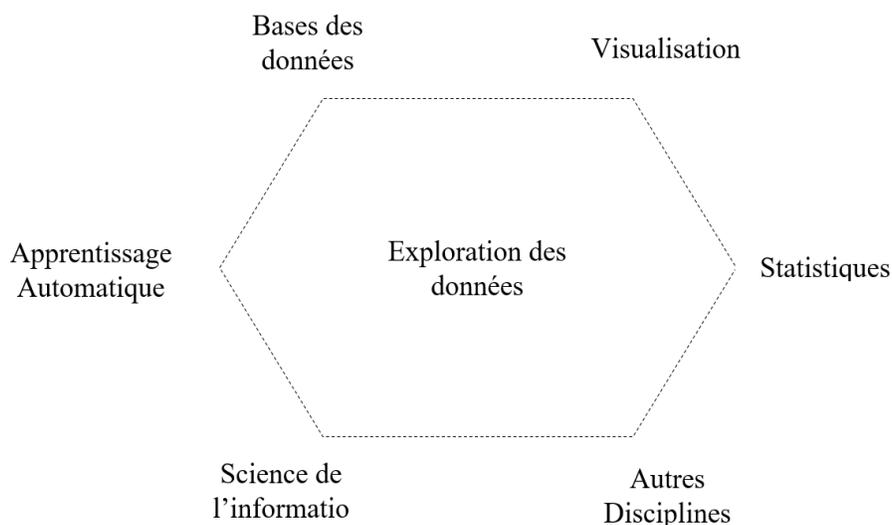


FIGURE 1.1 – Exploration des données comme confluent de plusieurs disciplines [46]

1.2 Étapes d'un projet d'exploration des données

L'exploration de données est plus qu'une simple collection d'algorithmes ; elle implique une série d'étapes qui sont brièvement décrites ci-dessous :

- **Identification du problème** : quel est l'objectif de l'étude ?, Comment transformer le problème en une requête pouvant faire l'objet d'une recherche sur l'exploration de données ? [48].
- **La collection des données** est une phase cruciale car l'algorithme s'en nourrit. La réussite du projet dépend de la collecte de données pertinentes, en quantité suffisante et de bonne qualité. Les données peuvent être structurées ou non selon différents types : données tabulaire ou textuelles, images, sons, enregistrements, etc ... [48].
- **Préparation et nettoyage des données** : l'élimination des données bruitées parce que certains attributs sont inutiles (les attributs ayant des valeurs invalides ou des valeurs manquantes) [46, 83].
 - Dans certaines situations, les données sont déjà sous forme convenable et au bon format, donc aucune modification n'est nécessaire.
 - La description d'une entité est un attribut. Il est parfois appelé variable ou caractéristique.
- **Sélection des données** : on ne conserve que les informations pertinentes pour la recherche en cours.
- **Construction et vérification des modèles** : sur la base des mesures, l'étape d'apprentissage consiste à découvrir des modèles strictement intéressants. Une autre série indépendante de données d'entraînement utilisée pour tester le modèle et assurer qu'il ne provoque pas de surapprentissage [48].

- **Optimisation et développement** : les modèles obtenus par des approches d'apprentissage supervisé, en particulier, ne peuvent pas être utilisés directement. Ils doivent être évalués. La procédure usuelle consiste à évaluer au mieux le taux d'erreur du modèle, ensuite l'utilisateur décide d'employer ou non le modèle de prédiction, tout en étant pleinement conscient des risques encourus. Il existe plusieurs techniques de validation par exemple : "validation croisée k-fold, exclusion d'un échantillon".

1.3 Domaines d'application

- **Marketing et gestion de la relation client** : établissement des listes de clients en se concentrant sur les clients potentiels et en explorant des secteurs émergents [68].
- **Environnement** : exemple étude géotechnique sur les données d'infrastructure et de fondations au travers d'un ensemble de phases successives les conditions morphologiques, géologiques, locales et régionales du site.
- **Finance** : minimisation des risques financiers.
- **Assurance** : identifiant divers groupes d'assurés associés à de nombreuses déclarations [68].
- **Bio-informatique** : étude de développement des médicaments, études des effets de ces médicaments sur la santé des êtres vivants.
- **Internet** : il y a plusieurs exemples parmi eux email spam, commerce électronique.
- **Diagnostic médical des causes et des symptômes des maladies** :
Exemple 1 : Le système peut catégoriser de nouveaux patients en fonction de points communs détectés avec les symptômes d'autres patients connus, ce qui permet d'analyser et estimer les risques et donne de bonnes directions sur les diagnostics.
Exemple 2 : localisation et classification de tumeur dans les différentes zones des corps.

1.4 Tâches d'exploration des données

- **La classification** : La construction de modèles (fonctions linéaires ou non linéaires) qui caractérisent et distinguent les idées pour les prédictions à venir nécessite d'associer les données à des groupements prédéterminés [46].
La classification est la méthode d'organisation en classes les objets qui partagent des caractéristiques similaires en fonction de classes ou de catégories, Les deux catégorie de classification sont supervisés et non supervisés [50]. La première catégorie d'approches permet de faire une classification à l'aide d'une

base de données d'apprentissage et une autre pour test et validation, alors que la seconde catégorie de méthodes permet de faire la classification sans étape d'apprentissage.

- **Clustering ou apprentissage non supervisé(e)** : Elle est semblable à la classification, seulement les groupes ne sont pas prédéfinis à l'avance [46]. Clustering, également appelé segmentation, est le rassemblement d'objets en catégories plus homogènes, le clustering permet de transformer un ensemble des objets d'un espace quelconque de départ en un espace d'arriver de groupes appelé Cluster. Par conséquent, un Cluster est constitué d'une collection d'objets qui sont semblables entre eux.
L'objectif des techniques du Clustering est de regrouper des objets similaires dans la même catégorie [13], de sorte que les éléments comparables d'un même groupe soient aussi semblables que possible, tandis que les éléments dissemblables de deux groupes distincts soient aussi différents que possible.
En traitement d'images, ce genre de classification est souvent utilisé pour la segmentation : routes, villes, rues, organes humaines.
- **La prédiction** : utilise des variables connues pour prévoir les valeurs futures des mêmes variables (séries temporelles) ou des autres variables (régression). Bien qu'à une échelle de temps différente, la prédiction est similaire à la classification [13]. Elle est basée sur le passé et le présent, tout comme les emplois précédents, mais le résultat se situe dans un futur assez précis. Il ya plusieurs moyen d'évaluer la précision d'une prédiction en utilisant des indicateurs de performance comme standard déviation.

1.5 Différents types d'apprentissage automatique

Plusieurs types des méthodes d'apprentissage automatique ont été créés pour traiter des problèmes dans différents domaines et sujets.

Selon la méthode de formation, ces méthodes de machine learning sont divisées en quatre catégories : apprentissage supervisé, non-supervisé, semi-supervisé, renforcé [23].

1.5.1 Caractérisation des méthodes : apprentissage supervisé

Le processus d'apprentissage supervisé est assez semblable à l'apprentissage humain. Rappelons que les gens apprennent des nouvelles compétences en trouvant des nouvelles solutions à des problèmes réels :

- Choisir un défi d'entraînement. Utiliser les informations actuelles pour trouver des solutions. Comparer la vraie solution avec le résultat.
- Si le résultat est faux, adapter les conditions et les paramètres d'optimisations actuels.

c. Pour tous les problèmes, essayer de répéter les deux premières étapes a) et b). Les deux types d'application les plus courants de l'apprentissage supervisé sont la classification et la régression. Ces mots peuvent sembler inconnus, mais ne sont en réalité pas si difficiles. Commençons par la classification.

C'est peut-être l'application la plus répandue de l'apprentissage automatique. Le problème de classification se concentre sur la recherche littérale des classes auxquelles appartiennent les données. Quelques exemples peuvent aider [66].

Service de filtrage des courriers indésirables → Classe les courriers par courrier ordinaire ou spam.

Service de reconnaissance de chiffres → Classe l'image numérique dans l'un des 0-9.

Service de reconnaissance faciale → Classe l'image du visage dans l'un des utilisateurs enregistrés.

Nous avons abordé dans la section précédente que l'apprentissage supervisé nécessite paires d'entrée et de sortie correctes pour les données d'apprentissage. De même, les données d'entraînement du problème de classification ressemble à ceci :

{entrée, classe} [66].

Dans le problème de classification, nous voulons savoir à quelle classe appartient l'entrée. Ainsi, la paire de données a la classe à la place de la sortie correcte correspondant à l'entrée. L'apprentissage supervisé est subdivisé en deux catégories : régression et classification, suivant la manière dont le modèle utilisé. La régression prévoit les valeurs futur des mêmes variables dans la série temporelle. Cependant, d'autre part, la classification indique à quel groupe doivent appartenir les éléments d'entrée. Le résultat d'une classification correcte est présenté sous une forme de classes [66].

- **But** : calculer la relation entre l'entrée X et la sortie Y à partir des données $\{(x_i, y_i) \in X \times Y, i = 1 \dots N\}$. Étant donné que les y_i sont destinés pour diriger le processus d'estimation, on parle d'apprentissage supervisé.
- **Techniques** : SVR, NN, Arbre de décision, régression logistique, k-plus proche voisin, régression linéaire.

1.5.2 Apprentissage non-supervisé

- **But** : la sortie Y n'existe pas, c'est-à-dire seulement les entrées $\{x_i \in X, i = 1 \dots N\}$ sont disponibles. L'objectif est de décrire l'organisation des données et d'obtenir des sous-ensembles homogènes.

Par conséquent, l'apprentissage non supervisé ne contient que des données entrées sans sorties réelles.

L'apprentissage non-supervisé est habituellement utilisé pour analyser les caractéristiques des variables et les prétraitements. L'idée est comparable à celle d'un élève qui n'apprend pas à répondre aux problèmes et qui se contente de les classer en fonction de leur construction et de leurs attributs [66].

- **Méthodes** : K-means, extractions de caractéristiques, analyse en composante

principale (réduction de dimension), Classification hiérarchique.

1.5.3 Caractérisation des méthodes : apprentissage semi-supervisé

- **Objectifs** : le but est identique à celui de l'apprentissage supervisé, mais nous préférons utiliser des données sans étiquette.
- **Techniques** : méthodes bayésiennes.

1.5.4 Caractérisation des méthodes : apprentissage par renforcement

L'apprentissage par renforcement utilise des ensembles d'entrées, de certaines sorties et de notes comme données de formation. Il est généralement utilisé lorsqu'une interaction optimale est requise, comme pour le contrôle et les jeux. {entrée, sortie, note pour cette sortie} [23, 66].

1.6 Définitions et concepts

Afin de rendre ce travail plus facile à comprendre, nous présentons ici quelques rappels sur les concepts essentiels :

1.6.1 Définitions ISO 5725

AFNOR, NF ISO 5725-1 : Principes généraux et définitions de la statistique, Partie 1 : Exactitude des résultats et techniques de mesure, sont abordés dans le livre AFNOR publié à Paris en décembre 1994, qui comprend 24 pages.

AFNOR, NF ISO 5725-2 : Règles fondamentales de statistique, l'exactitude des résultats et les techniques de mesure sont abordées dans la partie 2 de l'ouvrage AFNOR, publié à Paris en décembre 1994, qui comprend 54 pages.

AFNOR, NF ISO 5725-3 : le même que ISO 5725-2, mais les chercheurs ajoutent une 3^e Partie concernant " Les mesures moyennes de la fidélité d'une approche de mesure normalisée", AFNOR, Ed. Paris, Décembre 1994, 35 pages.

AFNOR, NF ISO 5725-4 : le même que la version précédente, sauf il y a une partie intéressante c'est la 4, technique fondamentale permettant de définir la justesse d'une approche de mesure normalisée, AFNOR, Ed. Paris, Décembre 1994, 32 pages.

AFNOR, NF ISO 5725-5 : le même que 5725-4, mais les chercheurs ajoutent une 5^e Partie consacrée aux : Techniques différentes permettant de définir la fidélité d'une métrologie normalisée, de l'ouvrage AFNOR, Ed. publié à Paris en décembre 1994, 66 pages.

AFNOR, NF ISO 5725-6 : le même que 5725-5, bien que les experts ont ajouté une 6^e partie dédiée l'utilisation concrète des valeurs d'exactitude, AFNOR, Ed. Paris,

Décembre 1994, 53 pages.

AFNOR, FD ISO/TS 21748 : Guide utilisé pour les évaluations de la répétabilité, reproductibilité et de la justesse dans un contexte d'estimation de l'incertitude de mesure, de l'ouvrage AFNOR, publié à Paris en Juin 2004, 32 pages.

1.6.2 Répétabilité

La définition de la répétabilité est la suivante : la même technique de mesure est utilisée de manière répétée sur une courte période par le même expert, avec le même outil de mesure, au même endroit et dans les mêmes conditions d'utilisation. La variance de répétabilité, l'écart-type de répétabilité ou toute autre mesure statistique décrivant la dispersion des résultats peuvent être utilisés pour indiquer la répétabilité, qui se caractérise par une dispersion plus ou moins grande des résultats de mesure. On note ici par σ_r^2 la variance de répétabilité, c'est la variance d'erreur « pure » [14, 41, 26].

1.6.3 Reproductibilité

La variation d'au moins une des conditions qui ont été énumérées pour la répétabilité, les autres conditions restant inchangées, est connue sous le nom de reproductibilité. Une variance de reproductibilité, également appelée écart type de reproductibilité, est un moyen d'exprimer la reproductibilité. La variance de reproductibilité σ_R^2 sera notée ici.

σ_R^2 : Variance de reproductibilité, c'est la sommation de la totalité des causes de variation de la réponse, les effets d'interaction entre les facteurs sont considérés, à l'exclusion de la variance de reproductibilité σ_r^2 .

La "justesse" et la "fidélité" sont les deux termes utilisés par l'ISO 5725 pour décrire l'exactitude d'une technique de mesure. L'étroitesse de la concordance entre la valeur moyenne arithmétique produite à partir d'une grande série de résultats de tests et la valeur de référence réelle est appelée « justesse ». L'étroitesse de la concordance entre les résultats de tests est connue sous le nom de "fidélité" [14, 41, 26].

1.6.4 Fidélité

Le degré de concordance entre les différents résultats des tests est appelé fidélité. La prise en compte de la fidélité découle du fait que les tests effectués sur des produits en apparence identiques dans des conditions supposées identiques n'aboutissent généralement pas à des résultats équivalents [14, 41].

Cela est dû aux défauts aléatoires inhérents à tout processus de mesure ; toutes les variables affectant le résultat d'une mesure ne peuvent être totalement contrôlées. Cette variabilité doit être prise en compte lors de l'interprétation pratique des résultats des mesures.

Par conséquent, les deux opposés polaires de la fidélité sont la répétabilité et la reproductibilité. Les résultats sont plus variés lorsqu'ils sont reproductibles que lorsqu'ils sont répétés. La manière conventionnelle d'exprimer la fidélité est en fonction d'écart-type.

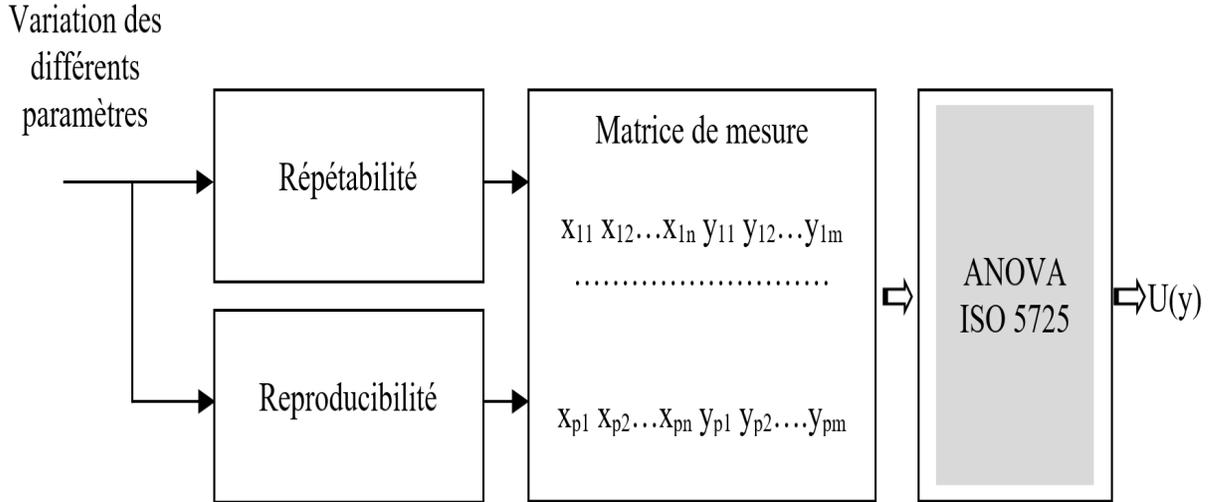


FIGURE 1.2 – Schéma des différents notions de base

Modèle de Base :

$$Y = m + b + e \quad (1.1)$$

Le seuil de test est la moyenne de m .

Dans plusieurs disciplines, le niveau de test est limité à la méthode de mesure et le concept de valeur réelle autonome ne s'applique pas.

Le terme b est censé être constant au cours des tests effectués dans le cadre de la répétabilité, bien que sa valeur puisse varier pour les tests effectués dans d'autres conditions. Quand les résultats des tests sont régulièrement analysés par les deux mêmes centres, il est nécessaire pour eux de déterminer leur biais relatif, soit à partir de leurs valeurs individuelles de biais déterminées pendant une expérience d'exactitude, soit en mettant eux-mêmes un essai particulier [14, 41].

Le terme e signifie qu'une erreur aléatoire se produit pour chaque résultat de test.

1.7 Termes métrologiques généraux

Métrologie : La métrologie est constituée de l'ensemble des méthodes et des connaissances qui assurent la possibilité d'effectuer des mesures et d'avoir une fiabilité adéquate dans leurs produits. La mesure est indispensable à tout savoir-faire, à toute prise de responsabilité et à toute activité.

La philosophie de chaque activité est la suivante :

Observer et mesurer, ensuite comprendre, et prévoir/agir, et enfin mesurer/vérifier.

Le concept incertitude : Le concept « incertitude » fait référence au doute. Par conséquent, « incertitude de mesure » dans le sens le plus large. Comme il existe peu de termes pour décrire à la fois l'idée générale d'incertitude et les quantités particulières qui fournissent des estimations quantitatives de cette idée, telles que l'écart-type, le mot "incertitude" est nécessaire dans les deux sens [43].

Mesurande : la quantité que nous voulons mesurer, par exemple la longueur d'une barre en état d'équilibre à la température extérieure de 23°C, sera différente de la longueur à la température déterminée de 20°C, qui est le mesurande, ce qui nécessite dans ce cas une fixation [43].

Mesurage : C'est le groupe des opérations visant à obtenir la valeur du mesurande, c'est-à-dire la valeur de la quantité à mesurer. Par conséquent, un mesurage débute avec une bonne description correcte et convenable du mesurande, de la technique et procédure de mesure [43].

Modélisation du mesurage : dans plusieurs cas, un mesurande Y n'est pas encore clairement mesuré mais il est obtenu à partir de différentes autres grandeurs X_1, X_2, \dots, X_n à travers une fonction $f : Y = f(X_1, X_2, \dots, X_n)$.

Erreur de mesure : C'est la différence entre la valeur réelle du mesurande et la valeur calculée du mesurage [43].

Sources d'incertitude : Dans la pratique, il existe plusieurs causes potentielles d'incertitude dans un mesurage, telles que :

- 1) définition incorrecte ou inadéquate du mesurande.
- 2) mauvaise exécution de la définition du mesurande.
- 3) échantillonnage non crédible ou non significatif.
- 4) connaissance trop limitée des influences des zones du milieu de mesurage.
- 5) les normes incorrectes, c'est-à-dire les valeurs des matériaux de référence et des étalons imprécises.
- 6) valeurs incorrectes des paramètres de sources extérieures et utilisés dans les algorithmes et les méthodes d'exploration de données.
- 7) les changements entre des observations répétées du mesurande dans des circonstances presque semblables [43].

1.8 Techniques d'évaluation d'incertitude

Il y a plusieurs estimateurs, on peut citer :

- 1-Estimation de l'incertitude à l'aide de la méthode conventionnelle(GUM).
- 2-Estimation des incertitudes à l'aide de la simulation de Monte Carlo.
- 3-Utilisation combinée de la simulation Monte Carlo et Chaîne de Markov.

1.9 Facteurs de l'importance des incertitudes

La mesure de l'incertitude pendant les essais, et l'étalonnage est une opération fondamentale et importante en science et ingénierie et particulièrement en laboratoire. La capacité de mesure est définie par l'importance des incertitudes de mesure qui sont affectées par les facteurs suivants :

- La performance de la méthode utilisée ou développée.
- Les performances de l'équipement, c'est-à-dire la capacité de mesure d'étalonnage de différents instruments et équipement. (Calibration Measurement Capability (CMC))
- Le moyen comme la compétence des ressources humaines.
- Les conditions et paramètres d'environnement.
- La performance des matériaux utilisés.

La quantité d'incertitude est une valeur de l'écart type ; il est défini comme une somme des composantes quadratiques.

1.10 Termes et concepts statistiques fondamentaux

Il y a plusieurs concepts, on peut citer :

Variable aléatoire : variable qui peut prendre n'importe quelle valeur parmi un intervalle de valeurs définies, et à laquelle est liée une loi de probabilité.

La variable aléatoire est un " concept " qui traduit les chances d'observer une valeur par rapport à toutes les valeurs possibles.

Loi de probabilité : fonction qui détermine la possibilité qu'une variable aléatoire prenne une certaine valeur ou qu'elle existe dans un groupe de valeurs données.

Fonction de répartition : généralement notée F , c'est la définition mathématique de la signification de la possibilité d'observer une grandeur par rapport à toutes les grandeurs disponibles. Ainsi, la fonction exprime la possibilité que la variable aléatoire pour toute valeur x , la probabilité que la variable aléatoire X soit inférieure ou égale à x : $F(x) = Pr(X \leq x)$ [43].

Corrélation : Lien entre deux ou plusieurs variables aléatoires dans une loi. La moyenne qui exprime plus précisément le centre d'une distribution [43].

Variance : La dispersion d'une variable aléatoire X par rapport à sa valeur moyenne est représentée par sa variance [43, 46].

La notion d'incertitude est exprimée par la variance. Plus cette variance est faible, moins incertain est le résultat de l'expérience aléatoire.

Écart-type : il est défini par [43, 46]

$$\sigma = \sqrt{V(x)} = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}} \quad (1.2)$$

Matrice de Corrélation Il y a 3 types :

- Une corrélation positive élevée implique que si une variable augmente, l'autre aug-

mente également.

-Une corrélation négative élevée signifie que si une variable augmente, en revanche, l'autre diminue.

-Une corrélation presque 0 indique que les variables ne sont pas liées.

Test de signification pour ajuster la corrélation

Les paramètres fondamentaux les plus couramment utilisés pour évaluer les performances de prédiction sont le coefficient de corrélation R et l'erreur quadratique moyenne (MSE). L'évaluation de R est citée dans les estimations suivantes [71].

- $|R| \geq 0.8$: corrélation élevée.
- $0.2 \leq |R| < 0.8$: la corrélation existe.
- $|R| < 0.2$: faible corrélation.

Distribution :

Une distribution (aussi appelée fonction généralisée) est un outil qui permet de formaliser les concepts de fonction et de mesure, en théorie de probabilité, c'est une loi « loi normal, loi de Poisson.. » qui définit les probabilités de chacune des variables aléatoires discrètes. Il y a deux catégories de distribution :

Distributions continues : par exemple « uniformes, gaussiennes, multi-gaussiennes, de Dirichlet,... »

Distributions discrètes : par exemple « binomiales, multinomiales, etc. ».

La majorité des distributions continues sont aussi définies sous forme discrète.

Fonction de vraisemblance :

C'est une fonction de probabilités conditionnelles, connue sous le nom de $L(x_1, x_2, \dots, x_n | \theta_1, \theta_2, \dots, \theta_k)$, définit les paramètres θ_j d'une loi statistique en fonction des valeurs x_i considérées être connues.

Elle est exprimée par la fonction de densité $f(x; \theta)$ par [63] :

$$L(x_1, \dots, x_n) = \prod_{i=1}^n f(x_i; \theta)$$

Techniques d'exploration des données

Dans ce chapitre, on essaie de donner le principe de différentes techniques d'apprentissages automatiques et d'optimisation en vue de prédire ou de décrire les phénomènes en instance. Parmi ces techniques on a : réseau de neurones artificiels (NN), moindres carrés récursives (RLS), moindres carrés partielles (PLS), séparateur à vaste marge (SVR), vraisemblance résiduelle maximum, ainsi que les versions adaptatives de chaque méthode en utilisant les fenêtres glissantes. Ces méthodes, qui ont trouvé leur succès dans des applications industrielles, permettent de développer rigoureusement les modèles disponibles même dans des systèmes compliqués. Par conséquent, ces approches sont adaptées à presque tous les problèmes pour lesquels de grandes quantités de données mesurées du système sont accessibles.

2.1 Réseau de neurones artificiels

2.1.1 Historique

Les réseaux de neurones ont été inventés par Mc Culloch et Pitts en 1950, et rendus populaires par Hopfield [38], fondés sur des mathématiques et des approches algorithmiques appelées logique de seuil. En tant que représentation de mappage non linéaire entre les ensembles d'entrée et de sortie, le NN émule les systèmes humains (cerveau), qui se caractérisent par l'adaptation et l'auto-organisation. De plus, ils ont la capacité d'apprendre de l'expérience et de la généralisation de l'échantillon précédent pour résoudre de nouveaux problèmes.

2.1.2 Définition et principe des neurones artificiels

Les réseaux de neurones artificiels sont des réseaux de processeurs élémentaires très connectés fonctionnant simultanément. Chaque processeur élémentaire calcule une sortie unique sur la base des informations qu'il reçoit, puis cette sortie se divise en un nombre variable de neurones en aval.

Tout neurone artificiel se présente comme un processeur fondamental. Il reçoit diverses entrées des neurones en amont ou des capteurs de la machine dont il fait partie. Un poids représentant la force de la connexion est associé à chaque entrée. Ils sont composés d'un nombre de composants de traitement extrêmement simples appelés neurones (voir Diagramme 1). Un neurone se compose généralement de quatre composants : (1) des données d'entrée, (2) un groupe de poids, (3) des noeuds, et (4) une fonction d'activation non linéaire f , telle que la sigmoïde. Notez que chaque neurone a une fonction d'activation. Les entrées x_i sont connectées aux noeuds de la couche d'entrée, et les sorties \hat{y} sont tirées de la couche de sortie.

Il existe une ou plusieurs couches cachées entre les couches d'entrée et de sortie. Les valeurs d'entrée du réseau sont transmises aux neurones de la première couche cachée par les neurones de la couche d'entrée.

Chaque noeud d'une couche est lié à chaque noeud de la couche suivante. La force de connexion entre la sortie d'un noeud i avec un noeud j est donnée par un poids w_{ij} . Les poids sont des coefficients de régression à estimer à partir d'un échantillon de données. Le terme de biais est comparable à l'interception du modèle de régression conventionnel, c'est-à-dire qu'il nous permet d'ajouter de la flexibilité lors de l'apprentissage [84].

Le modèle peut s'écrire comme suit :

$$u = \sum_{i=1}^N w_i x_i + b \tag{2.1}$$

$$\hat{y} = \phi(u) \tag{2.2}$$

où x_i représente la valeur d'entrée, \hat{y} le paramètre de sortie du processus, $W = \sum_{i=1}^N w_i x_i$ est le poids reliant l'entrée i à un neurone caché i , et ϕ est non linéaire appelée fonction d'activation. Diverses fonctions d'activation sont possibles, notamment sigmoïde, tangente hyperbolique et gaussienne.

De nombreuses méthodes existent pour déterminer l'architecture idéale d'un réseau de neurones.

Ces méthodes sont généralement de nature assez complexe et difficile à mettre en oeuvre. En outre, il n'existe pas de critères rigoureux concernant le choix du nombre de noeuds et de couches cachées à utiliser dans une application. Habituellement, quelques essais et erreurs sont nécessaires pour déterminer la meilleure combinaison pour minimiser l'erreur. Par conséquent, les poids de connexion sont auto-ajustés à l'aide d'algorithmes d'optimisation non linéaire efficaces, à savoir l'algorithme d'apprentissage de base de rétro-propagation (BBP). E est une fonction de coût quantifiant la différence entre les valeurs connues initiales de la fonction approchée sur l'ensemble discret de données et leurs approximations NN correspondantes, écrit sous la forme :

$$E = \frac{1}{n} \sum_{k=1}^n (e_k)^2 = \frac{1}{n} \sum_{k=1}^n (y_k - \hat{y}_k)^2 \tag{2.3}$$

où n est le nombre de termes d'erreur.

L'algorithme de Levenberg-Marquardt (LM), le plus largement utilisé dans un algorithme d'optimisation, peut être considéré comme une modification de la région de confiance à la méthode de Gauss Newton (GN) ou plus robuste comme un pont entre le GN et l'algorithme de descente de gradient. Il a été aisément montré dans de nombreux cas, qu'elle converge même si la surface d'erreur est beaucoup plus complexe que dans la situation quadratique. Le LM est similaire au rétro-propagation (BBP) dans le sens où il ne nécessite que le calcul du vecteur gradient, si le LM calcule en plus le Jacobien. L'algorithme LM peut être représenté comme suit :

$$W_{k+1} = W_k - (J_k^T J_k + \mu I)^{-1} J_k^T E; \quad (2.4)$$

où I est la matrice d'identité, et μ un paramètre d'apprentissage qui doit être ajusté plusieurs fois à chaque itération afin que le résultat avec la plus grande réduction d'erreur soit sélectionné. Lorsque la valeur μ est très grande, l'algorithme LM devient le plus raide décent ou BBP, tandis que lorsqu'il est égal à zéro, il s'agit de la méthode de Newton.

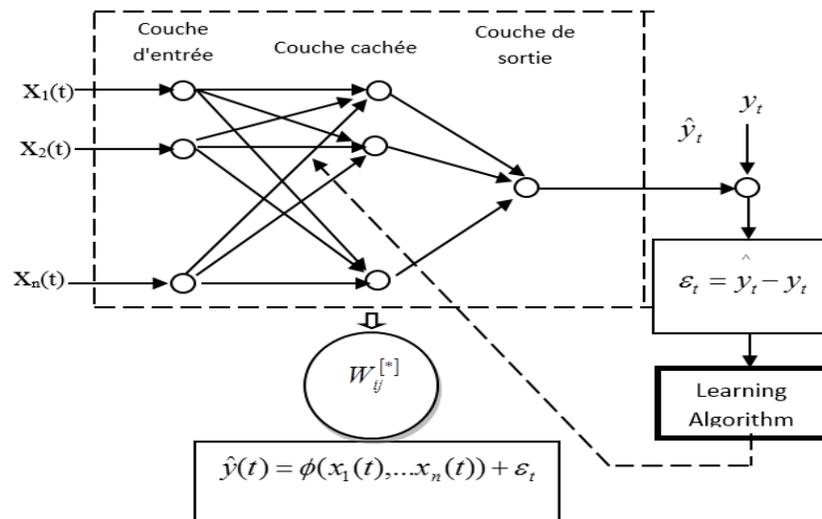


Diagramme 1 : Principe d'identification du modèle par la méthode NN

2.1.3 Types de fonctions d'activation

Le tableau 2.1 énumère de nombreuses fonctions de seuillage ϕ (ou bien fonction de transfert) qui peuvent être utilisées comme fonctions d'activation des neurones (à noter que chaque neurone a besoin d'une fonction d'activation) typiquement, la même fonction d'activation est utilisée pour tous les neurones dans une couche particulière d'un réseau de neurones. Les trois fonctions les plus utilisées sont les fonctions "linéaires" , "sigmoïdes" et "tangente hyperbolique".

Catégorie	Nom de la Fonction	Nom	Equation
Seuil	Heaviside(seuil)	Hardlim	$\phi(net) = \begin{cases} 1 & \text{si } net > 0 \\ 0 & \text{si } net < 0 \end{cases}$
	Signe	Hardlim	$\phi(net) = \begin{cases} 1 & \text{si } net > 0 \\ -1 & \text{si } net < 0 \end{cases}$
Linéaire	Identité	Purelin	$\phi(net) = net$
	Linéaire saturée	Satlin	$\phi(net) = \begin{cases} 0 & \text{si } net < -1 \\ net & \text{si } -1 \leq net \leq 1 \\ 1 & \text{si } net \geq 1 \end{cases}$
sigmoïde	sigmoïde	Logsig	$\phi(net) = \frac{1}{1 + e^{-net}}$
Tangente hyperbolique	Tangente hyperbolique	Tansig	$\phi(net) = \frac{e^{-net} - e^{net}}{e^{-net} + e^{net}}$

TABLE 2.1 – Fonctions de transfert [9, 11]

2.1.4 Réseau de neurones artificiels de type fonction à base radiale (RBANN)

Le RBANN est un type spécial de réseaux neuronaux supervisés non linéaires universels à couches, qui se compose généralement de trois couches, appelé couche d'entrée, sortie et une seule couche cachée composée d'un certain nombre d'unités d'activation non linéaires du RBANN. Les fonctions d'activation dans RBANN sont généralement définies comme des fonctions gaussiennes. En outre, RBANN bénéficie d'avantages importants tels que la précision, la robustesse et la vitesse d'apprentissage. Pour plus de détails, voir ([21], [79]).

L'équation suivante est utilisée pour déterminer la sortie de la couche cachée du réseau de la $k^{\text{ème}}$ fonction d'activation G_k :

$$G_k(\|x - c_k\|) = \exp\left(-\frac{\|x - c_k\|^2}{\gamma^2}\right), \quad k = 1, 2, \dots, n \quad (2.5)$$

ici, $\| \cdot \|$ est la norme euclidienne, c_k est le centre, et γ est la largeur de la couche cachée. La sortie \hat{y}_l du noeud l de la couche de sortie du réseau est calculée comme suit :

$$\hat{y}_l(x) = \sum_{k=1}^n w_{lk} G_k + b_l, \quad (2.6)$$

où w_{lk} est le poids entre le $k^{\text{ème}}$ neurone de la couche cachée et le $l^{\text{ème}}$ neurone de la couche de sortie, et b_l est le biais $l^{\text{ème}}$ dans la couche de sortie.

Le modèle identifié à l'aide de RBANN est estimé par un algorithme robuste appelé

Levenberg-Marquardt. La forme récursive est donnée par l'équation suivante :

$$W_{kl}^t = W_{kl}^{t-1} - (H_{t-1} + \lambda_{t-1}I)^{-1}J_{t-1}^T E, \quad (2.7)$$

où I est la matrice d'identité, λ_{t-1} est une constante. H_{t-1} et J_{t-1} sont respectivement le hessien et le jacobien de E . De plus, le jacobien J est défini par :

$$J = \begin{bmatrix} \frac{\partial E_1}{\partial W_1} & \frac{\partial E_1}{\partial W_2} & \cdots & \frac{\partial E_1}{\partial W_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial E_i}{\partial W_1} & \frac{\partial E_i}{\partial W_2} & \cdots & \frac{\partial E_i}{\partial W_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial E_m}{\partial W_1} & \frac{\partial E_m}{\partial W_2} & \cdots & \frac{\partial E_m}{\partial W_n} \end{bmatrix}$$

où n est le nombre de poids et m est le nombre de modèles.

2.1.5 Lissage exponentiel combiné à RBANN(ES-RBANN)

Pour la méthode de lissage exponentiel, nous référons à [61] avec les équations suivantes :

$$\varepsilon_{t-1} = y_{c,t-1} - y_{p,t-1}, \quad (2.8)$$

$$\varepsilon_t = y_{c,t} - y_{p,t}, \quad (2.9)$$

Nous essayons ici de calculer la première valeur de lissage y_{pn} et le paramètre α tel que

$$\varepsilon_{t+1} = \alpha\varepsilon_t + (1 - \alpha)\varepsilon_{t-1}, \quad (2.10)$$

$$y_{pn,t+1} = y_{p,t+1} + \varepsilon_{t+1}. \quad (2.11)$$

où ε est l'erreur, $y_{c,t}$ est la sortie calculée par la distribution de probabilité au temps t . $y_{p,t+1}$ et $y_{pn,t+1}$ sont les résultats des techniques RBANN et ES-RBANN au temps $t + 1$, respectivement. Le paramètre α est appelé paramètre de relaxation vérifiant $0 < \alpha < 1$.

2.1.6 Algorithmes

Il existe plusieurs algorithmes de formation par rétropropagation (Backpropagation). Ils ont des exigences différentes en matière de calcul et de stockage, et aucun algorithme n'est le mieux adapté à tous les emplacements. La liste suivante résume les algorithmes de formation inclus dans la boîte à outils :

Acronyme	Algorithme
LM	trainlm-Levenberg Marquardt
BR	trainbr-bayesian Regularization
BFG	trainbfg-BFGS Quasi Newton
SCG	trainscg- Scaled conjugate gradient
CGB	traincgb-Conjugate Gradient with powell/beale Restarts
CGF	Traincgf-Fletcher-Powell Conjugate Gradient
CGP	traincgp - Polak-Ribière Conjugate Gradient
OSS	trainoss - One-Step Secant
GDX	traingdx - Variable Learning Rate Backpropagation

Algorithme NN

La structure du modèle peut être définie comme suit $f \rightarrow NN$

et son identification est simplifiée par les étapes suivantes :

Initialiser les poids du réseau $W_{ij}^0 = [-0.5 \quad a \quad +0.5]$ et définir la boucle de calcul comme :

Pour $k = 1 : n$

Étape 1. Acquisition des entrées/sorties.

Étape 2. Calculer u en utilisant l'équation (2.1)

Étape 3. Calculer la sortie du modèle $\hat{y}(k)$ à partir de (2.2)

Étape 4. Calculer l'erreur de modélisation $e(k) = y(k) - \hat{y}(k)$

a) Si $e(k) \approx 0$, $W_{ij}^k = W_{ij}^{k-1} G(k)e(k) \rightarrow$ arrêt : $W_{ij}^k = W_{ij}^{[*]}$

b) Sinon, ajuster les poids NN en utilisant l'algorithme récursif : $W_{ij}^k = W_{ij}^{k-1}$ par l'algorithme de Levenberg-Marquardt c'est-à-dire l'équation(2.4).

Fin k .

Avantages et inconvénients :

Avantages : Les réseaux neuronaux pouvant en principe approcher n'importe quelle fonction continue, le chercheur n'a pas nécessairement besoin de formuler des hypothèses sur le modèle sous adjacent [68].

Inconvénients : demande beaucoup de temps d'apprentissage, beaucoup de données pour l'entraînement et lorsqu'une erreur se produit dans les résultats, l'utilisateur n'a aucune connaissance du fonctionnement interne.

2.1.7 Apprentissage profond (Deep learning)

Depuis 2006, l'apprentissage profond est apparu comme un nouveau domaine de recherche de Machine Learning. Ces dernières années, les méthodes développées dans l'apprentissage profond ont déjà eu un impact sur les travaux de l'analyse de données et traitement d'image, y compris sur certains aspects de l'apprentissage automatique et l'intelligence artificielle. Ce qui différencie l'apprentissage profond de la technique d'apprentissage réseau de neurone artificiel est l'utilisation de plusieurs couches cachées ce qui explique le concept de profondeur.

Cela permet d'augmenter considérablement l'abstraction du processus et de rendre plus efficace le traitement des problèmes dits "compositionnels", c'est-à-dire ceux qui pourraient être divisés en plusieurs éléments. L'apprentissage de la machine c'est proche de l'apprentissage de l'être humain. On apprend des choses élémentaires avant de comprendre des choses plus compliquées; ce qui rend cette façon de procéder intéressante. L'intelligence artificielle, l'apprentissage automatique et l'apprentissage profond sont distincts; mais ils sont liés comme suit : "Le Deep Learning est une partie de Machine Learning, et le Machine Learning est une partie d'intelligence Artificielle".

2.2 Méthode des moindres carrés récursive(RLS)

Les approches des moindres carrés sont largement développées et utilisées dans de nombreuses références (voir par exemple [10, 33, 39, 57]), il en existe une simple appelée moindres carrés qui nécessite une inversion de matrice; cependant la plus populaire est RLS pour lequel on surmonte l'inversion matricielle. Les systèmes de mesure étant multivariés, une version multivariable de RLS est développée pour un système de mesure dynamique; avec la possibilité de considérer simultanément les paramètres du modèle et l'estimation de l'état.

Généralement, un système de mesure multivariable est défini par un système multivariable avec MISO comme :

$$y(k) = f(y(k-1), \dots, y(k-N_y), u_1(k-1), u_1(k-2), \dots, u_1(k-N_{u1}), u_p(k-1), \dots, u_p(k-N_{up})) + e(k) \quad (2.12)$$

où $y(k-1)$ est le vecteur des sorties du processus au temps discret $k-1$, $u(k-1)$ est le vecteur de ses entrées au même instant, p est le nombre d'entrées, et N_y, N_{u1}, N_{up} sont trois entiers caractérisant le modèle.

À noter que : $N_y = 0$ si le processus est supposé statique.

Pour un système linéaire, (2.12) devient :

$$y(k) = a_1 y(k-1) + \dots + a_{N_y} y(k-N_y) + b_{11} u_1(k-1) + b_{12} u_1(k-2) + \dots + b_{1N_{u1}} u_1(k-N_{u1}) + \dots + b_{p1} u_p(k-1) + \dots + b_{pN_{up}} u_p(k-N_{up}) + e(k) \quad (2.13)$$

Le vecteur θ et le vecteur $\varphi(k)$ d'information connue sont définis par :

$$\theta = [a_1, \dots, a_{N_y}, b_{11}, \dots, b_{1N_{u1}}, \dots, b_{p1}, \dots, b_{pN_{up}}] \in \mathbb{R}^{N_y+N_{u1}+\dots+N_{up}} \quad (2.14)$$

$$\varphi(k) = [y(k-1), \dots, y(k-N_y), u_1(k-1), \dots, u_1(k-N_{u1}), \dots, u_p(k-1), \dots, u_p(k-N_{up})] \in \mathbb{R}^{N_y+N_{u1}+\dots+N_{up}} \quad (2.15)$$

A partir de (2.13) et (2.14), on obtient le modèle d'identification suivant

$$y(k) = \theta \varphi^T(k) + e(k), \quad (2.16)$$

$$\hat{y} = \theta \varphi^T(k). \quad (2.17)$$

Soit $i = 1, 2, \dots, k$ une étape itérative, L_k est la longueur des données et θ_i une estimation de θ à l'itération i donnée par la minimisation des critères suivants :

$$J(\theta) = \min \left(\sum_{k=1}^{L_k} [y(k) - \theta \varphi^T(k)]^2 \right) \Rightarrow \theta = \hat{\theta}. \quad (2.18)$$

Le principe de minimisation de l'erreur de modélisation e qui est corrélée aux différentes erreurs de perturbation (ε) est schématisé dans diagramme 2 :

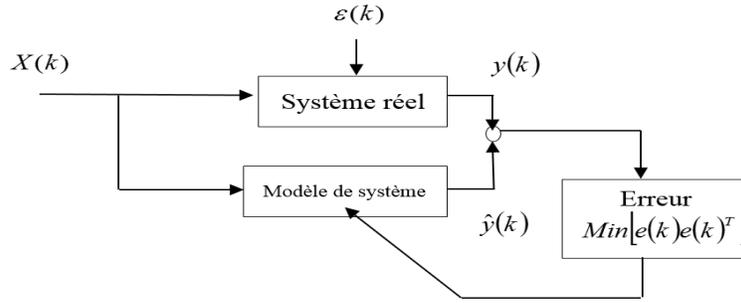


Diagramme 2 : Principe d'identification du modèle

En minimisant la fonction critère $J(\theta)$, l'estimation des moindres carrés de θ pour le temps global L_k et à chaque itération i est obtenue comme suit :

$$\hat{\theta}(k) = \left[\sum_{j=1}^k \varphi(j) \varphi(j)^T \right]^{-1} \left[\sum_{j=1}^k \varphi(j) y(j) \right]. \quad (2.19)$$

Pour chaque temps discret k (2.19) peut aussi s'écrire comme :

$$\hat{\theta}(k) = \left[\sum_{j=1}^k \varphi(j) \varphi(j)^T \right]^{-1} \left[\sum_{j=1}^{k-1} \varphi(j) y(j) + \varphi(k) y(k) \right], \quad (2.20)$$

$$\hat{\theta}(k) = \left[\sum_{j=1}^k \varphi(j) \varphi(j)^T \right]^{-1} \left(\left[\sum_{j=1}^{k-1} \varphi(j) \varphi(j)^T \right] \hat{\theta}(k-1) + \varphi(k) y(k) \right), \quad (2.21)$$

$$\hat{\theta}(k) = \left[\sum_{j=1}^k \varphi(j)\varphi(j)^T \right]^{-1} \left(\left[\sum_{j=1}^k \varphi(j)\varphi(j)^T - \varphi(k)\varphi^T(k) \right] \hat{\theta}(k-1) + \varphi(k)y(k) \right), \quad (2.22)$$

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \left[\sum_{j=1}^k \varphi(j)\varphi(j)^T \right]^{-1} \left(\varphi(k)y(k) - \varphi(k)\varphi^T(k)\hat{\theta}(k-1) \right). \quad (2.23)$$

Il faut noter que :

$$P(k) = \left[\sum_{j=1}^k \varphi(j)\varphi(j)^T \right]^{-1} \quad (2.24)$$

Eq(2.23) devient :

$$\hat{\theta}(k) = \hat{\theta}(k-1) + S(k) \left(y(k) - \varphi^T(k)\hat{\theta}(k-1) \right) \quad (2.25)$$

avec

$$S(k) = P(k)\varphi(k) \quad (2.26)$$

En se basant sur le lemme d'inversion matricielle [81] :

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(DA^{-1}B + C^{-1}DA^{-1}). \quad (2.27)$$

D'où :

$$P^{-1}(k) = P^{-1}(k-1) + \varphi(k)\varphi^T(k). \quad (2.28)$$

On obtient la forme itérative de $P(k)$ suivante :

$$P(k) = P(k-1) - \frac{P(k-1)\varphi(k)\varphi^T(k)P(k-1)}{1 + \varphi(k)P(k-1)\varphi^T(k)}. \quad (2.29)$$

Comme le montre l'équation (2.16), l'algorithme RLS traite le modèle linéaire lié aux paramètres, la sortie du modèle est une relation linéaire entre les entrées et les paramètres du modèle $\hat{y} = \theta\varphi^T(k)$.

Algorithme RLS

L'algorithme RLS développé sous sa forme multivariable est implémenté comme suit :

Initialisation : définir la structure du modèle et définir différentes variables : $N_y, N_{u1}, \dots, N_{up}$

$$N = N_y + N_{u1} + \dots + N_{up}$$

$$\theta_1 = [0 \ 0 \dots 0]$$

Pour $k = N : L_k$

1- Acquérir les données d'entrées/sorties.

2- Calculer le vecteur d'observation $\varphi(k)$ en utilisant (2.15).

- 3- Calculer le gain proportionnel $S(k)$ en utilisant (2.26).
 - 4- Utiliser la forme itérative et calculer la prochaine estimation $\hat{\theta}(k)$ en utilisant (2.23).
 - 5- Calculer la prochaine valeur $P(k)$ en utilisant (2.29).
- Fin k ,
- 6- Si $\hat{\theta}(k) - \hat{\theta}(k - 1) < \delta$ arrêter.
- La structure globale de l'identification du modèle est donnée dans le Diagramme 3 :

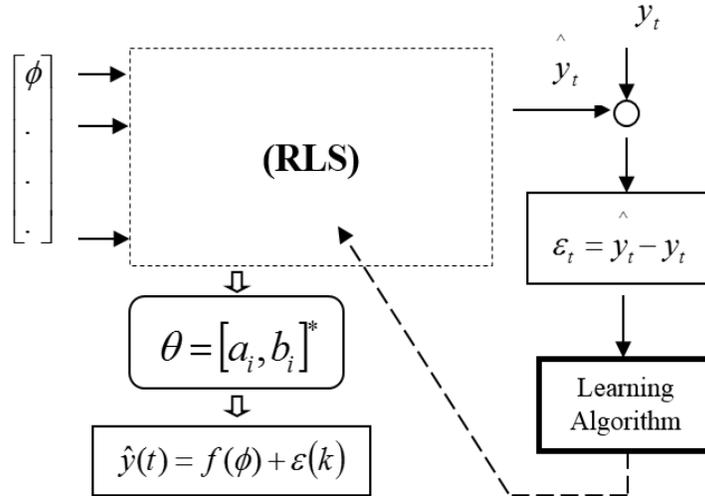


Diagramme 3 : Principe du modèle d'identification utilisant la méthode RLS

2.3 Méthode de Moindres carrés partiels (PLS)

L'étape clé de la modélisation de processus consiste à définir les variables d'entrée les plus importantes et à prévoir la réponse du processus à partir des données collectées. Néanmoins, la taille élevée et la colinéarité de ces données rendent complexe la construction d'un modèle de processus robuste. La nécessité d'expliquer la qualité du processus de ces données a conduit au développement d'une analyse multivariée pour les processus complexes. Afin d'atteindre cet objectif, nous avons essayé d'utiliser différentes stratégies parmi lesquelles les techniques des moindres carrés partiels (PLS), en tant que méthodologie de réduction de dimension supervisée qui a été inventée en 1983 par Herman Wold [77], il a été pris pour choisir un sous-ensemble optimal de variables d'entrée appelées "variables latentes", où son but est de construire de nouvelles variables prédictives (variables latentes) en tant que combinaisons linéaires des variables d'origine résumées dans une matrice X de variables prédictives (caractéristiques) et un vecteur y de variables de réponse (étiquettes de classe). Il se concentre

sur la maximisation de la covariance entre les facteurs extraits des ensembles d'entrée et de sortie des ensembles de processus de la manière suivante :

Soit $X \in \mathbb{R}^m$ un espace m -dimensionnel de variables et $Y \in \mathbb{R}^n$ un espace n -dimensionnel d'autres variables. Avec N échantillons observés de chaque espace $x \in X$ et $y \in Y$, nous obtenons deux blocs de variables $X \in \mathbb{R}^{N \times m}$ et $Y \in \mathbb{R}^{N \times n}$, et écrivons la formulation générale sous la forme :

$$X = \sum_{i=1}^k s_i \cdot p_i^T + \varepsilon_X = S \cdot P^T + \varepsilon_X, \quad (2.30)$$

$$Y = \sum_{i=1}^k u_i \cdot q_i^T + \varepsilon_Y = U \cdot Q^T + \varepsilon_Y, \quad (2.31)$$

où S et $U \in \mathbb{R}^{N \times p}$ sont des p matrices de facteurs extraites (score, variables latentes), $P \in \mathbb{R}^{m \times p}$ et $Q \in \mathbb{R}^{n \times p}$ sont des matrices de chargement, $\varepsilon_X \in \mathbb{R}^{N \times m}$ et $\varepsilon_Y \in \mathbb{R}^{N \times n}$ sont des termes d'erreur (matrices de résidus). La méthode PLS reste populaire en raison de ses points forts distinctifs parmi lesquels :

$$\operatorname{argmax}\{cov(X_{k-1}c_k, Y_{k-1}q_k)\}, \quad (2.32)$$

Sous les contraintes : $\|c\| = 1$. (*)

On cherche à maximiser en même temps la variance $s^{(1)}$ et la corrélation entre $s^{(1)}$ et Y . On doit donc chercher un vecteur normalisé $c^{(1)}$ maximisant :

$$\langle s^{(1)}, Y \rangle = \langle Xc^{(1)}, Y \rangle = \|Xc^{(1)}\| \cdot \|Y\| \cdot \operatorname{cor}(Xc^{(1)}, Y). \quad (2.33)$$

Nous résolvons le problème d'optimisation sous la contrainte (*) en utilisant l'approche des multiplicateurs de Lagrange dans le but de trouver la formule de $c^{(1)}$:

$$L(c^{(1)}, \lambda) = cov(Y, Xc^{(1)}) - \lambda(c^{T(1)}c^{(1)} - 1) = c^{T(1)}X^TY - \lambda(c^{T(1)}c^{(1)} - 1), \quad (2.34)$$

Où $\lambda \in \mathbb{R}^+$. Si nous fixons les dérivées premières de L par rapport à λ et $c^{(1)}$ à zéro, on écrit

$$\frac{\partial L}{\partial \lambda} = -(c^{T(1)}c^{(1)} - 1) = 0 \text{ ou } c^{T(1)}c^{(1)} = 1, \quad (2.35)$$

$$\frac{\partial L}{\partial c^{(1)}} = X^TY - 2\lambda c^{(1)} = 0 \text{ ou } X^TY = 2\lambda c^{(1)}. \quad (2.36)$$

En multipliant (2.36) par $c^{T(1)}$ et en utilisant (2.35), on obtient : $c^{T(1)}X^TY = 2\lambda$. Soit $\beta \in \mathbb{R}$, par symétrie on obtient :

$$\beta = 2\lambda = \langle s^{(1)}, Y \rangle = c^{T(1)}X^TY = Y^TXc^{(1)}. \quad (2.37)$$

De (2.36) et (2.37), on a :

$$(X^TY)Y^TXc^{(1)} = (\beta c^{(1)})\beta = \beta^2 c^{(1)}. \quad (2.38)$$

Par conséquent, $c^{(1)}$ est le vecteur propre correspondant à la valeur propre β^2 de la matrice $X^T Y Y^T X$; et la maximisation de $\langle X c^{(1)}, Y \rangle$ consiste à prendre β^2 pour la valeur propre maximale de $X^T Y Y^T X$.

Nous pouvons obtenir une formule pour $c^{(1)}$ et la valeur propre correspondante.

Nous écrivons :

$X^T Y Y^T X c^{(1)} = \lambda_1 c^{(1)}$ ou $\lambda_1 = \langle X c^{(1)}, Y \rangle^2 = (c^{T(1)} X^T Y)^T (c^{T(1)} X^T Y) = Y^T X X^T Y$.
Parce que $Y^T X X^T Y \in \mathbb{R}$, on obtient $X^T Y Y^T X c^{(1)} = (X^T Y) Y^T X (c^{(1)}) = \lambda_1 c^{(1)} = Y^T X X^T Y c^{(1)} = c^{(1)} Y^T X X^T Y = (c^{(1)}) Y^T X (X^T Y) \Rightarrow c^{(1)} = X^T Y$.

De la contrainte (*) nous avons $\|c^{(1)}\| = 1$ de sorte que $c^{(1)} = \frac{X^T Y}{\|X^T Y\|}$.

Pour vérifier si la 1^{ère} composante $s^{(1)} = X c^{(1)}$ peut exprimer adéquatement la totalité des variables explicatives, nous réalisons deux régressions X sur $s^{(1)}$ et Y sur $s^{(1)}$ pour obtenir :

$$X = X_{[0]} = s^{(1)} p^{T(1)} + X_1, \quad (2.39)$$

$$Y = Y_{[0]} = w^{(1)} s^{(1)} + Y_1, \quad (2.40)$$

avec $p^{(1)} = \frac{X_{[0]}^T s^{(1)}}{s^{T(1)} s^{(1)}}$ et $w^{(1)} = \frac{Y_{[0]}^T s^{(1)}}{s^{T(1)} s^{(1)}}$.

D'autres vecteurs de poids sont calculés de manière itérative par la méthode PLS, de sorte que lorsque $c^{(1)}$ et $w^{(1)}$ sont disponibles, les vecteurs de score peuvent être calculés par $s^{(1)} = X c^{(1)}$, $u^{(1)} = Y w^{(1)}$ et les chargements, c'est-à-dire que les premières colonnes de P et Q peuvent être calculées par $p^{(1)} = \frac{X_{[0]}^T s^{(1)}}{s^{T(1)} s^{(1)}}$ et $q^{(1)} = \frac{Y_{[0]}^T s^{(1)}}{u^{T(1)} u^{(1)}}$ respectivement. Les matrices de données X et Y sont ensuite dégonflés en soustrayant leurs approximations de rang un.

$$X \leftarrow X - s^{(1)} p^{T(1)} \text{ et } Y \leftarrow Y - u^{(1)} q^{T(1)} \quad (2.41)$$

Les nouveaux X et Y sont utilisés pour calculer $c^{(2)}$ et $w^{(2)}$ en fonction de

$$X^T Y Y^T X c^{(1)} = \lambda_1 c^{(1)} \text{ et } Y^T X X^T Y w^{(1)} = \lambda_1 w^{(1)} \quad (2.42)$$

Ce processus est répété jusqu'à ce que les résidus soient suffisamment petits ou un nombre prédéfini de vecteurs de poids $\{c^{(1)}, \dots, c^{(k)}\}$ et $\{w^{(1)}, \dots, w^{(k)}\}$ sont obtenus (voir [7, 15, 74]).

Algorithme PLS

Entrée : deux matrices X et Y , arbitraire c avec $\|c\| = 1$

Sortie : vecteurs de poids c et les résidus associés, vecteur de chargement S et vecteurs de score (s, u) .

Étape 1. Étant donné un vecteur de départ u_1 , généralement l'une des colonnes de Y .

Étape 2. Calculer les X -poids c , par $c_1 = \frac{X^T u_1}{\|X^T u_1\|}$

Étape 3. Calculer X -scores S , par $s_1 = X c_1$

Étape 4. Calculer les Y -poids L , par $L_1 = \frac{u_1^T s_1}{\|u_1^T s_1\|}$

Étape 5. Mettre à jour les Y -scores U , par $u_1 = Y L_1$

Étape 6. Mettre à jour B en fonction de s par : $B_1 = \frac{X^T s_1}{s_1^T t_1}$, $B \leftarrow \frac{B}{\|B\|}$, $s \leftarrow \frac{s}{\|s\|}$, $c \leftarrow \frac{c}{\|s\|}$.

Étape 7. Trouver le coefficient de régression pour la relation interne : $b_1 = \frac{u_1^T s_1}{\|s_1^T s_1\|}$

Étape 8. Les résidus, ε_x et ε_y , sont calculé par : $\varepsilon_x = X - s_1 B_1^T$, $\varepsilon_y = X - s_1 L_1^T$

Étape 9 . Continuer avec le composant suivant jusqu'à ce qu'il n'y en ait plus d'informations significatives [7].

2.4 Méthode de séparateur à vaste marge (SVR)

La régression à vecteurs de support a été introduite par Vapnik au début de l'année 1994 (voir [45]). Depuis, elle est devenue populaire et largement appliquée dans de nombreux domaines. Elle est fermement ancrée dans le cadre de la théorie de l'apprentissage statistique qui a conduit à un grand développement, fortement recommandé pour résoudre divers problèmes de classification et de prédiction, c'est aussi la clé pour construire la fonction de Lagrange à partir d'une fonction objective en convertissant le problème de minimisation en un problème dual [25, 72]. Supposons que l'ensemble d'apprentissage (échantillons) soit donné par :

$$E = \{(x_1, y_1), \dots, (x_i, y_i), \dots, (x_L, y_L)\}, \quad (2.43)$$

où $\forall i = 1, \dots, L$, $x_i \in \mathbb{R}^n$ est l'entrée de l'échantillon d'apprentissage, et $y_i \in \mathbb{R}$ est la valeur cible. Notre objectif est pour déterminer une fonction qui peut approximer les valeurs futures avec précision. Le modèle est alors donné par :

$$y = w \cdot \varphi(x) + b, \quad (2.44)$$

où $w \in \mathbb{R}^n$ sont les poids (c.à.d le vecteur des coefficients), $b \in \mathbb{R}$ est une constante, et φ une fonction de carte supposée être une transformation non linéaire de \mathbb{R}^n à une dimension supérieure espace de fonction.

Le but est de trouver le poids w et le biais b , tels que x est calculé en minimisant le risque de régression défini comme :

$$r_{reg}(f) = C \sum_{j=1}^m \Gamma(f(x_j) - y_j) + \frac{1}{2} \|w\|^2, \quad (2.45)$$

où $\Gamma(\cdot)$ est la fonction coût et C est une constante.

La solution qui minimise cette fonction coût satisfait un problème d'optimisation convexe avec une fonction de perte de marge souple, qui est représentée comme suit :

$$\min \frac{1}{2} \|w\|^2 + C \cdot \sum_j^m (\xi_j + \xi_j^*), \quad (2.46)$$

avec les contraintes

$$\begin{cases} \langle w, \phi(x_j) \rangle + b - y_j \leq \zeta + \xi_j^*, \\ y_i - \langle w, \phi(x_j) \rangle - b \leq \zeta + \xi_j, \\ \xi_j, \xi_j^* \geq 0, \end{cases} \quad (2.47)$$

où ζ est le paramètre insensible ; ξ_j, ξ_j^* sont des variables d'écart qui sont introduites pour relâcher les contraintes d'optimisation et le vecteur w peut s'écrire sous la forme :

$$w = \sum_{j=1}^m (\alpha_j - \alpha_j^*) \cdot \phi(x_j). \quad (2.48)$$

En substituant l'équation (2.48) dans (2.44), l'équation générique peut être réécrite comme :

$$y = f(x) = \sum_{j=1}^m (\alpha_j - \alpha_j^*) \cdot \langle \phi(x_j), \phi(x) \rangle + b = \sum_{j=1}^m (\alpha_j - \alpha_j^*) \cdot K(x_j, x) + b. \quad (2.49)$$

Ici $K(x_j, x)$ est la fonction noyau. Les fonctions noyaux permettent de calculer le produit scalaire dans un espace de grande dimension à partir de données d'entrée de petite dimension sans connaître la transformation ϕ . La fonction coût la plus largement utilisée est la fonction de perte insensible ζ qui a la forme :

$$\Gamma(f(x) - y) = \begin{cases} |f(x) - y| - \zeta & \text{pour } |f(x) - y| \geq \zeta, \\ 0 & \text{sinon.} \end{cases} \quad (2.50)$$

En résolvant le problème d'optimisation quadratique (2.46), le coût de la régression dans l'équation (2.45) et la fonction insensible ζ dans (2.50) peuvent être minimisés. On écrit :

$$\frac{1}{2} \sum_{j,l=1}^m (\alpha_j - \alpha_j^*)(\alpha_l - \alpha_l^*) K(x_j, x_k) - \sum_{j=1}^m \alpha_j^*(y_j - \zeta) - \alpha_j(y_j + \zeta), \quad (2.51)$$

Sous le contrainte :

$$\sum_{j,l=1}^m (\alpha_j^* - \alpha_j) = 0 \quad , \quad \alpha_j, \alpha_j^* \in [0, C], \quad (2.52)$$

où α_j^* et α_j sont les multiplicateurs de Lagrange, représentant les solutions au problème quadratique qui agissent comme des forces poussant les prédictions vers les valeurs cibles y_j .

Algorithme SVR

Initialiser les poids du réseau $W_{ij}^0 = [-0,5 \text{ à } +0,5]$ et définir la boucle de calcul comme suit :

Pour $k = 1 : n$

Étape 1 : Calculer ϕ

Étape 2 : Calculer la sortie du modèle \hat{y} à partir de (2.49)

Étape 3 : Calculer l'erreur de modélisation comme suit : $e(k) = y(k) - \hat{y}(k)$

Si $e(k) \approx 0$, $[w, b]^k = [w, b]^{k-1}$, alors arrêter : $[w, b]^k = [w, b]^{[*]}$

Sinon, ajuster les poids du SVR en utilisant l'algorithme récursif de programmation quadratique.

Fin k .

Avantages

-les méthodes séparatrices à vaste Marge possèdent une base mathématique rigoureuse très utile.

-Les SVR sont efficaces dans les espaces de grande dimension.

-Différentes fonctions noyaux peuvent être spécifiées.

2.5 Techniques adaptatives

Les méthodes PLS, SVR et NN présentées précédemment peuvent être utilisées sous une forme adaptative, ce qui est particulièrement nécessaire lorsque l'on considère que le système doit être modélisé dans le temps. Dans cette situation, la forme adaptative est réalisée par l'utilisation d'algorithmes de fenêtre glissante (Diagramme 4).

Si l'on considère un système défini par des données d'entrée et de sortie mesurées $X(t) \in \mathbb{R}^k$ et $y(t) \in \mathbb{R}$, pour prendre en compte la variabilité de $\langle X(t), y(t) \rangle$, on introduit une fenêtre glissante de largeur N et les données d'entrée et de sortie deviennent $\langle X(t - N : t), y(t - N : t) \rangle$ pendant un temps important $t = 1 : M$, M étant la longueur maximale de la fenêtre.

L'erreur de modélisation réelle est $e[t - N : t] = y_p[t - N : t] - y[t - N : t]$

La sortie du modèle prédit récursif en ligne qui prend en compte les erreurs de prédiction passées est donnée par $y[t - N + 1 : t + 1] = f(e[t - N : t], x[t - N + 1 : t + 1])$, où f est une fonction obtenue par un algorithme d'apprentissage SVR basé sur les entrées-sorties (voir [4]).

Algorithmes PLS, SVR et NN adaptatifs

La procédure de calcul est résumée comme suit :

Étape 1. Acquisition des données $\langle X(t), y(t) \rangle$, u_b , Z_b , choisir N (taille de la fenêtre) et saisir le maximum de la taille de la matrice en fonction du nombre de lignes et de colonnes M .

Étape 2. Calculer avec la boucle adaptative/sélectionner la méthode de calcul : PLS, SVR, NN

For $t = N : M$

$X(t - N : t), y(t - N : t)$

Exécuter la procédure de calcul de $\{PLS, SVR, NN\} \longrightarrow \hat{y}$ et $\varepsilon = y - \hat{y}$

Estimer les paramètres du modèle : $\min(\varepsilon^2) \longrightarrow \{w, c\}$

Fin.

La structure globale de ces techniques est donné par le diagramme suivant :

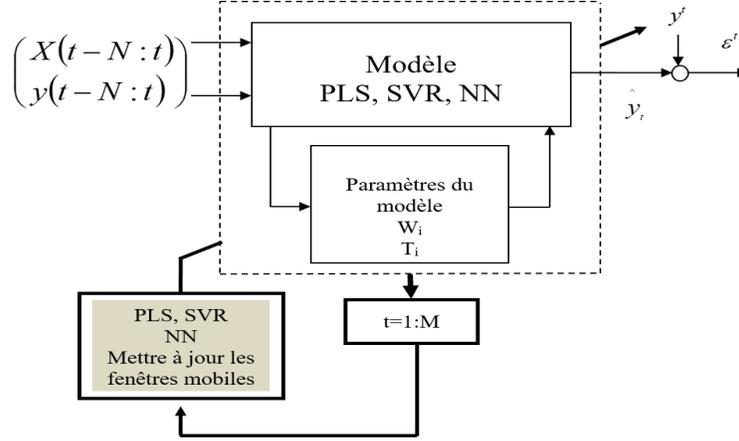


Diagramme 4 : Principe d'identification du modèle à l'aide des formes adaptatives PLS, SVR et NN

2.6 Méthode de vraisemblance résiduelle maximum

La méthode développée de vraisemblance résiduelle maximum connu sous l'appellation "Residual Estimation Maximum Likelihood" (REML) est utilisée pour modéliser les relations complexes entre les données de mesure qui sont collectées dans les conditions de répétabilité-reproductibilité [14, 75].

Considérons un système complexe caractérisé par différentes mesures qui est affecté par un comportement aléatoire. La fonction de distribution est définie comme suit :

$$f(y_i|\mu, \sigma^2 = 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y_i - \mu)^2}{2}}, \quad (2.53)$$

où y_i représente les valeurs observées répétitives ($i = N$ répétitions), σ est l'écart type, μ est la valeur moyenne et f est la fonction de distribution associée à la valeur observée. La fonction de vraisemblance pour toutes les observations d'un échantillon est le produit des vraisemblances :

$$L(\mu|y, \sigma^2) = \prod_{i=1}^N f(y_i|\mu, \sigma^2). \quad (2.54)$$

L'estimation de vraisemblance maximum (ML) de μ et σ est la valeur qui maximise cette fonction.

Cependant, il n'est pas pratique de prendre le produit d'un grand nombre de termes,

aussi, à la place, nous utilisons les propriétés de la fonction logarithmique pour obtenir :

$$\ln L(\mu|y, \sigma^2) = \sum_{i=1}^N \ln f(y_i|\mu, \sigma^2). \quad (2.55)$$

Extension de (2.55) à un modèle de régression de la forme $y = X\beta + \varepsilon$; où X est la matrice correspondante ($N \times p$) des p variables explicatives relatives aux N éléments de l'observation y , et β le vecteur de coefficients ($p \times 1$) de ces variables à effets fixes. En outre, le processus de génération des données est supposé être de la forme $Y_i \sim N(\mu, \sigma^2)$, c'est-à-dire un modèle linéaire gaussien.

On suppose que la valeur $y_i = X_i\beta$, et la variance σ^2 sont constantes mais inconnues, telles que :

$$L(\beta, \sigma^2|y, X) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - X_i'\beta)^2}{2\sigma^2}} = \frac{1}{(\sqrt{2\pi\sigma^2})^N} e^{-\frac{\sum (y_i - X_i'\beta)^2}{2\sigma^2}} \quad (2.56)$$

$$= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} e^{-\frac{(y - X\beta)'(y - X\beta)}{2\sigma^2}}. \quad (2.57)$$

En prenant les logarithmes, on obtient :

$$\ln(L(\beta, \sigma^2|y, X)) = -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} (y - X\beta)'(y - X\beta) \quad (2.58)$$

Si on met à zéro les dérivées premières de L par rapport à β et σ^2 , on écrit :

$$\frac{\partial L(\beta, \sigma^2|y, X)}{\partial \beta} = \frac{-1}{2\sigma^2} (2X'X\beta - 2X'y) = 0, \quad (2.59)$$

$$\frac{\partial L(\beta, \sigma^2|y, X)}{\partial \sigma^2} = \frac{-N}{2} \left(\frac{1}{\sigma^2} \right) + \frac{1}{2\sigma^4} (y - X\beta)'(y - X\beta) = 0. \quad (2.60)$$

Donc on obtient :

$$\beta = (X'X)^{-1}X'y, \quad (2.61)$$

$$\sigma^2 = \frac{(y - X\beta)'(y - X\beta)}{N} = \frac{\varepsilon'\varepsilon}{N}. \quad (2.62)$$

2.6.1 Vraisemblance maximum de la composante de variance dans le modèle mixte général

Soit Z la matrice de l'effet aléatoire, Z' sa transposée et G la matrice des composantes de la variance ; si l'on désigne par R les écarts résiduels, le modèle mixte général est donné par :

$$y = X\beta + Zu + e, \quad (2.63)$$

et nous supposons que $u \sim MVN(0, G)$ et $e \sim MVN(0, R)$ (MVN signifie distribution normale multivariée).

Dans le cadre de ce modèle, le processus de génération des données pour Y_{ij} est également supposé être normal multivarié avec une moyenne égale à $X\beta$ et la matrice de covariance Y_{ij} correspond aux valeurs observées obtenues dans les conditions de répétabilité et de reproductibilité.

La matrice de covariance globale a la forme :

$$V = ZGZ' + R, \quad (2.64)$$

la densité de probabilité de la donnée y est :

$$p(y|X\beta, V) = \left(\frac{1}{\sqrt{2\pi}}\right)^N \exp\left[-\frac{1}{2}(y - X\beta)'V^{-1}(y - X\beta)\right]. \quad (2.65)$$

L'étape suivante consiste à prendre le logarithme naturel, ce qui donne le logarithme de vraisemblance de β et V .

$$L(\beta, V|X, y) = -\frac{1}{2}[N\ln(2\pi) + \ln|V|(y - X\beta)'V^{-1}(y - X\beta)]. \quad (2.66)$$

Les composantes de variance que nous essayons d'estimer sont intégrées dans G et R , et nous supposons que $G = \sigma_A^2 A$; où A est la matrice de relation génétique additive, et $R = \sigma_E^2 I$; c'est-à-dire que les écarts résiduels des différents individus sont indépendants et homoscedastiques (c'est-à-dire que toutes ses variables aléatoires ont la même variance finie) .

En prenant la dérivée première par rapport à β , on obtient :

$$\frac{\partial L(\beta, V|X, y)}{\partial \beta} = \frac{\partial L [(y - X\beta)'V^{-1}(y - X\beta)]}{\partial \beta} = -X'V^{-1}(y - X\beta) - X(y - X\beta)'V^{-1}, \quad (2.67)$$

$$(y - X\beta)'V^{-1}(y - X\beta) = (y - X\hat{\beta})'V^{-1}(y - X\hat{\beta}) + (\hat{\beta} - \beta)'X'V^{-1}X(\hat{\beta} - \beta), \quad (2.68)$$

où $\hat{\beta}$ est l'estimation de β de la forme :

$$\hat{\beta} = \left(X'\hat{V}^{-1}X\right)^{-1} X'\hat{V}^{-1}y. \quad (2.69)$$

2.6.2 Estimation de vraisemblance résiduelle maximum (REML)

À partir du plan d'expérience et des valeurs observées Y_{ij} , nous définissons Z pour chaque Y_{ij} et la structure du modèle devient $Y = X\beta + Zu$, en utilisant le maximum de vraisemblance pour estimer β par (2.69).

Le schéma suivant définit la procédure de calcul du REML :

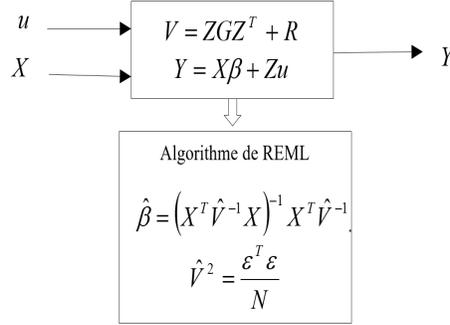


Diagramme 5 : Principe du modèle d'identification utilisant REML

Une forme adaptative du REML, basée sur des fenêtres mobiles dans le temps, peut être appliquée. Elle est généralement utilisée pour surmonter le système de variance temporelle. Dans ce cas, les entrées X et la sortie y sont mises en mémoire tampon comme suit :

$$X_b = X [t - N : t], y_b = y [t - N : t] \text{ et } u_b = u [t - N : t]. \quad (2.70)$$

Le modèle mixte s'écrit comme suit :

$$y_b = X_b \beta_b + Z_b u_b + \varepsilon_b. \quad (2.71)$$

De la même manière que celle définie dans la formulation ci-dessus, l'algorithme REML peut être facilement appliqué pour trouver V_b et β_b à partir de (2.64) et (2.69).

2.7 Distribution de l'évaluation de la fiabilité

2.7.1 Distribution normale

La distribution normale est l'une des distributions de probabilités continues les plus connues et les plus importantes utilisées en statistique ainsi que dans d'autres domaines scientifiques. Elle est souvent appelée distribution gaussienne [49, 63].

La fonction de densité de probabilité (pdf), $f(t)$ pour la distribution normale est définie par :

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \left(\frac{t - \mu}{\sigma} \right)^2}, \quad (-\infty < t < +\infty) \quad (2.72)$$

où μ et σ sont les paramètres de la distribution (μ = moyenne, σ = écart-type), et la fonction de densité cumulative (CDF) $F(t)$ pour la distribution normale est donnée

par :

$$F(t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \quad (2.73)$$

La fonction de fiabilité $R(t)$ pour une distribution normale est définie par l'équation suivante :

$$R(t) = \int_t^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \quad (2.74)$$

La transformation z est utilisée dans cette situation pour transformer la fonction de densité de probabilité (pdf) normale en une distribution dite pdf normale standard qui a une moyenne de zéro ($\mu = 0$) et un écart type de 1 ($\sigma = 1$). La transformation z est définie comme suit :

$$z = \frac{t - \mu}{\sigma}, \quad (2.75)$$

pour obtenir la fonction dite de densité cumulative

$$\varphi(Z) = \int_{-\infty}^Z \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \quad (2.76)$$

2.7.2 Probabilité de conformité et niveaux de risque associés

Pour évaluer la conformité des paramètres spécifiés, plusieurs éléments sont nécessaires, notamment : un intervalle $[T_L \ T_U]$ de valeurs admissibles de la quantité est spécifié (où T_L et T_U sont respectivement les limites de tolérance inférieure et supérieure, comme indiqué dans le diagramme 6) ; la capacité de mesurer la propriété et d'exprimer les résultats de la mesure conformément aux principes du Guide sur la mesure de l'incertitude [16, 60].

La probabilité que les mesures x soit dans l'intervalle $[a, b]$ est

$$Pr(a \leq x \leq b) = \int_a^b f(t)dt = F(b) - F(a), \quad (2.77)$$

où x est la variable d'entrée (température de la fonte, pression circulaire, température du vent chaud). A partir des étapes conduisant à l'expression (2.77), la probabilité pour $a \leq x \leq b$

$$Pr(a \leq x \leq b) = \varphi\left(\frac{b-x}{u}\right) - \varphi\left(\frac{a-x}{u}\right), \quad (2.78)$$

où u est l'incertitude.

Intervalles de tolérance unilatéraux avec des PDFs normaux

Limite inférieure de tolérance

D'après l'expression (2.78), avec $a = T_L$, $b \rightarrow \infty$; et en notant que $\varphi(\infty) = 1$, la probabilité de conformité est :

$$p_c = 1 - \varphi\left(\frac{T_L - x}{u}\right) \tag{2.79}$$

Comme $\varphi(t) + \varphi(-t) = 1$, la probabilité (2.79) est exprimée comme suit :

$$p_c = \varphi\left(\frac{x - T_L}{u}\right) \tag{2.80}$$

Limite supérieure de tolérance

D'après l'expression (2.78), avec $a \rightarrow \infty$, $b = T_U$, et en notant que $\varphi(-\infty) = 0$, la probabilité de conformité est :

$$p_c = \varphi\left(\frac{T_U - x}{u}\right) \tag{2.81}$$

Intervalles de tolérance bilatéraux avec des PDF normaux

L'utilisation de l'expression (2.78) avec $b = T_U$ et $a = T_L$ donne la probabilité de conformité :

$$p_c = \varphi\left(\frac{T_U - x}{u}\right) - \varphi\left(\frac{T_L - x}{u}\right) \tag{2.82}$$

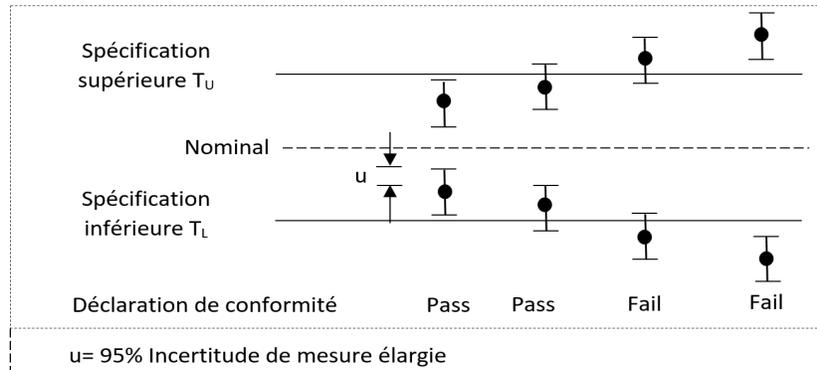


Diagramme 6 : Limite de tolérance bilatérale

Résultats mathématiques et numériques pour le contrôle qualité du métal chaud dans le haut fourneau

3.1 Introduction

Le haut fourneau est un système non linéaire extrêmement compliqué composé de plusieurs éléments spécifiques tels que l'équipement de chargement ; le circuit de refroidissement, toute la production de vent chaud, et le grand arbre cylindrique four.

Leur fonctionnement principal est le suivant : les matières premières solides comme le minerai de fer et le coke sont chargées en un mélange de couches successives par le haut du four. Sous l'action de leur poids particulier, elles descendent progressivement vers le fond du foyer jusqu'à ce qu'elles fondent ; cependant, les gaz de combustion chauds s'élèvent à travers la colonne des matériaux de combustion. Le métal en fusion, composé de fonte et de scories, s'écoule dans le creuset. À la fin de la combustion, les matériaux se divisent en deux éléments : le laitier en fusion d'un côté et le métal chaud en fusion de l'autre, qui s'accumulent en fonction de leurs propres masses spécifiques. Le laitier est évacué par le trou de coulée du laitier qui est plus haut que celui de la fonte [31, 65].

La complexité des processus de transfert de chaleur et de masse, combinée à une grande variété de réactions chimiques et à une pression élevée, rendent la modélisation des hauts fourneaux un problème extrêmement difficile [25]. Malgré les efforts des meilleurs chercheurs pour résoudre ce type de problèmes, des défis subsistent. La prédiction et le contrôle de la qualité des produits par la technologie de détection virtuelle est un outil clé en temps réel, et a été largement utilisé dans de nombreux processus de fabrication (voir [86, 90]). Habituellement, elle est divisée en deux groupes principaux, à savoir les méthodes pilotées par les données et les modèles de premier principe. Dans ce cas, il est très difficile de construire le premier

principe en raison de la grande taille du réacteur et des environnements opérationnels internes complexes, tandis que le modèle piloté par les données ne nécessite pas de connaissances préalables sur le fonctionnement complet du processus et est basé directement sur les données du processus [27]. Par conséquent, il a été appliqué efficacement à divers processus industriels, comme des exemples typiques des modèles de séries temporelles par [69], approche statistique multivariée (PLS, PCR).

Pour un aperçu des modèles discrets temporels pilotés par les données pour la prédiction dynamique de la teneur en silicium du métal chaud, le lecteur est invité à se reporter à [25] et [69], où ils ont proposé un modèle pour prédire le changement d'état thermique de la sole du haut fourneau avec une machine à vecteurs de support ; tandis que [44] ont développé un modèle adaptatif utilisant la régression par vecteurs de support (SVR) lisse à fenêtres glissantes pour un système de haut fourneau non linéaire.

De nombreux systèmes dynamiques sont caractérisés par un comportement dynamique non linéaire, tels que le haut fourneau, où des modèles non linéaires sont alors nécessaires. En effet, il a été démontré que le SVR et le réseau de neurones (NN) peuvent s'approcher des non-linéarités continues, et ont été appliqués à la modélisation de systèmes non linéaires complexes, dont la complexité est souvent due au nombre élevé de poids dans le réseau ; en plus du principe d'identification du modèle en utilisant le classique Recursive Least Square (RLS) et sa version adaptative. Plus de détails sur ces méthodes peuvent être trouvés dans différents documents, voir par exemple [25, 44] et leurs références. Dans cette étude, nous proposons une nouvelle approche pour le contrôle de la qualité de mesure du haut fourneau en utilisant des techniques de modélisation et d'identification non linéaires robustes connues sous le nom de méthodes guidées par les données ; avec une tentative de réduire les erreurs de modélisation et d'identification concernant la structure et les paramètres du modèle qui est inclus dans le budget d'incertitude [4].

3.2 Modélisation et identification

Ce type de systèmes est généralement modélisé par une relation théorique complexe entre leurs données d'entrée et de sortie, qui pourrait être donnée par :

$$\gamma = \phi(\chi) \tag{3.1}$$

où χ est l'espace d'entrée et γ la variable de sortie. Il convient de souligner qu'en raison du grand nombre de paramètres, la fonction ϕ a une forme complexe. Par souci de simplification, l'équation (3.1) peut être approximée par un modèle définissant la dépendance entre une donnée d'entrée et une donnée de sortie (X^t), (y^t), respectivement, tel que

$$y^t = f(X^t) + e^t \tag{3.2}$$

3.3. APPLICATION AU CONTRÔLE DE LA QUALITÉ DU MÉTAL CHAUD

où t est le temps, et $e^t \sim N(0, \sigma^2)$ est distribution normale de moyenne nulle du terme d'erreur obtenu dans l'étape de modélisation et l'écart type σ .

Le modèle f est entraîné, testé et développé à l'aide de données réelles. Les paramètres du modèle PLS, SVR, NN adaptatifs sont obtenus par une méthode d'identification utilisant une procédure adaptative basée sur les techniques de fenêtres mobiles.

3.3 Application au contrôle de la qualité du métal chaud

Dans cette section, nous présentons le schéma, le processus et la structure d'un haut fourneau comme le montrent les diagrammes 7 et 8. Le modèle des entrées et sorties expérimentales est présenté dans le tableau 3.1 et la figure 3.1 tableau 3.1 définit les paramètres de processus d'entrée et des données de sortie. La structure du modèle donnée dans la figure 3.1 montre l'influence de l'espace d'entrée sur la sortie. Les entrées du modèle sont le débit de gaz naturel, le débit de vent thermique et la pureté de l'oxygène comme indiqué sur la figure 3.1 (Fig.1a-Fig.1c), tandis que le modèle de sortie est la température du métal chaud (voir figure 3.1 Fig1d) [4].

Schéma du haut fourneau avec ses zones et résultats expérimentaux.

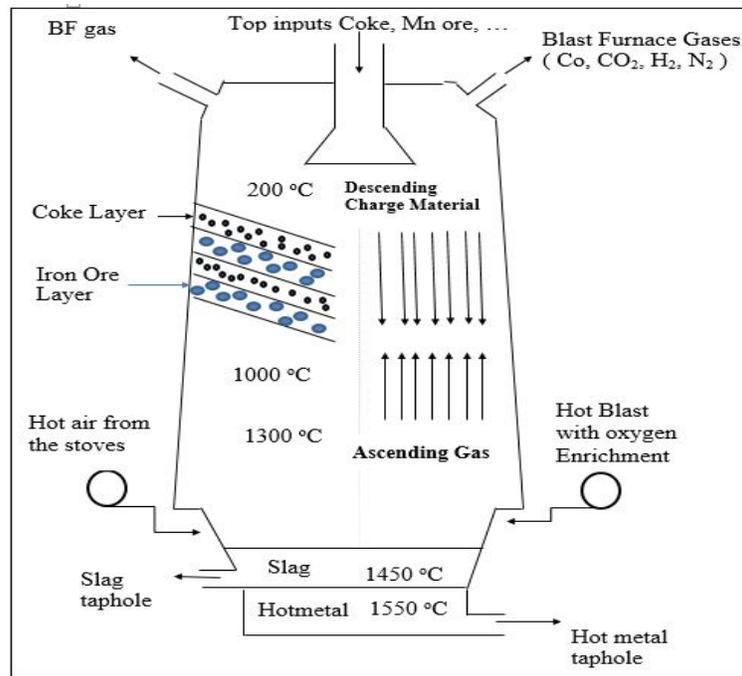


Diagramme 7 : Schéma du haut fourneau

3.3. APPLICATION AU CONTRÔLE DE LA QUALITÉ DU MÉTAL CHAUD

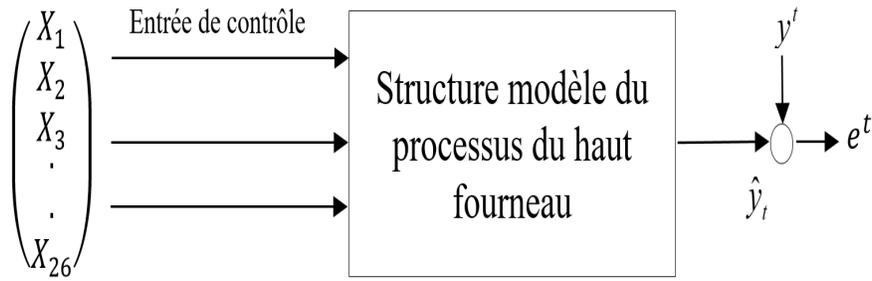


Diagramme 8 : Structure de Modèle

Variables	Désignation	Unité	Variables	Désignation
Paramètres de marche			Résultats d'analyse	
X_1	Pression circulaire	bar	X_{11}	<i>Fe</i> agglo
X_2	Contre pression	bar	X_{12}	<i>Fe</i> pellets
X_3	T. du vent chaud	°C	X_{13}	<i>Fe</i> min. cal.
X_4	T. de la fonte	°C	X_{14}	<i>Fe</i> charge
X_5	Vapeur	T/h	X_{15}	<i>Fe</i> fonte
X_6	T. de la flamme	°C	X_{16}	<i>Fe</i> fonte+perles
X_7	Vitesse de vent	m/s	X_{17}	H_2O coke
			X_{18}	H_2O pellets
Injections entrées			X_{19}	<i>CaO</i> agglo
X_8	Gaz naturel		X_{20}	<i>SiO₂</i> agglo
X_9	Débit du vent chaud	$Kn\text{m}^3/h$	X_{21}	<i>CaO</i> pel
X_{10}	O_2		X_{22}	<i>SiO₂</i> pel
			X_{23}	<i>CaO</i> min. cal
			X_{24}	<i>SiO₂</i> min. cal
			X_{25}	<i>CaO</i> Laitier
			X_{26}	<i>SiO₂</i> Laitier
Sortie				
y	T. de la fonte			

TABLE 3.1 – Nomenclature des données d'entrée et de sortie

Où T. désigne la température.
 Notre objectif est de trouver un modèle mathématique d'entrée/sortie $y = f(X_1, \dots, X_n)$.

3.3. APPLICATION AU CONTRÔLE DE LA QUALITÉ DU MÉTAL CHAUD

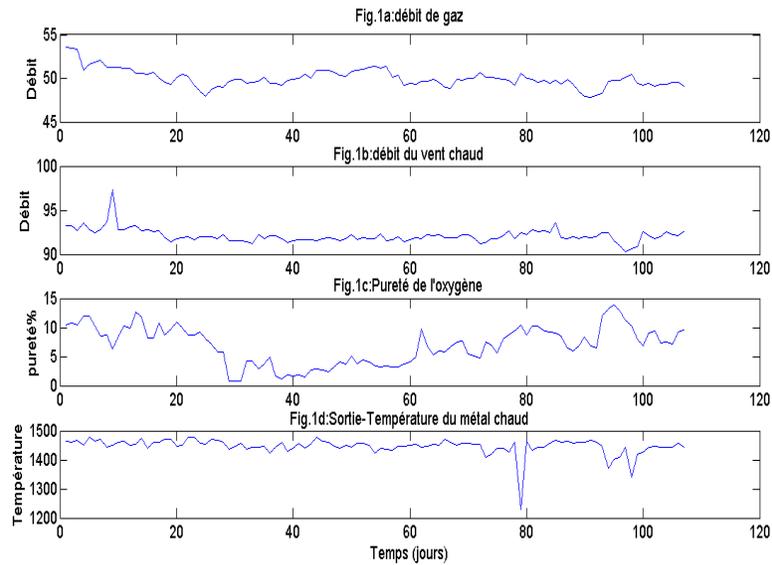


FIGURE 3.1 – Entrées et sorties expérimentales des modèles

	Gaz naturel	Débit de vent chaud	Pureté de l'oxygène	T. de la fonte
Moyenne	109.58	145420.31	6.93	1446.50
Médiane	122.34	150385	7.12	1451
Mode	168.96	160000	8.64	1443
Écart-type	48.61	16997.06	3.22	29.04
Variance	2363.48	288900367.5	10.43	843.45
Rang	168.96	109167	13.11	247
Minimum	0	55833	0.7	1231
Maximum	168.96	165000	13.81	1478

TABLE 3.2 – Description statistique des données collectées

3.3.1 Expérimentation numérique

Dans cette section, nous présentons les résultats obtenus par les techniques qui ont été discutées dans le chapitre précédent.

Résultats obtenus

Modélisation : l'étape de modélisation est réalisée par une série de données de mesure pour identifier les paramètres du modèle qui seront utilisés pour la prédiction. Les résultats de modélisation des différentes méthodes PLS, SVR, NN et leurs versions adaptatives sont présentées dans la figure [3.2](#).

3.3. APPLICATION AU CONTRÔLE DE LA QUALITÉ DU MÉTAL CHAUD

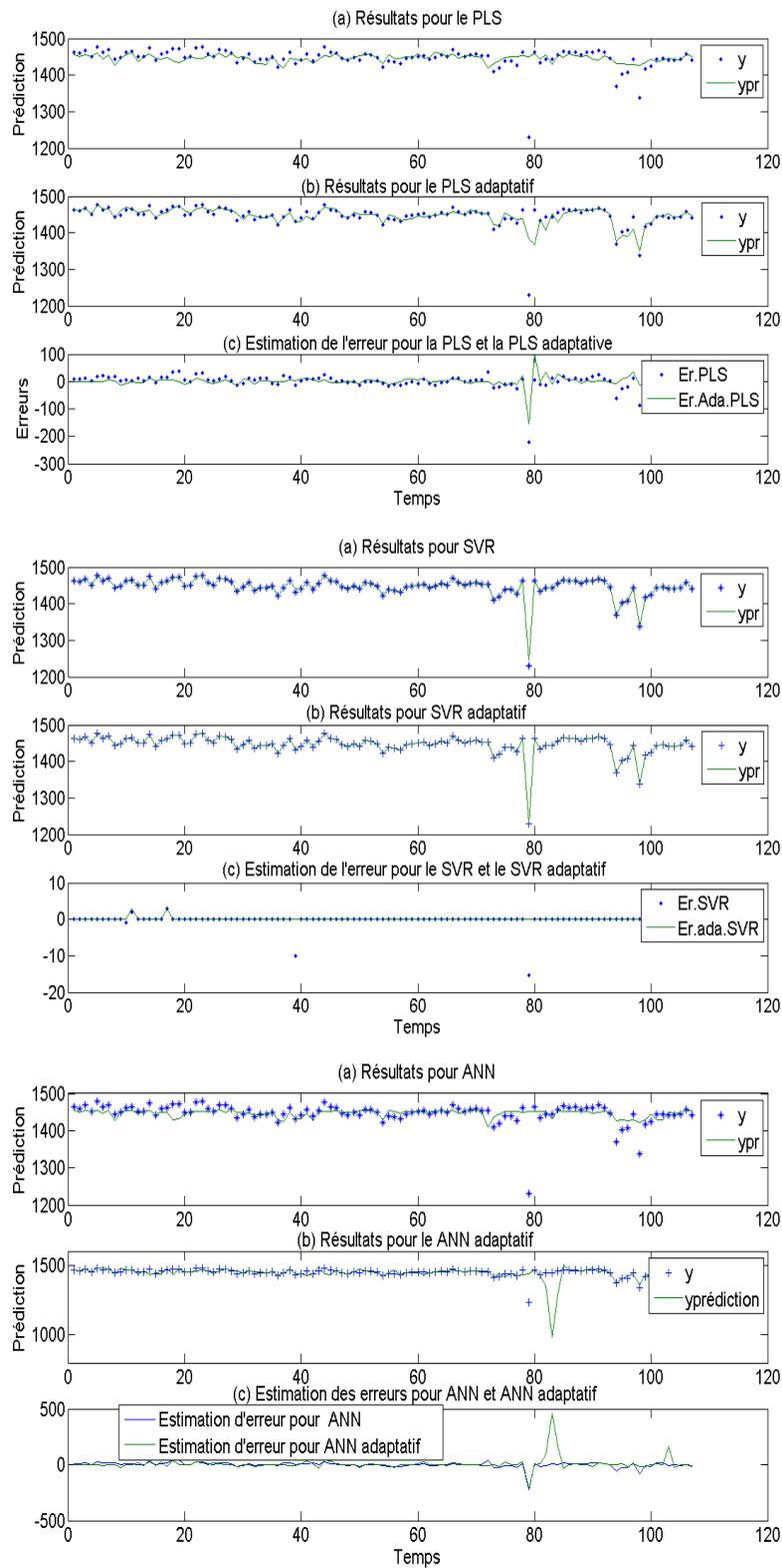


FIGURE 3.2 – Prédiction de la température de la fonte et des erreurs

3.3. APPLICATION AU CONTRÔLE DE LA QUALITÉ DU MÉTAL CHAUD

	Gaz naturel	Débit de vent chaud	Pureté de l'oxygène	T. de la fonte
Gaz naturel	1	0.7286	0.0526	0.2949
Débit de vent chaud	0.7286	1	0.0165	0.2790
Pureté de l'oxygène	0.0526	0.0165	1	-0.1265
T. de la fonte	0.2949	0.2790	-0.1265	1

TABLE 3.3 – Matrice de corrélation

Prédiction

La structure du modèle et les paramètres qui ont été déterminés dans le processus de modélisation pour les différents méthodes NN, SVR, PLS et leurs versions adaptatives ont été appliqués à de nouveaux ensembles de données pour la prédiction (données stables). Les résultats de la prédiction de la température du métal chaud sont illustrés dans la figure 3.3, et les performances de modélisation et de prédiction, caractérisées par l'écart type des erreurs et d'autres indicateurs de performance, sont présentées dans le tableau 3.4.

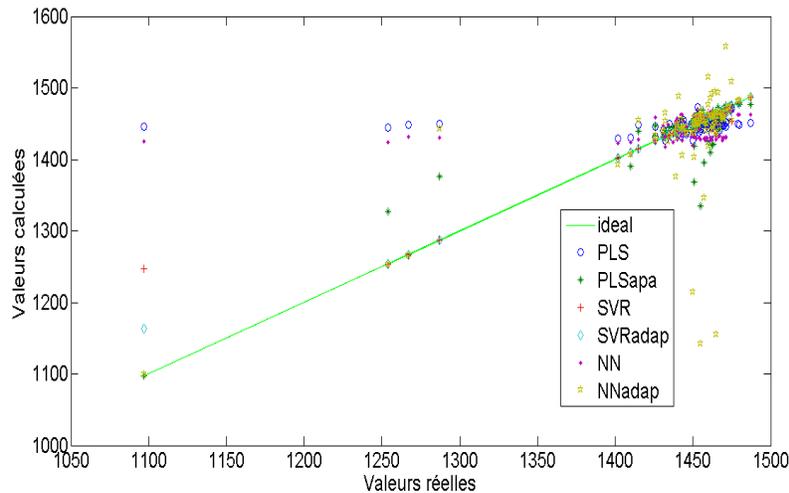


FIGURE 3.3 – Résultats calculés et prédits pour des données stables

Commentaires

Les résultats de la prédiction de la température de la fonte en fonction du débit d'oxygène, de la pureté de l'oxygène et du débit de vent chaud pour le processus du haut fourneau sont donnés dans la figure 3.2 et la figure 3.3, de haut en bas comme suit :

La prédiction de la température est effectuée par le modèle PLS et sa version adapta-

3.3. APPLICATION AU CONTRÔLE DE LA QUALITÉ DU MÉTAL CHAUD

Méthodes	STD Apprentissage	STD Prédiction	Observations
PLS	27.5812	47.4724	Double apprentissage
PLSadap	19.6558	21.9312	presque équivalent à l'apprentissage
SVR	1.8029	14.6761	plus important que d'apprendre
SVRadap	0.4082	6.4565	Pas grande différence
NN	26.2949	46.3315	double apprentissage
NNadap	25.0469	34.7740	Pas grande différence

TABLE 3.4 – Indicateurs de performances des différentes méthodes

tive correspondante. Cette dernière approche donne de meilleurs résultats de l'erreur prédite qui est plus faible que le PLS conventionnel. Ceci est évident car le modèle adaptatif effectue une correction immédiate en obtenant une nouvelle projection à chaque itération.

Comme l'approche SVR est basée sur l'algorithme de Vapnik, sa version adaptative est une exécution de cet algorithme conventionnel à chaque instant en utilisant un mode de fenêtre glissante. Comme pour l'algorithme PLS, la version adaptative du SVR se comporte mieux. Cependant, par rapport au PLS, l'approche SVR donne de meilleurs résultats dans les deux cas (conventionnel et adaptatif), ce qui est lié à la nature de l'algorithme de Vapnik qui est basé sur la programmation quadratique complexe, alors que le PLS est juste une technique de projection.

Pour la technique NN, la fonction d'activation est LOGSIG et la taille de la couche cachée utilisée dans le réseau neuronal artificiel est de 5 neurones, c'est-à-dire une seule couche cachée du réseau neuronal. Le modèle de réseau de neurones artificiels adaptatif ne fonctionne pas très bien, ce qui conduit à des valeurs très bruitées, l'implémentation numérique est encore à l'étude.

Sur la base des résultats obtenus (voir figure 3.2 et 3.3 et tableau 3.4), le modèle SVR donne la plus grande précision de prédiction par rapport aux approches proposées PLS et NN et il peut être qualifié de meilleur outil. Le modèle NN peut ne pas être recommandé, en particulier dans le cas de données bruitées, car il induit une très grande incertitude 4.

Les versions adaptatives sont implémentées de la même manière en utilisant des fenêtres glissantes comme le montre le Diagramme 4.

Il faut noter que dans le tableau 3.4, PLSadap, SVRadap, et NNadap sont les versions adaptatives de PLS, SVR et NN, respectivement.

3.4 Application de l'analyse multivariée des risques

Dans cette section, nous présentons la structure du modèle d'évaluation des risques, comme le montrent les diagrammes 9 et 10. Le modèle des données expérimentales est présenté dans la figure 3.4. Les tableaux 3.5 et 3.6 présentent respectivement les statistiques descriptives des échantillons collectés et la matrice de corrélation.

Risque additif : Il s'agit d'une opération d'addition de la probabilité de défaut des différentes causes, appliquée si au moins une cause peut générer un défaut,

Risque multiplicatif : Il s'agit d'une opération de multiplication de la probabilité de défaut des différentes causes, qui s'applique si toutes les causes peuvent générer un défaut,

Risque global : il s'agit d'une relation combinée entre les risques additifs et/ou multiplicatifs.

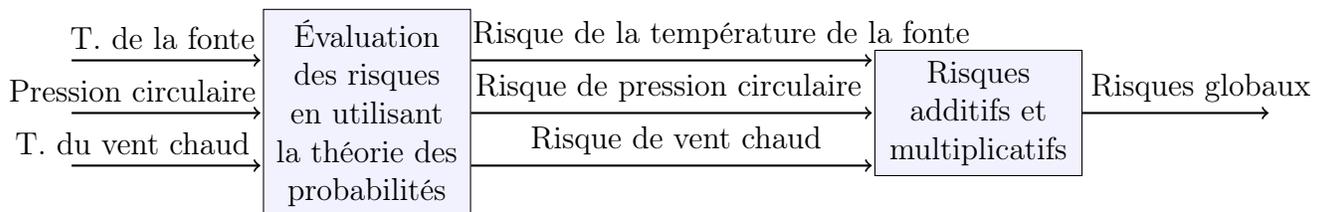


Diagramme 9 : Structure du modèle en utilisant les distributions de probabilité

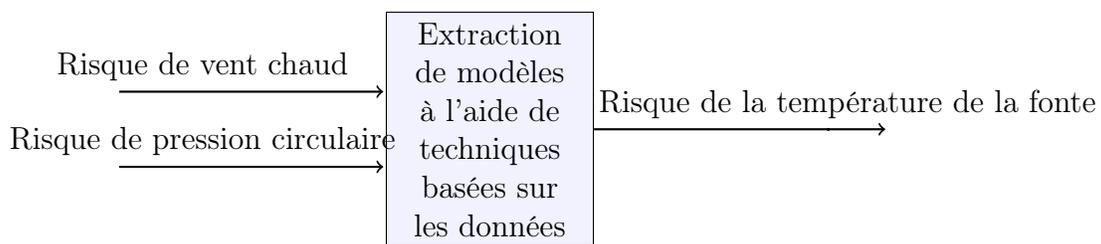


Diagramme 10 : Structure du modèle en utilisant des techniques basées sur les données

3.5 Expérimentation numérique

Dans cette section, nous présentons les résultats obtenus par les techniques qui ont été discutées dans les sections précédentes.

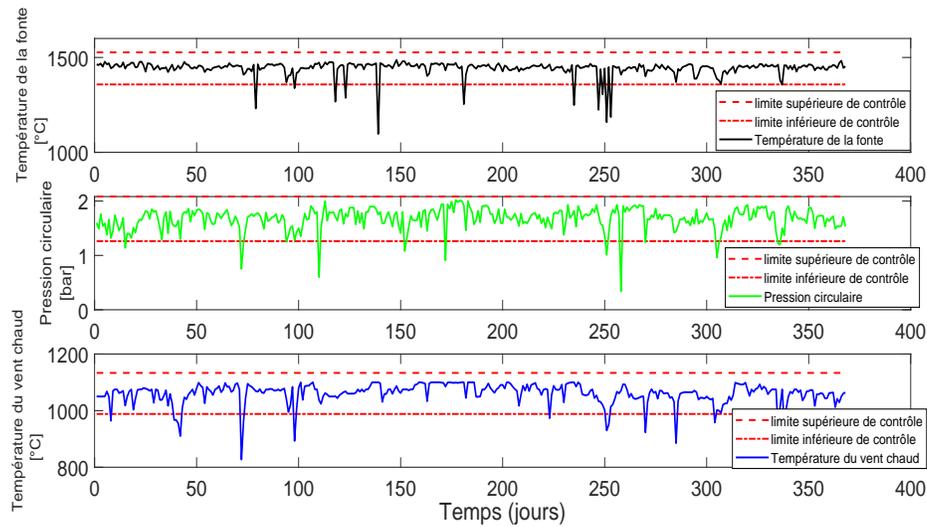


FIGURE 3.4 – Modèle des entrées expérimentales

Statistiques descriptives	Paramètres		
	Température de métal chaud	Pression circulaire	Température du vent chaud
Moyenne	1446.50	1.6	1055.20
Médiane	1451	1.66	1070
Mode	1443	1.54	1075
Écart-type	29.04	0.24	43.69
Variance	843.45	0.059	1909.03
Rang	247	1.91	273
Minimum	1231	0	827
Maximum	1478	1.91	1100

TABLE 3.5 – Statistiques descriptives des échantillons collectés.

3.5.1 Résultats et discussion

Discussion : Cette étude d'évaluation des risques a été élaborée par une combinaison de plusieurs techniques :

1. La première approche consiste à évaluer les différents risques qu'il y a à dépasser les limites en utilisant la théorie des probabilités et en supposant que les lois de distribution sont normales pour chaque paramètre,
2. Comme le processus est complexe et qu'il contient plusieurs paramètres, cette approche est complétée par la recherche d'une relation complexe entre les différents risques générés et la situation finale, c'est-à-dire le risque global,

Les notions de risques additifs ou multiplicatifs sont introduites pour générer le risque global en fonction du type de processus (parallèle ou en série), c'est-à-dire qu'un risque peut résulter de la combinaison d'autres risques liés à la complexité du système.

3.5. EXPÉRIMENTATION NUMÉRIQUE

	T. de metal chaud	Pression circulaire	T. du vent chaud
T. de metal chaud	1	0.1173	0.2453
Pression circulaire	0.1173	1	0.4675
T. du vent chaud	0.2453	0.4675	1

TABLE 3.6 – Matrice de corrélation

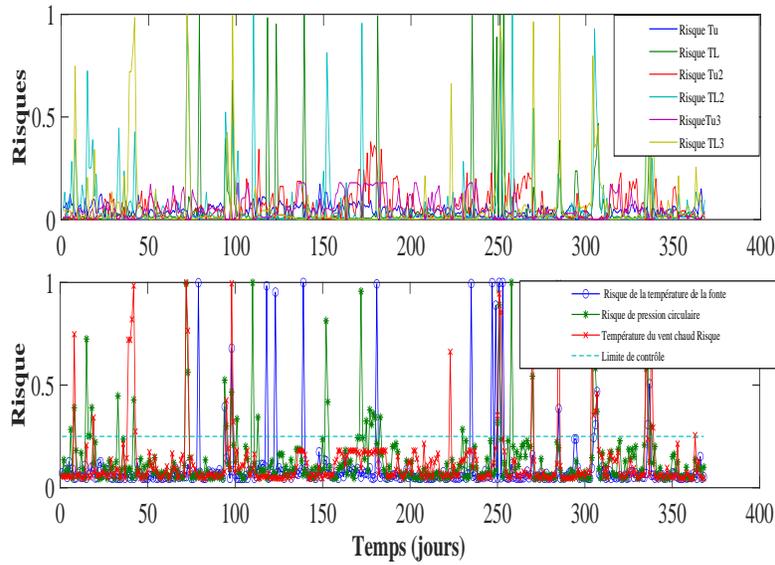


FIGURE 3.5 – Risques des différents paramètres

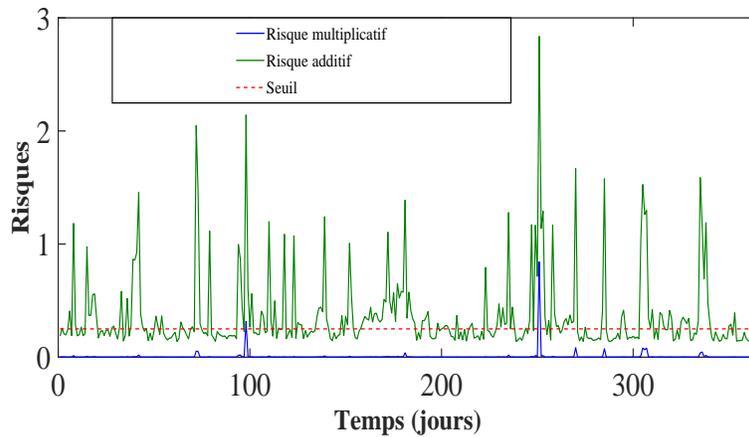


FIGURE 3.6 – Risques globaux additifs et multiplicatifs

Risques	Décision	Niveau de risque
T. de metal chaud	Non conformité	risque élevé
Pression circulaire	Non conformité	risque élevé
T. du vent chaud	Non conformité	risque élevé

TABLE 3.7 – Décision de conformité et de non-conformité

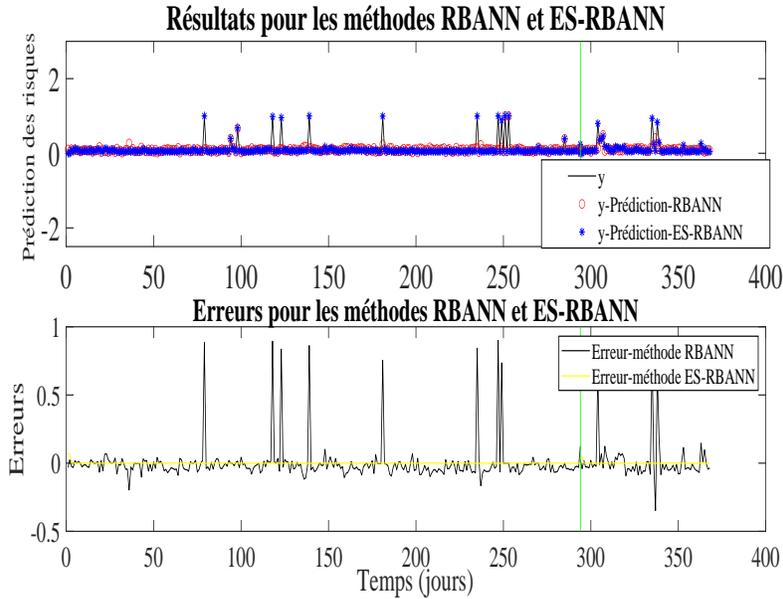


FIGURE 3.7 – Modélisation et prévision du risque et des erreurs de température de la fonte

Méthodes	Indicateurs de performance	
	STD(apprentissage)	STD(Test)
RBANN	0.1339	1.2815
ES-RBFANN	0.0057	0.0085

TABLE 3.8 – Indicateurs de performance des méthodes RBANN et ES-RBFANN

Le but majeur de ce travail est d'évaluer le risque multivarié dans le processus du haut fourneau. En outre, un autre objectif de cette étude est d'étudier la performance des méthodes conventionnelles et avancées pour construire des modèles crédibles qui peuvent facilement prévoir le risque lié à la température de la fonte, car il s'agit évidemment de l'un des paramètres cruciaux nécessaires pour estimer la qualité de la fonte dans le processus du haut fourneau. Pour cette raison, deux techniques différentes RBANN et ES-RBANN ont été appliquées pour générer des modèles fiables,

afin d'améliorer les phases de formation et de test, un nombre modeste d'échantillons (365 tests) et plusieurs variables d'entrée ont été utilisés pour cette étude. Enfin, une étude comparative entre ces techniques est réalisée et le meilleur modèle est sélectionné à l'aide de l'indicateur de performance de standard déviation (STD) (voir tableau 3.8).

Les données d'entrée/sortie utilisées dans la modélisation sont présentées dans la figure 3.4. RBANN et ES-RBANN sont des techniques multivariées efficaces, utilisées pour traiter des ensembles de données complexes, comme décrit dans les sections 4.1 et 4.2.

Les figures 3.5 et 3.6 présentent les risques supérieur/inferieur et multiplicatif/additif global, respectivement. La figure 3.7 présente les résultats de l'analyse de régression en deux parties : modélisation et prédiction à l'aide de différentes méthodes développées par exemple RBANN et ES-RBANN.

Pour la technique RBANN, les fonctions d'activation étaient Radbas et la taille de la couche cachée utilisée dans le réseau neuronal artificiel est de 25, c'est-à-dire une seule couche cachée du réseau de neurones, où l'algorithme utilisé est celui de Levenberg Marquardt, cependant dans le cas de la technique ES-RBANN, l'alpha est de 0,001.

3.6 Conclusion

Dans ce chapitre, on a exploré trois techniques et leurs versions adaptatives, où on présente les prédictions de température et leurs estimations d'erreur correspondantes. La prédiction de la température de la fonte brute en fonction du débit d'oxygène, de la pureté de l'oxygène et du débit de vent chaud pour le processus de haut fourneau est la solution de l'équation (2.44), où les résultats sont présentés dans le tableau 3.4 et les figures 3.2 et 3.3. Comme indiqué dans la section précédente, les approches adaptatives donnent de meilleurs résultats de l'erreur prédite que les approches conventionnelles dans les deux techniques PLS et SVR ; cependant, le SVR est plus compétitif que le PLS. Les résultats de la prédiction de la température de la fonte à l'aide de l'approche SVR sont obtenus par l'utilisation d'une version modifiée basée sur l'algorithme de Vapnik [85]. Une version adaptative de ce dernier est détaillée dans le diagramme 4. Comme dans le cas du PLS, elle donne également de meilleurs résultats, mais une grande incertitude est constatée. Les indicateurs pour les approches proposées sont donnés dans le tableau 3.4 et la figure 3.2. Par conséquent, nous concluons que les approches de modèles adaptatifs sont plus précises et peuvent être recommandées, plus particulièrement elles conduisent à une plage d'incertitudes réduite dans la prédiction par rapport au modèle des moindres carrés partiels car tous les changements d'entrée sont pris en compte ; alors que le modèle NN peut ne pas être recommandé, particulièrement dans le cas de données bruitées. En outre, les modèles peuvent être combinés avec une analyse de sensibilité

par l'utilisation de la technique de simulation de Monte Carlo.

Dans la section consacrée à l'évaluation des risques, une distribution des probabilités et une nouvelle technique d'apprentissage automatique sont proposées pour l'estimation des risques multivariés dans le processus du haut fourneau. En particulier, les risques liés à la température de la fonte, à la pression circulaire et à la température du vent chaud. Ce travail est une contribution au système de contrôle pour détecter le niveau de risque, par conséquent promouvoir la précaution en matière de sécurité et réduire le système de défaillance. Nos principales déductions sont les suivantes :

1. L'estimation des risques à plusieurs variables et le système de défaillance sont réalisés à l'aide de modèles de distribution statistique comme la distribution normale, combinés à un apprentissage automatique amélioré,
2. Les résultats obtenus indiquent que la technique ES-RBANN proposée donne une incertitude plus précise par rapport à la méthode RBANN,
3. Les modèles développés sont validés à l'aide de données réelles.

Méthode d'analyse de variance pour les données de contrôle qualité

4.1 Introduction

Les modèles de variance pour le contrôle et l'estimation de la qualité est un outil fondamental et important pour l'analyse des données de mesure et l'estimation des incertitudes. Des mesures brutes fournies dans le monde réel à la détermination des propriétés finales de la série de mesures, l'analyse et l'évaluation des données nécessitent une amélioration continue pour la prises de leurs décisions. Dans le domaine du contrôle qualité, le processus de mesure est analysé et modélisé à l'aide des outils et des méthodes appropriés. En fonction de leur complexité, le schéma de variance est défini et testé pour des données d'un problème agro-industriel visant à caractériser certaines propriétés lors du mélange de farine. Il y a plusieurs applications des approches conventionnelles basées sur l'analyse de la variabilité ; notre contribution en est une amélioration.

4.1.1 Quelques approches mathématiques

Il existe de nombreux outils mathématiques largement utilisés, tels que les procédures standard et de calcul pour évaluer la variance des mesures. Les plus populaires sont les normes ISO 5725 et ISO 21748 (voir [40] et [41]) et leurs techniques connexes comme le Guide de la mesure d'incertitude (GUM) [47]. Dans ce travail, la nouvelle version de la norme ISO 5725-2 basée sur le maximum de vraisemblance résiduelle (REML) [53], [33] et [57] est analysée et mise en oeuvre, où un modèle mixte entre des facteurs fixes déterministes et des facteurs aléatoires est considéré. Une telle approche améliore l'analyse de la variance en détectant les états de mesure constants et aléatoires.

Ce modèle mixte, pris comme une méthode d'analyse de la variabilité, pour laquelle une importance particulière est accordée aux changements de données et à l'analyse

de la corrélation entre les variables clés. Il s'agit généralement d'un outil puissant utilisé pour trouver des relations de cause à effet.

Le contrôle statistique multivarié des processus (MSPC) et les méthodes connexes sont des techniques souvent utilisées pour l'analyse de la variabilité, cependant des méthodes spécifiques telles que l'analyse en composantes principales (PCA) et les moindres carrés partiels (PLS) sont des méthodes qui se sont avérées utiles pour l'analyse des données, dans de nombreuses applications industrielles avec succès, voir par exemple [42] et [80].

Dans cette partie de notre travail, nous introduisons un modèle mixte et utilisons l'algorithme REML pour sélectionner le modèle de processus de mesure [5], [59] et [76]; pour lequel une analyse et une étude comparative avec la méthode conventionnelle d'analyse de variance sont présentées et des résultats compétitifs sont obtenus. La méthode REML développée est utilisée pour modéliser les relations complexes entre les données de mesure recueillies dans des conditions de répétabilité et de reproductibilité. Il est important de souligner qu'en raison de facteurs exogènes affectant le processus de mesure, les données répétées et reproductibles sont aléatoires, ce qui conduit à un mélange entre les facteurs statiques et dynamiques. Notre motivation à utiliser le modèle mixte est principalement justifiée par :

1. En pratique, les systèmes sont multivariés, ce qui nécessite l'utilisation d'un formalisme approprié,
2. La nature de l'application industrielle présente des non-linéarités, nécessitant l'utilisation de degrés de non-linéarité appropriés,
3. L'implémentation numérique devient relativement facile.

Le modèle mixte proposé est réalisé en appliquant les étapes suivantes :

- Collection de données : les données principales et importantes sont collectées à partir du processus de mesure selon un plan d'expérience, tous les facteurs d'influence sont pris en compte; les données sont visualisées et analysées, et les valeurs aberrantes sont exclues par une correction expérimentale,
- Recherche de corrélations entre les variables : Les données d'entrée sont utilisées pour la recherche de corrélation par le paramétrage du modèle REML en tenant compte des interactions possibles,
- Simulation numérique : le modèle mixte basé sur l'algorithme REML est mis en oeuvre.
- Évaluation globale : une comparaison des résultats obtenus par les deux méthodes développées (RLS, REML) est effectuée.

Dans ce qui suit on présente le développement des méthodes RLS et REML pour l'identification et l'analyse de la variabilité du gonflement de la pâte. La structure du modèle associé à l'analyse de la variabilité est établie. Les principaux facteurs ayant un impact sur la justesse ont été identifiés; les résultats de la simulation sont présentés. Une étude comparative entre les résultats obtenus par l'approche conventionnelle basée sur les techniques d'identification du modèle par RLS et la

méthode REML avec des résultats numériques efficaces est visualisée graphiquement (figures 4.1-4.4).

4.2 Application au contrôle qualité en agro-industrie

Dans cette section, on va présenter le processus de test de la farine, Nomenclature des données d'entrée et de sortie, descriptions statistiques des échantillons collectés, et les résultats obtenus par les trois méthodes détaillées ci-dessus. La structure du modèle de la farine d'essai est illustrée dans le Diagramme 11.

Les entrées du modèle sont la possibilité de gonflement de la pâte L02*, la résistance de la pâte à la déformation P02*, et la mesure de l'équilibre entre la ténacité et la capacité d'étirement P02/L02*. L'évolution typique de ces variables est illustrée par la figure 4.1 (Fig. 1a-1c), tandis que la sortie du modèle est le gonflement de la pâte (Voir Fig.1d). Le processus de la farine d'essai est illustré dans le Diagramme 11 :

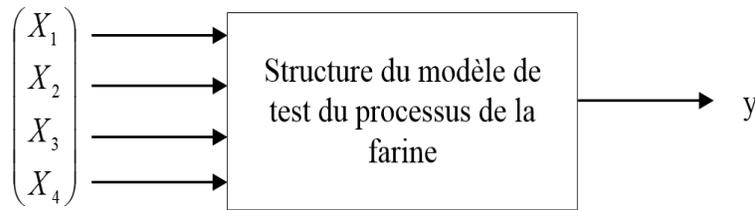


Diagramme 11 : Structure du modèle

Variables	Désignation	Nomenclature
Entrées	X ₁ Humidité	EAU02*
	X ₂ Résistance de la pâte à la déformation	P02*
	X ₃ Possibilité de gonflement de la pâte	L02*
	X ₄ Mesure de l'équilibre entre la ténacité et l'extensibilité	P/L02*
Sortie	Y Gonflement de la pâte	G02*

TABLE 4.1 – Nomenclature des données d'entrée et de sortie

Le but est de trouver un modèle mathématique $y = f(X_1, \dots, X_n)$.

4.2.1 Expérimentation numérique

Dans cette section, nous présentons les résultats obtenus par les techniques qui ont été discutées dans les sections précédentes.

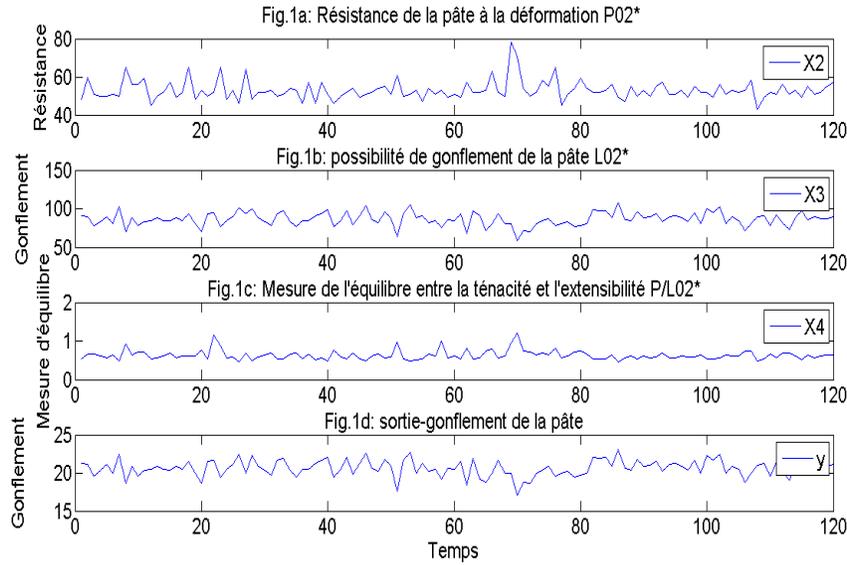


FIGURE 4.1 – Modèle des entrées et sorties expérimentales

	P02*	L02*	P02*/L02*	G02*
Moyenne	53.0083	86.4167	0.6290	20.6033
Médiane	52	86.5	0.6	20.6
Mode	52	88	0.6	20
Écart-type	5.0932	8.8428	0.1264	1.0725
Variance	25.9411	78.1947	0.016	1.1502
Asymétrie	1.7479	-0.2522	1.9773	-0.3623
Minimum	43	59	0.45	17.1
Maximum	78	108	1.2	23.1

TABLE 4.2 – Descriptions statistiques des échantillons collectés.

Résultats obtenus

Modélisation : l'étape de modélisation est effectuée par une série de données de mesure afin d'identifier les paramètres du modèle qui seront utilisés pour la prédiction. Les résultats de la modélisation pour RLS et REML sont présentés dans la figure 4.2 et la figure 4.3, respectivement.

Prédiction : de nouvelles séries de données ont été utilisées pour la prédiction, en utilisant la structure du modèle et les paramètres qui ont été identifiés dans l'étape de la modélisation pour les différents algorithmes RLS, REML et REML adaptatif. Les résultats obtenus pour le gonflement de la pâte sont résumés dans la figure 4.4 et les performances obtenues qui sont caractérisées par l'écart type des erreurs et des

4.2. APPLICATION AU CONTRÔLE QUALITÉ EN AGRO-INDUSTRIE

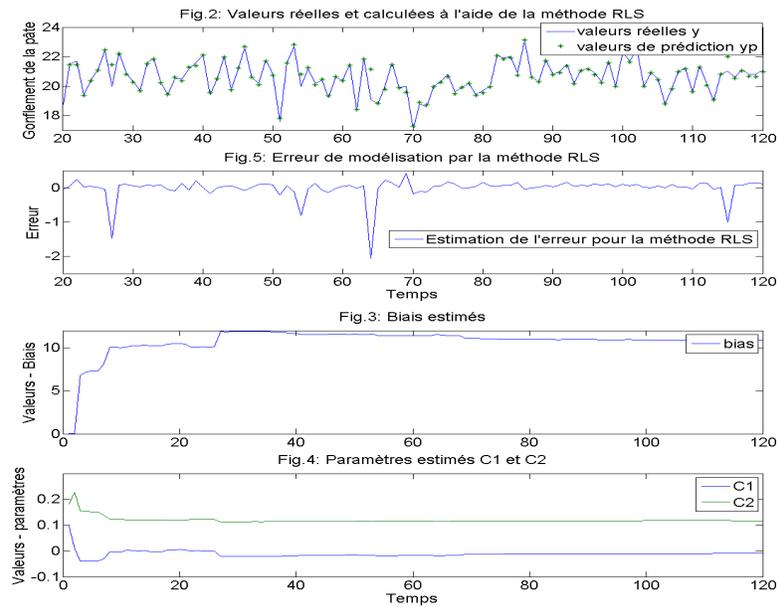


FIGURE 4.2 – Résultats de prédiction et erreur de modélisation de la méthode RLS

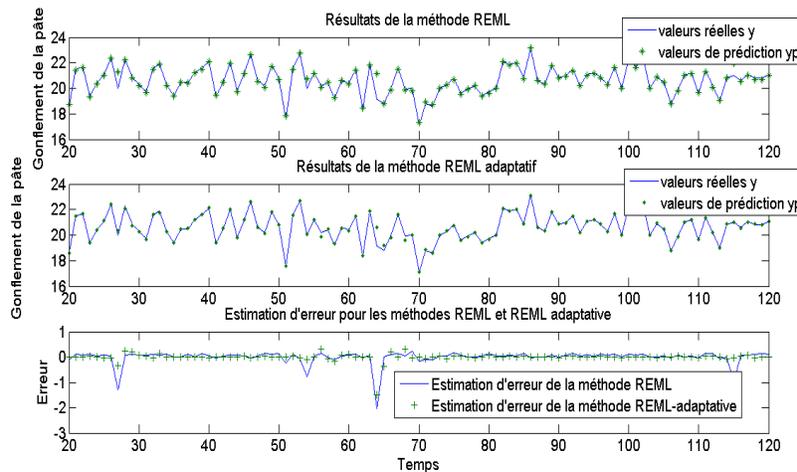


FIGURE 4.3 – Résultats de prédiction et erreur de modélisation des méthodes REML et REML adaptative

autres indicateurs de performance dans la modélisation et dans la prédiction sont reprises dans le tableau 4.3. Le tableau 4.3 présente les résultats des tests des différents modèles, la précision a été évaluée à travers divers indicateurs de performance statistique. Le modèle REML adaptatif a la meilleure précision dans la modélisation, suivi par le REML et le RLS. Les indicateurs statistiques sont standard deviation (STD), Efficacité de Nash-Sutcliffe (NSE), Index de l'accord (IOA), et Index de dispersion (IOS). Ils sont exprimés dans les équations suivantes :

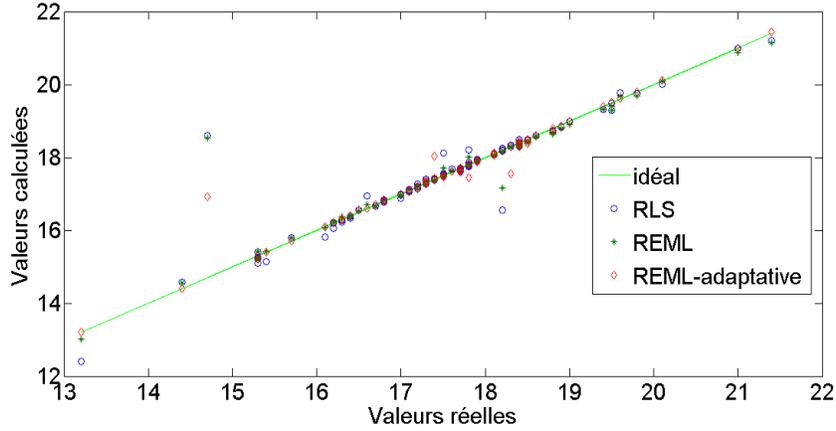


FIGURE 4.4 – Valeurs réelles et calculées du gonflement de la pâte

Méthodes	RLS	REML	adaptive REML
Incertitude de modélisation (STD)	0.2996	0.2836	0.1722
Efficacité de Nash-Sutcliffe (NSE)	0.9278	0.9354	0.9761
Indice de l'accord (IOA)	0.9997	0.9997	0.9999
Indice de diffusion (IOS)	0.0145	0.0137	0.0083
Incertitude de prédiction (STD)	0.4438	0.4040	0.2473
Efficience de Nash-Sutcliffe (NSE)	0.8735	0.8953	0.9607
Index de l'accord (IOA)	0.9994	0.9995	0.9998
Index de dispersion (IOS)	0.0251	0.0228	0.0140

TABLE 4.3 – Indicateurs de performance statistique des différentes méthodes.

1- Standard deviation (STD) :

$$STD = \left(\frac{1}{N} \sum_{i=1}^N (y_{pred,i} - y_{out,i})^2 \right)^{\frac{1}{2}}, (0 < STD < \infty) \quad (4.1)$$

où N est le nombre d'échantillons, $y_{out,i}$ est la valeur réelle. et $y_{pred,i}$ est la valeur prédite.

2- Efficience de Nash-Sutcliffe (NSE) :

$$NSE = 1 - \frac{\sum_{i=1}^N (y_{pred,i} - y_{out,i})^2}{\sum_{i=1}^N (y_{pred,i} - \bar{y}_{pred})^2}, (-\infty < NSE < 1) \quad (4.2)$$

3- Index de l'accord (IOA) :

$$IOA = 1 - \frac{\sum_{i=1}^N (y_{pred,i} - y_{out,i})^2}{\sum_{i=1}^N \left(\sum_{i=1}^N |y_{out,i} - \overline{y_{pred}}| + \sum_{i=1}^N |y_{pred,i} - \overline{y_{pred}}| \right)^2}, (0 < IOA < 1) \quad (4.3)$$

4- Index de dispersion (IOS) :

$$IOS = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (y_{pred,i} - y_{out,i})^2}}{\overline{y_{pred}}}, (0 < IOS < \infty) \quad (4.4)$$

Commentaires : pour le processus du test de la farine, les données d'entrée/sortie typiques utilisées dans la modélisation sont présentées dans la figure 4.1. Les résultats pour la prédiction du gonflement de la pâte en fonction de la possibilité de gonflement de la pâte L02*, de la résistance de la pâte à la déformation P02*, et de la mesure de l'équilibre entre la résistance et l'extensibilité P/L02*. Les estimateurs correspondants sont donnés dans la figure 4.2 pour la technique RLS et la figure 4.3 pour REML et sa forme adaptative. Afin de valider l'approche de développement basée sur des données expérimentales, ces données ont été fournies à partir de la situation normale d'un service de contrôle de la qualité. En outre, une nouvelle série de données a été fournie dans les mêmes conditions d'expérimentation que l'étape de modélisation, pour laquelle les résultats de prédiction sont présentés dans la figure 4.4

Les erreurs globales de modélisation et de prédiction sont estimées par STD (voir tableau 4.3).

Pour le RLS, nous avons considéré un modèle de premier ordre avec des constantes C1 et C2. Le biais est estimé comme une valeur constante, où la convergence a été obtenue après une trentaine d'itérations.

La figure 4.3 se concentre sur les principaux résultats obtenus par l'algorithme REML et sa forme adaptative. Dans la forme adaptative, l'algorithme est exécuté en utilisant un mode de fenêtre glissante, ce qui donne des meilleurs résultats à la fois dans la modélisation et la prédiction. Ceci s'explique par le fait que la boucle adaptative suppose une grande adaptation des paramètres pour garantir une erreur minimale. Sur la base des résultats obtenus présentés dans les figures 4.2-4.4, et des incertitudes calculées indiquées dans le tableau 4.3, le modèle REML adaptatif donne la plus grande précision de prédiction par rapport au REML et RLS, de sorte qu'il peut être qualifié de meilleur outil ; cependant, la pire estimation est obtenue en appliquant le RLS. Connaissant la fonctionnalité de ces algorithmes, les résultats obtenus étaient prévisibles, et il faut souligner que la meilleure précision obtenue reste suffisante pour ce type d'application. De plus, le modèle de régression linéaire est moins précis que le modèle non linéaire basé sur les trois techniques ci-dessus, ce qui est dû à la simplification du modèle de processus. En réalité, il existe de nombreuses

interactions non linéaires qui n'apparaissent pas facilement et qui ne peuvent être détectées que par des calculs numériques [5].

4.3 Conclusion

Dans ce chapitre, de nouvelles méthodes basées sur des techniques avancées ont été explorées et mises en oeuvre. On présente les gonflements de la pâte et leurs estimations d'erreur correspondantes, ainsi que la prédiction du gonflement de la pâte, en fonction de la possibilité de gonflement de la pâte L02*, de la résistance de la pâte à la déformation P02* et de la mesure de l'équilibre entre la ténacité et l'extensibilité P/L02*. La procédure de modélisation et d'estimation a été réalisée par les modèles RLS, REML et REML adaptatif. Une étude comparative entre ces trois approches est donnée. Nos principales déductions sont les suivantes :

- Une procédure de calcul utilisant le RLS, le REML et le REML adaptatif est développée et appliquée avec succès pour des données de contrôle de qualité utilisant un modèle mixte.
- Des modèles de prédiction sont obtenus et testés.
- Basé sur le résidu généré par ces méthodes efficaces proposées, le contrôle de la variabilité utilisant la méthode REML adaptative est plus précis que les techniques RLS et REML.

Modèles multivariés et méthodes de mesure

5.1 Introduction

L'analyse chimique d'un matériau est une activité fondamentale et cruciale dans les processus de fabrication et de contrôle de la qualité. Le contrôle de la qualité et les essais nécessitent une évaluation continue pour la prise de décision, de la matière première au produit fini. De nombreuses caractéristiques doivent être reconnues dans le domaine des sciences et de l'ingénierie des matériaux afin d'obtenir un produit sans défauts ou avec des défauts acceptables. L'instrument et l'équipement appropriés peuvent être utilisés pour caractériser ces matériaux dans un laboratoire statique ou lors de mesures en ligne. Les performances de l'instrument et de l'équipement de mesure sont affectées non seulement par la capacité et la précision de l'étalonnage, mais aussi par de nombreux facteurs tels que les compétences humaines, les conditions environnementales et d'autres facteurs d'influence. En outre, l'incertitude des mesures dépend de nombreux facteurs aléatoires qui ne peuvent être quantifiés. L'impact global de ces variables pourrait être résumé dans la relation complexe entre ces facteurs et les résultats du modèle considéré. Les outils statistiques tels que l'écart type et d'autres paramètres de variabilité sont largement utilisés [24] et [73].

Il existe plusieurs approches existantes qui pourraient être appliquées pour évaluer la précision, on peut citer :

1. Les approches statistiques conventionnelles et les méthodes associées par exemple, la corrélation etc. [54], [55] et [78],
2. Contrôle de processus statistique multivarié, gestion et exploration des données [18], [20] et [62].

Dans ce chapitre, une nouvelle idée basée sur les techniques multivariées, par exemple Multi-Layer perceptron neural network (MLP), Partial least square (PLS) et Support vector regression (SVR) ; est introduite pour l'évaluation de la précision avec une application spécifique à la mesure de la teneur en acier à l'aide de la fluorescence X (XRF). Cette approche est constituée des étapes suivantes :

La première étape consiste à modéliser et à identifier la relation entre les valeurs de

référence et les valeurs de mesure à l'aide de différents modèles mathématiques dans l'analyse par fluorescence Rayon X. Il est important de rappeler que ces technologies de mesure sont basées sur une manière comparative de la réponse instrumentale. Les modèles directs et inverses sont largement développés.

La deuxième étape consiste à valider le modèle obtenu en termes d'erreur de modélisation à l'aide de l'écart-type d'incertitude (STD) et d'autres indicateurs de performance.

La motivation pour utiliser cette nouvelle approche est principalement justifiée par :

1. L'application considérée est un système multivariable, ce qui nécessite un formalisme des interactions.
2. Le principe de mesure présente des réactions d'état non linéaires et instables, la forme non linéaire est donc plus appropriée.
3. L'algorithme proposé est relativement simple à mettre en oeuvre.
4. Une incertitude de mesure précise et réelle est réalisée.

Afin d'obtenir des valeurs précises de la composition chimique de l'échantillon analysé, l'équation du modèle d'étalonnage est obtenue par une approche d'ajustement de courbe en utilisant différentes données de mesure par l'application de différents étalons appelés matériaux de référence certifiés (CRM).

Ce document est organisé comme suit. Dans la section 2, le modèle inférentiel proposé utilisant l'exploration de données pour l'analyse de la variabilité est développé. Dans cette partie, l'identification du modèle est réalisée et l'erreur de modélisation est calculée.

La section 3 comprend des applications typiques pour la précision de la mesure des éléments chimiques, où la structure du modèle est sélectionnée, la mise en oeuvre numérique est effectuée et les performances sont présentées et commentées. Nous terminons par une étude comparative entre la méthode conventionnelle sans interactions entre les éléments, où les effets mutuels entre les éléments sont négligés, et le cas le plus réel avec interactions où les intégrations ne peuvent être évitées.

5.2 Mesure et analyse des matériaux par spectrométrie

5.2.1 Principe de mesure

L'analyse de la fluorescence X de l'acier est un système complexe caractérisé par des interactions multivariées entre de nombreux facteurs influençant la mesure, tels que les conditions ambiantes, la géométrie de l'échantillon de métal et l'état de l'équipement. Pour obtenir des résultats fiables, ces facteurs doivent être traités. Un spectromètre XRF utilise généralement le rayonnement primaire d'un tube à rayons X afin d'exciter l'émission secondaire de rayons X d'un échantillon. Le rayonnement

qui émerge de l'échantillon contient les pics de rayons X caractéristiques des éléments majeurs et traces présents dans l'échantillon. La détection de ces éléments présents dans l'échantillon est rendue possible par la dispersion de ces rayons X secondaires dans un spectre, ce qui est généralement réalisé par diffraction des rayons X [1].

Pour la plupart des éléments, la hauteur de chaque pic de rayons X caractéristique est corrélée à la concentration de l'élément correspondant dans l'échantillon, permettant une analyse quantitative de l'échantillon. Les échantillons solides sont des pastilles de poudre pressées qui sont examinées.

Il existe deux manières essentielles d'effectuer une analyse spectrale :

1. Analyse des ondes dispersives par spectrométrie de fluorescence X WD-XRF.
2. Analyse dispersive en énergie par spectrométrie de fluorescence X ED-XRF.

L'échantillon émet des raies de fluorescence caractéristiques lorsqu'il est excité par un faisceau de rayons X de longueur d'onde suffisamment courte. La longueur d'onde de l'énergie émise par un élément excité est caractéristique de cet élément et l'intensité est proportionnelle au nombre d'atomes excités.

La fluorescence diffusée est collimatée et guidée sur la surface plane du cristal d'analyse par la fente d'entrée du goniomètre. Les radiations linéaires, réfléchies selon la condition de Bragg $m\lambda = 2d\sin(\theta)$ (où m est l'ordre de diffraction, λ est la longueur d'onde, d est la distance interréticulaire, et θ est l'angle de Bragg), passent à travers une fente de sortie vers le capteur, où l'énergie des quanta de rayons X est transformée en impulsions électriques ou en comptages. Le principe de mesure est donné par le schéma dans la figure 5.1 :

A noter que $w(t)$ est égal à la puissance d'entrée de l'excitation des rayons X.

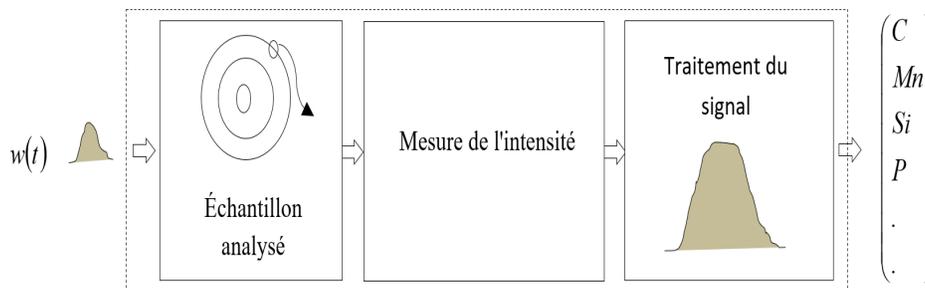


FIGURE 5.1 – Principe de l'analyse d'échantillons métalliques par fluorescence X

5.3 Application au contrôle de qualité des matériaux

Dans cette section, nous présentons dans figures 5.2 et 5.3 la structure de la fluorescence X. Le modèle des entrées et sorties expérimentales est présenté dans les

tableaux 5.1 et 5.2 et figures 5.2, 5.3 et figure 5.4. Dans le tableau 5.1 définit les paramètres d'entrées et de sortie. Tableau 5.2 présente les statistiques descriptives des échantillons collectés.

Modèle 1. L'entrée du modèle sans interaction est la concentration de carbone certifiée (voir figure 5.4, Fig.1a) et le modèle de sortie est la concentration de carbone mesurée (voir figure 5.4, Fig.1e); alors que les entrées du modèle avec interaction sont la concentration de carbone certifiée, la concentration de silicium certifiée, la concentration de manganèse certifiée et la concentration de soufre certifiée comme indiqué dans la figure 5.4 (Fig.1a-Fig.1d), et le modèle de sortie est la concentration de carbone mesurée (voir figure 5.4, Fig.1e).

Modèle 2. L'entrée du modèle sans interaction est la concentration de silicium certifiée (voir figure 5.4, Fig.1b) et le paramètre de sortie de processus est la concentration de silicium mesurée (voir figure 5.4, Fig.1f); alors que les entrées du modèle avec interaction sont les mêmes que dans le modèle 1, et le paramètre de sortie de processus est la concentration de silicium mesurée (voir figure 5.4, Fig.1f).

Modèle 3. L'entrée du modèle sans interaction est la concentration de manganèse certifiée (voir figure 5.4, Fig.1c) et le paramètre de sortie de processus est la concentration de manganèse mesurée (voir figure 5.4, Fig.1g), alors que les entrées du modèle avec interaction sont les mêmes que dans le modèle 1, et le modèle de sortie est la concentration de manganèse mesurée (voir figure 5.4, Fig.1g).

Modèle 4. L'entrée du modèle sans interaction est la concentration de soufre certifiée (voir figure 5.4, Fig.1d) et la sortie modèle est la concentration de soufre mesurée (voir figure 5.4, Fig.1h). alors que les entrées du modèle avec interaction sont les mêmes que dans le modèle 1, et le paramètre de sortie de processus est la concentration de soufre mesurée (voir figure 5.4, Fig.1h).

Principe des différents modèles

a-Cas 1 : entrée-sortie unique (E.S.U)

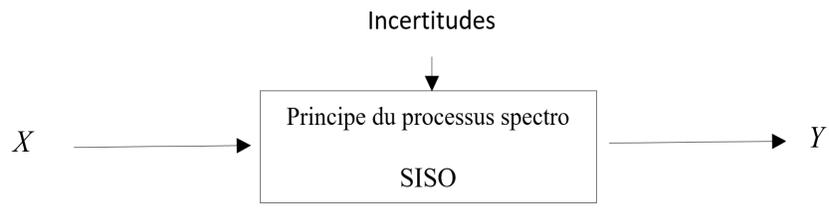


FIGURE 5.2 – E.S.U : structure du modèle sans interaction entre les concentrations

b-Cas 2 : entrées-sorties multiples(E.S.M)

Notre objectif est de trouver des modèles mathématiques d'entrées/sorties $Y_t = f_t(X_t) + \varepsilon$, où t est le temps et $\varepsilon \in N(0, \sigma)$ est une erreur résiduelle de la prédiction qui est normalement distribuée avec une valeur moyenne de zéro et un écart type σ .

5.3. APPLICATION AU CONTRÔLE DE QUALITÉ DES MATÉRIAUX

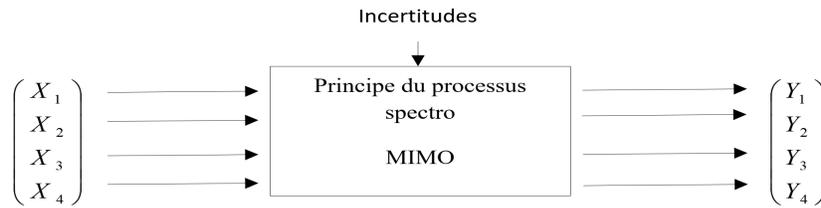


FIGURE 5.3 – E.S.M : structure du modèle avec interaction entre les concentrations

Variable		Désignation	Unité
Entrées	X_1	Concentration de carbone certifiée C	Wt%
	X_2	Concentration de silicium certifiée Si	Wt%
	X_3	Concentration de manganèse certifiée Mn	Wt%
	X_4	Concentration de soufre certifiée S	Wt%
Sorties	Y_1	Concentration de carbone mesurée C	Wt%
	Y_2	Concentration de silicium mesurée Si	Wt%
	Y_3	Concentration de manganèse mesurée Mn	Wt%
	Y_4	Concentration de soufre mesurée S	Wt%

TABLE 5.1 – Nomenclature des données d'entrée et de sortie.

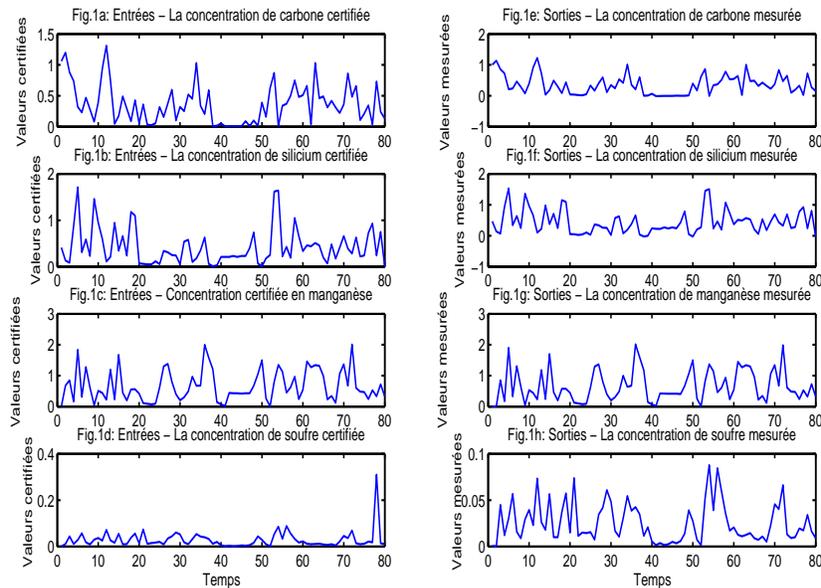


FIGURE 5.4 – Entrées et sorties expérimentales

	C_{Cert}	Si_{Cert}	Mn_{Cert}	S_{Cert}	C_M	Si_M	Mn_M	S_M
Moyenne	0.3667	0.4223	0.6685	0.0268	0.3560	0.4326	0.6614	0.0361
Médiane	0.3200	0.2650	0.5200	0.0235	0.3231	0.3028	0.5068	0.0203
Mode	0.1	0.21	0.43	0.009	-0.0117	-0.0433	0.0026	0.0391
Std	0.3184	0.3919	0.5063	0.0211	0.3159	0.3794	0.513	0.0611
Var	0.1014	0.1536	0.2563	0.000447	0.0998	0.1439	0.2631	0.0037
Rang	1.308	1.7096	2.005	0.0825	1.2367	1.5747	2.0129	0.4253
Min	0.003	0.0004	0.005	0.0005	-0.0117	-0.0433	0.0026	0.0003
Max	1.311	1.71	2.01	0.083	1.225	1.5315	2.0155	0.4256

TABLE 5.2 – Statistiques descriptives des échantillons collectés.

5.3.1 Résultats et discussion

a- Modélisation : un ensemble de données de mesure est utilisé pour identifier les paramètres du modèle qui seront utilisés pour la prédiction au cours du processus de modélisation. Les figures 5.5, 5.6, 5.7 et 5.8 illustrent les résultats de la modélisation pour les différentes méthodes et cas proposés.

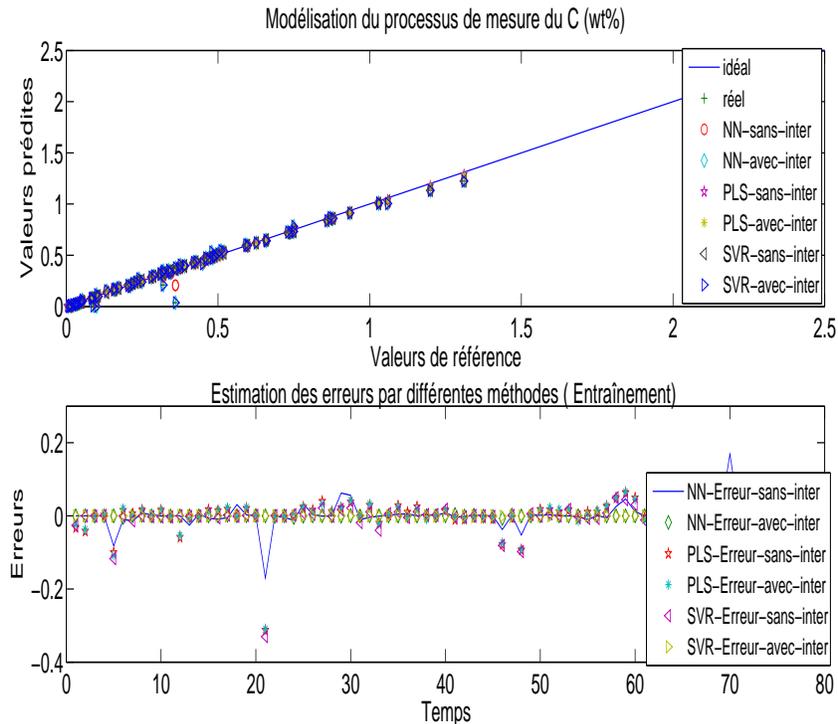


FIGURE 5.5 – Prédiction de la concentration de carbone et erreurs

5.3. APPLICATION AU CONTRÔLE DE QUALITÉ DES MATÉRIAUX

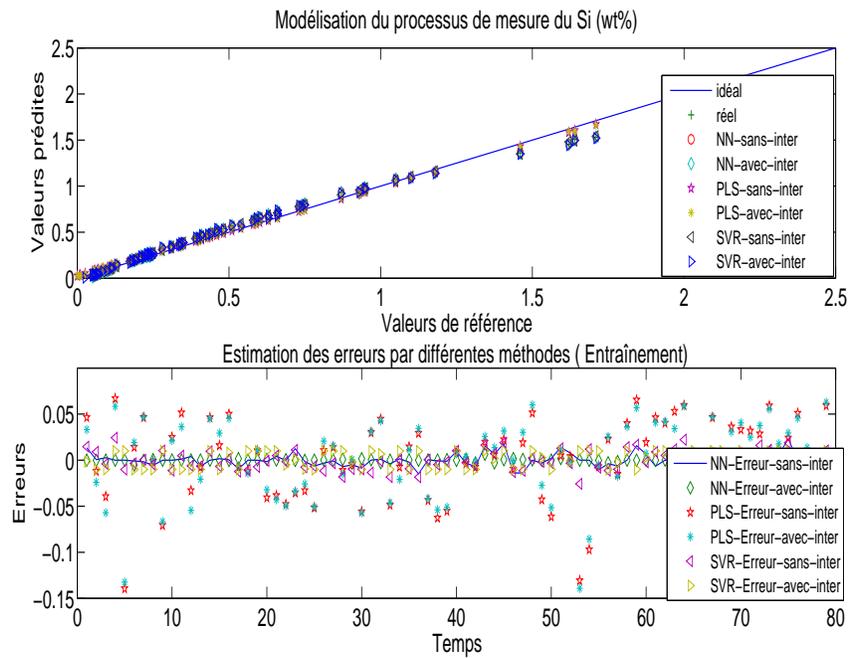


FIGURE 5.6 – Prédiction de la concentration en silicium et erreurs

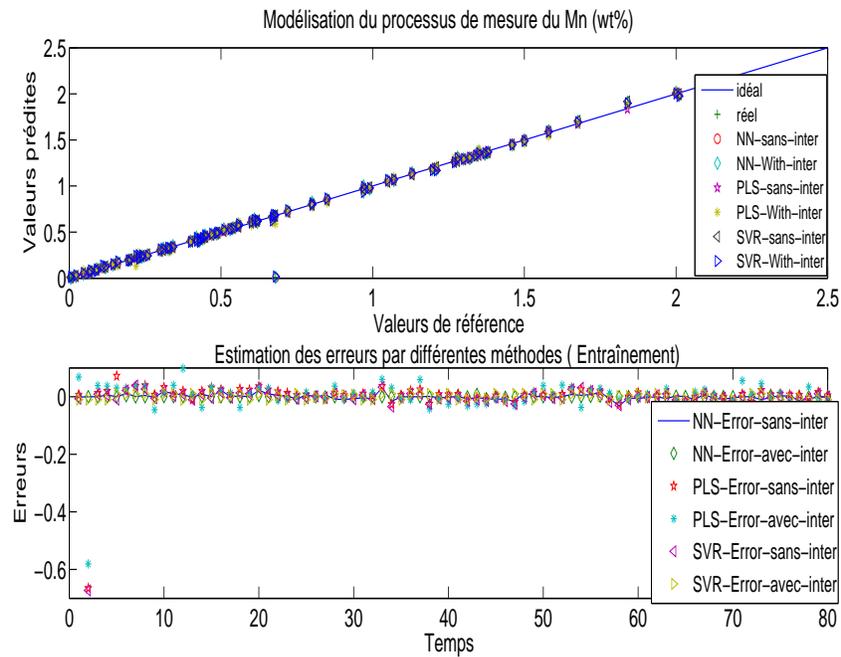


FIGURE 5.7 – Prédiction de la concentration en manganèse et erreurs

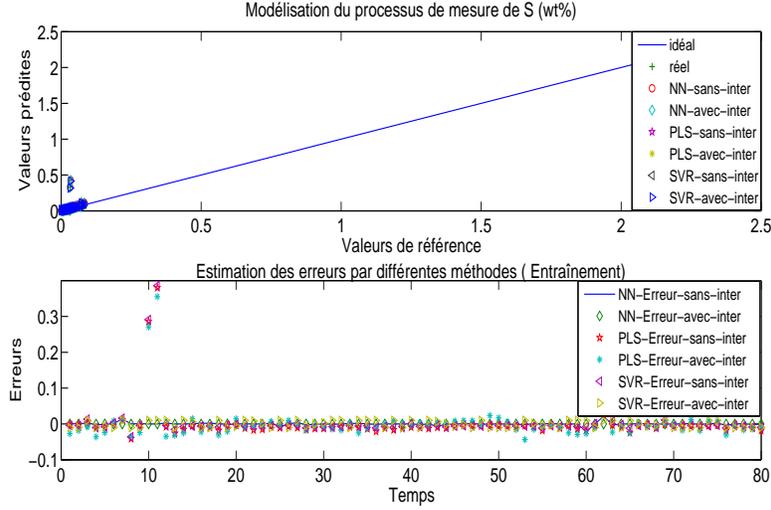


FIGURE 5.8 – Prédiction de la concentration de soufre et erreurs

En utilisant les algorithmes MLP, PLS, SVR et les données de mesure, les modèles suivants sont identifiés pour chaque méthode :

$$C_M(\%) = f_C(C_{Cert}(\%)) + \varepsilon_C \quad (5.1)$$

$$C_M(\%) = f_C(C_{Cert}(\%), Si_{Cert}(\%), Mn_{Cert}(\%), S_{Cert}(\%)) + \varepsilon_C \quad (5.2)$$

$$Si_M(\%) = f_{Si}(Si_{Cert}(\%)) + \varepsilon_{Si} \quad (5.3)$$

$$Si_M(\%) = f_{Si}(Si_{Cert}(\%), C_{Cert}(\%), Mn_{Cert}(\%), S_{Cert}(\%)) + \varepsilon_{Si} \quad (5.4)$$

$$Mn_M(\%) = f_{Mn}(Mn_{Cert}(\%)) + \varepsilon_{Mn} \quad (5.5)$$

$$Mn_M(\%) = f_{Mn}(Mn_{Cert}(\%), C_{Cert}(\%), Si_{Cert}(\%), S_{Cert}(\%)) + \varepsilon_{Mn} \quad (5.6)$$

$$S_M(\%) = f_S(S_{Cert}(\%)) + \varepsilon_S \quad (5.7)$$

$$S_M(\%) = f_S(S_{Cert}(\%), C_{Cert}(\%), Si_{Cert}(\%), Mn_{Cert}(\%)) + \varepsilon_S \quad (5.8)$$

Où nous désignons par M et $Cert$ les valeurs mesurées et certifiées respectivement et f_d est la fonction reliant les entrées et les sorties avec d désignant l'atome (composant), c'est-à-dire Carbone(C), Manganèse(Mn), Soufre(S), Silicium(Si).

5.3. APPLICATION AU CONTRÔLE DE QUALITÉ DES MATÉRIAUX

Éléments		STD (MLP)	STD (PLS)	STD (SVR)
Modèle sans interaction entre Concentration	Carbone	0.032411	0.044334	0.043535
	Silicium	0.007509	0.043721	0.011465
	Manganèse	0.009154	0.077163	0.076592
	Soufre	0.004149	0.054779	0.055033
Modèle avec interaction entre Concentration	Carbone	0.001710	0.044097	0.000205
	Silicium	0.001741	0.042997	0.009214
	Manganèse	0.001711	0.071806	0.009093
	Soufre	0.000122	0.053119	0.009086

TABLE 5.3 – Indicateurs statistiques de performance des différentes méthodes dans l'étape de modélisation.

b- Prédiction : nouvelles séries de données ont été utilisées pour la prédiction, en utilisant la structure du modèle et les paramètres qui ont été identifiés dans l'étape de modélisation pour les différents algorithmes PLS, SVR, et MLP. Une brève description de ces trois techniques est donnée dans les sections 2.1, 2.3 et 2.4. Les résultats obtenus pour la prédiction des concentrations de carbone, de silicium, de manganèse et de soufre et les erreurs sont résumés dans les figures 5.9, 5.10, 5.11, 5.12 respectivement et les performances obtenues qui sont caractérisées par STD dans la prédiction sont présentées dans le tableau 5.4.

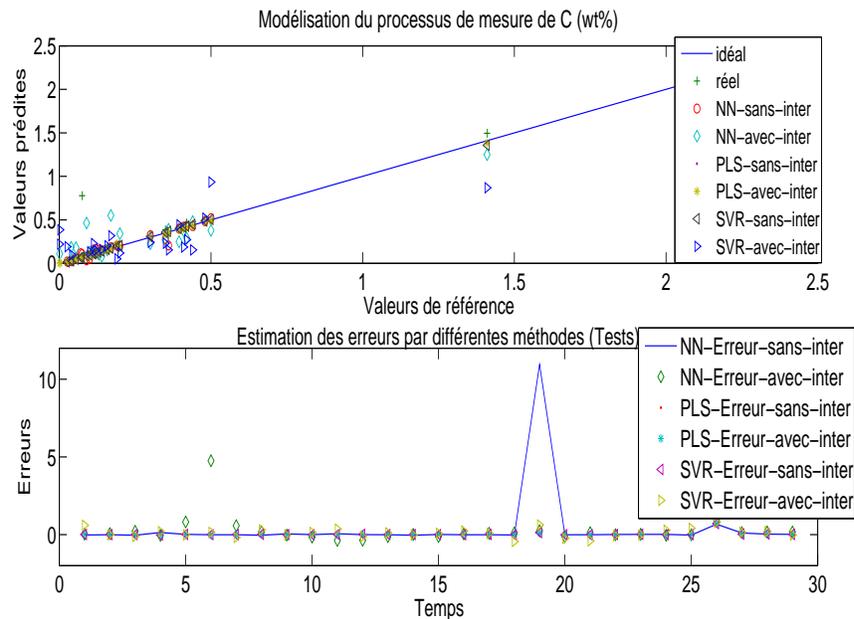


FIGURE 5.9 – Prédiction de la concentration de carbone et erreurs (test)

5.3. APPLICATION AU CONTRÔLE DE QUALITÉ DES MATÉRIAUX

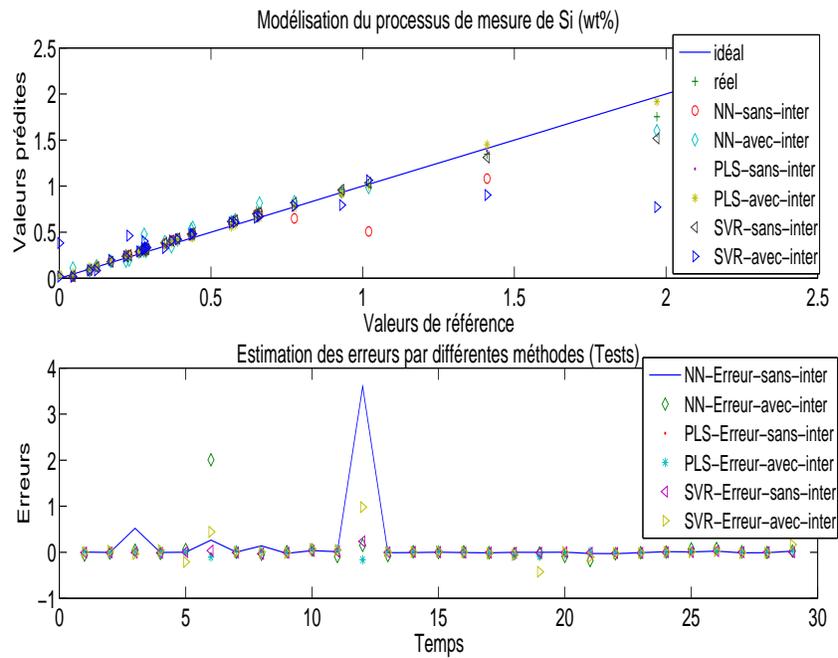


FIGURE 5.10 – Prédiction de la concentration de silicium et des erreurs (Test)

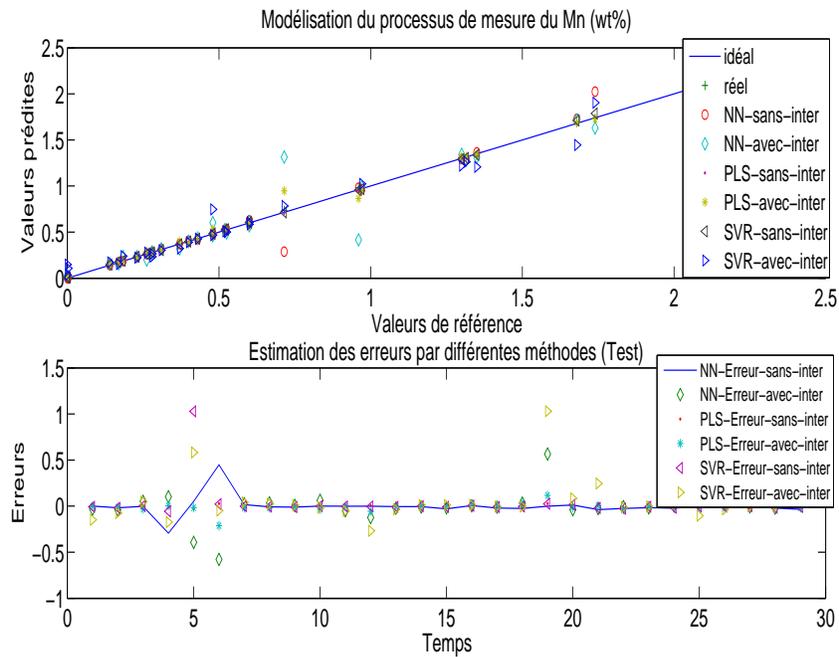


FIGURE 5.11 – Prédiction de la concentration de manganèse et des erreurs (Test)

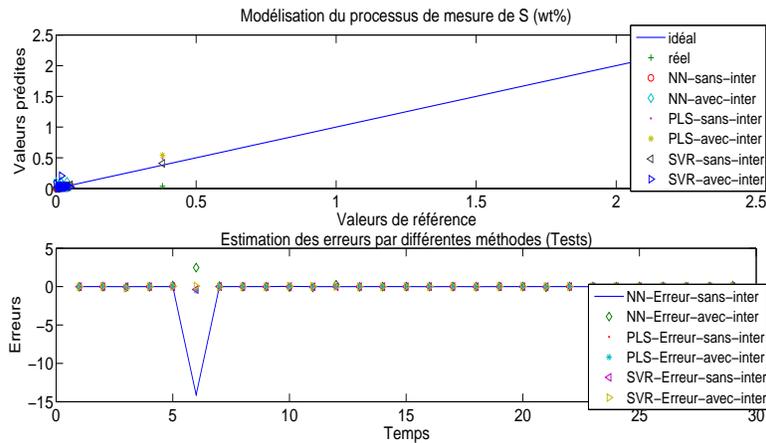


FIGURE 5.12 – Prédiction de la concentration en soufre et des erreurs (Test)

Éléments		STD (MLP)	STD (PLS)	STD (SVR)
Modèle sans interaction entre Concentration	Carbone	2.032933	0.130724	0.132237
	Silicium	0.671160	0.048132	0.045865
	Manganèse	0.103053	0.012893	0.192810
	Soufre	2.619091	0.083265	0.068677
Modèle avec interaction entre Concentration	Carbone	0.909436	0.131777	0.227164
	Silicium	0.378553	0.053930	0.224300
	Manganèse	0.174373	0.048758	0.237490
	Soufre	0.462510	0.094203	0.046209

TABLE 5.4 – Indicateurs statistiques de performance des différentes méthodes dans l'étape de prédiction.

Les tableaux 5.3 et 5.4 présentent les résultats des tests des différents modèles. La précision est évaluée à l'aide de l'écart type (STD). L'indicateur statistique de l'écart-type exprimé dans l'équation suivante est utilisé pour évaluer les différents modèles générés :

$$STD = \left(\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \right)^{\frac{1}{2}} \quad (5.9)$$

où n est le nombre d'échantillons, y_i est la valeur réelle et \hat{y}_i est la valeur prédite.

c- Discussion et commentaires

Pour le processus de fluorescence X, les données d'entrée/sortie typiques utilisées sont présentées dans la figure 5.4 et les statistiques descriptives des échantillons collectés sont présentées dans le tableau 5.2

Les tableaux 5.3 et 5.4 présentent les performances statistiques, où les indicateurs

ont été utilisés pour comparer les modèles proposés afin de sélectionner le meilleur d'entre eux. Il est à noter que STD dans l'équation (5.9) est utilisé pour un but comparatif, c.à.d les mêmes données sont utilisées pour construire les modèles.

L'étude comparative a été réalisée en utilisant ces de trois méthodes :

1. MLP : Réseaux de neurones perceptron multicouches,
2. PLS : Moindres carrés partiels,
3. SVR : Séparateur à vaste marge.

MLP, PLS et SVR sont de méthodes puissantes de calibration multivariée qui sont appliquées pour le traitement de données complexes. Pour la technique MLP, la fonction d'activation était TANSIG et la taille de la couche cachée utilisée dans le réseau de neurones artificiels est de 25, c'est-à-dire une seule couche cachée du réseau de neurones, et l'algorithme est celui de Levenberg Marquardt ; Par contre pour la technique PLS, le nombre de composantes est de 4. Enfin, pour l'approche SVR basée sur l'algorithme de Vapnik, le paramètre de régulation est une Série de C (Limite des multiplicateurs lagrangiens) et la fonction de noyau une gaussienne.

Pour chaque méthode, nous avons extrait 8 modèles, donc 24 modèles ont été implémentés. La prédiction des concentrations de carbone, de silicium, de manganèse et de soufre est effectuée par les trois différentes méthodes. L'approche MLP donne de meilleurs résultats en ce qui concerne l'erreur prédite, voir les figures 5.5, 5.6, 5.7, 5.8 et le tableau 5.3 dans l'étape de modélisation, les figures 5.9, 5.10, 5.11, 5.12 et tableau 5.4 dans l'étape de prédiction, ce qui est inférieur au PLS et au SVR, Les figures 5.5 - 5.12 présentent la relation entre les valeurs mesurées et les valeurs certifiées en utilisant les approches citées ci-dessus, ainsi que leurs erreurs associées. En plus, les résultats indiquent clairement que le modèle avec interaction entre les concentrations possède la meilleure précision dans les méthodes MLP, PLS et SVR, suivi par le modèle sans interaction entre les concentrations.

Selon les résultats obtenus présentés dans les figures 5.5-5.12 et les incertitudes calculées données dans les tableaux 5.3 et 5.4, la technique MLP présente la méthode la plus appropriée. En outre, le modèle avec interaction entre les concentrations donne la plus grande précision de prédiction par rapport à celui sans interaction, par exemple la concentration de silicium, de manganèse et de soufre a un effet sur la concentration de carbone.

5.4 Conclusion

Dans ce chapitre, différentes méthodes basées sur des techniques avancées ont été explorées et appliquées à l'évaluation de l'incertitude du processus de mesure de l'analyse métallique utilisant la fluorescence des rayons X. Les modèles de corrélation sont réalisés par régression multivariée. Ici, il est considéré comme un modèle à éléments multiples caractérisé par 8 équations (5.1 - 5.8) et figures 5.5 à 5.12.

La procédure de modélisation et d'estimation a été réalisée sur des données réelles provenant de l'usine Arcelor-Metal Factory-Annaba-Algérie, à l'aide de tests d'éta-lonnage et d'échantillons de mesure. Une étude comparative entre trois techniques d'intelligence artificielle a été réalisée et nos principales déductions sont les suivantes :

1. Les résultats obtenus indiquent que le MLP donne une incertitude plus précise que les autres méthodes,
2. Le modèle avec interaction entre les concentrations donne la meilleure précision, suivi par le modèle sans interaction.

Conclusion générale et Perspectives

Le travail de cette thèse a été principalement consacré à l'étude de trois problèmes industriels à l'aide de plusieurs méthodes d'apprentissage automatique telles que : les réseaux de neurones artificiels (NN), la régression par vecteurs de support (SVR), moindres carrés partiels (PLS), moindres carrés récursifs (RLS), la méthode de vraisemblance maximum (REML), et leurs versions adaptatives en utilisant les fenêtres glissantes. Chacune de ces méthodes a été mise en oeuvre à l'aide du logiciel MATLAB. Ces méthodes ont été appliquées à trois problèmes différents en utilisant des données réelles :

- Le premier provient de l'industrie sidérurgique, se rapportant aux Hauts Fourneaux.
- Le deuxième concerne un problème agro-industriel.
- Le troisième concerne l'application au contrôle qualité par machine d'analyse à rayons X.

Les résultats ont été illustrés à l'aide des différents indicateurs de performances tels que Standard deviation (STD), Index de l'accord (IOA), Index de dispersion (IOS), efficacité de Nash-Sutcliffe (NSE). Une étude comparative entre les méthodes, y compris leurs versions adaptatives, et d'autres techniques a été réalisée. Les résultats obtenus confirment plusieurs avantages de cette étude :

1. Un nouvel algorithme basé sur la méthode de vraisemblance maximum (REML adaptative) a été proposé.
2. Des modèles de prédiction ont été développés et testés.
3. Les approches adaptatives ont donné de meilleurs résultats par rapport aux approches conventionnelles dans les trois problèmes.
4. La méthode adaptative SVR s'est révélée plus précise que les autres approches dans le problème des Hauts Fourneaux.
5. Une estimation des risques multivariés, utilisant des méthodes hybrides basées sur la théorie des probabilités et l'apprentissage automatique, a été réalisée.
6. La méthode adaptative REML s'est révélée plus précise que les autres approches dans le problème agro-industriel.

-
7. Les méthodes sont efficaces dans les espaces de grande dimension.
 8. Les versions adaptatives sont implémentées de la même manière en utilisant des fenêtres glissantes comme le montre le diagramme 4. Il est à noter qu'en plus de la précision, les versions adaptatives sont plus compétitives et effectives.

Perspective

Les résultats obtenus par les différentes méthodes dans cette thèse ont conduit à des questions ouvertes, concernant l'analyse de sensibilité et les plans d'expériences, nous pouvons mentionner parmi ces questions par exemple : Quelles sont les paramètres d'entrées qui ont une grande influence sur la sortie y ?, ainsi est ce qu'il y a des interactions entre ces paramètres?

Bibliographie

- [1] Agarwal B. K. ; *X-ray spectroscopy : an introduction*, Springer 15, 2013.
- [2] Asradj Z. ; *Identification des systemes non linéaires par les réseaux de neurones*, Thèse de doctorat. Université Abderrahmane Mira-Bejaia, 2009.
- [3] Augusto V. ; *Data Mining*, Ecole nationale supérieure des Mines de Saint-Etienne 2012-2013.
- [4] Azzedine A., Nouri F.Z., and Bouhouche S. ; *Mathematical and numerical results for quality control of hot metal in blast furnace*, Journal of mathematical and computational science, vol. 11, pp 2914-2933, Avril 2021.
- [5] Azzedine A., Nouri F.Z., and Bouhouche S. ; *A Variance Analysis by an Adaptive Maximum Likelihood Approach*, Journal of applied probability and statistics, vol. 18, no. 2, pp 1-17, 2023.
- [6] Azzedine A., Nouri F. Z., and Bouhouche S. ; *A new hybrid approach based on probability distribution and an improved machine learning for multivariate risk assessment*. Diagnostyka, vol. 25, no. 1, 2024.
- [7] Bastien P., Vinzi V. E., and Tenenhaus M. ; *PLS generalised linear regression*, Computational Statistics and data analysis, vol. 48, no. 1, pp 17-46, 2005.
- [8] Ben Amina M. ; *Extraction Automatique des Connaissances à partir des Données : Etude de cas des organisées*, Université d'Oran1-Ahmed Ben Bella, 2007-2008.
- [9] Benbouras M. A. ; *L'utilisation des SIG sur les données d'infrastructure et de fondations : Application dans la zone d'Alger*. Thèse de doctorat. Ecole Nationale Polytechnique, 2018.
- [10] Bengt C. ; *Recursive identification*, Uppsala university, Department of information technology Sept 20, 2011.
- [11] Benchia H. ; *Modélisation d'un module photovoltaïque par les réseaux de neurones artificiels*, Mémoire de master. Université Larbi Ben M'hidi.06/2017.
- [12] Ben Messoud R. ; *Data mining*, Laboratoire ERIC., Institut universitaire de technologie lumière, 2007.

-
- [13] Berry M. J., and Linoff G. S.; *Data mining techniques : for marketing, sales, and customer relationship management*, John Wiley & Sons, Book Second Edition, 2004.
- [14] Bertrand L.; *Les essais inter-laboratoires en microbiologie des aliments Inter-laboratory studies in food microbiology*, Chemical Sciences. INAPG (AgroParisTech). English. ffNNT : 2004INAP0017ff. ffpastel-00001258, 2004.
- [15] Binard C.; *Introduction à la régression PLS*, université de Nice Sophia Antipolis, 16/11/2012.
- [16] Bipm, iec, ifcc, ilac, iso, iupac, iupap and oiml; *Joint Committee for Guides in Metrology - JCGM 106 :2012*, Evaluation of Measurement Data The Role of Measurement Uncertainty in Conformity Assessment, 2012.
- [17] Chen W., Wang B. X., and Han H. L.; *Prediction and control for silicon content in pig iron of blast furnace by integrating artificial neural network with genetic algorithm*, Ironmaking and Steelmaking, vol. 37, no. 6, pp 458-463, 2010.
- [18] Chunhui, Wang F., and Yingwei Z.; *Nonlinear process monitoring based on kernel dissimilarity analysis*, Control Engineering Practice, vol. 17, no. 1, pp 221-230, 2009.
- [19] David Garson G.; *Partial least squares regression and Structural Equations Models*, book, 2016.
- [20] De A. E., Melquiades F. L., Joao M. P., Margu E., and de Carvalho H. W.; *Determination of the polymeric thin film thickness by energy dispersive X-ray fluorescence and multivariate analysis*, Spectrochimica Acta Part B : Atomic Spectroscopy, vol. 167, pp 105818, 2020.
- [21] Faris H., Aljarah I., and Mirjalili; *Evolving radial basis function networks using moth-flame optimizer*, Handbook of neural computation, pp 537-550, 2017.
- [22] Fausett I.v.; *Fundamentals of neural networks : architectures, algorithms and applications* , 2006.
- [23] Francesco P.; *Online Support Vector Regression*, vol. 69, Italy 2007.
- [24] Jcgm, iso; *Evaluation of measurement data Guide to the expression of uncertainty in measurement, JCGM 100 :2008, GUM 1995 with minor corrections. First edition September, 2008.*
- [25] Gao C., Jian L., and Luo S.; *Modeling of the Thermal State Change of Blast Furnace Hearth With Support Vector Machines*, IEEE Trans. Ind. Electron, vol. 59, no 2, pp 1134-1145, 2012.
- [26] Gauchi J. P.; *Estimation de la répétabilité et de la reproductibilité des mesures expérimentales. Utilisation des procédures VARCOMP et NESTED du logiciel SAS/STAT (version 6.12). [Rapport Technique] RT 2000-4, auto-saisine, 44 p. fihal-01604565, 2000.*

-
- [27] Ge Z., Song Z., Ding S.X., and Huang B.; *Data Mining and Analytics in the Process Industry : The Role of Machine Learning*, IEEE Access, vol. 5, pp 20590-20616, 2017.
- [28] Gianluca B.; *Modélisation et simulation INFO-F 305*, Département d'informatique Boulevard de Triomphe, 2022.
- [29] Gilles G.; *Introduction au Data-Mining*, INSA Rouen -Département ASI Laboratoire LITIS 1, 13 septembre 2016.
- [30] Goodfellow I., Bengio Y, and Courville A.; *Deep learning*, goodfellow2016deep, MIT press, 2016.
- [31] Ghomari F.; *Science des matériaux de construction*, Département de génie civil, Université Aboubeker Belkaid.
- [32] Gurney K.; *An introduction to neural networks*, publisher : CRC press, 2018.
- [33] Hafezi Z., and Arefi M.; *Recursive generalized extended least squares and RML algorithms for identification of bilinear systems with ARMA noise*, ISA transactions, vol. 88, pp 50-61, 2019.
- [34] Hagan M. T., and Howard B.; *Demuth and Mark Beale, 2002 Neural Network Design*, publisher : Boston, Thomson Learning, 2002.
- [35] Hartigans J. H.; *Clustering algorithms*, publisher : John Wiley and Sons, Inc., 1975.
- [36] Haykin S.; *Neural Networks a comprehensive foundation*, second edition book, Canada, 1999.
- [37] Hengky L., and Richard N.; *Partial least squares path modeling-Basic concepts, methodological issues and application*, Springer, 2017.
- [38] Hopfield J.; *Neural networks and physical systems with emergent collective computational abilities*, Proceedings of the National Academy of Sciences of the United States of America, vol. 79, no. 8, pp 2554-2558, 1982.
- [39] Islam S. A. U., and Bernstein D. S.; *Recursive least squares for real-time implementation [lecture notes]*, IEEE Control Systems Magazine, vol. 39, no. 3, pp 82-85, 2019.
- [40] Iso.; *Guidance for the Use of Repeatability, Reproducibility and Trueness Estimates in Measurement Uncertainty Evaluation*, 2017.
- [41] Iso.; *Accuracy (trueness and precision) of measurement methods and results-Part 2 : Basic method for the determination of repeatability and reproducibility of a standard measurement method*, 2019.
- [42] Iuming Z.; *A Tutorial Restricted maximum likelihood estimation in linear regression and linear Mixed effect model*, 2015.
- [43] Jcgm J., Int. Organ. Stand; *Evaluation of measurement data-Guide to the expression of uncertainty in measurement*, Int. Organ. Stand. Geneva ISBN, vol. 50, pp 134, 2008.

-
- [44] Jian L., Gao C., and Xia Z. ; *A Sliding-window Smooth Support Vector Regression Model for Nonlinear Blast Furnace System*, Steel Res. Int., vol. 82, no. 3, pp 169-179, 2011.
- [45] Jianli X., Chao W., and Yuncai L. b. ; *Speed estimation of traffic flow using multiple kernel support vector regression*, Physica A : Statistical Mechanics and its Applications, vol. 509, pp 989-997, 2018.
- [46] Jiawei H., Michelin k., and Jian p. ; *Data mining : Concepts and techniques*, journal : Morgan Kaufmann, vol. 10, pp 559–569, 2006.
- [47] Kacker R. N. ; *Measurement uncertainty and its connection with true value in the GUM versus JCGM documents*, Measurement, vol. 127, pp 525-532, 2018.
- [48] Kadous D. ; *Utilisation des réseaux de neurones comme outil du data mining : Génération de modèle comportemental d'un processus physique à partir de données*, Laboratoire d'étude et Prédiction de Matériau Unité de recherches Matériaux et Energies Renouvelables, mémoire de master, 2012.
- [49] Kara Y., Canal MR S., and Boran FE. ; *Selecting and Analyzing Appropriate Probability Distributions for Reliability of Electrical Transmission Lines*, Gazi University Journal of Science Part C : Design and Technology, vol. 9, no. 1, pp 108-121, 2021.
- [50] Labiad A. ; *Sélection des mots clés basée sur la classification et l'extraction des règles d'association*, Université du Québec à Trois-Rivières, 2017.
- [51] Laouti N. ; *Diagnostic de défauts par les Machines à Vecteurs Supports : application à différents systèmes multivariés nonlinéaires*, thèse de doctorat Université Claude Bernard-Lyon I, 2012.
- [52] Larose D.T. ; *Discovering Knowledge In Data : An Introduction to Data Mining*, Central Connecticut State University, 2005.
- [53] Lavoie F.B., Langlet A., Muteki K., and Gosselin R. ; *Likelihood Maximization Inverse Regression : A novel non-linear multivariate model*, Chemometrics and Intelligent Laboratory Systems, vol. 194, pp 103844., 2019.
- [54] Li F., Zhixing G., Liangquan G., Sun D., Xutao D., Shun W., Bo H., and Jingru X. ; *Application of artificial neural networks to X-ray fluorescence spectrum analysis*, X-Ray Spectrometry, vol. 48, no. 2, pp 138-150, 2019.
- [55] Li W., Guang L., and Dongxiao Z. ; *An adaptive ANOVA-based PCKF for high-dimensional non linear inverse modeling*, Journal of Computational Physics, vol. 258, pp 752-772, 2014.
- [56] Ling J., and Chuanhou G. ; *Binary Coding SVMs for the Multiclass Problem of Blast Furnace System*, IEEE Transactions on Industrial Electronics, vol. 60, no. 9, pp 3846-3856, 2012.

-
- [57] Liu Q., Ding F., Wang Y., Wang C., and Hayat T.; *Auxiliary model based recursive generalized least squares identification algorithm for multivariate output error autoregressive systems using the decomposition technique*, Journal of the Franklin Institute, vol. 355, no. 15, pp 7643-7663, 2018.
- [58] Md Sazzad H., Zhi Chao O., Zubaidah I., Siamak N., and Shin Y. K.; *Artificial neural networks for vibration based inverse parametric identifications : A review*, Applied Soft Computing, vol. 52, pp 203-219, 2017.
- [59] Meyer K., and Kirkpatrick M.; *Restricted maximum likelihood estimation of genetic principal components and smoothed covariance matrices*, Genetics Selection Evolution, vol. 37, no. 1, pp 1-30, 2005.
- [60] Malengo A., and Walter B.; *Conformance probability in the assessment of Calibration and Measurement Capabilities*, Measurement, vol. 192, pp 110865, 2022.
- [61] Messis A., Adjebli A., R., Ayeche, Ghidouche A. and Ait-Ali D.; *Forecasting daily confirmed COVID-19 cases in Algeria using ARIMA models*, medRxiv, 2020.
- [62] Ma M. D., Ko J. W., Wang S. J., Wu M. F., Jang S. S., Shieh S. S., and Wong D. S. H.; *Development of adaptive soft sensor based on statistical identification of key variables*, Control Engineering Practice, vol. 17, no. 9, pp 1026-1034, 2009.
- [63] Modarres M., Kaminskiy M. P., and Krivtsov V.; *Reliability engineering and risk analysis : a practical guide*, CRC press, 2017.
- [64] Nemiche M.; *Data mining*, Master Masi faculté des sciences d'Aghadir, 2015.
- [65] Ostermark R., and Saxen H.; *VARMAX-modelling of blast furnace process variables*, Eur. J. Oper. Res., vol. 90, pp 85-101, 1996.
- [66] Phil K.; *MATLAB Deep Learning*, Journal : With Machine Learning, Neural Networks and Artificial Intelligence, publisher : springer , vol. 130, no. 21, 2017.
- [67] Ping W., and Chun-Jie Y.; *Identification and Control of Blast Furnace Gas Top Pressure Recovery Turbine Unit*, ISIJ International, vol. 52, no. 1, pp 96-100, 2012.
- [68] Prabhu S., and Venkatesan N.; *Data Mining and Warehousing, New Age International (P) Ltd.*, Publishers : New Delhi, 2007.
- [69] Saxen H., Gao C., and Gao Z.; *Data-Driven Time Discrete Models for Dynamic Prediction of the Hot Metal Silicon Content in the Blast Furnace-A Review*, IEEE Trans. Ind. Inf., vol. 9, no. 4, pp 2213-2225, 2012.
- [70] Shi L., Zhi-ling L., Tao Y., and jiang-peng L.; *Model of Hot Metal Silicon Content in Blast Furnace Based on Principal Component Analysis Application and Partial, Least Square*, Journal of iron and steel research, international, vol. 18, no. 10, pp 13-16, 2011.

-
- [71] Smith G. N. ; *Probability and statistics in civil engineering*, Collins professional and technical books, vol. 244, 1986.
- [72] Tang XL, Zhang L., and Jian C. J. ; *Prediction of silicon content in hot metal using support vector regression based on chaos particle swarm optimization*, Expert Syst Appl, vol. 36, no. 9, pp 11835-11857, 2009.
- [73] Velychko O., and Gordiyenko T. ; *The use of guide to the expression of uncertainty in measurement for uncertainty*, International Journal of Greenhouse Gas Control, vol. 3, no. 4, pp 514-517, 2009.
- [74] Wang Q., Chen F., Xu W., and Yang M. H. ; *Object tracking via partial least squares analysis*, IEEE Transactions on Image Processing, vol. 21, no. 10, pp 4454-4465, 2012.
- [75] Witkovsky V. ; *MATLAB algorithm mixed. m for solving Henderson's mixed model equations.*, Science Direct Working Paper, no. S1574-0358, pp 04, 2002.
- [76] Wentzell P. D., and Lohnes M. T. ; *Maximum likelihood principal component analysis with correlated measurement errors : theoretical and practical considerations*, Chemometrics and Intelligent Laboratory Systems, vol. 45, no. 1-2, pp 65-85, 1999.
- [77] Wold H. ; *Estimation of principal components and related models by iterative least squares*, in : P.R. Krishnaiah (Ed.), Multivariate Analysis, Academic Press, pp. 391-420, 1966.
- [78] Xiao J., Wei C., and Liu Y. ; *Speed estimation of traffic flow using multiple kernel support vector regression*, Physica A : Stat. Mech. Appl., vol. 509, pp 989-997, 2018.
- [79] Xie T., Hao Y., and Bogdan W. ; *Comparison between traditional neural networks and radial basis function networks*, IEEE international symposium on industrial electronics, pp 1194-1199, 2011.
- [80] Xu A., Wang Y., Zheng S., and Cai F. ; *Bias reduction in the two-stage method for degradation data analysis*, Applied Mathematical Modelling, vol. 77, pp 1413-1424, 2020.
- [81] Xu X. ; *Generalization of the Sherman-Morrison-Woodbury formula involving the Schur complement*, Applied Mathematics and computation, vol. 309, pp 183-191, 2017.
- [82] Yang H. , Jie L., Xiao Le. Ya., Wei Xing L., and Yu-Zhu Z. ; *Dynamic Prediction Research of Silicon Content in Hot Metal Driven by Big Data in Blast Furnace Smelting Process under Hadoop Cloud Platform*, Hindawi, 2018.
- [83] Zaiane O. ; *Principles of Knowledge Discovery in databases, chapter 8 : Data Clustering, lecturing slides for CmPUT 690*, University of Alberta, 1999.
- [84] Zhang G., Eddy Patuwo B., and Hu M. Y. ; *Forecasting with artificial neural networks*, International journal of forecasting, vol. 14, no. 1, pp 35-62, 1998.

- [85] Zhang Q.X., Tao Y., and Cui Z.B.; *The Temperature Prediction in Blast Furnace Base on Fuzzy Least Squares Support Vector Machine*, Applied Mechanics and Materials, vol. 336, pp 566-569, 2013.
- [86] Zhang X., Kano M., and Li Y.; *Locally weighted kernel partial least squares regression based on sparse nonlinear features for virtual sensing of nonlinear time-varying processes*, Computers and Chemical Engineering, vol. 104, pp 164-171, 2017.
- [87] Zhang X., HU L., and WANG Z.; *Multiple Kernel Support Vector Regression for Economic Forecasting*, Computers and Chemical Engineering, October 2018.
- [88] Zhang X., Kano M., and Matsuzaki S.; *Ensemble pattern trees for predicting hot metal temperature in blast furnace*, Computers and Chemical Engineering, Elsevier, vol. 121, pp 442-449, 2019.
- [89] Zhou B., Ye H., Zhang H., and Li M.; *Process monitoring of iron-making process in a blast furnace with PCA based methods*, Control engineering practice, vol. 47, pp 1-14, 2016.
- [90] Zhou P., Song H., Wang H., and Chai T.; *Data-Driven Nonlinear Subspace Modeling for Prediction and Control of Molten Iron Quality Indices in Blast Furnace Ironmaking*. IEEE Transactions on Control Systems Technology, vol. 25 , pp 1761-1774, 2017.