

وزارة التعليم العالي والبحث العلمي

BADJI MOKHTAR UNIVERSITY -ANNABA-
UNIVERSITE BADJI MOKHTAR -ANNABA-



جامعة باجي مختار
- عنابة -

Faculté : Sciences de l'Ingénieur -Année 2009-

Département : Informatique

MEMOIRE

Présentation en vue de l'obtention du diplôme de magister

Prédiction des Paramètres de Pollution de l'air:
Application à la Région d'Annaba.

Option

Texte Parole et Image

Par

Sabri GHAZI

DIRECTEUR DE MEMOIRE: Khadir Med Tarek Maitre de conférences Univ. Annaba

DEVANT LES JURY

PRESIDENT: Dr. Seridi Hassina Maitre de conférences Univ. Annaba

EXAMINATEURS:

Dr. Farah Nadir Maitre de conférences Univ. Annaba

Dr. Bahi Halima Maitre de conférences Univ. Annaba

Remerciements

Je remercie tout d'abord *Allah le tous puissant*,

Je tiens à remercier tous mes enseignants qui ont participé à ma formation, spécialement *Mr Khadir* qui a accepté de m'encadrer dans ce travail, merci d'être si compétant et toujours compréhensible.

Je remercie aussi tous mes amis spécialement *Mohamed, Tarek, Sofiane* qui m'ont encouragé, et guidé à compléter ce travail.

Je remercie aussi mes parents ainsi que mes frères et ma sœur.

Ps : à part la première phrase, l'ordre n'a pas d'importance.

Sabri,

ملخص

يعتبر تلوث الهواء من بين الإشكاليات التي تواجه كثير من مدن العالم، و إنه لمن الأهمية بمكان استعمال أنظمة تنبؤ بتركيز الملوثات في الهواء. تأثر كثير من المتغيرات في هذه الظاهرة، أهمها : العوامل المناخية، طبوغرافية المكان، النشاط الصناعي و السكاني. هذا ما يجعل تلوث الهواء صعب المحاكاة، و تصميم أنموذجات تحاكيه صعب للغاية، هناك الكثير من النماذج التي اقترحت أنموذج الرياضي للانبعاث، الأنموذج الخطي و الشبكات العصبية الاصطناعية. هذا ما يسمح بإعطاء تنبؤات عن تركيز الملوثات في الهواء و يسمح بالتعامل مع حالات التلوث القصوى. في هذا العمل نقترح نظام قادر على إعطاء تنبؤات دقيقة لحالة التلوث و تركيز الملوثات في الهواء بناحية عنابة شمال شرق الجزائر. نماذج متنوعة جربت : خطية ARX ، شبكات إلمان، RBF، PMC، MLP تراجعية (NARX). من اجل إعطاء تنبؤات متوسطة المدى قبل 12 ساعة. و باختيار النموذج الأكثر فعالية قمنا بتطوير نموذج عام للتنبؤ بدرجة نوعية الهواء. النماذج RBF و إلمان كانت الأكثر فعالية لكن تطويرا هو الأكثر تعقيدا، نماذج MLP كانت الأسرع من حيث التطوير و تقدم تنبؤات فعالة، أما NARX لها أهمية لأنها تقدم تنبؤات متعددة ما يسمح بالتعرف على تطور تركيز الملوثات، نماذج ARX هي الأسرع من حيث التطور لكنها الأقل فعالية.

كلمات مفتاحية :

تنبؤ بنوعية الهواء، تنبؤ بتركيز الملوثات في الهواء، شبكات العصبية الاصطناعية، RBF، إلمان، MLP، ARX، NARX.

Résumé

La pollution de l'air est un problème qui prend de plus en plus d'importance dans beaucoup de villes dans le monde. Plusieurs facteurs influencent ce phénomène, dont les plus importants sont : les conditions climatiques, la topologie, et l'activité urbaine. Ceci rend la pollution de l'air très difficile à modéliser. Beaucoup de modèles de prédiction de pollution de l'air ont été proposés (modèle mathématique d'émission, modèles linéaires, modèles basés réseaux de neurones artificiels, modèles hybrides logique flou et RNA) dans le but d'avoir un système efficace de prédiction et de contrôle de qualité de l'air, qui aidera à limiter l'influence des périodes de pics en informant la communauté, et en prenant les précautions nécessaires.

Dans ce travail on présente une étude de plusieurs approches donnant un système qui peut donner des prédictions efficaces concernant la concentration des polluants dans l'air. Ceci dans le but de concevoir un système de prédiction de l'indice de qualité de l'air dans la région d'Annaba dans le nord-est de l'Algérie.

Différents modèles ont été testés : linéaire ARX, Modèles basés RNA (Elman, RBF, PMC, PMC récurrent appelé aussi NARX) pour donner des prédictions sur la concentration de : SO₂, NO₂, PM10, O₃ et CO₂ à moyen terme c'est-à-dire 12 heures en avance. Nous avons comparé les performances que donne chaque modèle, et nous avons choisi les plus performants pour concevoir le modèle globale de prédiction de l'indice de qualité de l'air.

Les modèles basés RBF et Elman ont donné les meilleures performances. Ceux basés PMC donnent aussi des prédictions efficaces. Les modèles basés NARX ont l'avantage de donner des prédictions multi-pas ce qui permet de voir le développement future de la pollution, mais leurs performances se détériorent à chaque fois que le nombre de pas de prédiction dépasse 6 heures. Cependant leur utilisation reste intéressante pour prédire les pics et pour des prédictions à moyen et court terme. Les modèles basés ARX sont très utilisés dans ce domaine, ceux sont les moins consommant en terme de temps de conception mais leurs performances par rapport aux autres modèles est les moins efficaces.

Mots clés :

Prédiction de la pollution de l'air, prédiction de la qualité de l'air, réseaux de neurones artificiels, réseaux d'Elman, RBF, réseaux de neurones récurrents, ARX, PMC.

Abstract

Air pollution is a major concern in many cities in the world, many factors influences in air pollutants concentrations the most important are: climatical conditions, topology and population density. This makes air pollution difficult to model. Many air pollution prediction models have been studied (mathematical emission models, linear models, Artificial Neural Networks based models) in order to design an air quality prediction system to control the air pollution, and limit the influence of peaks periods by informing community, and taking the necessary precaution.

In this work we present a system that can make efficient prediction about air pollutant concentration, in order to predict the air quality indices for the region of Annaba north-east of Algeria. Different models are presented, linear, nonlinear and ANN based models. All these models are created in order to predict the SO₂, NO₂, PM10, O₃ and CO₂ concentration in 12 hours ahead. We have compared the performances of these models and by choosing the most efficient we have build another model that can use there in order to make prediction about air quality indices.

The RBF and Elman based neural networks models gives the best performance. MLP also gives efficient prediction. Those based on recurrent MLP (NAX) models which give a multi-steps ahead prediction have a good accuracy in the short terms prediction, but when the number of steps become greater than 6 steps (hours) their performance starts falling down.

The models based on recurrent MLP ARX are the least consuming in training and creation time, but gives the worst performances compared with the others models.

Keys words:

Air pollution prediction, Air quality prediction, artificial neural networks, Elman networks, RBF, recurrent neural networks, MLP, ARX, forecasting.

Sommaire

Remerciements	2
ملخص	3
Résumé	4
Abstract	5
Sommaire	6
Index des figures	10
Index des tableaux	12
Introduction générale	15
Prédiction des paramètres de pollution atmosphérique	18
Introduction	19
1.1 L'atmosphère	20
1.1.2 Composition chimique de l'atmosphère	20
1.1.3 La structure verticale de l'atmosphère	21
1.1.4 Pollution atmosphérique	22
1.1.5 Différents types de polluant	22
1.1.5.1 Polluants primaires	23
1.1.5.2 Polluants secondaires	24
1.2 Système de prédiction et contrôle de la pollution de l'air	24
1.2.1 Module d'un SPCPA	26
1.2.2 Modèles de prédiction de pollution de l'air (état de l'art)	26
Données utilisées et région ciblée	43
2.1 Région ciblée	44

2.2 Données utilisées dans l'étude	46
2.2.1 Les polluants mesurés par Sama Safia	47
2.2.2 Paramètres climatiques	47
2.2.3 Propriétés statistiques des données	48
PM10	48
Ozone (O ₃)	49
NO _x	50
NO	51
CO ₂	52
NO ₂	53
Humidité	53
Température	54
Vitesse du vent	55
Les Réseaux de neurones artificiels	57
3.1 Réseau de neurones artificiels	58
3.1.1 Le neurone formel	58
3.1.2 Connexion entre unité	59
3.1.3 Type de neurones simples	59
3.1.3.1 Neurone de Mc Culloch-Pitt	59
3.1.3.2 Perceptron	60
3.2 La règle d'apprentissage de Hebb	61
3.3 La règle d'apprentissage du Perceptron	61
3.2.1 Convergence de la règle de perceptron	62

3.3 Différents type de réseaux de neurones artificiels	63
3.3.1 Perceptron multicouches (PMC)	63
3.4.2 PMC pour l'approximation des fonctions	65
3.4.3 Apprentissage par retro-propagation	66
3.3.5 Réseau à fonction de base radiale	70
3.4.6 Réseau de neurone récurrent	71
3.4.6.1 Réseau de neurones faiblement récurrent	72
3.4.6.1 Réseau de neurones fortement récurrent	72
Développement du modèle	74
4.1 Prétraitement des données	76
4.1.2 Nettoyage des données	77
4.1.3 Normalisation des données	77
4.1.4 Résultat de l'étape de prétraitement	79
4.1 Modèles linéaires	84
4.2 Modèles basés réseaux de neurones artificiels	87
4.2.1 Modèles basé PMC	87
4.2.2 Modèle basé PMC récurrent	97
4.2.3 Modèle basé réseaux d'Elman	105
4.2.4 Modèle basé réseau RBF	115
4.3 Comparaison des résultats obtenus	116
Modèle global de prédiction de la qualité de l'air	121
5.1 Modèle global de prédiction de la qualité de l'air	122
Conclusion générale	126

Références	127
Annexe A : Calcul de performance	132
Annexe A : Calcul de performance	133

Index des figures

Figure 1.1. Profile vertical de température dans l'atmosphère.	21
Figure 1.2 Echelle spatio-temporelle de chaque polluant, source (Boubel et al).....	23
Figure-1.3- : Architecture générale d'un système de prédiction et de contrôle de pollution de l'air.	25
Figure -1.4- : Schéma de modèle hybride de prédiction de qualité de l'air.	40
Figure -1.5- : Modèle en box de pollution de l'air.	41
Figure 2.1 : Carte géographique de Annaba (Google ©).	44
Figure 2.2 : Topographie de la région d'Annaba (Google Earth ©)	45
Figure 2.3 : Concentration PM10 en 2003.	49
Figure 2.4 : Concentration P10 en 2004.....	49
Figure 2.5 : Concentration O ₃ en 2004.....	50
Figure 2.6 : Concentration NO _x en 2003.	50
Figure 2.7 : Concentration NO _x en 2004.	51
Figure 2.8 : Concentration du NO en 2003.	51
Figure 2.9 : Concentration du NO en 2004.	52
Figure 2.10 : Concentration du CO ₂ en 2003.	52
Figure 2.11 : Concentration du NO ₂ en 2003.....	53
Figure 2.12 : Humidité en 2003.	54
Figure 2.13 : Humidité en 2004.	54
Figure 2.14 : Température en 2003.	55
Figure 2.15 : Température en 2004.	55
Figure 2.16 : Vitesse du vent en 2003.....	56

Figure 2.17 : Vitesse du vent en 2004.....	56
Figure 3.1 : Schéma d'un neurone formel.....	59
Figure 3.2 : Neurone de Mc Culloch-Pitt.....	60
Figure 3.3 : Le perceptron.....	60
Figure 3.4 : Règle d'apprentissage du perceptron.....	62
Figure 3.5: Perceptron Multicouches	63
Figure 3.6 : PMC justifiant le théorème de Cybenko.....	65
Figure 3.7 : Deux neurones de différentes couches dans un réseau de type PMC.....	67
Figure 3.8 : Trois neurones de différentes couches dans un PMC à deux couches cachées....	68
Figure 3.9 : Schéma général d'un réseau à fonction de base radiale.....	70
Figure 3.10 : Un réseau de neurones récurrent.....	71
Figure 3.11: Schéma d'un réseau d'Elman typique.	72
Figure 4.1 : Construction des données d'apprentissage et de test.....	81
Figure 4.2 : Comparaison de résultat des modèles pour la prédiction du PM10.....	117
Figure 4.3 : Comparaison du résultat des modèles pour la prédiction de l'O ₃	117
Figure 4.4 : Comparaison du résultat des modèles pour la prédiction de l'NO ₂	118
Figure 4.5 : Comparaison du résultat des modèles pour la prédiction de l'CO ₂	118
Figure 4.6 : Comparaison du résultat des modèles pour la prédiction de l'SO ₂	119
Figure 4.7 : Comparaison du résultat des modèles pour la prédiction de PM10.....	119
Figure 5.1 : Architecture d'un modèle global de prédiction de la qualité de l'air.	123
Figure 5.2 : Résultat du modèle global de prédiction de l'indice de qualité de l'air	125

Index des tableaux

Tableau 1.1 Composition chimique de l'air	21
Tableau -1.3- : Un extrait des règles construites.....	39
Tableau 2.1 : Les polluants mesurés par Sama Safia.....	47
Tableau 2.2 : Les paramètres climatiques mesurés par Sama Safia.....	47
Tableau 2.3 Paramètres météorologiques moyens dans la région d'Annaba.....	48
Tableau 2.3 : Propriétés statistiques de mesure de concentration de PM10.....	49
Tableau 4 : Propriétés statistiques de mesure de concentration du l'Ozone.....	50
Tableau 5 : Propriétés statistiques de mesure de concentration du NO _x	51
Tableau 6 Propriétés statistiques de mesure de concentration du NO.....	52
Tableau 7 : Propriétés statistiques de mesure de concentration du CO ₂	53
Tableau 8 : Propriétés statistiques de mesure de concentration du NO ₂	53
Tableau 9 : Propriétés statistiques de mesure de l'humidité.....	54
Tableau 10 : Propriétés statistiques de mesure de température.....	55
Tableau 11 : Propriétés statistiques de mesure de la vitesse du vent.....	56
Tableau 3.1 : Les étapes de l'algorithme de rétro-propagation de l'erreur.....	69
Tableau 4.1 : Les données disponibles.....	77
Tableau 4.2 : Normalisation des données.....	78
Tableau 4.3 : Résultat de l'étape de prétraitement.....	80
Tableau 4.4 : Les paramètres du modèle basé ARX.....	84
Tableau 4.5 : Les modèles ARX testés	87
Tableau 4.6 : Les paramètres du modèle basé PMC	87

Tableau 4.7 : Les performances du modèle basé PMC pour le SO ₂	88
Tableau 4.8 : Les courbes d'erreur de modèle PMC pour le SO ₂	89
Tableau 4.9 : Les performances du modèle basé PMC.....	89
Tableau 4.10 : Les courbes d'erreur des modèles construits.....	90
Tableau 4.11 : Performances des modèles basés PMC pour le CO ₂	91
Tableau 4.12 : Les courbes d'erreur de modèle.	92
Tableau 4.13 : Performance du modèle pour le NO ₂	93
Tableau 4.14 : Les courbes d'erreur de modèle.	94
Tableau 4.15 : Performance du modèle pour l'O ₃	95
Tableau 4.16 : Les courbes d'erreur de modèle PMC.....	97
Tableau 4.17 : Paramètres du modèle PMC récurrent.....	97
Tableau 4.18 : Les résultats du modèle PMC récurrent pour le polluant SO ₂	98
Tableau 4.19 : Les courbes d'erreur des modèles basés PMC récurrent.....	100
Tableau 4.20 : Performance du modèle PMC récurrent pour le NO ₂	101
Tableau 4.21 : Les courbes d'erreur des modèles basés PMC récurrent.....	102
Tableau 4.22 : Performance du modèle PMC récurrent pour le PM ₁₀	103
Tableau 4.23 : Performance du modèle PMC récurrent pour le O ₃	104
Tableau 4.24 : Performance du modèle PMC récurrent pour le CO ₂	105
Tableau 4.25 : Paramètres du modèle Elman.....	105
Tableau 4.26 : Performance du modèle Elman pour le PM ₁₀	106
Tableau 4.27 : Courbes d'erreurs des modèles basés Elman pour le PM ₁₀	107
Tableau 4.28 : Performance du modèle basé Elman pour le NO ₂	108
Tableau 4.29 : Courbes de l'erreur du modèle basé Elman pour le NO ₂	109

Tableau 4.30 : Performances du modèle basé Elman pour le CO ₂	110
Tableau 4.31 : Courbes de l'erreur du modèle basé Elman pour le CO ₂	111
Tableau 4.32 : Performance du modèle basé Elman pour le SO ₂	112
Tableau 4.33 : Courbes de l'erreur du modèle basé Elman pour le SO ₂	113
Tableau 4.34 : Performances du modèle basé Elman pour le O ₃	113
Tableau 4.35 : Courbes de l'erreur du modèle basé Elman pour le O ₃	115
Tableau 4.36 : Paramètres du modèle basé RBF.....	115
Tableau 4.37 : Performance du modèle basé RBF	116
Tableau 5.1 : L'indice de qualité de l'air utilisée par SamaSafia.....	124

Introduction générale

Dans ce travail nous proposons des solutions et des modèles qui permettent de mieux prédire la concentration des polluants dans l'air, afin de pouvoir donner des prédictions efficaces sur la qualité de l'air dans la région de Annaba nord-est de l'Algérie. Ce qui permettra de mieux gérer les périodes des pics en informant la communauté et les personnes souffrants de maladie chronique tel que l'asthme.

La pollution atmosphérique est un véritable problème mondial de santé publique, spécialement dans les zones urbaines à grande concentration tel que les grandes villes algériennes (Alger, Constantine, Oran, Annaba) dans lesquels l'activité industriel est très proche des zones urbaines.

Le problème reste posé dans les zones d'industrie pétrochimique dans le sud du pays, en fait malgré que les raffineries pétrolières ne sont pas proche des zones urbaines, quelques villes notamment Ouargla dans laquelle l'expansion urbaine devient plus en plus proche des zones industrielles, ce qui fait augmenter les risques de pollution de tout type (pollution de sol, pollution de l'air, pollution d'eau).

La création d'entreprises et l'élargissement du marché économique étaient le souci majeur des décideurs de la période postindépendance. Ces investissements, malheureusement, n'ont pas été accompagnés d'un dispositif de lutte contre la pollution, aucun décret spécifique de protection contre la pollution atmosphérique n'ayant été élaboré. C'est en 2004 seulement, après que la qualité de l'air s'est détériorée dans la plupart des grandes villes, qu'il y a prise de conscience générale.

La pollution est définie comme étant un phénomène ou un élément perturbateur d'un équilibre établie et plus particulièrement si cet élément est nuisible à la vie et la santé. La pollution peut être anthropique (c'est-à-dire causé par l'homme et ses activités), ou d'origine non humaine (tel que l'activité volcanique de la terre).

La pollution est l'un des problèmes le plus critique qu'affronte l'humanité, en fait comment assurer les besoins de plus de six milliards tout en gardant une certaine harmonie avec la nature et l'environnement dans lequel nous vivons. Le problème de pollution se pose partout, surtout dans les pays en développement tel que l'Algérie.

La prédiction et le contrôle de la qualité de l'air reste la seule arme pour lutter contre les pics de pollution qui cause des effets nocifs sur la santé et l'écosystème. La reproduction des

comportements de phénomène de pollution de l'air permet de mieux étudier, comprendre, identifier les paramètres les plus influant, afin d'avoir les éléments nécessaires pour prendre des décisions allant jusqu'à l'évacuation des villes et la suspension de certaines activités industrielles. Un exemple de telle situation est les jeux olympique de pékin en 2008 qui selon les organisateurs ont été influencé par la pollution ce qui à conduit les autorités à suspendre l'activité industrielles dans les zones de band lieux de pékin.

Le mémoire est organisé comme suit :

La première partie du premier chapitre présente les systèmes de prédiction et de contrôle de la pollution de l'air, et aussi les différents modules qui composent ce genre de système, dans la deuxième partie un état de l'art des modèles utilisées pour la prédiction de pollution atmosphérique est présenté.

La description géographique et topologique de la région concernée par cette étude et les données utilisées est faite dans le deuxième chapitre.

Les différents types de réseaux de neurones qui ont été utilisés dans ce travail sont présentés dans le troisième chapitre.

La première partie du chapitre quatre décrit les étapes de prétraitement qui ont été appliquées aux données. Dans la deuxième partie les architectures et les performances en termes de résultat sont présentés.

Le cinquième chapitre présente un modèle de prédiction de l'indice de la qualité de l'air, ce dernier a été développé en se basant sur les résultats obtenus dans les étapes précédentes cité dans le quatrième chapitre.

Chapitre 1

Prédiction des paramètres de pollution atmosphérique

Introduction

La prédiction de la pollution atmosphérique a fait l'objet de plusieurs études, plusieurs modèles et approches ont été proposées pour prédire la concentration des différents polluants, ainsi que pour reproduire les comportements de ce phénomène physique-chimique qui influence d'une façon directe sur nos vies et notre environnement.

L'évolution de l'informatique a permis l'utilisation des approches qui demandent une puissance de calcul importante, et aussi à la conception des systèmes de simulation et de prédiction de la pollution atmosphérique à grande échelle (régional, national ou même mondial (Cihan *et al* 2006), (Gitte *et al* 2005).

L'intérêt de pouvoir fournir des informations sur la pollution de l'air et due au fait que ces informations possèdent une importance pour plusieurs acteurs : autorité publique, utilisateur simple, expert du domaine qui cherche à trouver des outils qui lui permettent l'étude de la pollution de l'air. La génération de ces informations se fait en se basant sur des ressources matériels et logiciels qui forment un système de prédiction et de contrôle de pollution de l'air. Ce type de système informatique tient leurs importances du fait qu'ils facilitent l'étude et la production des informations concernant la pollution de l'air. En fait c'est grâce à ce genre de système que le monde a pu prédire les effets de l'échauffement climatique.

Donc le but des systèmes de prédiction et de contrôle de la pollution atmosphérique peut être résumé en : Prédire la pollution, simuler et étudier le phénomène de pollution.

1.1 L'atmosphère

Définie comme étant l'enveloppe gazeuse de la terre, l'atmosphère est l'environnement dans lequel la vie subsiste, beaucoup de transformation chimique, en particulier d'origine photochimique surviennent à ce niveau.

L'air qui est le fluide gazeux qui constitue l'atmosphère, est indispensable à la vie car il participe au processus de la respiration et à la photosynthèse des végétaux, cet environnement très sensible et qui subit le plus d'influence de la part de l'activité urbaine.

1.1.2 Composition chimique de l'atmosphère

L'azote moléculaire (N₂) est le composant principal de l'air, suivie de l'oxygène moléculaire (O₂), ainsi que quelques gazes moins abondants comme Ne, He, Kr, et Xe, qui apparaissent aussi en quantité bien déterminée.

Chaque espèce¹ possède un temps de résidence allant de quelques minutes jusqu'à plusieurs années, ce qui prouve le dynamisme de la structure chimique de l'air.

Le tableau ci-dessous illustre les différentes concentrations des constitués de l'atmosphère, ainsi que le temps de résidence de chaque espèce :

<i>Gaz</i>	<i>Concentration</i>	<i>Temps de résidence</i>
Azote (N ₂)	78.084 %	-
Oxygène (O ₂)	20.946 %	-
Argon (Ar)	0.934 %	-
Eau (H ₂ O)	[0.4..400] x 10 ² ppm	10 jours
Dioxyde Carbone CO ₂	370 ppm	4 ans
Néon (Ne)	18.18 ppm	-
Hélium (He)	5.12 ppm	2 * 10 ⁶ ans

¹ Dans le domaine d'étude de la pollution et la géologie, une espèce représente un élément chimique ou solide.

Méthane (CH ₄)	1.75 ppm	10 ans
Krypton (Kr)	1.14 ppm	-
Hydrogène (H ₂)	0.4 ppm	-
Xénon (Xe)	0.87 ppm	-

Tableau 1.1 Composition chimique de l'air

1.1.3 La structure verticale de l'atmosphère

Schématiquement l'atmosphère peut être découpée en quatre zones selon l'altitude :

- Troposphère : de 0 jusqu'à 12 kilomètres.
- Stratosphère : jusqu'à 50 kilomètres.
- Mésosphère : jusqu'à 80 kilomètres.
- Enfin la thermosphère : jusqu'à 150 kilomètres.

Chaque couche se distingue par la fonction température/altitude, et c'est en se basant sur la température que le découpage a été défini, la figure suivante illustre la courbe de la température selon l'altitude :

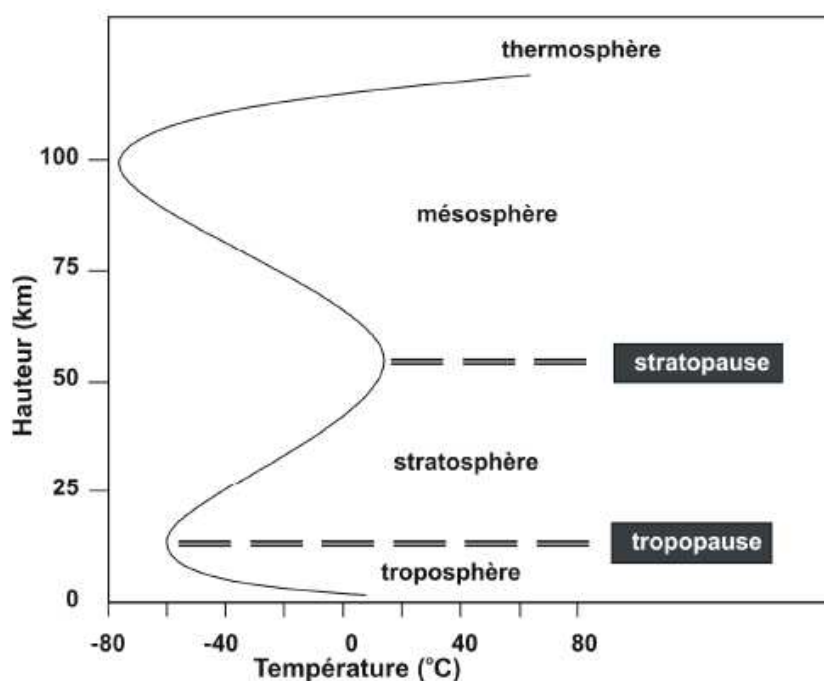


Figure 1.1. Profil vertical de température dans l'atmosphère.

1.1.4 Pollution atmosphérique

La pollution atmosphérique (ou de l'air) est l'altération de la composition chimique de l'air, par l'augmentation de la concentration de certaines espèces (chimique ou solide), cet altération durera une certaine période de temps pour créer des effets toxiques, et qui aura aussi des effets nocif sur l'écosystème et la santé (Barker et Al 1963).

La pollution est dite locale (à l'échelle régional tel que : les pics d'ozone, le smog), globale à l'échelle planétaire (exemple l'effet de serre, destruction de la couche d'ozone).

La pollution qui touche la couche troposphérique constitue le type le plus important, du fait que ce dans cette couche que l'humain ainsi que beaucoup espèce vivent, les autres types de pollution sont moins importants, sauf ceux qui influencent sur le climat, tel que l'effet de serre qui à lieu dans la couche thermosphère.

1.1.5 Différents types de polluant

La pollution atmosphérique se distingue selon les espèces considérées, la liste suivante contient les principales espèces dites 'polluant' :

- SO_2 : dioxyde de soufre, lié à la combustion du charbon.
- O_3 : l'ozone et les oxydes d'azote (NO , NO_2),
- COV : composés organiques volantes.
- Métaux lourds, tel que : le plomb, le zinc, est qui sont lies à certaines type de processus industriels, sous forme particulaire.
- Mercure, émis sous forme gazeuse
- Aérosols, sous forme de particule et qui englobe le sulfate SO_2 , NH_4 , sable, sels de mer etc.
- Polluants Organiques Persistants (POPs), chaînes organiques complexes avec de grands temps de vie (pesticides, dioxine).

Les particules (particulate matter) :

Les particules regroupent également une variété importante de substances. Le terme de particules désigne en général la fraction des composants (liquides ou solides) en suspension dans le milieu gazeux. Souvent, les particules sont classées en fonction de leur granulométrie,

par exemple les particules dont le diamètre aérodynamique moyen est inférieur à 10 μm est noté PM10, de même pour les PM2.5.

Chaque polluant considéré va donc se caractériser par une classe d'espèces particulières et des propriétés spécifiques : sources d'émission, mécanismes physico-chimiques, temps caractéristiques, la figure suivante illustre le temps de résidence dans l'atmosphère de chaque polluant :

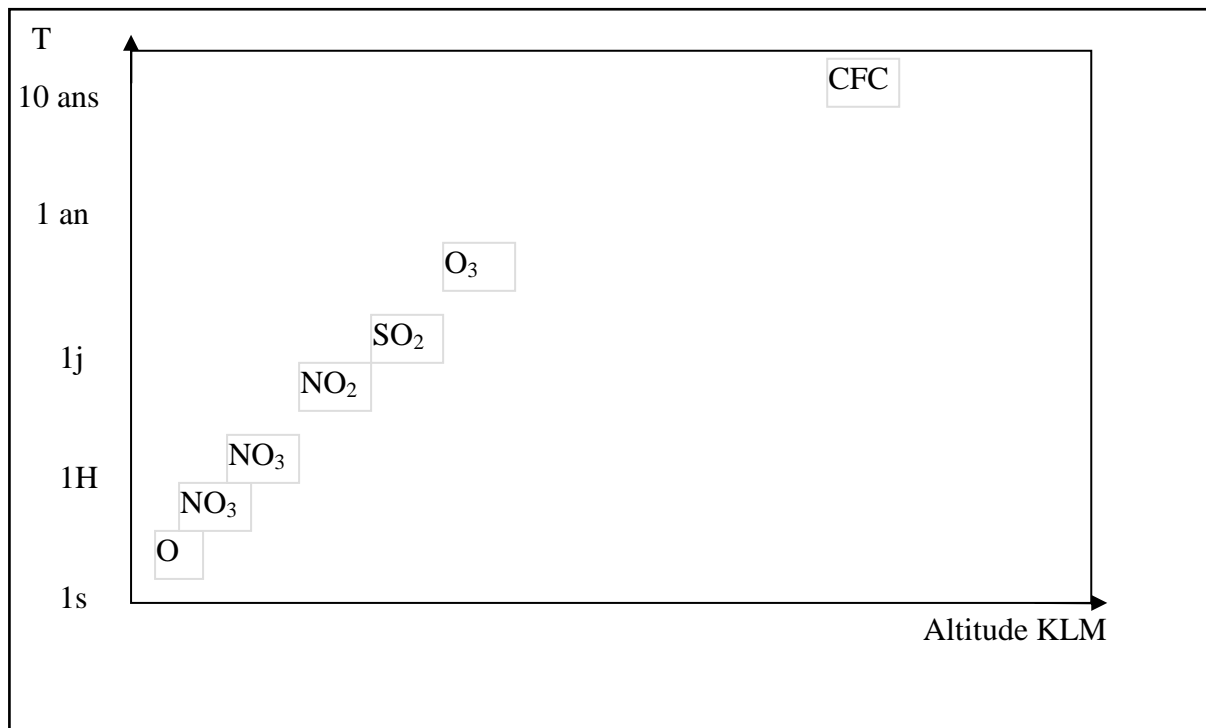


Figure 1.2 Echelle spatio-temporelle de chaque polluant, source (Boubel et al).

1.1.5.1 Polluants primaires

Les polluants primaires, sont ceux qui sont émis directement dans l'air, ces émissions sont catégorisées en deux types :

- Biogénique : liées à l'activité naturelle tel que l'activité photosynthétique, sels de mer, volcaniques etc.
- Anthropique : liées à l'activité humaine tel que les processus industriels, transport, production de l'énergie etc.

On peut citer comme exemple de polluant primaire l'oxyde de soufre qui est émis par l'opération de combustion de charbon.

1.1.5.2 Polluants secondaires

Les polluants secondaires résultent des réactions photochimiques, dans l'air, dans lesquelles les polluants primaires jouent un rôle très important, un exemple de ce genre de polluant est l'ozone, ce dernier résultent de la photochimie des oxydes de l'azote(NOx).

1.2 Système de prédiction et contrôle de la pollution de l'air

Un Système de Prédiction et de Contrôle de Pollution de l'Air (SPCPA), comme schématisé dans la figure 1.3 est un ensemble de ressources matériels et logiciels mise a la disposition des utilisateurs afin de :

- Donner des informations concernant la concentration des différents polluants, ce qui permettra le lancement des alertes, afin de permettre aux autorités à prendre des mesures de précaution pour contrôler la situation allant jusqu'à la suspension des activités polluantes.
- Permettre l'étude des relations qui existent entre la concentration des polluants et les paramètres climatiques. Cela se fait par la création des cas non réels et la simulation, Les experts peuvent à l'aide d'un SPCPA voir les paramètres climatiques qui influencent plus sur la concentration des polluants.
- Faire des notifications périodiques concernant la qualité de l'air, les rapports concernant le climat se font d'une façon journalière, cela aussi doit se faire pour les paramètres de qualité de l'air.
- Déclencher des alertes en cas des pics de pollution, d'une façon automatique le système doit offrir un module de gestion des alertes. En cas où la concentration d'un polluant dépasse un certain seuil une alerte doit être déclenchée.
- Vu que le domaine de pollution de l'air traite de grosse quantité de données, un système de prédiction et de contrôle de pollution de l'air doit posséder une interface homme-machine riche qui permet de présenter les informations d'une façon utile.
- Visualisation 3 D des scènes, facilitant la présentation des informations concernant la dispersion de polluant dans l'air, ce qui permettra de bien localiser les zones les plus polluées, et voir le rapport entre la topographie de celles-ci avec la concentration des

polluants, et aussi voir les effets des activités urbaines exercées dans ces zones sur la concentration des polluants dans l'air.

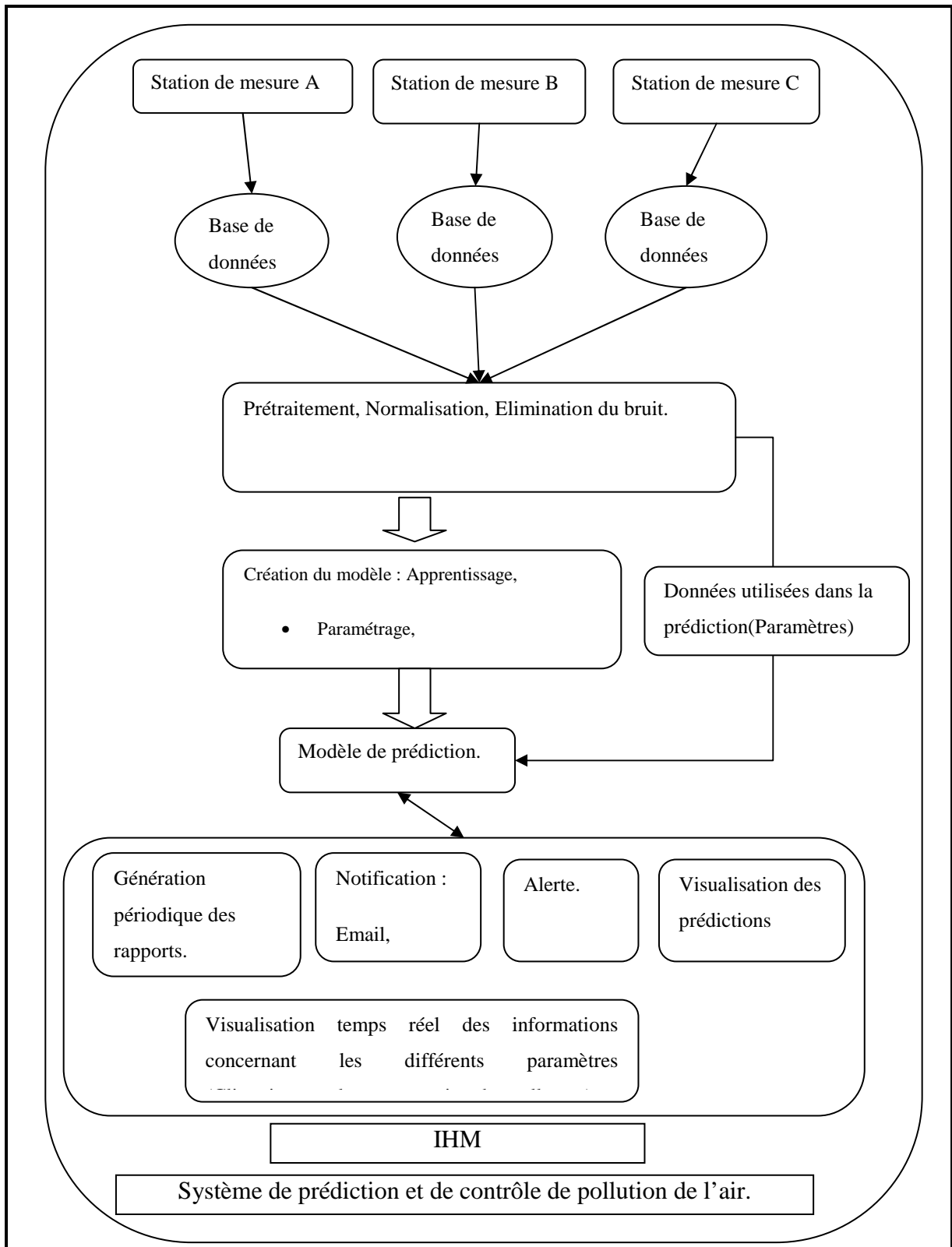


Figure-1.3- : Architecture générale d'un système de prédiction et de contrôle de pollution de l'air.

1.2.1 Module d'un SPCPA

Comme illustré dans le schéma, un SPCPA se constitue de plusieurs modules, chaque module assure une fonctionnalité précise, on peut citer :

1. Module de mesure : ce module se compose de station de surveillance fixe et mobile, placé d'une manière à avoir des données représentatives de toutes les zones qui constitue la région étudiée,
2. Module de Stockage des mesures : toutes les mesures doivent être sauvegardés et traités, car la quantité des données disponibles influence énormément sur l'efficacité des prédictions,
3. Module de prétraitement des données : les mesures sont issue de différentes machines et capteurs, et doivent être traitées et normalisées afin qu'elles soient utilisables facilement,
4. Module de création et mise à jour de modèle : C'est un composant qui permet aux experts de créer et de faire l'apprentissage de modèle, afin qu'il soit utilisable par le système,
5. Modèle de prédiction : c'est le composant le plus important du système. A partir des données disponibles (paramètres) le modèle doit avec un certain taux de succès, donner des prédictions,
6. Module d'interface homme-machine : l'interface avec laquelle le système génère les sorties et reçoit les données (paramétrage etc.). L'IHM donne aux utilisateurs un accès aux différentes fonctionnalités du système. L'IHM doit aussi contenir des fonctionnalités dédiées aux experts, pour permettre la création des modèles, la mise-à-jour des modèles, et aussi celles des simulations des cas non réels.

1.2.2 Modèles de prédiction de pollution de l'air (état de l'art)

L'un des composants les plus importants des SPCPA est le modèle de prédiction, qui détermine l'efficacité et l'utilité de tout le système, pour trouver le modèle le plus efficace plusieurs études ont été effectuées, dans la partie qui suit on cite quelque unes.

Les travaux dans le domaine de prédiction de la qualité de l'air peuvent être catégorisés selon plusieurs critères :

- Le polluant étudié (O₃, PM10, PM2.5, CO, CO₂, etc.).

- Le modèle utilisé : linéaire (ARIMA, AR, ARX, etc.), non linéaire, RNA ou d'autres modèles issues de l'IA tel que la logique floue, Algorithmes Génétiques et des modèles hybrides incluent plusieurs approches (ex : ANN et logique floue).
- Aspect temporel de la prédiction : des modèles génèrent des prédictions à court terme et autres génèrent des prédictions à long terme ou même à moyen terme.
- Les paramètres utilisés : qui consiste à inclure ou non les paramètres climatiques, ou d'autres données relatives à l'émission des polluants (tel que l'activité routière et industriel).
- L'implication ou non du processus de construction/ destruction du polluant, en fait quelque modèle prend en considération les interactions chimiques qui ont lieu dans l'air. Ces interactions entre polluants sont souvent modélisées par des formules mathématiques.
- Type et taille de la zone étudiée.

Les réseaux de neurones artificiels sont des alternatives aux méthodes statistiques classiques utilisées dans les stations de surveillance de la qualité de l'air. Ces méthodes qui se basent sur l'analyse des séries temporelles des différents polluants, pour détecter les tendances de chaque série, et donner des prédictions. Il serait très efficace d'utiliser les grosses quantités d'informations disponibles à partir des stations de surveillance pour la conception des modèles de prédictions efficaces dans le but d'avoir des systèmes de contrôle et de surveillance de pollution atmosphérique plus efficaces et utiles. Et qui permettent d'étudier le phénomène de pollution de l'air. Par exemple : savoir quels sont les paramètres qui influencent le plus sur la concentration des polluants dans l'air.

La relation entre les paramètres météorologiques et la pollution est très étroite, et complexe à modéliser. Dans leurs études (Malkar et Boznar 1994), présentent un modèle basé PMC (Perceptron Multi Couches), pour la prédiction de la concentration du dioxyde de soufre SO₂. Le modèle reçoit en entrée la concentration de SO₂ et les autres paramètres météorologiques, ainsi que les paramètres relatifs à l'émission des stations de production de l'énergie.

Le modèle a montré une bonne qualité de prédiction, non seulement pour les cas alarmants, mais aussi à avoir une baisse du nombre de fausse alarme, et dépasse le modèle ARIMA conçue avec les mêmes paramètres, car la tendance d'un modèle à faire des fausses alarmes (faire des prédictions surestimées) est un critère à prendre en considération pendant le développement d'un modèle de prédiction, notamment ceux de prédiction de la pollution

atmosphérique, pour que les utilisateurs ont confiance au résultat que donne un système de prédiction.

La sélection des paramètres influence beaucoup sur l'efficacité du modèle, (Mlakar et Josef 1997) présentent une approche qui s'inspire de la reconnaissance des formes, qui se base sur la formulation (1.1) pour modéliser la prédiction et les paramètres :

$$B = \{V_i, O_i\}..i = 1,2,\dots, m \quad (1.1)$$

Où

B : est un ensemble universel, la prédiction et les paramètres.

V_i : est le pattern numéro i , la prédiction.

O_i : est la description du pattern i (le vecteur des paramètres).

Le vecteur des paramètres V est défini comme suit : $V_i = [x_1, x_2, \dots, x_r]^T$. Avec le vecteur de sortie $V_0 = [Y_1(t, \tau)] = SO_2(t + \tau)$ où $\tau = 1/2h$ (une demi heure pour la prédiction à court terme). La réduction des paramètres consiste à réduire les paternes de r dimensions vers n dimensions où $r \gg n$, en utilisant un réseau de neurone de type *Kohanen* pour extraire les paramètres pertinents. Cela a permis d'optimiser le temps de l'apprentissage et aussi à augmenter l'efficacité du modèle.

Pour trouver le modèle le plus adapté au problème de la prédiction de la qualité de l'air. (Jorquera et al 1998) présentent une comparaison entre trois modèles, réseaux de neurones artificiels, modèle linéaire, et modèle flous. La comparaison est faite en se basant sur les critères suivants :

- La capacité du modèle à prédire un épisode alarmant de la concentration de l'ozone.
- La tendance du modèle à déclarer des épisodes qui n'auraient pas lieu, (faire une prédiction alarmante qui ne se produise pas) ce dernier critère doit être le plus faible possible.

Un modèle mathématique a été utilisé pour modéliser l'émission de l'ozone, ce modèle se base sur l'équation de diffusion atmosphérique suivante :

$$\frac{\partial C}{\partial t} = -V.\Delta C + \Delta.(K.\Delta C) + Q(x, y, z, t).R\{C_k\}, T, t) + L(x, y, z, t) \quad (1.2)$$

Où

$C(x, y, z, t)$: La concentration des l'espèce (polluants).

$V = (u, v, w)$: Le vecteur du paramètre vent (direction et force).

$K = \text{diag}(Kx, Ky, Kz)$: Diffusion de l'espèce (polluant).

Q : Le tau d'émission.

R : La balance entre la production et la destruction de l'espèce.

Dans le cas de l'ozone le paramètre émission peut être éliminé, car l'ozone n'est pas un polluant primaire. Il résulte des réactions chimiques des autres polluants spécialement le *NO*.

Le modèle définie par (1.2) peut être simplifié en :

$$\frac{\partial C}{\partial t} = R(\{C_k\}, T, t) \quad (1.3)$$

L'intégration de l'équation (1.3), entre le début de la journée (t_0) et à t_{0+3h} donne :

$$O_3^t \max - O_3^t \min = \int_{t_0}^{t_0+3h} R(\{C_k\}, T, t) dt \quad (1.4)$$

Donc le modèle résultant est :

$$O_{3,t+1} = \alpha O_{3,t} + \beta T_{t+1} + \gamma T_t + \delta \quad (1.5)$$

Où : α est positive, β , γ sont des paramètres de différent signe, δ a une petite valeur et T c'est la température. Cette étude a permet de conclure que le modèle flou est le plus adéquat pour la prédiction de l'ozone, car il a montré des performances dépassant les autres modèles, notamment dans le nombre de fausse alarme (le plus faible par rapport aux autres). Cela ne veut plus dire qu'il le soit pour autre région, car chaque région possède ses propres paramètres qui affecteront les performances du modèle.

(Gardner et Drooling 1999a) présentent un modèle pour la prédiction de la concentration de NO_2 , avec une comparaison entre un régresseur linéaire et un modèle neuronal de type PMC,

le modèle neuronal a dépassé significativement le modèle linéaire, utilisant les mêmes paramètres. Dans (Gardner et Droling 1999b), les tendances de la concentration de l'ozone est modélisés à l'aide d'un réseau de neurone de type PMC, la série temporelle de la concentration de l'ozone est modélisé comme suit:

$$O_3(t) = E(t) + S(t) + W(t) \quad (1.6)$$

Où

$E(t)$: le composant long terme.

$S(t)$: le composant saisonnier.

$W(t)$: le composant court terme.

Le composant long terme représente la variation de la concentration de certains gaz qui participent dans le cycle de production / destruction de l'ozone, et le changement climatique.

Le composant saisonnier représente le changement annuel de la radiation solaire et de direction de vent.

Le composant court terme représente les changements météorologiques d'un jour à l'autre.

Cette étude a permis de concevoir un modèle qui permet de détecter les tendances à long terme de la concentration de l'ozone pour la région de Londres.

La prédiction à court terme de la qualité de l'air est de l'ordre de quelques heures, cela est dû au temps de résidence des polluants dans l'air, un exemple de prédiction à court terme est présenté dans (Perez et al.2002) avec une comparaison entre trois modèles : réseau de neurones, régression linéaire et le modèle persistant. Tous ces modèles ont été utilisés pour la prédiction à court terme (avant quelques heures) de la concentration de PM 2.5 (les particules volantes dont leur diamètre est 2.5 micromètre). Cette étude a montré que le modèle neuronal est le plus performant.

La prédiction à long terme dans le domaine de la prédiction de la pollution de l'air ne dépasse pas 30 heures. (Pérez et Reyes 2002) présentent un modèle basé PMC pour la prédiction de la concentration de PM10 avant 30 heures. Le modèle reçoit en entrée les mesures de la concentration de PM10 à 9 : AM (dans la matinée), et aussi les paramètres météorologiques prédits. Le modèle proposé a montré une grande efficacité à prédire les journées polluées par rapport à un modèle linéaire, qui est le réseau lui-même avec la fonction d'activation neutre $f=1$.

Dans (Rob et *al.* 2001) une approche de détection de non linéarité est présentée, qui se base sur l'analyse de la série temporelle et l'application de la méthode DVS (Delay Vector) qui consiste à calculer pour chaque vecteur une distance et d'extraire les K vecteurs les plus proches (dont leurs distance est plus faible). Ce modèle est utilisé pour calculer les prédictions en utilisant le sous-ensemble de test, et les erreurs sont calculés en variant le nombre de voisins pris (K), jusqu'à avoir une valeur qui donne le plus d'optimalité pour l'erreur. Cette valeur indique la non-linéarité de la série temporelle comme suit :

- Si K est proche au nombre total des vecteurs de retardement, alors le modèle linéaire est plus adaptable pour la série, et le modèle est équivalent à un AR
- Si K est plus petit ou moyen le modèle non linéaire est plus adéquat.

En utilisant cette méthode une analyse a été faite sur la série temporelle de la concentration de l'ozone, une fois que la non linéarité a été détecté deux modèles ont été conçus. Un de topologie Feed-Forward et un autre de topologie récurrent dite NARMA(p,q) (Nonlinear Autoregressive Integrated Moving Average) . Cette étude a montré que le modèle NARMA avec Autorégressive d'ordre p et de « moving average » (MA) d'ordre q , est un outil très efficace pour la prédiction de la concentration de l'ozone, et que les deux topologies ont des performances très semblables.

L'influence des paramètres météorologiques sur la précision des prédictions a été bien étudiée dans (Jef et *al.* 2005). Pour chaque paramètre les auteurs ont mis en évidence son influence sur les performances du modèle, cela s'est fait en construisant pour chaque paramètre un réseau de neurone et utiliser la concentration de PM10 précédente et le paramètre en question. Ce qui a permis de concevoir 5 modèles pour comparer les résultats et pour détecter le paramètre qui a le plus d'influence sur les performances du modèle. Le paramètre qui a montré le plus d'importance est le BLH (la masse atmosphérique de la journée). Ce paramètre influence directement sur le temps de séjour des particules dans l'atmosphère), reste à dire que cette conclusion n'est valable que pour la région concerné par l'étude. Une autre méthode proposée par (Giorgio Corani 2005) qui consiste à éliminer le paramètre qui possède le poids le plus faible dans un réseau à feed-forward, par l'application de la méthode OBD (Optimal Brain Damage) (Strok et Hassibi 1993) Qui fait tendre les poids les moins influents à zéro, éliminant les neurones les moins significatives. Ceci permettra d'obtenir une topologie adéquate pour le PMC. Dans cette étude, les auteurs présentent un autre modèle Lazy learning qui se base sur la notion de point de requête (Query-point), qui consiste à un vecteur $Q(t) = [Q1(t), Q2(t), \dots, Qn(t)]$, qui correspond aux valeurs des

paramètres pendant le temps (t), et un autre point dit voisin qui est un vecteur U disponible dans la matrice des données, où $U(\bar{t}) = [u_1(\bar{t}), u_2(\bar{t}), \dots, u_n(\bar{t})]$, qui correspond à la sortie $Y(T+1)$ déjà connus, en calculant une distance entre le point de requête est toutes les points voisins, cela permet de trouver les K points les plus proches, c'est-à-dire avoir à prendre $K \min < k < k \max$, et à prendre la valeur adéquate par rapport au point de requête. Le principe de ce modèle est très semblable au raisonnement à base de cas CBR. Comparé avec un réseau de neurone de type FeedForward, le modèle LL a été le plus performant pour prédire le PM10 et L'Ozone, mais comme pour toute les études cela n'est valable que pour la région de Milan en Italie.

La plupart des modèles proposés génèrent la valeur de la concentration du polluant étudié, dans (Jorquera et al 1998) le modèle proposé doit fournir une décision concernant la classe de qualité de l'air (A: good, B: bad, C: critical). Le modèle possède trois sorties chacune d'entre elle représente une classe de la qualité de l'air, le modèle utilisé est un réseau de neurones artificiels de type PMC, ce modèle a été comparé avec un modèle linéaire (un réseau de neurone sans fonction d'activation qui est équivalent à un modèle ARX), les deux possèdent le même nombre de paramètre, et ont montré des prédictions très affinés, mais le modèle non linéaire (le PMC) dépasse significativement le modèle linéaire équivalent. Ceci a été aussi montré par (Bianchini et al.2006) après création de N réseaux de neurones artificiels de type PMC, où N est le période de la cyclo-stationnarité de la série temporelle de la concentration de NO₂, le modèle a donné des prédictions efficaces.

Plusieurs RNA de différents types (PMC, RBF, PMC quadratique) sont comparés dans (Ordieres et al 2005), ces trois modèles ont été utilisés pour la prédiction de la concentration des PM2.5. Le modèle qui a montré le plus d'efficacité a été le RBF (Radial Basis Function). (Sousa et al. 2007) appliquent un traitement ACP (Analyse en Composante Principale) aux vecteurs de paramètres pour réduire la complexité de modèle, l'utilisation de cette méthode a permis d'améliorer les performances des prédictions d'un modèle neuronal de type PMC.

Les modèles proposés dans (Cigizoglu et al 2006) permettent de donner des estimations sur la concentration de dioxyde de soufre (SO₂) dans la ville d'Istanbul en Turquie, en prenant en paramètres la concentration mesurée de monoxyde de charbon et la concentration mesurée du monoxyde du nitrate. Dans cette étude trois modèles sont proposés : Un modèle se base sur un réseau de neurones de type RBF (Radial Basis Function) (Broomhead and Lowe 1988), un modèle basé PMC, et un modèle basé régresseur linéaire. Ces trois modèles ont été utilisés pour faire des prédictions long-terme de SO₂, les deux modèles neuronaux ont montré des

bonnes performances, quoique le modèle basé RBF a présenté des comportements de surestimations des pics. Cela peut être due au fait que les pas de mesure des données est d'une journée, ce qui donnera l'effet de prendre en considération que les pics. Les mêmes modèles ont été utilisés dans (Cigizoglu et al 2006), pour l'étude de l'éventuel influence des concentrations de CO et du NO sur la concentration de PM10. Quatre modèles ont été conçus un basé PMC prenant comme paramètres la concentration de CO et un autre de même topologie mais prenant en paramètre la concentration mesurée de NO, et un autre Modèle basé RBF prenant en paramètre la concentration mesurée de CO et un autre de même type mais avec la concentration NO comme entrée. Les choix des paramètres a été validé par l'application de l'ACP (Analyse en Composante Principale) ce qui a permis de voir le degré de l'importance de chaque paramètres (polluant). Les quatre modèles ont permis de voir clairement la relation et l'influence des concentrations des polluants primaires tels que CO et NO sur la concentration des PM10.

Une autre approche d'étude de relation entre la concentration des polluants secondaires et les concentrations des polluants primaires est présentée dans (Acciani et al 2006). Deux modèles basés réseau de neurones artificiels sont utilisés pour étudier les paramètres qui influencent le plus sur la concentration de l'Ozone O₃ et de PM10. Ces deux polluants sont de types polluants secondaires qui résultent de l'interaction chimique des polluants primaires. Pour voir le polluant primaire et le paramètre climatique le plus important, une méthode dite de partitionnement est appliquée. Cette dernière se base sur l'index nommé « index d'importance relative » RII (Relative Importance Index) qui permet de voir combien un élément d'entrée affecte-il la sortie. L'index d'importance relative se base sur les poids d'un neurone dans un réseau de type PMC. Après l'étape d'apprentissage et il est donné comme suit :

$$RI = \frac{\sum_{j=1}^h \left| \frac{w_{ij} w_{kj}}{\sum_{i=1}^n |w_{ji}|} \right|}{\sum_{i=1}^n \sum_{j=1}^h \frac{|w_{kj}| |w_{ji}|}{\sum_{i=1}^n w_{ji}}} \quad (1.7)$$

Où :

W_{ij} c'est le poids de connexion de l' i -ème neurone d'entrée avec le j -ième neurone de la couche cachée, et W_{kj} c'est le poids de connexion entre le j -ième neurone de la couche cachée avec le k -ième neurone de la couche de sortie. L'index d'importance relative RII est donc le degré d'importance qui lie le paramètre i avec la sortie k , passant par le neurone j de la couche cachée. Une sélection des paramètres est appliquée afin de prendre en considération que ceux qui ont un index d'importance relative important, un autre réseau de neurone est conçu prendra comme paramètre que les paramètres sélectionnés, ce qui permettra d'optimiser le modèle et augmenter ses performances. L'approche peut être résumée comme suit :

1. Conception d'un réseau de neurone de type PMC à une seule couche cachée avec tous les paramètres,
2. Calcule pour chaque paramètre l'index d'importance relative,
3. Sélection des paramètres qui ont le RII le plus important,
4. Conception d'un autre réseau de avec les paramètres sélectionnées en 3.

Paramètres utilisés dans l'étude :

paramètres	unités
SO ₂	µg/m ³
NO _x	µg/m ³
NO	µg/m ³
NO ₂	µg/m ³
CO	µg/m ³
Benzène	µg/m ³
Toluène	µg/m ³
O-xilème	µg/m ³
Direction du vent	Secteur
Vitesse du vent	m/s
DVG	Secteur
Radiation	W/m ²
Pluviosité	Mm

Tableau -1.2- paramètres utilisés.

Cette étude a montré que le CO, Toluène, vitesse du vent et la pluviosité sont les paramètres les plus influençant sur la concentration de l'Ozone O₃. Dans le cas de PM10 CO, le Benzène et la vitesse du vent sont plus important par rapport aux autres paramètres. Pour avoir une bonne prédiction de la qualité de l'air (Kolehmalnen et *al* 2000) ont proposé une approche de datamining, qui permet de construire un modèle pour chaque épisode de la série temporel de pollution. Deux méthodes de clustering ont été appliquées, une basée carte de Kohonen (Kohonen, 1995) et une autre basée carte de Sammon, ces deux méthodes ont permis d'avoir sept épisodes de pollution, pour chaque épisode un réseau de neurones de type PMC a été conçu pour prédire l'épisode, ce qui a permis d'avoir une prédiction multi-modèles où chaque modèle est spécialisé dans la prédiction d'une partie de la série temporelle.

La non-linéarité de la série temporelle de la concentration de l'ozone est bien étudiée dans (Bordignon et *al* 2002). Trois modèles ont été conçus pour faire des prédictions à court-terme de la concentration de l'ozone O₃, un modèle basé modèle Additif, un modèle basé MARS (Multivariate Additive Regression Splines) (Friedman, 1991), et un modèle basé réseau de neurones artificiels. L'approche adoptée consiste à faire une représentation mathématique et à chercher le modèle le plus performants qui simule le plus fidèlement possible la relation entre Y les prédictions et le vecteurs des paramètres X . Les trois modèles ont montré de bonne performance, la seule critique c'est que les méthodes non-linéaires ne permettent pas une étude explicative du phénomène malgré leurs résultats incontestables.

La prédiction à court-terme des concentrations de dioxyde de soufre SO₂ et dioxyde de nitrogène NO₂ et PM10 a été proposée dans (Sharma et *al* 2008), l'approche Box-Jenkins a été adoptée pour concevoir des modèles basés ARIMA (autoregressive integrated moving average). L'approche peut être résumée comme suit :

- Identification des spécifications préliminaires du modèle,
- Estimation des paramètres du modèle,
- Diagnostique et validation de l'adaptabilité du modèle,
- Prédiction.

La série temporelle est composée en plusieurs parties : autorégressive notée AR, intégration I, Moving Average notée MA. Supposant que Z_t est notre série temporelle (définie par une séquence de N observations qui sont équidistantes dans le temps (z_1, z_2, \dots, z_N)), l'opérateur d'intégration Δ et Δ_s définie comme suit :

$$\Delta z_t = z_t - z_{t-1} \text{ et } \Delta_s z_t = z_t - z_{t-s} \quad (1.8)$$

Où s est la période ou le pas du mesure des observations.

Mettant l'opérateur B comme suit :

$$Bz_t = z_{t-1} \quad \text{Et} \quad B^s z = z_{t-s} \quad (1.9)$$

La partie AR notée Φ se constitue d'un polynôme d'ordre p de B comme suit:

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (1.10)$$

La partie MA notée θ définie comme suit :

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (1.11)$$

La première étape consiste à appliquer une analyse de la série temporelle par l'utilisation d'ACF (autocorrelation function) qui permet de mesurer le degré de dépendance linéaire entre les observations. L'application de PACF (Partial Autocorrelation Function) permet aussi de voir l'importance de la partie AR dans le modèle. Après application des règles citées dans (Box et al 1994). Le modèle issue est de classe ARIMA (p, d, q) est peut être formalisé comme suit :

$$\phi_p(B)\theta_q(B)\Delta_s^d \Delta^d x_t = \theta_q(B)\Theta(B^s)a_t + c \quad (1.12)$$

Après application de l'algorithme MSQ (Marquardt least-square) les paramètres de modèle ont été identifiés et le modèle est formalisé comme suit :

$$\begin{aligned} [w_{t+1}] = & \phi'_1[w_{t+l-1}] + \phi'_2[w_{t+l-2}] + \dots + \phi'_{p+sp}[w_{t+l-p-sp}] + [a_{t+1}] - \\ & \theta'_1[a_{t+l-1}] - \theta'_2[a_{t+l-2}] - \dots - \theta'_{q+sq}[a_{t+l-q-sq}] + c \end{aligned} \quad (1.13)$$

Où :

- L : est le temps de prédiction,
- W_{t+l} et Φ'_1 sont les paramètres AR généralisés,
- θ'_l sont les paramètres MA.

Les résultats obtenus dans cette étude ont montrés que les modèles stochastiques linéaires tel que ARIMA permettent d'avoir des descriptions quantitatives de la pollution de l'air. L'approche permet d'avoir des prédictions efficaces avec moins d'effort de calcul. Un autre

modèle linéaire est présenté dans (Kim et al 2005) un modèle basé ARX (Autoregressive a entrée eXogène) est proposé pour la prédiction à court-terme de la concentration de l'ozone O₃ dans la ville de Cincinnati, OH, en Etats Unies d'Amérique (USA).

Le modèle ARX formalisé comme suit :

$$y_t = \mu + \sum_{j=1}^p \phi_j y_{t-j} + \beta' x_t + w_t, t = q+1, \dots, n \quad (1.14)$$

Avec $\Phi \neq 0$, et p l'ordre d'autorégressive, $X_t = (x_{1t}, x_{2t}, \dots, x_{lt})$ est les vecteurs des entrées exogènes.

Dans cette étude plusieurs modèles ont été conçus :

- FARX (Fonctional Coefficient AutoRegressive with eXogenous inputs) : permet d'ajouter les paramètres direction du vent et temps de la journée (nuit ou matin),
- TARX (Tong AutoRegressive with eXogenous inputs) ,
- AR modèle autoregressive,
- ARX. (AutoRegressive with eXogenous inputs)

Chaque modèle a montré une certaine capacité à prédire correctement la concentration du polluant, quoique le modèle basé FARX a été le plus performant.

Dans (Castellano et al 2009) une étude statistique des séries temporelles de NO_x et l'Ozone est présentée. Dans cette étude une approche statistique d'indentification des épisodes de pollution a été suivie, l'analyse statistique a été appliquée sur les données des paramètres de pollution de trois zones en Espagne : Guarda, Corrubedo et Verín (deux ville côtières et une ville intérieure). Pour chaque ville un modèle basé ARIMA a été conçu pour donner des prédictions à long-terme sur la concentration de O₃, parmi les paramètres inclus dans le modèle on peut citées :

- La concentration de SO₂
- La concentration de SH₂
- La concentration de NO
- La concentration de NO₂
- La concentration de CO

- La concentration de l'Ozone,
- La concentration des PM10
- Mesure de la température,
- Vitesse et direction du vent,
- La pluviométrie
- La pression atmosphérique,
- Humidité et rayonnement solaire,

Dans (Carmichael et al. 2008) une approche CTM (Chemical Transport Models) a été utilisée pour la conception d'un système de contrôle de l'émission des polluants dans l'air, qui consiste à trouver un modèle mathématique qui permet d'inclure les éléments suivants : émission, transport, transformation et destruction des pollutions dans l'air. En se basant sur les paramètres chimiques du polluant et de son processus de transformation et aussi de transport (notamment les paramètres météorologiques) un modèle ARIMA a été conçu pour la prédiction de la concentration de l'Ozone et de PM10. Le modèle a présenté des résultats intéressants pour la prédiction de la concentration des polluants secondaires, tel l'Ozone (O₃). Mais l'adaptation d'un tel modèle pour un polluant primaire reste une tâche très complexe surtout quand les données sur l'émission de ce polluant ne sont pas disponibles. Une approche hybride est présentée dans (CIPRIAN et al 2002), dans laquelle un modèle Neuro-fuzzy a été intégré dans un système à base de connaissance dans le but de prédire la qualité de l'air. Un modèle connexionniste (Neuronal) est créé en se basant sur les données disponibles afin d'extraire les connaissances implicites, en plus un modèle basé logique flou est créé en se basant sur les connaissances explicites des experts du domaine de pollution de l'air, le schéma global de modèle est illustré dans la figure 1.4. A chaque fois un réseau de neurones artificiels est créé. Ensuite et après les étapes d'apprentissage des méthodes d'extraction des connaissances sont appliquées pour extraire les informations symboliques du réseau créé. Un extrait des règles produites est montré dans le tableau -1.3- :

IF RainToday is YES THEN NO2After10 is L1 (74.80%)
IF InversionToday is L0 THEN NO2After10 is L1 (58.95%)
IF InversionToday is L1 THEN NO2After10 is L2 (50.76%)
IF InversionToday is L2 THEN NO2After10 is L2 (33.32%)

IF InversionToday is L3 THEN NO2After10 is L2 (61.16%)
IF SolRadAt13 is LOW THEN NO2After10 is L2 (95.10%)
IF SolRadAt13 is MED THEN NO2After10 is L2 (100.00%)
IF SolRadAt13 is HIGH THEN NO2After10 is L2 (45.70%)
IF WindDirect_10 is NV THEN NO2After10 is L3 (35.56%)
IF WindDirect_10 is NE THEN NO2After10 is L3 (47.47%)
IF WindSpeed_10 is LOW THEN NO2After10 is L2 (34.07%)
IF WindSpeed_10 is MED THEN NO2After10 is L1 (61.32%)
IF WindSpeed_10 is HIGH THEN NO2After10 is L3 (60.81%)
IF NO_10 is MED THEN NO2After10 is L2 (31.54%)
IF NO_10 is HIGH THEN NO2After10 is L1 (54.82%)
IF NO_10 is HIGH THEN NO2After10 is L2 (44.99%)
IF NO2_10 is L2 THEN NO2After10 is L1 (99.40%)
IF NO2_10 is L3 THEN NO2After10 is L1 (41.75%)
IF NO2_10 is L3 THEN NO2After10 is L2 (45.33%)
IF NO2_10 is L4 THEN NO2After10 is L1 (73.23%)

Tableau -1.3- : Un extrait des règles construites.

Le modèle a été utilisé pour la prédiction de pics de concentration de NO₂ et a montré de bonnes performances. Un autre modèle hybride est présenté dans (Maruf et al 2008), ce modèle consiste à l'utilisation de HMM (Hidden Markov Model) et la logique floue (HMM-fuzzy) pour prédire la concentration du NO_x et NO₂. L'approche consiste à appliquer les étapes suivantes :

1. L'HMM est utilisé pour partitionner le sous-ensemble d'apprentissage,
2. Un Algorithme Top-Down est appliqué pour extraire le nombre minimum de règles (logique floue) qui permettent d'atteindre l'erreur calculée dans l'étape 1,
3. L'algorithme descente gradient est appliqué pour extraire les paramètres de modèle.

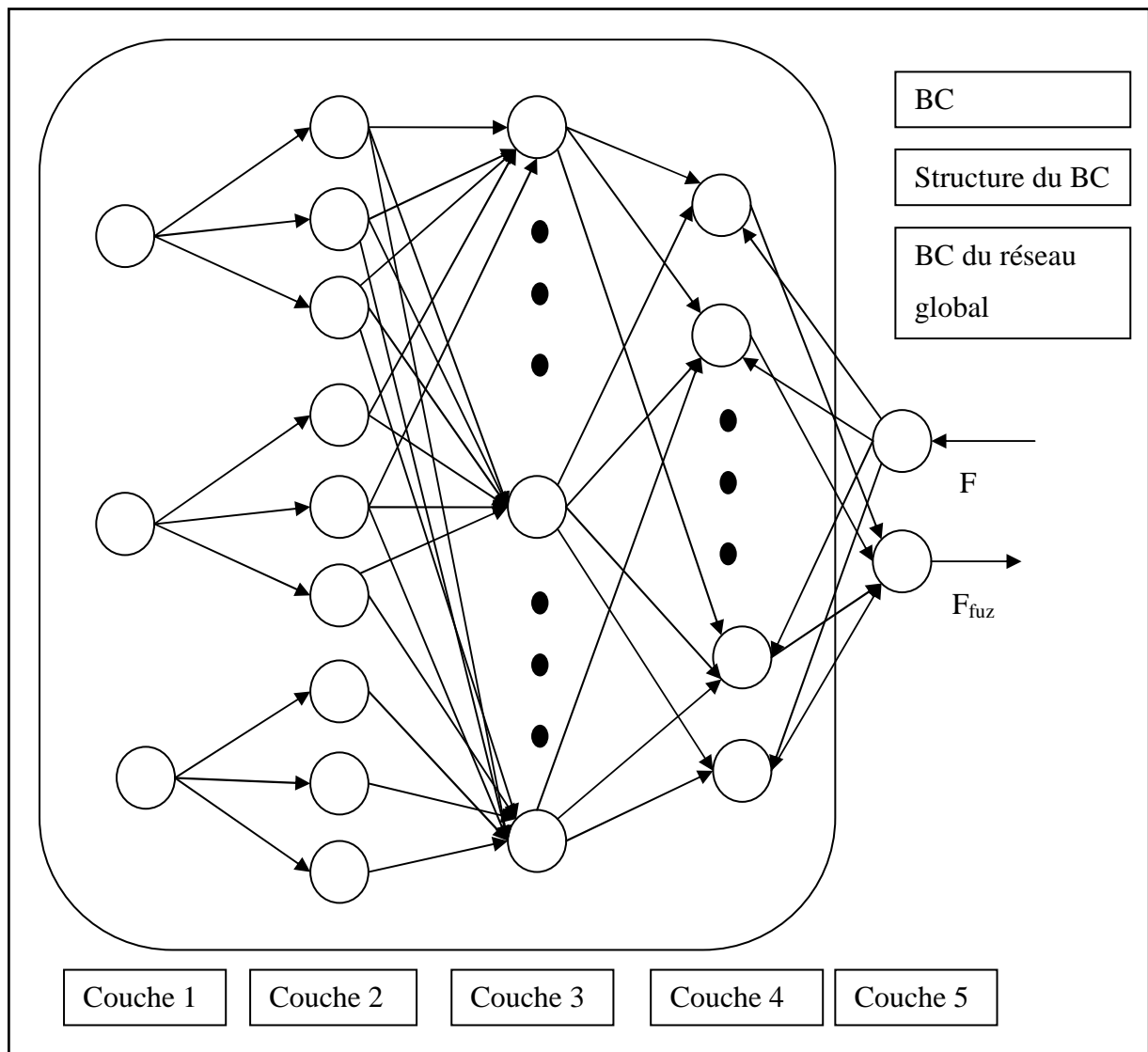


Figure -1.4- : Schéma de modèle hybride de prédiction de qualité de l'air.

Le modèle prend en entrée les paramètres de l'émission, les paramètres métrologiques, et fournit en sortie la prédiction de la concentration de NO_2 . Le modèle a montré de bonne performance. Dans la plupart des études un seul polluant est étudié, dans (İçağa et Sabah 2008) la relation entre deux polluants est étudiée. Un modèle basé PMC a été conçu pour détecter l'éventuelle relation entre dioxyde de soufre SO_2 et les PM (particule en suspension dans l'air). Cette étude a permis de conclure que l'humidité est l'un des facteurs les plus importants à affecter la relation entre ces deux polluants. Ce qui a permis de tirer des conclusions à propos de l'influence de SO_2 et l'humidité sur le processus de transformation / destruction de PM_{10} . Dans (Shuiyuan et al 2006) un modèle de prédiction et de simulation basée 3-D multi-box (3DMB) est proposé. Le modèle prend en considération les paramètres relative à l'émission, c'est pour ça que la zone étudiée a été découpée en 4 sous-zone : une

zone industriels, une zone semi-industriel (a activité industriel légère), une zone urbaine, et une zone touristique et culturelle. Pour chaque sous-zone un modèle 3DMB est conçu pour prédire la concentration des TSP (Total Suspend Particulate) et l'SO₂ et les PM10. Ce type de modèle se base sur :

- Les données d'émission,
- Le principe de conservation de masse (la capacité de l'air à garder les composants polluants),
- Dispersion tridimensionnelle des polluants,
- Distribution de polluant dans l'air,
- Interaction chimique entre les composants de l'ait et le polluant.

Un exemple de Box-Gaussien peut être illustré comme dans la figure -1.5-

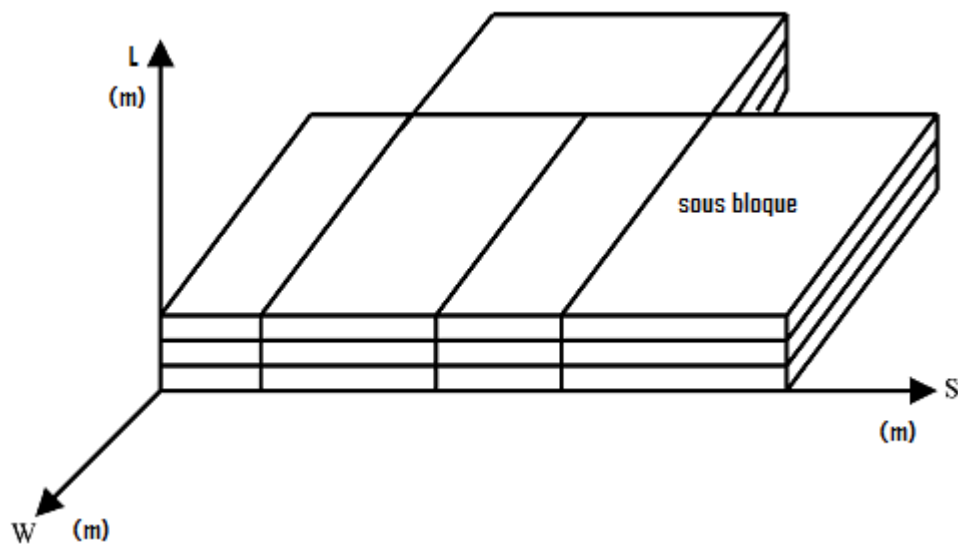


Figure -1.5- : Modèle en box de pollution de l'air.

Chaque sous-cube est formulé comme (1.15). Les sous-cube permettent de rendre la prédiction plus précise en indiquant à chaque fois la sous-zone qui donne le plus de concentration et permet de bien localiser les pics de pollutions. Ce qui permettra aux utilisateurs d'avoir plus de détails sur le phénomène pour prendre les mesures nécessaires pour limiter et contrôler la pollution.

$$\begin{aligned}
& L_5 H U_{n3} C_{n03} - L_5 H U_{n3} C_{n15} + E L_5 W_5 (C_{n10} - C_{n15}) / H \\
& + 2 E' W_2 H (C_{n12} - C_{n15}) / (L_2 + L_5) \\
& + 2 E' W_3 H (C_{n13} - C_{n15}) / (L_3 + L_5) \\
& - L_5 H U_{n3} C_{n15} [1 - \exp(-K W_5 / U_{n3})] \\
& - L_5 W_5 C_{n15vg} - L_5 W_5 W_D C_{n15} P / 4 = 0
\end{aligned} \tag{1.15}$$

Ou:

L_i , W_i : respectivement la longueur et la largeur de sous-cube, M : la moyenne de vitesse du vent, C_n : la concentration de polluant dans la n -ième sous-cube, E : coefficient de dispersion vertical, H : coefficient de dispersion horizontal, P : pression. Ce modèle a permis de donner de bons résultats, et permet d'avoir la possibilité de bien localiser la pollution en divisant les zones étudiées en sous-zones, mais ce type de modèle n'est valable que pour les zones dont les paramètres et les données relatives à l'émission sont disponibles. Une étude comparative entre l'approche linéaire et non-linéaire est présentée dans (Comrie 2007), cette étude a été basée sur les données relatives à la concentration de l'Ozone (O_3) qui est un polluant secondaire (résulte de l'interaction photochimique des autres polluants), ce qui a favorisé le modèle linéaire, car pour d'autres polluants primaires le modèle risque de ne pas avoir les paramètres relatifs à la transformation (construction/ destruction) du polluant. Dans (Brunelli et al 2006), un modèle basé sur un réseau d'Elman est présenté, le modèle donne des prédictions concernant la concentration des SO_2 dans les zones industrielles de Priolo en Italie, le modèle reçoit comme paramètres la concentration passée du SO_2 et aussi les valeurs de l'émission des sites industriels ciblés par l'étude. Le modèle a donné de bonnes performances de prédiction.

Chapitre 2

Données utilisées et région ciblée

2.1 Région ciblée

Le modèle développé dans cette étude concerne la région d'Annaba dans le nord-est de l'Algérie. Comme illustré dans la carte géographique dans la Figure 2.1. Annaba est l'une des plus grandes villes en Algérie connue aussi comme étant l'un des pôles de l'industrie Algérienne, surtout dans le domaine de sidérurgie. En fait l'usine d'Arcelor-Metal (ex EL-Hadjar) est l'une des plus grande usine en Afrique avec un taux de production avoisinant le 1 million de tonnes par ans. Malgré les efforts effectués, l'influence de cette usine sur l'entourage et l'écosystème reste de grande rigueur. Dans le processus de production de l'acier beaucoup d'organes chimiques sont utilisés tel que le charbon qui est émis dans l'air lors de l'étape de la mise des composants chimiques dans les grands fours, ce qui cause la dispersion dans l'air de ces composants chimiques, et sous certaine condition climatique ces dispersions forment des nuages de pollution qui constituent des pics de pollution de l'air.

Les stations pétrochimiques (Naftal, Asmidal) qui constituent des sources de pollutions les plus critiquées malgré les efforts effectués dans le but de réduire l'émission de ces stations, l'influence reste important dans la pollution de l'air à Annaba.



Figure 2.1 : Carte géographique de Annaba (Google ©).

La topographie de la région d'Annaba (comme illustrée dans la figure 2.2) est en forme de cuvette favorisant ainsi la stagnation de l'air et la formation d'inversions de températures. Ces situations permettent l'accumulation des polluants et l'élévation des taux de concentration qui en résulte. Les effets des brises de mer, terre, et pente concourent au transport des nuages de polluants. En effet, les nuages de polluants sont entraînés par la brise de terre la nuit vers la mer, et de jour. Ces nuages de polluants retournent sur la ville par effet de brise de mer en longeant la montagne de Edoug. Les nuages tournent sur la ville sous une forme de cercle. Les polluants se déposent lentement par gravité et l'on assiste à une pollution affectant les trois récepteurs (mer, terre, air). L'industrie est l'élément moteur de dégradation de la qualité d'air, cette industrialisation a assurément permis de répondre aux besoins des populations et du pays en produits sidérurgiques, engrais azotés, constructions ferroviaires et autres industries de transformations. A l'inverse, elle a suscité une urbanisation démesurée de la ville avec tous ses corollaires et une pollution de l'atmosphère et des sols, suivies de conséquences néfastes sur le biotope et la société



Figure 2.2 : Topographie de la région d'Annaba (Google Earth ©)

2.2 Données utilisées dans l'étude

Les données utilisées pour ce travail, proviennent du réseau national de contrôle de qualité de l'air nommé Samaa Safia. Ce dispositif d'évaluation du degré réel de la pollution urbaine et industrielle dans les grandes villes entre dans le cadre de l'accord d'association avec l'Union européenne (UE) qui exige certaines normes internationales à respecter et l'intégration dans la politique atmosphérique euro-méditerranéenne.

En date du 19 octobre 2005, le Conseil de Gouvernement a examiné et adopté un décret exécutif définissant les valeurs limites, les seuils d'alerte et les objectifs de qualité de l'air, en cas de pollution atmosphérique, présenté par le Ministre de l'Aménagement du Territoire et de l'Environnement.

Pris en application de la loi relative à la protection de l'environnement dans le cadre du développement durable, le texte vise à préciser les seuils d'alerte à partir desquels doivent être déclenchées les mesures d'urgence visant à réduire les activités polluantes par les collectivités locales concernées, l'objectif étant de prévenir des situations où une exposition, même de courte durée, qui pourrait entraîner des effets négatifs sur l'homme et sur l'environnement.

Ce réseau possède plusieurs stations de mesure des différents polluants, ces stations sont installées dans les différentes régions de l'Algérie. Ces stations sont opérationnelles depuis 2003. Elles sont reliées à un réseau commun au niveau du ministère de tutelle. Un réseau similaire est en cours d'étude à Oran et d'autres villes algériennes. Nous avons focalisé notre travail sur les données provenant des stations implantées dans la wilaya de Annaba constitué de stations installées dans les régions qui enregistre des taux alarmant de pollution de l'air et qui contient des zones industrielles trop proches aux zones urbaines telles que la commune de Sidi Amar et El-Bonni.

Les données qu'on a pu récupérer sont des mesures horaires des années 2003 et 2004 à partir de la station numéro 4 exactement dans la commune de Sidi Amar (voir figure 2.1), qui consiste en des matrices pour chaque mois avec quelques plages vides qui sont causés par des pannes et des journées dans lesquelles les stations n'ont pas mesurer ou les données ont été perdus.

2.2.1 Les polluants mesurés par Sama Safia

Les stations de Sama Safia surveillent en temps réel la concentration des polluants suivants :

Polluant	Description
SO ₂ : Dioxyde de soufre	Proviens généralement des combustions fossiles contenant du soufre (fuel, gasoil). Station de production de l'énergie, Véhicule à moteur.
NO ₂ : Dioxyde d'azote	Provient des installations de combustion (gaz naturel) et des véhicules automobiles.
CO : Monoxyde de carbone	Résulte de la combustion incomplète des combustibles fossiles (essence, fuel, charbon, bois).
O ₃ : Ozone	Polluant secondaire provenant de la transformation de certains polluants (NO ₂ et hydrocarbure) sous l'action du soleil.
PM10 : particulate matter	Poussière fines en suspension, émis par l'activité naturelles et humaines (moteur diesel, industrie etc).

Tableau 2.1 : Les polluants mesurés par Sama Safia.

2.2.2 Paramètres climatiques

En plus des polluants Sama Safia effectue des mesures des paramètres climatiques, car les mesures des concentrations des polluants seules ne permettent pas de donner des indices sur la qualité de l'air. Les paramètres climatiques mesurés par Sama Safia sont décrits dans le tableau suivant:

Paramètres	Description
Humidité	L'humidité influence énormément sur la capacité des composants chimiques de l'air à s'inter-réaction, de se déplacer d'une zone à une autre.
Température	La température est l'un des paramètres qui donne l'énergie nécessaire aux réactions photochimiques qui résulte des polluants secondaires
Vitesse du vent	Le vent déplace les masses de l'air d'une zone géographique à une autre. Avec une vitesse de vent faible les polluants se cumule dans la même zone ce qui créera une période de pics.

Tableau 2.2 : Les paramètres climatiques mesurés par Sama Safia.

	<i>Temperature</i>	<i>Humidité</i>	<i>Vitesse du vent</i>
	(°C)	(%)	(nœuds)
<i>Janvier</i>	12,285	75,973	6,8733
<i>Février</i>	11,891	77,938	7,5507
<i>Mars</i>	13,318	72,782	7,612
<i>Avril</i>	15,327	72,725	7,5625
<i>Mai</i>	19,144	74,219	7,1205
<i>Juin</i>	22,783	72,092	7,93
<i>Juillet</i>	24,786	71,135	8,4846
<i>Aout</i>	26,158	71,791	8,0258
<i>Septembre</i>	23,476	72,847	7,4396
<i>Octobre</i>	19,931	73,097	6,8879
<i>Novembre</i>	15,521	75,514	7,5688
<i>Décembre</i>	12,925	75,669	7,7968

Tableau 2.3 Paramètres météorologiques moyens dans la région d'Annaba.

2.2.3 Propriétés statistiques des données

En analysant statistiquement les données qu'on a pu récupérer, permet de mieux comprendre le phénomène afin de créer et concevoir le modèle le plus adéquate.

PM10

Ce sont petites particules contiennent une multitude de substances dont des métaux, des acides, du carbone et des hydrocarbures aromatiques polycycliques. Certaines articles sont émises dans les gaz d'échappement des véhicules tandis que d'autres sont formées dans l'atmosphère par des réactions chimiques entre les polluants émis dans les gaz d'échappement. Les particules fines aggravent les symptômes des personnes atteintes de maladies respiratoires

ou cardiovasculaires.les deux figures 2.3 et 2.4 montrent les propriétés de PM10 en 2003 et en 2004

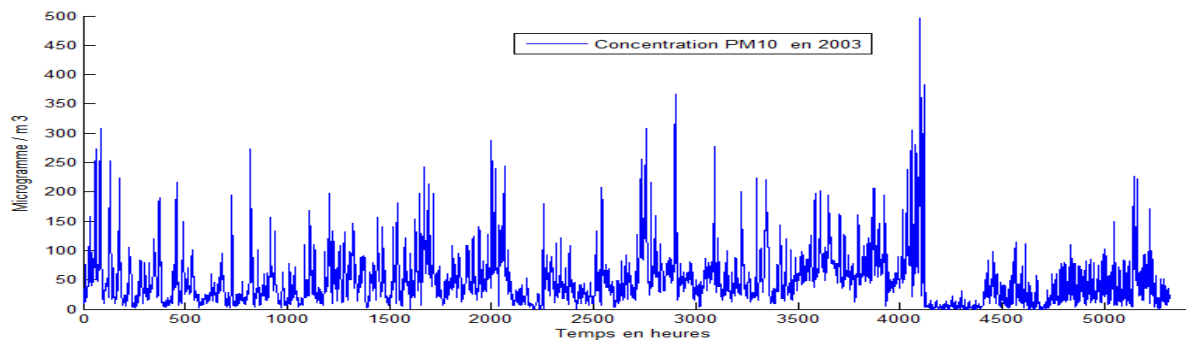


Figure 2.3 : Concentration PM10 en 2003.

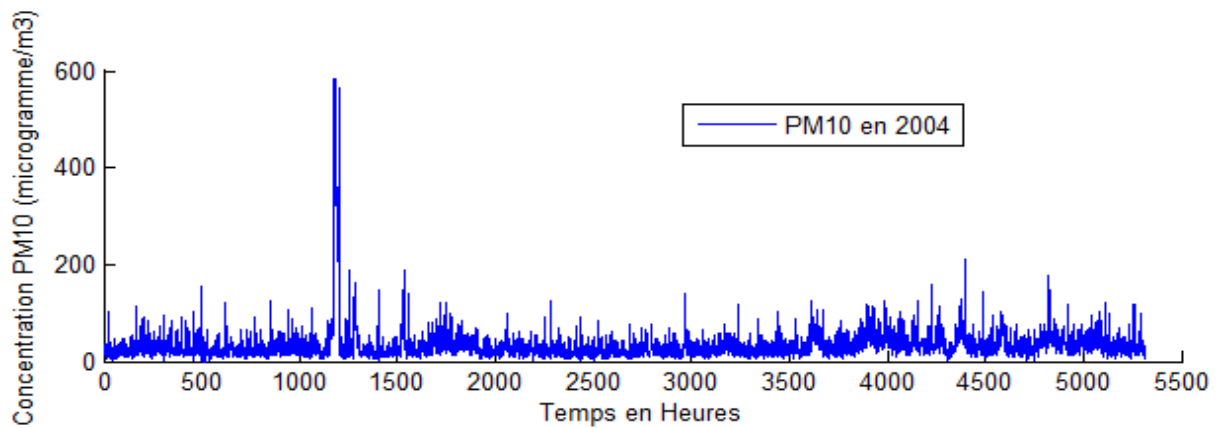


Figure 2.4 : Concentration P10 en 2004.

PM10	2003	2004
Moyenne (microgramme /m3)	49.0995	27.6830
Covariance	43.0445	26.3846
Valeur maximum (microgramme /m3)	496	508
Valeur minimum (microgramme / m3)	1.0200	13.0200

Tableau 2.3 : Propriétés statistiques de mesure de concentration de PM10.

Ozone (O₃)

Ce polluant n'a pas été pris en considération dans les mesures de l'année 2003, seules les données de 2004 sont disponibles, comme le montre la Figure 2.5.

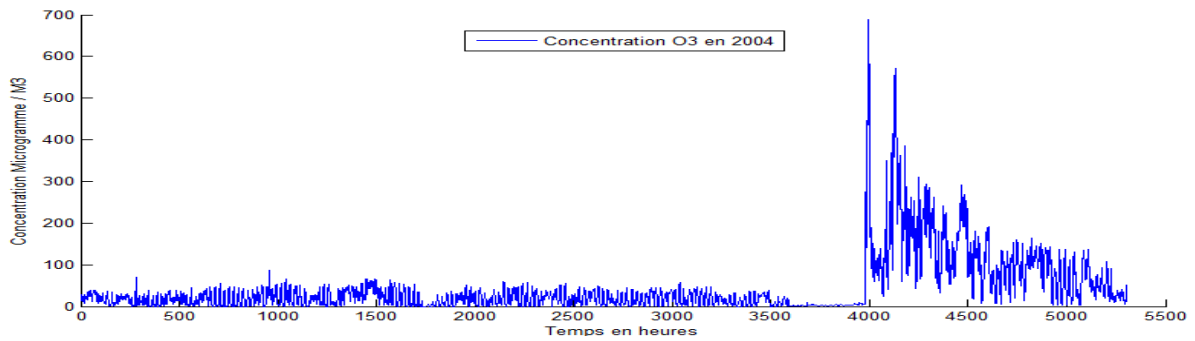


Figure 2.5 : Concentration O₃ en 2004.

Ozone	2004
Moyenne (microgramme / m3)	42.2740
Covariance	64.5807
Valeur maximum (microgramme / m3)	688
Valeur minimum (microgramme / m3)	3,000

Tableau 4 : Propriétés statistiques de mesure de concentration du l'Ozone.

NO_x

Ce vecteur englobe les mesures du NO₂ et du NO, comme montré dans les Figure 2.6 et 2.7.

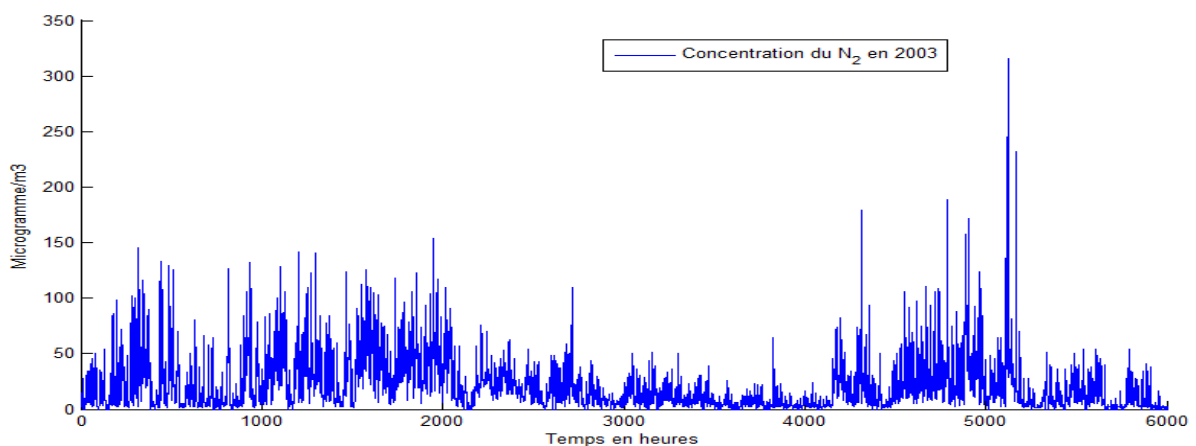


Figure 2.6 : Concentration NO_x en 2003.

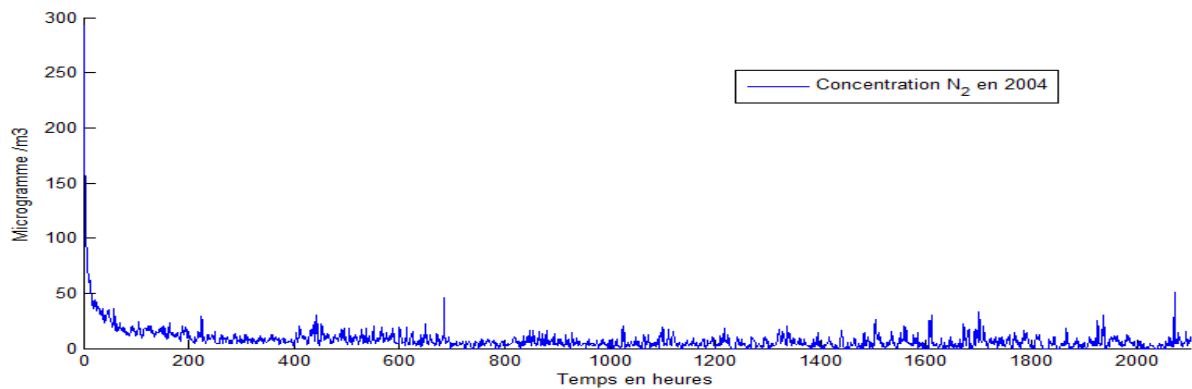


Figure 2.7 : Concentration NO_x en 2004.

NO _x	2003	2004
Moyenne	20.4698	7.2341
Covariance	22.8553	11.0005
Valeur maximum	316	293
Valeur minimum	2,0000	1,0000

Tableau 5 : Propriétés statistiques de mesure de concentration du NO_x.

NO

Les deux Figures 2.8 et 2.9 montent la série temporelle du NO.

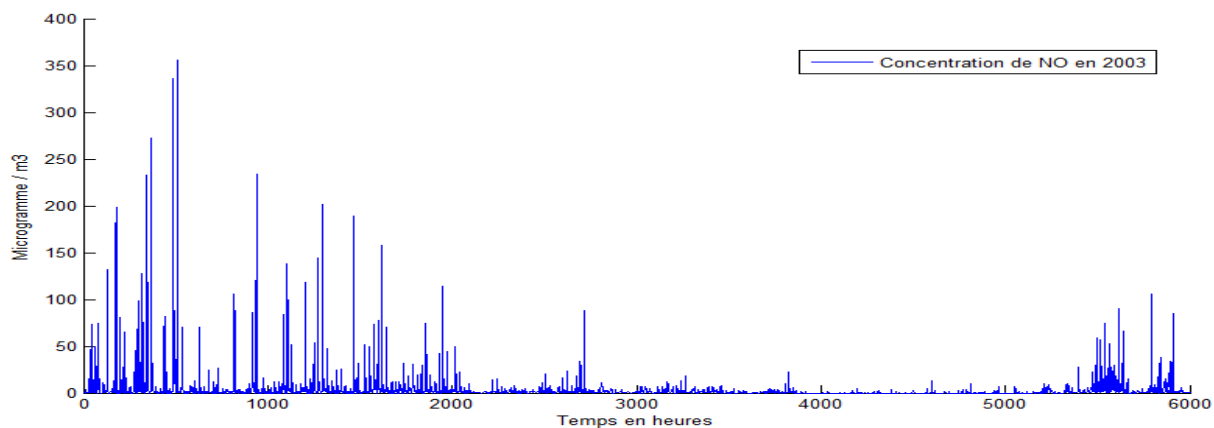


Figure 2.8 : Concentration du NO en 2003.

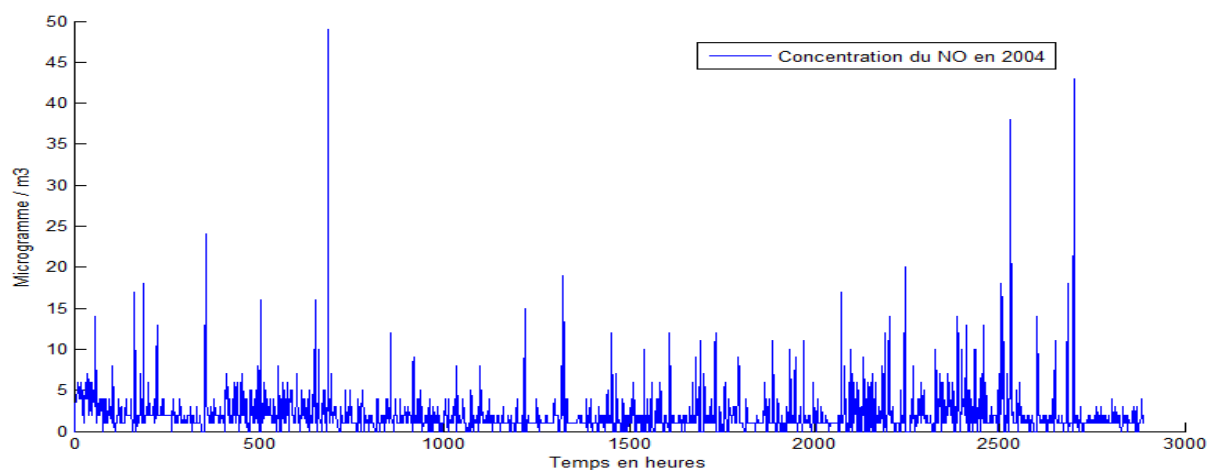


Figure 2.9 : Concentration du NO en 2004.

NO	2003	2004
Moyenne	3.7251	2.0768
Covariance	15.0672	2.6069
Valeur maximum	356	49
Valeur minimum	1,0000	1,0000

Tableau 6 Propriétés statistiques de mesure de concentration du NO.

CO_2

Pour ce polluant on n'a que les données du 2003, montré dans la Figure 2.10.

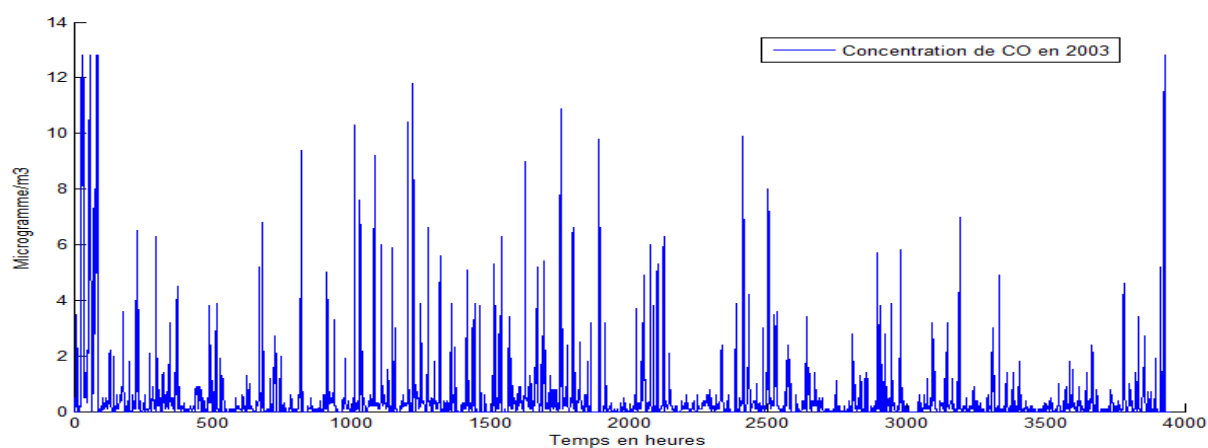


Figure 2.10 : Concentration du CO_2 en 2003.

CO ₂	2003
Moyenne	0.5279
Covariance	1.3287
Valeur maximum	12.8000
Valeur minimum	1,0000

Tableau 7 : Propriétés statistiques de mesure de concentration du CO₂.

NO₂

Pour ce polluant on n'a que les données du 2003, montré dans la Figure 2.11.

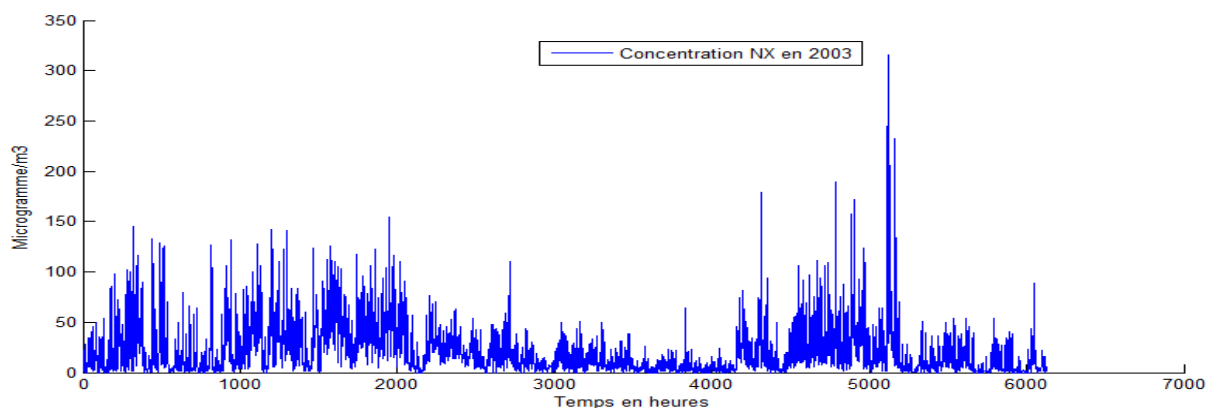


Figure 2.11 : Concentration du NO₂ en 2003.

NO ₂	2003
Moyenne	14.5021
Covariance	25.0133
Valeur maximum	435
Valeur minimum	1,0000

Tableau 8 : Propriétés statistiques de mesure de concentration du NO₂.

Humidité

Ce paramètre est marqué par la variation entre jour et nuit, comme montré dans les deux figures 2.12 et 1.13.

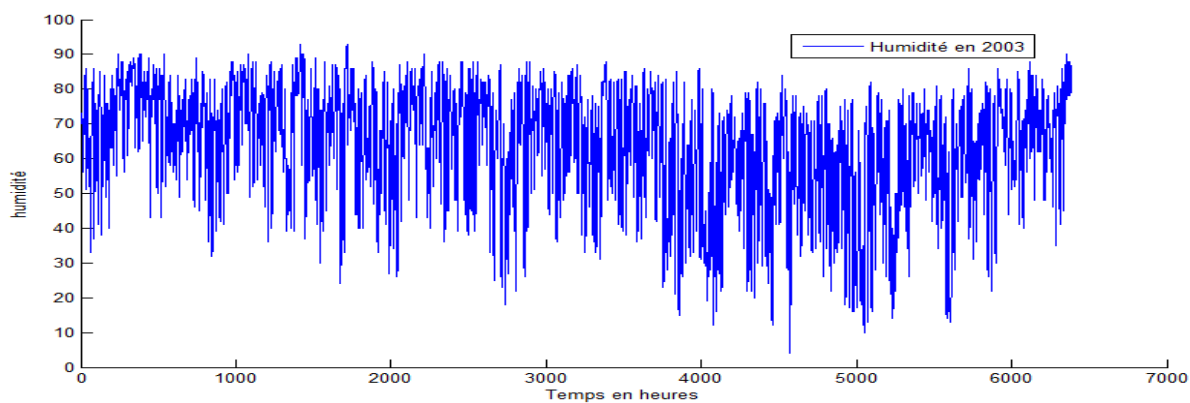


Figure 2.12 : Humidité en 2003.

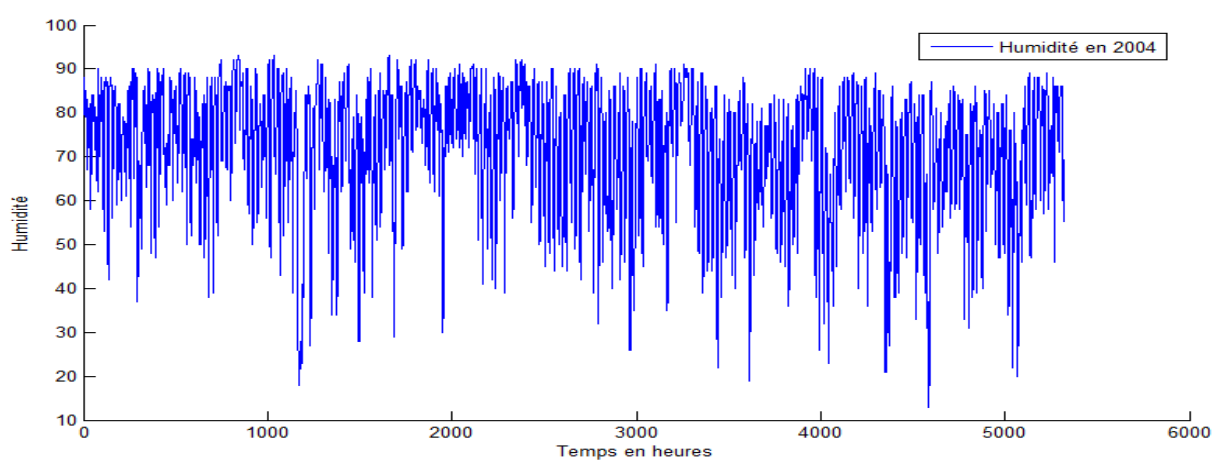


Figure 2.13 : Humidité en 2004.

humidité	2003	2004
Moyenne	63.5222	71.9237
Covariance	16.5087	14.3361
Valeur maximum	93	93
Valeur minimum	4	13

Tableau 9 : Propriétés statistiques de mesure de l'humidité.

Température

Les variations entre été et hiver est un aspect marquant ce paramètre, comme le montrent les deux figures 2.14, 2.15.

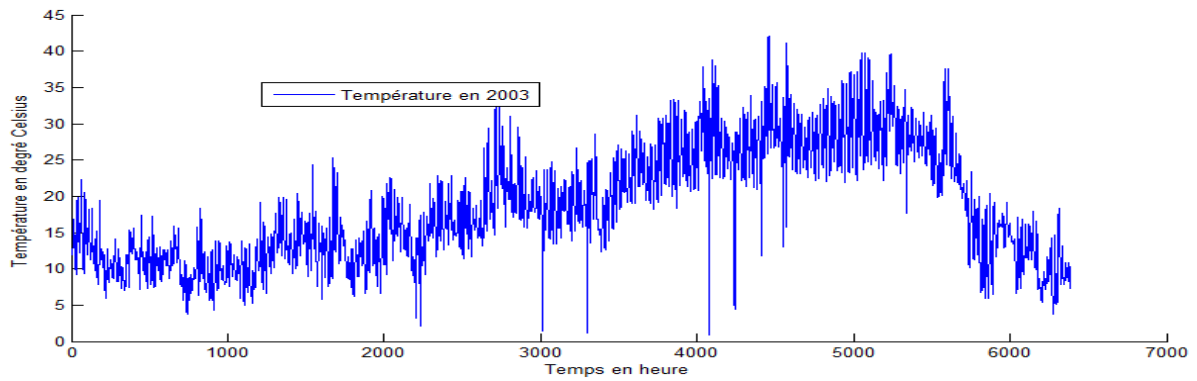


Figure 2.14 : Température en 2003.

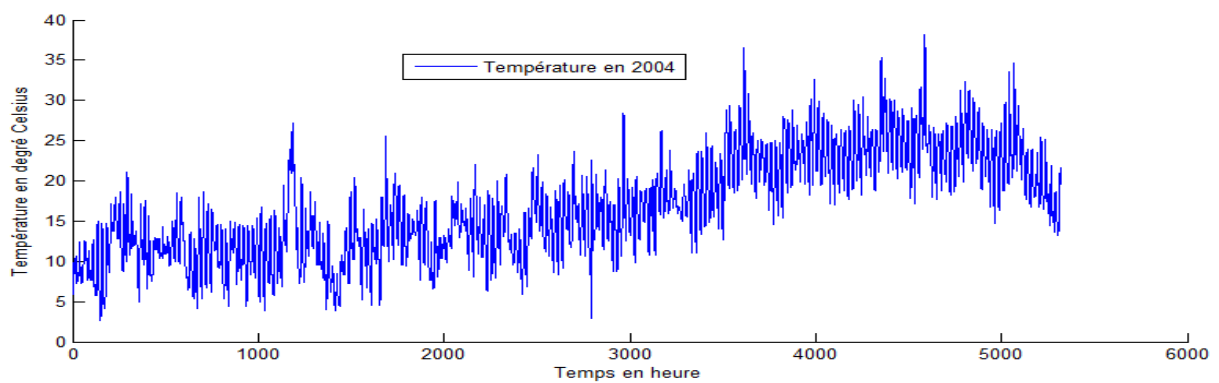


Figure 2.15 : Température en 2004.

Température	2003	2004
Moyenne	18.6942	16.8210
Covariance	16.5087	6.3003
Valeur maximum	663	38.2000
Valeur minimum	1	2.7000

Tableau 10 : Propriétés statistiques de mesure de température.

Vitesse du vent

Comme le montrent les deux Figures 2.16 et 2.17 la variation entre jour et nuit est l'une des propriétés majeures de ce paramètre.

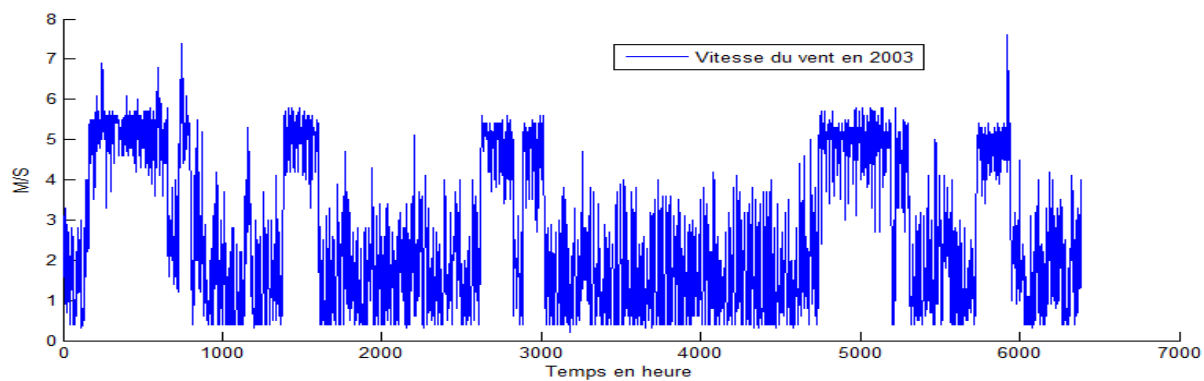


Figure 2.16 : Vitesse du vent en 2003.

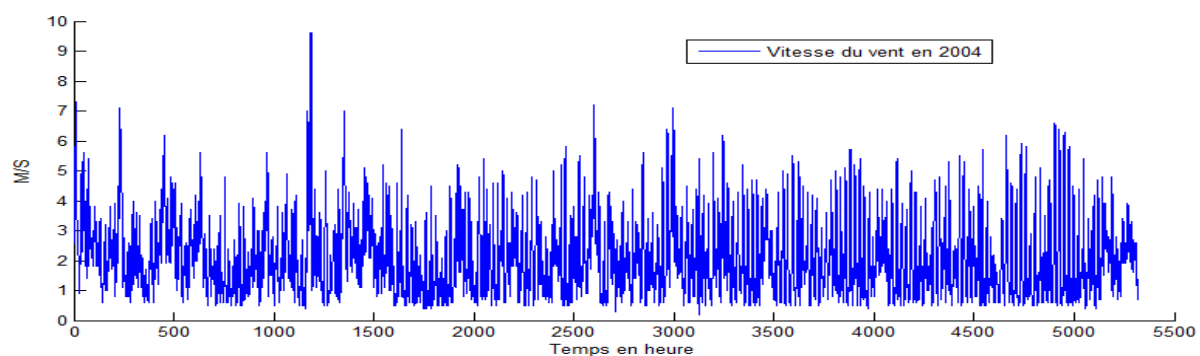


Figure 2.17 : Vitesse du vent en 2004.

Vitesse du vent	2003	2004
Moyenne	2.6599	2.1283
Covariance	1.7814	1.2749
Valeur maximum	7.6000	9.6000
Valeur minimum	0.2000	0.2000

Tableau 11 : Propriétés statistiques de mesure de la vitesse du vent.

Chapitre 3

Les Réseaux de neurones artificiels

3.1 Réseau de neurones artificiels

Depuis la présentation de neurone formel par McCulloch et Pitts (McCulloch et Pitts 1943), le modèle connexionniste a émergé comme étant une approche intéressante de résolution de problème lié à l'IA. Le neurone formel a été présenté comme un modèle de neurone biologique très promettant. Quand Minsky et Papert publient leur ouvrage intitulé « Perceptron » en 1961 (Minsky et Papert 1961), dans lequel ils ont présenté les insuffisances des modèles basé perceptron, beaucoup de recherches ont été abandonnées ou redirigés vers d'autres approches. Seuls quelques chercheurs ont continués à explorer le modèle neuronal tel que : Teuvo Kohonen, Stephen Grossberg, James Anderson, et Kunihiko Fukushima.

Dans les années quatre-vingt les réseaux de neurones artificiels s'ont ré-émergés, cela est due à deux points :

- Des travaux ont montrés des résultats très importantes (par exemple : la découverte de l'algorithme rétro-propagation de l'erreur).
- Les développements des processeurs sophistiqués qui permettent des puissances de calcul importantes ce qui permettra la mise en œuvre des modèles de calcul très coûteux en terme de puissance de calcul.

Le réseau de neurones artificiels est un modèle qui peut être caractérisé comme étant un modèle de calcul (computation model) avec des propriétés spécifiques tel que la puissance d'apprentissage, d'adaptation, de généralisation, et de clusterer les données.

3.1.1 Le neurone formel

Les réseaux de neurones artificiels, sont un ensemble d'unités de calcul qui s'intercommuniquent à travers un ensemble de connexions avec poids.

Les aspects majeurs de l'approche connexionniste selon Rumelhart et McClelland 1980 sont, comme illustré dans la figure 3.1:

- Ensemble d'unités de calcul (neurone, cellule),
- Etat d'activation pour chaque unité de calcul : y_k qui est équivalent a la sortie de l'unité,

- Connexions entre unité, avec des poids w_{jk} ce dernier détermine l'influence du signal de la j -ième unité sur la sortie de l'unité k ,
- Fonction d'activation F_k ,
- Entrée externe appelée θ ,

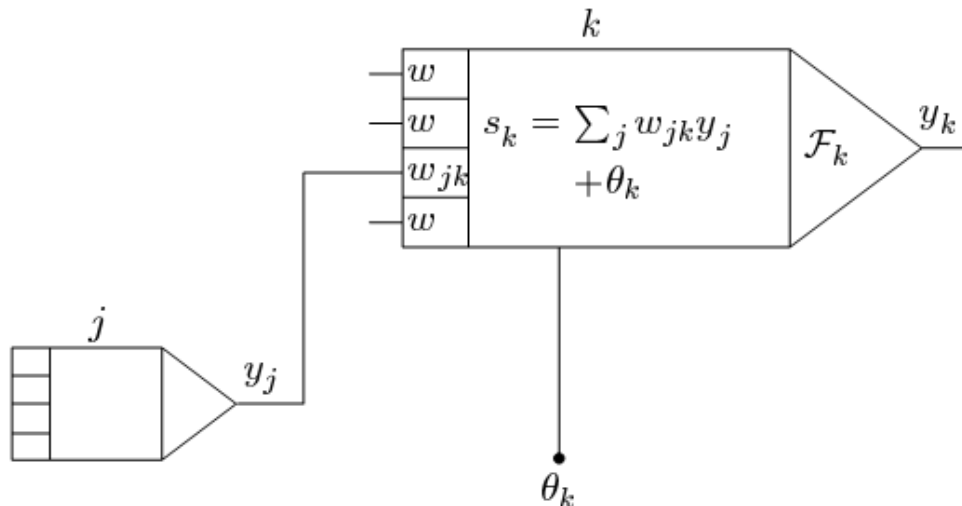


Figure 3.1 : Schéma d'un neurone formel.

3.1.2 Connexion entre unité

L'entrée à l'unité K est la somme total des multiplications de chaque entrée par le poids, plus le biais θ :

$$s_k(t) = \sum_j w_{jk}(t) y_j(t) + \theta_k(t). \quad (3.1)$$

La contribution d'une valeur positive W_{jk} est considérée comme étant une *excitation*, et pour les valeurs négatives comme étant *inhibition*.

3.1.3 Type de neurones simples

3.1.3.1 Neurone de Mc Culloch-Pitt

Se caractérise principalement par la fonction d'activation de type $\text{signe}(0,1)$, comme illustré dans la figure 3.2

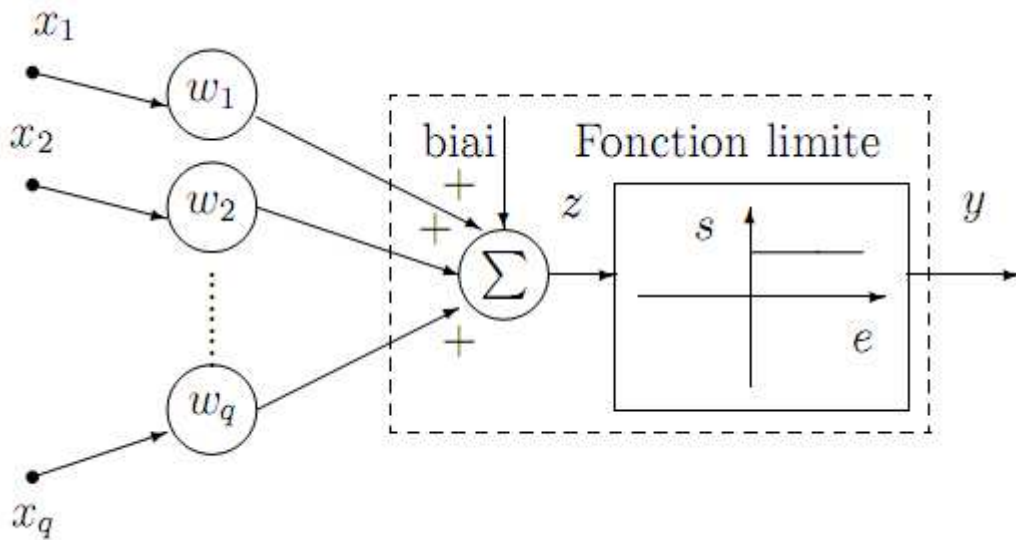


Figure 3.2 : Neurone de Mc Culloch-Pitt.

3.1.3.2 Perceptron

Comme illustré dans la figure 3.3, se caractérise principalement par :

- Une fonction d'activation continue de type : $y = \frac{2}{1 + e^{-\beta z}} - 1$.
- Fonction d'activation de type squashing, $-1 < y < 1$.
- La fonction d'activation est différentiable.

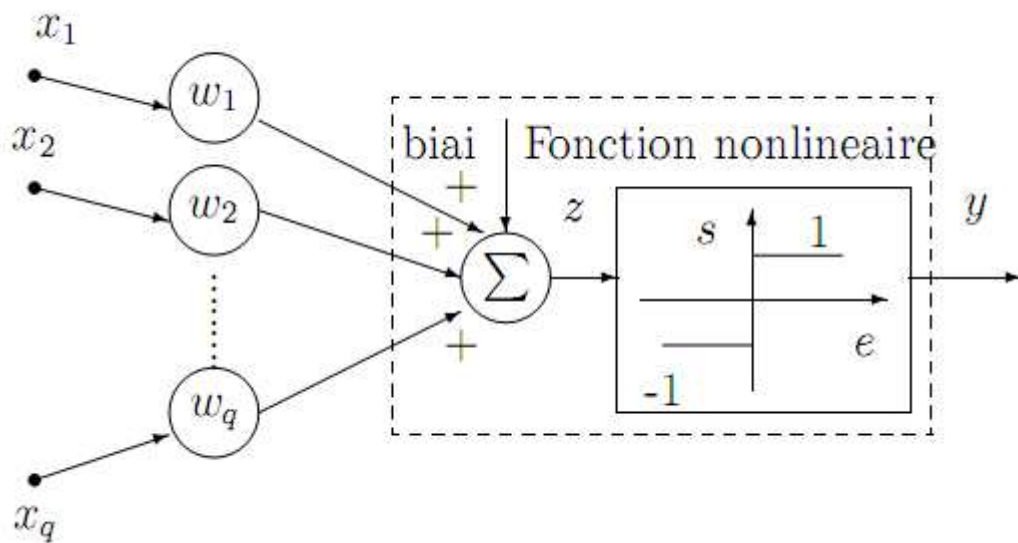


Figure 3.3 : Le perceptron.

3.2 La règle d'apprentissage de Hebb

Hebb (1945) :

Quand l'axone de la cellule A est proche d'exciter la cellule B ceci se traduit par un décharge (fire), un changement métabolique ou évolution se produit dans l'une ou l'autres cellule. En d'autres termes, l'efficacité de déchargement de A sur B est augmentée.

Algorithme :

$$y = f(w^T x) \quad (3.1),$$

$$\Delta w = \mu y x = \mu f'(w^T x) x \quad (3.2)$$

Pour le neurone i et l'entrée j , on a :

$$\Delta w_{ij} = \mu y x = \mu f'(w_i^T x_j) x_j \quad (3.3).$$

Avec

- w_0 est initialisé avec des valeurs aléatoires,
- L'apprentissage est purement a propagation avant (feedforward) et non supervisé.
- Les poids w_i , seront fortement influencés par des entrées fréquentes.
- Une évolution sans contraintes des poids w_i .
- Une règle de mise à jour basée sur la corrélation.

3.3 La règle d'apprentissage du Perceptron

La règle d'apprentissage du perceptron (Rosenblatt 1948) est caractérisée par :

- La règle est applicable au neurone de type signe.
- Le signal d'apprentissage est la différence entre l'actuelle et la réponse désirée.
- La nécessité d'un signal désiré implique un apprentissage supervisé.

Soit le perceptron, avec le signal d'apprentissage d :

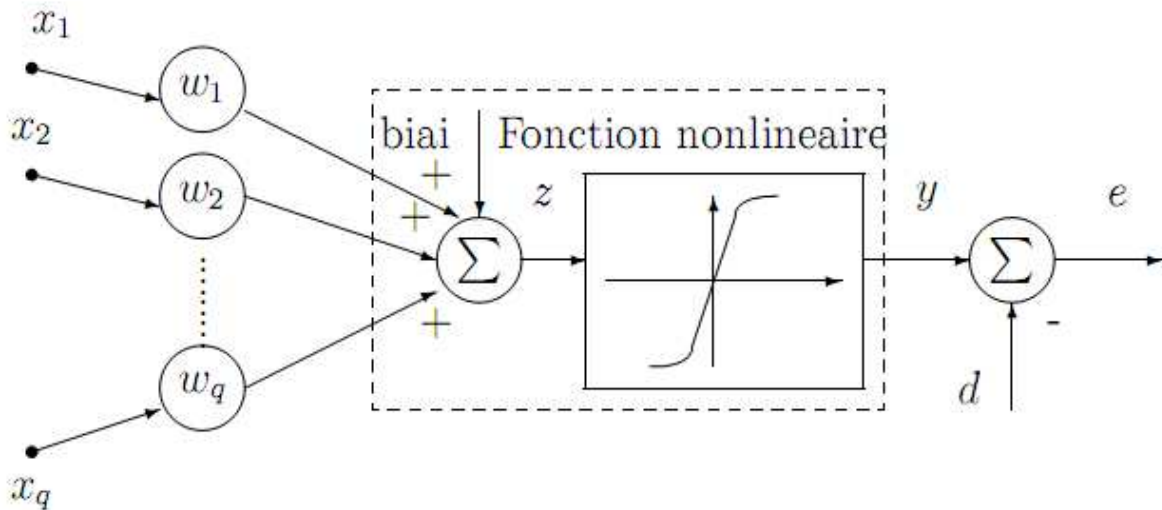


Figure 3.4 : Règle d'apprentissage du perceptron.

Algorithme :

$$e = d - y, \quad y = \text{sgn}(w^T x) \quad (3.4)$$

$$\Delta w = \mu e x = \mu [d - \text{sgn}(w^T x)] x \quad (3.5)$$

Etant donné que $d = 1$ ou $d = -1$,

$$\Delta w = \pm 2\mu x \quad (3.6)$$

3.2.1 Convergence de la règle de perceptron

Théorème : Si il existe un vecteur de poids, qui classe correctement l'ensemble d'apprentissage (supposé linéairement séparable), donc la règle d'apprentissage trouvera un, et un seul vecteur de poids, w^* , dans un nombre fini d'itérations.

Assomptions :

- Il existe au moins un vecteur des poids w^* .
- Il existe un nombre fini de vecteurs d'apprentissage (ensemble d'apprentissage limité).
- La fonction d'activation est de type unipolaire (0 ou 1).

3.3 Différents type de réseaux de neurones artificiels

3.3.1 Perceptron multicouches (PMC)

Comme démontré par Cybenko, une fonction continue quelconque peut être arbitrairement bien approximée par réseau de neurones Feed-forward à couche cachée unique. Ou chaque neurone dans la couche cachée a comme fonction d'activation une fonction non linéaire sigmodal (le perceptron). Ce type de neurones utilisés dans un réseau ce qui est nommé le Perceptron Multicouches.

Ce dernier constitue le réseau de neurones le plus utilisé pour la résolution de problème d'approximation de fonction (telle que la prédiction).

Suite à l'incapacité démontrée du Perceptron à résoudre des problèmes intéressants, ce qui a causé un ralentissement notable dans la recherche sur le connexionnisme. Mais certains travaux ont montré que l'on pouvait quand même dépasser ces limitations en intercalant, entre la couche d'entrée et la couche de sortie, une ou plusieurs couches à la suivantes (réseaux feed forward), comme illustré dans la figure 3.5 . Dans ce cas n'importe quelle fonction peut être efficacement approximée.

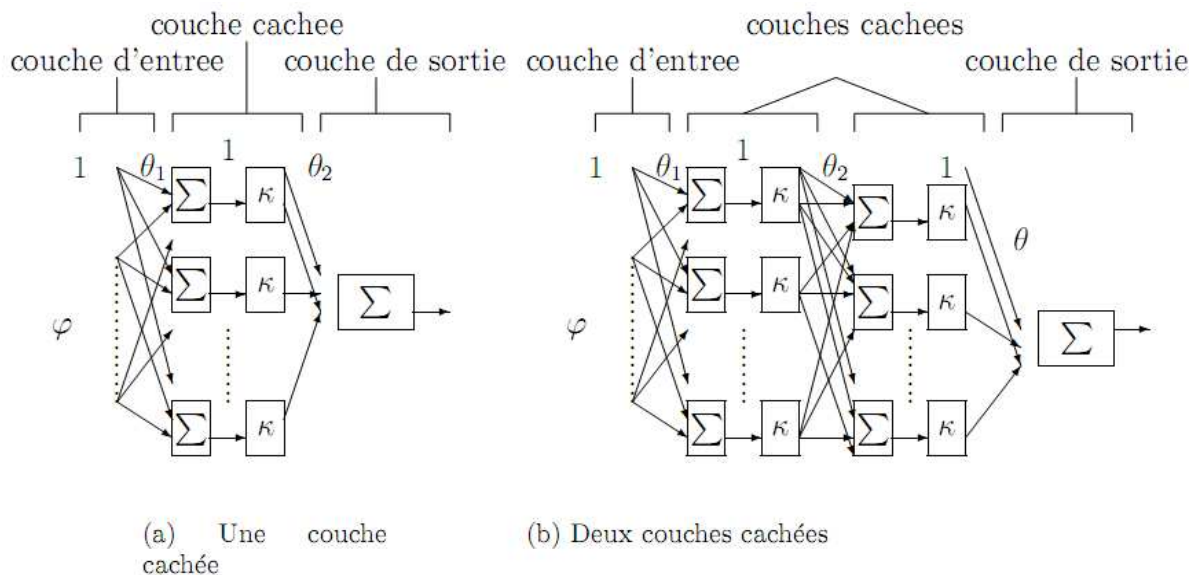


Figure 3.5: Perceptron Multicouches

Où φ est le vecteur d'entrée (appelé regresseur), et θ est la fonction d'activation, des neurones utilisés dans le réseau, fonction sigmodal pour le PMC.

Largement inspiré de la théorie des systèmes linéaires (Ljung 1987), et (Box and Jenkins 1989), un nombre de régresseurs a été choisi pour être utilisés dans le cas non linéaire, et notamment avec le PMC.

Ceci est pratique car comme établi pour le neurone simple, la non linéarité n'apparaît que dans la liaison du régresseur à l'espace de sortie. Si la sortie estimée est \hat{y} au temps k , elle est donnée par :

$$\hat{y}(k/\theta) = g(\rho(k), \theta) \quad (3.7)$$

Les valeurs passées de \hat{y} sont en fonction de $g(\rho(k))$ et donc en fonction de θ , c'est donc la raison pour laquelle \hat{y} au temps k est donné par $\hat{y}(k/\theta)$. Il est donc possible de distinguer les régresseurs suivants :

-Structure à réponse impulsionnelle finie :

$$\rho = [u(k-d)u(k-1-d)\dots u(k-m-d)]^T \quad (3.8)$$

-Structure autorégressive avec entrées exogènes (Autoregressive with exogenous inputs model structure NARX)

$$\rho = [y(k-1)/\theta y(k-2)/\theta \dots y(k-n)u(k-1-d)u(k-2-d)\dots u(k-m-d)]^T \quad (3.9)$$

-Structure d'erreur de sortie (output model structure NOE) :

$$\rho = [y(k-1/\theta)y(k-2/\theta)\dots y(k-n/\theta)u(k-1-d)u(k-1-d)\dots u(k-m-d)]^T \quad (3.10)$$

-Structure autorégressive à moyenne ajustée (auto-regressive moving average model ARMAX) :

$$\rho = [y(k-1)y(k-2)\dots y(k-n)u(k-1-d)u(k-2-d)\dots u(k-m-d)\xi(k-\tau)]^T \quad (3.11)$$

Où $\xi_u(k) = y(k) - \hat{y}_u(k/\theta)$, $\hat{y}_u(k/\theta)$ sont les sorties simulées n'utilisant que les valeurs passées de u .

3.4.2 PMC pour l'approximation des fonctions

Le PMC est très utile dans le cas de problème d'approximation des fonctions. D'après le théorème de Cybenko (Cybenko 1989) : Soit une fonction F continue et bornée et monotone. Soit I_p donnant l'hyper-cube $[0,1]^p$ de dimension p . L'espace de la fonction continue sur I_p est donnée par $C(I_p)$.

Alors, si $f \in C(I_p)$ et $\varepsilon > 0$, il existe un entier M et un nombre de constantes réelles α_i, β_i et w_{ij} ou $i=1, \dots, p$, tel que :

$$F(x_1, \dots, x_p) = \sum_{i=1}^M \alpha_i f\left(\sum_{j=1}^p w_{ij} x_j - \beta_i\right) \quad (3.12)$$

Comme une approximation de la fonction f , ou :

$$\left| F(x_1, \dots, x_p) - f(x_1, \dots, x_p) \right| < \varepsilon \quad (3.13)$$

Pour tout x_1, \dots, x_p .

La figure montre un PMC à une seule sortie conformément à l'équation 3.12.

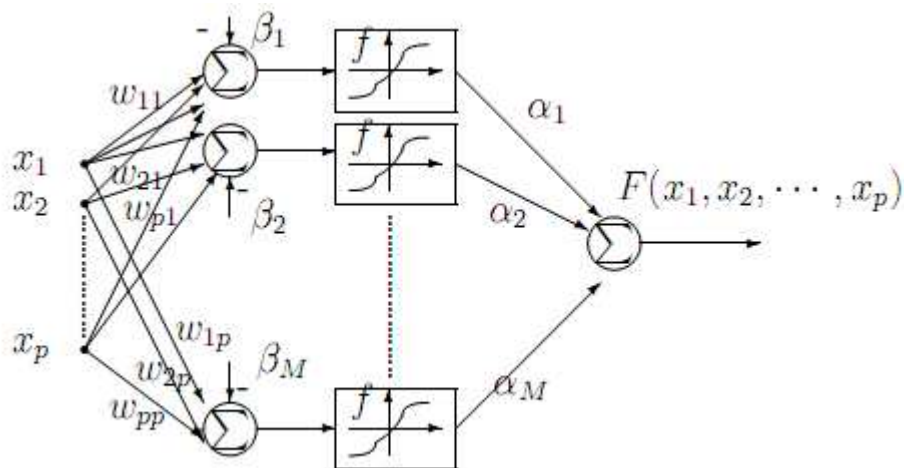


Figure 3.6 : PMC justifiant le théorème de Cybenko.

3.4.3 Apprentissage par retro-propagation

Deux phases peuvent être distinguées dans l'apprentissage par retro-propagation. Dans la première, on présente un exemple à l'entrée, et on propage les signaux de couche en couche jusqu'à la sortie, à travers les poids connectant les sorties des neurones d'une à ceux de la couche suivante. On peut alors mettre en évidence les erreurs apparaissant sur les neurones de la couche de sortie. Dans la seconde phase, on applique l'algorithme de mise à jour expliqué dans la section précédente, dans la dernière couche de poids, puis on propage les erreurs de sortie à travers cette couche de poids sur l'avant-dernière couche, ce qui va permettre de passer une itération l'algorithme pour mettre à jour l'avant-dernière couche, jusqu'à arriver à la couche d'entrée.

Algorithme :

Mathématiquement :

En se basant sur la règle d'apprentissage delta suivante :

$$w_{k+1} = w_k + 2\mu e f'(z) X \quad (3.14)$$

Pour chaque neurone, aucun problème ne se pose pour appliquer cette règle car l'erreur ($e = d - y$) est directement disponible. Par contre pour les neurones de la couche cachée (si l'on considère un réseau à une couche), on est dans l'obligation de rétro-propager l'erreur de la sortie pour obtenir une erreur effective à la sortie de chaque neurone de la couche cachée comme illustrée dans la figure 3.7 :

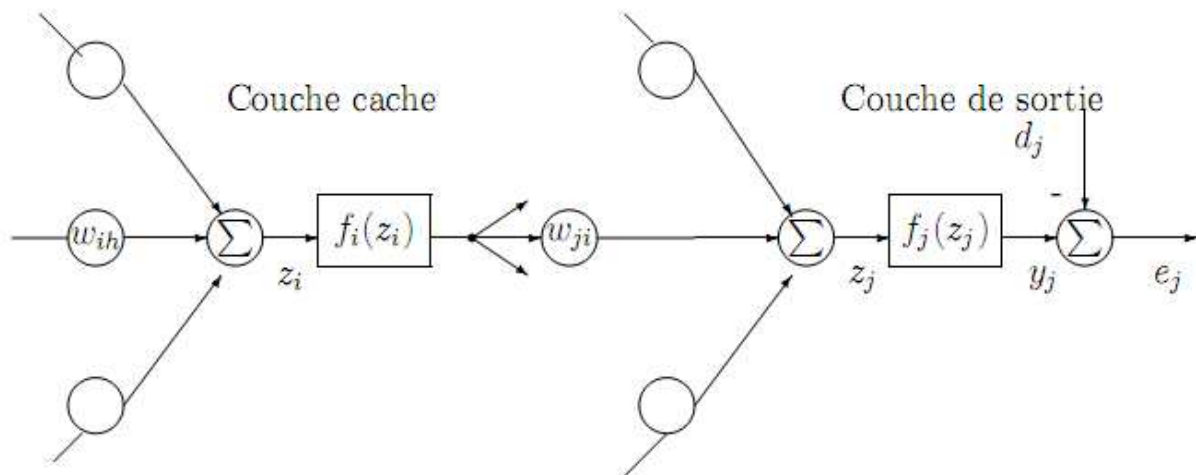


Figure 3.7 : Deux neurones de différentes couches dans un réseau de type PMC.

La fonction cout est donnée par :

$$J = \sum_{j=1}^N e_j^2 \quad (3.15)$$

$$J = \sum_{j=1}^N J_j \quad (3.16)$$

Ou chaque J_j dépend seulement du neurone j , La minimisation de chaque J_j indépendamment de l'ensemble, deviens possible :

Le gradient pour le neurone j dans la couche de sortie est donnée par :

$$\frac{\delta J}{\delta w_{ij}} = \frac{\delta J \delta e_j \delta z_j}{\delta e_j \delta z_j \delta w_{ij}} \quad (3.17)$$

On a :

$$\frac{\delta J}{\delta e_j} = 2e_j \quad (3.18)$$

$$\frac{\delta e_j}{\delta z_j} = -f'(z_j) \quad (3.19)$$

Si on définit :

$$\delta_j = -\frac{\delta J}{\delta z_j} = 2e_j f'(z_j) \quad (3.20)$$

Ce qui peut être interprété comme le gradient local à travers le neurone j , d'où :

$$\frac{\delta J}{\delta w_{ij}} = -2e_j f'(z_j) y_j = -\delta_j y_i \quad (3.21)$$

Si le neurone j se trouve dans une couche cachée, figure 3.9, alors la fonction cout devient :

$$J = \sum_{k=1}^M e_k^2 \quad (3.22)$$

L'évaluation se fait par l'utilisation du gradient $\frac{\delta J}{\delta w_{ij}}$ pour l'évaluation de la mise à jour des poids w_{ij} . A noter que maintenant w_{ij} affecte toutes les erreurs de sortie e_1, e_2, \dots, e_m .

$$\frac{\delta J}{\delta w_{ij}} = \frac{\delta J \delta z_j}{\delta z_j \delta w_{ij}} \quad (3.23)$$

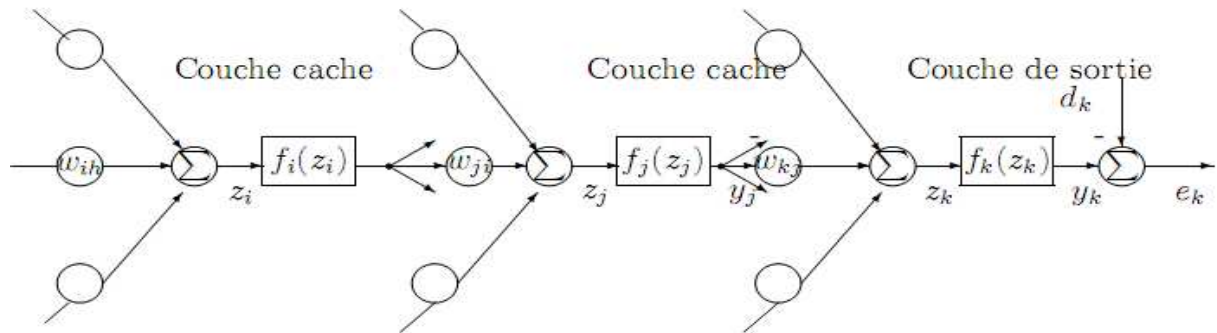


Figure 3.8 : Trois neurones de différentes couches dans un PMC à deux couches cachées.

$$\Delta w_{ji} = -\mu \frac{\delta J}{\delta w_{ij}} = \mu \delta_j y_i \quad (3.24)$$

$$\Delta w_{ji} = 2\mu f'_j(z_j) \frac{\delta J}{\delta w_{ij}} = \mu f'_j(z_j) \sum_{k=1}^M \delta_k w_{kj} y_i \quad (3.25)$$

On peut résumer par :

$$\Delta w_{ji} = -\mu \delta_j y_i \quad (3.26)$$

- Si le neurone j est dans une couche de sortie alors : $\delta_j = 2f'_j(z_j)e_j$.
- Si le neurone j est dans une couche cachée : $\delta_j = 2f'_j(z_j) \sum_{k=1}^M \delta_k w_{kj}$.

Donc l'algorithme de retro-propagation peut être résumé en :

Première Passe (propagation)	
1	Appliquer le vecteur d'entrée ainsi que son correspondant le vecteur de sortie.
2	Propager l'entrée vers tous les neurones de différentes couches, jusqu'à arriver à calculer la sortie.
3	Calculer l'erreur pour chaque neurones de sortie J , $E_j = y_j \dots O_j$, ou O c'est la sortie calculer et y c'est la sortie désirée.
deuxième passe (retro-propagation)	
4	Ajusté les poids entre les neurones de la couche cachée et la couche de sortie, selon la formule :2.25
5	Calculer l'erreur E_i pour tous les neurones des couches précédentes.
6	Propager en arrière l'erreur jusqu'aux neurones de la couche d'entrée.

Tableau 3.1 : Les étapes de l'algorithme de rétro-propagation de l'erreur.

3.3.5 Réseau à fonction de base radiale

Les réseaux à fonction de base radiale (Radial Basis Function Networks, RBF) ont été proposés et utilisés dans plusieurs études (Moody Darken 1989, Renals and Rohwer 1989), dans (Chen et al 1991) une présentation plus détaillée des RBF peut être trouvée. Une architecture générale des RBF est présentée dans la Figure 3.9.

La première couche consiste en n neurones d'entrées, Ces derniers sont totalement connectés aux neurones de la couche cachée. Les neurones de la couche cachée ont la fonction à base radiale comme étant une fonction d'activation. Un RBF peut être formulé comme dans (3.27).

$$f(x) = e^{(-x-M)^2/2\delta^2} \quad (3.27)$$

Ou :

M est la moyenne et δ est la covariance de la variable d'entrée X ,

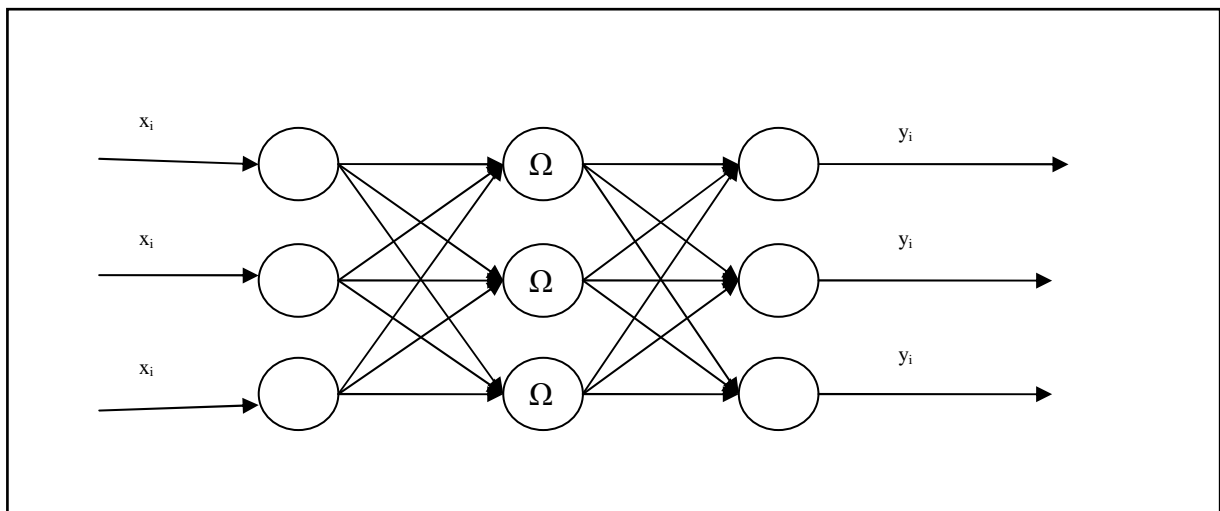


Figure 3.9 : Schéma général d'un réseau à fonction de base radiale

Avantage des RBF par rapport à MLP

- Temps d'apprentissage plus rapide par rapport à la taille de réseaux.

- Bonne généralisation.
- Pas de problème de minimum local.

Les réseaux de type RBF ont un inconvénient majeur qui est le problème de dimensionnalité, surtout dans la couche cachée.

Les réseaux de type RBF peuvent avoir une boucle de rétroaction, ce qui donne un système de type RBF récurrent appelé aussi RRBF.

3.4.6 Réseau de neurone récurrent

Un réseau de neurones est dit récurrent s'il possède au moins une boucle de rétroaction. A partir d'une sortie d'un neurone d'une couche à un neurone d'une couche précédente ou la couche d'entrée). Ce genre de réseaux se basent sur le concept que à un moment donné, la sortie ne dépend pas que de l'entrée et les poids de connections seulement, mais aussi de la sortie passée ($t-1$), un réseau de neurone récurrent typique est présenté dans la Figure 3.10. Ce genre de modèle a été largement utilisé dans la prédiction : prédiction des flux des eaux (Coulibay et *al* 1999), (Wunsch et *al* 1999), prédiction dans le domaine des finances (Giles et *al* 1997), charge électrique (Costa et *al* 1999), aussi dans la reconnaissance des formes (Saad et *al* 1999).

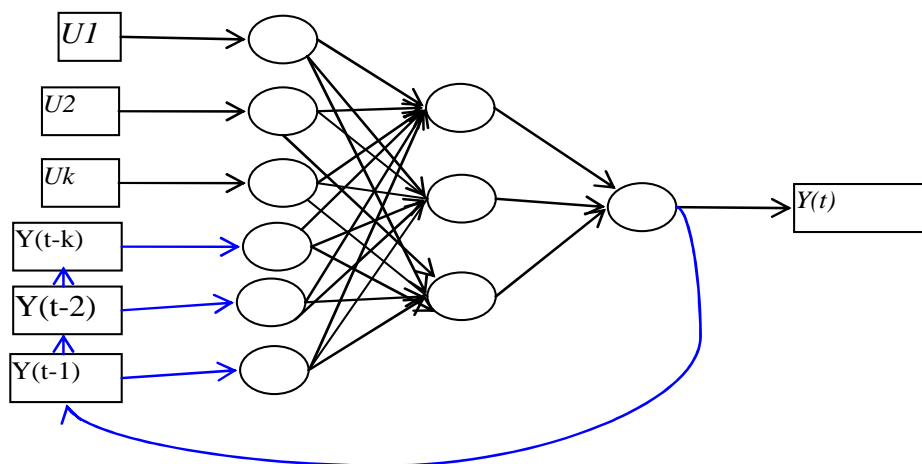


Figure 3.10 : Un réseau de neurones récurrent.

Le réseau prend en considération les prédictions déjà faites, c'est-à-dire qu'il reçoit en paramètre les K anciennes prédictions.

L'importance de ce genre de réseau vient de fait qu'il prend en compte l'aspect temporel du phénomène « mémoire », dans notre cas la valeur de prédiction d'un polluant à temps t , dépend de la valeur en $t-1$ jusqu'à $t-k$. En incluant ces anciennes valeurs, la performance du réseau s'améliore par rapport à un réseau feed-forward.

On peut distinguer des réseaux type : réseaux de neurones partiellement récurrents, réseaux de neurones complètement récurrents.

3.4.6.1 Réseau de neurones faiblement récurrent

Dans ce type de réseau, la récurrence peut être utilisée dans la couche cachée, comme le cas des réseaux d'Elman (Elman 1990), la Figure 3.11 illustre un réseau de type Elman.

Les réseaux d'Elman appelés aussi réseaux à tampons, permettent de prendre en considération la valeur passée du réseau et à sauvegarder de l'information pour l'utilisation future dans le pas suivant.

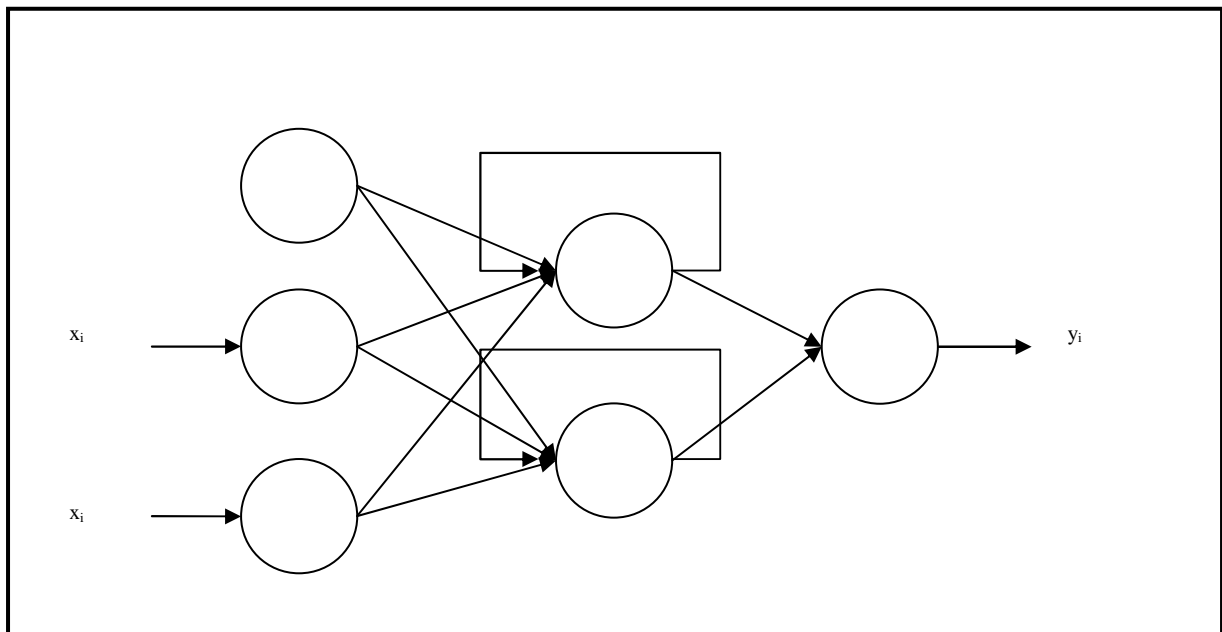


Figure 3.11: Schéma d'un réseau d'Elman typique.

3.4.6.1 Réseau de neurones fortement récurrent

Le NARX est un type de réseaux de neurone qui consiste à inclure la sortie du réseau dans le vecteur d'entrée pour la prochaine étape de prédiction. Cette classe de modèle nommé

Nonlinear AutoRegressive models with eXogenous inputs (**NARX**) peut être formulée comme dans (3.28).

$$y(t) = f(u(t - D_u), u(t - 1), u(t), \dots, y(t - D_y), \dots, y(t - 1)) , \quad (3.28).$$

Où :

u : le vecteur d'entrée.

y : la sortie de modèle.

D_u et D_y : respectivement l'ordre de vecteurs d'entrées et de sortie.

f : est une fonction non linéaire qui peut être approximé par un Perceptron Multicouches

Le système résultant est un NARX.

Chapitre 4

Développement du modèle

Ce chapitre se compose de trois parties. La première partie présente les étapes de prétraitement qui ont été appliquées aux données avant la création des modèles.

La deuxième partie décrit les modèles qu'on a testés, pour avoir le modèle le plus adéquat et qui reproduit avec le plus en fidélité les comportements de phénomène physique qui est la pollution de l'air. Nous avons testé des modèles issues de différents paradigmes (linéaire, non-linéaire, réseau de neurones artificiels de type PMC, réseau de neurones artificiels récurrent, réseau d'Elman, réseau de type RBF).

Les modèles ont été testés pour le cas de tous les polluants ;

Chaque modèle testé a présenté des résultats selon les données de tests et de validation qu'on a. Le chapitre est finalisé par une conclusion résumant une comparaison entre les modèles testés et présentant les performances que donne chaque modèle.

4.1 Prétraitement des données

Comme présenté dans la section 2.2, dans cette étude les mesures de 2003 et 2004 des différents paramètres climatiques ainsi que la concentration des différents polluants ont été utilisées. Ces données ont été récupérées de la station numéro 4 du réseau national de contrôle de pollution de l'air SAMASAFIA. Le Tableau 4.1 résume d'une façon générale la disponibilité des données pour chaque paramètres ou polluant pendant la période du 1 janvier 2003 jusqu'à 31 décembre 2003 et du 1 janvier 2004 jusqu'à 31 décembre 2004.

Polluant / paramètre climatique	2003	2004
SO ₂ : Dioxyde de soufre	Disponible, nombre de mesures : 6185	Non disponible
NO ₂ : Dioxyde d'azote	Disponible, nombre de mesures : 6133	Disponible, nombre de mesures : 2243
CO : Monoxyde de carbone	Disponible, nombre de mesures : 3951	Non disponible
O ₃ : Ozone	Non disponible	Disponible, nombre de mesures : 3506
PM ₁₀ : particulate matter	Disponible, nombre de mesures : 5318	Disponible, nombre de mesures : 5030
Humidité	Disponible, nombre de mesures : 6388	Disponible, nombre de mesures : 5318

Température	Disponible, nombre de mesures : 6388	Disponible, nombre de mesures : 5318
Vitesse du vent	Disponible, nombre de mesures : 6388	Disponible, nombre de mesures : 5318

Tableau 4.1 : Les données disponibles.

4.1.2 Nettoyage des données

Les données dans leurs format brut ont été sauvegardées sous format de fichier MS-Excel , chaque trois mois ensemble, avec l'existence de plage vide due aux ruptures ou pannes dans les capteurs, ou perdue à cause de conversion de leurs format original vers Excel .

Pour éliminer ces plages nous avons procédé à implémenter une fonction sous **Matlab**, qui consiste à nettoyer les plages vides et les erreurs de type NAN (not a number), cette fonction nous a permet d'avoir un ensemble de matrice pour chaque paramètres ou polluant qui ne contient que des valeurs logiques. Quoique y avait des mesures négatives ou nulle qui ont été aussi éliminées. Dans le cas de vecteur de température nous avons appliqué l'équation 4.1 pour éliminer les valeurs négatives :

$$T = T + (T_{\max} - T_{\min}) \quad 4.1$$

4.1.3 Normalisation des données

Pour pouvoir faire des prédictions sur la qualité de l'air et la concentration des polluants de l'air, nous avons choisi de se basé sur les pourcentages par rapport aux maximums (Pics), c'est-à-dire des valeurs qui ne sont entre 1 et 0, ce qui permet de voir les prédictions faites par rapport aux pics. Donc nous avons précédé à appliquer une normalisation des valeurs de chaque matrice obtenus dans l'étape de nettoyage.

La normalisation se base sur la formulation 4.2 suivante

$$V_p = \frac{V_p}{\max(V_p) - \min(V_p)} \quad 4.2$$

Où :

V_p : est le vecteur du paramètre p, min et max se sont respectivement des fonctions qui permettent d'avoir le minimum et le maximum du vecteur passé en paramètres. Le tableau 4.2 contient les valeurs max et min de chaque vecteur utilisé.

Polluant / paramètre climatique	Max	Min
SO2 : Dioxyde de soufre	190	0,002
NO2 : Dioxyde d'azote	316	0,07
CO : Monoxyde de carbone	12,80	0,01
O3 : Ozone	688	0,077
PM10 : particulate matter	423,3333	0,0045
Humidité	93	0,008
Température	42	0,90
Vitesse du vent	9,60	0,003

Tableau 4.2 : Normalisation des données.

4.1.4 Résultat de l'étape de prétraitement

Après application des étapes citées (nettoyage et normalisation), les données ont été sauvegardées sous format de fichier *data_Pollution.mat* (matlab), ce dernier a été utilisé dans tous nôt expérimentations. Le tableau 4.3 décrit les vecteurs résultants de l'étape de prétraitement.

Vecteur	Description
CO_2003	Taille : 1 x 3951 Ce vecteur contient les mesures horaires de CO : Monoxyde de carbone, de l'année 2003.
NO_2003	Taille :1 x 6133 Vecteur des mesures horaires de NO2 : Dioxyde d'azote de l'année 2003
NO_2004	Taille :1 x 2243 Vecteur des mesures horaires de NO2 : Dioxyde d'azote de l'année 2004
O3_2004	Taille :1 x 5306 Vecteur des mesures horaires O3 : Ozone de l'année 2004
PM2003	Taille :1 x 5318 Vecteur des mesures horaires PM10 : particulate matter, de l'année 2003
PM2004	Taille :1 x 5318 Vecteur des mesures horaires PM10 : particulate matter, de l'année 2004

HU_2003	Taille :1 x 6388 Vecteur des mesures horaires de l'humidité, de l'année 2003
HU_2004	Taille :1 x 5318 Vecteur des mesures horaires de l'humidité, de l'année 2004
TE_2003	Taille :1 x 6388 Vecteur des mesures horaires de la température de l'année 2003
TE_2004	Taille :1 x 5318 Vecteur des mesures horaires de la température de l'année 2004
VV_2003	Taille :1 x 6388 Vecteur des mesures horaires du vitesse du vent de l'année 2003
VV_2004	Taille :1 x 5318 Vecteur des mesures horaires de la vitesse du vent de l'année 2004

Tableau 4.3 : Résultat de l'étape de prétraitement.

4.1.5 Préparation des données de test et d'apprentissage

Le choix des vecteurs à inclure dans les deux ensembles de l'apprentissage et de test affecte beaucoup sur la puissance de généralisation du modèle. C'est pour ça qu'on a procédé a construire deux ensembles un pour les testes et un pour l'apprentissage en tenons compte des points suivants :

- Construire un ensemble d'apprentissage significatif, c'est-à-dire qui inclut les données de toutes tranches temporelles, cela aide à prendre en considération la variation entres

saisons et aussi des week-ends, ainsi que les jours et la nuit, car les mesures sont fait d'une façon horaire.

- Pour les polluant qui ont deux années de mesure, nous avons pris des parties des deux années, cela nous aidera à prendre en considération la variation entre les deux années (2003 et 2004).

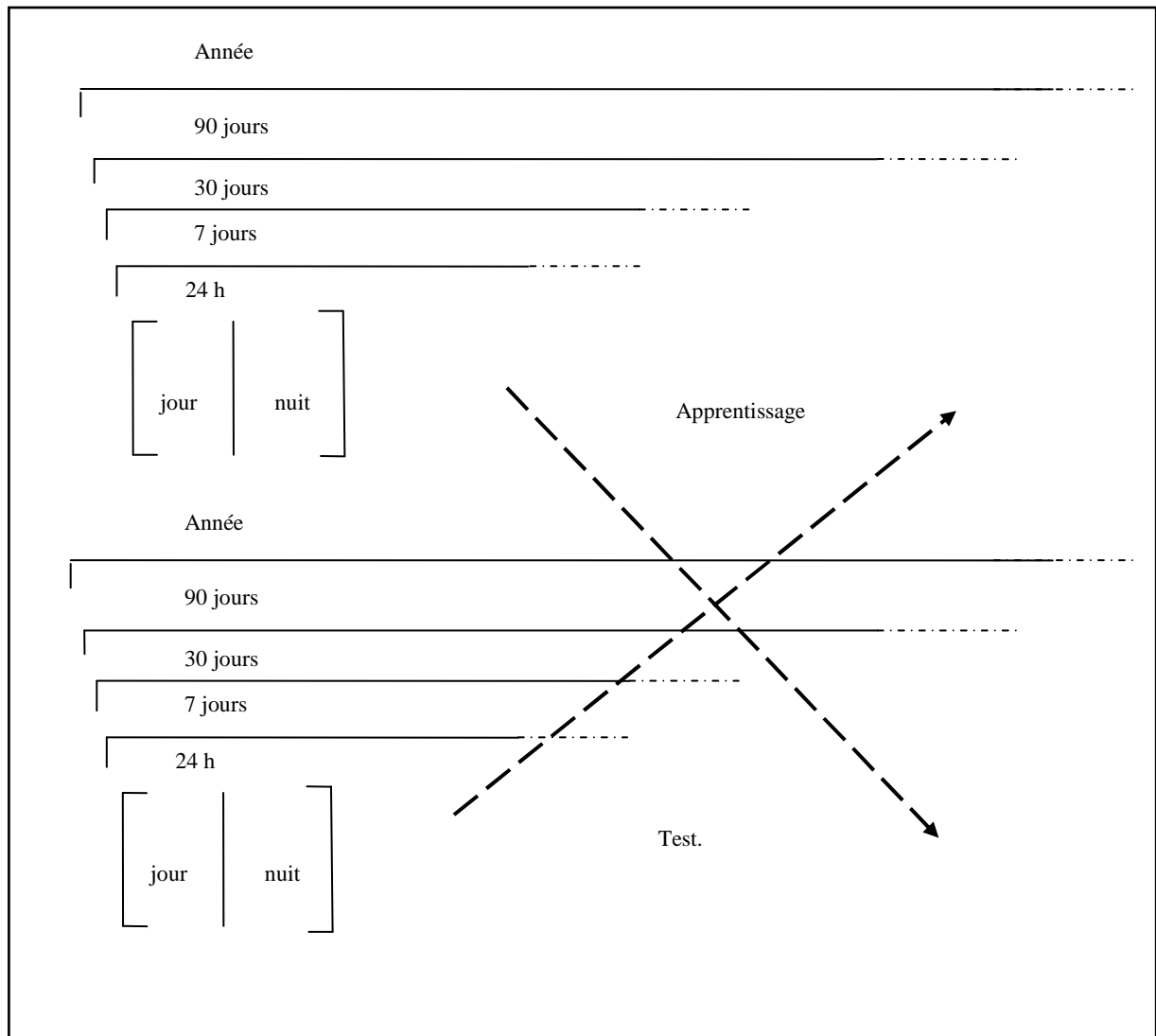


Figure 4.1 : Construction des données d'apprentissage et de test

L'algorithme utilisé dans la construction des ensembles de teste et d'apprentissage peut être résumé comme suit :

On définit quatre variables : JN_x , W_x , M_x , S_x , ces quatre sont initialisés à la valeur 'Changé'.

$JN_x = \text{'Changé'}$;

$W_x = \text{'Changé'}$;

$M_x = \text{'Changé'}$;

$S_x = \text{'Changé'}$;

$Indice = 0$;

Vecteur d'Apprentissage = vide ;

Vecteur de Teste = vide ;

Pour $i=0$ jusqu'à N faire :

Si $JN_x = \text{Changé}$ et $indice = 0$

Permuter et ajouter ces 12 valeurs dans le Vecteur d'Apprentissage ;

$JN_x = \text{Non Changé}$;

Sinon

Permuter et ajouter ces 12 valeurs dans le Vecteur de Teste ;

$JN_x = \text{Changé}$

Si $W_x = \text{changé}$ et $indice = 1$

Permuter et ajouter ces 24 valeurs dans le Vecteur d'Apprentissage ;

$W_x = \text{non Changé}$

Sinon

Permuter et ajouter ces 24 valeurs dans le Vecteur de Teste

Wx=Changé

Si Mx=Changé et indice=2

Permuter et ajouter ces 15 jours de valeur dans le vecteur d'apprentissage ;

Mx=non changé

Sinon

Permuter et ajouter ces 15 jours de valeurs dans le vecteur de teste ;

Mx= changé ;

Si Sx =Changé et indice=3

Permuter et ajouter les 45 jours de valeurs dans le vecteur d'apprentissage

Sx= non Changé ;

Sinon Permuter et ajouter les 45 jours dans le vecteur de teste

Sx= Changé ;

Initialiser indice ;

La variable indice permet de changer à chaque itération une partie (jour/nuit, Weekend, Saison, Année).

Cet algorithme a été implémenté dans une fonction MatLab, et appliqué pour tous les vecteurs issus de l'étape de prétraitement.

En se basant sur ces vecteurs on a fait l'apprentissage et le test des modèles présentés dans la section suivante.

D'autres algorithmes ont été utilisés pour la préparation des vecteurs utilisés dans l'apprentissage et la validation, tel que l'algorithme de rotation, qui consiste à adapté les vecteurs selon la longueur de pas de prédiction. Comme le but est d'avoir une prédiction moyen terme $k=12$ heures ; l'algorithme de rotation a été utilisé pour faire des rotations à 12 heures en arrière.

4.1 Modèles linéaires

Le modèle basé ARX reçoit en entrée les paramètres présentés dans le tableau suivant :

Paramètres	Nombre de valeurs passées à chaque époque	Description
Polluant	1 valeur	La concentration du polluant
Température	1 valeur	La température
Humidité	1 valeur	L'humidité
Vitesse du vent	1 valeur	La vitesse du vent

Tableau 4.4 : Les paramètres du modèle basé ARX.

Pour chaque polluant un modèle est créé, le tableau 4.5, montre les performances et les détails de chaque modèle.

Polluant	IA		MSE		Modèle
PM10	0.80657	0.52973	0.69647	1.7972	$A(q)y(t) = B(q)u(t) + e(t)$ $A(q) = 1 - 0.6784 q^{-1} - 0.07605 q^{-2} - 0.0007767 q^{-3} - 0.05722 q^{-4}$ $B1(q) = 0.8159 q^{-1} - 0.9238 q^{-2} + 0.296 q^{-3} - 0.09594 q^{-4}$ $B2(q) = 0.06965 q^{-1} - 0.1209 q^{-2} + 0.01417 q^{-3} + 0.05091 q^{-4}$ $B3(q) = 0.2028 q^{-1} - 0.3833 q^{-2}$

					$+ 0.02085 q^{-3} + 0.1639 q^{-4}$ $B4(q) = -0.02146 q^{-1} + 0.009059 q^{-2} - 0.001827 q^{-3} + 0.003353 q^{-4}$
SO ₂	0.6935	0.4793	0.34113	0.69259	$A(q)y(t) = B(q)u(t) + e(t)$ $A(q) = 1 - 0.3423 q^{-1} - 0.09082 q^{-2} - 0.1175 q^{-3} - 0.09583 q^{-4}$ $B1(q) = 0.4027 q^{-1} - 0.3893 q^{-2} - 0.0287 q^{-3} + 0.1523 q^{-4}$ $B2(q) = 0.1377 q^{-1} + 0.07207 q^{-2} - 0.1545 q^{-3} + 0.1093 q^{-4}$ $B3(q) = -0.1293 q^{-1} + 0.5286 q^{-2} - 0.3717 q^{-3} + 0.0001021 q^{-4}$ $B4(q) = -0.04964 q^{-1} + 0.1069 q^{-2} - 0.007732 q^{-3} - 0.06325 q^{-4}$
O ₃	0.7521	0.53752	1.6754	2.7073	$A(q)y(t) = B(q)u(t) + e(t)$ $A(q) = 1 - 1.184 q^{-1} + 0.3296 q^{-2} + 0.01366 q^{-3} + 0.02091 q^{-4}$ $B1(q) = -0.00842 q^{-1} + 0.09556 q^{-2} - 0.2812 q^{-3} + 0.1955 q^{-4}$ $B2(q) = 0.09201 q^{-1} - 0.1566 q^{-2} + 0.2279 q^{-3} - 0.1236 q^{-4}$ $B3(q) = 0.6078 q^{-1} - 0.215 q^{-2} - 0.1721 q^{-3} - 0.09981 q^{-4}$

					$B4(q) = -0.0694 q^{-1} + 0.1269 q^{-2} + 0.08899 q^{-3} - 0.1864 q^{-4}$
CO ₂	0.75363	0.48015	0.088266	4.063204	$A(q)y(t) = B(q)u(t) + e(t)$ $A(q) = 1 - 0.8928 q^{-1} + 0.2236 q^{-2} + 0.03242 q^{-3} - 0.002602 q^{-4}$ $B1(q) = 0.2208 q^{-1} - 0.1375 q^{-2} - 0.1828 q^{-3} + 0.1335 q^{-4}$ $B2(q) = 0.1154 q^{-1} - 0.2096 q^{-2} + 0.1987 q^{-3} - 0.1033 q^{-4}$ $B3(q) = 0.6284 q^{-1} - 1.134 q^{-2} + 0.7087 q^{-3} - 0.2076 q^{-4}$ $B4(q) = 0.02377 q^{-1} + 0.00737 q^{-2} - 0.01511 q^{-3} + 0.009867 q^{-4}$
NO ₂	0.719	0.56996	1.1952	18.891	$A(q)y(t) = B(q)u(t) + e(t)$ $A(q) = 1 - 0.9058 q^{-1} + 0.2661 q^{-2} - 0.08974 q^{-3} + 0.1899 q^{-4}$ $B1(q) = 0.1704 q^{-1} - 0.02832 q^{-2} - 0.2386 q^{-3} + 0.3712 q^{-4}$ $B2(q) = -0.1054 q^{-1} + 0.1625 q^{-2} + 0.2084 q^{-3} - 0.283 q^{-4}$ $B3(q) = -0.09861 q^{-1} - 0.1501 q^{-2} + 0.6552 q^{-3} - 0.379 q^{-4}$ $B4(q) = -0.021 q^{-1} + 0.03523 q^{-2}$

					$2 + 0.0113 q^{-3} - 0.05797 q^{-4}$
--	--	--	--	--	--------------------------------------

Tableau 4.5 : Les modèles ARX testés

4.2 Modèles basés réseaux de neurones artificiels

4.2.1 Modèles basé PMC

Paramètres	Nombre de valeurs passées à chaque époque	Description
Polluant	1 valeur	La concentration du polluant
Température	1 valeur	La température
Humidité	1 valeur	L'humidité
Vitesse du vent	1 valeur	La vitesse du vent

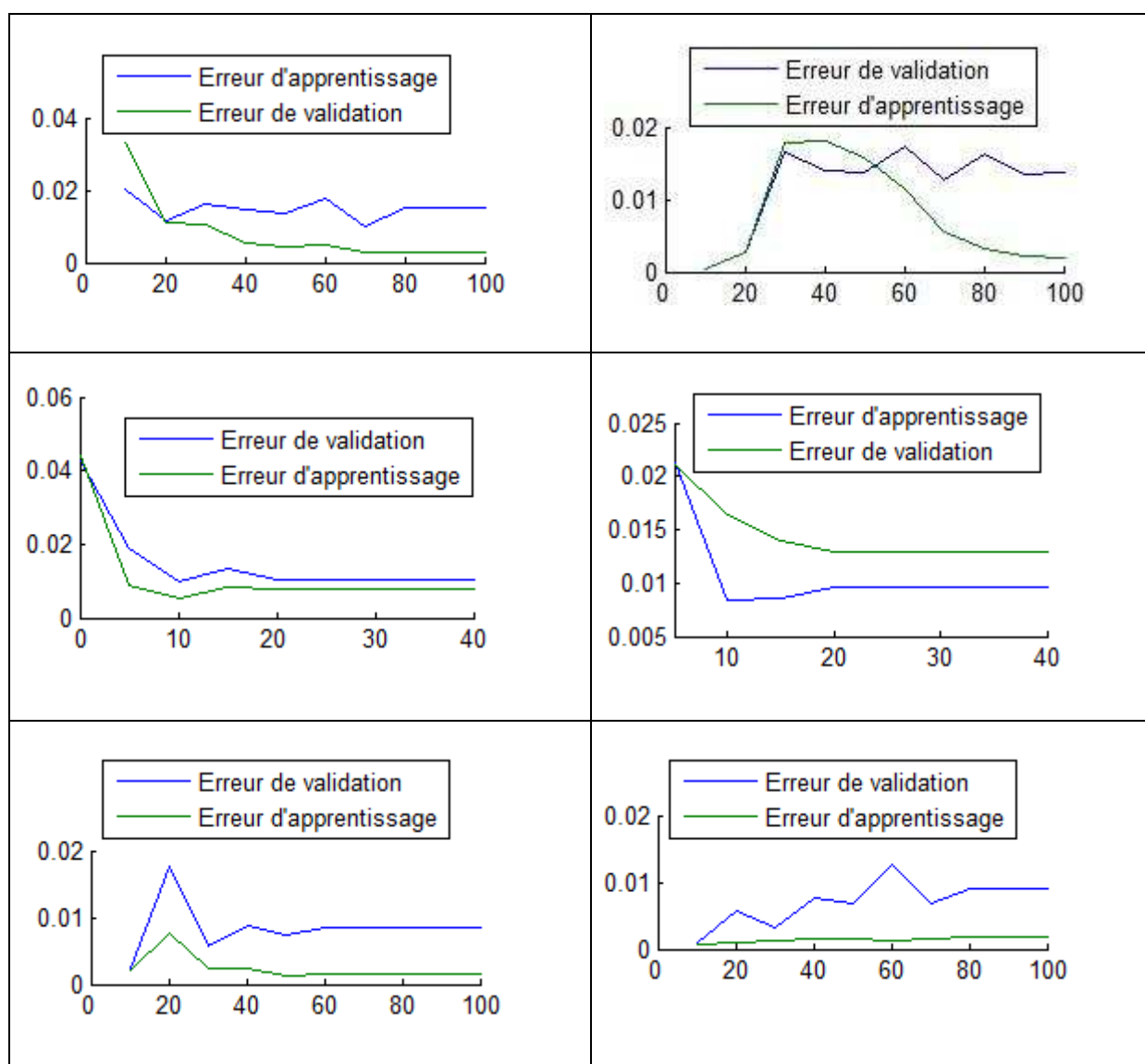
Tableau 4.6 : Les paramètres du modèle basé PMC

SO₂ :

<i>Architecture</i>	<i>IA</i>		<i>MSE</i>	
	<i>Apprentissage</i>	<i>Validation</i>	<i>Apprentissage</i>	<i>Validation</i>
[10-8-1]	0.95167	0.82367	0.011247	3.3742
[10-9-1]	0.98485	0.44482	0.57801	4.1276
[10-10-1]	0.99054	0.85353	0.080178	1.6124

[10-11-1]	0.99486	0.66077	0.81405	2.6498
[10-12-1]	0.98681	0.4569	0.42897	2.5619
[10-13-1]	0.99164	0.52494	0.56511	2.718
[10-14-1]	0.99241	0.57103	0.63092	2.6115

Tableau 4.7 : Les performances du modèle basé PMC pour le SO₂



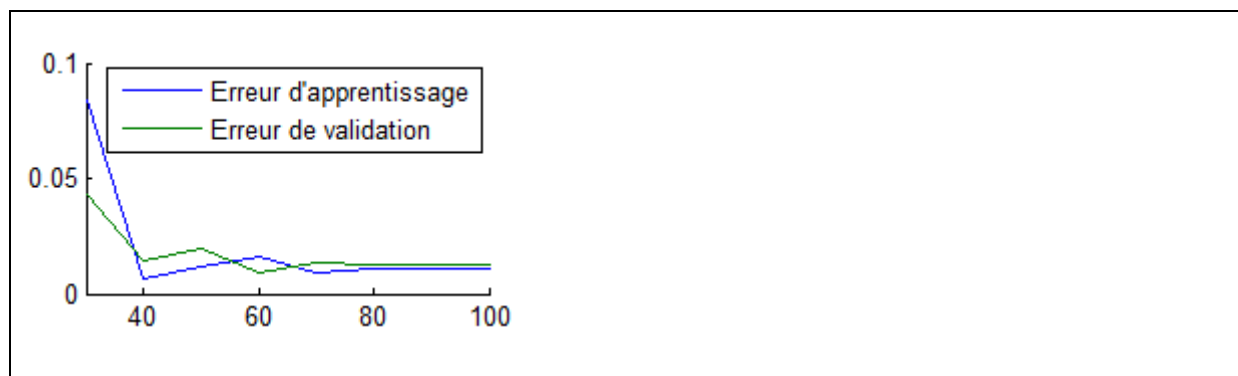


Tableau 4.8 : Les courbes d'erreur de modèle PMC pour le SO2.

PM10 :

<i>Architecture</i>	<i>IA</i>		<i>MSE</i>	
	<i>Apprentissage</i>	<i>Validation</i>	<i>Apprentissage</i>	<i>Validation</i>
[10-8-1]	0.99057	0.91845	0.11364	1.503
[10-9-1]	0.99817	0.98991	0.10993	1.7283
[10-10-1]	0.99747	0.98955	0.1169	1.6949
[10-11-1]	0.98862	0.93102	0.11039	1.4207
[10-12-1]	0.99718	0.90709	0.11807	1.3848
[10-13-1]	0.98854	0.97362	0.14687	2.88
[10-14-1]	0.99329	0.92135	0.15597	3.1412

Tableau 4.9 : Les performances du modèle basé PMC.

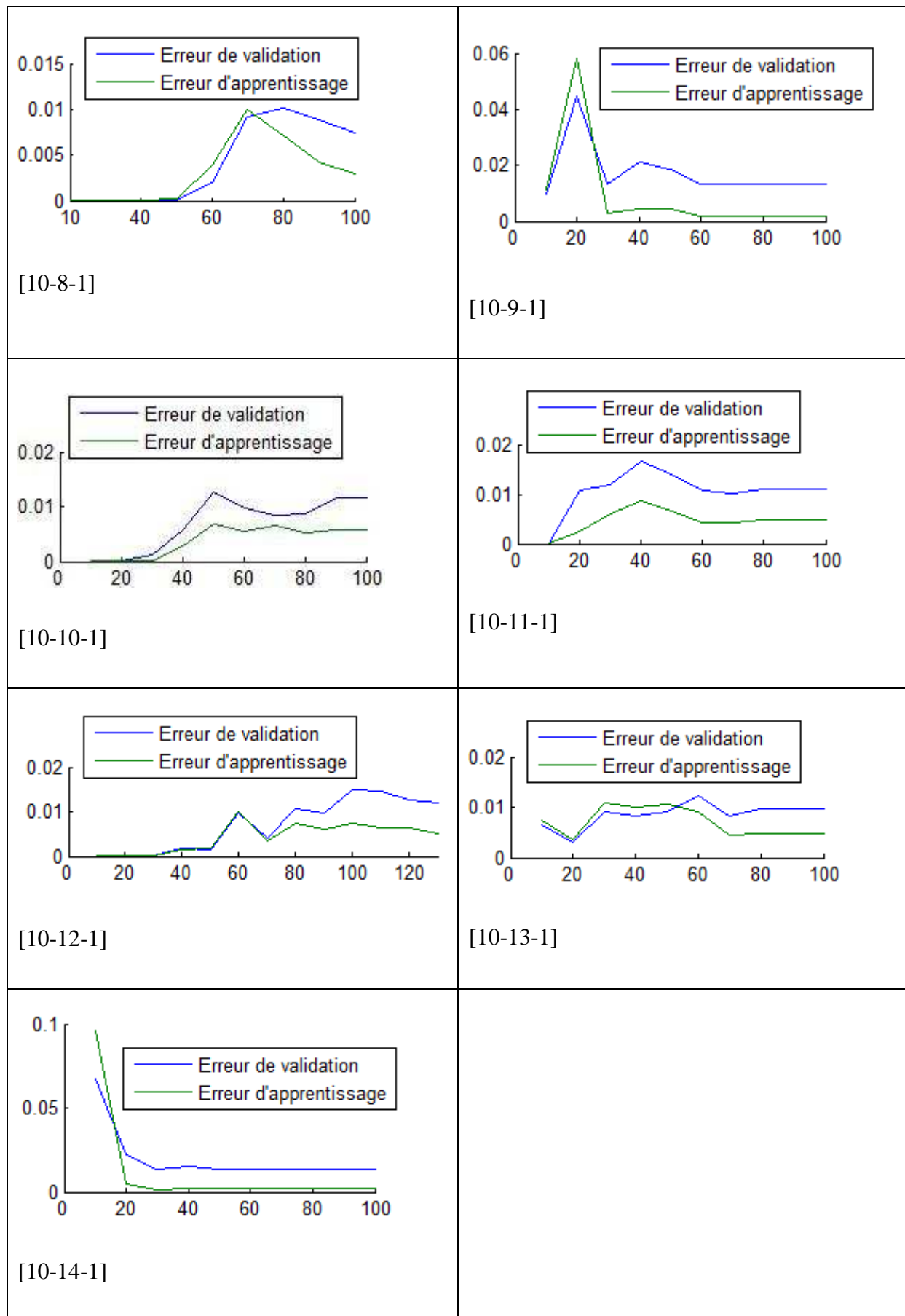


Tableau 4.10 : Les courbes d'erreur des modèles construits.

CO₂ :

<i>Architecture</i>	<i>IA</i>		<i>MSE</i>	
	<i>Apprentissage</i>	<i>Validation</i>	<i>Apprentissage</i>	<i>Validation</i>
[10-8-1]	0.99746	0.89764	0.019115	2.6929
[10-9-1]	0.99734	0.89855	0.040206	1.1773
[10-10-1]	0.99869	0.89851	0.0204	1.2118
[10-11-1]	0.9981	0.89724	0.020275	1.2038
[10-12-1]	0.99803	0.89566	0.023112	1.5701
[10-13-1]	0.99835	0.89801	0.023638	1.0495
[10-14-1]	0.9978	0.89667	0.022604	1.6763

Tableau 4.11 : Performances des modèles basés PMC pour le CO₂

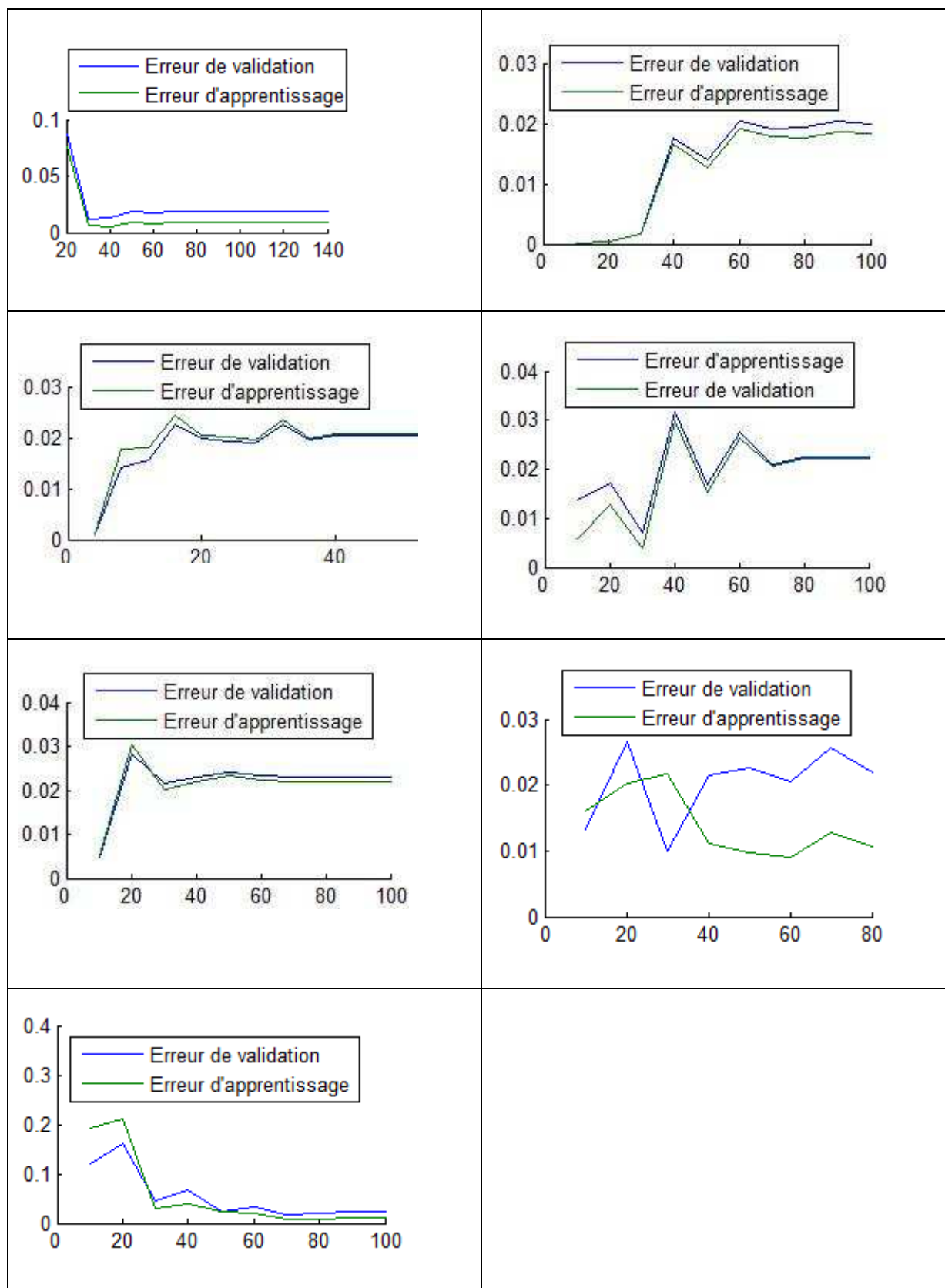
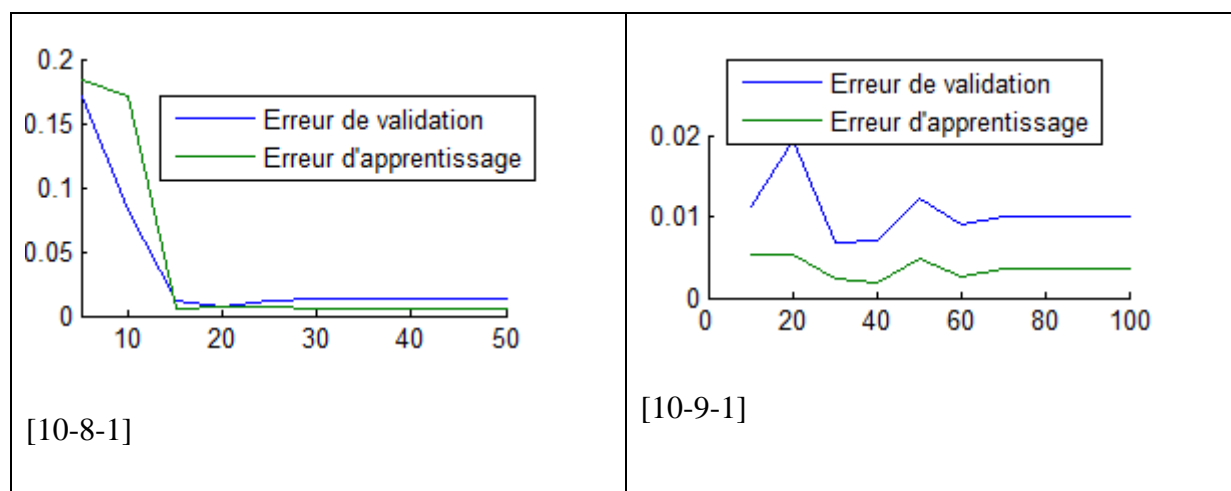


Tableau 4.12 : Les courbes d'erreur de modèle.

NO₂ :

<i>Architecture</i>	<i>IA</i>		<i>MSE</i>	
	<i>Apprentissage</i>	<i>Validation</i>	<i>Apprentissage</i>	<i>Validation</i>
[10-8-1]	0.99475	0.93403	0.56642	3.3359
[10-9-1]	0.99485	0.97661	0.81197	1.4694
[10-10-1]	0.99228	0.97287	0.84224	1.013
[10-11-1]	0.99378	0.96099	0.89675	2.9787
[10-12-1]	0.99243	0.95968	0.91582	2.9213
[10-13-1]	0.99863	0.97285	0.96878	3.1005
[10-14-1]	0.99576	0.93233	0.65192	4.1167

Tableau 4.13 : Performance du modèle pour le NO₂



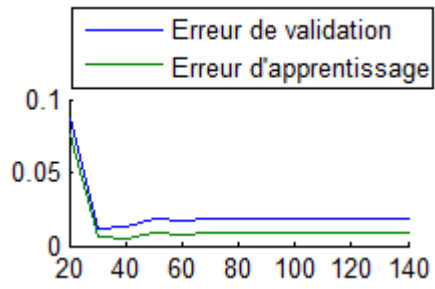
<p>[10-10-1]</p>	<p>[10-11-1]</p>
<p>[10-12-1]</p>	<p>[10-13-1]</p>
<p>[10-14-1]</p>	

Tableau 4.14 : Les courbes d'erreur de modèle.

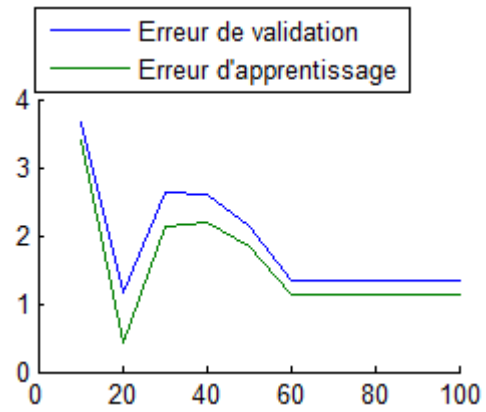
O₃ :

<i>Architecture</i>	<i>IA</i>		<i>MSE</i>	
	<i>Apprentissage</i>	<i>Validation</i>	<i>Apprentissage</i>	<i>Validation</i>
[10-8-1]	0.90656	0.64555	0.30058	1.2355
[10-9-1]	0.98934	0.78262	0.34027	1.0169
[10-10-1]	0.90177	0.88828	0.349	1.4003
[10-11-1]	0.96802	0.57875	0.60714	4.1772
[10-12-1]	0.91045	0.7335	0.65813	5.5467
[10-13-1]	0.98961	0.76696	0.69532	6.8611
[10-14-1]	0.93615	0.75475	0.83022	6.0648

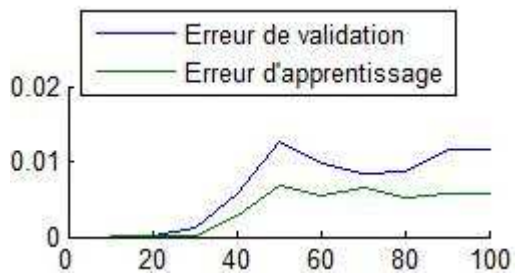
Tableau 4.15 : Performance du modèle pour l'O₃.



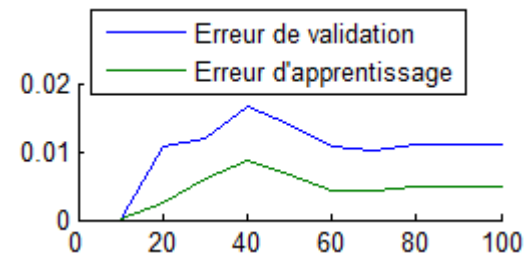
[10-8-1]



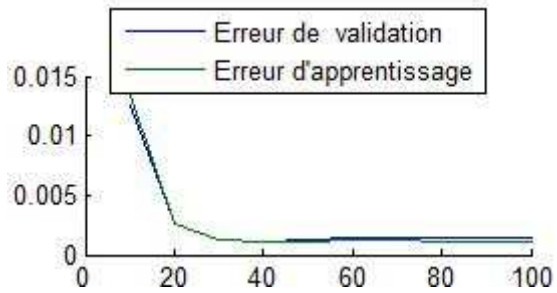
[10-9-1]



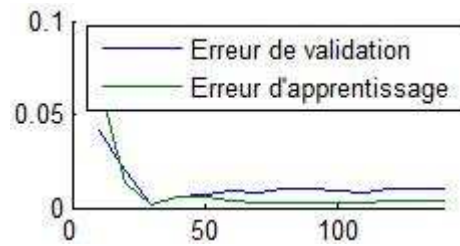
[10-10-1]



[10-11-1]



[10-12-1]



[10-13-1]

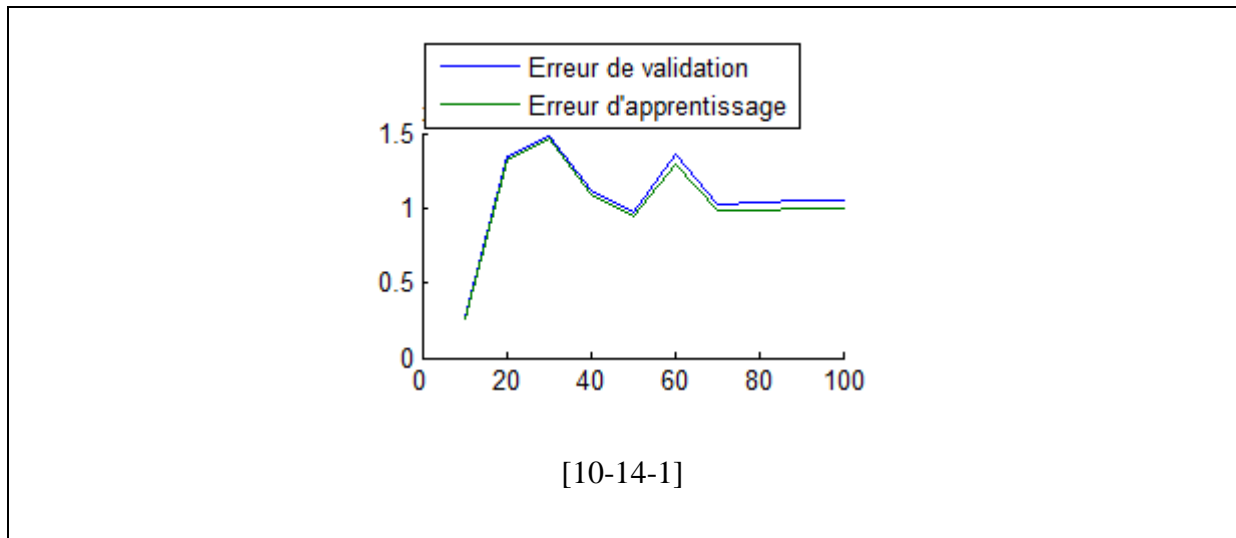


Tableau 4.16 : Les courbes d'erreur de modèle PMC.

4.2.2 Modèle basé PMC récurrent

Le modèle PMC récurrent comme présenté dans le chapitre 3. Présenté aussi dans (Ghazi et Khadir 2008) et (Ghazi et Khadir 2009a) et dans (Ghazi et Sabri 2009b). Il peut donner des prédictions à plusieurs pas, c'est-à-dire qu'au lieu d'avoir une seule sortie, le modèle est capable de donner un vecteur de k valeurs. Nous avons procédé à varier le nombre des pas k de 2 jusqu'à 12 (heurs), ce qui permet d'avoir un modèle de prédiction multi-pas.

Le modèle a été appliqué pour chaque polluant, en prenant en considération les paramètres climatiques, comme présenté dans le tableau 4.17.

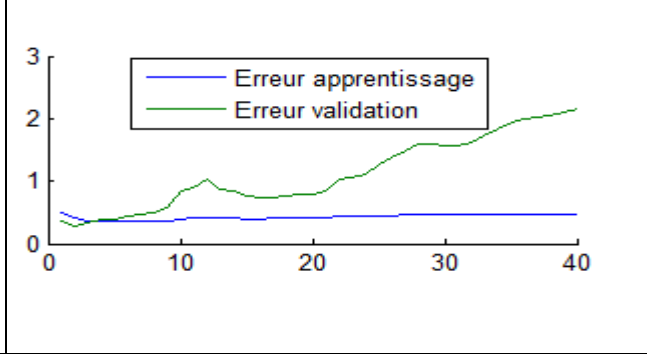
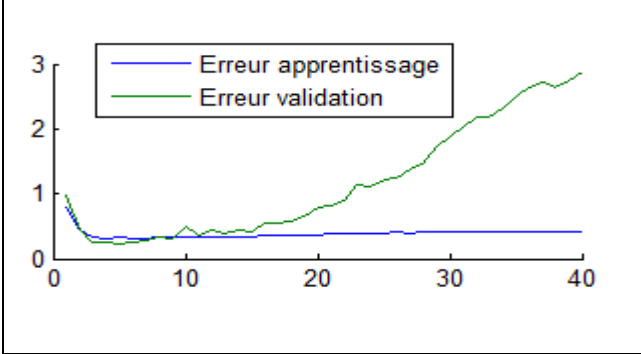
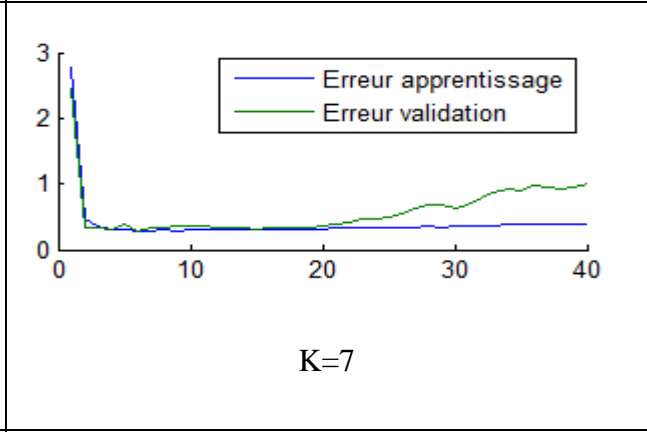
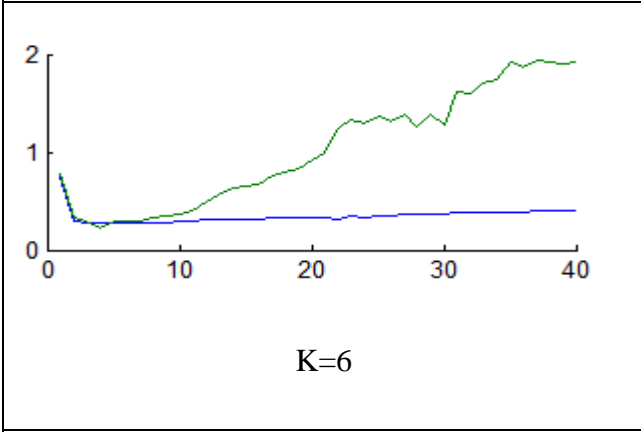
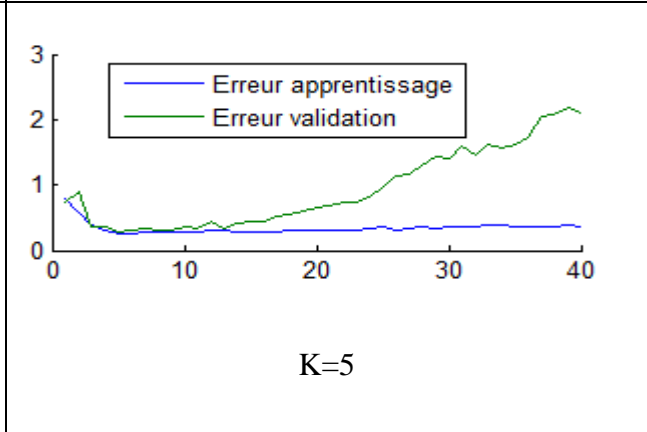
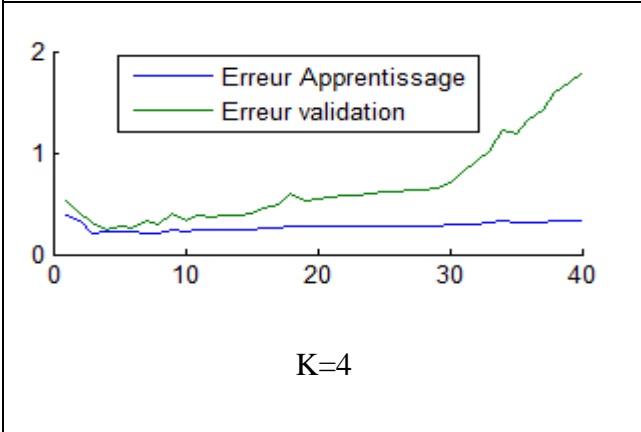
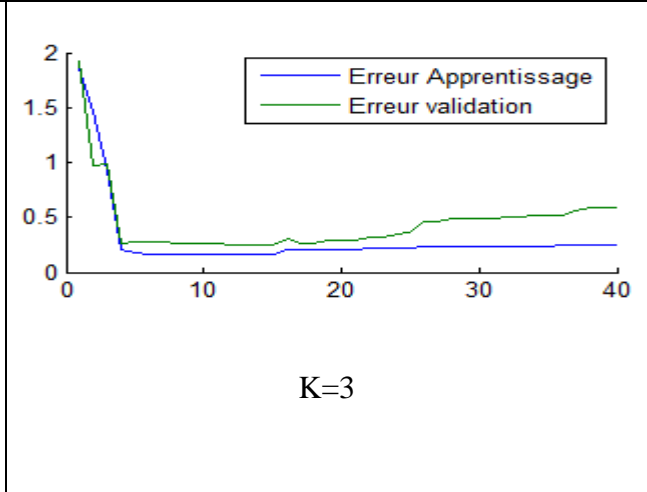
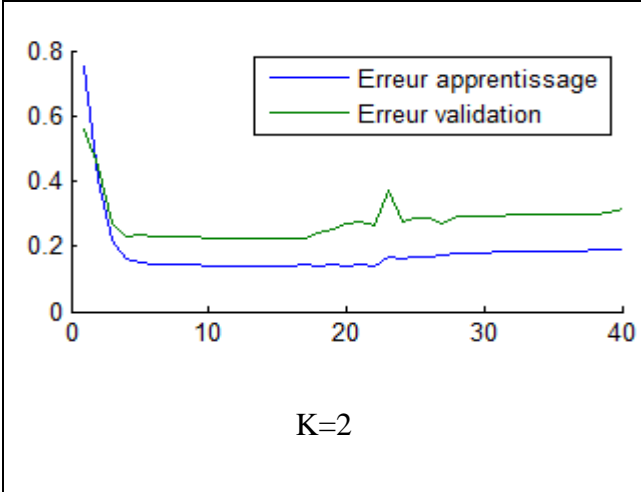
Paramètres	Nombre de valeurs passées à chaque époque	Description
Polluant	K valeurs	La concentration du polluant
Température	K valeur	La température
Humidité	K valeurs	L'humidité
Vitesse du vent	K valeurs	La vitesse du vent

Tableau 4.17 : Paramètres du modèle PMC récurrent.

SO₂ :

<i>K le nombre des pas</i>	<i>Architecture</i>	<i>IA</i>		<i>MSE</i>	
		<i>Apprentissage</i>	<i>validation</i>	<i>Apprentissage</i>	<i>Validation</i>
2	10-12-1	0.77401	0.75966	0.3614	1.9234
3	10-12-1	0.88266	0.86618	0.2160	1.9120
4	10-12-1	0.7785	0.70047	0.2629	2.262
5	10-12-1	0.85527	0.84211	0.2511	2.6680
6	10-12-1	0.7307	0.59187	0.3751	3.1947
7	10-12-1	0.67088	0.40478	0.4522	3.1440
8	10-12-1	0.69281	0.52756	0.3321	3.2922
9	10-12-1	0.6152	0.24102	0.4671.8	3.4562
10	10-12-1	0.6737	0.57069	0.4587.1	3.987.2
11	10-12-1	0.6403	0.53882	0.3775.3	4.3870

Tableau 4.18 : Les résultats du modèle PMC récurrent pour le polluant SO₂.



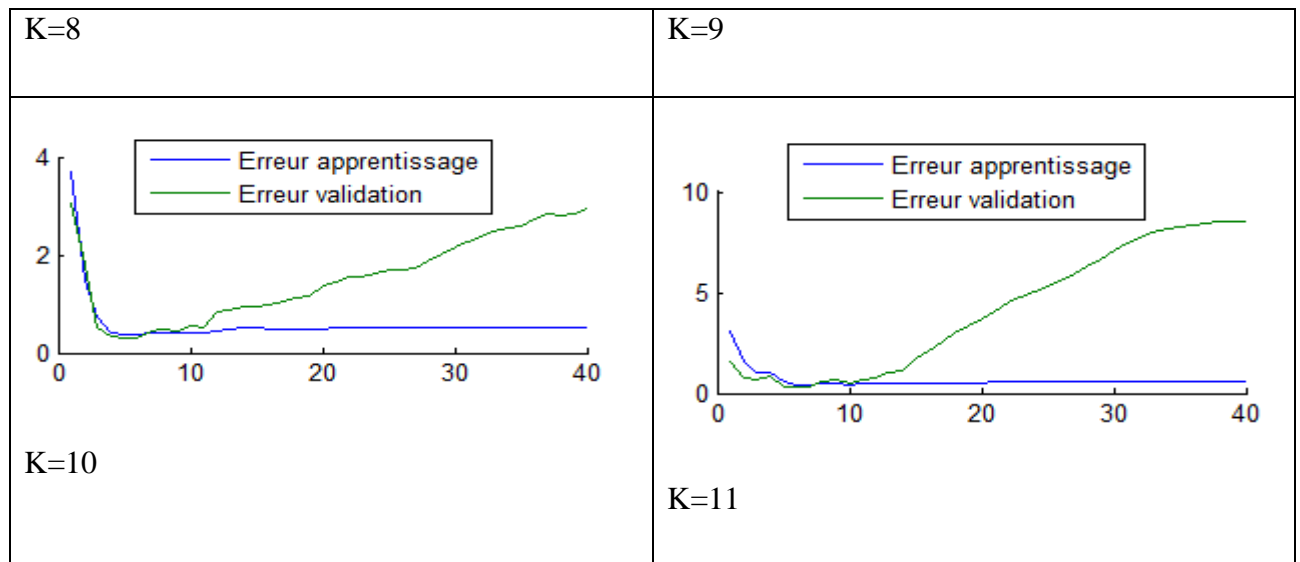


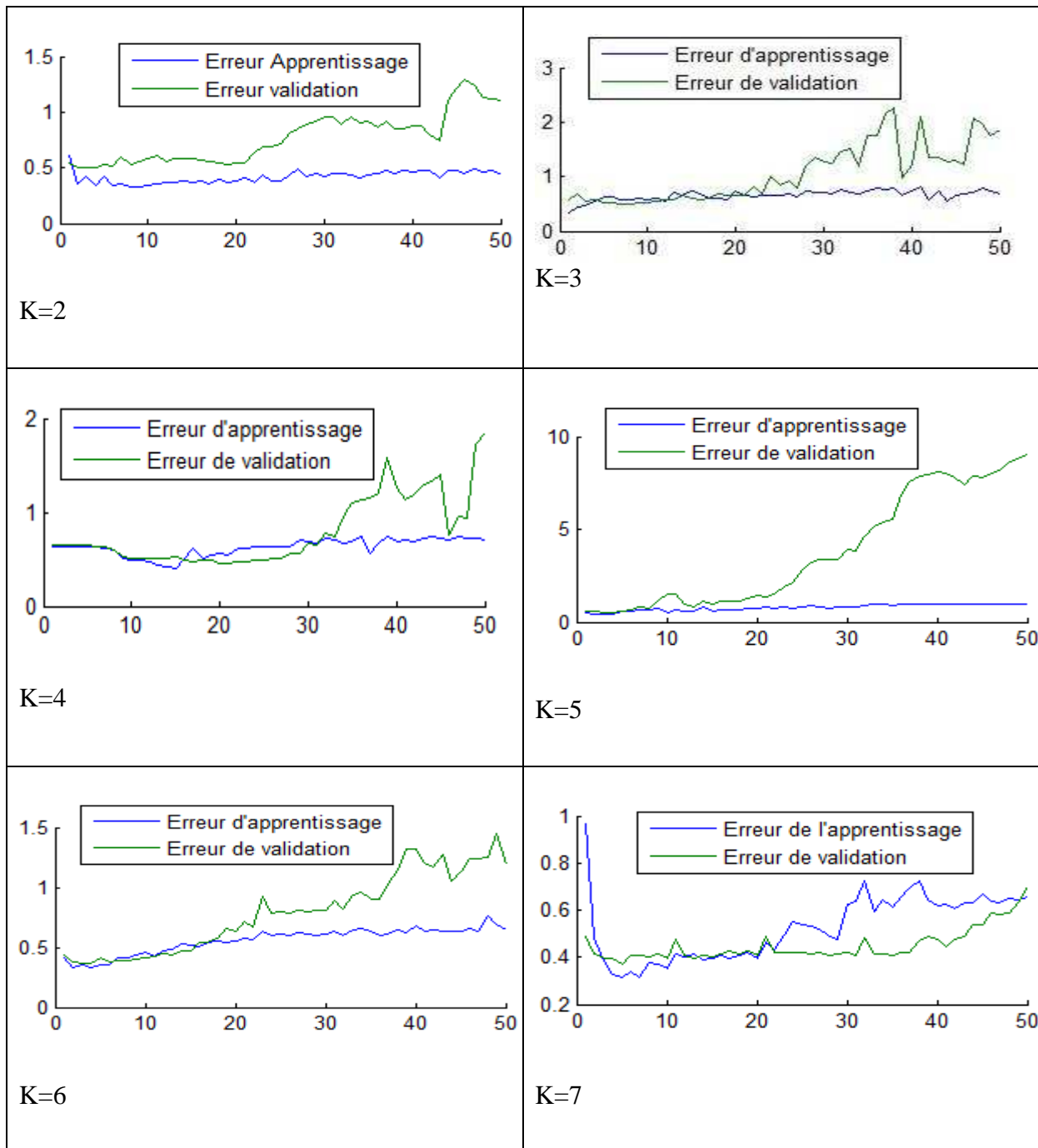
Tableau 4.19 : Les courbes d'erreur des modèles basés PMC récurrent.

NO₂ :

<i>K le nombre des pas</i>	<i>Architecture</i>	<i>IA</i>		<i>MSE</i>	
		<i>Apprentissage</i>	<i>validation</i>	<i>Apprentissage</i>	<i>Validation</i>
2	10-12-1	0.77401	0.75966	0.3614	1.9234
3	10-12-1	0.88266	0.86618	0.2160	1.9120
4	10-12-1	0.94406	0.76639	0.6453	1.4728
5	10-12-1	0.9495	0.8778	0.595	1.3535
6	10-12-1	0.83704	0.60696	0.7032	1.8034
7	10-12-1	0.83704	0.7068	0.6703	1.0834
8	10-12-1	0.76116	0.43876	0.40523	1.4507

9	10-12-1	0.50493	0.37708	0.16757	1.5028
10	10-12-1	0.8054	0.55736	0.19101	1.06922
11	10-12-1	0.68054	0.56808	0.4803	1.07990

Tableau 4.20 : Performance du modèle PMC récurrent pour le NO₂.



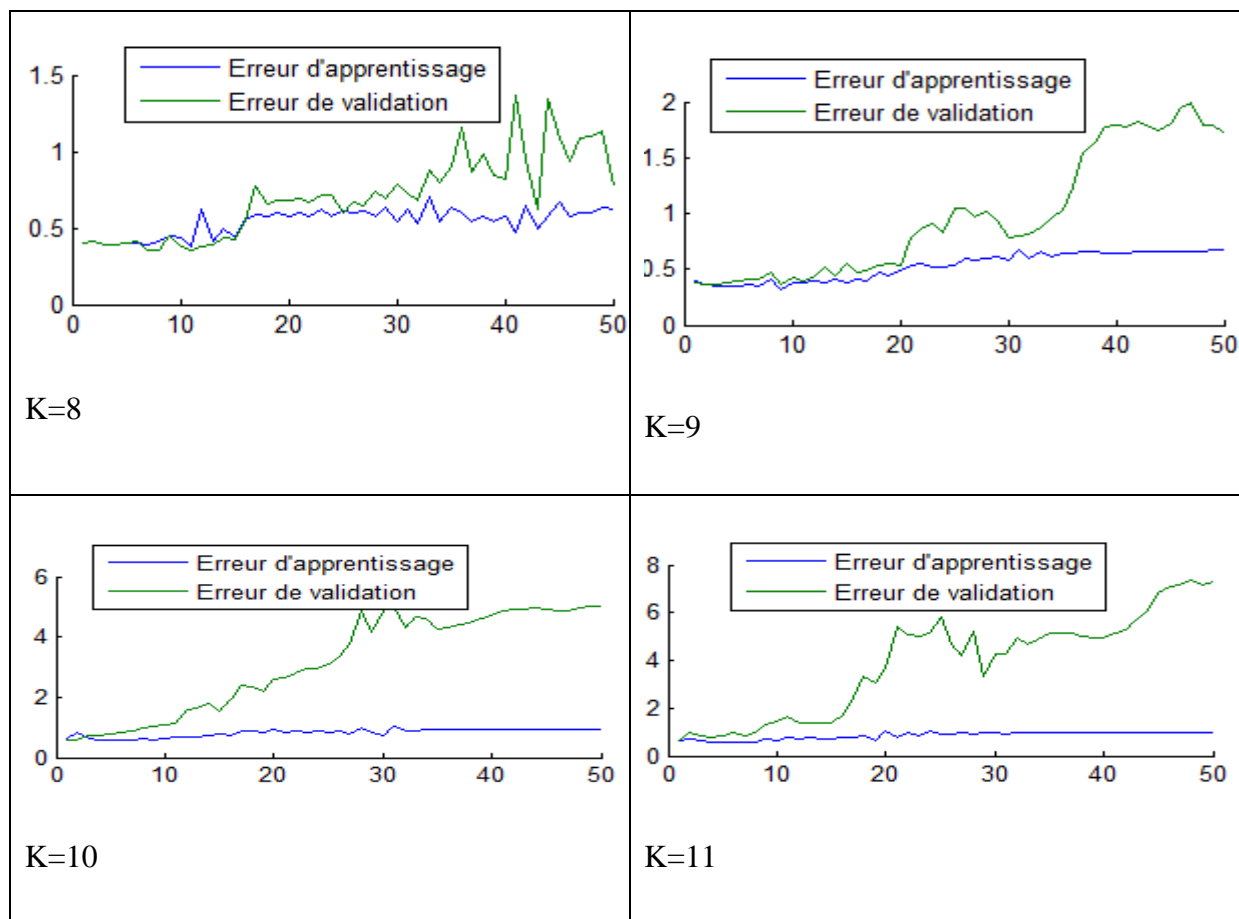


Tableau 4.21 : Les courbes d'erreur des modèles basés PMC récurrent.

PM10

<i>K le nombre des pas</i>	<i>Architecture</i>	<i>IA</i>		<i>MSE</i>	
		<i>Apprentissage</i>	<i>validation</i>	<i>Apprentissage</i>	<i>Validation</i>
2	10-12-1	0.91597	0.53994	0.046998	2.269
3	10-12-1	0.9187	0.59345	0.041058	1.5153
4	10-12-1	0.84592	0.62931	0.053112	1.49767
5	10-12-1	0.91974	0.61875	0.05229	1.53124

6	10-12-1	0.87899	0.51023	0.051036	1.55688
7	10-12-1	0.80632	0.65334	0.045467	1.56769
8	10-12-1	0.78181	0.5142	0.046156	1.60137
9	10-12-1	0.78838	0.44384	0.048335	1.66984
10	10-12-1	0.67307	0.44848	0.049239	1.57534
11	10-12-1	0.61482	0.62975	0.050824	1.64203

Tableau 4.22 : Performance du modèle PMC récurrent pour le PM10.

O₃

<i>K le nombre des pas</i>	<i>Architecture</i>	<i>IA</i>		<i>MSE</i>	
		<i>Apprentissage</i>	<i>validation</i>	<i>Apprentissage</i>	<i>Validation</i>
2	10-12-1	0.86324	0.57889	0.029405	1.099421
3	10-12-1	0.8902	0.79588	0.0443	1.023173
4	10-12-1	0.84178	0.81654	0.022994	1.024793
5	10-12-1	0.87777	0.58396	0.028946	1.032385
6	10-12-1	0.80643	0.6704	0.042839	1.040415
7	10-12-1	0.6472	0.32785	0.080114	1.52495

8	10-12-1	0.71826	0.6563	0.040328	2.054053
9	10-12-1	0.68287	0.65467	0.029163	1.053346
10	10-12-1	0.61454	0.5509	0.032692	1.052929
11	10-12-1	0.59317	0.57044	0.031878	1.057712

Tableau 4.23 : Performance du modèle PMC récurrent pour le O₃.

CO₂

<i>K le nombre des pas</i>	<i>Architecture</i>	<i>IA</i>		<i>MSE</i>	
		<i>Apprentissage</i>	<i>validation</i>	<i>Apprentissage</i>	<i>Validation</i>
2	10-12-1	0.76699	0.42856	0.052578	1.011234
3	10-12-1	0.78569	0.482	0.018955	1.017596
4	10-12-1	0.82023	0.6864	0.027448	1.029253
5	10-12-1	0.8535	0.52406	0.07091	1.028716
6	10-12-1	0.71953	0.55889	0.011316	1.017691
7	10-12-1	0.79308	0.44979	0.024344	1.0248
8	10-12-1	0.65953	0.4553	0.013326	1.021549
9	10-12-1	0.82473	0.44017	0.095564	1.053707

10	10-12-1	0.60641	0.3796	0.032812	2.036023
11	10-12-1	0.71051	0.3937	0.16118	2.026202

Tableau 4.24 : Performance du modèle PMC récurrent pour le CO₂.

4.2.3 Modèle basé réseaux d'Elman

Le modèle basé Elman comme présenté dans le chapitre 2, permettent de bien prédire, avec un temps d'apprentissage très court.

Paramètres	Nombre de valeurs passées à chaque époque	Description
Polluant	1 valeur	La concentration du polluant
Température	1 valeur	La température
Humidité	1 valeur	L'humidité
Vitesse du vent	1 valeur	La vitesse du vent

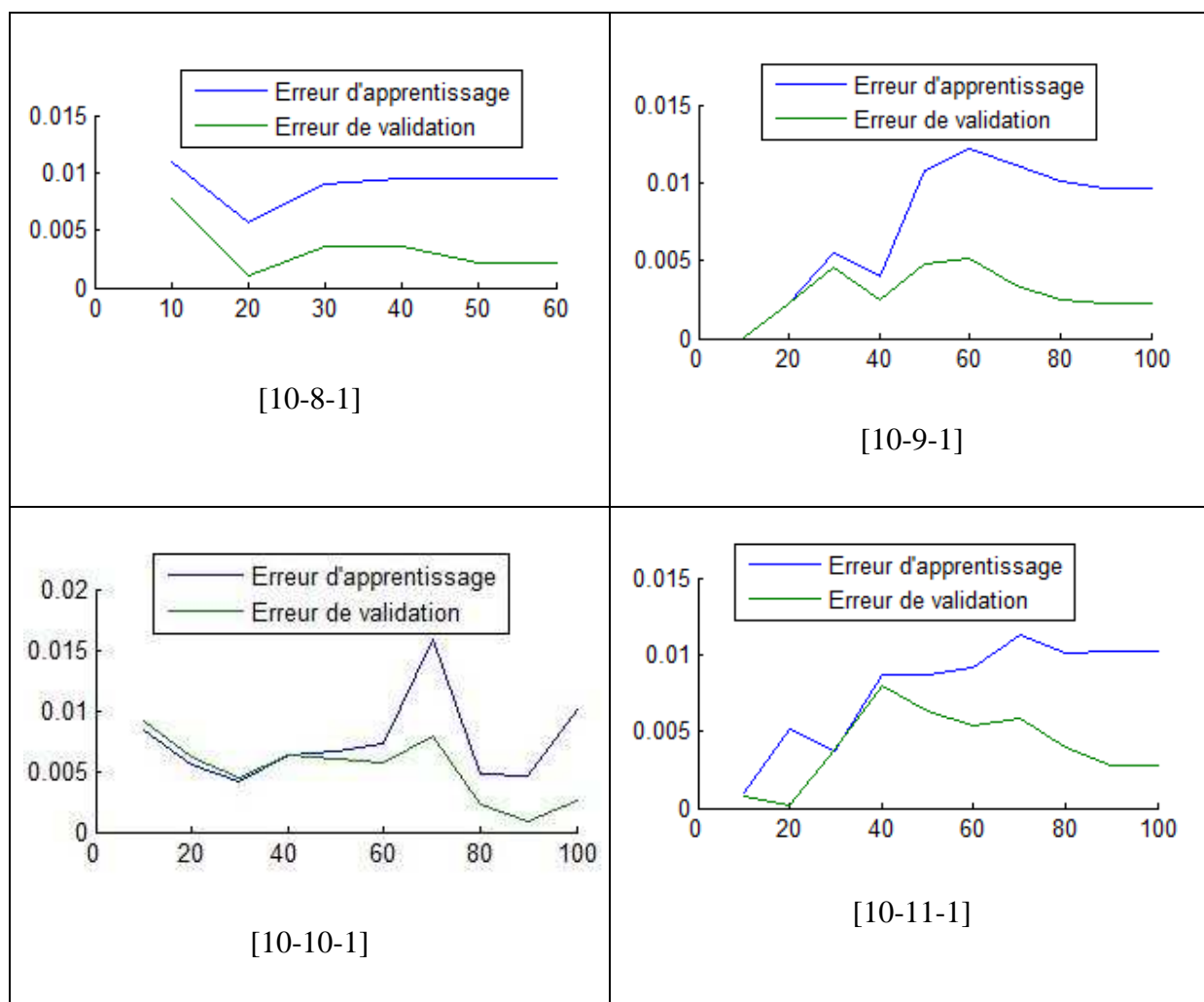
Tableau 4.25 : Paramètres du modèle Elman.

PM10

<i>Architecture</i>	<i>IA</i>		<i>MSE</i>	
	<i>Apprentissage</i>	<i>Validation</i>	<i>Apprentissage</i>	<i>Validation</i>
[10-8-1]	0.99349	0.90899	0.76642	4.4045

[10-9-1]	0.99342	0.90552	0.7635	4.3394
[10-10-1]	0.99336	0.88777	0.79398	4.4285
[10-11-1]	0.9947	0.89638	0.75991	4.3207
[10-12-1]	0.99693	0.91781	0.75835	4.4583
[10-13-1]	0.99277	0.88837	0.7754	4.4093
[10-14-1]	0.99492	0.93139	0.82503	4.3747

Tableau 4.26 : Performance du modèle Elman pour le PM10.



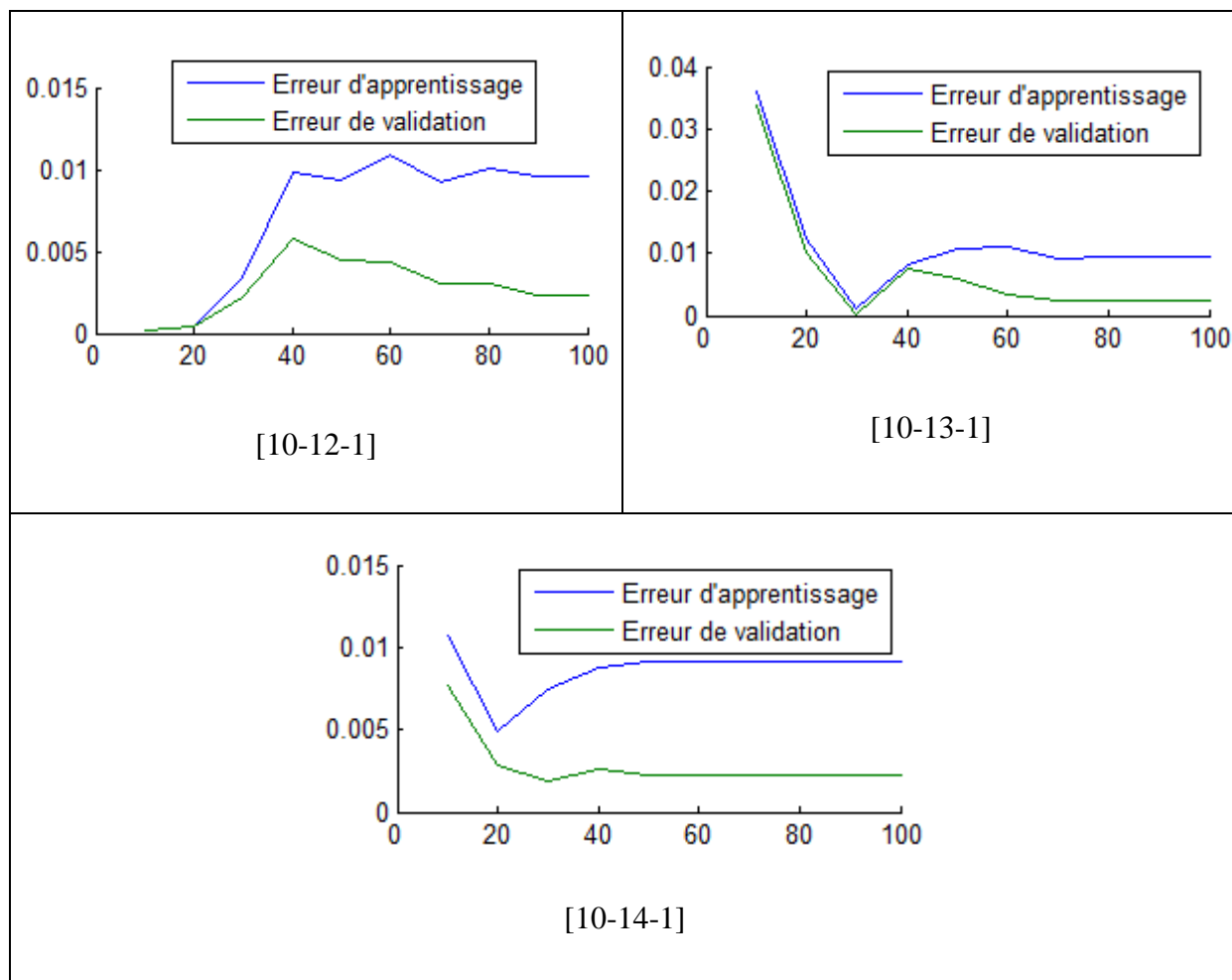


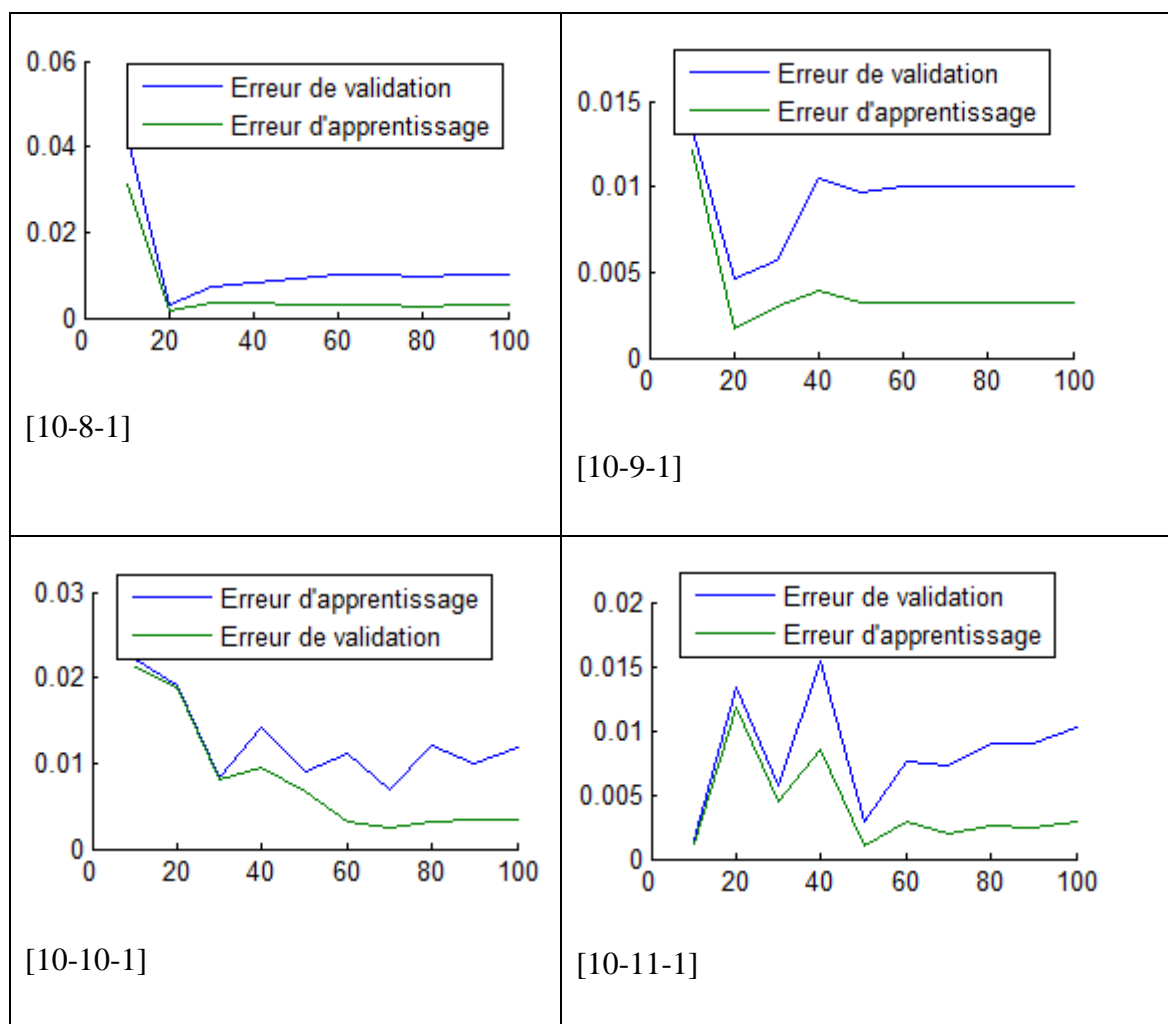
Tableau 4.27 : Courbes d'erreurs des modèles basés Elman pour le PM10.

NO₂

<i>Architecture</i>	<i>IA</i>		<i>MSE</i>	
	<i>Apprentissage</i>	<i>Validation</i>	<i>Apprentissage</i>	<i>Validation</i>
[10-8-1]	0.99174	0.88322	0.80465	3.0779
[10-9-1]	0.99196	0.93024	0.87902	3.0905
[10-10-1]	0.99403	0.95582	0.91718	3.058

[10-11-1]	0.9927	0.94002	.88212	2.9005
[10-12-1]	0.99829	0.97431	0.91844	2.986
[10-13-1]	0.99332	0.93578	0.86236	2.9583
[10-14-1]	0.99268	0.95521	0.93671	2.9897

Tableau 4.28 : Performance du modèle basé Elman pour le NO₂.



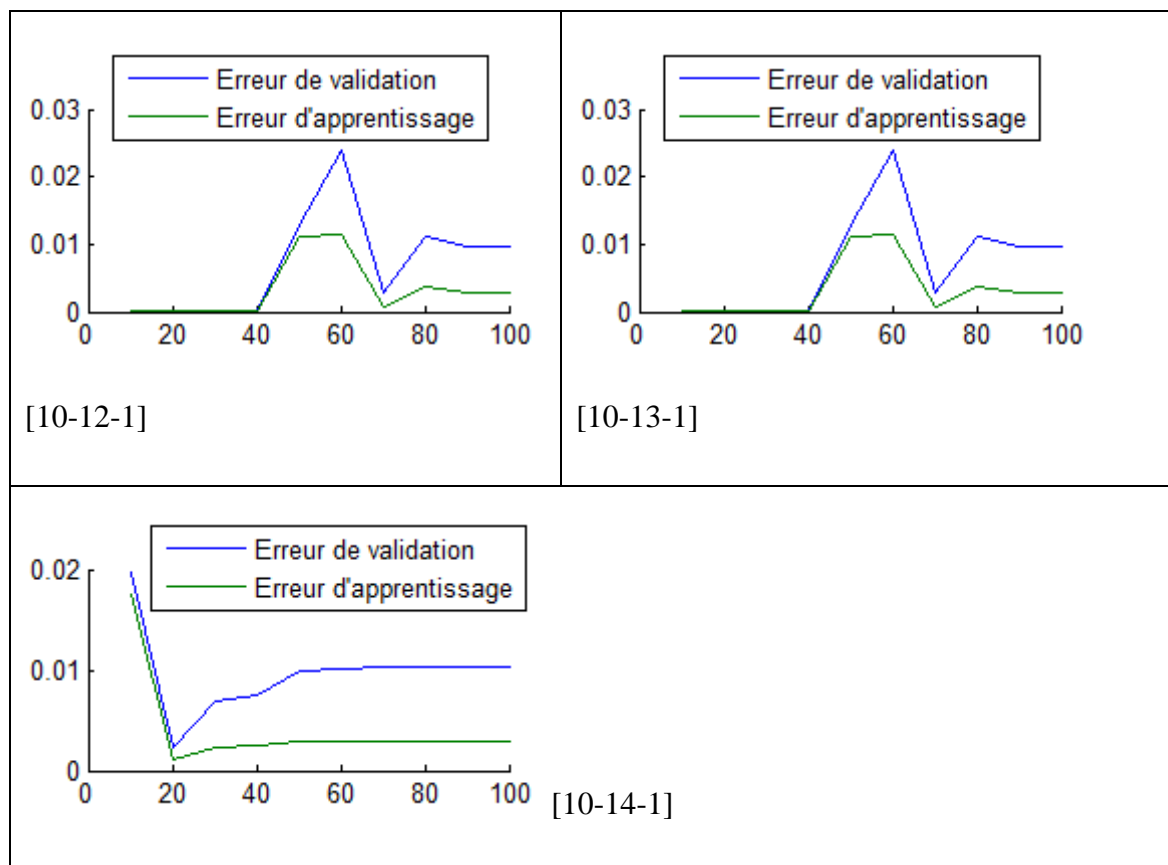


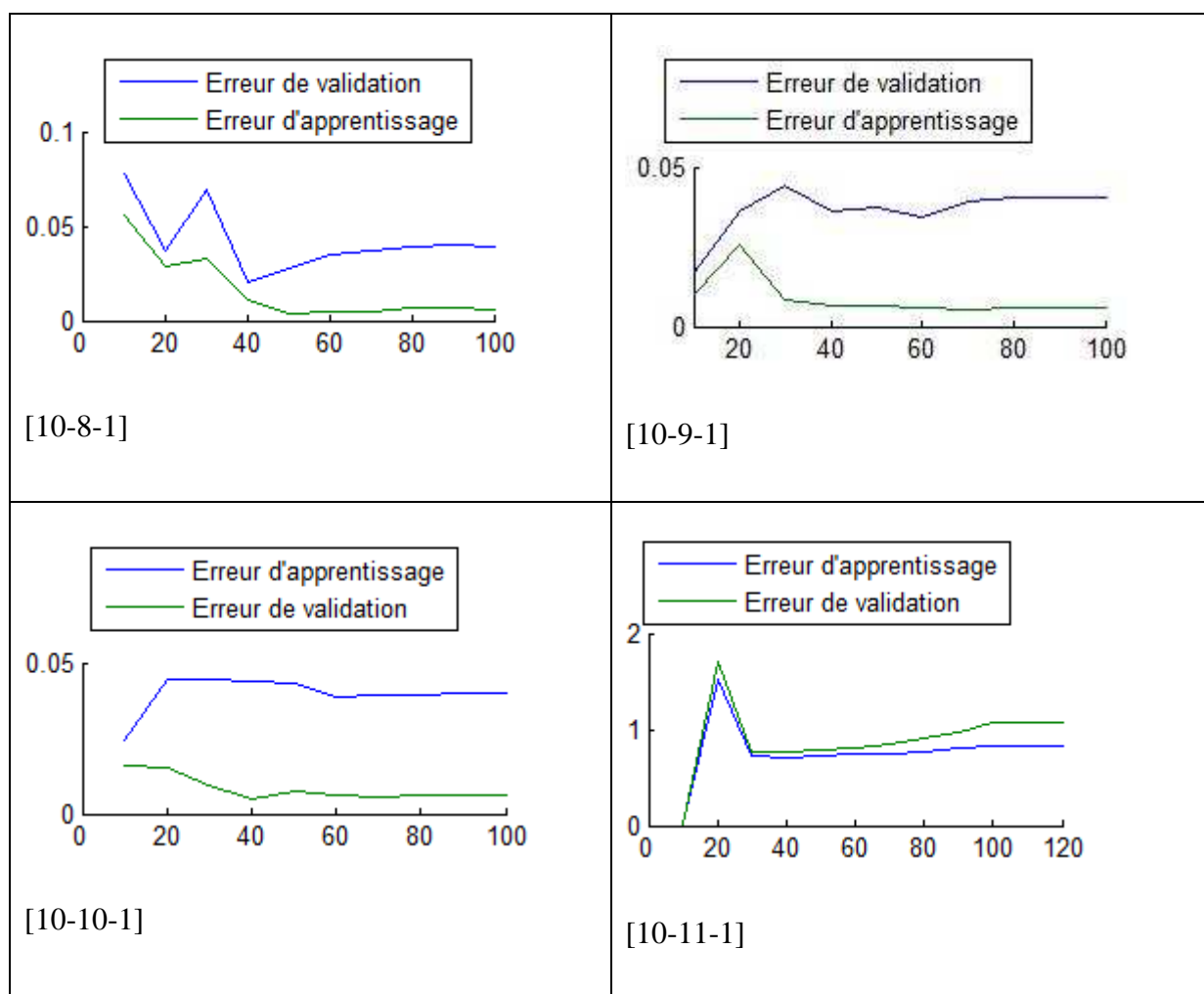
Tableau 4.29 : Courbes de l'erreur du modèle basé Elman pour le NO₂.

CO₂

<i>Architecture</i>	<i>IA</i>		<i>MSE</i>	
	<i>Apprentissage</i>	<i>Validation</i>	<i>Apprentissage</i>	<i>Validation</i>
[10-8-1]	0.99862	0.89041	0.039798	1.8186
[10-9-1]	0.99863	0.89067	0.04013	1.7937
[10-10-1]	0.99861	0.88857	0.039862	1.7877
[10-11-1]	0.99858	0.88985	0.040243	1.7888

[10-12-1]	0.99877	0.89460	0.040856	1.9675
[10-13-1]	0.9476	0.8796	0.01985	2.8313
[10-14-1]	0.99744	0.8642	0.019557	2.7331

Tableau 4.30 : Performances du modèle basé Elman pour le CO₂.



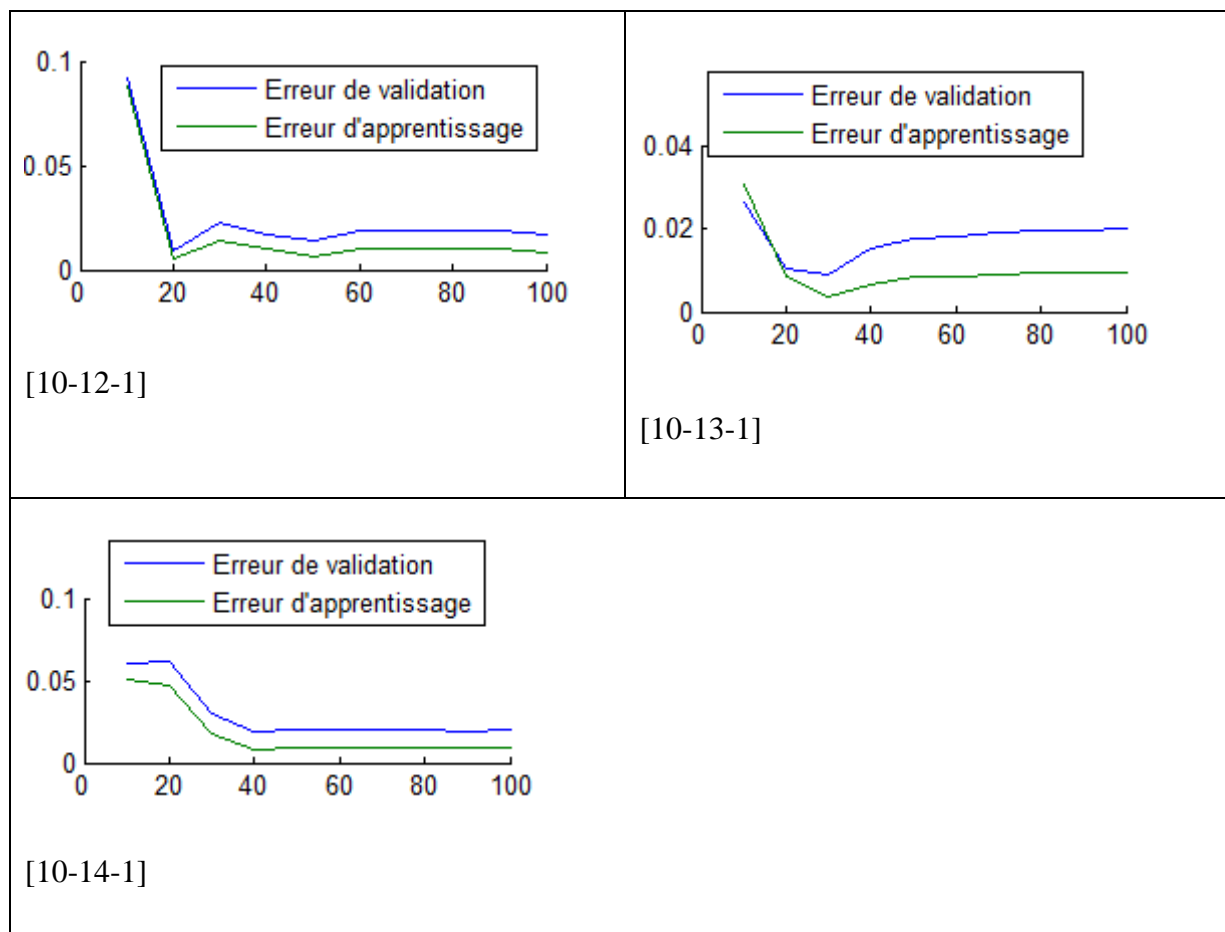


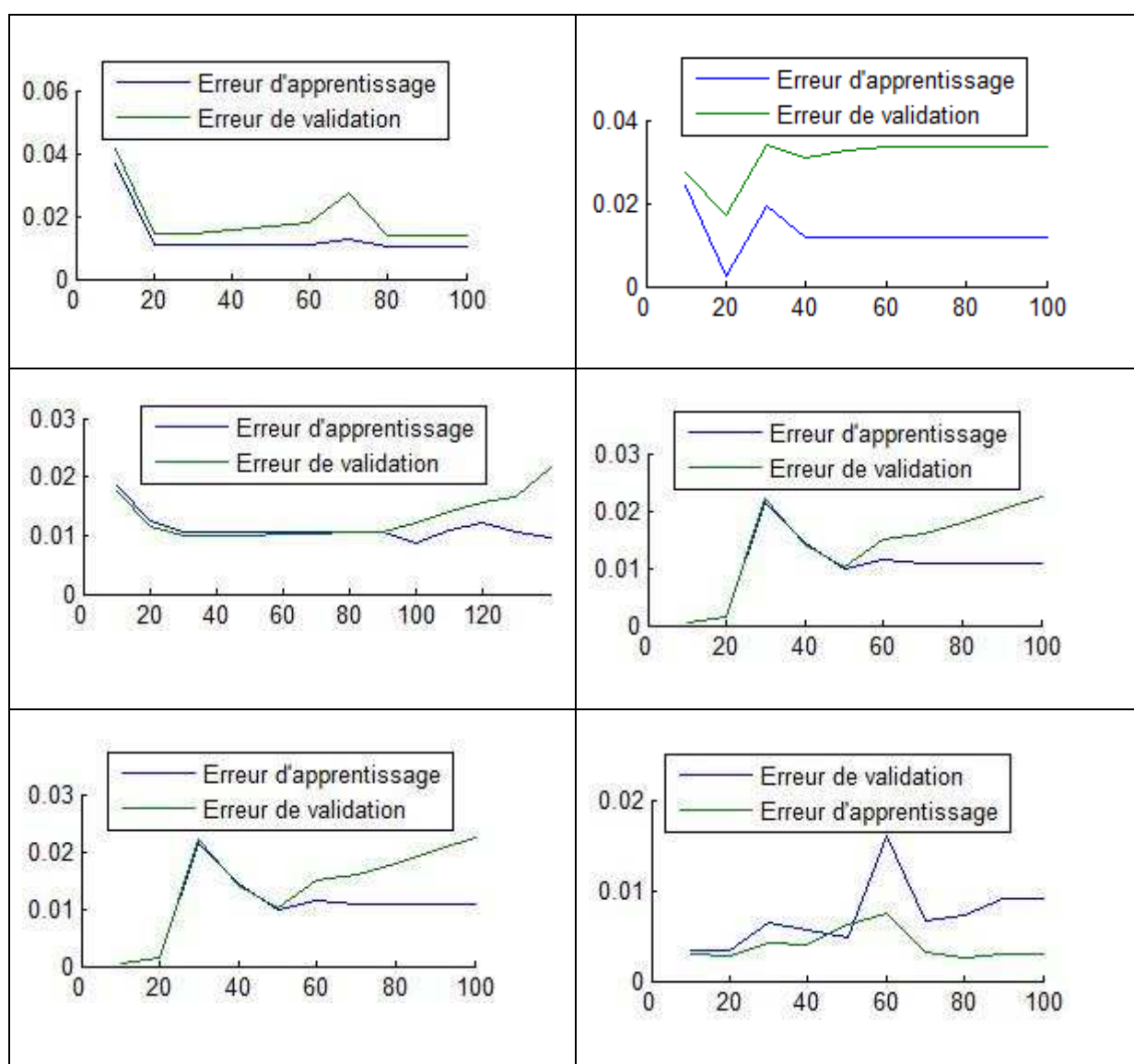
Tableau 4.31 : Courbes de l'erreur du modèle basé Elman pour le CO₂.

SO₂

<i>Architecture</i>	<i>IA</i>		<i>MSE</i>	
	<i>Apprentissage</i>	<i>Validation</i>	<i>Apprentissage</i>	<i>Validation</i>
[10-8-1]	0.94616	0.78174	0.95153	1.2471
[10-9-1]	0.9318	0.75325	0.11236	4.0211
[10-10-1]	0.94182	0.85809	0.9028	3.0094
[10-11-1]	0.90899	0.88464	0.87420	6.5363

[10-12-1]	0.93061	0.88014	0.99268	3.3089
[10-13-1]	0.96114	0.90964	0.6681	3.1694
[10-14-1]	0.93362	0.92574	0.012658	1.9379

Tableau 4.32 : Performance du modèle basé Elman pour le SO₂.



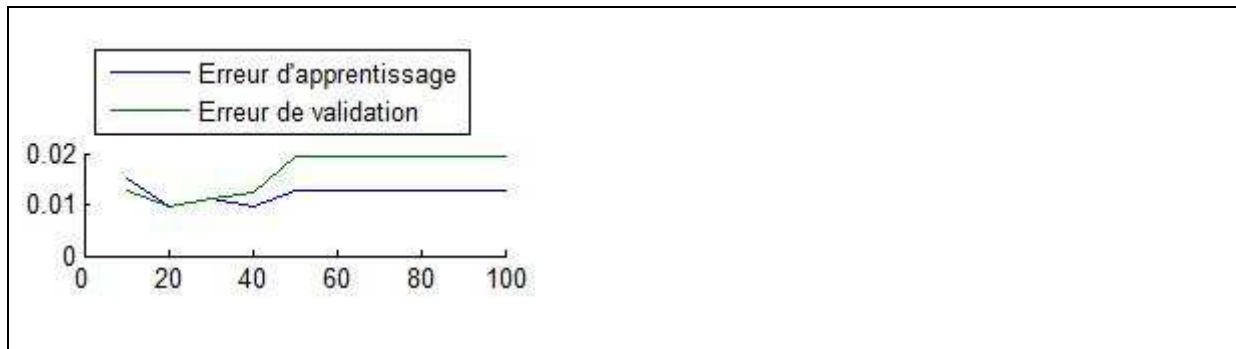
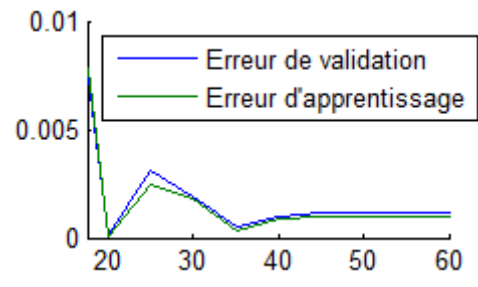


Tableau 4.33 : Courbes de l'erreur du modèle basé Elman pour le SO₂.

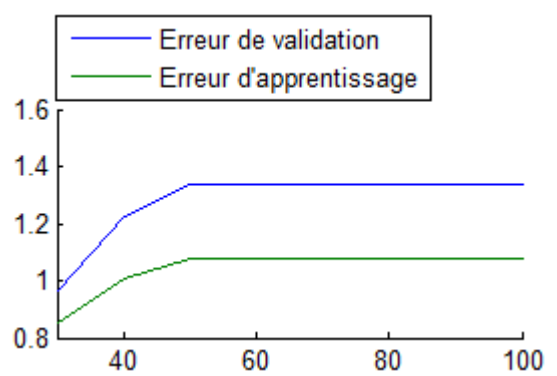
O₃

<i>Architecture</i>	<i>IA</i>		<i>MSE</i>	
	<i>Apprentissage</i>	<i>Validation</i>	<i>Apprentissage</i>	<i>Validation</i>
[10-8-1]	0.96492	0.95115	0.35213	0.84336
[10-9-1]	0.94475	0.90174	0.40207	0.97151
[10-10-1]	0.91906	0.8751	0.89083	0.68721
[10-11-1]	0.91424	0.89577	0.42359	0.54183
[10-12-1]	0.89835	0.87712	0.33695	0.41822
[10-13-1]	0.90107	0.74896	0.37845	0.27922
[10-14-1]	0.91066	0.85135	0.43139	0.5477

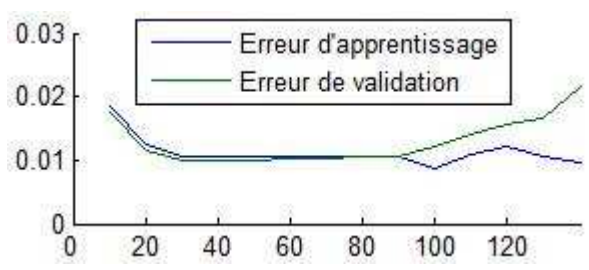
Tableau 4.34 : Performances du modèle basé Elman pour le O₃.



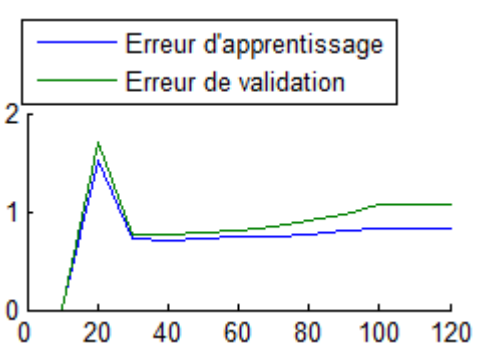
[10-8-1]



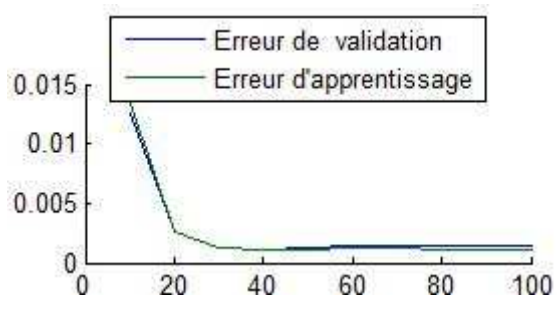
[10-9-1]



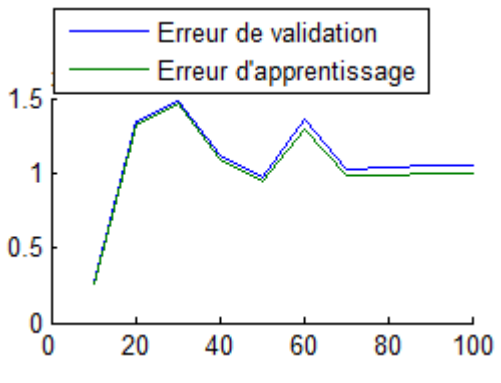
[10-10-1]



[10-11-1]



[10-12-1]



[10-13-1]

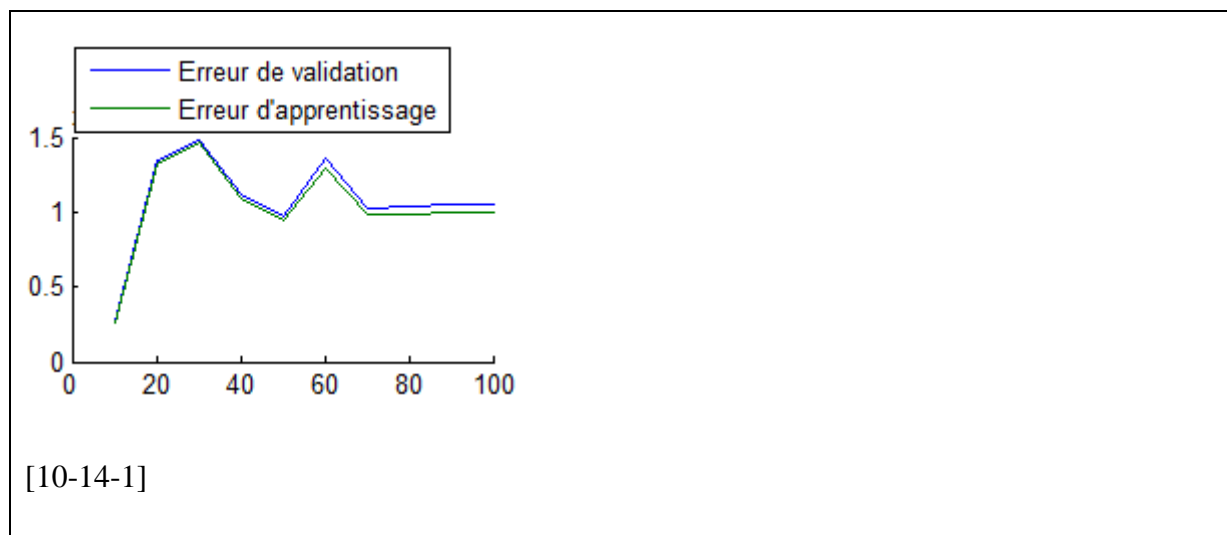


Tableau 4.35 : Courbes de l'erreur du modèle basé Elman pour le O₃.

4.2.4 Modèle basé réseau RBF

Paramètres	Nombre de valeurs passées à chaque époque	Description
Polluant	1 valeur	La concentration du polluant
Température	1 valeur	La température
Humidité	1 valeur	L'humidité
Vitesse du vent	1 valeur	La vitesse du vent

Tableau 4.36 : Paramètres du modèle basé RBF

Au fur et mesure de l'apprentissage, à chaque époque le modèle basé RBF doit ajuster le nombre de neurones dans la couche cachée, cela se fait à l'aide d'un paramètre qui indique le nombre de neurones à ajouter à chaque époque.

Ce nombre permet de paramétrer le modèle. En changeant ce paramètre de 5 jusqu'à 15 trois modèles ont été créés pour chaque polluant. Cela permet d'avoir l'architecture la plus optimale qui donne les meilleurs résultats.

<i>Architecture</i>	<i>Polluant</i>	<i>IA</i>		<i>MSE</i>	
		<i>Apprentissage</i>	<i>Validation</i>	<i>Apprentissage</i>	<i>Validation</i>
[10-320]	PM10	0.99999	0.99873	0.0004691	0.5508
[10-180]	O3	0.99999	0.69220	0.022969	1.2571
[10-105]	NO2	0.99667	0.48666	0.13581	2.7515
[10-90]	SO2	0.99762	0.59305	0.099593	2.9806
[10-45]	CO2	0.99004	0.97164	0.10908	1.0513

Tableau 4.37 : Performance du modèle basé RBF

4.3 Comparaison des résultats obtenus

Les modèles basés RBF et Elman ont montré les meilleurs résultats en suite viennent ceux basés PMC

En dernier ceux basés PMC récurrent permettent d'avoir toute une plage de prédiction 12 valeurs ce qui permet d'avoir une idée sur le futur développement de la concentration du polluant. Les modèles PMC récurrent aussi montrent quelques défaillances à prédire les pics. Cela est dû au fait que si quelques pics ayant lieu dans les derniers pas de prédiction dans lesquels la précision est moins efficace le modèle risque de sous-estimer les pics.

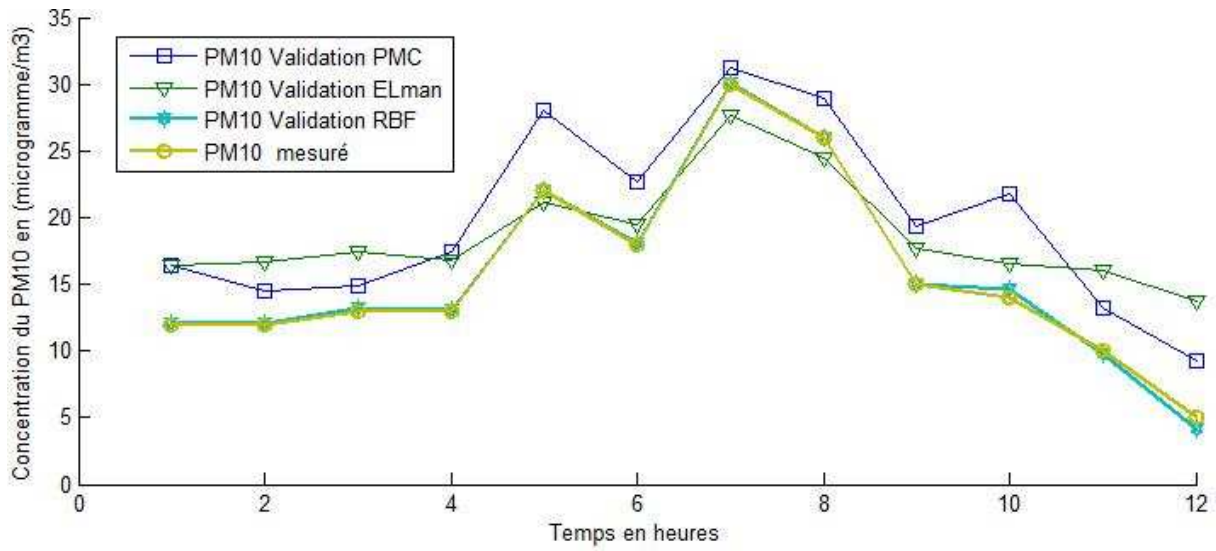


Figure 4.2 : Comparaison de résultat des modèles pour la prédiction du PM10

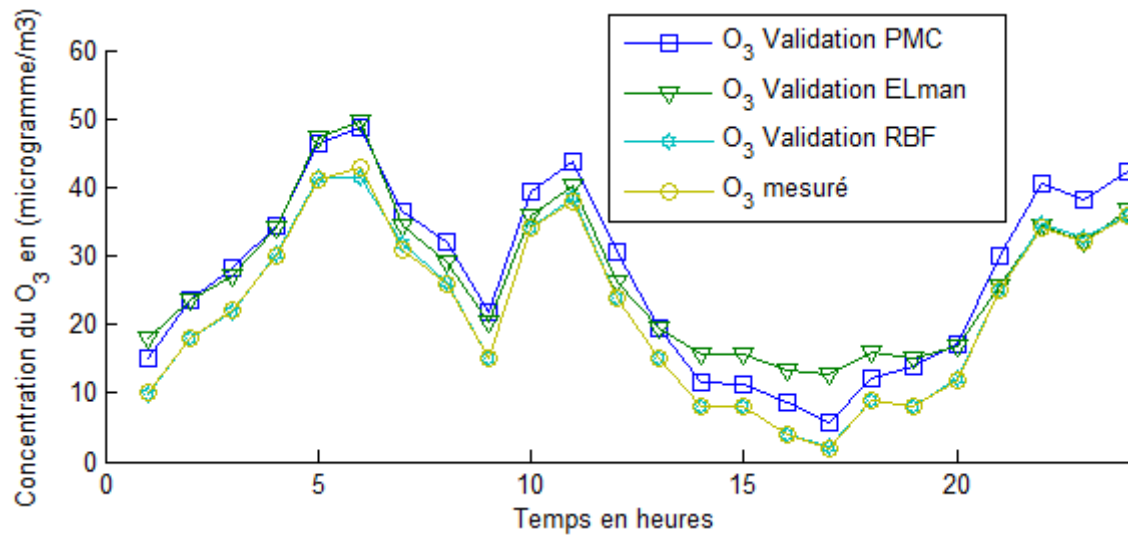


Figure 4.3 : Comparaison du résultat des modèles pour la prédiction de l'O₃

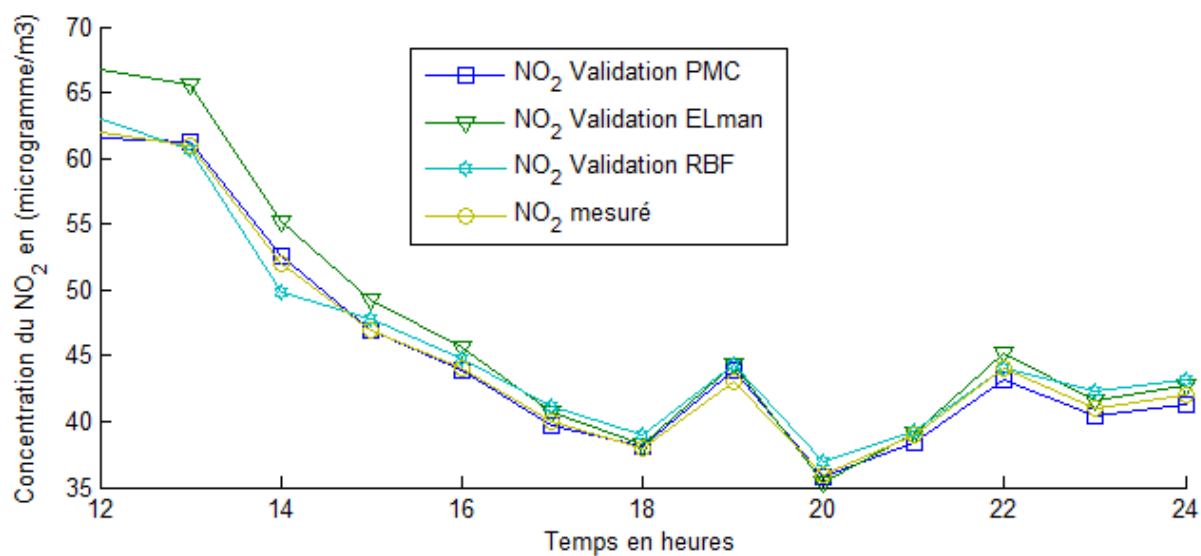


Figure 4.4 : Comparaison du résultat des modèles pour la prédiction de l'NO₂.

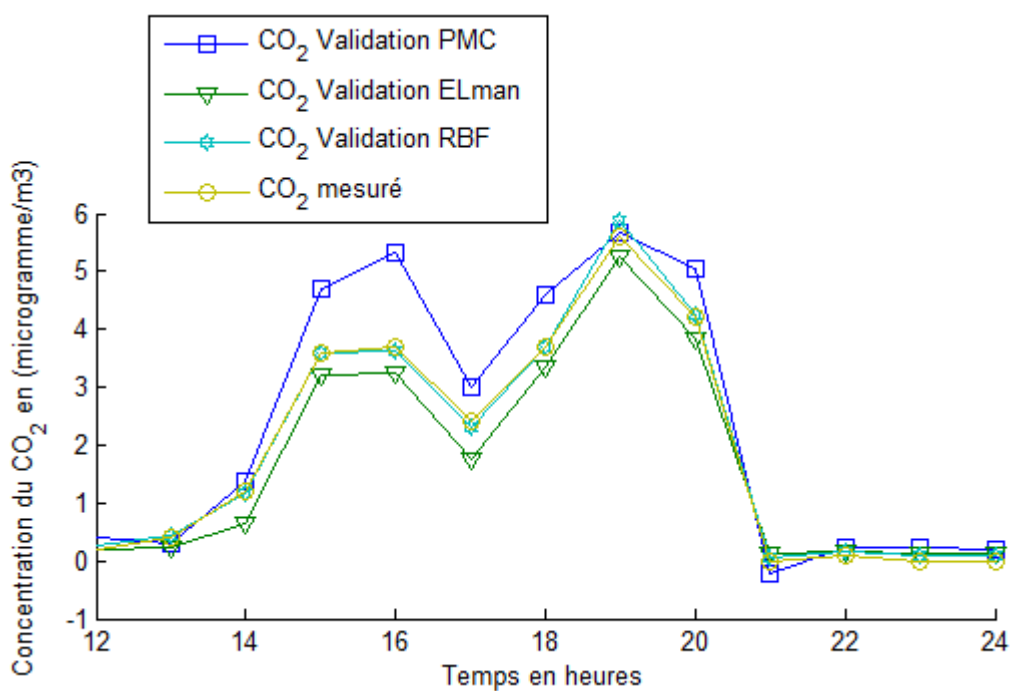


Figure 4.5 : Comparaison du résultat des modèles pour la prédiction de l'CO₂.

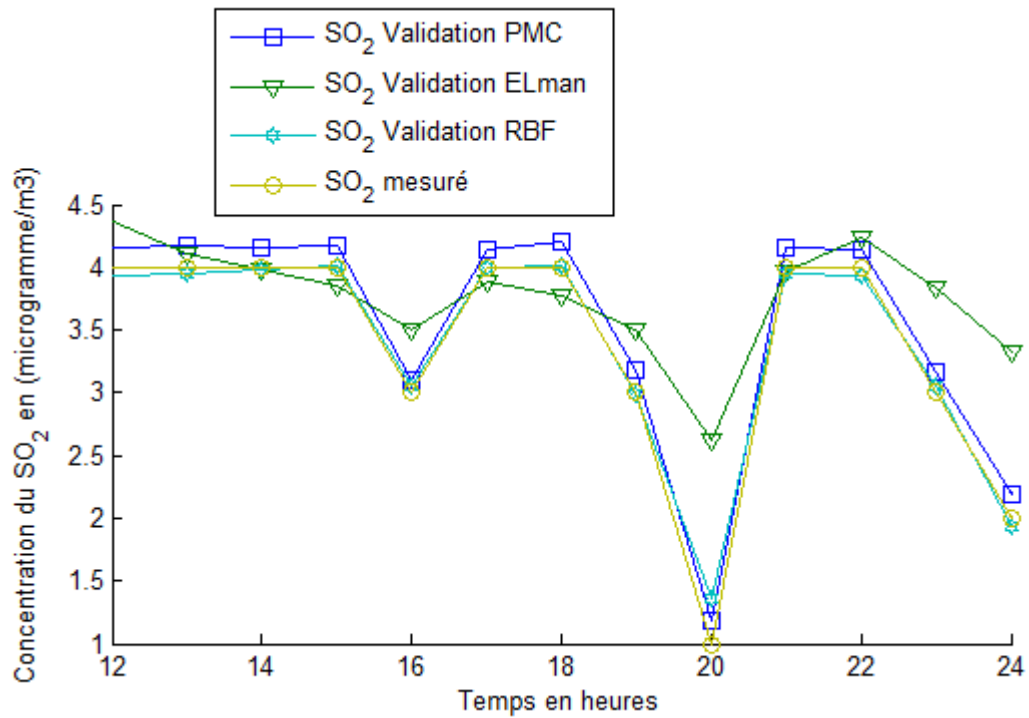


Figure 4.6 : Comparaison du résultat des modèles pour la prédiction de l'SO₂.

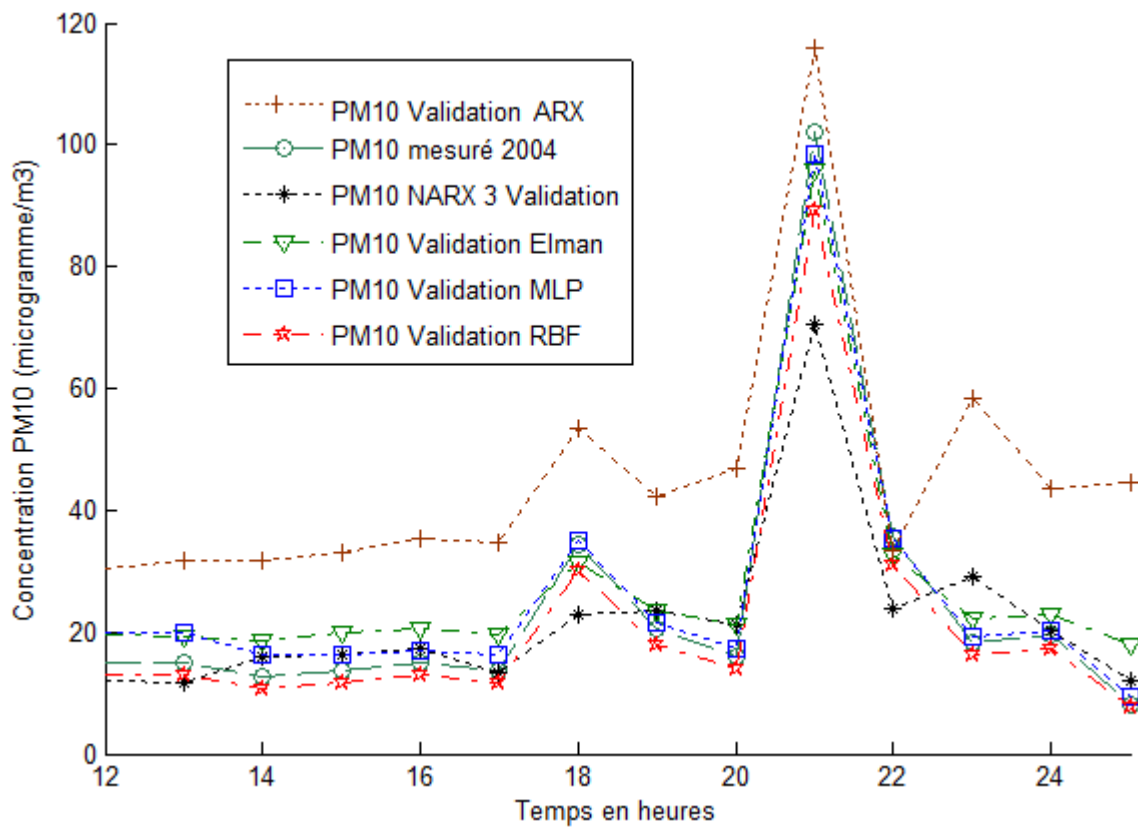


Figure 4.7 : Comparaison du résultat des modèles pour la prédiction de PM10.

Comme le montrent les figures 4.2, 4.3, 4.4, 4.5, 4.6 et 4.7 les modèles basés RBF ont été les mieux à prédire efficacement la concentration des polluants, quoique pour les polluants qui ont deux années de mesure les modèles montrent quelques défaillances. Ceci est logique car le comportement de la série temporelle est différent d'une année à l'autre.

Du point de vue performance les modèles basés réseaux d'Elman permettent aussi de bien prédire la concentration des polluants, mais dans le cas du SO₂ ils montrent des défaillances.

Les modèles basés PMC donnent aussi des performances convaincantes mais présentent des défaillances dues au sur-apprentissage surtout pour les polluants qui n'ont pas assez de données.

Les PMC récurrents donnent des prédictions efficaces pour des nombres de pas moyens mais dès que ce nombre dépasse 6 heures leurs performances se dégradent, et sous-estiment quelques pics ou surestiment quelques uns. Leur utilisation reste intéressante pour avoir des prédictions de type (moyenne de la journée ou la valeur maximale)

Dans le cas des PMC récurrents quand les pics ont lieu dans les derniers pas de prédiction, pour lesquels le modèle donne des résultats moins efficaces, le modèle sous-estime le pic. Ce comportement peut être corrigé par l'utilisation d'un nombre de pas moyen ou petit.

Comme le montre la Figure 4.7 les modèles basés ARX sont les moins efficaces, et ne peuvent pas généraliser en dehors des données utilisées dans l'apprentissage. Cela est dû à la non-linéarité des séries temporelles de concentration des polluants de l'air.

Chapitre 5

Modèle global de prédiction de la qualité de l'air

5.1 Modèle global de prédiction de la qualité de l'air

En se basant sur les modèles de prédiction de la concentration des polluants dans l'air, un modèle global de prédiction de qualité de l'air est conçu.

Ce modèle doit donner des décisions sur la qualité de l'air (indice) selon les normes et les règlements en vigueur. En fait selon la concentration des différents polluants une décision peut être donnée classant la qualité de l'air pour les prochaines 12 heures. Le modèle reçoit en entrée les paramètres climatiques ainsi que la concentration en cours des différents polluants et donne en sortie la classe de la qualité de l'air de les prochaines 12 heures, le schéma présenté dans la figure 5.1 illustre la structure de modèle.

Le modèle se constitue de plusieurs modèles dont un pour évaluer la qualité de l'air, ce dernier est un simple PMC qui donne une valeur entre 1 et 10 qui signifient : 1 : très bonne, 2 : bonne, 3 : moyenne, 4 : mauvaise, 5 très mauvaise, etc. Il reçoit en entrée 5 paramètres qui sont :

- La concentration prédite des PM10,
- La concentration prédite de SO₂,
- La concentration prédite de CO₂,
- La concentration prédite de NO₂,
- La concentration prédite d'O₃.

Ces concentrations ont été prédites par les modèles de prédiction des concentrations présentés dans la section précédente. Un modèle basé PMC a été développé, avec 10 neurones dans la couche cachée, en se basant sur le tableau 5.1 l'ensemble d'apprentissage a été conçu, avec des sorties désirées bien déterminer qui sont des valeurs de l'indice entre 1 et 10, comme le modèle génère des valeurs réels une fonction d'arrondissement a été appliquée aux sorties de modèle pour avoir des valeurs fixes sans virgule.

Une autre fonction qui permet de donner l'appréciation a été développée qui consiste à convertir la valeur de l'indice prédite par le modèle en des appréciations de type (bonne, mauvaise etc). Ces valeurs sont faciles à comprendre par des utilisateurs simples qui n'ont pas de connaissance sur la signification des concentrations des polluants.

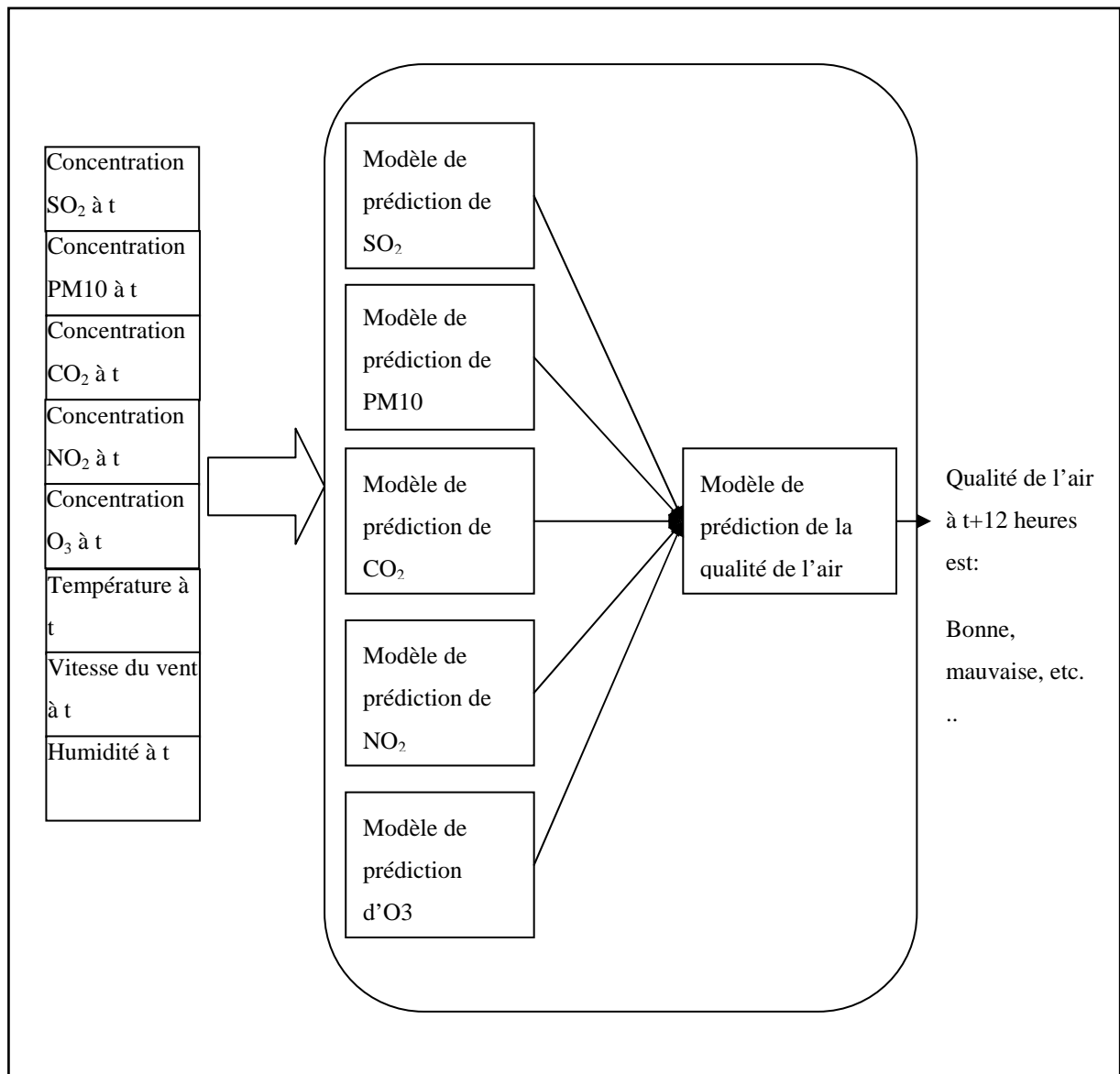


Figure 5.1 : Architecture d'un modèle global de prédiction de la qualité de l'air.

Ces indices sont basés sur les teneurs en ozone, dioxyde d'azote, dioxyde de soufre et les particules PM_{10} . Ils ne tiennent donc pas compte des composés en métaux lourds ou organiques dont les effets résultent plutôt (mais pas toujours) d'une exposition à long terme. Les indices portent donc sur les principaux polluants irritants. Pour chacun des polluants, on définit une échelle de concentrations pour calculer un sous-indice. Les échelles de concentrations sont basées à la fois sur les valeurs limites imposées par les directives internationales. L'indice de qualité est alors déterminé comme étant le sous-

indice le plus élevé, donc la qualité la plus médiocre. Si l'un des sous-indices manques, l'indice global n'est pas calculé et il faut au minimum 50 % de données valides pour calculer un sous-indice.

Le tableau suivant présent l'indice de qualité de l'air utilisé par SMASAFIA en 2005 :

<i>SO2</i>	<i>NO2</i>	<i>O3</i>	<i>PM10</i>	<i>Indice</i>	<i>Appréciation</i>
0 - 15	0-25	0-30	0-10	1	<i>Excellente</i>
15 - 30	14-45	30-45	10-20	2	<i>Très bonne</i>
45	60	60	30	3	<i>Assez bonne</i>
60	80	80	40	4	<i>Bonne</i>
80	110	100	50	5	<i>Moyen</i>
100	150	120	70	6	<i>Médiocre</i>
125	200	150	100	7	<i>Très médiocre</i>
165	270	200	150	8	<i>Mauvaise</i>
250	400	270	200	9	<i>Très mauvais</i>
>250	>400	>270	>200	10	<i>Exécrable</i>

Tableau 5.1 : L'indice de qualité de l'air utilisée par SamaSafia².

Le modèle a donné de très bonnes performances, comme montré dans la Figure 5.2, qui présenté une page de prédiction de l'indice de qualité de l'air pour la région de Annaba en 2004. Le modèle prédit d'une façon efficace l'indice de qualité de l'air.

² Ce tableau est disponible sur le site web du réseau national SAMASAFIA www.samasafia.dz/SmasafiaDoc/Decret/FP3.pdf

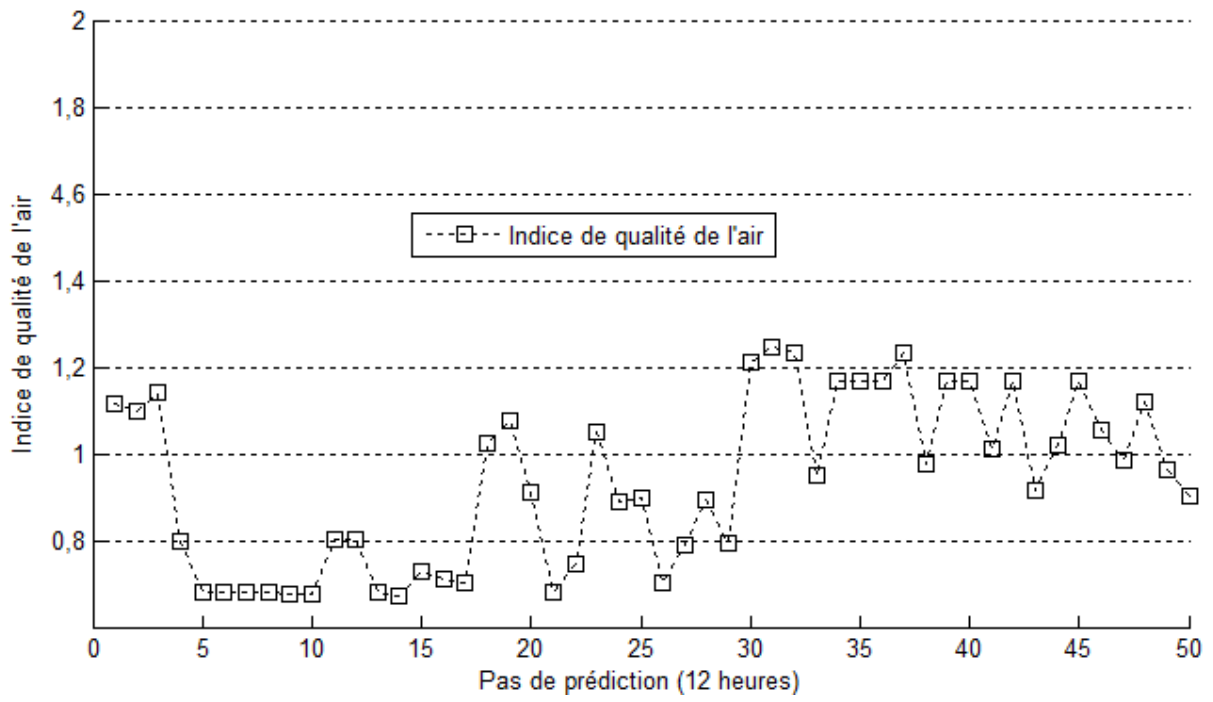


Figure 5.2 : Résultat du modèle global de prédiction de l'indice de qualité de l'air

Conclusion générale

Les systèmes de prédiction et de contrôle de la pollution de l'air constituent des outils efficaces de lutte contre les périodes des pics de pollution. Utilisés avec les autres outils de mesure et d'analyse ils permettent de bien prédire la concentration des polluants dans l'air. Les modèles de prédiction sont l'un des plus important module d'un système de prédiction et de contrôle de la pollution de l'air. Dans ce travail les modèles sont conçus dans le but de donner des prédictions à moyen terme (12 heures en avance) sur la concentration des cinq polluants mesurés par le réseau national de contrôle de la pollution de l'air SAMASAFIA dans la région d'Annaba. Plusieurs types de modèles ont été testés. Les modèles basés RBF et réseaux d'Elman ont donné les meilleurs résultats, mais avec des temps d'apprentissage et de conception important. Ceux basés PMC ont été les moins difficile à réaliser, et donnent des prédictions intéressantes. Les modèles basés PMC récurrent (NARX) ont l'avantage de donner des prédictions multi-pas ce qui permet de voir l'évolution de la concentration de polluant pendant 12 heures, ce type de modèle a donné des prédictions performantes pour des nombres de pas moyen (6 heures en avance) mais quand le nombre de pas deviens important tel que 12 heures leurs performances commence a se décroître, mais leurs utilisation reste utile pour avoir les pics par le calcul de la valeur maximal des k prédictions faites. A la fin viennent ceux basés ARX, qui sont très utilisés dans ce domaine, et sont les plus faciles à réaliser mais ne donne pas de bonne prédiction surtout pour les polluants qui possèdent deux années de mesure. Ces résultats ne restent valides que pour la région d'Annaba, dans le cas d'autre région la topologie et les variations climatiques affectent d'une façon directe la qualité des prédictions. En se basant sur les modèles qui ont donné les meilleurs résultats, un modèle global qui permet de prédire l'indice de la qualité de l'air de la région d'Annaba est présenté, ce modèle permet d'avoir des prédictions sur l'indice de la qualité de l'air selon les normes imposées par les lois algériennes et internationales.

A noter qu'il est nécessaire d'avoir suffisamment de données pour avoir des modèles capables à généraliser et à donner des prédictions efficaces, dans notre cas on a les données de mesure des différents paramètres de pollution de l'air et aussi des paramètres climatiques. L'existence des données relatives à l'émission permettra aussi d'améliorer la qualité de la prédiction notamment pour les polluants primaires (ceux qui sont émis dans l'air par les activités industriels).

Références

Acciani G. , Chiarantoni E. , and Fornarelli G. *A Neural Network Approach to Study O3 and PM10 Concentration in Environmental Pollution*, International Conference on Artificial Neural Networks, Part II, LNCS 4132, pp. 913-922, 2006.

Andrew C. Comrie, *Comparing Neural Networks and Regression Models for Ozone Forecasting*, Journal of the Air & Waste Management Association, volume 47, 653-663.

Barker K., Cambi F., Catcott E. J., *La pollution de l'air*, Organisation Mondiale de la Santé, Série des monographies, vol 46, 1963

Bordignon S. , Gaetan C. , Lisi F., *Nonlinear models for ground-level ozone forecasting*, Statistical Methods & Applications, vol 11, pp. 227-245, 2002.

Box, G. E. P., Jenkins, G.M., Reinsel, G. C. *Time series analysis: Forecasting and control* (3rd edn.). Englewood Cliffs, NJ: Prentice Hall. 1994.

Brunelli U., Piazza V., Pignato L., Sorbello F., Vitabile S. *Two-days ahead prediction of daily maximum concentrations of SO2, O3, PM10, NO2, CO in the urban area of Palermo, Italy* , Environmental Atmospheric environment, vol. 41, no14, pp. 2967-2995, 2007.

Carmichael Gregory R., Adrian Sandu, Tianfeng Chai, Dacian N. Daescu, Emil M. Constantinescu and Youhua Tang, *Predicting Air Quality: Current Status and Future Directions*, Air Pollution Modeling and Its Application XIX, Business Media B.V., pp. 481- 495, 2008.

Castellano M. , Franco A., *Identification of NOx and Ozone Episodes and Estimation of Ozone by Statistical Analysis*, Water Air Soil Pollution vol 2, 95-110, 2009.

Chelani A. B., Gajghate S. M., Tamhane S. M., Hasan M. Z. . *Statistical modeling of ambient air pollutants in Delhi*. Water, Air, and Soil Pollution, vol 132, 315–331, 2001.

Chen, S., C.F.N. Cowan, and P. M. Grant, *"Orthogonal Least Squares Learning Algorithm for Radial Basis Function Networks,"* IEEE Transactions on Neural Networks, vol. 2, no 2, pp. 302-309, March 1991.

Cigizoglu H. K., Alp K., Kömürcü M. *Estimation of air pollution parameters using artificial neural networks*, Advances in Air Pollution Modeling for Environmental Security, 63–75, 2005.

Cigizoglu H. K., Alp K., Kömürcü M. *Two Neural Network Methods in Estimation of Air Pollution Time Series*. Environmental Simulation Chambers: Application to Atmospheric Chemical Processes, 421–431, 2006.

Cihan Sahin, Ashish Thandavan and Vassil N. Alexandrov, *Grid Enablement of the Danish Eulerian Air Pollution Model*, Lecture Notes in Computer Science, High Performance Computing and Communications, 2006

Ciprian-Daniel Neagu, Nikolaos Avouris, Vasile Palade, *Neural and Neuro-Fuzzy Integration in a Knowledge-Based System for Air Quality Prediction*, Kluwer Academic Publishers, Applied Intelligence, vol 17, pp. 141–169, 2002

Costa, M., Pasero, E., Piglione, F, and Radasanu, D. *Short term load forecasting using a synchronously operated recurrent neural network*, Proceedings of the International Joint Conference on Neural Networks, 1999.

Coulibay, P., Anctil, F., and Rousselle, J., *Real-time short-term water inflows forecasting using recurrent neural networks*, Proceedings of the International Joint Conference on Neural Networks, 1999.

Cybenko, G. *Approximation by superposition of sigmoidal function*, Math. Cont. Sig. syst,2, 303-314,1989.

Elman, J. L., "*Finding structure in time*," Cognitive Science, vol. 14, pp. 179-211, 1990.

Friedman J., *Multivariate adaptive regression spline*. The Annals of Statistics. vol 19, pp. 1-50, 1991.

Gardner and S. Dorling, *Neural network modelling and prediction of hourly NOx and NO2 concentration in urban in London*, Atmospheric Environment, vol. 33, pp. 709-719, 1999.

Gardner, M.W. and Dorling, S.R, *Meteorologically adjusted trends in UK daily maximum ozone concentrations*, Atmospheric Environment, vol. 34, Issues 2, pages 171-176, 1999.

Ghazi Sabri , Khadir Mohamed Tarek, *Recurrent neural network for air pollution peaks prediction for the region of Annaba -Algeria-*, International Joint Conference: Intelligent Information Systems, Kraków, Poland, June 15-18, 2009.

Ghazi Sabri, Khadir Mohamed Tarek, *Recurrent Neural Network for Multi-steps ahead prediction of PM10*, Journal of Automation and systems engineering, vol 2, 2009

Ghazi Sabri, Khadir Mohamed Tarek, *Recurrent Neural Network for Multi-steps ahead prediction of PM10 concentration*, proceeding of international conference on electrical engineering design and technologies, Tunisia, 2008.

Giles, C. L., Lawrence, S., Tsoi, A.-C., *Rule inference for financial prediction using recurrent neural networks*, IEEE Conference on Computational Intelligence for Financial Engineering, IEEE Press, 253,1997.

Giorgio Corani, *Air quality prediction in Milan: feed-forward neural networks, pruned neural networks and lazy learning*, Ecological Modelling, Volume 185, Issues 2-4 , Pages 513-529, 10 July 2005.

Gitte B. Hedegaard, Jørgen Brandt, Jesper H. Christensen, Lise M. Frohn, Camilla Geels, Kaj M. Hansen and Martin Stendel, *Impacts of Climate Change on Air Pollution Levels in the Northern Hemisphere with Special Focus on Europe and the Arctic*, NATO Science for Peace and Security Series C: Environmental Security, Air Pollution Modeling and Its Application XIX, 2005

J. B. Ordieres, a, E. P. Vergaraa, R. S. Capuzb and R. E. Salazarc, *Neural network prediction model for fine particulate matter (PM2.5) on the US-Mexico border in ElPaso*, Environmental Modelling and Software, Volume 20, Issue 5, Pages 547-559, May 2005.

Jef Hooyberghs, Clemens Mensink, Gerwin Dumont, Frans Fierens and Olivier Brasseur, *A neural network forecast for daily average PM10 concentrations in Belgium*, Atmospheric Environment, Volume 39, Issue 18, Pages 3279-3289 , June 2005.

Jorquera Hector, Ricardo Prez, Aldo Cipriano, Andrs Espejo, M. Victoria Letelier and Gonzalo Acua, *Forecasting ozone daily maximum levels at Santiago, Chile*, Atmospheric Environment, Volume 32, Issue 20, Pages 3415-3424 , 25 September 1998.

Kim S. E., Kumar A., *Accounting seasonal nonstationarity in time series models for short-term ozone level forecast*. Stoch Environ Res Risk Assess, vol 19, 241–248, 2005.

Kolehmalnen M. , Martikainen H. , Hiltunen T. and Ruuskanen J. , *Forecasting air quality parameters using hybrid neural network modeling*, Environmental Monitoring and Assessment, vol 65, pp. 277–286, 2000.

Li, S., Wunsch II, D. C., O'Hair, E., and Giesselmann, M. G., *Wind turbine power estimation by neural networks with Kalman filter training on a SIMD parallel machine*, Proceedings of the International Joint Conference on Neural Networks, 1999.

M. Maruf Hossain, Md. Rafiul Hassan, and Michael Kirley, *Forecasting Urban Air Pollution Using HMM-Fuzzy Model*, T. Washio et al. (Eds.): PAKDD 2008, LNAI 5012, pp. 572–581, 2008.

Malkar P., Josef S., *Determination of Features for Air Pollution Forecasting Models*, 3rd ed. International Conference on Intelligent Information Systems (IIS '97) IEEE, 1997.

Mlakar P., and Boinar M., *Short-term air pollution prediction on the basis of artificial neural network*. 2nd International Conference on Air pollution, Barcelona, Air Pollution Volume 1, Computer Simulation Southampton, Boston, Computational Mechanics Publ, pp.545-552, 1994.

Monica Bianchini, Ernesto Di Iorio, Marco Maggini, Chiara Mocenni, Augusto Pucci, *A Cyclostationary Neural Network Model for the Prediction of the NO₂ Concentration*, ESANN'2006 proceedings European Symposium on Artificial Neural Networks Bruges (Belgium), d-side publi., ISBN 2-930307-06-4, 2006.

Nikhil, Bestamin Özkaya, Ari Visa, Chiu-Yue Lin, Jaakko A. Puhakka, and Olli Yli-Harja, *An Artificial Neural Network Based Model for Predicting H₂ Production Rates in a Sucrose-Based Bioreactor System*, proceedings of World Academy of Science, Engineering and Technology, volume 27,1307-1312, 2008.

Nikola K. Kasabov, *Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering*, MIT Perss.

Patricio Perez , Trier, A., Reyes, J, *Prediction of PM_{2.5} concentrations several hours in advance using neural networks in Santiago, Chile.*, Atmospheric Environment 34, 1189-1196,2000.

Perez P., Reyes J., *Prediction of maximum of 24-h average of PM₁₀ concentrations 30 h in advance in Santiago, Chile*, Atmospheric Environment 36 4555-4561, 2002.

Richard W. Boubel, *Fundamentals of air pollution*, Academic Press, ISBN 0-12-118930-9.

Rob Foxall, Igor Krcmar, Gavin Cawley, Steve Dorling, Damilo P.Mandic, *Nonlinear modelling of air pollution times series*, IEEE 0-7803-7041, 2001.

S.I.V. Sousaa, F.G. Martins, a, , M.C.M. Alvim-Ferraza and M.C. Pereira *“Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations”*, Environmental Modelling & Software, Volume 22, Issue 1, Pages 97-103, January 2007.

Saad, E. W., Caudell, T. P., and Wunsch II, D. C., *Predictive head tracking for virtual reality*, Proceedings of the International Joint Conference on Neural Networks, 1999.

Sharma P., Chandra A. Kaushik S., *Forecasts using Box–Jenkins models for the ambient air quality data of Delhi City*. Environ Monit Assess, vol 8, pp. 520-528, 2008.

Shuiyuan Cheng, Jianbing Li, Beng Feng, Yuquan Jin, · Ruixia Hao, *A gaussian-box modeling approach for urban air quality management in a northern chinese city—I.model development*, Water Air Soil Pollut, 178:37–57, DOI 10.1007/s11270-006-9120-3, 2006.

Stork, D. and Hassibi, B., *Second order derivatives for network pruning: Optimal Brain Surgeon*, In T. J. Sejnowski G. E. Hinton and D. S. Touretzky, editors *Advances in Neural Information Processing Systems (NIPS)*, 5, pages 164-171, 1993.

U. Brunelli¹, V. Piazza¹, L. Pignato¹, F. Sorbello², and S. Vitabile³ *Hourly Forecasting of SO₂ Pollutant Concentration Using an Elman Neural Network* (Eds.): WIRN/NAIS, LNCS 3931, pp. 65–69, 2006.

Y. İçağa & E. Sabah, *Statistical Analysis of Air Pollutants and Meteorological Parameters in Afyon, Turkey*, *Environ Model Assess*, DOI 10.1007/s10666-008-9139-5, 2007

Annexe A : Calcul de performance

Annexe A : Calcul de performance

Les performances sont calculé en terme de :

Index d'agrément IA:

$$IA = 1 - \frac{\overline{(C_p - C_0)^2}}{\overline{|C_p - \overline{C_0}|^2 + |C_0 - \overline{C_0}|^2}}$$

Erreur Moyenne Quadratique EMQ (MSE):

$$\sum_t^n \frac{(C_{pt} - C_{0t})^2}{n}$$

Où:

C_p : prédit, C_0 : mesuré (réel).