

UNIVERSITE AKLI MOHAND OULHADJ - BOUIRA
Faculté des Sciences et Sciences Appliquées



THESE DE DOCTORAT
En Chimie des Matériaux et de l'Environnement

Présentée par :

MAZOUZI CHOUAIB

THÈME

**Contribution à l'étude de la pollution atmosphérique de la
région algéroise par de nouvelles techniques de
modélisation**

Soutenue publiquement le :14/07/2022 devant le jury composé de:

Président	Mr. A. KETTAB	Professeur, U. Bouira
Examineur	Mr. N. NASRALLAH	Professeur, USTHB
Examineur	Mr. S. KADOUCHE	MCA, U. Tizi Ouzou
Examineur	Mr. R. KAOUA	MCA, U. Bouira
Examineur	Mr. S. KHEDIDJI	MCA, U. Bouira
Co-Directeur de thèse	Mme. Z. Bouhila Khodja	Chercheur Sénior, CRND/COMENA
Directeur de thèse	Mr. H. LOUNICI	Professeur, UAMO, Bouira

2021/2022

Université de Bouira. Rue Drissi Yahia, Bouira, Algérie

Dédicace



A ma mère en témoignage de mon affection et mon admiration, pour son courage et sa patience et à qui je ne trouverai jamais les mots pour qualifier cette étoile et à qui je ne saurai jamais exprimer toute ma reconnaissance...

A mon père qui m'a tout donné et qui a su avec de ses précieux conseils m'inspirer le sens de la dignité, le courage et le travail sans rebrousser chemin, qu'il trouve ici l'accomplissement de tous ses vœux...

A mes sœurs, Imene et Khaoula, à qui je souhaite tellement de succès et de réussites, bonne chance dans votre vie, que la vie vous sourira.

A tous les enseignants qui ont participé de près ou de loin à ma formation.

A tous mes amis et mes proches en témoignages de mes plus profondes amitiés et ma sincère sympathie.

A tous ceux qui ont contribué de près ou de loin à la réalisation de ce travail et pour que cette thèse voit la lumière.

A tous ceux qui me sont chers...

Mazouzi Chouaib

Remerciements

Dieu merci de m'avoir donné la force et la volonté d'accomplir ce mémoire. Merci de m'avoir aidé de manière irréprochable à affranchir les nombreux obstacles dressés dans mon chemin, merci de m'avoir donné la détermination quand tout semblaient perdus. Merci pour votre protection et pour les nombreux signes que vous m'avez envoyé.

Il me sera très difficile de remercier tout le monde car c'est grâce à l'aide de nombreuses personnes que j'ai pu mener cette thèse à son terme.

Je tiens à remercier vivement Madame Z. Khodja, qui en tant que promotrice au Centre de Recherche Nucléaire de Draria, d'avoir accepté d'encadrer ce travail, pour tous les précieux conseils qu'elle m'a donné, pour sa disponibilité tout au long de la réalisation de cette thèse et pour l'aide et le temps qu'elle a bien voulu me consacrer. Veuillez trouver ici ma profonde reconnaissance et mes profonds respects.

Je tiens également à remercier Professeur Mr H. Lounici, en tant que directeur de thèse, pour son oreille attentive qu'il m'a toujours accordé, pour sa confiance et de m'avoir orienté vers le chemin de la recherche scientifique, mille merci pour l'aide et le temps que vous avez bien voulu me consacrer. Merci pour la liberté que vous m'avez toujours accordé. Soyez assurés, cher professeur, de mon estime et de ma profonde gratitude.

Je voudrais exprimer mon entière reconnaissance au Professeur Monsieur A. Kettab de m'avoir consacré de son temps, pour l'honneur qu'il m'a fait en acceptant de présider mon jury de thèse, de juger notre travail et pour sa remarquable pédagogie.

Je tiens aussi à remercier vivement l'ensemble des membres de mon jury : Monsieur N. Nasrallah, Mr S. Kadouche, Mr R. Kaoua, Mr S. Khedidji pour l'honneur qu'ils m'ont fait en acceptant de juger notre travail.

Pour finir, je souhaite remercier ma famille qui m'a bien évidemment très fortement soutenu tout au long de ce travail, merci pour votre aide et votre soutien inconditionnel et indéfectible.

Merci à tous ceux qui m'ont aidé de quelque façon que ce soit, de près comme de loin.

Table des matières

Liste des abréviations	iii
Liste des figures	ii
Liste des tableaux.....	v
Introduction générale.....	vii

Chapitre I : La pollution atmosphérique

I.1. Définition	4
I.2. Composition de l'air	5
I.3. La pollution par les particules	5
I.4. Sources des particules fines.....	6
I.4.1. Particules de source naturelle	7
I.4.2. Particules de source artificielle.....	7
I.4.2.1. Les sources fixes	7
I.4.2.2. Les sources mobiles :	8
I.5. Les Eléments Traces (ET).....	9
I.5.1. Les impacts des Eléments Traces	11
I.5.1.2. Impact sur les animaux.....	11
I.5.1.3. Impact sur la santé.....	12
I.6. Méthode d'échantillonnage des TSP.....	13
I.6.1. Le choix du site de prélèvement	13
I.6.2. Le choix du filtre	13
I.6.2.1. Filtre en fibre de verre.....	14
I.6.2.2. Filtres membranaires.....	14
I.7. Normes et réglementation.....	14
I.8. Évaluation de la pollution de l'air en Algérie	15

Chapitre II : La modélisation

Introduction	16
II.1.L'analyse factorielle exploratoire (AEF).....	17
II.1.1. Présentation générale de l'AEF.....	17
II.1.2. Description sommaire de la technique	18
II.1.3. Particularité des matrices réduites.....	19
II.1.4. Nombre de facteurs communs à extraire.....	20

II.2. Le modèle ARIMA de Box-Jenkins.....	20
II.2.1. La différenciation	21
II.2.2. Auto-régression	23
II.2.3. Moyenne mobile.....	23
II.2.4. Les paramètres du modèle ARIMA	24
II.2.5. Termes AR	25
II.2.6. Termes MA	27
II.2.7. Racines unitaires	28
II.2.8. Evaluation des modèles	28
II.2.9. Modèles saisonniers	29
II.2.10. Séries chronologiques interrompue	30
II.3. Les réseaux de neurones artificiels	31
II.3.1. Historique	31
II.3.2. Modèle mathématique	32
II.3.2.1. Limites d'un neurone formel	34
II.3.3. Propriétés des réseaux de neurones	34
II.3.3.1. La non linéarité	35
II.3.4. Architecture des réseaux de neurones artificiels	35
II.3.4.1. Les réseaux monocouche : le perceptron	35
II.3.4.2. Les réseaux multicouches	36
II.3.5. Structure des réseaux de neurones artificiels	37
II.3.5.1. Les réseaux feed forward	37
II.3.5.2. Les réseaux récurrents : feedback	38
II.3.6. Apprentissage des réseaux de neurones artificiels	39
II.3.6.1. Apprentissage supervisé	39
II.3.6.2. Apprentissage non supervisé	40
II.3.6.3. Apprentissage semi-supervisé.....	41
II.3.7. Algorithmes d'apprentissage des RNA	41
II.3.7.1. L'apprentissage du perceptron.....	41
II.3.7.2. La propagation de l'erreur	42
II.3.7.4. La régularisation Bayésienne.....	45
II.3.8. Domaine d'application des réseaux de neurones artificiel.....	46
II.3.9. La mise en œuvre des réseaux de neurones artificiel	46
II.3.10. Avantages des RNA	47
II.4. La dispersion de la pollution : Modèle Gaussien	47

II.4.1. Le transport	48
II.4.2. La diffusion	48
II.4.3. Le dépôt.....	49
II.4.4. Description du modèle Gaussien.....	49
II.4.5. Détermination des paramètres du modèle	51
II.4.5.1. Méthode de Pasquill-Gifford	51
II.4.5.2. Méthode de Caraway	53
II.5. L'Arduino au service de l'environnement	54
II.5.1. Présentation des cartes Arduino	55
II.5.2. Principe du fonctionnement	56
II.5.3. Capteurs de type MQ.....	56
Chapitre III : Méthodologie expérimentale	
III.1. Description de la zone d'étude	59
III.2. Echantillonnage des TSP	60
III.3. Echantillonnage des polluants gazeux.....	60
III.3.1. Capteurs utilisés	61
III.3.1.1. Capteur MQ ₂	61
III.3.1.2. Capteur MQ ₃	63
III.3.1.3. Capteur MQ ₇	63
III.3.1.4. Capteur MQ ₁₃₅	64
III.3.1.5. Capteur DHT ₁₁	64
III.3.2. Montage de la station de prélèvement	65
III.3.3. Algorithme de la station de prélèvement	67
III.4. Analyse des TSP et ET associés	68
III.4.1. Analyse des TSP	68
III.4.2. Analyse des ET associés aux TSP	69
III.5. Traitement des données brutes	70
III.5.1. Intérêt d'une série temporelle	70
III.5.2. Détection des valeurs singulières.....	71
III.6. Modélisation des TSP et ET associés par le modèle ARIMA	72
III.7. Modélisation des TSP et ET associés par les réseaux de neurones artificiels	73
III.7.1. Détermination des entrées du modèle	73
III.7.2. Stratégie adoptée	74
III.8. Critères de validation des modèles obtenus.....	76
III.9. La prévision des TSP et ET associés.....	76

III.9.1. Prévision par le modèle ARIMA	77
III.9.2. Prévision des TSP par les réseaux de neurones	77
III.9.3. Prévision des ET par les réseaux de neurones	79
III.10. Critères de validation de la prévision obtenue	80
III.11. La cartographie de la pollution	80

Chapitre IV : Résultats et discussions

IV.1. Etude de la pollution par les gaz	81
IV.1.1. La station de prélèvement	81
IV.1.2. Analyse de la pollution atmosphérique par les gaz	83
IV.1.3. Répartition journalière	91
IV.1.4. Répartition hebdomadaire	92
IV.1.5. Répartition mensuelle	94
IV.2. Etude de la pollution par les TSP et ET associés	95
IV.3. Modélisation de la pollution par la méthode ARIMA	98
IV.4. Modélisation de la pollution par les réseaux de neurones artificiels	104
IV.5. Comparaison entre les deux méthodes de modélisation	105
IV.6. Prévision de la pollution atmosphérique par la méthode ARIMA	107
IV.7. Prévision de la pollution atmosphérique par les réseaux de neurones artificiels	113
IV.8. Comparaison entre les deux méthodes de prévision	123
IV.9. La cartographie de la pollution de l'air par les TSP	130
Conclusion générale	135
Références bibliographiques	137
Annexes	

*L*iste des abréviations

ACP	Analyse en composantes principales
AEF	Analyse factorielle exploratoire
ANOVA	Analyse of variance
ARIMA	Auto-Regressive Integrated Moving Average
BR	Régularisation bayésienne
C_e	Concentration de l'élément dans le standard ($\mu\text{g/g}$ de standard).
C_s	Concentration de l'élément dans l'échantillon (ng/m^3).
Cste	Constante
Distrib.	Distribution
ET	Eléments traces
FAC	Factorial analysis correlogram
FACP	Factorial analysis correlogram partial
J	Jour
LM	Algorithme de Levenberg-Marquardt
Max	Maximum.
Min	Minimum.
MQx	Capteur de type Message Queuing
ms	Masse du standard (mg).
MSE	Mean Squar Error
NET $_e$	Net de l'élément dans l'échantillon.
NET $_s$	Net de l'élément dans le standard.
PM $_x$	Périmètre moyen de x micromètre
r	Coefficient de corrélation
RDM	Relative distance to minimum
RMSE	Root Mean Squar Error
RNA	Réseaux de neurones artificiels
tde	Temps de décroissance de l'élément dans l'échantillon (s).
tds	Temps de décroissance de l'élément dans l'échantillon (s).
TSP	Particules totales en suspension
V	Volume d'air aspiré (m^3).
Vari.	Variance
y^*	Valeur calculée ou modélisée
y_i	Valeur observée

Liste des figures

Figure I.1: Agrandissement des PM_{10} et PM_2	5
Figure I.2: Comparaison entre PM_2 , grain de sable et un cheveu humain.....	11
Figure I.3: Pénétration des PM_2 dans le système respiratoire	12
Figure II.1: Modélisation 3D de la qualité de l'air à l'intérieur d'une chambre.....	16
Figure II.2: Algorithme du modèle ARIMA.....	22
Figure II.3: Neurone formel de McCulloch et Pitts.....	32
Figure II.4: La correspondance entre le neurone biologique et le neurone artificiel.....	33
Figure II.5: Fonctions de transfert d'un neurone artificiel	34
Figure II.6: Réseau de neurones monocouche : le perceptron	36
Figure II.7: Réseau de neurones multicouches.....	37
Figure II.8: Réseau de neurone feedforward	38
Figure II.9: Réseau de neurone récurrent.....	39
Figure II.10: Principe de la retro-propagation de l'erreur.....	43
Figure II.11: Comparaison entre l'apprentissage avec l'algorithme de LM seul et avec l'algorithme LM + régularisation bayésienne	45
Figure II.12: Système de coordonnées du modèle gaussien	50
Figure II.13: coefficient de diffusion vertical ∂_z de Pasquill-Gifford en fonction de la distance x.....	52
Figure II.14: coefficient de diffusion vertical ∂_y de Pasquill-Gifford en fonction de la distance x.....	52
Figure II.15: Carte Arduino UNO R3	55
Figure II.16: La nomenclature des broches (pin) d'une carte Arduino du type UNO R3.....	56
Figure III.1: Capteur de type MQ	57
Figure III.2: schéma électrique du capteur de type MQ	57
Figure III.3: Situation géographique du site de prélèvement.....	59
Figure III.4: Capteur mq2	62
Figure III.5: Broches de connexion d'un capteur type MQ.....	62
Figure III.6: Élément de détection tubulaire d'un capteur type MQ.....	63
Figure III.7: vue intérieure de la bobine de détection.....	63
Figure III.8: Capteur dht11	64
Figure III.9: Capteur DHT ₁₁ décortiqué.....	64
Figure III.10: Montage du circuit électrique de la station de prélèvement des gaz polluants	66
Figure III.11: Circuit imprimé à double couche de la station de prélèvement. Circuit jaune : face recto, circuit orangé : face verso.	66
Figure III.12: Algorithme de la station de prélèvement des gaz polluants.....	68
Figure III.13: simulation d'écoulement d'air autour du système d'échantillonnage.....	69
Figure III.14: Schéma de réseau de neurones pour la modélisation des TSP et ET associés.....	74
Figure III.15: Algorithme de la modélisation par les réseaux de neurones artificiels	75
Figure III.16: Schéma de la méthode itérative par les réseaux de neurones pour la prévision	78
Figure III.17: algorithme de la prévision des TSP et ET associés par les réseaux de neurones artificiels.....	79
Figure III.18: Schéma du réseau de neurone combiné pour la prévision des TSP et ET associés.....	80

Figure IV.1: Circuit imprimé de la station montée sur la carte Arduino Uno.....	81
Figure IV.2: simulation de profil de vitesse et la direction de l'air autour des capteurs de prélèvement.....	82
Figure IV.3: Emplacement des capteurs dans le support	82
Figure IV.4: rapport R_s/R_0 enregistrés par les capteurs en fonction du temps du prélèvement.....	83
Figure IV.5: Concentration moyenne en ppm des polluants gazeux analysés au niveau de notre site d'étude	85
Figure IV.6: variation de la concentration du toluène, Alcool, NH_4 , CO_2 durant la période d'échantillonnage	86
Figure IV.7: Variation de la concentration du propane et GPL durant la période d'échantillonnage	87
Figure IV.8: Variation de la concentration de l'acétone, CO et H_2 durant la période d'échantillonnage.....	87
Figure IV.9: Variation de la concentration du benzène durant la période d'échantillonnage	88
Figure IV.13: Rose des vents enregistrée au niveau de la station de Dar El Beida durant la période d'étude ...	89
Figure IV.11: Variation de la température, pression atmosphérique et la vitesse du vent durant la période d'étude	90
Figure IV.12: Variation de la température, la précipitation et les rafales du vent durant la période d'étude....	90
Figure IV.13: Répartition journalière de la concentration en ppm des polluants gazeux analysés au niveau du site d'étude	91
Figure IV.14: répartition hebdomadaire de la concentration en PPM des polluants gazeux analysés au niveau du site d'étude.....	93
Figure IV.15: Répartition mensuelle de la concentration en ppm des polluants gazeux analysés au niveau du site d'étude	94
Figure IV.16: Eléments trace associés aux TSP (ng/m^3), déterminés par INAA dans les échantillons d'air de la zone d'étude.....	96
Figure IV.17: Modélisation des TSP par le modèle arima.....	98
Figure IV.18: Modélisation du Scandium (Sc) par le modèle ARIMA.....	99
Figure IV.19: Modélisation d'Ytterbium (Yb) par le modèle ARIMA.....	99
Figure IV.20: Modélisation du Brome (Br) par le modèle ARIMA	100
Figure IV.21: Modélisation du Sodium (Na) par le modèle ARIMA	100
Figure IV.22: Modélisation de l'Antimoine (Sb) par le modèle ARIMA.....	101
Figure IV.23: Modélisation du Hafnium (Hf) par le modèle ARIMA.....	101
Figure IV.24: Modélisation du Chrome (Cr) par le modèle ARIMA.....	102
Figure IV.25: Modélisation du Samarium (Sm) par le modèle ARIMA	102
Figure IV.26: Modélisation du Césium (Ce) par le modèle ARIMA.....	103
Figure IV.27: Modélisation du Lanthane (La) par le modèle ARIMA.....	103
Figure IV.28: Modélisation des TSP par les RNA	100
Figure IV.29: Modélisation du Sodium (Na) par les réseaux de neurones artificiels.....	100
Figure IV.30: Modélisation du Scandium (Sc) par les réseaux de neurones artificiels.....	101
Figure IV.31: Modélisation d'Ytterbium (Yb) par les réseaux de neurones artificiels	101
Figure IV.32: Modélisation du Brome (Br) par les réseaux de neurones artificiels.....	102
Figure IV.33: Modélisation du Samarium (Sm) par les réseaux de neurones artificiels.....	102
Figure IV.34: Modélisation du Hafnium (Hf) par les réseaux de neurones artificiels	103
Figure IV.35: Modélisation de l'Antimoine (Sb) par les réseaux de neurones artificiels.....	103
Figure IV.36: Modélisation du Cérium (Ce) par les réseaux de neurones artificiels	104
Figure IV.37: Modélisation du Chrome (Cr) par les réseaux de neurones artificiels.....	104
Figure IV.38: Modélisation du Lanthane (La) par les réseaux de neurones artificiels.....	105
Figure IV.39: Comparaison du coefficient de la régression r^2 des deux méthodes de modélisations.....	106
Figure IV.40: Comparaison du RMSE des deux méthodes de modélisations.....	106
Figure IV.41: Prévision des TSP par le modèle arima	108
Figure IV.42: Prévision du Sodium(Na) par le modèle ARIMA	108

Figure IV.43: Prévission du Samarium (Sm) par le modèle ARIMA.....	109
Figure IV.44: Prévission du Scandium (Sc) par le modèle ARIMA.....	109
Figure IV.45: Prévission d'Ytterbium (Yb) par le modèle ARIMA.....	110
Figure IV.46: Prévission du Césium (Ce) par le modèle ARIMA.....	110
Figure IV.47: Prévission du Chrome (Cr) par le modèle ARIMA.....	111
Figure IV.48: Prévission du Hafnium (Hf) par le modèle ARIMA.....	111
Figure IV.49: Prévission du Lanthane (La) par le modèle ARIMA.....	112
Figure IV.50: Prévission de l'Antimoine(Sb) par le modèle ARIMA.....	112
Figure IV.51: Prévission du Brome (Br) par le modèle ARIMA.....	113
Figure IV.52: Prévission des TSP par les réseaux de neurones artificiels.....	117
Figure IV.53: Prévission du Brome (Br) par les réseaux de neurones artificiels.....	117
Figure IV.54: Prévission du Césium (Ce) par les réseaux de neurones artificiels.....	118
Figure IV.55: Prévission du Chrome (Cr) par les réseaux de neurones artificiels.....	118
Figure IV.56: Prévission du Hafnium (Hf) par les réseaux de neurones artificiels.....	119
Figure IV. 57: Prévission du Lanthane (La) par les réseaux de neurones artificiels.....	119
Figure IV.58: Prévission du Sodium (Na) par les réseaux de neurones artificiels.....	120
Figure IV.59: Prévission de l'Antimoine (Sb) par les réseaux de neurones artificiels.....	120
Figure IV.60: Prévission du (Sc) par les réseaux de neurones artificiels.....	121
Figure IV.61: Prévission du Samarium (Sm) par les réseaux de neurones artificiels.....	121
Figure IV.62: Prévission d'Ytterbium (Yb) par les réseaux de neurones artificiels.....	122
Figure IV.63: comparaison de la prévission des TSP par les deux méthodes de prévission.....	123
Figure IV.64: Comparaison de la prévission du Chrome (Cr) par les deux méthodes de prévission.....	123
Figure IV.65: Comparaison de la prévission du Hafnium (Hf) par les deux méthodes de prévission.....	124
Figure IV.66: Comparaison de la prévission de l'Antimoine (Sb) par les deux méthodes de prévission.....	124
Figure IV.67: Comparaison de la prévission du Scandium (Sc) par les deux méthodes de prévission.....	125
Figure IV.68: Comparaison de la prévission du Samarium (Sm) par les deux méthodes de prévission.....	125
Figure IV.69: Comparaison de la prévission d'Ytterbium (Yb) par les deux méthodes de prévission.....	126
Figure IV.70: Comparaison de la prévission d Césium (Ce) par les deux méthodes de prévission.....	126
Figure IV.71: Comparaison de la prévission du Lanthane (La) par les deux méthodes de prévission.....	127
Figure IV.72: Comparaison de la prévission du Sodium (Na) par les deux méthodes de prévission.....	127
Figure IV.73: Comparaison de la prévission du Brome (Br) par les deux méthodes de prévission.....	128
Figure IV.74: Comparaison du r^2 de la prévission des TSP et ET associes par les deux méthodes de prévission.....	129
Figure IV.75: Comparaison du RMSE de la prévission des TSP et ET associés par les deux méthodes de prévission.....	129
Figure IV.76: Rose des vents enregistre au niveau de la station de Dar El Beida durant la période d'étude.	130
Figure IV.77: Exemple du profil vertical et horizontal de la dispersion TSP par le modèle plume.....	131
Figure IV.80: Schéma représentatif de la méthode de l'élaboration de la cartographie de la pollution par les TSP.....	132
Figure IV.79: Cartographie de la pollution de l'air par les tsp au niveau de notre zone d'étude.....	132
Figure IV.80: cartographie de la pollution de l'air par les TSP au niveau de notre zone d'étude avec géolocalisation.....	133

Liste des tableaux

Tableau I.1: Taille, origine, composition et source des particules	6
Tableau I.2: Les organes affectés par les métaux lourds	12
Tableau I.3: Caractéristique des filtres membranaire	14
Tableau II.1: La correspondance entre le niveau de la stabilité, classe et description.....	51
Tableau II.2: La classe de la stabilité en fonction de la vitesse du vent	51
Tableau II.3: Coefficients du modèle de Caraway pour la détermination du ∂_z	53
Tableau II.4: Coefficients du modèle de Caraway pour la détermination du ∂_y	53
Tableau IV.1: Données statistique de la concentration des polluants gazeux détectés au niveau de notre zone d'étude en ppm.....	85
Tableau IV.2: Teneurs en mg/l du benzène et toluène dans certaines villes	86
Tableau IV.3: Matrice de corrélation de Pearson des gaz polluants analysés au niveau de notre site d'étude	88
Tableau IV.4: Répartition journalière de la concentration en ppm des polluants gazeux analysés au niveau du site d'étude	91
Tableau IV.5: répartition hebdomadaire de la concentration en ppm des polluants gazeux analysés au niveau du site d'étude.....	93
Tableau IV.6: Répartition mensuelle de la concentration en ppm des polluants gazeux analysés au niveau du site d'étude	94
Tableau IV.7: La concentration de polluants à modéliser détectés au niveau de notre zone d'étude, TSP en mg/m ³ et les ET en ng/m ³	95
Tableau IV.8: Matrice de corrélation de Spearman des concentrations des éléments trace associés aux TSP, déterminées par INAA.....	97
Tableau IV.9: Tableau récapitulatif de la modélisation par le modèle ARIMA	98
Tableau IV.10: Variance totales expliquée	104
Tableau IV.11: Matrice des composantes principales.....	104
Tableau IV.12: Tableau récapitulatif de la modélisation des TSP et les ET associés par les réseaux de neurones artificiels	99
Tableau IV.13: comparaison de la modélisation par les deux méthodes.....	105
Tableau IV.14: Tableau récapitulatif de la prévision des TSP et les ET associés par le modèle ARIMA	107
Tableau IV.15: Tableau récapitulatif de la prévision des TSP et des ET associés par les réseaux de neurones artificiels	114
Tableau IV.16: comparaison du résultat de la prévision par les deux méthodes de prévisions	128

*I*ntroduction générale

*I*ntroduction générale

Depuis les origines du monde, la décomposition des matières végétales et des débris animaux, les incendies de forêts et l'érosion des roches étaient précurseurs de la pollution de l'air. Au 12^{ème} siècle, à Londres, les fumées produites par la combustion du charbon et du bois étaient devenues insupportables, le roi de l'Angleterre Edward II ordonna de punir la personne prise en train de souiller l'air avec des fumées de charbon (*Wilson et al. 1996*).

Avec l'avènement de l'ère industrielle et les avancées technologiques, la demande croissante en énergie a imposé l'utilisation du charbon comme principale source d'énergie, par conséquent la pollution de l'air a pris de l'ampleur (*Kaygusuz, 2012*). Cependant, la combustion du charbon génère principalement du NO_x, du SO₂, du mercure et de la poussière, etc. (*Sui et al. 2018 ; Wang et al. 2014*) qui est l'une des principales causes de catastrophes environnementales telles que les pluies acides, la brume et le smog photochimique (*Atkinson et al. 1984*). Au cours des trente dernières années écoulées, la notion de la pollution atmosphérique s'est largement complexifiée, avec une population atteignant plus de six milliards d'individus sur terre et une consommation croissante de ressources naturelles et énergétique, le phénomène de pollution de l'air a pris aujourd'hui une ampleur sans précédent (*Yin and Harrison, 2008 ; Nguyen and Byeong-Kyu, 2010*). Mise à part cette large vague d'industrialisation, l'intérêt croissant porté aux divers aspects de la pollution, s'explique aussi par les événements dramatiques survenus lors des catastrophes de Donora (Pennsylvania) en 1948 et Londres en 1952 (*Nemery, et al. 2001*) et qui ont causé la mort de milliers de personnes (*Bell and Davis, 2001 ; National Public Radio, 2002*).

La pollution de l'air intéresse aujourd'hui nombreux chercheurs dont les travaux relèvent de divers domaines, cette pollution due aux rejets anthropiques incontrôlés entraîne des perturbations au niveau des êtres vivants et des compartiments abiotiques fondamentaux des milieux (*Bi et al. 2020 ; Brauer et al. 2016 ; Chen et al. 2020 ; Chan et al. 2017 ; Cohen et al. 2017 ; Fuertes et al. 2020*). Elle est aussi à l'origine d'effets néfastes qui affectent le plan environnemental, sanitaire, économique et politique (*Chan et al. 2019 ; Liu et al. 2019 ; Seow et al. 2014 ; Zhang et al. 2021*). En Algérie, dans ce contexte de récents travaux ont montré que le Grand Alger fait objet d'une intense pollution atmosphérique (*Oucher 2006 ; Bitouche 2008*). Plusieurs études effectuées dans divers régions du monde comme en Algérie ont montré que le trafic routier constitue l'une des principales causes de la pollution de l'air ambiant (*Aghaei et al. 2016 ; Boughedaoui et al. 2004 ; Bouhila et al. 2015 ; Kerbachi et al. 1998, 2006 ; Martins et al. 2007 ; Skov et al. 2001*).

Dans le cadre de notre étude on s'intéressera particulièrement à la pollution de l'air par les particules totales en suspension (TSP), les éléments traces (ET) associées et quelques gaz polluants parmi les plus dominants. Il est clair de nos jours que les particules fines ont une relation avec la présence des métaux toxiques dans l'air particulièrement en milieu urbain et à proximité des zones industrielles (*Boughedaoui et al. 2004*), ces polluants très dangereux de

part leurs stabilités, constituent un risque sanitaire important aussi bien pour l'homme que les divers écosystèmes de son environnement (*Chen et al. 2020*).

A cet effet, des initiatives visant à prévoir la détérioration de la qualité de l'air ont été prises. Tout d'abord, les chercheurs ont développés des méthodes d'analyses physicochimiques se basant sur une analyse directe des polluants émis par le biais des capteurs physicochimiques. Cette méthode, bien qu'elle fournisse des données quantifiées, c'est vite avérée : onéreuse sur le plan économique, et non indicatrice sur les changements aux niveaux biologiques sur le plan environnemental.

Dans cette étude nous avons modélisé la pollution atmosphérique par les TSP et les ET associés au niveau de Draria afin de créer des modèles exploitables qui aideront à la prise de décisions rapides, l'étude qualitative et quantitative de cette pollution a fait l'objet d'une étude antérieure (*Bouhila et al. 2015*). La présente étude vise les objectifs suivants :

- Elaborer pour chaque élément trace un modèle capable de modéliser les ET à partir des TSP uniquement ;
- Prévoir la pollution par les TSP et les ET associés au niveau de notre zone d'étude sur un horizon de quatre jours ;
- Etablir une cartographie de la pollution par les TSP au niveau de notre zone d'étude sur un diamètre d'un kilomètre ;
- Echantillonnage et analyse de la pollution de l'air par les gaz polluants.

De ce fait, nous avons envisagé une modélisation non conventionnelle préventive en faisant appel à une technique qui ne nécessite pas la connaissance explicite des lois physiques liées aux phénomènes à expliquer mais qui sollicite des données numériques. L'intelligence artificielle par les réseaux de neurones artificiels, utilisée avec succès dans la modélisation des séries chronologiques, offre l'avantage d'être un outil non linéaire, multi-variable, paramétrable par apprentissage sur des données expérimentales représentatives, et dans lequel il est possible d'inclure tout ou une partie de la connaissance physique disponible (*Viotti et al., 2002*).

Pour cela, il est nécessaire d'envisager l'utilisation des modèles récurrents ou bouclés dont le retard sera équivalent à une période d'échantillonnage, nous avons donc développé un modèle sophistiqué qui associe autant de réseaux que d'éléments, chacun d'eux ayant été élaboré pour estimer un élément bien particulier.

Le résultat final de ce travail nous permettra de :

- Estimation de la pollution atmosphérique par les TSP et les ET associés sans les analyser ;
- Gagner en temps opératoire (échantillonnage et analyse) ;
- Minimiser les coûts d'analyses ;
- Création des scénarios prévisionnistes ;
- Réfléchir aux mesures à prendre et à la formulation d'une stratégie de prévention de la pollution de l'air.

Chapitre I

La pollution atmosphérique

« Quand ils auront coupé le dernier arbre, pollué le dernier ruisseau, pêché le dernier poisson. Alors ils s'apercevront que l'argent ne se mange pas. »

(Sitting Bull 1831- 1890)



Chapitre I : La pollution atmosphérique

Introduction

La pollution de l'air est liée directement aux effets néfastes qu'elle engendre. Sa perception évolue en fonction des progrès réalisés en recherche physico-chimique mais surtout en toxicologie et épidémiologie.

La notion de pollution atmosphérique a connu de profondes mutations, elle évolue et change d'échelle ; elle intègre des phénomènes complexes et interdépendants allant de l'échelle globale à l'échelle locale, du changement climatique planétaire aux modifications dans la biodiversité.

Les enjeux de la pollution atmosphérique sont devenus tels que les pouvoirs publics au plus haut niveau, les chercheurs et la société civile en font une de leurs principales préoccupations (**Blanchard, 2004**). Des conventions internationales de réduction de la pollution et une réglementation de plus en plus sévère et restrictive sont le résultat de ces préoccupations.

Malgré toutes les mesures prises, la consommation croissante de ressources naturelles et surtout énergétiques a fait que la qualité de l'air ne cesse de se dégrader, et ce en particulier dans les pays peu développés.

Cette pollution atmosphérique se caractérise par des émissions primaires qui sont soit d'origine naturelle, soit anthropogéniques c'est à dire produites par l'homme. A ces émissions primaires viennent s'ajouter les émissions secondaires qui résultent de la transformation dans l'atmosphère de polluants primaires en polluants secondaires dont le potentiel de nuisance peut être encore plus élevé.

I.1. Définition

La pollution de l'air a été définie en 1949 selon American Medical Association comme étant : «*La concentration excessive de produits étrangers aux biens*» (**Yamani, 2006**). Cette définition a évolué en 1956 vers la définition suivante (**Oregon, 1956**) : «*présence dans l'atmosphère extérieure de substances ou contaminants, apportés par l'homme, en quantités ou à des concentrations et pendant des périodes telles qu'ils gênent une proportion importante des habitants d'un secteur, ou nuisent à la santé publique, à la vie humaine, végétale ou animale, ou aux biens, ou portent atteinte à l'agrément de l'existence ou à la jouissance des biens dans l'état, ou dans les provinces ou les zones de l'état qui sont touchés*».

Le Conseil de l'Europe dans sa déclaration de mars 1968 propose cette définition : «*Il y a pollution atmosphérique lorsque la présence d'une substance étrangère ou une variation importante dans la proportion de ses composants est susceptible de provoquer un effet nocif, compte tenu des connaissances scientifiques du moment, ou de créer ou une nuisance ou une gêne.*»

Selon la loi sur l'air et l'utilisation rationnelle de l'énergie de 1996, la pollution atmosphérique est définie comme : «*l'introduction par l'homme, directement ou indirectement, dans l'atmosphère et les espaces clos, de substances ayant des conséquences*

préjudiciables de nature à mettre en danger la santé humaine, à nuire aux ressources biologiques et aux écosystèmes, à influencer sur les changements climatiques, à détériorer les biens matériels, à provoquer des nuisances olfactives”.

En corrélation à tous ces exposés, la pollution de l’air présentement est définie comme « La présence de toutes substances étrangères ou modifications dans la concentration des différents composants de l’air naturel pur ».

1.2. Composition de l’air

L’homme inhale quotidiennement environ 15,000 litres d’air. Cet air est composé majoritairement d’azote (78%) et d’oxygène (21%) accompagné d’un peu d’argon (0.9%) et de dioxyde de Carbone (0.035%). Le solde est constitué de gaz rares (Hélium, Krypton, Xénon) et d’hydrogène. L’air typique de l’environnement terrestre est souvent humide car il contient de la vapeur d’eau. Il peut aussi contenir du dioxyde de soufre, des oxydes d’azote, de fines substances en suspension sous forme d’aérosol, des poussières et des micro-organismes.

Les émissions de polluants dans l’atmosphère peuvent être divisées en deux grandes catégories (*Yamani, 2006*) : les sources anthropiques liées directement à l’activité humaine (transports, industries, chauffage des bâtiments, incinération des déchets, etc.) et les sources naturelles (émissions volcaniques, plantes produisant le pollen, foudre, etc.) qui existent depuis la création de la terre. Leurs émissions sur la surface de la terre à de faibles concentrations, et du fait des conditions météorologiques, les dispersent dans l’atmosphère et finissent par disparaître sans causer de dégâts. Pour cela, la pollution anthropique sera essentiellement traitée dans ce qui suit (*Yamani, 2006*).

1.3. La pollution par les particules

Les aérosols particulaires se caractérisent par la suspension dans l’air de fines particules solides ou liquides, dont la dimension est de l’ordre de 10 μm au plus dans l’air au repos, et de quelques centaines de microns dans l’air en mouvement (tempête, orages, ...) (*Herrebaut et al., 1976*).

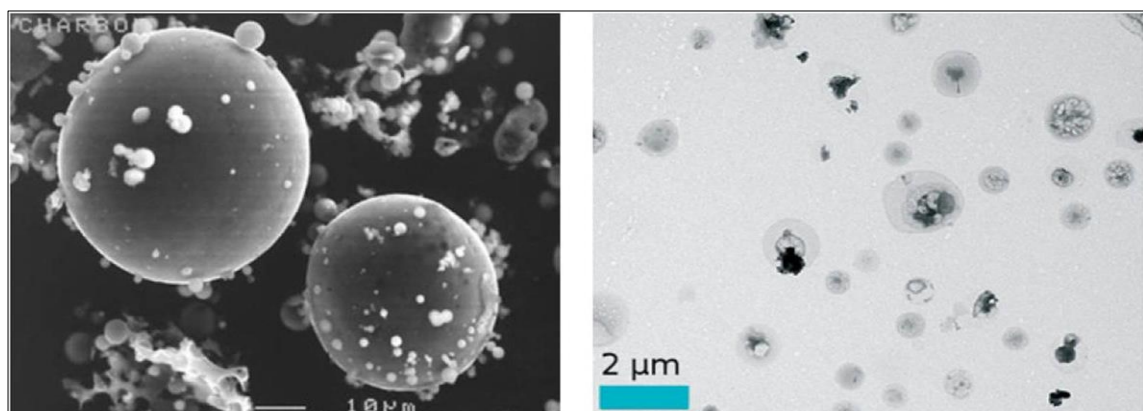


Figure I.1: Agrandissement des PM_{10} et PM_2
(*Jourdain, 2007*)

L'aérosol de pollution se définit par sa nature chimique, sa forme, sa densité, sa granulométrie et sa concentration en nombre (particules par m³ d'air) ou en poids (milligrammes ou microgrammes par m³) (*Herrebaut et al., 1976*). La taille des particules est déterminante quant aux effets sur la santé car leur cheminement dans l'appareil respiratoire dépend de leur granulométrie (*Priest, 1973 ; Ontario, 2000 ; Magdelaine, 2001 ; Detrie, 1969*).

Dans la littérature, et pour traduire la grande diversité des particules dans leur origine, leur composition et leur comportement aérodynamique, on utilise différents termes pour qualifier la pollution particulaire. Les termes les plus employés sont (*Renoux, 1990 ; Stern et al., 1984*) : aérosols, fumés noirs, suies, poussières, brumes, cendres volantes et particules.

I.4. Sources des particules fines

Les particules fines proviennent de certains procédés industriels, des gaz d'échappement des véhicules et du chauffage au bois, les sources naturelles comme l'érosion éolienne des sols, les feux de forêts, les embruns océaniques et les éruptions volcaniques sont aussi des sources non négligeables. De plus, les particules les plus fines sont surtout des particules secondaires résultant de réactions chimiques de gaz et de particules déjà présentes dans l'atmosphère.

L'atteinte de certains objectifs dans l'air ambiant pourrait exiger des programmes de contrôle très variés car les sources sont nombreuses autant naturelles qu'anthropiques.

Le tableau I.1 résume l'origine et la composition des particules (*Mohamedou et al., 2003*).

Tableau I.1: Taille, origine, composition et source des particules

	Particules de diamètre : 2,5-10 µm	Particules de diamètre : < 2,5 µm
Origine	<ul style="list-style-type: none"> - Procédés mécaniques (broyage, meulage). - Évaporation d'aérosol. - Recirculation de poussières. - Combustion incomplète. 	<ul style="list-style-type: none"> - Condensation de vapeur et de gaz. - Coagulation de très petites particules. - Réactions chimiques diverses. - Nucléation.
Composition chimique	<ul style="list-style-type: none"> - Cristaux de silice, de fer, de sel. - Cendres volatiles, Pollens, Fibres. - Débris végétaux et animaux. 	<ul style="list-style-type: none"> - Sulfates, Nitrates, Ammoniac. - Hydrocarbures aromatiques polycycliques.
Sources	Transport, agriculture, mines, construction, démolition, combustion de charbon et huile, biotechnologie, océans	<ul style="list-style-type: none"> - Combustion (industries, transport). - Transformation atmosphérique de précurseurs (NO_x, SO₂). - Procédés industriels à très haute température.(métallurgie, incinération)

I.4.1. Particules de source naturelle

Les particules naturelles comprennent les poussières du sol et les particules minérales soulevées par le vent, la poussière volcanique, les embruns salins et les matières biologiques comme le pollen, les spores, les bactéries et les débris provenant des feux de forêt. En général, les sources naturelles produisent des particules grossières mais, sous l'effet de températures très élevées (dans le cas de feux de forêt, par exemple) certaines d'entre elles produisent des particules fines (*Masclat, 2005*).

Les quatre principales sources naturelles sont (*Masclat, 2005*): source terrigène, source marine, source volcanique et a source biogénique.

I.4.2. Particules de source artificielle

Les sources artificielles liées à toutes les activités humaines produisent à la fois (*Kerbachi, 2005*) :

- **Particules fines** : issues des processus de combustion et de diverses activités industrielles.
- **Particules grossières** : issues des champs agricoles, des routes, des chantiers de construction, ...

Les sources artificielles peuvent être classées d'après plusieurs critères selon :

- qu'elles résultent de phénomènes de combustion ou de mécanismes plus spécifiques.
- qu'elles soient ponctuelles ou diffuses.
- qu'elles soient fixes ou mobiles.

I.4.2.1. Les sources fixes

Tous les processus de combustion et la plupart des branches industrielles sont génératrices de particules dans l'air.

L'industrie thermo énergétique: La combustion induit dans l'atmosphère plusieurs variétés minérales, des métaux et des particules carbonées (suies), ainsi que des gaz qui se transforment dans l'atmosphère en particules tels que (les NO_x et le SO_2) (*Dang et al., 1990*).

L'industrie sidérurgique: Les principaux polluants issus de ce domaine sont des poussières de composition chimique variable et de granulométrie entre 2 et 100 μm (SiO_2 , AlO_3 , CaO_2 , K_2O , ZnO ,...) et des particules fines entre autre les fumées d'oxyde de fer. Ces émissions sont le résultat de différents processus de production à savoir : l'agglomération et « frittage » des minerais, la cokéfaction, la fusion dans les hauts fourneaux, le chauffage, le transport, etc (*Kerbachi, 2005*).

L'industrie des matériaux non ferreux: Elle participe à l'émission d'importantes quantités de Pb, Zn, Cu, Cd, Ni, Hg,... Lors du raffinage de ces métaux au cours des opérations de fusion, des vapeurs métalliques se répandent dans l'air, s'oxydent et se subliment en fines particules, qui sont entraînées sur de longues distances par la circulation de l'air (*Kerbachi, 2005*).

L'industrie des matériaux de construction: La production des matériaux de construction contribue fortement à la pollution particulaire à cause des procédés d'extraction, de transport, de broyage et de fusion des minerais. Les cimenteries sont des grandes émettrices de poussières, du fait de l'importance des quantités de ciment produites et des roches utilisées pour sa production (*Kerbachi, 2005*).

L'industrie chimique: ne présente pas une importante source de poussière dans l'air, mais participe d'une manière indirecte dans la pollution par l'émission des gaz qui se transforment en particules. Certains métaux (Cd) et certains hydrocarbures aromatiques polycycliques sont produits dans l'air par les raffineries de pétrole (*Brumer, 1998*).

L'incinération des déchets: engendre l'émission de grandes quantités de particules. Ces particules peuvent contenir d'importantes quantités de métaux (Fe, Zn, Cd, Pb, Hg) et des oxydes métalliques divers dont la toxicité est considérable. En plus de ces particules, on trouve les gaz qui dans l'air se transforment en particules (SO₂) (*Magdelaine, 2002*).

I.4.2.2. Les sources mobiles :

Le parc automobile, dont le volume croît continuellement, constitue en milieu urbain la principale source d'émission de particules (*Brumer, 1998*). La pollution issue du trafic routier comprend les émissions suivantes :

Les particules routières: Le trafic automobile émet surtout des halogénures (bromures et chlorures) de plomb, volatils à la température du gaz à l'échappement, mais qui ensuite, par hydrolyse de ces sels, engendrent dans l'atmosphère des acides chlorhydriques (HCl) et bromhydriques (HBr) qui ne restent pas en général sous forme gazeuse mais se retrouvent adsorbés par les poussières atmosphériques. Ces rejets se présentent alors dans la gamme granulométrique des PM_{2,5} (*Magdelaine, 2001*).

Les particules diesel: Les moteurs diesel émettent des particules toxiques capables d'atteindre les alvéoles pulmonaires. Ainsi, leur diamètre aérodynamique moyen est centré sur 0,1 µm en sortie d'échappement. Le diesel dégage une pollution spécifique visible, des fumées noires et des suies constituées de particules de carbone et imprégnées de diverses substances, en particulier d'hydrocarbures aromatiques polycycliques (HAP).

Ainsi, les particules diesel sont principalement constituées :

- de résidus solides sous forme de suies (carbone élémentaire et carbone organique) ;
- d'une partie soluble constituée d'hydrocarbures imbrûlés, principalement les HAP, issus de la combustion incomplète du carburant diesel ;
- d'une fraction insoluble (carbone non fonctionnalisé, etc.).

Ce sont en fait, des particules solides avec une structure fractale qui s'apparente à une éponge. Des mesures effectuées en site proche du trafic révèlent qu'environ 75% du carbone suie se trouve dans la fraction granulométrique inférieure à 1 µm, le reste étant à associer à des particules plus grossières. Ce qui démontre bien l'importance et la spécificité de la pollution diesel.

1.5. Les Eléments Traces (ET)

Eléments traces (ET), anciennement appelés les métaux lourds ou éléments traces métalliques (ETM), existent naturellement mais en quantités très faibles dans les sols, l'eau et l'air (***Blin et al., 2004***). Ils désignent en général les métaux dont le poids atomique est supérieur à celui du fer.

Certaines activités humaines, comme la combustion du charbon, du pétrole, des déchets et certains procédés industriels en rediffusent en revanche en grande quantité dans l'environnement. Ils sont souvent liés aux particules fines PM_{2,5}, à l'exception du mercure, qui est principalement gazeux (***Blin et al., 2004***).

Si une partie des métaux lourds peut retomber aux alentours de la source d'émissions, certains métaux lourds peuvent voyager sur de longues distances (***Blin et al., 2004***).

Les éléments traces ne posent pas seulement un problème pour la pollution de l'air : ils sont bio-persistants, perturbent les écosystèmes, détériorent les sols, les eaux de surface, les forêts et les cultures et s'accumulent dans la chaîne alimentaire. Certains sont cancérigènes pour l'homme (***Blin et al., 2004***).

On considère en général les ET suivants: Arsenic (As), Cadmium(Cd), Chrome (Cr), Cuivre (Cu), Mercure (Hg), Nickel (Ni), Plomb (Pb), Sélénium (Se), Zinc (Zn).

Le mercure (Hg) est le seul métal liquide à température ambiante. Il se combine très aisément avec d'autres composés et a une volatilité importante. Pour le mercure métallique (Inorganique), on le retrouve sous forme gazeuse, liquide ou ionique. Mais le mercure peut également se combiner avec du gaz carbonique et on parle alors de mercure organique. Le mercure est très sensible à l'acidité du milieu. Il est extrêmement toxique et a des effets sur le système nerveux. La volatilité importante du mercure en fait un polluant important dans les études de pollution transfrontière (***Bur et al., 2008***).

Le plomb (Pb) est une source importante des émissions de plomb dans l'atmosphère a été le transport car le plomb a été pendant longtemps additionné à l'essence du fait de son pouvoir antidétonant. De ce fait, il contamine souvent les terrains en bordures d'axes routiers. L'ingestion de plomb déclenche le saturnisme (***Bur et al., 2008***).

Le cadmium (Cd) Il provient surtout de l'incinération des déchets, ainsi que de procédés industriels (métallurgie,...). Il a des effets sur le système respiratoire et gastro-intestinaux (***Blin et al. 2004***).

L'antimoine (Sb) est un métalloïde de couleur gris métallique, présentant deux niveaux d'oxydation (III et V). Il est utilisé comme retardateur de flamme, dans les semi-conducteurs et dans des alliages métalliques. Cet élément est répandu dans l'environnement par la combustion du charbon, les activités minières ou métallurgiques (***Gebel et al., 1998; Bur et al., 2008***).

L'arsenic (As) est un métalloïde de couleur gris métallique, possédant deux états d'oxydation (III et V). Il est souvent associé aux activités minières et à la combustion de ressources organiques fossiles. Il a été utilisé en agriculture en tant que pesticide (arséniate de

plomb), en tant que durcisseur dans les cartouches de chasse, ou associé à Ga (Galium) ou In (Indium) pour former des matériaux supraconducteurs (*Blin et al., 2004*).

C'est un élément non essentiel et un poison puissant dont la toxicité varie avec l'état d'oxydation : les arsénites (III) sont plus toxiques et plus mobiles dans les sols que les arséniates (V).

Le chrome (Cr) est un métal dur, blanc grisâtre, possédant 3 degrés principaux d'oxydation (II, III, VI) et cinq états d'oxydation inhabituels (-II, -I, 0, IV, V). L'état d'oxydation conditionne la stabilité et la toxicité de Cr dans les sols (Cr (III) plus stable mais moins toxique que Cr (VI)). Il est principalement utilisé dans la fabrication de l'acier et d'autres alliages, de pigments et dans le traitement du cuir, du bois et de l'eau.

C'est un élément essentiel pour l'homme (maintien du métabolisme du glucose), mais sa nécessité n'a pas été démontrée pour les végétaux (*Blin et al., 2004*).

Le cobalt (Co) est un métal de transition de couleur blanc-argenté, présentant deux états d'oxydation (II et III). Il est utilisé dans les alliages, les aimants, les médias d'enregistrement, et comme catalyseur dans l'industrie chimique et pétrolière. C'est un élément essentiel entrant, par exemple, dans la composition de la vitamine B12 mais présentant un caractère toxique dû à ses propriétés chimiques (*Marcel et al., 2006*).

Le cuivre (Cu) est l'un des métaux les plus anciennement connus car il existe à l'état natif (comme Ag, As, Au, Sb, Pb). Il est principalement utilisé (seul ou sous forme d'alliage) dans la confection de feuilles métalliques, de câbles électriques ou de tuyaux (très largement utilisé pour les conduites d'eau) (*Marcel et al., 2006*).

Le Cu est un oligo-élément, il est plus communément étudié pour les problèmes de carence qu'il peut engendrer chez les végétaux, le bétail et l'homme, que pour sa toxicité.

Le sélénium (Se) est un élément non métallique de couleur gris-noir, présentant 3 états d'oxydation (II, IV, VI). La gamme dans laquelle il est présent dans les systèmes biologiques en quantité suffisante pour éviter une carence, mais sans être toxique, est étroite : de 0.04 à 4µg.g⁻¹ (*Marcel et al., 2006*).

Le zinc (Zn) est un métal, moyennement réactif, qui se combine avec l'oxygène de l'air pour former une couche d'oxyde blanc imperméable. On le trouve généralement à l'état d'oxydation (II) (*Blin et al., 2004*).

Il est principalement extrait sous forme de sulfure (ZnS; sphalérite) dont l'oxydation est généralement accompagnée d'une oxydation de la pyrite créant un contexte acide favorable à la mise en solution de différents ET.

Zn est largement utilisé pour la galvanisation de l'acier, mais aussi dans des produits cosmétiques (crèmes, shampoings), batteries, pigments et peintures.

Le Hg, Pb et Cd sont tous très toxiques (effets sur le système nerveux) et ont une durée de vie très grande et une conductivité électrique élevée. Les métaux ne posent pas seulement un problème pour la pollution de l'air, mais aussi pour celle de l'eau et des sols.

I.5.1. Les impacts des Eléments Traces

La présence des métaux lourds dans l'environnement est connue depuis bien des années. La forte concentration de ces derniers présente un risque majeur pour la santé humaine et pour l'environnement par leurs propriétés (*Seraghni, 2007*):

- Ils ne sont pas dégradables.
- Ils sont enrichis au cours de processus minéraux et biologiques, puis s'accumulent dans la nature.
- Ils peuvent être absorbés directement par le biais de la chaîne alimentaire, entraînant des effets aigus ou chroniques.

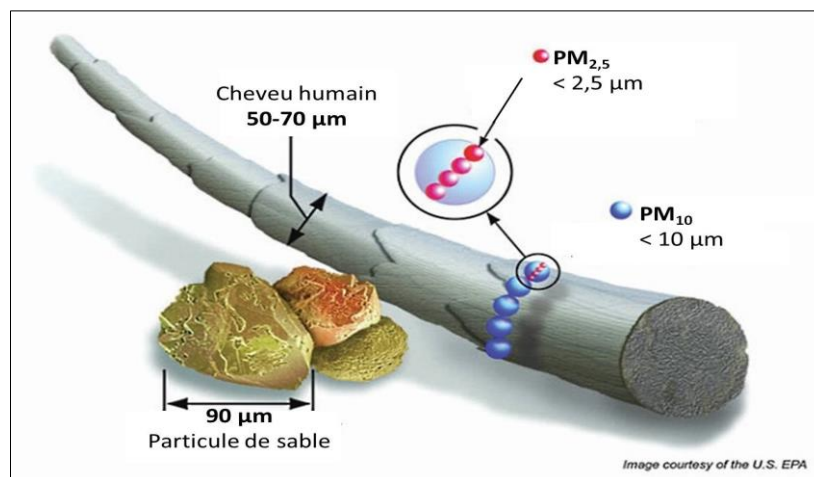


Figure I.2: Comparaison entre PM_i, grain de sable et un cheveu humain
([Http://www.deq.state.or.us/air/planning](http://www.deq.state.or.us/air/planning))

I.5.1.1. Impact sur la végétation

Les trop fortes teneurs en métaux lourds diminuent la densité, la biodiversité et l'activité de la microflore et de la microfaune du sol (*Bert et al., 2010*).

Une forte concentration des métaux lourds entraîne un déséquilibre métabolique. En effet, l'accumulation des métaux lourds dans l'organisme des végétaux se traduit par le remplacement des éléments essentiels par leur pouvoir oxydant et leur forte stabilité. Ceci conduit non seulement à la nécrose des racines et par la suite la mort des végétaux mais aussi le transfert de ces métaux dans la chaîne alimentaire (*Vanobberghen, 2010*).

I.5.1.2. Impact sur les animaux

Les animaux peuvent être contaminés par contact avec la peau, par inhalation de poussières toxiques et par ingestion de l'eau et d'aliments pollués. La dispersion des métaux constitue un danger pour la faune sauvage, le bétail et l'Homme, qui se situe en bout de chaîne alimentaire. Dans certaines zones très contaminées, le pâturage, les cultures fourragères ou destinées à l'alimentation humaine sont interdites (*Bert, 2010*).

I.5.1.3. Impact sur la santé

Les métaux lourds dans l'atmosphère sont sous forme de particules en suspensions à l'exception de certains métaux qui sont volatils à température ambiante (EX. le mercure). Par conséquent, l'homme est touché par cette pollution via les voies respiratoires qui se traduit par l'apparition de nombreuses maladies (cancers, maladies respiratoires...etc.) (**Boldo et al., 2006 ; Oberdorster et al., 2005**). Les divers impacts sont résumés dans le tableau I.2.

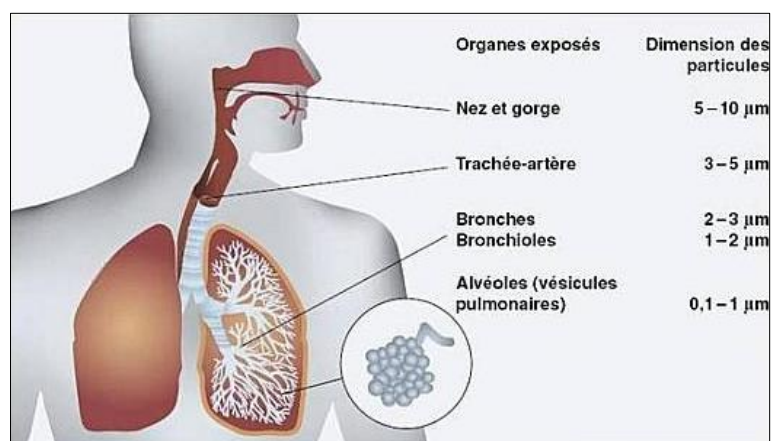


Figure I.3: Pénétration des pmi dans le système respiratoire
(http://a398.idata.over_blog.com/540x362/0/00/49/42)

Tableau I.2: Les organes affectés par les métaux lourds

<i>Organe</i>	<i>Arsenic</i>	<i>Cadmium</i>	<i>Cuivre</i>	<i>Mercure</i>	<i>Plomb</i>	<i>Thalliu</i>
Mutations génétiques	-	-	-	X	-	-
Système immunitaire	-	X	-	-	-	-
Système nerveux	X	-	-	X	X	X
Lésions cérébrales	-	-	X	X	X	-
Peau	X	-	-	-	-	X
Poumons système respiratoire	X	X	-	-	-	X
Système digestif et gastro-intestinal	X	-	-	X	-	-
Foie	-	-	-	X	-	-
Reins	X	X	-	X	X	-
Sang	-	-	-	-	X	-
Cancer	X	X	-	-	-	-

I.6. Méthode d'échantillonnage des TSP

L'échantillonnage des aérosols est l'opération qui permet d'obtenir un échantillon de matière dispersée dans l'air sous forme de particules solides ou liquides. Cette opération peut répondre à plusieurs objectifs :

- Déterminer les caractéristiques physico-chimiques de l'aérosol en un point.
- Mesurer l'évolution d'un ou de plusieurs paramètres au cours du temps.
- Évaluer l'exposition de personnes à des substances nocives ou toxiques à un poste de travail particulier.

Un système d'échantillonnage d'aérosol comprend généralement plusieurs parties distinctes ayant chacune leur fonctionnalité. L'aérosol est aspiré à travers un orifice ou une fente à l'aide d'une pompe d'aspiration menée d'un volumètre ou d'un débitmètre, qui assure le captage de l'aérosol ambiant (*Larssen et al., 1999*).

L'aérosol aspiré est piégé dans un filtre placé au niveau de l'orifice d'aspiration (généralement l'aérosol échantillonné correspond aux particules collectées sur ce support de collecte). Si les particules contiennent des substances volatiles, celles-ci peuvent subir une vaporisation partielle sur le support. Pour mesurer de telles substances, il convient d'utiliser soit dans la cassette (support du filtre), soit à l'intérieur d'un dispositif extérieur associé, un ou plusieurs substrats capables de fixer la phase vapeur (filtre imprégné de réactif, support adsorbant,...).

Les particules de l'aérosol subissent tout au long de leur cheminement, depuis l'air ambiant jusqu'au filtre, une série d'évènements qui influent sur leur probabilité d'être échantillonnées effectivement sur le support final. Le mouvement des particules de l'aérosol au voisinage et à l'intérieur d'un échantillonneur est uniquement guidé par des paramètres physiques liés aux particules (taille, forme, densité,...), à la géométrie de l'échantillonneur et aux conditions aérauliques de son environnement immédiat (vitesse d'air extérieur, intensité de turbulence,...).

I.6.1. Le choix du site de prélèvement

Le choix d'un site d'étude ou de prélèvement est en fonction du type de pollution atmosphérique à surveiller et du risque présentés sur ce site. Quand il s'agit de la santé humaine, l'installation de stations de surveillance de la qualité de l'air au niveau des centres urbains à densité importante est plus que nécessaire (*Elichegaray et al., 2010*).

I.6.2. Le choix du filtre

Le choix des filtres pour la collecte des TSP doit répondre aux trois exigences suivantes (*Oikawa 1999*) :

- Une efficacité de collection d'au moins 99% pour les particules de diamètre égal ou supérieur à 0,3µm ($d = 0,3\mu\text{m}$).
- Une hygroscopicité faible.
- Un taux d'impureté aussi faible que possible en éléments métalliques pour éviter les interférences lors de l'analyse quantitative des particules collectées sur le filtre.

I.6.2.1. Filtre en fibre de verre

Avec une porosité de $0,1 \div 4\mu\text{m}$, ces filtres présentent une efficacité de rétention de 99,9% pour les $\text{PM} > 0,3\mu\text{m}$, mais ils ne sont pas fiables pour l'analyse qualitative des particules retenues vu que plusieurs éléments métalliques rentrent dans leur composition (Oikawa 1999).

I.6.2.2. Filtres membranaires

Ce sont des filtres nitrocellulose, acétate et ester de cellulose qui sont solubles dans les éthers et les alcools caractérisés par leur très grande efficacité. Ils ne sont pas utilisés pour les prélèvements à haut débit car ils présentent une grande résistance à l'écoulement à cause de leur fine porosité. Les propriétés et les caractéristiques de ces filtres sont résumées dans le tableau I-3

Tableau I.3: Caractéristique des filtres membranaires (Hinds, 2012).

Filtre	Type	Matériel	Epaisseur (mm)	Porosité (μm)	Résistance (mm)	Efficacité (%)
Whatman4	Fibre	Cellulose	0,19	3 à 20	0,35	72
Microsorbant	Fibre	Polystyrène	1,50	0,70	0,04	99,50
MSA1106B	Fibre	Verre	0,23	0,1 à 4	0,10	99,93
Millopore AA	Membrane	Ester cellulose	0,15	0,8	0,19	99,98

I.7. Normes et réglementation

Vu les nombreux risques provoqués par les ETM et les TSP sur la santé humaine, la faune et la flore, plusieurs conventions internationales portent sur la pollution atmosphérique.

Le journal officiel de la République Algérienne du 08 Dhou El Hidja 1426 / 8 janvier 2006 consacre les articles 03, 04, 05, 06, 07, 08, 09 et 10 (JORA N°1, 2006) pour la surveillance de la qualité de l'air. L'extrait de ces articles est le suivant :

- La surveillance de la qualité de l'air concerne CO_2 , SO_2 , O_3 , TSP
- Les valeurs limites ainsi que les objectifs de qualité de l'air sont fixés comme suit :
 - * Objectif de qualité ou valeur cible annuelle des TSP: $50 \mu\text{g}/\text{m}^3$.
 - * Valeur limite annuelle des TSP: $80 \mu\text{g}/\text{m}^3$.
- Lorsque les seuils d'information et les seuils d'alerte sont atteints ou risquent de l'être, les responsables prennent toutes les mesures visant à protéger la santé humaine et l'environnement ainsi que les mesures de réduction et/ou de restriction des activités polluantes.

Les valeurs limites fixées par l'OMS pour les TSP sont $100 \mu\text{g}/\text{m}^3$ comme valeur limite annuelle et $250 \mu\text{g}/\text{m}^3$ comme valeur limite journalière.

Concernant les ETM et selon l’OMS et le Journal Officiel de l’Union Européenne, la réglementation actuelle ne touche que certains métaux lourds (*WHO, 2000*).

Tableau I.4: Valeurs cibles et valeurs limites de quelques ETM (*Piervittori and Maffei, 2001*)

Polluant	Union Européenne		OMS
	Valeur limite	Valeur cible	Valeur limite
Plomb	0,5µg/m ³	-	0,5 µg/ m ³
Arsenic	-	6 ng/ m ³	-
Cadmium	-	5 ng/ m ³	5ng/ m ³
Nickel	-	20 ng/ m ³	20ng/ m ³
Mercuré	50ng/ m ³	-	-

1.8. Évaluation de la pollution de l’air en Algérie

La pollution de l’air par les métaux suscite un grand intérêt au sein de la communauté scientifique (*Contini et al., 2012 ; Calvo et al., 2013 ; Feng et al., 2013 ; Quentin et al., 2015*). En effet, les métaux, qui sont des constituants normaux de l’environnement à l’état de traces, contrairement à de nombreux contaminants (pesticides), sont tous toxiques au dessus d’un certain seuil (*Kucuksezgin et al., 2006*). Ainsi, les éléments traces métalliques dits essentiels (rôle important dans les processus biologiques: le cuivre, le fer et le zinc) peuvent produire des effets toxiques au même titre que ceux dits non essentiels (aucun rôle dans les processus biologiques: le mercure, le plomb et le cadmium) lorsque leur concentration dépasse un certain seuil d’acceptabilité (*Chiffolleau et al., 2001 ; Lane et al., 2015 pour le cadmium*).

Depuis le début des années 90, en Algérie, les recherches sur la surveillance de la contamination particulaire et métallique du milieu atmosphérique se multiplient ; mais de nombreuses difficultés d’ordre méthodologique (liées au fait que les travaux portent sur des substances invisibles, contenues dans un air toujours en mouvement, émises par des millions de sources mobiles, elles-mêmes très hétérogènes) restent à résoudre.

En effet, depuis le début des années 70, on constate une hausse du trafic routier et le parc de véhicules diesel a augmenté très rapidement, ce qui a suscité le démarrage des études d’impact du trafic routier dans les milieux urbains et périurbains (*Boukadoum, 2005*). L’intérêt pour les nuisances liées à la circulation automobile s’est peu à peu transformé en une véritable préoccupation de notre Société. Au départ centré essentiellement sur les risques sanitaires liés aux rejets des véhicules et sur le bruit provoqué par le trafic, cette préoccupation s’est peu à peu élargie pour englober aujourd’hui quasiment tous les éléments

de notre environnement (risque pour la santé, nuisances sonores, dégradation visuelle des paysages par les routes, rejets et accumulation dans la nature d'éléments dangereux pour les écosystèmes eux-mêmes.

Dans ces conditions, il nous a semblé difficile de ne pas proposer un rapide point sur les connaissances actuelles, parfois accompagné d'une analyse détaillée des travaux axés sur la pollution atmosphérique qui sont en relation directe avec notre travail.

En effet, on recense plusieurs travaux réalisés dans ce contexte au centre, à l'est et à l'ouest d'Alger qui relèvent presque tous des études de la pollution atmosphérique par les particules PM_{10} , $PM_{2.5}$ et PM_1 ainsi que de certains éléments traces qui leur sont associés. A l'est et plus exactement au niveau du site implanté à l'ENP, une étude de niveau granulométrique PM_{10} , $PM_{2.5}$ et PM_1 (2002-2003) a été réalisée et les valeurs moyennes annuelles recueillies (respectivement 75.2, 36.1 et 25.9 $\mu\text{g}/\text{m}^3$) dépassent largement les normes en vigueur (*Oucher and Kerbach, 2012*). L'étude d'éléments traces qui leur sont associés montre que le plomb puis le fer sont les éléments les plus abondants (*Oucher et al., 2015*).

Un autre site situé à Bab Ezzouar classé comme urbain a aussi fait l'objet d'une étude granulométriques des différentes matières particulaires (2004-2005), les résultats sont très similaires à ceux constatés dans le site de l'ENP (*Bitouche, 2011*). Au centre d'Alger, les données enregistrées dans deux études (Place du 1^{er} mai en 2006 et Bach-Djerrah en 2009) (*Berkouki, 2007*), ont fait l'objet d'un rapport annuel sur la qualité de l'air établi par le Ministère de l'Aménagement du Territoire et de l'Environnement, qui montre des valeurs annuelles en PM_{10} excédant largement la valeur limite annuelle (40 $\mu\text{g}/\text{m}^3$) et celle de l'objectif de qualité (30 $\mu\text{g}/\text{m}^3$) (*Samasafia, 2003*).

A l'ouest, une étude récente (2011-2012) sur une zone rurale de Tipaza a enregistré des teneurs en PM_{10} , $PM_{2.5}$ et PM_1 (respectivement 38.7, 12.1 et 2.7 $\mu\text{g}/\text{m}^3$) nettement inférieures à celles enregistrées dans les autres zones étudiées. En effet, ce site n'est pas exposé directement aux émissions des grands axes routiers à faible densité de population de la commune de Larhat de la wilaya de Tipaza (7359 habitants en 2011), ainsi l'analyse de la composition élémentaire associée montre que le plomb, après le cuivre et le fer, est l'élément le plus abondant (*Talbi, 2010*).

Ainsi, les seules études qui concernent les aérosols prélevés en mode matière particulaire en suspension (TSP) au niveau de la capitale concernent en premier lieu le site de Bab El Oued qui est caractérisé par la présence d'un hôpital et d'un incinérateur de déchets médicaux ainsi que par la forte densité de la circulation automobile. En effet, les résultats expérimentaux en TSP recueillis en ce site, dépassent la valeur cible (50 $\mu\text{g}/\text{m}^3$). Cette étude a prouvé l'implication des TSP dans le transport des éléments traces reconnus toxiques tels que : le Cd, Pb, Cr et Ni (*Belamri and Benrachedi, 2009*).

Le deuxième site concerné par l'étude des TSP et éléments associés, est celui de l'hôpital Mustapha Pacha, les résultats expérimentaux montrent que 33 % de la teneur journalière en TSP varient dans une gamme en dessus de la valeur limite (80 $\mu\text{g}/\text{m}^3$). Dans cette étude, les éléments tels que : Na, Mg, Cl, Sc, Cr, Tl, V, Fe, Co, Cu, Zn, Se, Br, Ag, Sb, Ce, La, Hf, Ta et Hg ont été identifiés et quantifiés à l'aide de la technique d'analyse par

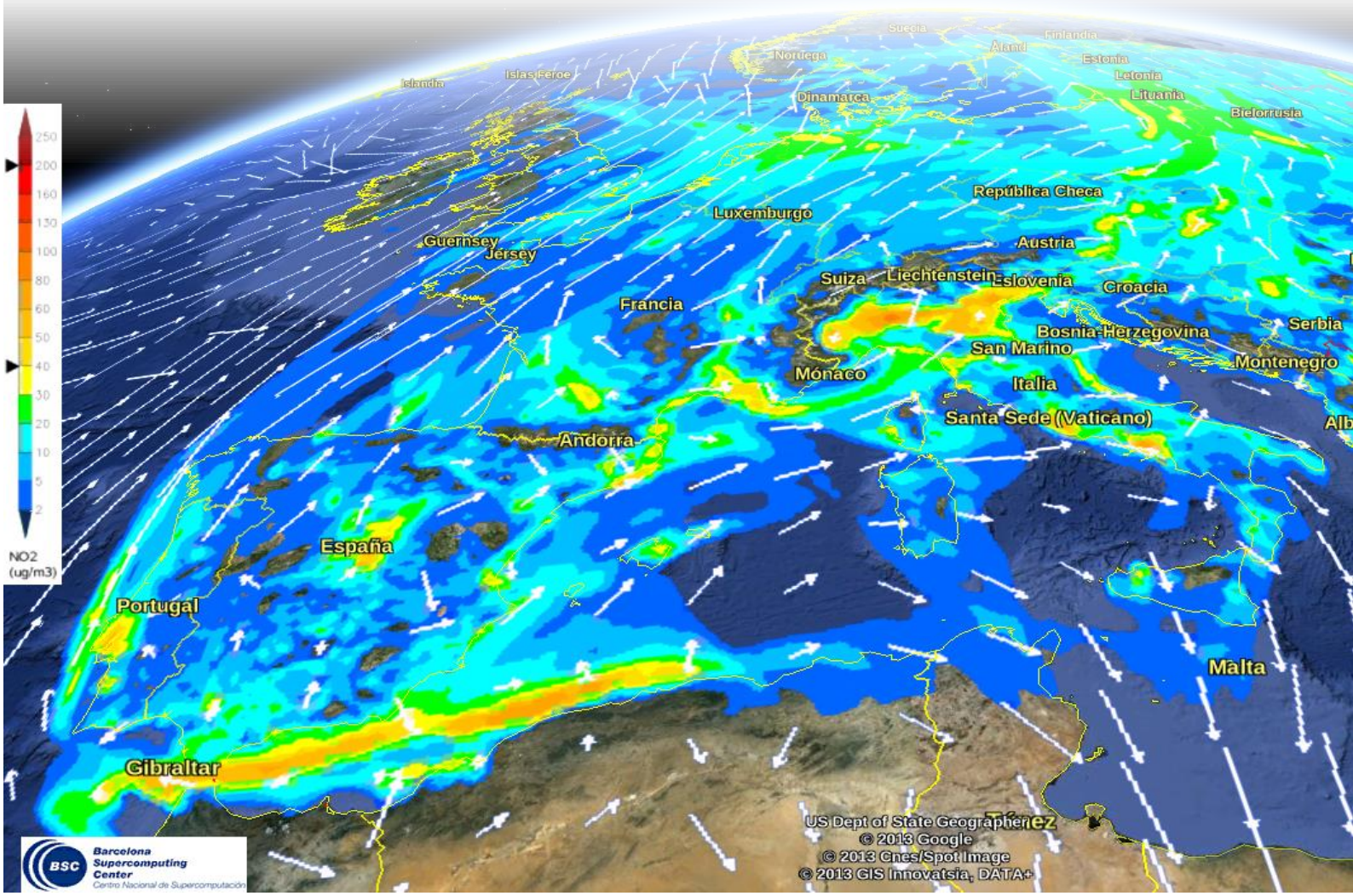
activation neutronique. Cette étude a montré que la limite supérieure des valeurs pour les poussières en suspension totales et les concentrations élevées pour certains éléments toxiques trouvés sont dus principalement aux conditions météorologiques et à l'intensité de la circulation routière autour du site de collecte (*Belamri and Benrachedi, 2010*).

D'autre part, plusieurs campagnes de suivi de la pollution de l'air ont été menées à Alger pour évaluer le niveau de contamination par les pesticides et polluants organiques persistants (POP) (*Moussaoui et al., 2012*), ainsi que par les BTEX (*Kerchich, 2012*) et les CO, HC, NO_x et de CO₂ (*Chikhi, 2014*). Toutes ces études montrent que la pollution à Alger est plus importante que dans les villes occidentales, où le taux de motorisation est bien plus élevé.

Chapitre II

La modélisation

« Toute loi physique doit avoir une beauté mathématique. »
(Paul Dirac 1902- 1984)



Chapitre II : La modélisation

Introduction

La modélisation est la représentation d'un système par un autre plus facile à appréhender. Il peut s'agir d'un système mathématique ou physique, le modèle sera alors numérique ou analogique.

La modélisation est la simplification du phénomène réel, à des échelles de temps et d'espace différents du réel. Il s'agit généralement de trouver des équations reliant les différents paramètres du phénomène entre eux.

Gérard Dreyfus en 2002 a défini le modèle comme étant une représentation de la réalité visible ou observable. Selon Jean Perrin, modéliser c'est remplacer du visible compliqué par l'invisible simple.

Les modèles sont divisés en trois catégories comme suit :

- **Modèles de connaissance** consistent à mettre en œuvre des équations décrivant les phénomènes étudiés (biologiques, physiques, chimiques...) en se basant sur des connaissances théoriques sur ces derniers. Ils ont une valeur prédictive et explicative.
- **Modèles boîte noire**, qui ne nécessitent que la possession des mesures, n'ont qu'une valeur prédictive sur un domaine de validation.
- **Modèles semi physique** qui reposent sur l'élaboration des équations basées sur une théorie, et des équations purement empiriques qui résultent d'une modélisation de type boîte noire.

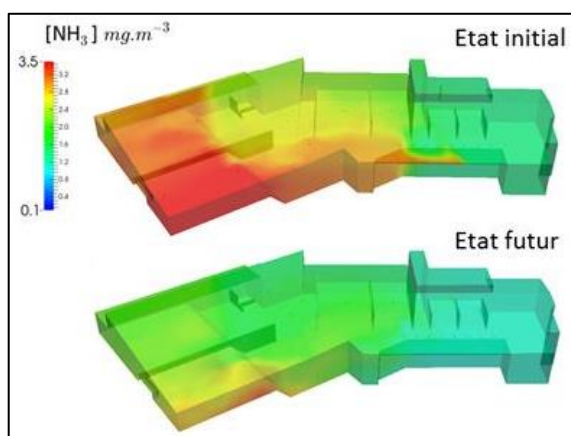


Figure II.1: Modélisation 3D de la qualité de l'air à l'intérieur d'une chambre
(<https://www.egis.fr/action/actualites/modelisation-3d-pour-la-qualite-de-lair-interieur>)

Pourquoi modéliser ? :

Particulièrement utilisée dans les différents métiers de l'environnement, la modélisation permet de (*Contamine 2005*) :

- Réduire le nombre, la durée et les coûts des essais expérimentaux.
- Réduire les coûts de construction et d'exploitation des procédés.

L'avantage principal de la modélisation permet de tester différents scénarios sans contrainte d'échelles (pilote, prototype, ouvrage industriel) et de faisabilité.

La modélisation numérique constitue une nouvelle méthode d'approximation des systèmes complexes, particulièrement utiles lorsque ses systèmes sont difficilement modélisables à l'aide des méthodes statistiques classiques.

Ces dernières années une nouvelle génération de modèles est apparue, basée sur le concept des réseaux de neurones humains (*Jodouin, 1994 ; French et al., 1992*) l'ont utilisé pour la première fois dans la prévision spatiale et temporelle des précipitations, alors que (*Zealand et al., 1999*) l'ont appliqué pour la prévision de l'écoulement des rivières. Ces modèles ont trouvés une vaste utilisation dans la simulation et la prévision des séries hydrologiques et notamment dans la modélisation pluie-débit (*Lek et al., 1996, Hsu et al., 1995, Minns et al., 1996, Sajikumar et al., 1999, Shamseldin, 1997 ; Coulibaly et al., 1999*).

II.1.L'analyse factorielle exploratoire (AEF)**II.1.1. Présentation générale de l'AEF**

L'analyse factorielle exploratoire (AFE) est une technique qui permet de mettre en évidence la structure latente d'une masse de données. On entend par structure latente, la présence d'un certain nombre de facteurs (ou de dimensions sous-jacentes) permettant d'expliquer pourquoi certaines de nos variables sont inter-corrélées, alors que d'autres variables ne le sont pas. Les variables latentes (communément appelées facteurs) ne sont pas directement observables, mais elles sont inférées en tenant compte du patron de corrélation observé entre nos variables. L'AFE ressemble beaucoup à l'analyse en composantes principales (ACP) et on peut faire de nombreux parallèles entre les deux techniques : dans les deux cas, les analyses portent sur des matrices de corrélation pouvant comporter un nombre élevé de variables; de plus, dans les deux cas, il s'agit d'expliquer une portion importante de variance présente dans les données en ayant recours à un nombre limité de dimensions.

La différence fondamentale entre les deux techniques concerne justement la portion de variance que l'on cherche à expliquer. Alors que l'ACP s'intéresse à la variance totale présente dans la matrice de corrélation, l'analyse factorielle exploratoire quant à elle se concentre strictement sur la portion de variance commune partagée par certaines variables.

L'analyse factorielle exploratoire est une analyse conduite dans le but de découvrir quelles sont les variables latentes (facteurs) sous-jacentes à un ensemble de variables ou de mesures. Cette technique est généralement mise en contraste avec l'analyse factorielle confirmatoire qui permet de tester formellement des théories et de mettre à l'épreuve des hypothèses concernant les facteurs anticipés (*Vogt, 1993*).

II.1.2. Description sommaire de la technique

Le postulat fondamental à la base de l'analyse factorielle est le suivant : si des variables sont corrélées les unes avec les autres dans les données, c'est parce qu'elles subissent l'influence de certains facteurs qui leur sont communs.

L'analyse a pour objectif de mettre en évidence ces facteurs communs, qui ne sont malheureusement pas directement observables, mais qui pourront être estimés. Le modèle mathématique à la base de l'analyse factorielle s'exprime dans un ensemble d'équations s'apparentant à des équations de régression multiple et ayant la forme suivante :

$$\begin{aligned} var_1 &= r_1 F_1 + r_2 F_2 + r_3 F_3 \dots + r_i F_i + U_1 \\ var_2 &= r_1 F_1 + r_2 F_2 + r_3 F_3 \dots + r_i F_i + U_2 \\ var_3 &= r_1 F_1 + r_2 F_2 + r_3 F_3 \dots + r_i F_i + U_3 \\ &\dots \\ &\dots \end{aligned}$$

Quelques observations importantes doivent être faites à l'égard de cet ensemble d'équations. Tout d'abord à remarquer que les équations sont identiques dans leur forme pour chacune des variables soumises à l'analyse ($var_1, var_2, var_3, \dots, var_n$). Cependant, il faut bien comprendre que les coefficients de régression r_i correspondent à des paramètres qui seront déterminés lors de l'analyse et qui prendront dans les faits des valeurs différentes d'une variable à l'autre. Chaque variable sera donc prédite à l'aide des mêmes facteurs communs mais recevant des pondérations différentes. Une deuxième observation importante concerne les facteurs F_i et suivants qui ne sont pas des variables prévisionnelles individuelles comme on est habitué de les voir en analyse de régression. Ce sont plutôt des facteurs latents, non directement observables, mais définis eux-mêmes par différents regroupements de variables (*Rherib, 2021*).

L'une des tâches du chercheur sera justement d'identifier ou de nommer ces facteurs en examinant les variables qui les composent. Enfin, chaque équation comporte également un terme U correspondant à une source de variation unique pour chaque variable. La source de variance U est très importante en analyse factorielle puisqu'elle permet justement de distinguer cette technique de l'analyse en composantes principales. En effet, dans le modèle de l'analyse factorielle, chaque variable possède une portion de variance unique qui est indépendante de toutes les autres sources de variance et qui n'a rien à voir avec la variance commune; cette variance unique est elle-même constituée de deux portions, à savoir (*Brigitte et al., 2008*) :

- La portion de variance réelle de la variable, non assujettie à l'influence des facteurs communs et donc appelée variance spécifique,
- La variance d'erreur de la variable. L'analyse factorielle met de côté la variance unique des variables et se concentre sur la portion de variance commune pour en extraire les dimensions sous-jacentes.

En contraste que l'analyse en composantes principales ne fait pas ces nuances entre variance commune et variance unique, ne fait aucune place à la variance d'erreur, assume que la variance totale est importante et l'utilise donc au complet pour en extraire des composantes.

Pour ces raisons, il faut reconnaître que l'AFE et l'ACP sont fort différentes tant au niveau conceptuel qu'au niveau des modèles mathématiques qui les sous-tendent.

Une conséquence importante découle de ce qui vient d'être dit : en analyse factorielle, le chercheur doit préciser dès le début pour chacune des variables analysées, la portion de variance commune sur laquelle il souhaite se pencher.

Or, paradoxalement, cette information ne sera disponible qu'au terme de l'analyse factorielle. La solution à ce dilemme consiste à fournir au logiciel d'analyse un estimé de la variance commune de chaque variable. L'approche la plus courante consiste à utiliser le r^2 comme estimé de la variance commune. Avant l'analyse proprement dite, les estimés de départ des variances communes sont donc placés dans la diagonale principale de la matrice de corrélation à être analysée, remplaçant ainsi les valeurs 1 que l'on retrouve habituellement le long de cette diagonale. On parle désormais d'une matrice de corrélation réduite, puisque la variance totale de chacune des variables n'est plus 1, mais bien une valeur réduite r^2 correspondant à la variance commune estimée.

L'analyse factorielle de la matrice réduite peut alors démarrer et les facteurs sont extraits selon le nombre requis. Cette première extraction mène à une solution dans laquelle les variances communes de chaque variable sont calculées. Or, il est bien possible que ces variances communes calculées diffèrent des variances communes estimées au départ; si c'est le cas, les anciennes valeurs r^2 de la diagonale principale de la matrice analysée sont remplacées par les nouvelles valeurs calculées et le processus d'extraction des facteurs recommence. Ce processus itératif continue autant de fois que nécessaire, c'est-à-dire jusqu'à ce qu'il n'entraîne plus de changements importants dans le calcul des variances communes. On dit alors que la solution a convergé, c'est-à-dire qu'elle a atteint une phase stable et qu'elle peut maintenant être interprétée.

II.1.3. Particularité des matrices réduites

En analyse factorielle, la matrice de corrélation réduite comporte une diagonale principale correspondant à des estimés de la variance commune de chaque variable. Les coefficients r multiple au carré r^2 sont généralement utilisés à cet effet. En les retraçant dans la diagonale de la matrice que l'on s'apprête à analyser, la somme des valeurs de cette diagonale correspond à la variance qui sera analysée.

Il va sans dire que les matrices de corrélation réduites soumises à une AFE doivent posséder les mêmes qualités globales que celles notées dans le cas de l'ACP. Ainsi, une variable qui ne serait corrélée à aucune autre devrait certainement être retirée de l'analyse, puisque nous nous intéressons à la variance commune partagée entre les variables. À cet égard, le calcul des mesures d'adéquacité de l'échantillonnage de Kaiser-Meyer-Olkin est tout indiqué (*Kaiser, 1960*).

$$KMO = \frac{\sum_{j \neq k} \sum_{jk} r^2}{\sum_{j \neq k} \sum_{jk} r_{jk}^2 + \sum_{j \neq k} \sum_{jk} p_{jk}^2}$$

r^2 : la corrélation entre les variables en question

p^2 : la corrélation partielle entre les variables

Rappelons les valeurs repères suggérées par Kaiser :

- Inacceptable en dessous de 0.5
- Médiocre entre 0.5 et 0.6
- Moyen entre 0.6 et 0.7
- Bien entre 0.7 et 0.8,
- Très bien entre 0.8 et 0.9
- Excellent au-delà de 0.9

Aussi, les situations de multi-colinéarité modérée ne sont pas problématiques en AFE, mais il faut à tout prix éviter la situation de singularité où une variable serait entièrement prédite par une combinaison de plusieurs variables.

II.1.4. Nombre de facteurs communs à extraire

La détermination du nombre de facteurs communs à extraire dans une analyse factorielle se fait en s'appuyant sur les mêmes critères dans le cas de l'ACP. Essentiellement, le chercheur peut appuyer sa décision en considérant les quatre points suivants:

- le critère de Kaiser indiquant la limite inférieure du nombre de facteurs à extraire
- l'inspection du graphique d'accumulation de variance de Cattell
- l'analyse parallèle de Horn
- la facilité à interpréter les facteurs retenus.

Preacher et MacCallum insistent sur le fait qu'aucun critère n'est absolu et que la décision quant au nombre de facteurs à extraire comportera toujours un élément important de subjectivité. Ils suggèrent de combiner l'analyse parallèle et l'inspection du graphique d'accumulation de variance de manière à prendre une décision éclairée. Dans le cas où les différents critères ne concordent pas, il est fortement recommandé de procéder à plusieurs analyses complètes en modifiant le nombre de facteurs extraits et en s'appuyant ultimement sur le critère plus subjectif relié à la facilité à interpréter la solution finale (*Preacher et al., 2003*).

II.2. Le modèle ARIMA de Box-Jenkins

Il existe deux catégories de modèles pour rendre compte d'une série temporelle. Les premiers considèrent que les données sont une fonction du temps ($y = f(t)$). Cette catégorie de modèle peut être ajustée par la méthode des moindres carrés, ou d'autres méthodes itératives. L'analyse des modèles par transformée de Fourier est une version sophistiquée de ce type de modèle.

Une seconde catégorie de modèles cherche à déterminer chaque valeur de la série en fonction des valeurs qui la précède ($y_t = f(y_{t-1}, y_{t-2}, \dots)$). C'est le cas des modèles ARIMA (Auto-Regressive – Integrated – Moving Average). Cette catégorie de modèles a été popularisée et formalisée par Box et Jenkins en 1976 (*Box et al. 1976*).

La méthodologie de Box & Jenkins vise à formuler un modèle permettant de représenter une chronique avec comme finalité de prévoir des valeurs futures. De ce fait, l'objet de cette méthodologie est de modéliser une série temporelle en fonction de ses valeurs passées et présentes afin de déterminer le processus ARIMA adéquat par principe de parcimonie. Cette méthodologie suggère une procédure à trois étapes (*Box et al. 1976*) :

- Identification du modèle
- Estimation du modèle
- Validation du modèle (Test de diagnostique)

A noter que le choix de l'un ou l'autre type de modèle est surtout théorique: est-il raisonnable de penser que dans un phénomène donné, les points sont fondamentalement fonction des points précédents et de leurs erreurs, plutôt qu'un signal, périodique ou non, entaché de bruit.

On peut noter cependant que souvent, on a recours à l'analyse de variance pour traiter les séries temporelles. Or une des assumptions majeures de l'ANOVA est que les résidus des différentes mesures ne sont pas auto-corrélés. Ce n'est évidemment pas le cas si la performance à l'essai t est liée à la performance réalisée à l'essai t-1.

Les processus autorégressifs supposent que chaque point peut être prédit par la somme pondérée d'un ensemble de points précédents, plus un terme aléatoire d'erreur. Le processus d'intégration suppose que chaque point présente une différence constante avec le point précédent.

Les processus de moyenne mobile supposent que chaque point est fonction des erreurs entachant les points précédents, plus sa propre erreur (*Box et al, 2015*).

Un modèle ARIMA est étiqueté comme modèle ARIMA (p,d,q), dans lequel:
p est le nombre de termes auto-régressifs
d est le nombre de différences
q est le nombre de moyennes mobiles.

II.2.1. La différenciation

L'estimation des modèles ARIMA suppose que l'on travaille sur une série stationnaire. Ceci signifie que la moyenne de la série est constante dans le temps, ainsi que la variance. La meilleure méthode pour éliminer toute tendance est de différencier, c'est-à-dire de remplacer la série originale par la série des différences adjacentes. Une série temporelle qui a besoin d'être différenciée pour atteindre la stationnarité est considérée comme une version intégrée d'une série stationnaire (d'où le terme *Integrated*) (*Papoulis, 2002*).

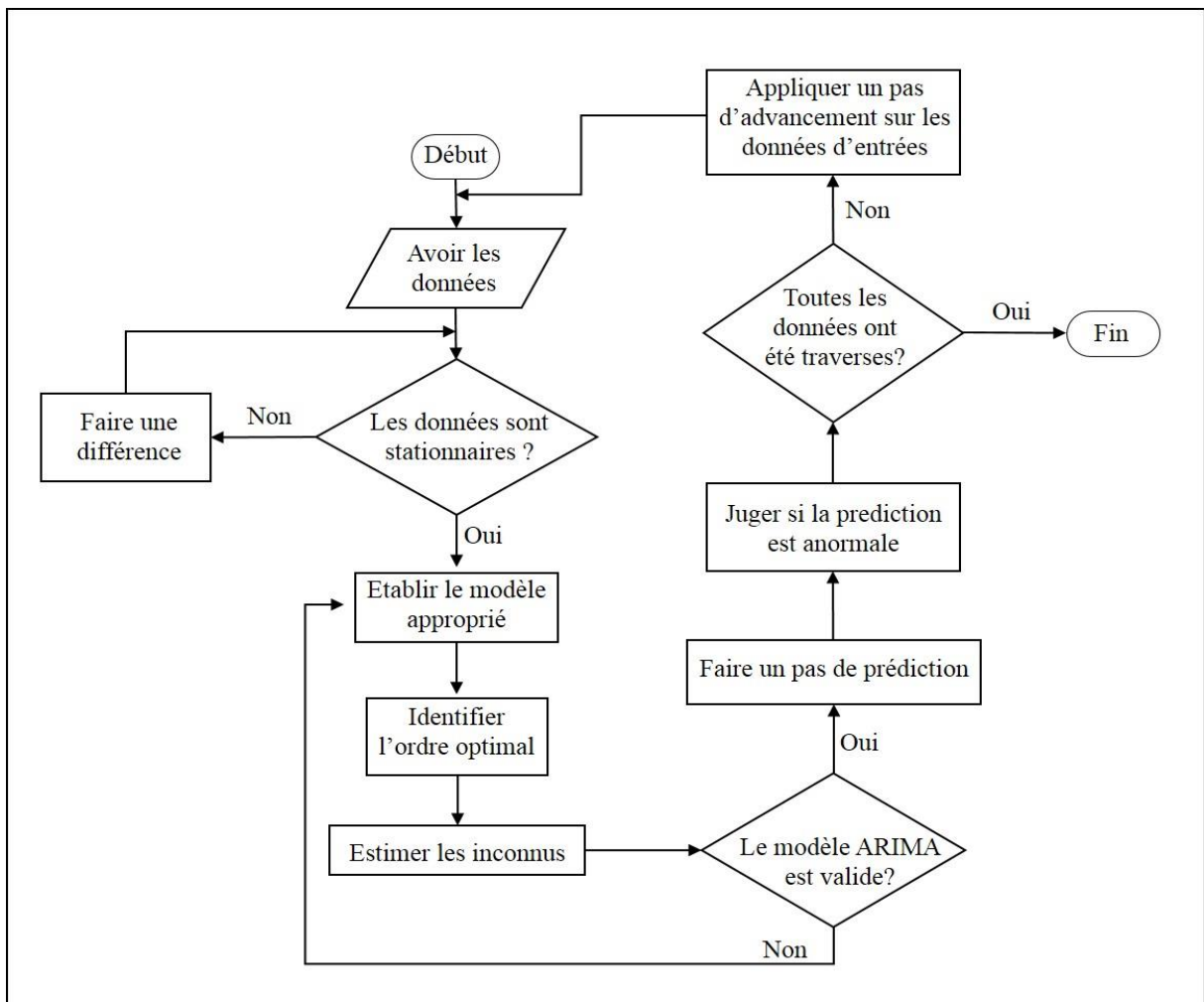


Figure II.2: Algorithme du modèle ARIMA

La correction d'une non-stationnarité en termes de variance peut être réalisée par des transformation de type logarithmique (si la variance croît avec le temps) ou à l'inverse exponentielle. Ces transformations doivent être réalisées avant la différenciation. Une différenciation d'ordre 1 suppose que la différence entre deux valeurs successives de y est constante.

$$y_t - y_{t-1} = \mu + \varepsilon_t$$

μ est la constante du modèle, et représente la différence moyenne en y . Un tel modèle est un ARIMA(0,1,0). Il peut être représenté comme un accroissement linéaire en fonction du temps. Si μ est égal à 0, la série est stationnaire.

Les modèles d'ordre 2 travaillent non plus sur les différences brutes, mais sur les différences de différence. La seconde différence de y au moment t est égale à :

$$(y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2}$$

Un modèle ARIMA (0,2,0) obéira à l'équation de prédiction suivante :

$$y_t - 2y_{t-1} + y_{t-2} = \mu + \varepsilon_t$$

ou encore:

$$y_t = \mu + 2y_{t-1} - y_{t-2} + \varepsilon_t$$

II.2.2. Auto-régression

Les modèles autorégressifs supposent que y_t est une fonction linéaire des valeurs précédentes.

$$y_t = \mu + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varphi_3 y_{t-3} + \varepsilon_t$$

Littérairement, chaque observation est constituée d'une composante aléatoire (choc aléatoire, ε) et d'une combinaison linéaire des observations précédentes. φ_1 , φ_2 et φ_3 dans cette équation sont les coefficients d'auto-régression

A noter que cette équation porte soit sur les données brutes, soit sur les données différenciées si une différenciation a été nécessaire. Pour un modèle ARIMA (1,1,0) on aura :

$$y_t - y_{t-1} = \mu + \varphi(y_{t-1} - y_{t-2}) + \varepsilon_t$$

Ce qui peut également être écrit:

$$y_t = \mu + y_{t-1} + \varphi(y_{t-1} - y_{t-2}) + \varepsilon_t$$

Notez qu'un processus autorégressif ne sera stable que si les paramètres sont compris dans un certain intervalle ; par exemple, s'il n'y a qu'un paramètre autorégressif, il doit se trouver dans l'intervalle $-1 < \varphi_1 < +1$. Dans les autres cas, les effets passés s'accumuleraient et les valeurs successives des x_t se déplaceraient infiniment vers l'avant, ce qui signifie que la série ne serait pas stationnaire. S'il y a plus d'un paramètre autorégressif, des restrictions similaires (générales) sur les valeurs des paramètres peuvent être posées (**Box et Jenkins, 1976 ; Montgomery, 1990**).

II.2.3. Moyenne mobile

Les modèles à moyenne mobile suggèrent que la série présente des fluctuations autour d'une valeur moyenne. On considère alors que la meilleure estimation est représentée par la moyenne pondérée d'un certain nombre de valeurs antérieures (ce qui est le principe des procédures de moyennes mobiles utilisées pour le lissage des données). Ceci revient en fait à considérer que l'estimation est égal à la moyenne vraie, auquel on ajoute une somme pondérée des erreurs ayant entaché les valeurs précédentes :

$$y_t = \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \theta_3 \varepsilon_{t-3} + \varepsilon_t$$

Littérairement, chaque observation est composée d'une composante d'erreur aléatoire (choc aléatoire, ε) et d'une combinaison linéaire des erreurs aléatoires passées. θ_1 , θ_2 et θ_3 sont les coefficients de moyenne mobile du modèle.

Comme précédemment cette équation porte soit sur les données brutes, soit sur les données différenciées si une différenciation a été nécessaire. Pour un modèle ARIMA (0,1,1) on aura :

$$y_t - y_{t-1} = \mu - \theta \varepsilon_{t-1} + \varepsilon_t$$

Ce qui peut également être écrit:

$$y_t = \mu + y_{t-1} - \theta \varepsilon_{t-1} + \varepsilon_t$$

Un modèle de moyenne mobile correspond à des séries exhibant des fluctuations aléatoires autour d'une moyenne variant lentement. Plutôt que de prendre comme précédemment la valeur précédente comme prédicteur, on utilise une moyenne de quelques observations précédentes, de manière à éliminer le bruit, et estimer plus précisément la

moyenne locale. Cette logique correspond au lissage exponentiel simple, qui considère chaque observation comme la résultante d'une constante b et d'un terme d'erreur ε , soit :

$$y_t = b + \varepsilon_t$$

La constante b est relativement stable sur chaque segment de la série, mais peut se modifier lentement au cours du temps. Si ce modèle est approprié, l'une des manières d'isoler la réelle valeur de b , et donc la partie systématique ou prévisible de la série, consiste à calculer une sorte de moyenne mobile, où les observations courantes et immédiatement précédentes « les plus récentes » ont une pondération plus forte que les observations plus anciennes. C'est exactement ce que fait un lissage exponentiel simple, où les pondérations les plus faibles sont affectées exponentiellement aux observations les plus anciennes. La formule spécifique de lissage exponentiel simple est (*Box et Jenkins, 1976 ; Montgomery, 1990*):

$$y_t = \alpha \hat{y}_t - (1-\alpha) y_{t-1}$$

Lorsqu'on l'applique de façon récurrente à chaque observation successive de la série, chaque nouvelle valeur prédite est calculée comme la moyenne pondérée de l'observation courante et de l'observation précédente prédite ; la précédente observation prédite était-elle-même calculée à partir de la valeur (précédente) observée et de la valeur prédite avant cette valeur (précédente), et ainsi de suite. Par conséquent, chaque valeur prédite est une moyenne pondérée des observations précédentes, où les poids décroissent exponentiellement selon la valeur des paramètres α . Si α est égal à 1 les observations précédentes sont complètement ignorées ; si α est égal à 0, l'observation courante est totalement ignorée, et la valeur prédite ne porte que sur les valeurs prédites précédentes (qui est calculée à partir de l'observation lissée qui lui précède, et ainsi de suite ; c'est pourquoi toutes les valeurs prédites auront la même valeur que la valeur initiale \hat{y}_0). Les valeurs intermédiaires de α produiront des résultats intermédiaires (noter que la valeur $1-\alpha$ correspond au θ des équations précédentes) (*Brigitte et al., 2008*).

On peut également envisager des modèles mixtes: par exemple un modèle ARIMA (1,1,1) aura l'équation de prédiction suivante:

$$y_t = \mu + y_{t-1} + \varphi(y_{t-1} - y_{t-2}) - \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

Néanmoins on préfère généralement utiliser de manière exclusive les termes AR ou MA.

II.2.4. Les paramètres du modèle ARIMA

L'objectif essentiel des modèles ARIMA est de permettre une prédiction de l'évolution future d'un phénomène. Son développement dans le domaine de l'économétrie est basé sur ce principe. On en verra plus loin une illustration.

Un autre intérêt, peut-être plus essentiel en ce qui concerne la recherche scientifique, est de comprendre la signification théorique de ces différents processus. Il est clair cependant que cette interprétation dépend de la nature du phénomène étudié, et des modèles dont le chercheur dispose pour en rendre compte.

- Un processus non différencié à bruit blanc ARIMA (0,0,0) suggère des fluctuations aléatoires autour d'une valeur de référence. Cette valeur de référence peut être

considérée comme une caractéristique stable du système étudié (trait de personnalité, mémoire, capacité stabilisée, etc.)

- Un processus de moyenne mobile suggère que la valeur de référence évolue d'une mesure à l'autre. Plus précisément, la valeur de référence est fonction de la valeur de référence précédente et de l'erreur ayant entaché la mesure précédente.
- Un processus autorégressif suggère que le phénomène étudié n'est pas déterminé par une valeur de référence. C'est la performance précédente (ou les performances précédentes) qui déterminent entièrement la performance présente.

Par exemple, Spray et Newell en 1986 analysent des données tirées d'une expérimentation portant sur le rôle de la connaissance des résultats dans l'apprentissage. Les sujets réalisent 77 essais dans une tâche manuelle. Le protocole comprenait plusieurs groupes, différenciés par des combinaisons spécifiques d'essais avec ou sans connaissance des résultats. Notamment, certains sujets disposaient de connaissance des résultats tout au long des 77 essais, pour d'autre la CR était supprimée au-delà de 17, 32 ou 52 essais. Un groupe n'avait pas de CR du tout (*Newell, 1986 ; Spray, 1986*).

Les résultats de la modélisation montrent que les séries avec CR (ou les portions de séries avec CR) peuvent être représentée par des processus à bruit blanc du type: $y_t = \mu + \varepsilon_t$

C'est-à-dire un modèle ARIMA (0,0,0). Cette équation suggère donc que les performances successives oscillent de manière aléatoire autour d'une valeur moyenne, sorte de référence interne construite par la connaissance des résultats.

Les séries sans CR (ou les portions de série sans CR) sont quant à elles modélisées selon un ARIMA (0, 1, 1) selon la formule:

$$y_t = \mu - \theta_1 \varepsilon_{t-1} + \varepsilon_t \text{ OU } y_t = r_t + \varepsilon_t$$

r_t représentant la valeur de référence, qui cette fois change à chaque essai. On peut dériver du modèle que $r_t = r_{t-1} - \theta_1 \varepsilon_{t-1}$

C'est-à-dire que la référence est une combinaison de la référence précédente et de l'erreur ayant entaché l'essai précédent. Ce modèle indique clairement que l'essai en cours est influencé par l'essai précédent, ce qui n'était pas le cas dans les essais avec CR.

L'analyse des données de Diggle (1977) suggère que la référence précédente est plus importante que la performance actuelle (*Diggle, 1977*).

On peut noter que pour les sujets ayant bénéficié de la CR durant 52 essais sur 77, la série demeure stationnaire et à bruit blanc jusqu'à la fin de l'expérimentation.

II.2.5. Termes AR

Après que la série ait été stationnaire, l'étape suivante consiste à identifier les termes AR et MA nécessaires pour corriger les auto- corrélations résiduelles. Cette analyse est basée sur l'examen des fonctions d'auto-corrélation et d'auto-corrélation partielle. Rappelons que l'auto-corrélation est la corrélation d'une série avec elle-même, selon un décalage (lag) défini.

L'auto-corrélation de décalage 0 est par définition égale à 1. La fonction d'auto-corrélation fait correspondre à chaque décalage l'auto-corrélation correspondante.

D'une manière générale, une corrélation partielle entre deux variables est la quantité de corrélations qui n'est pas expliquée par les relations de ces variables avec un ensemble spécifié d'autres variables. Supposons par exemple que l'on réalise la régression de Y sur trois variables X_1 , X_2 et X_3 . La corrélation partielle entre Y et X_3 contrôlant X_1 et X_2 est la quantité de corrélation entre Y et X_3 qui n'est pas expliqué par leurs relations communes avec X_1 et X_2 . Elle peut être calculée comme la racine carrée du gain de variance expliquée obtenu en ajoutant X_3 à la régression de Y sur X_1 et X_2 .

Dans le cas des séries temporelles, la corrélation partielle de décalage k est la corrélation entre y_t et y_{t-k} , contrôlant l'influence des $k-1$ valeurs interposées.

L'auto-corrélation de décalage 1 est la corrélation entre y_t et y_{t-1} . On suppose que c'est également la corrélation entre y_{t-1} et y_{t-2} . Si y_t et y_{t-1} sont corrélés, et que y_{t-1} et y_{t-2} le sont également, on peut supposer qu'une corrélation sera présente entre y_t et y_{t-2} . C'est-à-dire que la corrélation de décalage 1 se propage au décalage 2 et sans doute aux décalages d'ordre supérieurs. Plus précisément, la corrélation attendue au décalage 2 est le carré de la corrélation observée au décalage 1. L'auto-corrélation partielle de décalage 2 est donc la différence entre l'auto-corrélation de décalage 2 et la corrélation attendue due à la propagation de la corrélation de décalage 1.

Si l'on revient à la fonction d'auto-corrélation de l'exemple précédent (avant différenciation), on peut supposer que la présence d'auto-corrélations fortes pour un grand nombre de décalages successifs est lié à ce phénomène de propagation. Ceci est confirmé par l'examen de la fonction d'auto-corrélation partielle, qui n'a qu'une valeur significative au décalage 1 (notons que l'auto-corrélation partielle de décalage 1 est égale à l'auto-corrélation correspondante, aucune valeur n'étant intercalée).

Ceci indique clairement que les auto-corrélations de décalage supérieur à 1 ne sont dues qu'à la propagation de l'auto-corrélation de décalage 1.

Une extinction brutale de l'auto-corrélation partielle associée à un déclin plus progressif de l'auto-corrélation constitue la signature caractéristique d'un processus autorégressif. Plus particulièrement, l'auto-corrélation partielle de décalage k est égale au coefficient $AR(k)$ estimé dans un modèle contenant k termes AR . On pourrait d'ailleurs déterminer les coefficients AR par régression multiple, en prédisant $(y_t - y_{t-1})$ à partir de k échantillons représentant les k décalages.

Le décalage auquel l'auto-corrélation partielle disparaît indique le nombre de termes auto régressifs à inclure.

Généralement, ce pattern est associé à une auto-corrélation de décalage 1 positive, signe que la série demeure sous-différenciée. Une légère sous-différenciation peut donc être compensée par l'ajout d'un terme autorégressif.

On peut noter que l'auto-corrélation de décalage 1 est significative et positive, et que l'auto-corrélation partielle présente une extinction plus brutale que l'auto-corrélation. L'auto-corrélation partielle ne présente en fait que deux pics significatifs, alors que l'auto-corrélation en présente quatre. La série différenciée présente donc une signature de processus autorégressif d'ordre 2.

Les auto-corrélations au décalage 1 et 2 ont été supprimées, et aucune auto-corrélation n'apparaît aux décalages supérieurs. L'analyse montre que les coefficients AR sont significativement différentes de 0, et l'écart-type des résidus a été réduit de 10% (de 1.54 à 1.42 par l'ajout des termes AR). L'équation de prédiction a la forme suivante:

$$y_t = \mu + y_{t-1} + \varphi_1(y_{t-1} - y_{t-2}) + \varphi_2(y_{t-2} - y_{t-3}) \quad (19)$$

Cette équation permet la prédiction des données ultérieures à la période dite de validation du modèle.

II.2.6. Termes MA

La fonction d'auto-corrélation joue pour les processus de moyenne mobile le même rôle que la fonction d'auto-corrélation partielle pour les processus autorégressifs. Si l'auto-corrélation est significative au décalage k mais plus au décalage $k+1$, ceci indique que k termes de moyenne mobile doivent être ajoutés au modèle.

A noter que si les coefficients AR peuvent être estimés par une analyse en régression multiple, une telle démarche est impossible pour les coefficients MA. D'une part, parce que l'équation de prédiction est non-linéaire, et d'autre part les erreurs ne peuvent être spécifiées en tant que variable indépendantes. Les erreurs doivent être calculées pas à pas en fonction des estimations courantes des paramètres.

Une signature MA est généralement associée à une auto-corrélation négative au décalage 1, signe que la série est sur-différenciée. Une légère sur-différenciation peut donc être compensée par l'ajout d'un terme de moyenne mobile.

Ces fonctions sont typiques de la signature des processus de moyenne mobile. Le pic unique et négatif pour la fonction d'auto-corrélation indique un processus MA(1). Le modèle pertinent serait donc un ARIMA(0, 2, 1). L'équation de prédiction serait la suivante:

$$y_t = 2y_{t-1} - y_{t-2} - \theta_1 e_{t-1}$$

Ces deux modèles peuvent ajuster de manière alternative la série de départ. Sachant que les termes AR peuvent compenser une légère sous différenciation, et les termes MA une légère sous-différenciation, il est courant que deux modèles alternatifs soient possibles: un premier avec 0 ou 1 ordre de différenciation combiné avec des termes AR, et un autre avec le niveau de différenciation supérieur, combiné à des termes MA. Le choix d'un ou l'autre modèle peut reposer sur des présupposés théoriques liés au phénomène observé.

Les outils principaux utilisés lors de la phase d'identification sont donc les tracés de la série, les corrélogrammes d'auto-corrélation (FAC), et d'auto-corrélation partielle (FACP). La décision n'est pas simple et les cas les plus atypiques requièrent, outre l'expérience, de nombreuses expérimentations avec des modèles différents (avec divers paramètres

ARIMA). Toutefois, les composantes des séries chronologiques empiriques peuvent souvent être assez bien approchées en utilisant l'un des 5 modèles de base suivants, identifiables par la forme de l'autocorrélogramme (FAC) et de l'autocorrélogramme partiel (FACP). Puisque le nombre de paramètres (à estimer) de chaque type dépasse rarement 2, il est souvent judicieux d'essayer des modèles alternatifs sur les mêmes données (*Box et Jenkins, 1976 ; Montgomery, 1990*).

- **Un paramètre autorégressif (p)** : FAC - décomposition exponentielle ; FACP - pic à la période 1, pas de corrélation pour les autres périodes.
- **Deux paramètres autorégressifs (p)** : FAC - une composante de forme sinusoïdale ou un ensemble de décompositions exponentielles ; FACP - pics aux périodes 1 et 2, aucune corrélation pour les autres périodes.
- **Un paramètre de moyenne mobile (q)** : FAC - pic à la période 1, aucune corrélation pour les autres périodes ; FACP - exponentielle amortie.
- **Deux paramètres de moyenne mobile (q)** : FAC - pics aux périodes 1 et 2, aucune corrélation pour les autres périodes ; FACP - une composante de forme sinusoïdale ou un ensemble de décompositions exponentielles.
- **Un paramètre autorégressif (p) et un de moyenne mobile (q)** : FAC - décomposition exponentielle commençant à la période 1 ; FACP - décomposition exponentielle commençant à la période 1.

II.2.7. Racines unitaires

Si une série est largement sur- ou sous-différenciée (c'est-à-dire qu'un ordre complet de différenciation doit être ôté ou ajouté), on obtient généralement une racine unitaire (*unit root*) au niveau des coefficients AR ou MA.

Dans le cas d'un modèle ARIMA (1,k,0), AR(1) est égal à 1. Dans le cas plus général d'un modèle ARIMA (p,k,0), la somme des coefficients AR est égale à 1 (c'est-à-dire non significativement différente de 1). On peut considérer qu'alors les termes AR simulent une différence (on a vu dans l'exemple précédent que le premier terme autorégressif pouvait être substitué à la première différence). Le nombre de termes AR doit être réduit d'une unité et on doit accroître d'une unité l'ordre de différenciation (*Goodwin et al., 2009 ; Green et al., 2007, Harris et al., 2003, Hyndman et al., 2006*).

Le même raisonnement peut être tenu pour les coefficients MA. Dans le cas d'un modèle ARIMA (0,k,1), MA(1) est égal à 1, et d'une manière générale la somme des coefficients MA est égale à 1. Les termes MA dans ce cas tendent à annuler exactement une différence. Le nombre de termes MA doit être réduit d'une unité et on doit réduire d'une unité l'ordre de différenciation.

II.2.8. Evaluation des modèles

L'objectif de la modélisation est de déterminer combien de paramètres autorégressifs (p) et de moyennes mobiles (q) sont nécessaires pour obtenir un modèle effectif et parcimonieux du processus (parcimonieux signifie qu'il s'agit, parmi tous les modèles possibles, du modèle possédant le moins de paramètres et le plus grand nombre de degrés de

liberté pour ajuster les données). En pratique, le nombre de paramètres p ou q dépasse très rarement 2.

Estimation des Paramètres: l'analyse produit les valeurs approchées de t , calculées à partir des erreurs-types des paramètres. Si cette valeur de t n'est pas significative, le paramètre respectif peut être éliminé du modèle, sans affecter dans la plupart des cas l'ajustement global du modèle (*Goodwin et al., 2009 ; Green.et al., 2007, Harris et al., 2003, Hyndman et al., 2006*).

Un autre indicateur simple et courant pour mesurer la fiabilité du modèle est de comparer la précision des estimations basées sur des données partielles, afin que les prévisions puissent être comparées avec les dernières observations connues de la série (initiale).

Toutefois, un bon modèle ne doit pas seulement produire des prévisions suffisamment précises, il doit également être parcimonieux et produire des résidus statistiquement indépendants, ne contenant que du bruit, sans aucune composante régulière (par exemple, le corrélogramme des résidus ne doit pas révéler d'auto-corrélations). Un bon test du modèle consiste :

- à tracer les résidus et à les examiner pour voir s'il existe des tendances systématiques, et
- à étudier l'auto-corrélogramme des résidus (il ne doit pas y avoir d'auto-corrélations entre les résidus).

Ce que l'on cherche à savoir ici, c'est si les résidus ne sont pas distribués de façon systématique dans la série (par exemple, systématiquement négatifs dans une première partie de la série puis proches de zéro dans une seconde) ou s'ils ne sont pas autocorrélés ce qui tendrait à montrer que le modèle ARIMA est inadéquat. L'analyse des résidus de l'ARIMA constitue un test important du modèle. La procédure d'estimation postule que les résidus ne sont pas auto-corrélés et qu'ils sont distribués normalement (*Goodwin et al., 2009 ; Green.et al., 2007, Harris et al., 2003, Hyndman et al., 2006*).

La méthode ARIMA n'est appropriée que lorsque la série chronologique est stationnaire (c'est-à-dire que les moyennes, variances, et auto-corrélations doivent être sensiblement constantes au cours du temps). Il est également recommandé d'avoir au moins 50 observations dans le fichier de données. Les paramètres estimés sont considérés constants dans toute la série.

II.2.9. Modèles saisonniers

L'ARIMA saisonnière multiplicative est une généralisation et une extension de la méthode présentée aux paragraphes précédents, pour des séries où le même phénomène se répète au cours du temps, à chaque saison. Outre les paramètres non saisonniers, des paramètres saisonniers pour un décalage spécifié (repéré lors de la phase d'identification) doivent être estimés. Comme pour les paramètres de l'ARIMA simple, ce sont : les paramètres autorégressifs saisonniers (ps), de différenciation saisonnière (ds), et de moyenne mobile saisonnière (qs). Par exemple, le modèle $(0,1,2) (0,1,1)$ décrit un modèle sans

paramètre autorégressif, avec 2 paramètres de moyenne mobile normaux et 1 paramètre de moyenne mobile saisonnier. Le calcul de ces paramètres intervient après différenciation une fois de 1 période, et une fois saisonnièrement. Le décalage saisonnier utilisé pour les paramètres saisonniers est souvent déterminé lors de la phase d'identification et doit être explicitement spécifié (*Goodwin et al., 2009 ; Green et al., 2007, Harris et al., 2003, Hyndman et al., 2006*)

Les recommandations générales concernant la sélection des paramètres à estimer (basés sur les FAC et FACP) s'appliquent également aux modèles saisonniers (pour une illustration d'une ARIMA saisonnière, voir l'Exemple dans la section ARIMA du chapitre). La principale différence est que dans les séries saisonnières, les FAC et FACP auront des coefficients assez importants pour de nombreuses périodes saisonnières (en plus de leur composante globale reflétant la composante non saisonnière de la série).

II.2.10. Séries chronologiques interrompue

Une question qui se pose fréquemment dans les analyses de séries chronologiques est de savoir si un événement extérieur (exogène) a perturbé les observations ultérieures. Par exemple, la mise en œuvre d'une nouvelle politique économique a-t-elle permis d'améliorer les performances économiques ; une nouvelle loi anti-crime a-t-elle permis d'infléchir les taux d'agression suivants ; et ainsi de suite. En général, nous souhaiterions pouvoir évaluer l'impact d'un ou plusieurs événements discrets sur les valeurs de la série chronologique. Ce type d'analyse de série chronologique interrompue est décrit en détail par *McDowall, McCleary, Meidinger, et Hay (1980)*. McDowall distingue entre trois grands types d'impact possibles : soudain et permanent, permanent et graduel, soudain et temporaire.

Impact Soudain et Permanent : Un impact soudain et permanent implique simplement que la moyenne globale de la série chronologique se trouve modifiée après l'intervention ; le déplacement global est noté ω .

Impact Graduel et Permanent : La structure d'impact graduel et permanent implique que l'augmentation ou la diminution due à l'intervention soit progressive, et que l'impact permanent final ne devienne évident qu'après un certain temps :

$$Impact_t = \delta \cdot Impact_{t-1} + \omega \text{ (pour tout } t \text{ période d'impact, sinon } = 0 \text{).}$$

Cette structure d'impact est définie par deux paramètres δ (delta) et ω (oméga). Si δ est proche de 0 (zéro), l'impact permanent final deviendra évident après quelques observations supplémentaires ; si δ est proche de 1, l'impact permanent final de l'impact deviendra évident après de nombreuses observations supplémentaires. Tant que le paramètre δ demeure supérieur à 0 et inférieur à 1 (les limites de stabilité du système), l'impact sera graduel et produira une modification (déplacement) asymptotique de la moyenne globale par la quantité : $\omega/(1-\delta)$

Les programmes d'analyse calculent automatiquement la modification asymptotique des impacts graduels et permanents. Notez que lors de l'évaluation d'un modèle, il est important que les deux paramètres soient statistiquement significatifs ; sinon, on pourrait aboutir à des conclusions paradoxales. Par exemple, supposons que le paramètre ω ne soit pas significativement différent de 0 (zéro) mais que le paramètre δ le soit ; cela voudrait dire

qu'une intervention aurait causé une modification graduelle significative, sans que le résultat ne soit significativement différent de zéro (*Goodwin et al., 2009 ; Green et al., 2007, Harris et al., 2003, Hyndman et al., 2006*).

Impact Soudain et Temporaire : Un impact soudain et temporaire implique que l'augmentation ou la diminution initiale soudaine due à l'intervention s'estompe lentement, sans modification permanente de la moyenne de la série. Ce type d'intervention peut être synthétisé par les expressions :

- Avant l'intervention : $\text{Impact}_t = 0$
- Pendant l'intervention : $\text{Impact}_t = \omega$
- Après l'intervention : $\text{Impact}_t = \delta \cdot \text{Impact}_{t-1}$

Notez que cette structure d'impact est à nouveau définie par les deux paramètres δ (delta) et ω (oméga). Tant que le paramètre δ demeurera supérieur à 0 et inférieur à 1 (les limites de stabilité du système), l'impact initial soudain se décomposera graduellement. Si δ est proche de 0 (zéro) la décomposition sera très rapide, et l'impact aura complètement disparu après quelques observations seulement. Si δ est proche de 1 la décomposition sera lente, et l'intervention continuera à affecter la série sur de nombreuses observations. Notez que lors de l'évaluation d'un modèle, il est encore une fois important que les deux paramètres soient statistiquement significatifs ; sinon on pourrait aboutir à des conclusions paradoxales. Par exemple, supposons que le paramètre ω ne soit pas significativement différent de 0 (zéro) mais que le paramètre δ le soit ; cela voudrait dire qu'une intervention n'aurait pas causé de modification initiale soudaine, tout en montrant une décomposition significative (*Goodwin et al., 2009 ; Green et al., 2007, Harris et al., 2003, Hyndman et al., 2006*).

II.3. Les réseaux de neurones artificiels

II.3.1. Historique

Bien que le concept de la mémoire associative était introduite pour la première fois par le célèbre psychologue américain W. James en 1890 (*James, 1890*), qui propose ce qui deviendra une loi de fonctionnement pour l'apprentissage sur les réseaux de neurones connues plus tard sous le nom de la loi de Hebb, l'étude des réseaux de neurones, proprement dite, date des années 1940-50 (*Hebb, 1949*). Héritière de la révolution scientifique qui mobilisait les esprits à cette période, cette discipline trouve ses sources à la fois dans les connaissances neurobiologiques de l'époque et dans un nouvel outil permettant d'explorer par simulation le comportement des modèles théorique proposés, l'ordinateur (*Jodouin, 199, Krauss, 1997*).

Alan Turing et John Von Neumann étaient parmi les premiers à s'être intéressés au cerveau humain. Ils ont l'idée d'une machine basée sur le modèle de neurone artificiel proposé en 1943 par le neurophysiologiste Warren McCulloch et le logicien Walter Pitts qu'on décrira plus loin. Bien qu'il soit essentiellement connu pour la machine qui porte son nom, Alan Turing est aussi un des pères des réseaux de neurones : dès 1948 il les étudie dans un article intitulé *Intelligent Machinery* (publié en 1968, 14 ans après sa mort) où il décrit les principes du connexionnisme (*Schmitt, 2001*). Il introduit une machine constituée de neurones artificiels associée à des dispositifs permettant de modifier leurs connexions. La publication tardive de ces travaux fait que l'on ne redécouvre que plus tard les réseaux de neurones.

En 1949, Hebb souligna l'importance du couplage synaptique dans le processus d'apprentissage en se fondant son argumentation sur le comportement psychophysique; première règle d'apprentissage. Ainsi, les premiers réseaux de neurones artificiels basés sur la vision des systèmes biologiques apparaissent sous le terme perceptron proposé en 1958 par Frank Rosenblatt (*McCulloch et al, 1943*). Inspiré des idées de Hebb, Mc Culloch et Pitts, le perceptron était en mesure d'apprendre à calculer certaines fonctions logiques par l'exemple en modifiant ses connexions synaptiques.

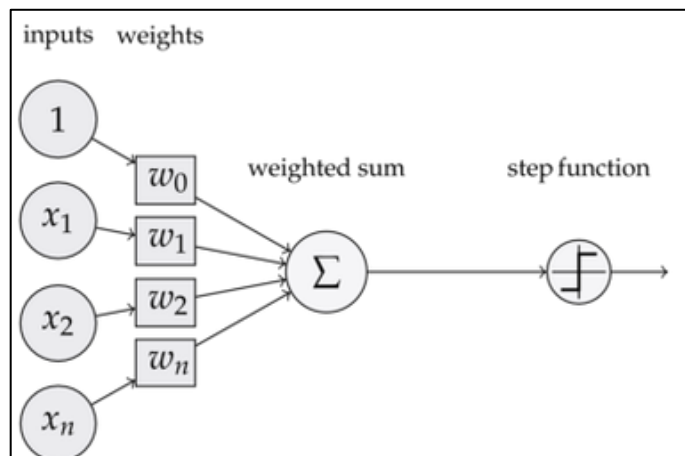


Figure II.3: Neurone formel de mc culloch et pitts
(<https://blog.knoldus.com/introduction-to-perceptron-neural-network>)

Cependant, certains tels Wilkie, Stonham et Aleksander persistent et parviennent en 1980 à résoudre les problèmes d'apprentissage difficiles en utilisant, malgré tout, des réseaux de neurones multicouches. D'autres études telles celles de Grossberg et Kohonen en 1982 contribuent elles aussi au regain d'intérêt pour les réseaux de neurones formels et leurs dérivés. Dans la même période Werbos invente l'idée de rétro-propagation en 1974 (*Grossberg, 1982 ; Kohonen, 1982*).

II.3.2. Modèle mathématique

Les réseaux de neurones biologiques réalisent facilement un certain nombre d'applications telles que la reconnaissance de formes, le traitement du signal, l'apprentissage par l'exemple, la mémorisation, la généralisation. Ces applications sont pourtant, malgré tous les efforts déployés en algorithmique et en intelligence artificielle, à la limite des possibilités actuelles.

C'est à partir de l'hypothèse que le comportement intelligent émerge de la structure et du comportement des éléments de base du cerveau que les réseaux de neurones artificiels se sont développés. Les réseaux de neurones artificiels sont des modèles, à ce titre ils peuvent être décrit par leurs composants, leurs variables descriptives et les interactions des composants (*Touzet, 1992*).

Le modèle mathématique le plus souvent utilisé est celui de Mc Culloch et Pitts représenté sur la figure II.3. On y reconnaît les divers éléments associés à un neurone biologique: (*Le Fablec, 1999*) figure II.3

- Des synapses que nous appellerons par la suite connexions sur lesquelles arrivent les entrées x_i
- Le poids : chaque signal x_i passant sur une connexion sera multiplié par le coefficient de pondération w_i
- Un mécanisme de sommation permettant de faire une combinaison linéaire des entrées pondérées précédentes
- Un biais b qui représente le seuil au-dessus duquel le neurone sera excité. Il sert à centrer la zone d'action du neurone
- Une sortie y qui peut être utilisée comme entrée pour d'autres neurones
- Une fonction d'activation f qui sert à limiter sa sortie et permet d'agir sur la forme de la zone sur laquelle agit le neurone

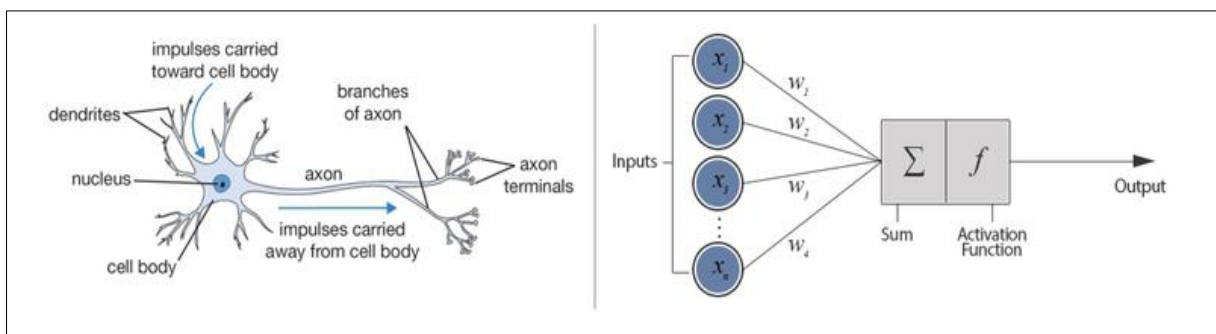


Figure II.4: La correspondance entre le neurone biologique et le neurone artificiel

Un neurone réel produit une impulsion à partir d'influx afférents provenant d'autres neurones. De la même manière, le neurone artificiel produit une sortie associée à un ensemble d'entrées : la sortie y est calculée à partir des entrées x_i avec :

$$f = \sum_{i=1}^n (x_i \cdot w_i + b) \quad (21)$$

A partir de la somme pondérée calculée à l'entrée, une fonction de transfert calcule la valeur de l'état du neurone. C'est cette valeur qui sera transmise aux neurones avals. Il existe de nombreuses formes possibles pour la fonction de transfert, les plus courantes sont présentées sur la figure II.4.

On remarquera qu'à la différence des neurones biologiques dont l'état est binaire, la plupart des fonctions de transfert sont continues, offrant une infinité de valeurs possibles comprises dans l'intervalle $[0, +1]$ (ou $[-1, +1]$).

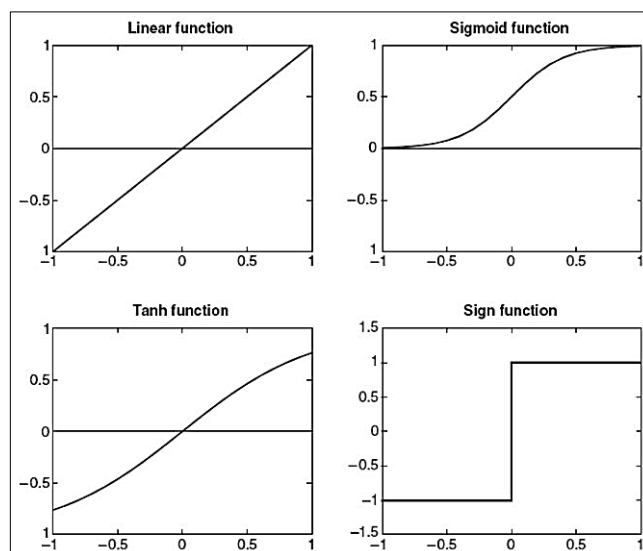


Figure II.5: Fonctions de transfert d'un neurone artificiel
(<https://scilogs.fr/intelligence-mecanique>)

II.3.2.1. Limites d'un neurone formel

Ce modèle n'a pas pour rôle de représenter toutes les caractéristiques du neurone biologique:

- Le processus de transport par l'axone et la modulation synaptique sont cruellement réduit à une simple multiplication de deux vecteurs $x.w$
- La dynamique électrochimique complexe, à la source de la propagation du signal, est ramenée à sa plus simple expression par une fonction statique f
- Il n'y a pas de dynamique interne au sein de ce modèle, pour des poids synaptiques fixes, le neurone formel accomplit une transformation statique.

Dans un sens, ce modèle ne capture que les propriétés spatiales du neurone biologique et passe outre les relations temporelles. La capacité représentative du neurone formel est fortement restreinte en raison de la forme rudimentaire de la fonction f utilisée (seuil), (*Yahiaoui et al., 2002*).

En résumé, les réseaux de neurones artificiels sont des réseaux fortement connectés de processeurs élémentaires fonctionnant en parallèle. Chaque processeur élémentaire calcule une sortie unique sur la base des informations qu'il reçoit. Toute structure hiérarchique de réseaux est évidemment un réseau. La capacité de traitement d'un réseau est stockée sous forme de *poids d'interconnexions* obtenus par un processus *d'adaptation* ou *d'apprentissage* à partir d'un ensemble d'exemples d'apprentissage.

II.3.3. Propriétés des réseaux de neurones

Un réseau de neurones se compose de neurones qui sont interconnectés de façon à ce que la sortie d'un neurone puisse être l'entrée d'un ou plusieurs autres neurones. Ensuite il y a des entrées de l'extérieur et des sorties vers l'extérieur (*Rumelhurt et al., 1986*).

Huit composants principaux d'un réseau de neurones :

- Un ensemble de neurones
- Un état d'activation pour chaque neurone (actif, inactif, ...)
- Une fonction de sortie pour chaque neurone $f(S)$
- Un modèle de connectivité entre les neurones (chaque neurone est connecté à tous les autres, par exemple)
- Une règle de propagation pour propager les valeurs d'entrée à travers le réseau vers les Sorties
- Une règle d'activation pour combiner les entrées d'un neurone (très souvent une somme pondérée)
- Une règle d'apprentissage
- Un environnement d'opération

Le comportement d'un réseau et les possibilités d'application dépendent complètement de ces huit facteurs et le changement d'un seul d'entre eux peut changer le comportement du réseau complètement.

Les réseaux de neurones sont souvent appelés des "boîtes noires" car la fonction mathématique qui est représentée devient vite trop complexe pour l'analyser et la comprendre directement. Cela est notamment le cas si le réseau développe des représentations distribuées (*Rumelhart et al., 1986*), c'est à dire que plusieurs neurones sont plus ou moins actifs et contribuent à une décision. Une autre possibilité est d'avoir des représentations localisées, ce qui permet d'identifier le rôle de chaque neurone plus facilement. Les réseaux de neurones ont quand même une tendance à produire des représentations distribuées (*Framling, 1992*).

II.3.3.1. La non linéarité

Cette notion est extrêmement importante puisque le fonctionnement des réseaux neuronaux ressemble beaucoup au fonctionnement d'autre système, si l'on élimine la fonction d'amplification non linéaire de chaque neurone. Cette non-linéarité intrinsèque, et parfois récurrente, autorise les réseaux neuronaux à modéliser des associations très complexes.

La non linéarité de type seuil est génératrice de capacité de généralisation : cela est simple à concevoir intuitivement, puisque la sortie seuillée d'un neurone ne peut prendre que deux valeurs alors que cette même sortie non seuillée est multivaluée. En vocabulaire mathématique, on l'exprime par le fait que le neurone de Mc Culloch et Pitts sépare l'espace d'entrée par un hyper plan. Cela est une capacité intrinsèque de généralisation (*Yahiaoui et al., 2002*)

II.3.4. Architecture des réseaux de neurones artificiels

II.3.4.1. Les réseaux monocouche : le perceptron

Introduit par Rosenbalt en 1958, sur la base du modèle du neurone formel de Mc Culloch et Pitts en 1943 ainsi de la première loi de Hebb en 1949, le perceptron représente le premier réseau de neurones artificiels (*Hebb, 1949*). Il est très utilisé pour le classement comme il est introduit dans les réseaux traitant la décision. Il se compose de deux couches de neurones formels appelées « rétine » et « couche de sortie » (*Jauffert, 2002*).

La couche rétinée est une contraction de la couche d'entrée et de la couche de traitement des informations. Cette contraction a été effectuée pour des raisons de simplification, car la couche d'entrée est une couche inerte qui ne participe pas directement au traitement de l'information. Les cellules de la rétine, lorsqu'elles sont excitées, répondent par « oui » ou par « non » (1 ou 0) à leur sortie, car elles sont régies par une fonction binaire à seuil (*Davalo, 1991*).

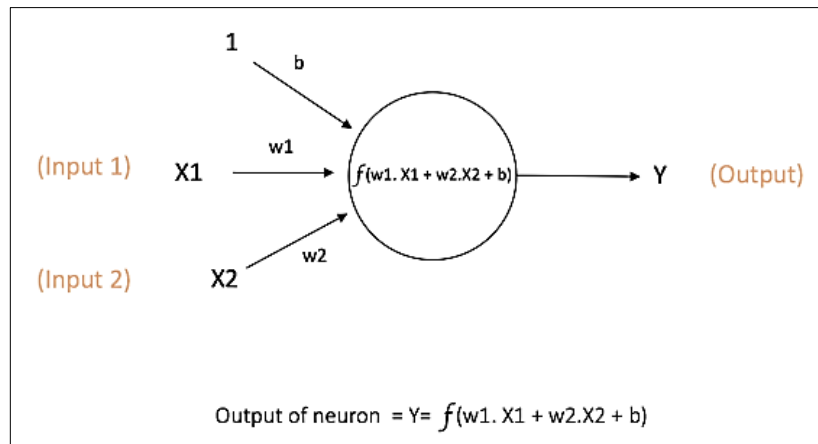


Figure II.6: Réseau de neurones monocouche : le perceptron
(<https://ujjwalkarn.me/2016/08/09/quick-intro-neural-networks>)

La fonction de transfert utilisée dans ce type de réseau est la fonction à seuil (hardlimit). La modification des poids du perceptron est effectuée avec un apprentissage de type supervisé en exploitant la loi d'apprentissage de Widrow-Hoff (*Lourakis & Manolis, 2005*).

II.3.4.2. Les réseaux multicouches

Contrairement aux deux modèles de réseaux de neurones précédents, les réseaux multicouches sont constitués de plusieurs couches disposées l'une après l'autre. La fonction d'activation utilisée dans ce type de réseau est généralement de type sigmoïde, ce qui leur donne un caractère non linéaire. L'apprentissage dans ce cas consiste à modifier des connexions ayant un nombre très grand, du fait de la multiplicité des couches ainsi que celle des neurones dans chaque couche. Après l'intégration de l'algorithme d'apprentissage de la rétro-propagation (back-propagation) dans les réseaux de neurones, la tâche du traitement de problèmes très complexes est devenue moins délicate.

La topologie de ce type de réseaux est telle qu'avant d'arriver à une sortie, chaque unité d'entrée est suivie par une succession de couches cachées et chaque unité d'une couche donnée reçoit son entrée à partir de la couche précédente seulement (*Mokhnache, 2004*).

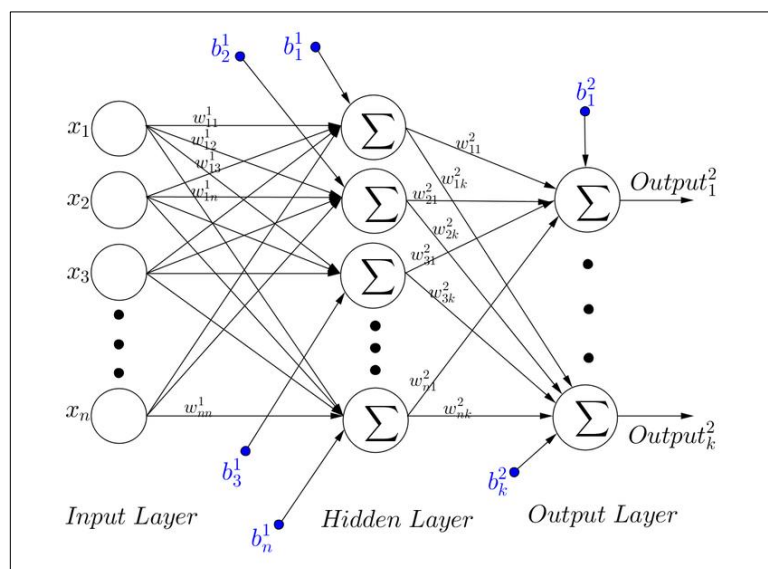


Figure II.7: Réseau de neurones multicouches
(https://www.researchgate.net/figure/Multilayer-neural-network-architecture_fig1_303875613)

II.3.5. Structure des réseaux de neurones artificiels

Définir l'architecture d'un RNA revient à faire un choix judicieux du vecteur d'entrée, de la taille du réseau (ou nombre total de couches et de neurones), de la structure du réseau (c.-à-d. le type d'interconnexion entre les couches de neurones) et des fonctions de transfert, sans disposer de procédure standard (à présent). Le choix de la taille du réseau incombe à l'utilisateur. Il n'existe pas dans la littérature d'éléments suffisants pour déterminer objectivement le nombre de neurones requis pour résoudre un problème donné. Les approches heuristiques proposées par différents auteurs (*Hecht-Nielsen, 1987 ; Baum et Haussler, 1989 ; Widrow et Lehr, 1990 ; Zurada, 1992 ; Murata et al., 1994 ; Kohavi, 1995 ; Jacobs et al., 1997 ; Rudolph 1997*) sont toutes expérimentales et ne s'appliquent qu'à des cas particuliers. La définition de la taille du réseau et de sa structure sont deux étapes difficiles de la construction d'un RNA (*Rumelhurt et al., 1986 ; Hassoum, 1995 ; Coulibaly et al., 1999*).

Selon leur structure, on peut classer les RNA en deux grandes catégories :

- Les réseaux « feedforward » (statiques)
- Les réseaux « feedback » (dynamiques ou récurrents)

II.3.5.1. Les réseaux feed forward

On dit d'un réseau qu'il est feedforward, ou encore à sens unique, lorsqu'il ne contient pas de boucle interne, c'est à dire lorsque l'information qui le traverse ne circule que de l'entrée vers la sortie. Les neurones sont organisés en couches successives, les neurones d'une couche étant reliés à ceux de la suivante. Parmi ce type de réseaux, il existe les réseaux à deux couches. Ils sont composés d'un ensemble de neurones répartis sur deux couches distinctes appelées couche d'entrée et de sortie, les neurones de la première étant connectés à ceux de la seconde (figure II.7).

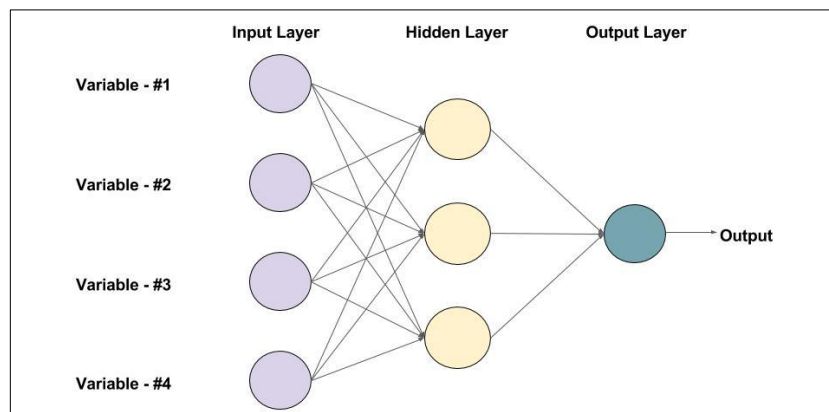


Figure II.8: Réseau de neurone feedforward

(<https://www.learnopencv.com/understanding-feedforward-neural-networks>)

Si l'on intercale entre la sortie et l'entrée une ou plusieurs couches (qui seront alors appelées couches cachées) (figure II.8), on parle alors de réseau feedforward multicouches (perceptron multicouches).

L'intérêt de rajouter au moins une couche cachée est d'augmenter le nombre de connexions, ce qui accroît la capacité d'un réseau à extraire l'information des données fournies en entrée (*Le Fablec, 1999*). Les perceptrons multicouches sont appelés souvent des approximateurs universels et ceci est grâce à leur capacité de faire l'approximation de n'importe quelle fonction mathématique.

II.3.5.2. Les réseaux récurrents : feedback

On parle de réseau de neurones récurrent lorsqu'il existe une boucle au moins dans l'ensemble des connexions. Il y a alors au minimum un feedback d'une couche sur la précédente. La figure II-8 en donne un exemple très simple : un neurone de la couche de sortie renvoie sa sortie sur la couche d'entrée avec un délai. Cela permet d'utiliser une des sorties à l'instant t comme entrée à l'instant $t+1$ (*Le Fablec, 1999*).

Les réseaux récurrents sont souvent utilisés pour la modélisation et la prévision des systèmes dynamiques ainsi que les séries temporelles. Certains auteurs ajoutent à ces deux architectures une troisième : les réseaux auto organisés.

Les réseaux de neurones auto organisés sont des réseaux qui changent leurs structures internes pendant l'apprentissage. Ainsi les neurones se regroupent topologiquement suivant la représentation des exemples. Ces réseaux sont des dérivées des modèles de Kohonen. (*Mokhnache 2004*)

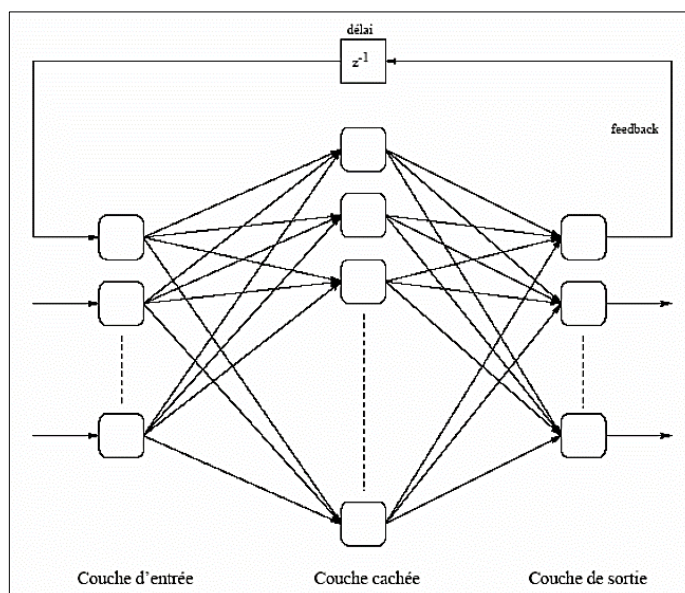


Figure II.9: Réseau de neurone récurrent

II.3.6. Apprentissage des réseaux de neurones artificiels

Dans le cas d'un cerveau humain, l'utilisation correcte des neurones nécessite au moins une phase d'apprentissage qui permettra d'avoir le fonctionnement escompté. Cet apprentissage consiste à jouer sur les synapses qui relient les neurones. De la même manière, les réseaux de neurones artificiels que nous venons de décrire ont besoin d'apprendre (*Le Fablec, 1999*).

L'apprentissage est une phase du développement d'un réseau de neurones. C'est un changement du comportement du réseau de façon à lui permettre de se rapprocher d'un but défini. Ce but est normalement l'approximation d'un ensemble d'exemples ou l'optimisation de l'état du réseau en fonction de ses poids pour atteindre l'optimum d'une fonction économique fixée a priori. Il faudra là aussi agir sur les synapses. Pour cela, on devra régler les valeurs des poids de chaque connexion entre neurones ainsi que les seuils (ou biais) (*Cornuéjols and MicletEyrolles, 2007*).

L'apprentissage des réseaux de neurones n'est donc qu'une procédure qui consiste à estimer les paramètres des neurones d'un réseau, afin que celui-ci remplisse au mieux la tâche qui lui est affectée.

Il existe trois types d'apprentissages principaux. Ce sont l'apprentissage supervisé, non supervisé, et semi supervisé.

II.3.6.1. Apprentissage supervisé

Il consiste à apprendre une association particulière entre deux ensembles de vecteurs. Durant une première phase, appelée phase d'apprentissage, on présente au réseau des vecteurs d'entrées en lui imposant, comme des contraintes, les vecteurs de sortie que l'on désire lui faire associer à ces entrées. Pendant une seconde phase, appelée phase de test, on présente au réseau de nouveaux vecteurs qui n'ont pas servi à l'apprentissage et l'on observe ses réponses. Les performances du réseau sont alors évaluées par le pourcentage d'erreurs.

L'intérêt des réseaux neuronaux réside précisément dans leur capacité intrinsèque à la généralisation, c'est-à-dire leur aptitude à bien se comporter sur des vecteurs qu'ils n'ont pas appris (*Yahiaoui et al., 2002*).

Pour ce type d'apprentissage la règle la plus utilisée est celle de Widrow-Hoff. D'autres règles d'apprentissage existent telle que la règle de Hebb, la règle du perceptron, la règle de Grossberg etc.

La loi de Hebb : Si deux unités connectées sont actives simultanément, le poids de leur connexion est augmenté ou diminué. ν est une constante positive qui représente la force d'apprentissage (learning rate).

La loi de Widrow-Hoff : La règle d'apprentissage de Widrow-Hoff est une règle qui permet d'ajuster les poids d'un réseau de neurones pour diminuer à chaque étape l'erreur commise par le réseau de neurones (à condition que le facteur d'apprentissage soit bien choisi). Un poids est modifié en utilisant la formule suivante:

$$w_{k+1} = w_k + \alpha \delta_k x_k \quad (22)$$

où: w_k est le poids à l'instant k ,

w_{k+1} le poids à l'instant $k+1$,

α est le facteur d'apprentissage,

δ_k caractérise la différence entre la sortie attendue et la sortie effective du neurone à l'instant k .

x_k la valeur de l'entrée avec laquelle le poids w est associé à l'instant k .

Ainsi, si δ_k et x_k sont positifs tous les deux, alors le poids doit être augmenté. La grandeur du changement dépend avant tout de la grandeur de δ_k mais aussi de celle de x_k . Le coefficient α sert à diminuer les changements pour éviter qu'ils deviennent trop grands, ce qui peut entraîner des oscillations du poids.

Deux versions améliorées de cet apprentissage existent, la version "par lots" et la version "par inertie" (momentum en anglais), dont l'une utilise plusieurs exemples pour calculer la moyenne des changements requis avant de modifier le poids et l'autre empêche que le changement du poids au moment k ne devienne beaucoup plus grand qu'au moment $k-1$ (*Framling, 1992*).

Loi de Grossberg : On augmente les poids qui entrent sur l'unité gagnante a_i s'ils sont trop faibles, pour les rapprocher du vecteur d'entrée a_j . C'est la règle d'apprentissage utilisée dans les cartes auto-organisatrices de Kohonen.

$$\Delta w_{ij} = R(a_j - w_{ij})a_i$$

II.3.6.2. Apprentissage non supervisé

Il consiste à découper l'ensemble des vecteurs d'entrées en classes d'équivalence, sans qu'il soit nécessaire de donner au réseau neuronal les noms de classes pour chaque exemple. Cette classification automatique ressemble à certaines méthodes utilisées en analyse de données.

La séparation en classes d'équivalence s'opère par mesure de ressemblance entre les vecteurs proposés. La contrainte que l'on doit imposer au réseau neuronal est le nombre de

classes d'équivalence. Pendant la phase d'apprentissage, le réseau neuronal construit une topologie de l'espace des vecteurs d'entrées.

L'intérêt de cette approche est que l'on n'a pas besoin de disposer d'exemples de problèmes résolus (c'est-à-dire de vecteurs de sortie correspondant à chaque vecteur d'entrée utilisé pour l'apprentissage) (*Cornuéjols and Miclet Eyrolles, 2007*).

L'inconvénient est que le critère de mesure de ressemblance ne correspond pas forcément avec ce que l'on veut (*Yahiaoui et al., 2002*).

II.3.6.3. Apprentissage semi-supervisé

C'est un apprentissage de type essai-erreur où le réseau donne une solution et est seulement évalué avec une information indiquant si la réponse était correcte ou si elle était au moins meilleure que la dernière fois.

II.3.7. Algorithmes d'apprentissage des RNA

II.3.7.1. L'apprentissage du perceptron

L'apprentissage se fait en trois étapes :

- Présentation du vecteur d'entrées e
- Calcul de l'erreur Δ en sortie par la formule suivante $\Delta = y - \sum_{j=0}^{j=n} (w_j \cdot e_j)$

Où Δ : erreur calculée

$\Sigma (w_j \cdot e_j)$: Entrée pondérée du neurone

y : Vecteur de la sortie désirée

- Modification des poids de connexions : $\Delta w_{ij} = K \cdot \Delta e_j$

Où Δw_{ij} : Modification à apporter aux poids

K : Constante numérique aléatoire

Δ : Erreur calculée en sortie

e_j : Activation du neurone j

L'inconvénient du perceptron est qu'il n'est efficace que pour les problèmes linéairement séparables, car l'algorithme d'apprentissage ne peut pas converger dans le cas contraire. C'est à-dire que les vecteurs d'entrées doivent être classés en deux classes. Si w_0 est le seuil du neurone, on aura donc :

La réponse = 1 si $\Sigma(w_j \cdot e_j) > w_0$

La réponse = 0 si $\Sigma(w_j \cdot e_j) < w_0$

Cette problématique a limité le champ d'application du perceptron monocouche. Par conséquent, la recherche dans le domaine des réseaux de neurone a été arrêtée pendant 20 ans, pour revoir la lumière dans les années 80 après le développement de l'algorithme de rétro-propagation de l'erreur.

II.3.7.2. La propagation de l'erreur

Cet algorithme est une généralisation aux réseaux de neurones formels multicouches du modèle de Widrow-Hoff. Il a été mis au point par deux équipes simultanément : une équipe française (Fogelman-Soulié, Gallinari, Lecun) et une équipe américaine (Rumelhart, Hinton, Williams). Son idée consiste en l'utilisation d'une fonction sigmoïde dérivable en remplacement de la fonction à seuil couramment usité. La fonction sigmoïde prise en général est : $f(x) = \frac{1}{1+e^x}$ (Davallo, 1991)

L'algorithme permet de réinjecter l'erreur, calculée en sortie, vers les couches internes, en suivant le chemin inverse du signal provenant de la couche d'entrée et allant vers la couche de sortie. Avec cette méthode, la réadaptation des poids se fait après passage de tous les exemples. Le réajustement se fait donc suivant la moyenne de tous les exemples ce qui rend insensibles aux bruits, et son application ne nécessite pas un apprentissage en temps réel. L'algorithme de base de la back-propagation est celui de la descente du gradient (Mokhnache, 2004).

L'algorithme se présente comme ceci :

- Soit un échantillon \vec{x} que l'on met à l'entrée du réseau de neurones et la sortie recherchée pour cet échantillon \vec{t}
- On propage le signal en avant dans les couches du réseau de neurones : $x_k^{(n-1)} \mapsto x_j^{(n)}$
- La propagation vers l'avant se calcule à l'aide de la fonction d'activation g , et des poids synaptiques $w_{jk}^{(n)}$ entre le neurone $x_k^{(n-1)}$ et le neurone $x_j^{(n)}$.
- Lorsque la propagation vers l'avant est terminée, on obtient à la sortie le résultat \vec{y}
- On calcule alors l'erreur entre la sortie donnée par le réseau \vec{y} et le vecteur \vec{t} désiré à la sortie pour cet échantillon. Pour chaque neurone i dans la couche de sortie, on calcule : $e_i^{sortie} = g'(h_i^{sortie})(t_i - y_i)$
- On propage l'erreur vers l'arrière $e_i^{(n)} \mapsto e_j^{(n-1)}$ grâce à la formule suivante :

$$e_j^{(n-1)} = g'^{(n-1)}(h_j^{(n-1)}) \sum_i w_{ij} e_i^{(n)}$$

$$e_j^{(n)} = \sum_i [t_i - y_i] \frac{\partial y_i}{\partial h_j^{(n)}}$$

On met à jour les poids dans toutes les couches en calculant $\Delta w_{ij}^{(n)} = \lambda e_i^{(n)} x_j^{(n-1)}$ où λ représente le taux d'apprentissage (de faible magnitude et inférieur à 1). La Figure II.10 schématise le principe de la rétro-propagation de l'erreur.

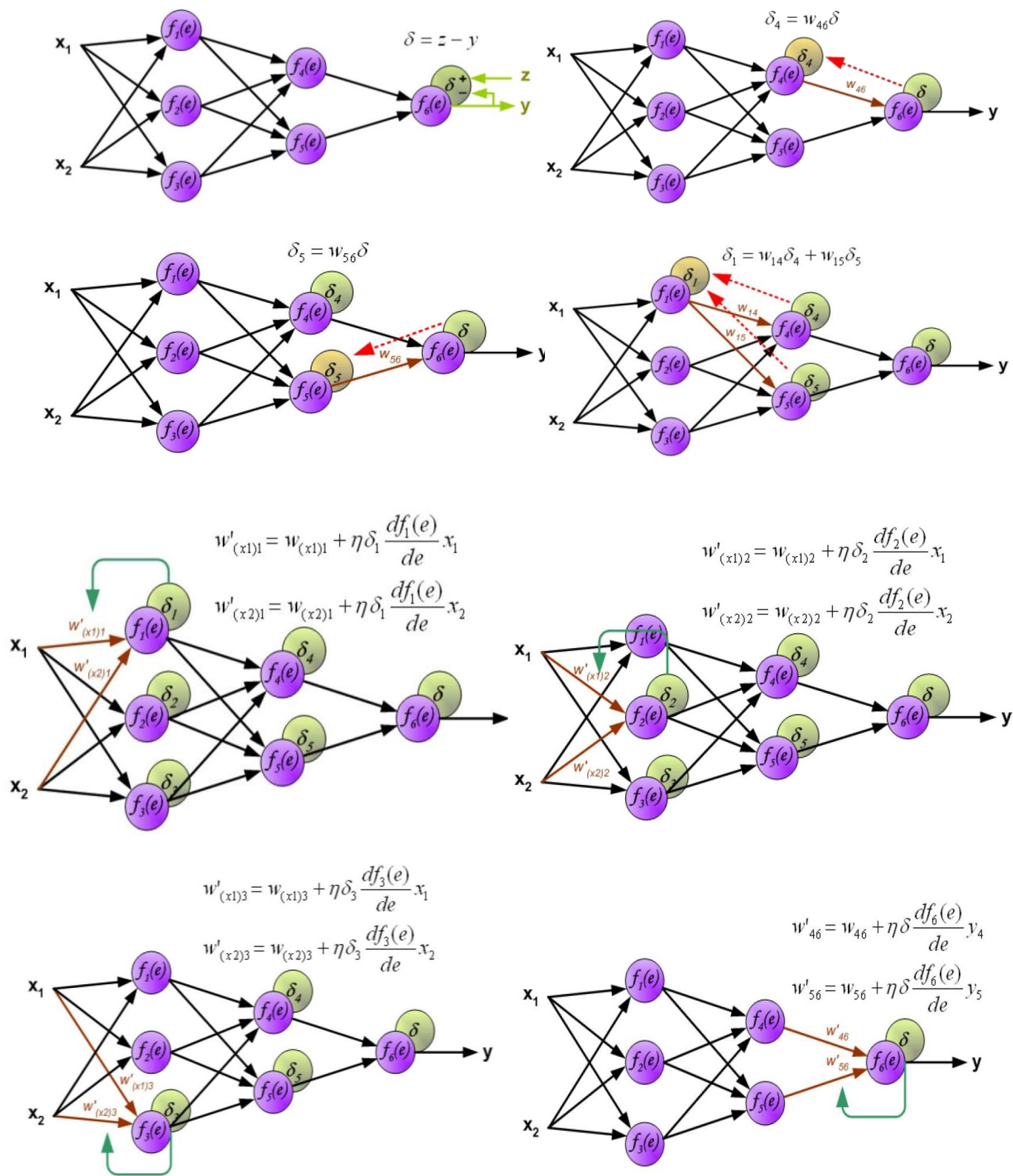


Figure II.10: Principe de la retro-propagation de l'erreur

II.3.7.3. Algorithme de Levenberg-Marquardt

La méthode d'optimisation de Levenberg-Marquardt est particulièrement robuste et efficace. Elle est devenue un algorithme de référence pour la minimisation de fonctions. Cependant, aucune implémentation en R n'existait (*Marquardt, 1963*)

L'algorithme itératif de Levenberg-Marquardt permet de trouver un minimum local (éventuellement global) d'une fonction continue dérivable deux fois. Lorsque l'estimation à l'itération courante est loin du minimum, la matrice *hessienne* est souvent non inversible, l'algorithme de Levenberg-Marquardt permet de gonfler la diagonale pour proposer malgré tout une direction (*Marquardt, 1963*).

À proximité du minimum, il correspond à l'algorithme de Newton-Raphson. Concernant les critères de convergence, deux critères secondaires (stabilisation des estimations et de la valeur de la fonction) et un critère principal (la Distance Relative au Minimum - RDM) sont implémentés. Leurs valeurs seuils sont modifiables par l'utilisateur. Le RDM (*Commenges et al., 2006*) est un critère original correspondant à la norme des gradients dans la métrique des paramètres à estimer divisée par le nombre de paramètres à estimer pour s'adapter à la dimension du problème.

Il est aussi interprétable comme le ratio entre l'erreur numérique et l'erreur statistique commise. Ainsi, sa valeur doit être aussi proche de zéro que possible et en tout cas inférieure à 1

La procédure de l'algorithme est itérative. On part d'un paramètre initial, que l'on suppose « assez proche » d'un minimum, et qui constituera le vecteur p de départ. Dans beaucoup de cas, un paramètre de départ « standard », tel que $p^T=(1,1,\dots,1)$ fonctionnera sans problème. Dans certains cas, il n'y a convergence que si le vecteur de départ n'est pas trop éloigné de la solution.

À chaque itération, on remplace p par une nouvelle estimation $p + q$. Afin de déterminer q , les fonctions $f_i(p + q)$ sont approchées en étant linéarisées :

$$f(p+q) \approx f(p) + J_q$$

où on a noté J la Jacobéenne de f en p .

À un minimum de la somme des carrés S , on a $\Delta q S = 0$. En dérivant le carré de l'expression de droite, qui s'annule donc, et en posant $y = f(p+q)$, on obtient :

$$(J^T J)q \approx J^T (y - f(p)) \quad \text{d'où l'on tire aisément } q \text{ en inversant } J^T J.$$

Dans l'inversion matricielle, tout va dépendre du rapport entre la valeur propre la plus grande en norme, et la valeur propre la plus petite ; ce rapport, appelé conditionnement de matrice, va concrètement refléter la robustesse de l'algorithme face au bruit.

Le point essentiel de l'algorithme de Levenberg-Marquardt est d'approcher cette équation, en l'« amortissant » un peu. On parle alors de « chargement de la diagonale » afin de contourner le mauvais conditionnement le cas échéant, problème que l'on retrouve avec l'algorithme de Capon et que l'on peut résoudre par une décomposition en valeurs singulières

$$J^T J + \lambda \text{diag}(J^T J) q \approx J^T (y - f(p))$$

Le facteur d'amortissement positif λ est ajusté à chaque nouvelle itération. Si la diminution de S est rapide, on peut utiliser une valeur plus faible - ce qui rapproche l'algorithme de celui de Gauss-Newton. Si en revanche une itération est peu efficace, on peut augmenter λ , ce qui rapproche cette fois l'algorithme de celui du gradient. Un tel facteur d'amortissement est utilisé par exemple dans la régularisation de Tikhonov, utilisée pour résoudre certains problèmes linéaires (*Marquardt, 1963*).

Si on a effectué plus d'un certain nombre d'itérations, ou bien que l'on s'est approché suffisamment d'un minimum, la procédure se termine et renvoie le paramètre p comme estimation de la solution.

II.3.7.4. La régularisation Bayésienne

La technique de régularisation consiste à imposer des contraintes, donc à apporter une information supplémentaire, sur l'évolution des poids du réseau de neurones. Par exemple, on peut volontairement pénaliser les poids trop grands selon la formule :

$$F_{\text{erreur}} = \frac{1}{N} \sum e_i^2 \text{ et } F_{\text{poids}} = \frac{1}{N} \sum w_i^2$$

Ce qui donne :

$$F_{\text{new}} = \gamma F_{\text{erreur}} + (1-\gamma) F_{\text{poids}}$$

Où γ est un paramètre d'optimisation. Mais le problème réside dans le choix de la valeur de ce paramètre (*Idier, 2001 ; Antoni, 2012*).

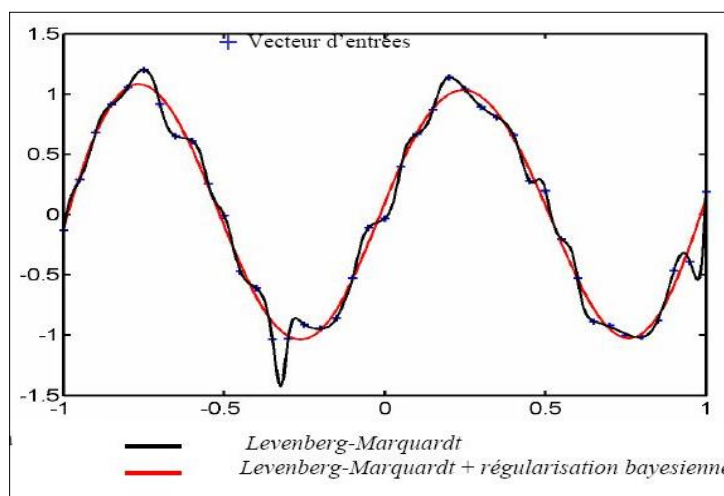


Figure II.11: Comparaison entre l'apprentissage avec l'algorithme de LM seul et avec l'algorithme LM + régularisation bayésienne (*Philippe, 1997*)

La régularisation bayésienne, qui suppose que les poids et les biais suivent des distributions spécifiques (les paramètres sont estimés au fur et à mesure de l'apprentissage) donne en général des résultats très satisfaisants (*Idier, 2001 ; Antoni, 2012*).

II.3.8. Domaine d'application des réseaux de neurones artificiel

Il est avantageux de mettre en œuvre les réseaux de neurones pour toute application nécessitant de trouver, par des méthodes statistiques, une relation non linéaire entre des données numériques.

Le problème posé par l'ingénieur doit satisfaire certaines conditions afin de le pouvoir résoudre à l'aide des réseaux de neurones :

- La première condition, nécessaire et non suffisante: il faut disposer d'un échantillon de taille suffisamment grande, et bien représentatifs.
- En l'absence de toute connaissance a priori sur l'intérêt d'un modèle non linéaire, l'ingénieur doit d'abord utiliser les méthodes simples et éprouvées d'élaboration d'un modèle linéaire (qui est simple et moins coûteux en temps de calcul que celui d'un réseau de neurone). S'il s'avère que la précision du modèle est insuffisante bien que tous les facteurs pertinents soient présents dans les entrées du modèle, alors le concepteur de modèle doit envisager la mise en œuvre de modèles non linéaires tels que les réseaux de neurones.
- Si les données sont disponibles et s'il on s'est assuré qu'un modèle non linéaire est utile, il faut s'interroger sur l'opportunité d'utiliser un réseau de neurone de préférence à une autre famille de fonction non linéaire, les polynômes par exemple. Il est prouvé que les réseaux de neurones sont plus avantageux que le nombre de variables est « grand », c'est-à-dire qu'il est supérieur ou égal à trois.

En résumé, les réseaux de neurones peuvent être utiles pour résoudre un problème à régression non linéaire possédant au moins trois variables, sachant que les données numériques disponibles sont suffisamment nombreuses et représentatives (*Dreyfus, 2002*).

II.3.9. La mise en œuvre des réseaux de neurones artificiel

Les réseaux de neurones réalisent des fonctions non linéaires paramétrées ; ils peuvent approcher toute fonction de régression si l'on trouve un réseau de neurones qui rend minimum la fonction de coût des moindres carrés. La mise en œuvre d'un réseau de neurones nécessite donc de :

- Déterminer les entrées pertinentes, c'est-à-dire les grandeurs qui ont une influence significative sur le phénomène que l'on cherche à modéliser, Collecter les données nécessaires à l'apprentissage et à l'évaluation des performances du réseau de neurones,
- Trouver le nombre de neurones cachés nécessaires pour obtenir une approximation satisfaisante,
- Estimer les valeurs des paramètres correspondant à un minimum de la fonction de coût, c'est-à-dire d'effectuer un apprentissage,
- Evaluer les performances du réseau de neurones à l'issue de l'apprentissage.
- En fonction des résultats obtenus, il peut être nécessaire d'effectuer plusieurs itérations de tout ou partie de cette procédure.

II.3.10. Avantages des RNA

Un des intérêts majeurs des réseaux de neurones artificiels (informatiques) est qu'ils sont organisés sous la forme d'une structure massivement parallèle (*Le Fablec, 1999*). De plus, on peut citer aussi les avantages suivants (similaires à ceux des réseaux de neurones biologiques) :

- Les neurones sont des composants non linéaires, par conséquent le réseau aussi : ceci est une propriété très importante car cela leur permet de traiter des problèmes physiques non linéaires
- Ils sont capables d'apprendre et de généraliser : après une période d'apprentissage qui consiste à leur présenter des entrées ainsi que les sorties que l'on voudrait qu'ils trouvent, ils sont capables de produire des sorties "raisonnables" sur des cas qu'ils n'ont pas appris
- S'il a appris à travailler dans un environnement précis, on peut réapprendre un réseau pour qu'il s'adapte à un autre environnement semblable
- Dans le cas d'un problème de classification un réseau de neurones ne se contente pas de faire un choix : il fournit aussi un niveau de confiance sur sa décision. Cela permet de rejeter des motifs ambigus et d'avoir une idée sur la précision du résultat
- Si l'on considère un réseau connecté physiquement qui subit un dommage qui n'est pas trop important alors les réponses ne sont pas trop dégradées. Le réseau est donc résistant car il continue à être utilisable bien qu'il soit endommagé.

II.4. La dispersion de la pollution : Modèle Gaussien

L'objectif de la modélisation de la dispersion atmosphérique est de connaître, en tout point de l'espace et à tous les instants, les concentrations en espèces relâchées. Il convient à ce niveau de distinguer deux types de relâchement atmosphérique :

- celui qui fait suite à un incident ou un accident (rejet d'une bouffée de polluant par exemple)
- et celui issu du fonctionnement normal d'une installation (rejet de « routine » lié au procédé de l'installation).

En modélisation, ces deux situations sont traitées de manières différentes car elles ne font pas appel aux mêmes types de données, notamment météorologiques. En effet, les rejets de routine sont modélisés à partir de statistiques météorologiques alors que les rejets accidentels sont modélisés à partir de données météorologiques locales observées ou prévues.

À cet objectif de description de la dispersion atmosphérique, ajoutons celui d'estimer l'impact des polluants dispersés sur les populations et leur environnement (les écosystèmes).

Un polluant rejeté dans l'atmosphère est soumis aux différents mécanismes régissant les écoulements de l'atmosphère. Il existe ainsi trois processus principaux qui sont le transport, la diffusion et le dépôt au sol du polluant. La dispersion d'une pollution par voie atmosphérique est très dépendante des conditions météorologiques.

Les paramètres météorologiques ayant une influence notable sur les mécanismes de la dispersion atmosphérique sont le vent, la turbulence de l'atmosphère et la distribution verticale de température, appelée aussi gradient thermique. Ces paramètres interviennent au sein même des trois processus de la dispersion atmosphérique que sont le transport, la diffusion et les dépôts au sol.

II.4.1. Le transport

Représente la partie physique de la dispersion atmosphérique liée au vent et aux caractéristiques propres des polluants lors de leur émission (vitesse, température, densité). Le transport est donc le déplacement du centre de gravité du nuage. Le vent oriente les panaches et les dilue plus ou moins en fonction de sa vitesse.

II.4.2. La diffusion

La diffusion est le déplacement relatif des volumes élémentaires de nuage par rapport à son centre de gravité. La diffusion est la partie de la dispersion imputable à la turbulence de l'atmosphère.

On distingue deux échelles de turbulences, petites et grandes, en fonction de leur initiateur.

- La turbulence de grandes échelles est due à des instabilités lointaines que l'on peut, par exemple, relier aux conditions météorologiques à l'échelle synoptique. Elle est donc indépendante des phénomènes locaux et n'a, par conséquent, pas d'effet sur la diffusion à l'échelle qui intéresse les études présentées ici. Néanmoins, en générant des vents, elle participe au processus de transport des polluants.
- La turbulence de petite échelle est, quant à elle, directement influencée par les facteurs modifiant les échanges mécaniques et thermiques entre le sol et l'air. On peut citer le frottement de l'air sur le sol, la présence de bâtiments, le gradient thermique par exemple. Elle est donc située à proximité du sol dans une zone appelée couche limite atmosphérique.

Cette zone est caractérisée par la présence de gradients verticaux de vitesse importants. Par conséquent, cette turbulence a la capacité de brasser l'air et ainsi d'influer sur le processus de diffusion du polluant.

Une grandeur météorologique importante agissant sur la turbulence est la stabilité de l'atmosphère. Cette grandeur est liée au gradient vertical de densité de l'air, lui-même lié au gradient vertical de température.

Lorsque l'air, au voisinage du sol, est plus chaud qu'en altitude, il a tendance à s'élever, emportant avec lui les polluants. La turbulence est importante et l'atmosphère est dite instable

En revanche, si l'air est plus froid au niveau du sol qu'en altitude, les polluants vont avoir tendance à stagner près du sol, c'est le phénomène d'inversion thermique. La turbulence verticale est alors faible et l'atmosphère est dite stable.

II.4.3. Le dépôt

On étudie deux grandes classes de polluants atmosphériques : les gaz et les particules solides (poussières). On considère en général que seules les particules solides se déposent au sol. On distingue alors deux types de dépôts, les dépôts secs et les dépôts humides.

- Sous l'appellation « dépôt sec » sont regroupés ;
- L'autre forme de dépôt, le dépôt humide, intervient lorsque l'air est chargé de gouttes d'eau qui, en traversant le nuage, emportent une fraction des particules de polluant sur le sol. Les dépôts conduisent à appauvrir le panache en polluant, l'appauvrissement est plus important en présence de pluie.

Les modèles les plus utilisés et les plus répandus sont les modèles dits gaussiens. Ces modèles reposent sur les solutions gaussiennes de l'équation de transport-diffusion du polluant, dans des conditions simplificatrices (uniformité du vent, loi de gradient de diffusion turbulente homogène, voire isotrope) pour lesquelles existe une solution analytique sous forme de gaussienne (*Crabot, 1999*). Les coefficients de la distribution gaussienne sont déduits d'expériences in situ ; ils dépendent des conditions de stabilité atmosphérique (classées selon des situations typiques) et soit, selon Doury (*Doury, 1976*), du temps de transfert du polluant, soit selon Pasquill (*Pasquill, 1961*), de la distance parcourue depuis le point d'émission. Dans ces modèles, la turbulence n'est pas explicitement simulée mais ses effets sur la dispersion sont intégrés au travers de la formulation des paramètres d'écart-types des lois gaussiennes. Cette formulation est fondée sur des expérimentations in situ, qui, pour la plupart, sont réalisées en terrain plat et sans constructions, afin d'avoir un champ de vent moyen unidirectionnel. En revanche, plusieurs conditions de stabilité atmosphérique ont été observées et la turbulence y est présente dans toutes ses échelles, contrairement à ce que des simulations sur modèle réduit en soufflerie permettent.

II.4.4. Description du modèle Gaussien

Les modèles gaussiens permettent de simuler la dispersion atmosphérique de polluants non-réactifs à proximité de la source. Les concentrations dans le panache sont représentées par l'équation suivante :

$$c(x) = \frac{q}{2\pi V_x \partial_y \partial_z} e^{-\frac{y^2}{2\partial_y^2} - \frac{z^2}{2\partial_z^2}}$$

Où : C : concentration de la pollution (g/m^3)

q : débit d'émission (g/s)

V : vitesse du vent (m/s)

$\partial_y \partial_z$: la déviation standard horizontale et verticale (m)

Z : hauteur (m)

Y : distance latérale (m)

L'ellipse ($4 \partial_y, 4 \partial_z$) contient 95,4% de la masse du panache gaussien

L'ellipse ($2 \partial_y, 2 \partial_z$) contient 68,2% de la masse du panache gaussien

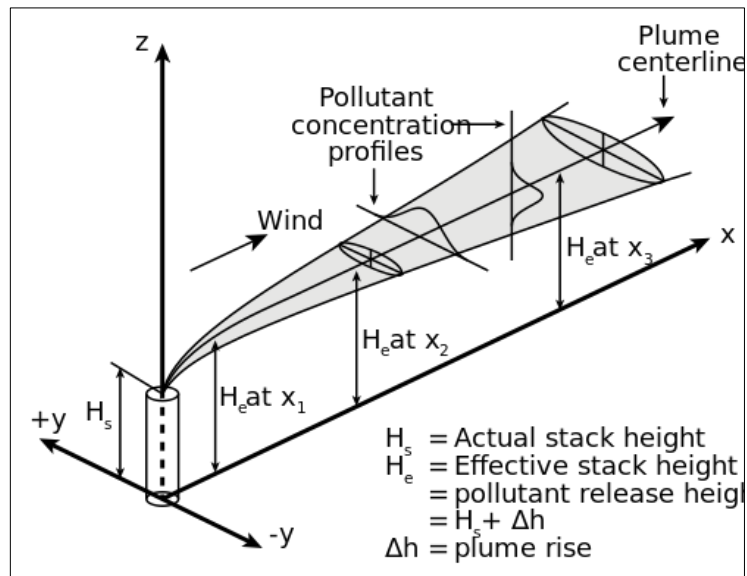


Figure II.12: Système de coordonnées du modèle gaussien
 (https://www.pngitem.com/middle/iTJxbii_gaussian-plume-model-hd-png-download)

La prise en compte de la hauteur initiale du panache est faite simplement par changement de coordonnée verticale : $z - z_0$

$$c(x) = \frac{q}{2\pi V_x \sigma_y \sigma_z} e^{-\frac{(y-y_0)^2}{2\sigma_y^2}} e^{-\frac{(z-z_0)^2}{2\sigma_z^2}}$$

Cette hauteur z_0 est la hauteur de la source (par exemple, cheminée) + la hauteur de surélévation du panache par rapport à la source

La stabilité atmosphérique influence la dispersion du panache :

- des conditions stables dispersent peu le panache ;
- des conditions instables dispersent fortement le panache.

La réflexion d'un panache sur le sol est représentée dans la formulation gaussienne en ajoutant une source virtuelle qui est symétrique de la source réelle par rapport au sol :

$$c(x) = \frac{q}{2\pi V_x \sigma_y \sigma_z} e^{-\frac{y^2}{2\sigma_y^2}} \left[e^{-\frac{(z-z_0)^2}{2\sigma_z^2}} + e^{-\frac{(z+z_0)^2}{2\sigma_z^2}} \right]$$

La réflexion d'un panache sous une couche d'inversion de température est représentée dans la formulation gaussienne en ajoutant une source virtuelle qui est symétrique de la source réelle par rapport à la base de cette couche d'inversion :

$$c(x) = \frac{q}{2\pi V_x \sigma_y \sigma_z} e^{-\frac{y^2}{2\sigma_y^2}} \left[e^{-\frac{(z-z_0)^2}{2\sigma_z^2}} + e^{-\frac{(z-z_0-z_1)^2}{2\sigma_z^2}} \right]$$

II.4.5. Détermination des paramètres du modèle

II.4.5.1. Méthode de Pasquill-Gifford

Le problème principal revient à déterminer les deux constantes ∂_y et ∂_z du modèle, la méthode de *Pasquill-Gifford* est une méthode graphique basée sur la classe de la stabilité, le tableau II.1 représente cette classe avec la description correspondante (*Gifford, 1961*).

Tableau II.1: La correspondance entre le niveau de la stabilité, classe et description

Classe	Niveau de stabilité	Description
A	Trop instable	Météo d'été trop ensoleillée
B	Instabilité modérée	Ensoleillé et chaud
C	Faible instabilité	partiellement froid
D	Neutre	Jour/nuit nuageux
E	Faible stabilité	Partiellement nuageux la nuit
F	Stabilité modérée	Nuit à ciel dégagé
G	Trop stable	Nuit froide à ciel dégagé

La détermination de la classe de la stabilité se fait en fonction de la vitesse du vent mesurée au niveau du sol, le tableau II.2 illustre cette classification.

Tableau II.2: La classe de la stabilité en fonction de la vitesse du vent

Vitesse du vent à la surface (m/s)	Jour avec ensoleillement			Nuit	
	Fort	Modéré	Faible	Nuages basses	Nuages
2	A	A-B	B	-	-
2-3	A-B	B	C	E	F
3-5	B	B-C	C	D	E
5-6	C	C-D	D	D	D
6	C	D	D	D	D

Une fois la classe est déterminée, grâce aux graphes représentés ci-dessous, les constantes ∂_y et ∂_z sont extraites par une simple projection de la distance x sur la classe de la stabilité correspondante.

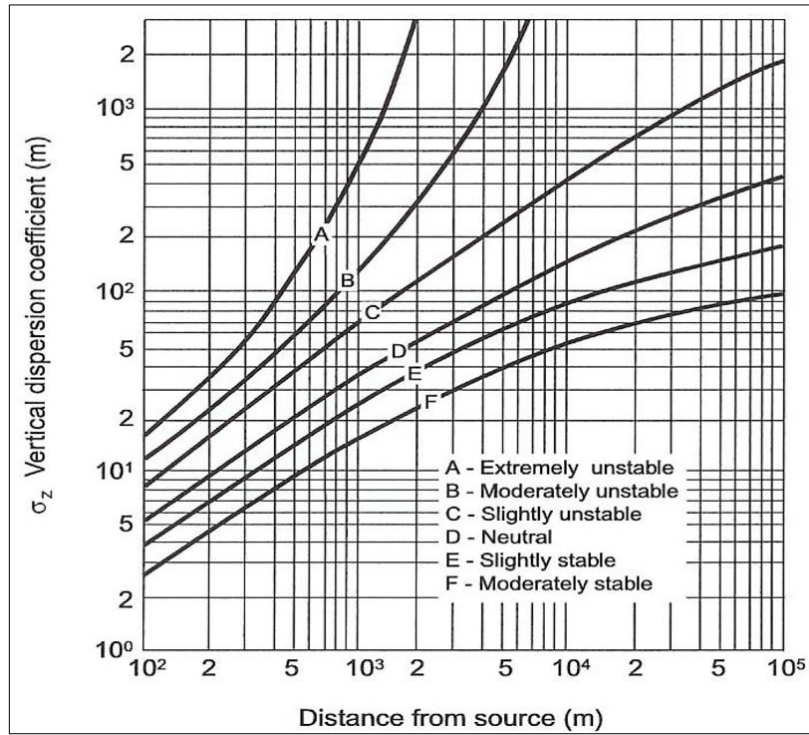


Figure II.13: Coefficient de diffusion verticale $\hat{\sigma}_z$ de Pasquill-Gifford en fonction de la distance x

(<https://slideplayer.com/slide/6977077>)

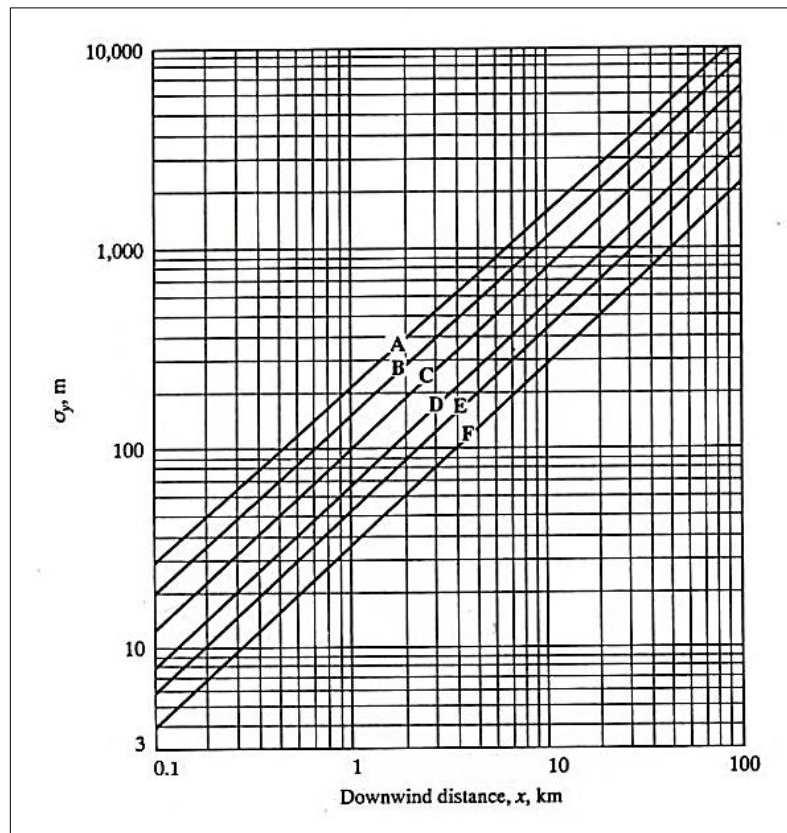


Figure II.14: Coefficient de diffusion verticale $\hat{\sigma}_y$ de Pasquill-Gifford en fonction de la distance x

(<https://slideplayer.com/slide/6977077>)

II.4.5.2. Méthode de Caraway

La méthode de Caraway repose sur la modélisation des courbes représentées dans la figure II.13 et la figure II.14, l'équation du lissage du modèle de Caraway pour déterminer la diffusion verticale est la suivante (*Weber, 1999*) :

$$\partial_z = a.x^b$$

Les valeurs des constantes de ces modèles (a , b) sont extraites du tableau suivant :

Tableau II.3: Coefficients du modèle de Caraway pour la détermination du ∂_z

Classe de stabilité atmosphérique	Distance x (m) $100 < x < 500$		Distance x (m) $500 < x < 5000$		Distance x (m) $5000 < x$	
	a	b	a	b	a	b
	A=1	,03830	1,28100	,00025	2,08900	,00025
B=2	,13930	,94670	,04936	1,11400	,04936	1,11400
C=3	,11200	,91000	,10140	,92600	,11540	,91090
DD=4	,08560	,8650	,25910	,68690	,73680	,56420
DN=5	,08180	,81550	,25270	,63410	1,29700	,44210
E=6	,10940	,76570	,24520	,63580	0,92040	,48050
F=7	,05645	,80500	,19300	,60720	1,50500	,36620

L'équation du lissage du modèle de Caraway pour déterminer la diffusion latérale est la suivant

$$\partial_y = c.x^d$$

Les valeurs des constantes de ces modèles (c , d) sont extraites du tableau suivant :

Tableau II.4: Coefficients du modèle de Caraway pour la détermination du ∂_y

Classe de stabilité atmosphérique	Distance x (m) $x < 10,000$		Distance x (m) $x \geq 10,000$	
	c	d	c	d
	A=1	0,495	0,873	0,606
B=2	0,31	0,897	0,523	0,84
C=3	0,197	0,908	0,285	0,867
DD=4	0,122	0,916	0,193	0,865
DN=5	0,122	0,916	0,193	0,856
E=6	0,934	0,912	0,141	0,868
F=7	0,0625	0,911	0,8	0,884

II.5. L'Arduino au service de l'environnement

Pour agir efficacement sur un processus physique, chimique ou biologique, naturel ou industriel, il importe, au préalable, de bien le connaître : la chaîne d'acquisition fournit au décideur (homme ou machine) les informations permettant d'orienter son action et de valider ses décisions. Encore faut-il que les informations acquises soient suffisantes et dignes de foi.

D'un monde réel complexe où s'enchevêtrent de multiples phénomènes, la chaîne d'acquisition permet d'extraire les informations simples et nécessaires à une représentation valable et utile. Ces informations caractérisent les grandeurs physiques qui sont les paramètres agissants des modèles représentatifs de la réalité : température, pression, débit, force, etc. Connaître une grandeur, c'est, pour le physicien ou l'ingénieur, lui affecter une valeur numérique ou une succession de valeurs numériques significatives qui résultent d'opérations de mesure : la chaîne d'acquisition est donc une chaîne de mesure. La grandeur objet de mesure ou mesurande est caractérisée par son intensité et par son évolution temporelle; ce sont ces informations que pour chaque mesurande la chaîne d'acquisition doit pouvoir délivrer sous la forme de signaux électriques représentatifs et exploitables (*Simon, 2013 ; Julien et al., 2013 ; Patrick et al., 2014*).

Les difficultés que présente l'acquisition de données ont deux causes principales :

- L'imbrication dans la réalité de multiples grandeurs qui rend souvent difficile l'isolement et la saisie du seule mesurande recherchée
- Les imperfections des dispositifs constitutifs de la chaîne qui entraînent une détérioration des signaux et donc de l'information.

C'est à ces deux niveaux que doit se porter l'attention du concepteur pour limiter l'imprécision des mesures et éviter les risques d'erreur.

Le capteur est la source première du signal électrique et donc de l'information traitée; le signal qu'il délivre doit être l'image électrique du mesurande qu'il traduit et donc contenir toute l'information recherchée. Or, du seul fait de son installation, un capteur est susceptible de modifier la valeur du mesurande qu'il est censé traduire et, de plus, les autres grandeurs physiques de son environnement risquent d'altérer sa réponse.

Les dispositifs dont l'association au capteur constitue la chaîne d'acquisition (amplificateur, filtre, échantillonneur-bloqueur, convertisseur analogique-numérique) ont pour fonction de traiter le signal de façon à le rendre approprié à l'exploitation prévue. Cependant, les caractéristiques métrologiques de ces dispositifs sont des limitations dont il faut tenir compte pour éviter une détérioration du signal et donc une perte d'information : dérives thermiques et temporelles, bruit de fond, sensibilité aux perturbations électromagnétiques, non linéarité, bande passante limitée ou temps de réponse fini. Concevoir une chaîne d'acquisition, c'est choisir les dispositifs et leur montage de façon que leurs limitations soient compatibles avec la sauvegarde de l'information liée au signal (*Simon, 2013 ; Julien et al., 2013 ; Patrick et al., 2014*).

La généralisation du rôle des ordinateurs dans le contrôle-commande des processus industriels et en particulier, leur association aux chaînes d'acquisition n'a pas supprimé les

problèmes précédents : elle donne, cependant, des moyens nouveaux, automatisés et rapides pour en tenir compte et les corriger, déchargeant ainsi l'opérateur des tâches fastidieuses de réglage, de compensation, d'annulation, etc. En outre, l'ordinateur est capable de remplir, en aval de la chaîne, les fonctions complémentaires à l'acquisition que sont la mémorisation, l'affichage des données, leur analyse et leur traitement en temps réel.

II.5.1. Présentation des cartes Arduino

Arduino, et son synonyme Genuino, est une marque qui couvre des cartes matériellement libres sur lesquelles se trouve un microcontrôleur, le microcontrôleur peut être programmé pour analyser et produire des signaux électriques, de manière à effectuer des tâches très diverses comme la domotique (le contrôle des appareils domestiques - éclairage, chauffage...), le pilotage d'un robot, de l'informatique embarquée, ...

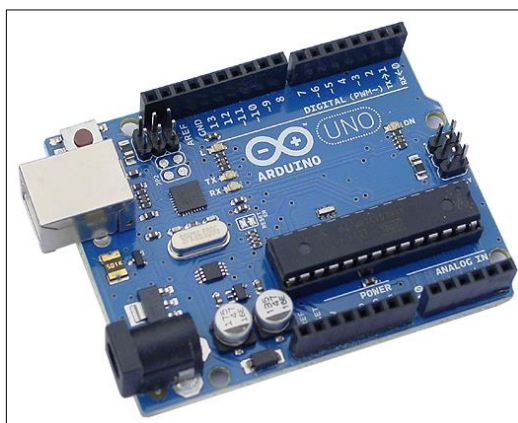


Figure II. 15: Carte Arduino UNO R₃

C'est une plateforme basée sur une interface entrée/sortie simple. Il était destiné à l'origine principalement mais pas exclusivement à la programmation multimédia interactive en vue de spectacles ou d'animations artistiques, ce qui explique en partie la descendance de son environnement de développement de Processing, lui-même inspiré de l'environnement de programmation Wiring (*Simon, 2013 ; Julien et al., 2013 ; Patrick et al., 2014*).

Un module Arduino est généralement construit autour d'un microcontrôleur Atmel AVR et de composants complémentaires qui facilitent la programmation et l'interfaçage avec d'autres circuits. Chaque module possède au moins un régulateur linéaire 5 Voltes et un oscillateur à quartz 16 MHz (ou un résonateur céramique dans certains modèles).

Le microcontrôleur est préprogrammé avec un bootloader de façon qu'un programmeur dédié ne soit pas nécessaire. Le logiciel de programmation des modules Arduino est une application Java, libre et multi-plateforme, servant d'éditeur de code et de compilateur, et qui peut transférer le firmware et le programme au travers de la liaison série RS-232, Bluetooth ou USB selon le module. Il est également possible de se passer de l'interface Arduino, et de compiler et uploader les programmes via l'interface en ligne de commande (*Simon, 2013 ; Julien et al., 2013 ; Patrick et al., 2014*).

Le langage de programmation utilisé est le C++, compilé avec avr-g++ 8, et lié à la bibliothèque de développement Arduino, permettant l'utilisation de la carte et de ses

entrées/sorties. La mise en place de ce langage standard rend aisé le développement de programmes sur les plates-formes Arduino à toute personne maîtrisant le C ou le C⁺⁺.

II.5.2. Principe du fonctionnement

Les différentes versions des Arduino fonctionnent sous le même principe général :

- Des entrées d'alimentation (7 V – 12 V)
- Des ports neutres (masses)
- Sortie 5 V
- Sortie 3,3 V
- Des broches ou pins *numériques* du type logique 0/1 ou tout ou rien, elles offrent en sortie du 5 V et acceptent en entrée du 5 V sur le même principe.
- Des broches ou pins *analogiques* pour contrôler un signal de 0 V à 5 V

L'ensemble des broches (pins) analogiques et numériques peuvent communiquer avec des différents capteurs en émettant et recevant des signaux électriques qui seront contrôlé et traduit par le biais du microcontrôleur selon le programme introduit.

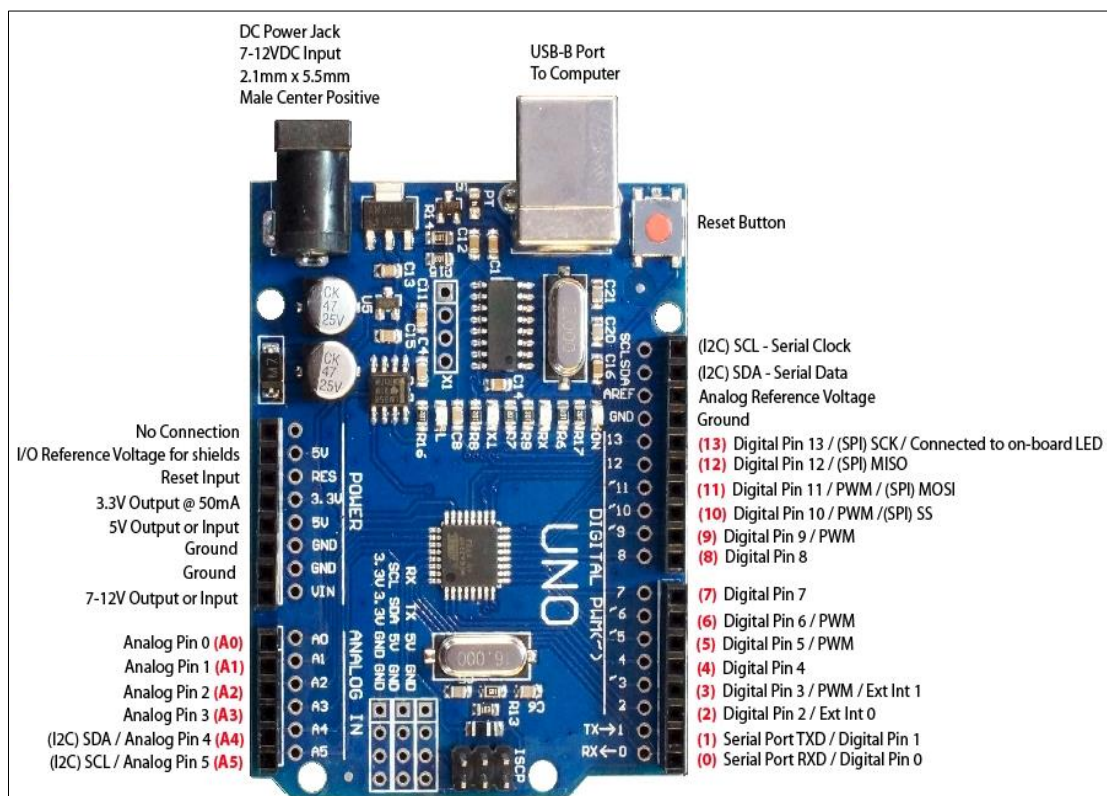


Figure II.16: La nomenclature des broches (pin) d'une carte Arduino du type UNO R₃ (<https://protosupplies.com/product/uno-r3-smd/>)

II.5.3. Capteurs de type MQ

Les capteurs de gaz de la série MQ utilisent un petit radiateur à l'intérieur doté d'un capteur électrochimique. Ils sont sensibles à une gamme de gaz et sont utilisés à l'intérieur à

température ambiante. Ils peuvent être plus ou moins calibrés mais une concentration connue du gaz ou des gaz mesurés est nécessaire pour cela.

La sortie est un signal analogique et peut être lue avec une entrée analogique de l'Arduino. Comme il ne contient aucun composant électronique, la plupart des capteurs peuvent donc être utilisés avec des tensions alternatives et continues.

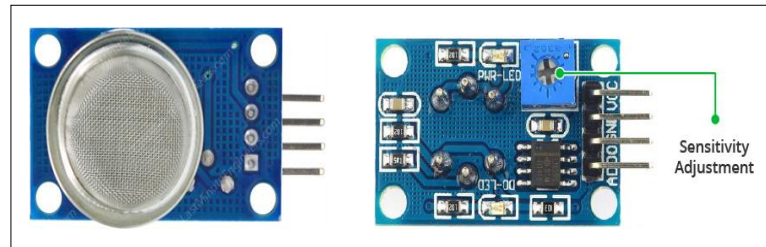


Figure III. 1: Capteur de type MQ
(<https://playground.arduino.cc/Main/MQGasSensors>)

Le câblage préféré consiste à connecter les deux broches «A» ensemble et les deux broches «B» ensemble. C'est plus sûr et on suppose que cela donne des résultats de sortie plus fiables. Bien que de nombreux schémas et fiches techniques indiquent le contraire, il est conseillé de connecter les deux broches «A» ensemble et les deux broches «B» ensemble.

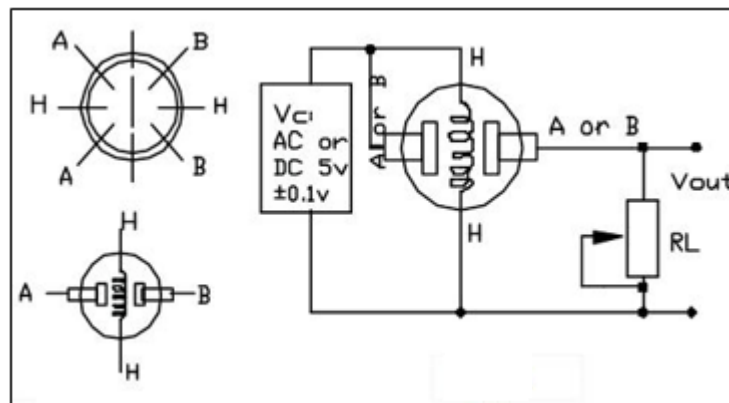


Figure III.2: schéma électrique du capteur de type MQ
(<https://playground.arduino.cc/Main/MQGasSensors>)

Sur cette figure, le chauffage est pour + 5V et est connecté aux deux broches «A». Cela n'est possible que si le chauffage nécessite une tension fixe de + 5V.

La résistance variable dans la figure est la résistance de charge et elle peut être utilisée pour déterminer une bonne valeur. Une résistance fixe pour la résistance de charge est utilisée dans la plupart des cas.

Le radiateur : La tension pour le chauffage interne est très importante. Certains capteurs utilisent +5V pour le chauffage, d'autres ont besoin de 2V. L'appareil de chauffage ne peut pas être connecté directement à une broche de sortie de l'Arduino, car il utilise trop de courant

pour cela. Certains capteurs nécessitent quelques temps pour le chauffage. Un transistor ou une diode devrait également être utilisés dans cette situation pour le chauffage. Les capteurs qui utilisent 5V ou 6V pour le chauffage interne se réchauffent peuvent facilement obtenir 50 ou 60 degrés Celsius. Après le "temps de rodage", le chauffage doit rester allumé pendant environ 3 minutes avant que les lectures ne deviennent stables.

Charge-résistance : Le capteur nécessite une résistance de charge à la sortie de la terre. Sa valeur pourrait être comprise entre 2kOhm et 47kOhm. Plus la valeur est basse, moins la sensibilité est grande. Plus la valeur est élevée, moins les concentrations de gaz sont précises. Si un seul gaz spécifique est mesuré, la résistance de charge peut être étalonnée en appliquant une concentration connue de ce gaz. Si le capteur est utilisé pour mesurer un gaz (comme dans un détecteur de qualité de l'air), la résistance de charge peut être réglée pour une valeur d'environ 1 V en sortie avec de l'air pur. Le choix d'une valeur correcte pour la résistance de charge n'est valable qu'après le temps de rodage.

Burn-in : Certaines fiches techniques utilisent le terme "préchauffer", mais c'est le moment de roder le capteur. Ceci est destiné à rendre les lectures du capteur plus cohérentes. Un temps de 12 ou 24 heures est généralement utilisé pour le temps de rodage. Le rodage est réalisé en appliquant une alimentation normale au capteur (à l'élément chauffant et avec les broches 'A' et 'B' connectées, et avec une résistance de charge). Dans certains cas particuliers, un rodage spécifique est nécessaire.

Chapitre III

Méthodologie expérimentale

« Le vrai but de la science n'est pas de créer de beaux objets : c'est une méthode de réflexion, un moyen d'appréhender l'univers et d'y trouver sa place. »

(Paul Auster 1947)



Chapitre III : Méthodologie expérimentale

III.1. Description de la zone d'étude

L'objectif de cette étude est la modélisation et la prévision de la pollution atmosphérique en TSP et en ET associés à ces TSP et par les polluants gazeux au niveau de Drariïa. Pour ce faire, le choix du site de prélèvement est primordial et doit répondre aux exigences suivantes :

- apport continu en air ;
- absence des contraintes (obstacles) entre l'air et le dispositif de prélèvement.

Le site d'étude se situe au sein du « Centre de Recherche Nucléaire de Drariïa » sise dans la circonscription de Drariïa, wilaya d'Alger, peuplée de 44.141 habitants et qui se trouve à 10 Km au Sud-ouest du centre de la capitale. La zone d'étude est localisée aux coordonnées géographiques suivantes : 36°43'32.18''Nord et 3°00'29.15''Est.

Le système de prélèvement a été installé sur le toit d'un bâtiment de trois étages, d'environ 17,5 mètres au-dessus du niveau du sol et 10 mètres loin de la route la plus proche ayant un trafic d'environ 30.000 véhicules par jour.

Le trafic est une source majeure de pollution de l'air dans cette région, principalement en raison de l'abrasion des garnitures de frein et pneus suivis par la remise en suspension de la poussière de route générée par des vents relativement forts marquant généralement cette région.

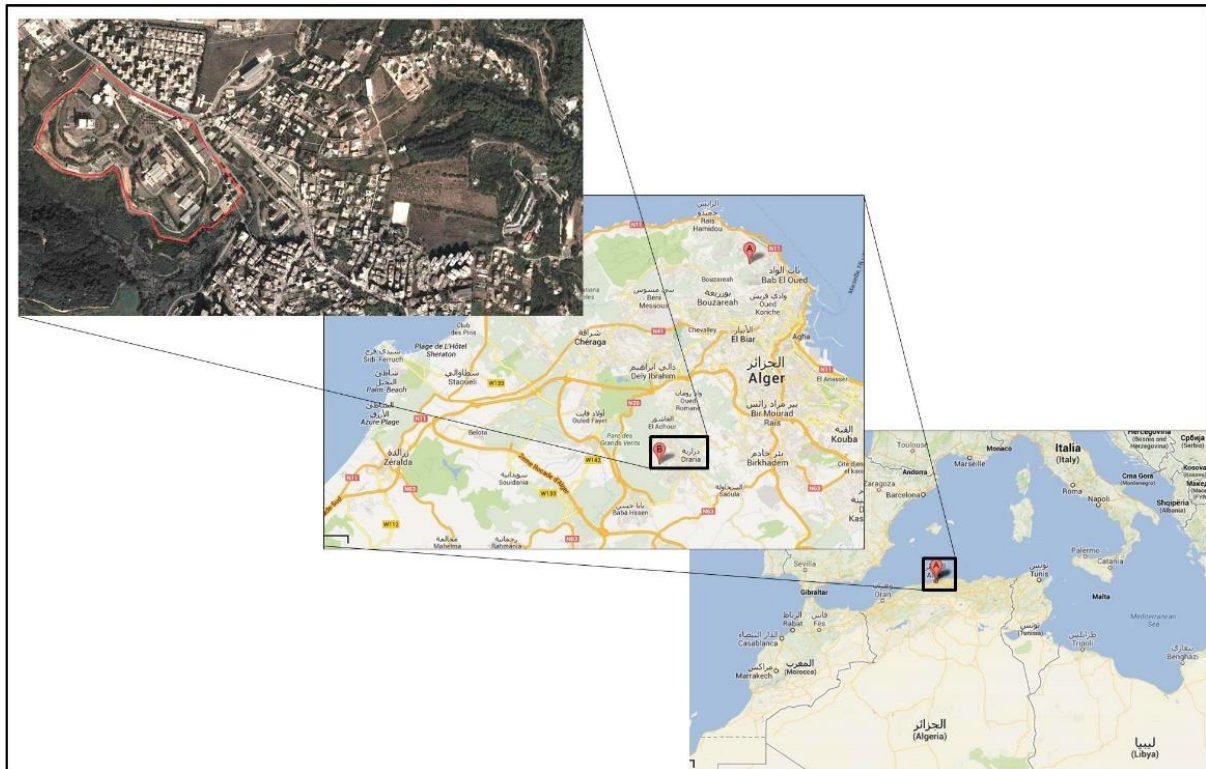


Figure III.3: Situation géographique du site de prélèvement

III.2. Echantillonnage des TSP

Le prélèvement de l'air a été effectué par un dispositif installé au niveau de la terrasse de la division DPAN (Division de Physique et Applications Nucléaires) sur une hauteur de 17,5 mètres par rapport au niveau du sol, ce dernier est à 200m d'altitude par rapport au niveau de la mer.

Le préleveur est composé des quatre équipements suivants :

1. Corps de protection : conçu pour fixer la tête d'aspiration ainsi pour la protection du filtre de prélèvement contre la pluie, oiseaux...
2. Porte filtre ou cassette fixé dans une tête de prélèvement en inox.
3. Pompe d'aspiration : pour aspirer l'air, réglée à un débit d'aspiration fixe.
4. Volumètre : quantification du volume d'air aspiré.

Le débit de prélèvement de l'air a été fixé à 1 m³/h, l'équivalent de la respiration humaine, pour pouvoir évaluer le risque de l'exposition des personnes à cette pollution vu que ce site d'étude s'agit d'un site urbain. Un prélèvement a été effectué chaque les 48 heures et parfois chaque 72 heures, durant la période du prélèvement toute remarque pouvant influencer sur les résultats (pluie, neige, vent violent...) a été mentionné (*Bouhila et al., 2015*).

La collecte des aérosols a été faite sur des filtres membranaires (en nitrate de cellulose) ouverts sans coupure granulométrique précise et de porosité de 0,8µm (type AA) dotés d'un diamètre de 37 mm. Ces filtres répondaient à tous les exigences requises (*Bouhila et al., 2015*).

III.3. Echantillonnage des polluants gazeux

Les capteurs électrochimiques des gaz polluants sont des détecteurs de gaz qui mesurent la concentration d'un gaz cible en oxydant ou en réduisant le gaz cible au niveau d'une électrode et en mesurant le courant résultant.

Les capteurs contiennent deux ou trois électrodes, parfois quatre, en contact avec un électrolyte. Les électrodes sont typiquement fabriquées en fixant un métal précieux de surface élevée sur la membrane hydrophobe poreuse. L'électrode de travail entre en contact avec l'électrolyte et l'air ambiant à surveiller généralement via une membrane poreuse. L'électrolyte le plus couramment utilisé est un acide minéral, mais des électrolytes organiques sont également utilisés pour certains capteurs. Les électrodes et le boîtier se trouvent généralement dans un boîtier en plastique contenant un trou d'entrée de gaz pour les contacts de gaz et électriques.

Le gaz diffuse dans le capteur par l'arrière de la membrane poreuse jusqu'à l'électrode de travail où il est oxydé ou réduit. Cette réaction électrochimique se traduit par un courant électrique qui traverse le circuit externe. En plus de mesurer, d'amplifier et d'exécuter d'autres fonctions de traitement du signal, le circuit externe maintient la tension aux bornes du capteur entre les électrodes de travail et de compteur pour un capteur à deux électrodes ou entre les électrodes de travail et de référence pour une cellule à trois électrodes. Au niveau de la contre-électrode, une réaction égale et opposée se produit, telle que si l'électrode de travail est une oxydation, la contre-électrode est une réduction.

L'intensité du courant est contrôlée par la quantité de gaz cible oxydée au niveau de l'électrode de travail. Les capteurs sont généralement conçus de telle sorte que l'alimentation en gaz soit limitée par la diffusion, de sorte que la sortie du capteur est linéairement proportionnelle à la concentration de gaz. Cette sortie linéaire est l'un des avantages des capteurs électrochimiques par rapport aux autres technologies de capteurs (infrarouge, par exemple), dont la sortie doit être linéarisée avant de pouvoir être utilisée. Une sortie linéaire permet une mesure plus précise des faibles concentrations et une calibration beaucoup plus simple (il suffit de connaître la ligne de base et un point).

Le contrôle de la diffusion offre un autre avantage. Changer la barrière de diffusion permet au fabricant du capteur d'adapter le capteur à une plage de concentration de gaz cible particulière. De plus, comme la barrière de diffusion est essentiellement mécanique, la calibration des capteurs électrochimiques tend à être plus stable dans le temps et les instruments à capteurs électrochimiques nécessitent donc beaucoup moins de maintenance que certaines autres technologies de détection. En principe, la sensibilité peut être calculée en fonction des propriétés de diffusion du circuit de gaz dans le capteur, bien que des erreurs expérimentales dans la mesure des propriétés de diffusion rendent le calcul moins précis que l'étalonnage avec un gaz de test.

Pour certains gaz tels que l'oxyde d'éthylène, la sensibilité croisée peut être un problème car l'oxyde d'éthylène nécessite un catalyseur d'électrode de travail très actif et un potentiel de fonctionnement élevé pour son oxydation. Par conséquent, les gaz qui s'oxydent plus facilement, tels que les alcools et le monoxyde de carbone, donneront également une réponse. Les problèmes de sensibilité croisée peuvent être éliminés grâce à l'utilisation d'un filtre chimique, par exemple des filtres qui permettent au gaz cible de traverser librement, mais qui réagissent avec les interférences courantes et les suppriment.

Les capteurs électrochimiques offrent de nombreux avantages, mais ils ne conviennent pas à tous les gaz. Comme le mécanisme de détection implique l'oxydation ou la réduction du gaz, les capteurs électrochimiques ne conviennent généralement que pour les gaz électrochimiquement actifs, bien qu'il soit possible de détecter les gaz électrochimiquement inertes indirectement si le gaz interagit avec une autre espèce dans le capteur qui produit alors une réponse. Les capteurs de dioxyde de carbone sont un exemple de cette approche et sont disponibles dans le commerce depuis plusieurs années.

III.3.1. Capteurs utilisés

III.3.1.1. Capteur MQ₂ (Annexe 1)

Le matériau sensible du capteur de gaz MQ₂ est le SnO₂, qui présente une conductivité inférieure dans de l'air pur. Quand les combustibles cibles existent, la conductivité du capteur est plus élevée avec la concentration de gaz en hausse.

La structure et la configuration du capteur de gaz MQ₂ sont : capteur composé d'un micro tube en céramique AL₂O₃, étain La couche sensible au dioxyde (SnO₂), l'électrode de mesure et le réchauffeur sont fixés dans une croûte en plastique et en acier inoxydable net. L'appareil de chauffage fournit les conditions de travail nécessaires au travail des composants sensibles. Le MQ₂ enveloppé a 6 broches, 4 d'entre eux sont utilisés pour récupérer des signaux, et 2 autres sont utilisés pour fournir un courant de chauffage.



Figure III.4: Capteur MQ2

(<https://lastminuteengineers.com/mq2-gas-senser-arduino-tutorial>)

Le capteur de gaz MQ₂ fonctionne sur une tension continue de 5 V et consomme environ 800 mW. Il peut détecter les concentrations de GPL, de fumée, d'alcool, de propane, d'hydrogène, de méthane et de monoxyde de carbone entre 200 et 10 000 ppm.

Le capteur est actuellement enfermé dans deux couches de maillage fin en acier inoxydable appelé réseau anti-explosion. Cela garantit que l'élément chauffant à l'intérieur du capteur ne provoquera pas d'explosion, car nous détectons des gaz inflammables.

Il protège également le capteur et filtre les particules en suspension, de sorte que seuls les éléments gazeux puissent passer à l'intérieur de la chambre. La maille est liée au reste du corps par un anneau de serrage plaqué cuivre.

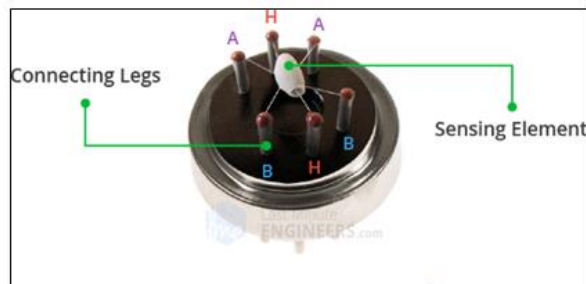


Figure III.5: Broches de connexion d'un capteur type MQ

(<https://lastminuteengineers.com/mq2-gas-senser-arduino-tutorial>)

Voici à quoi ressemble le capteur lorsque le maillage externe est retiré. La structure en forme d'étoile est formée par l'élément de détection et six branches de connexion qui s'étendent au-delà de la base en bakélite. Sur six conducteurs, deux (H) sont responsables du chauffage de l'élément de détection et sont connectés par l'intermédiaire d'une bobine en nickel-chrome, alliage conducteur bien connu.

Les quatre conducteurs restants (A & B) responsables des signaux de sortie sont connectés à l'aide de fils Platine. Ces fils sont connectés au corps de l'élément de détection et transmettent de faibles variations du courant qui traverse l'élément de détection.

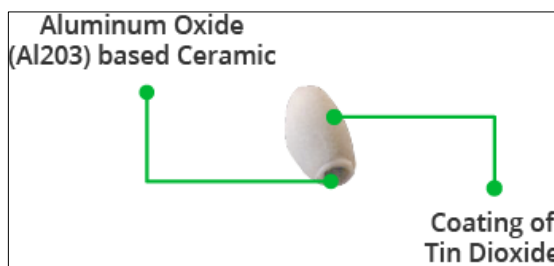


Figure III.6: Elément de détection tubulaire d'un capteur type MQ
(<https://lastminuteengineers.com/mq2-gas-senser-arduino-tutorial>)

L'élément de détection tubulaire est constitué de céramique à base d'oxyde d'aluminium (Al_2O_3) et est recouvert de dioxyde d'étain (SnO_2). Le dioxyde d'étain est le matériau le plus important sensible aux gaz combustibles. Cependant, le substrat en céramique augmente simplement l'efficacité du chauffage et garantit que la zone du capteur est constamment chauffée à une température de fonctionnement.

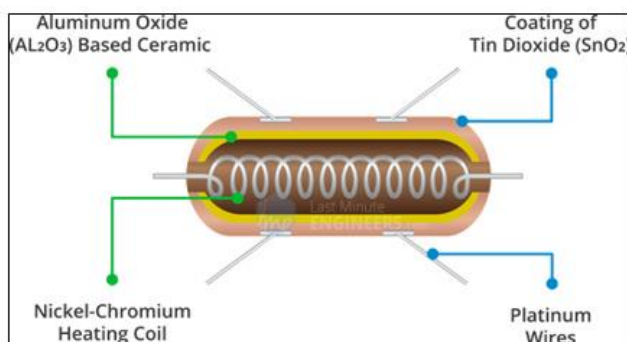


Figure III.7: Vue intérieure de la bobine de détection
(<https://lastminuteengineers.com/mq2-gas-senser-arduino-tutorial>)

Ainsi, la bobine à base de nickel-chrome et la céramique à base d'oxyde d'aluminium forment un système de chauffage ; tandis que les fils de platine et le revêtement de dioxyde d'étain forment un système de détection.

III.3.1.2. Capteur MQ₃ (Annexe 2)

La structure et la configuration du capteur de gaz MQ₃ est la suivante : capteur composé de micro tube en céramique Al_2O_3 , couche sensible au dioxyde d'étain (SnO_2), électrode de mesure et élément chauffant sont fixés dans croûte faite de plastique et filet en acier inoxydable. Le radiateur fournit les conditions de travail nécessaires au travail de composants sensibles. Le MQ₃ enveloppé a 6 broches, 4 d'entre elles sont utilisées pour récupérer les signaux, et 2 autres sont utilisées pour fournir un courant de chauffage

III.3.1.3. Capteur MQ₇ (Annexe 3)

La structure et la configuration du capteur de gaz MQ₇ est la suivante : le capteur est composé d'un micro tube en céramique Al_2O_3 , d'une couche sensible au dioxyde d'étain (SnO_2), d'une électrode de mesure et d'un chauffage dans une croûte en plastique et filet en acier

inoxydable. L'appareil de chauffage fournit les conditions de travail nécessaires au travail des composants sensibles.

III.3.1.4. Capteur MQ₁₃₅ (Annexe 4)

La structure et la configuration du capteur de gaz MQ₁₃₅ est la suivante : le capteur est composé d'un micro tube en céramique AL₂O₃, d'une couche sensible au dioxyde d'étain (SnO₂), d'une électrode de mesure et d'un chauffage dans une croûte en plastique et filet en acier inoxydable. L'appareil de chauffage fournit les conditions de travail nécessaires au travail des composants sensibles.

III.3.1.5. Capteur DHT₁₁ (Annexe 5)

Le DHT11 détecte la vapeur d'eau en mesurant la résistance électrique entre deux électrodes. Le composant de détection d'humidité est un substrat de rétention d'humidité avec des électrodes appliquées à la surface. Lorsque la vapeur d'eau est absorbée par le substrat, des ions sont libérés par le substrat, ce qui augmente la conductivité entre les électrodes. La variation de résistance entre les deux électrodes est proportionnelle à l'humidité relative. Une humidité relative plus élevée diminue la résistance entre les électrodes, tandis qu'une humidité relative plus basse augmente la résistance entre les électrodes.

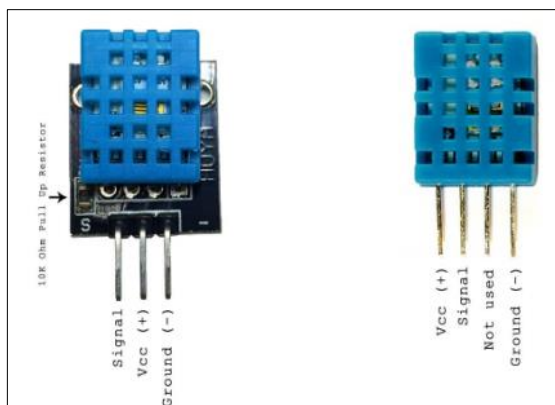


Figure III.8: Capteur DHT11

(<https://www.circuitbasics.com/how-to-set-up-the-dht11-humidity-sensor-on-an-arduino>)

Le DHT11 mesure la température à l'aide d'un capteur de température NTC (thermistance) intégré à l'appareil.



Figure III.9: Capteur DHT₁₁ décortiqué

(<https://www.circuitbasics.com/how-to-set-up-the-dht11-humidity-sensor-on-an-arduino>)

Un CI monté à l'arrière de l'unité convertit la mesure de résistance en humidité relative. Il stocke également les coefficients d'étalonnage et contrôle la transmission du signal de données entre le DHT11 et l'Arduino:

Le DHT11 utilise un seul câble de signal pour transmettre des données à l'Arduino. L'alimentation provient de +5V et de fils de terre séparés. Une résistance de charge de 10K Ohm est nécessaire entre la ligne de signal et la ligne +5V pour garantir le niveau de signal élevé par défaut.

III.3.2. Montage de la station de prélèvement

Dans le but de concevoir une station intelligente de prélèvement des gaz polluants, nous avons utilisé :

- Une carte programmable Arduino de type UNO R₃ ;
- Quatre capteurs de gaz (MQ₂, MQ₃, MQ₇ et MQ₁₃₅) ;
- Un capteur de température et humidité (DHT₁₁) ;
- Un module (Real Time Clock) pour mémoriser la date et l'heure du prélèvement ;
- Une carte mémoire de type micro SD ;
- Un transistor NPN ;
- Lampe témoin LED.

L'ensemble des capteurs et des modules sont alimentés via la broche +5V de la carte Arduino, la borne négative (-) de ces derniers à l'exception du module RTC sont reliés au collecteur du transistor, l'émetteur du transistor est branché dans la broche négative de la carte Arduino. Le module RTC utilise une broche négative (-) de la carte Arduino indépendamment des autres modules.

Les capteurs MQ sont reliés aux broches analogiques A0, A1, A2 et A3, le capteur DHT₁₁ à la broche digitale D7, la lampe LED qui indique que la station fonctionne selon l'algorithme injecté au microcontrôleur est branchée sur la broche digitale D3. La base du transistor est branchée sur la broche digitale D2

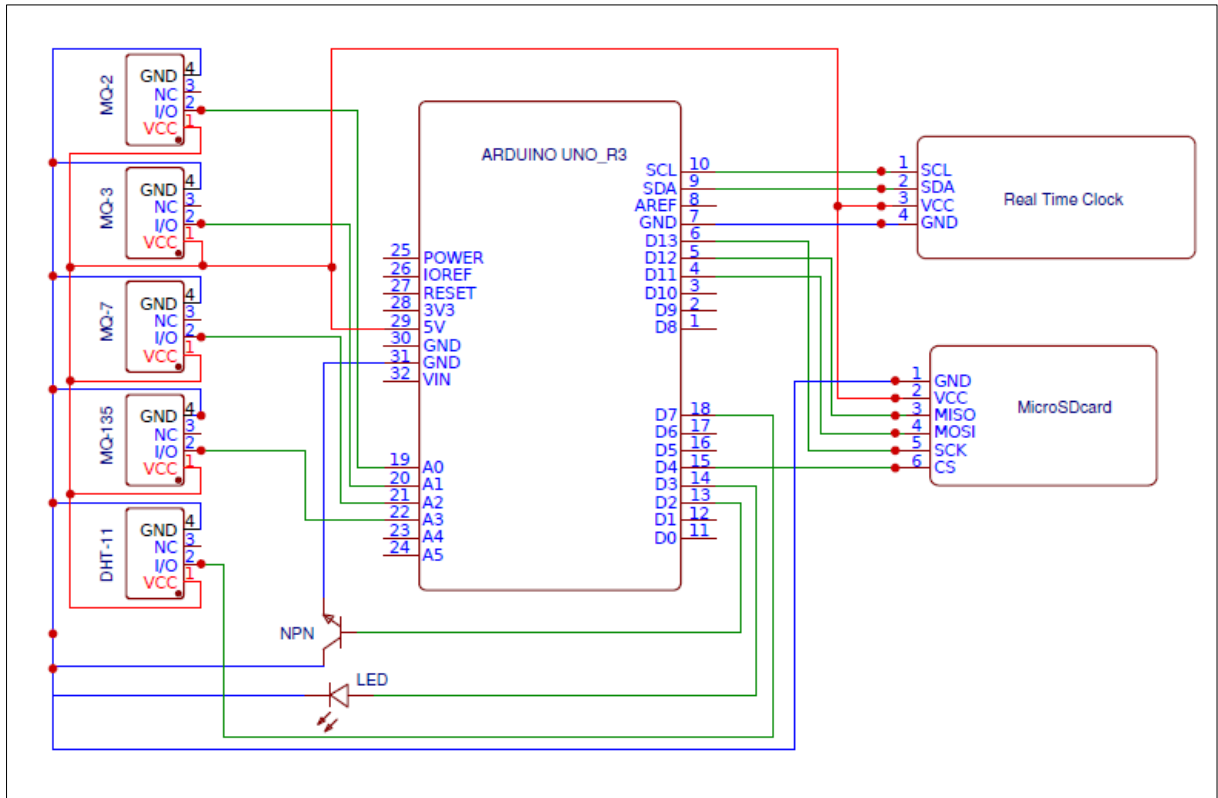


Figure III.10: Montage du circuit électrique de la station de prélèvement des gaz polluants

Le nombre trop important des fils reliant les différents composants et qui assurent la transmission des différents signaux nous a poussé à réfléchir comment les diminuer et les organiser, la réponse à cette problématique était d’opter pour les circuits imprimés à double couche. Dans la PCB nous avons créé une prise positive (+) et une prise négative (-) notées $V_{i(i=0,...6)}$ et $G_{i(i=0,...6)}$ simultanément. Ces prises servent comme broche d’alimentation pour chaque module et capteur

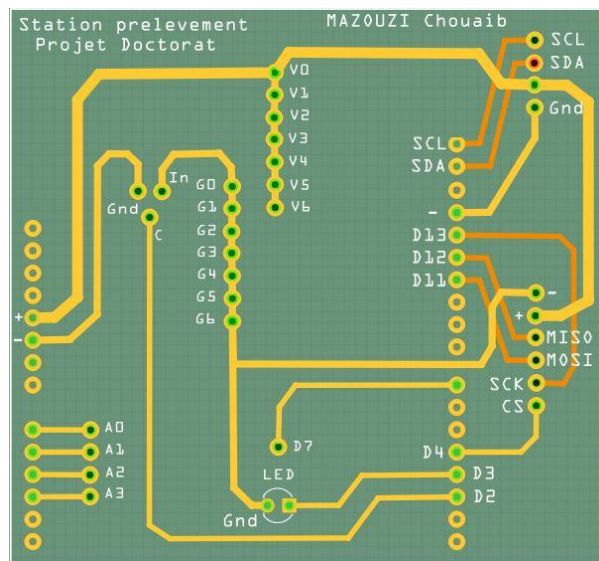


Figure III.11: Circuit imprimé à double couche de la station de prélèvement. Circuit jaune : face recto, circuit orangé : face verso.

III.3.3. Algorithme de la station de prélèvement

Notre station de prélèvement des gaz polluants a été placée au même endroit que celui du dispositif de prélèvement des TSP, nous avons fixé un intervalle d'échantillonnage équivalent à un prélèvement chaque les 03 heures.

La carte Arduino est branchée à un transformateur de courant qui délivre +12V qui assure son alimentation continue par l'électricité, la première étape consiste à allumer l'ensemble des capteurs et modules pendant un temps déterminé (15min) pour faire chauffer les capteurs, pour ce faire on envoie un signal électrique à la broche digitale D2 qui est reliée à la base du transistor, cette charge va activer le transistor du coup faire passer le courant du collecteur vers l'émetteur. Après le chauffage on passe à l'étape de l'acquisition des données captées par les différents capteurs puis enregistrer les valeurs obtenue sur la carte mémoire. Si tous les modules et les capteurs fonctionnent suivant l'algorithme désiré la lampe LED clignote une seule fois, comme dernière étape on procède à l'extinction de l'ensemble des modules et des capteurs en coupant le signal électrique de la broche digitale D2.

Après l'étape de l'extinction on calcul les différentes concentrations enregistrées, pour ce faire, on convertie le signal électrique en une valeur numérique exploitable suivant la fiche technique de chaque capteur. On procède au calcul de l'écart type pour détecter n'importe quelle anomalie et puis faire une investigation si le problème provient du capteur ou bien de l'échantillon lui-même. Si une anomalie est enregistrée (variation brutale de la concentration) on impose un nouveau délai d'échantillonnage égale au un dixième (1/10) du délai précédent. Une fois l'anomalie est résolue on réutilise l'ancien délai fixé à un prélèvement chaque les 3 heures

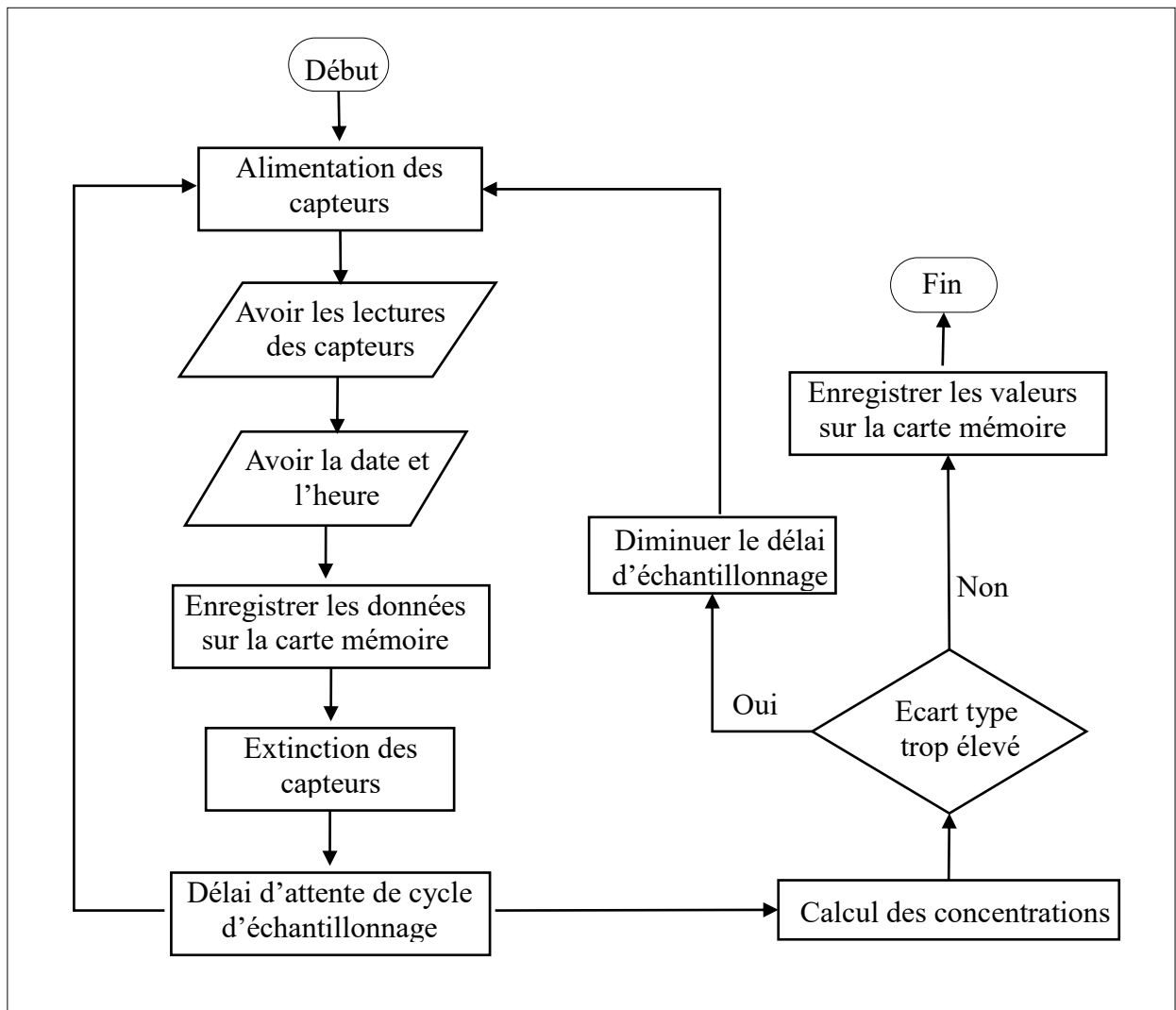


Figure III.12: Algorithme de la station de prélèvement des gaz polluants

III.4. Analyse des TSP et ET associés

III.4.1. Analyse des TSP

Le filtre neuf (avant prélèvement) qui va être placé dans le dispositif d'échantillonnage a été pesé, puis après l'étape de l'échantillonnage le filtre récupéré de l'échantillonneur est pesé à nouveau à l'aide d'une balance analytique de précision de 0,1mg. Le volume d'air aspiré a été noté au cours de la période d'échantillonnage (Bouhila et al., 2015).

La concentration des TSP dans l'air est calculée par l'équation suivant :

$$TSP(\mu\text{g}/\text{m}^3) = \frac{m_2 - m_1}{V} \cdot 10^3$$

TSP : Concentration des particules totales en suspension dans l'air ($\mu\text{g}/\text{m}^3$).

m_1 : Masse du filtre avant prélèvement (mg).

m_2 : Masse du filtre après prélèvement (mg).

V : Volume d'air aspiré.

L'application de cette relation nécessite la condition que les particules retenues sur le filtre pendant l'aspiration ne sont à l'origine qu'à cette dernière et non pas au vent infiltré dans le système de prélèvement et des dépôts accidentels. Pour ce faire, le fonctionnement de l'échantillonneur a été simulé en lui imposant une contrainte de vent égale à 60m/s dans les trois directions (x,y,z). Cette simulation a été faite avec le logiciel professionnel SolidWorks 2010–Flow Simulation la figure III.14 illustre le résultat de la simulation de la circulation d'air au niveau de site de prélèvement (*Bouhila et al., 2015*).

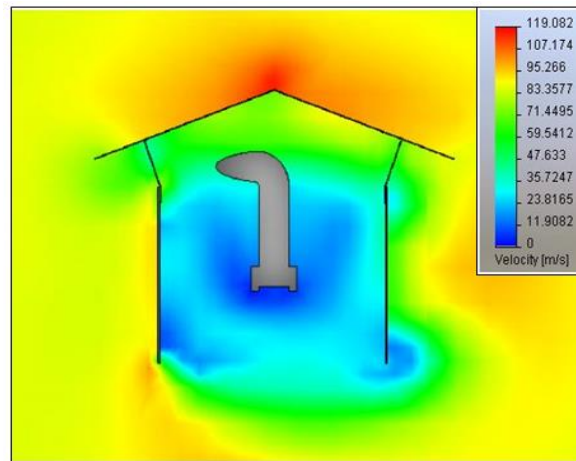


Figure III.13: Simulation d'écoulement d'air autour du système d'échantillonnage

III.4.2. Analyse des ET associés aux TSP

L'analyse des échantillons (filtres) ont été analysés par la méthode d'Analyse par Activation Neutronique en se basant sur ses nombreux avantages offerts par cette technique (la haute sensibilité, la non destruction de l'échantillon, multi-élémentaire...) et qui s'est déroulée au niveau des canaux d'irradiation des canaux d'irradiation du réacteur nucléaire NUR de Draria doté d'une puissance de 1 méga watt (*Bouhila et al., 2015*).

L'utilisation de la méthode absolue n'était pas pratique à cause de l'imprécision de la connaissance de la fluctuation du flux neutronique dans l'espace et dans le temps, la raison pour laquelle la méthode relative a été utilisée dont le principe de quantification des éléments retenus dans les filtres repose sur des standards qui servent comme références pour le calcul des concentrations en polluants dans les échantillons, l'emploi de plusieurs standards dans une telle expérience est très utile pour pouvoir confirmer la concentration d'un élément par rapport à plusieurs références (*Bouhila et al., 2015*).

Les trois standards qu'ils ont été utilisés sont :

- GSV4 : feuilles de thé chinois
- Soil7 : sol
- NIST : sol

A l'aide des abaques (*Adams and Dams, 1970*) qui associent chaque énergie à un radio-isotope, les spectres ont été ainsi traduits. Dans plusieurs cas, des problèmes d'identification et d'interprétation des résultats bruts peuvent être posés, citons à titre d'exemple le phénomène d'interférences. Une fois un élément déterminé, son NET et son temps de décroissance radioactive sont utilisés pour le calcul de sa concentration.

Pour le calcul, l'équation de la méthode relative est adaptée comme suit :

$$C_e (ng / m^3) = C_s \cdot m_s \cdot \frac{NET_e \cdot e^{\lambda \cdot tde}}{NET_s \cdot e^{\lambda \cdot tds}} \cdot \frac{2}{V}$$

C_e : Concentration de l'élément dans l'échantillon (ng/m³).

C_s : Concentration de l'élément dans le standard (µg/g de standard).

m_s : Masse du standard (mg).

NET_e : Net de l'élément dans l'échantillon.

NET_s : Net de l'élément dans le standard.

tde : Temps de décroissance de l'élément dans l'échantillon (s).

tds : Temps de décroissance de l'élément dans le standard (s).

V : Volume d'air aspiré (m³).

Facteur 2 : Car que la moitié du filtre a été analysée.

III.5. Traitement des données brutes

La performance d'une étude de modélisation est dépendante en premier lieu de la disponibilité et de la qualité des données nécessaires. Dans le cas de la prévision de la pollution atmosphérique, les modèles proposés exigent un historique de données complet et qui contient le maximum d'informations significatives. En effet, des tests statistiques ont été élaborés afin de traiter et d'analyser les séries temporelles pour en tirer le maximum d'informations. Ainsi, des méthodes peuvent être utilisées pour combler les lacunes des séries tronquées.

III.5.1. Intérêt d'une série temporelle

C'est la suite d'observations d'une seule variable x durant une période de temps t donnée. En pratique, c'est tout ce qui est chiffrable et varie en fonction du temps. La dimension temporelle est ici importante car il s'agit de l'analyse d'une chronique historique : des variations d'une même variable au cours du temps, afin de pouvoir comprendre la dynamique. La périodicité de la série n'importe en revanche pas : il peut s'agir de mesures quotidiennes, mensuelles, trimestrielles, annuelles, voire même sans périodicité.

Les types de problèmes le plus souvent rencontrés lors du traitement des données atmosphériques sont (*Lamas, 1985*) :

- Le manque de données : station non existante ou à l'arrêt ;
- Non continuité des données : jour férié ou weekend ;
- La mauvaise qualité des données, qui peut dépendre de deux causes :

- Données inconsistantes : lorsqu'elles sont entachées d'une erreur systématique due à des instruments de mauvaise qualité ou non calibrés.
- Données non homogènes : lorsqu'il se produit à un moment donné (rejet d'une usine, incendie...).

Afin de garantir la qualité des données, nous devons suivre les étapes suivantes :

- Comblent les lacunes des séries observées si elles existent ;
- Détecter et corriger les valeurs douteuses ;
- Vérifier l'homogénéité et l'indépendance des données.

Dans la pratique courante, de nombreuses causes concourent pour entraîner l'absence de certaines informations (absence de mesures, défaut de transmission, erreur de stockage, manipulation erronée, ...), ce qui fait que dans les fichiers disponibles certaines données sont manquantes.

Nous pouvons estimer les valeurs manquantes ou erronées en se basant sur plusieurs méthodes :

- Remplacer la valeur manquante par la moyenne des valeurs voisines : Cette méthode est utilisée lorsque les valeurs moyennes annuelles de la lecture à compléter ne diffèrent pas plus de 10 % des valeurs moyennes annuelles des échantillons de référence ;
- Remplacer la valeur manquante par une moyenne pondérée par la tendance annuelle des échantillons ;
- Remplacer la valeur manquante par la moyenne interannuelle de la série considérée : Une donnée journalière manquante, par exemple, peut être remplacée soit par la moyenne entre la valeur des jours précédant et suivant, soit par la moyenne du mois ou bien par la moyenne journalière interannuelle du jour considéré. S'il s'agit d'une valeur mensuelle, nous estimons la valeur manquante soit par la moyenne mensuelle des mois disponibles ou bien par la moyenne mensuelle interannuelle du mois considéré ;
- Les méthodes basées sur l'analyse de régressions et corrélations ;
- Méthode basée sur la simulation.

III.5.2. Détection des valeurs singulières

Il arrive fréquemment que des erreurs se glissent dans les données de base et en dénaturent complètement le contenu. Cela peut se produire au cours de la mesure, de la transmission, du codage, du stockage ou du décodage des observations, ou à cause des valeurs intégrées après le comblement des lacunes.

Il est important donc avant tout traitement de détecter les erreurs grossières de ce genre, afin que l'on puisse effectuer une analyse plus fine pour décider si cette donnée douteuse est aberrante et doit donc être éliminée, ou si c'est une donnée exceptionnelle, qui par conséquent doit être au contraire précieusement conservée, nous nous sommes basé sur trois tests :

- Test de Grubbs et Beck (*Grubbs et al., 1972*) ;
- Test d'homogénéité ;
- Test d'indépendance.

III.6. Modélisation des TSP et ET associés par le modèle ARIMA

La modélisation ARIMA utilise de façon très simple des données récentes ou plus anciennes afin de modéliser les données existantes et de réaliser des prévisions adéquates concernant un comportement futur. L'objectif est d'identifier un modèle sous-jacent qui explique le changement dans le procédé. Tout point qui s'éloigne de ce comportement prévu pourrait être considéré comme une cause spéciale, puisqu'il ne suit pas les mouvements généraux dans les données.

Pour créer un modèle ARIMA, les données doivent être stationnaires, c'est-à-dire que le procédé ne doit présenter aucune tendance, ni à la hausse, ni à la baisse. En d'autres termes, il faut atteindre une certaine stabilité dans la moyenne du procédé

Les données auxquelles nous avons appliqué une différence d'ordre 1 semblent être devenues stationnaires, sans tendance marquée à la hausse ou à la baisse. Cela signifie que nous pouvons utiliser ces données pour établir un modèle ARIMA. Tout d'abord, nous allons effectuer une analyse d'auto-corrélation et d'auto-corrélation partielle sur les données différenciée.

Nous avons une fonction d'auto-corrélation avec un schéma sinusoïdal (onde sinusoïdale) et des pics pour les décalages 1 à 3, ce qui suggère un modèle autorégressif d'ordre 3, ou AR(3). Le comportement sinusoïdal sur la fonction d'auto-corrélation partielle et les pics jusqu'au décalage 3 suggèrent un modèle de moyenne mobile d'ordre 3, ou MA(3). La modélisation de série chronologique peut se révéler un procédé itératif, voire hasardeux. Cependant, ces graphiques suggèrent que le modèle ARIMA (3,1,3) est un bon point de départ.

Chaque partie du modèle ARIMA joue un rôle dans les prévisions obtenues à partir de celui-ci. La partie autorégressive du modèle prévoit la valeur au temps t en prenant en compte les valeurs précédentes de la série aux temps t_1, t_2 , etc. La moyenne mobile utilise les valeurs résiduelles anciennes : les différences entre la valeur réelle et la valeur prévue en se basant sur le modèle au temps t .

Les valeurs de p sont seulement significatives au niveau des 10 % pour le coefficient de premier ordre de la partie autorégressive du modèle et le coefficient de 3^{ème} ordre de la partie moyenne mobile du modèle. En outre, la statistique du Khi deux Ljung-Box, qui teste le caractère aléatoire global du modèle, suggère qu'il peut y avoir un effet saisonnier d'ordre 1, au moins.

Par conséquent, nous allons essayer de comprendre ces données de façon plus approfondie en créant un modèle ARIMA (1,1,3) (1,0,0). Cette notation n'est pas aussi compliquée qu'elle le paraît. Le premier ensemble de parenthèses nous indique que les parties autorégressive (AR) et intégrée (1) du modèle seront basées sur des décalages 1, alors que la moyenne mobile (MA) sera basée sur le décalage 3. Le deuxième ensemble de parenthèses indique l'effet saisonnier, qui, nous supposons, suit un cycle de 12 mois, autrement dit, annuel, autour d'AR(1).

Le coefficient autorégressif de premier ordre, le coefficient saisonnier et le coefficient de moyenne mobile de 3^{ème} ordre sont tous significatifs au niveau d'alpha de 10 %, ce qui indique que le modèle pourrait être efficace. La somme des carrés qui mesure la somme des différences quadratiques entre chaque point de données d'origine et sa valeur estimée à l'aide de ce modèle ARIMA est assez faible. De plus, la statistique du Khi deux Ljung-Box ne met en évidence aucune corrélation entre les points avec une différence de 12 ou 24 décalages ; même le coefficient saisonnier les a exclus.

Il est crucial pour ARIMA, comme dans la régression ou la modélisation ANOVA, d'examiner le comportement des valeurs résiduelles pour voir si elles sont normales, aléatoires ou si elles ont une variation constante. Dans ce cas-là, les valeurs résiduelles représentent les différences entre la valeur observée au temps t et la valeur prévue en se basant sur le modèle ARIMA. Ces différences peuvent être négatives ou positives, et parfois nulles lorsque l'ajustement est parfait.

Les hypothèses sont assez bien satisfaites, sauf qu'il existe des variations non constantes sur la représentation graphique en fonction des valeurs ajustées. Cela s'explique par le fait que le modèle est mieux ajusté pour les anciens points de données que pour les plus récents.

III.7. Modélisation des TSP et ET associés par les réseaux de neurones artificiels

III.7.1. Détermination des entrées du modèle

L'analyse factorielle exploratoire a pour objectif de chercher les relations entre individus (entrées) par l'évaluation de leurs ressemblances, ainsi que les relations entre variables par l'évaluation de leurs liaisons. Dans le cadre de cette thèse, nous avons cherché la liaison existante entre l'historique de l'observation à modéliser, le volume d'air aspiré et la concentration des TSP.

L'AFE a révélé l'existence d'une forte liaison entre l'observation du jour J et $J-1$, la représentation graphique des deux composantes principales issues de l'ACP de chaque polluant à modéliser a montré qu'il y a deux nuages : le premier nuage regroupe l'observation du jour J et $J-1$, le deuxième nuage regroupe les observation des $J-i$ ($i=2,3,4$), la contribution des TSP dans l'explication du modèle est minime, voir négligeable.

Nous avons remarqué qu'il existe une forte corrélation entre l'ensemble des observations au jour J relatives à l'élément à modéliser et les observations des jours $J-i$ ($i=1,2,3,4$), c'est-à-dire que la concentration d'un élément au jour J peut être exprimée en fonction de ses concentrations aux jours $J-1$, $J-2$, $J-3$ et $J-4$.

Nous avons utilisé le logiciel IBM-SPSS Statistique v19 pour accomplir l'analyse factorielle exploratoire.

III.7.2. Stratégie adoptée

Le modèle des réseaux de neurones opté dans cette thèse est le perceptron multicouche à sens unique (*feedforward* en anglais). Nos essais et nos tests consistaient à élaborer plusieurs modèles de réseaux de neurones en modifiant le nombre de couches, le nombre de neurones dans chaque couche ainsi que les fonctions de transfert. Cette démarche a été adoptée aussi bien pour les TSP que pour les ET associés à ces TSP.

Ainsi, notre choix s'est limité pour des modèles ayant une fonction transfert de type sigmoïde (tangente sigmoïde ou logarithme sigmoïde) dans la couche cachée, et de type linéaire dans la couche de sortie. Des travaux précédents ont montré que ce couple de fonctions permet d'approximer quasiment tous les types de relations non linéaires (*Dreyfus et al., 2002*).

Notre objectif est de construire un réseau de neurones capable d'apprendre les données qu'on lui propose mais aussi qui offre une bonne généralisation à des fins de prédiction, en d'autres termes qui évite autant que possible l'étape du sur-ajustement lors de l'apprentissage. Pour ce faire, nous retirons 20 % d'observations de base qui serviront de test de validation et les 80% d'observations restantes seront utilisées pour l'étape de l'apprentissage.

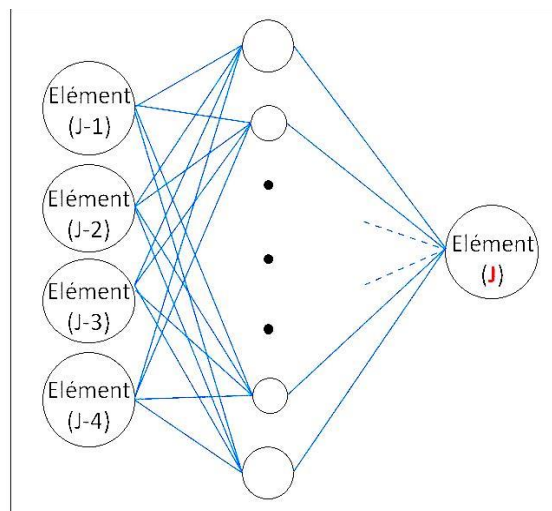


Figure III.14: Schéma de réseau de neurones pour la modélisation des TSP et ET associés

L'apprentissage des données est un apprentissage supervisé par l'algorithme de rétro propagation de l'erreur (*back propagation* en anglais), l'erreur étant estimée par la moyenne des carrés d'erreurs (MSE *Mean Squared Error* en anglais) dont la relation est donnée par :

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - y_i^*)^2$$

L'apprentissage est réalisé par l'algorithme de Levenberg-Marquardt qui donne de meilleurs résultats que les algorithmes usuels à pas constant. Pour permettre et rendre notre réseau plus souple et moins rigide, nous avons utilisé une méthode basée sur la technique de la régularisation bayésienne par modération des poids qui consiste à pénaliser les valeurs

élevées des poids en modifiant la fonction de coût (*Dreyfus et al., 2002*). Cette technique force les paramètres (les poids) à ne pas prendre des valeurs trop élevées et par conséquent, à éviter le sur-ajustement.

La mise en pratique d'une modélisation par les réseaux de neurones en tenant compte de tous les critères et les paramètres cités supra a été traduite en un algorithme puis en un programme informatique écrit et exécutés sous langage MATLAB v9.5 (2018).

L'algorithme de la modélisation est représenté dans la figure suivante :

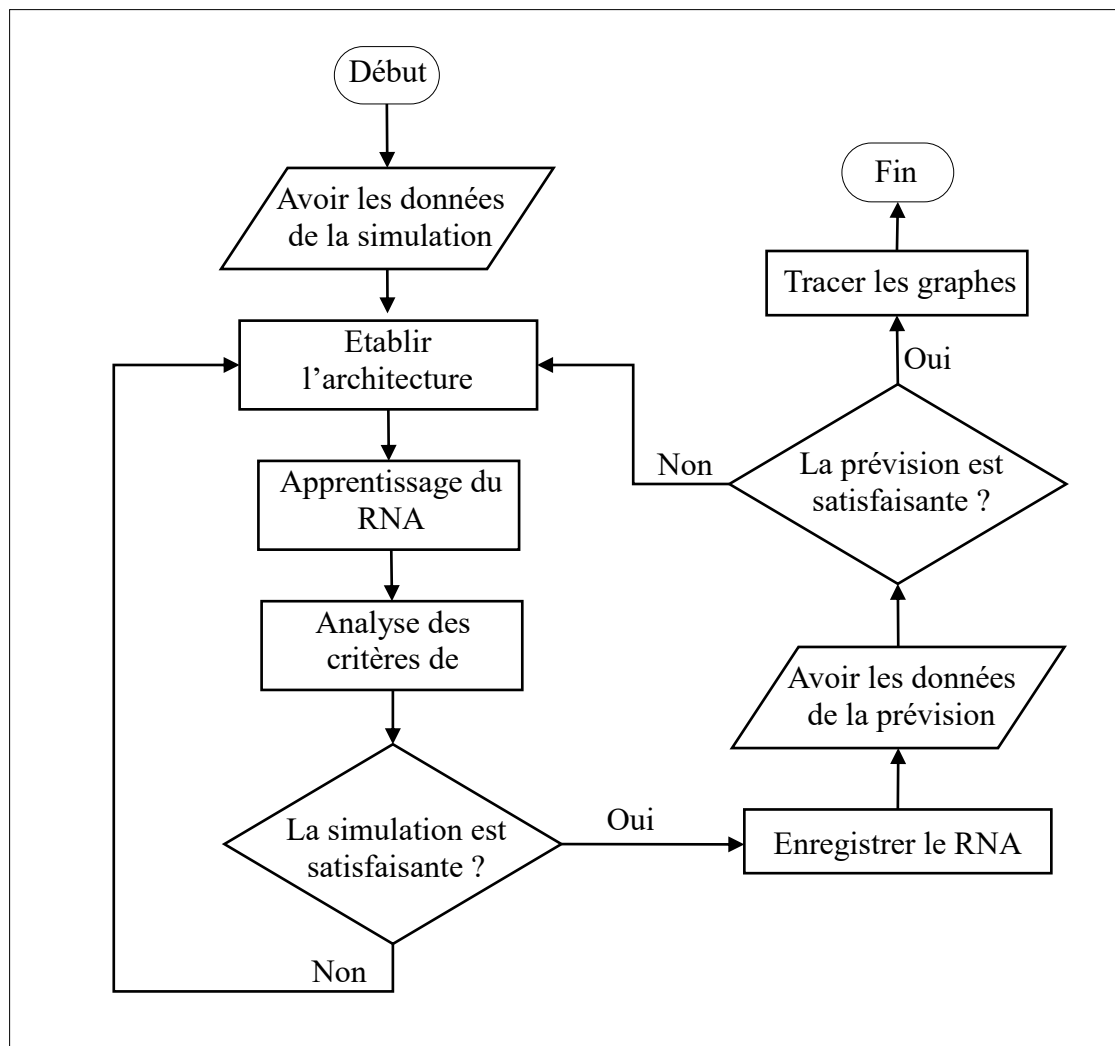


Figure III.15: Algorithme de la modélisation par les réseaux de neurones artificiels

III.8. Critères de validation des modèles obtenus

Pour évaluer la performance et la précision de nos modèles, nous avons utilisé les trois critères de validation suivants:

- L'erreur absolue sur la prédiction, donnée par l'équation (1), est définie par la différence entre la valeur réelle et la valeur modélisée, cette erreur est répartie sur trois classes :
 - Classe 1 : $-2,5\% < \text{erreur} < + 2,5\%$
 - Classe 2 : $-5\% < \text{erreur} < + 5\%$
 - Classe 3 : $\text{erreur} > + 7,5\%$
- Le coefficient de corrélation « r^2 » ; équation (2)
- La déviation quadratique moyenne ; équation (3)

$$e = y_i - y_i^* \quad (1)$$

$$r^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2 - \sum_{i=1}^n (y_i - y_i^*)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_i^*)^2} \quad (3)$$

Où e : Erreur absolue
 n : Nombre d'observations
 y_i : Observation réelle
 y_i^* : Observation modélisée

III.9. La prévision des TSP et ET associés

La prévision de la pollution atmosphérique est définie comme l'estimation des futurs polluants atmosphérique pour une période donnée, à partir des observations passées et actuelles (*Wmo, 1994*). Son objectif général est de fournir les meilleures estimations de ce qui peut arriver en un point donné à une date future précise (*Hipel et al., 1994*), contrairement à la prédiction qui vise l'estimation des conditions futures sans référence à un temps spécifique (*Lettenmaier et al., 1993*).

On peut noter que les éléments caractéristiques de la prévision sont :

- La variable à prévoir et les variables explicatives (c'est-à-dire l'ensemble des données disponibles) ;
- L'horizon de prévision ;
- Les méthodes de calcul ou d'estimation ;
- l'objectif de la prévision (alerte, planification, estimation, contrôle,...) ;
- Le type de résultats désirés (valeurs numériques, graphiques, distribution, ...).

Du point de vue mathématique, deux types d'approches ont été utilisés dans la prévision de la pollution atmosphérique :

- **L'approche déterministe (conceptuelle)** basée sur la simulation physique du système. Le modèle déterministe suppose qu'un calcul exact des paramètres est possible. Dans de nombreux cas, cette hypothèse s'avère peu réaliste, car les phénomènes naturels sont très aléatoires. Les modèles déterministes sont finalement tous limités par le grand nombre de paramètres à mesurer et par les limites des connaissances actuelles des systèmes naturels complexes.
- **L'approche empirique (stochastique)** permet de passer outre les limites des connaissances physiques du système. Elle prend seulement en compte l'aspect aléatoire des phénomènes. Contrairement aux modèles déterministes, les modèles empiriques fonctionnent comme des « boîtes noires », c'est à dire sans aucune considération de la structure interne du système. (*P. Coulibaly 1999*)

III.9.1. Prévision par le modèle ARIMA

Une fois les modèles obtenus sont approuvés et validés nous les avons utilisé pour prévoir les TSP et les ET associés, nous avons utilisé un algorithme simple, on applique le modèle déjà élaboré sur l'élément en question.

Pour ce faire, les quatre éléments caractéristiques de cette prévision sont :

- *La variable à prévoir* : TSP ;
- *Les variables explicatives* : Historiques des TSP noté TSP_{J-i} ;
- *L'horizon de prévision* : Quatre jours ;
- *Les méthodes de calcul* : ARIMA ;
- *L'objectif de la prévision* : Quantification quotidienne des TSP et ET associés ;
- *Le type de résultat désiré* : Valeurs numériques.

III.9.2. Prévision des TSP par les réseaux de neurones

Après avoir élaboré et validé la modélisation des TSP, nous avons procédé à la prévision de ces derniers, l'intérêt pratique de cette prévision c'est d'estimer la pollution quotidienne par les TSP au niveau de notre site d'étude (Drarîa) en effectuant que deux prélèvements par semaine.

Pour ce faire, les quatre éléments caractéristiques de cette prévision sont :

- *La variable à prévoir* : TSP ;
- *Les variables explicatives* : Historiques des TSP noté TSP_{J-i} ;
- *L'horizon de prévision* : Quatre jours ;
- *Les méthodes de calcul* : Réseau de neurones ;
- *L'objectif de la prévision* : Quantification quotidienne des TSP ;
- *Le type de résultat désiré* : Valeurs numériques.

Dans le but d'avoir la meilleure qualité de prévision, nous avons testé deux stratégies différentes :

La stratégie de bloc : un seul réseau qui a pour entrées les TSP aux jours $J-i$ ($i=1,2,3,4$) et les TSP $_{J+k}$ ($k=1,2,3,4$) à la sortie. Nous avons remarqué qu'il y a une divergence sur la précision au-delà $J+4$, la raison pour laquelle nous avons adopté un horizon de prévision de quatre jours.

La stratégie itérative (Figure III.17) a été la plus performante en termes de qualité de précision sur la prévision, le principe de la méthode est le suivant :

➤ *La première itération* : les entrées sont l'historique de la concentration des TSP durant les quatre derniers jours, noté TSP $_{J-i}$ ($i=1,2,3,4$); la sortie c'est la concentration du TSP du jour suivant, noté TSP $_J$.

➤ Le résultat de la 1^{ère} itération (TSP $_J$) combiné avec les quatre dernières valeurs des TSP forment les entrées de la 2^{ème} itération ; la sortie représente la concentration des TSP au jour J_2

➤ On continue de la même manière jusqu'au J_4

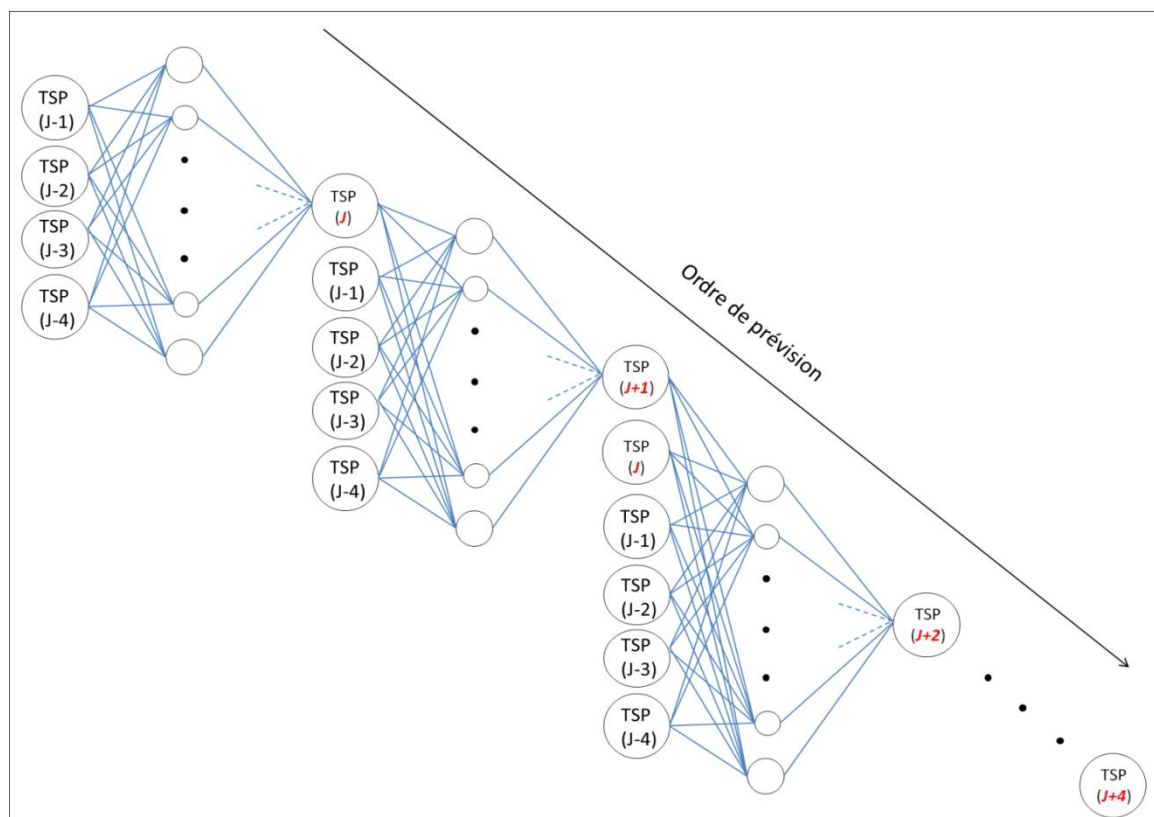


Figure III.16: Schéma de la méthode itérative par les réseaux de neurones pour la prévision

La mise en pratique de cette stratégie a été faite par le MatLab 2018 suivant l'algorithme suivant :

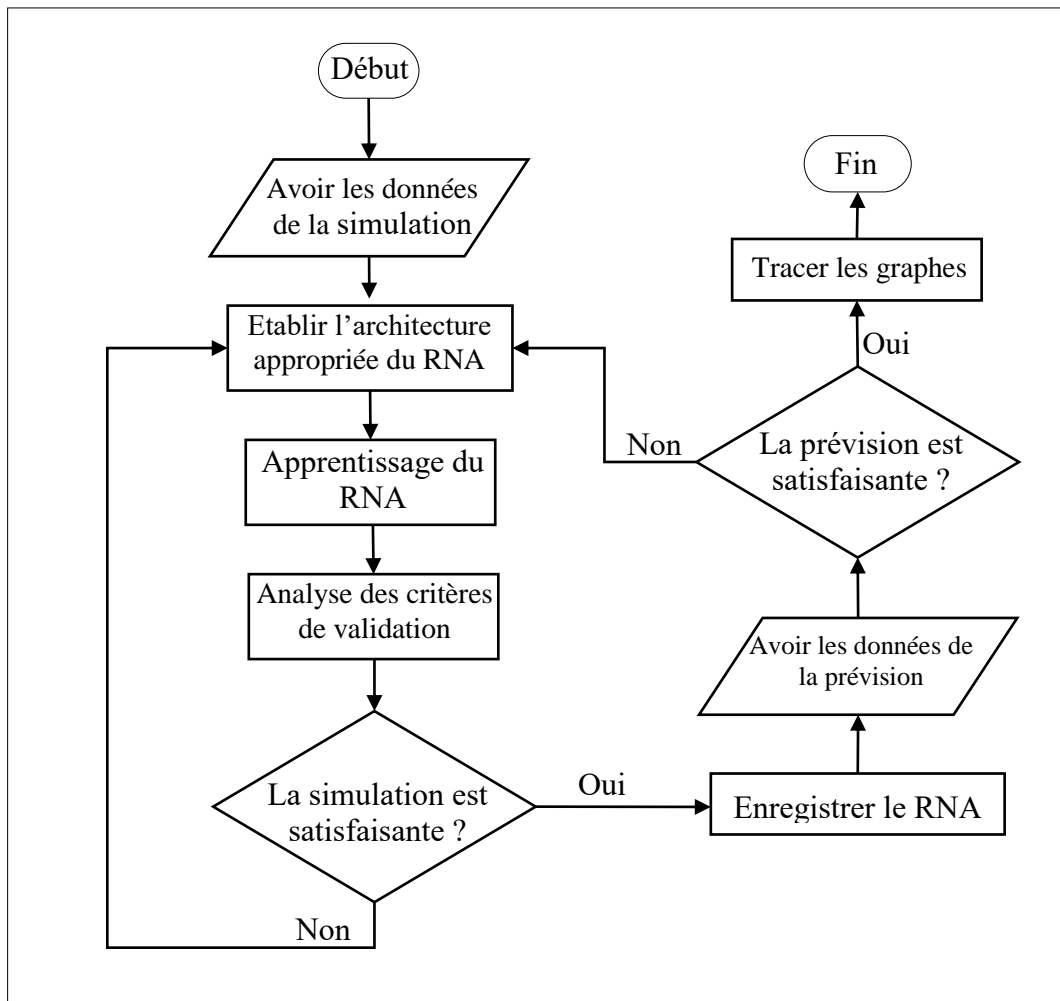


Figure III.17: Algorithme de la prévision des TSP et ET associés par les réseaux de neurones artificiels

III.9.3. Prévision des ET par les réseaux de neurones

En pratique, pour avoir les concentrations des ET d'une seule capsule il faut compter entre 30 et 40 jours (échantillonnage, conditionnement, irradiation, décroissance radioactive, comptage et interprétation), cette problématique nous a poussé à poser la question :

« Comment quantifier les ET sans les analyser tout en exploitant nos modèles ? ».

Puisque nous cherchons à quantifier les ET associés aux TSP, en première étape, nous avons entraîné un réseau de neurone qui détermine la concentration d'un ET au jour J à partir de l'historique des TSP aux jours $J-i_{(i=1,2,3,4)}$ puis nous avons combiné ce réseau avec celui de la prévision des TSP en utilisant toujours la méthode itérative (Figure III.18).

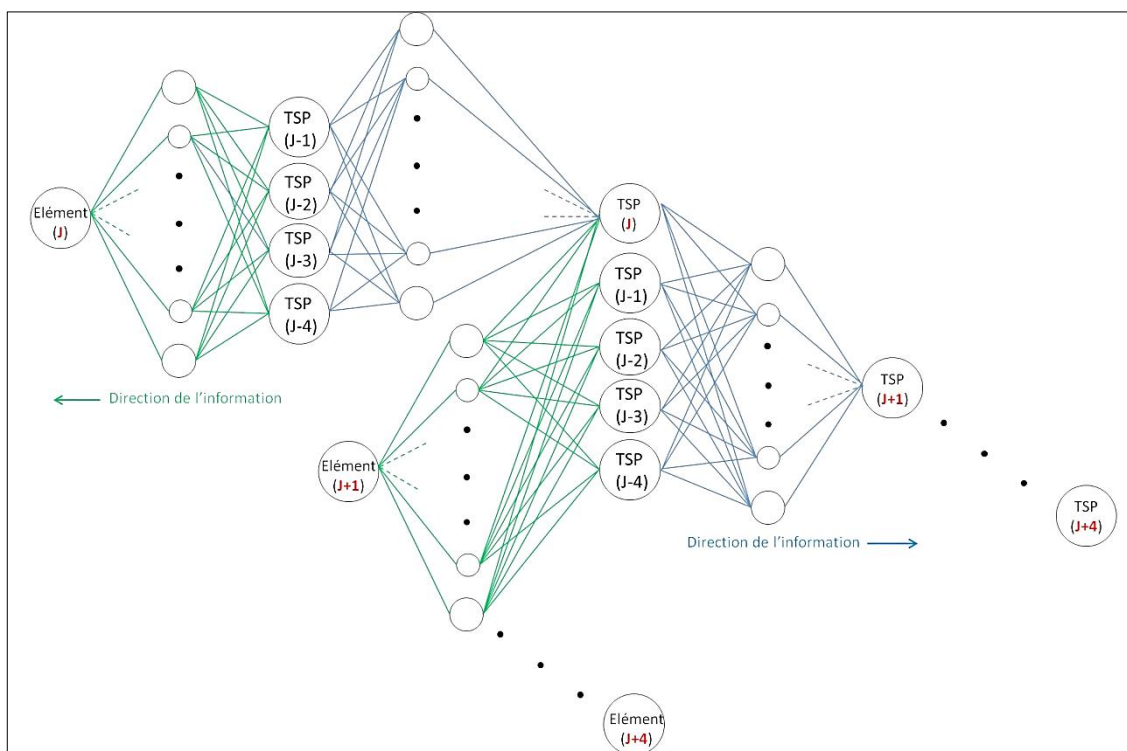


Figure III.18: Schéma du réseau de neurone combiné pour la prédiction des TSP et ET associés

III.10. Critères de validation de la prédiction obtenue

Pour tester la fiabilité des méthodes utilisées ainsi que les stratégies adoptées et les réseaux de neurones construits, nous avons utilisé comme base de comparaison les concentrations réelles des TSP et des ET associés enregistrées au niveau de notre zone d'étude, et qui se sont déroulées dans les mêmes conditions d'échantillonnages et d'analyses que les échantillons utilisés pour la modélisation.

III.11. La cartographie de la pollution

Après avoir traité les données et une fois les concentrations de la pollution par les TSP ont été calculées et la modélisation des TSP a été validée et approuvée nous avons procédé à l'application du modèle gaussien de la dispersion de la pollution, comme nous avons expliqué dans le chapitre précédent, l'application de ce modèle nécessite la connaissance de la vitesse et la direction du vent au moment du prélèvement, ces données ont été gracieusement offertes par l'Office Nationale de Météorologie (l'ONM).

Nous avons construit une matrice de trois dimensions qui englobe : la dispersion verticale, la dispersion horizontale, concentration en ces points. Le principe de notre cartographie est le comme suit : le modèle gaussien offre la concentration des TSP en un point x, y bien déterminé, nous avons mémorisé la concentration en chaque point de notre cartographie qui s'étend sur un rayon de 500 mètres par rapport à la station de prélèvement, après nous avons calculé la moyenne de la concentration en chaque point x, y de cette carte ou matrice.

Chapitre IV

Résultats et discussions

« Le monde que nous avons créé est le résultat de notre niveau de réflexion, mais les problèmes qu'il engendre ne sauraient être résolus à ce même niveau »

(Albert Einstein 1879- 1955)

« Que la stratégie soit belle est un fait, mais n'oubliez pas de regarder le résultat »

(Winston Churchill 1874- 1965)



Chapitre IV : Résultats et discussions

IV.1. Etude de la pollution par les gaz

IV.1.1. La station de prélèvement

La figure IV.1 illustre la partie électronique de notre station de prélèvement des gaz polluants, le circuit imprimé a été fabriqué par PCBWAY en Chine et les différentes pièces électronique et modules ont été montés et soudés par nous-même.



Figure IV.1: Circuit imprimé de la station montée sur la carte Arduino UNO

Afin d'assurer le bon fonctionnement de la station et afin d'avoir les meilleurs prélèvements en terme de précision nous avons simulé le montage de la station par le logiciel SolidWorks 2018, nous avons appliqué une contrainte de vent de vitesse et de directions différentes pour analyser le sens et la direction de l'écoulement de l'air à travers les capteurs (Figure IV.2), le but de cette simulation a pour objectif d'optimiser l'emplacement des capteurs afin de permettre à chaque capteur de recevoir un air frais et pour que l'écoulement de ce dernier ne soit pas gêné par les autres capteurs

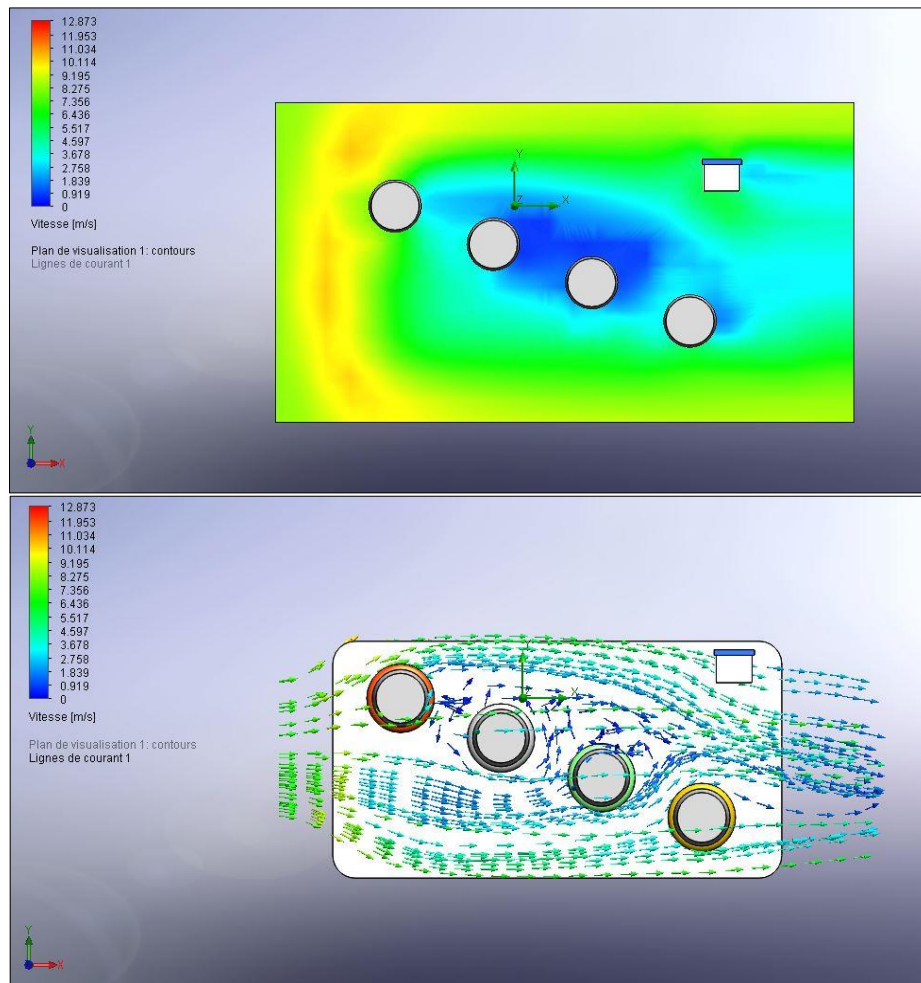


Figure IV.2: Simulation de profil de vitesse et la direction de l'air autour des capteurs de prélèvement

On remarque bien qu'un écoulement libre d'air est assuré autour des capteurs, l'air passe librement entre ces derniers en apportant à chaque passage de l'air frais (pas de redondance).



Figure IV.3: Emplacement des capteurs dans le support

Note : Sur toutes les présentations graphiques l'axe des abscisses (temps) représente la répartition des prélèvements dans le temps durant la période d'échantillonnage qui s'est déroulée du 03 juin 2018 au 28 novembre 2018.

IV.1.2. Analyse de la pollution atmosphérique par les gaz

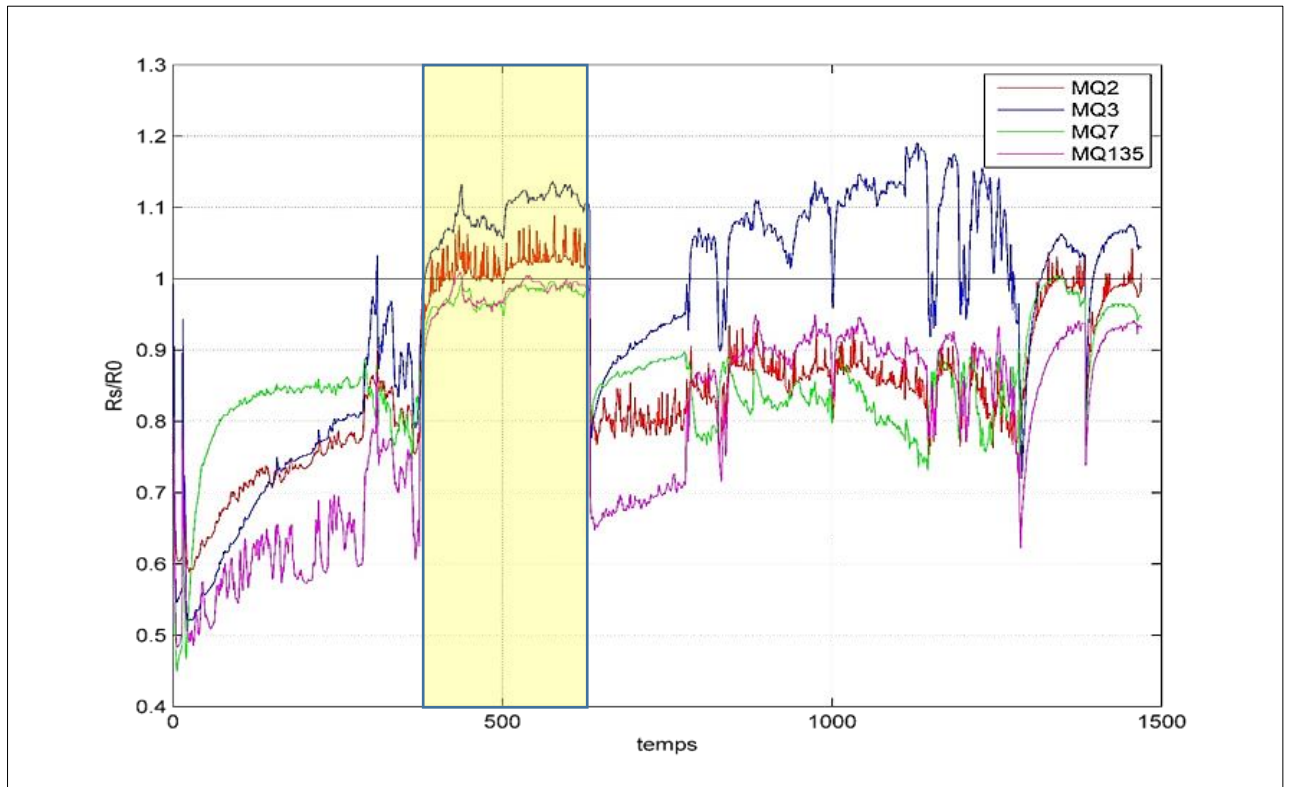


Figure IV.4: Rapport R_s/R_0 enregistrés par les capteurs en fonction du temps du prélèvement

La figure IV.4 représente les signaux enregistrés par les capteurs durant la période d'échantillonnage allant du 03 Juin 2018 au 28 Novembre 2018. Le programme informatique élaboré dont l'algorithme est présenté dans le chapitre précédent nous a permis de présenter graphiquement directement le rapport R_s/R_0 , ce rapport est utilisé pour convertir le signal enregistré par chaque capteur en une concentration exprimée en ppm.

On remarque qu'en date du 20 Juin au 10 Juillet 2018 y'a une déflexion (représentée par le cadre jaune) anormale dans la gamme de mesure, cette déflexion sera interprétée plus tard (voir page 90).

Nous avons pu capter par le biais de cette station 11 gaz polluant, chaque capteur était programmé pour analyser un gaz bien déterminé en se basant sur son affinité envers le gaz en question :

- MQ₂ : GPL, Propane, CH₄
- MQ₃ : Alcool, Benzène
- MQ₇ : H₂, CO
- MQ₁₃₅ : Toluène, Alcool, NH₄, Acétone

La méthode d'analyse par la série des capteurs type MQ est expliquée dans le chapitre précédent; puisque nous avons plus de 1400 analyses à calculer nous avons opté pour la méthode numérique au lieu de celle basée sur la projection graphique, pour ce faire, nous avons convertie les graphes de lecture de chaque capteur en une équation de lissage afin d'injecter seulement le rapport R_0/R_s déjà calculé pour avoir la concentration correspondante en ppm.

Les équations de lissages de chaque capteur et pour chaque gaz sont les suivantes :

Capteur MQ2

$$GPL = 625,68 \cdot \left(\frac{R_0}{R_s}\right)^{-2,113}$$

$$Propane = 685,74 \cdot \left(\frac{R_0}{R_s}\right)^{-2,104}$$

$$CH_4 = 4380,8 \cdot \left(\frac{R_0}{R_s}\right)^{-2,628}$$

Capteur MQ3

$$Alcool = 0,4275 \cdot \left(\frac{R_0}{R_s}\right)^{-1,501}$$

$$Benzène = 4,9767 \cdot \left(\frac{R_0}{R_s}\right)^{-2,691}$$

Capteur MQ7

$$H_2 = 72,978 \cdot \left(\frac{R_0}{R_s}\right)^{-1,348}$$

$$CO = 103,88 \cdot \left(\frac{R_0}{R_s}\right)^{-1,5}$$

Capteur MQ135

$$Toluène = 47,697 \cdot \left(\frac{R_0}{R_s}\right)^{-3,325}$$

$$NH_4 = 103,3 \cdot \left(\frac{R_0}{R_s}\right)^{-2,556}$$

$$CO_2 = 115,26 \cdot \left(\frac{R_0}{R_s}\right)^{-2,887}$$

$$Acétone = 38,043 \cdot \left(\frac{R_0}{R_s}\right)^{-3,282}$$

Tableau IV.1: Données statistique de la concentration des polluants gazeux détectés au niveau de notre zone d'étude en PPM

	Min	Moyenne	Max	Ecart type
GPL	522,826	908,376	1919,954	267,27558
Propane	573,451	993,822	2094,228	291,06939
Toluène	46,308	124,992	536,597	2447,3643
Alcool	71,376	164,899	300,000	97,27696
NH₄	100,979	185,689	300,000	86,89584
CO₂	112,340	258,786	942,687	74,51328
Acétone	36,949	98,154	414,799	171,12726
CO	103,173	131,976	345,296	75,24851
H₂	72,531	90,350	214,778	27,25940
CH₄	5000	7238,551	17667,422	16,30739
Benzène (mg/l)	3,117	6,700	28,874	4,95309
Alcool (mg/l)	0,329	0,482	1,140	0,17040

Par rapport aux normes internationales régissant le benzène, on note que l'objectif de qualité de l'air de $2 \mu\text{g}/\text{m}^3$ en vigueur en France est largement dépassé à Draria. Le seuil limite de $10 \mu\text{g}/\text{m}^3$ en moyenne annuelle est toutefois respecté.

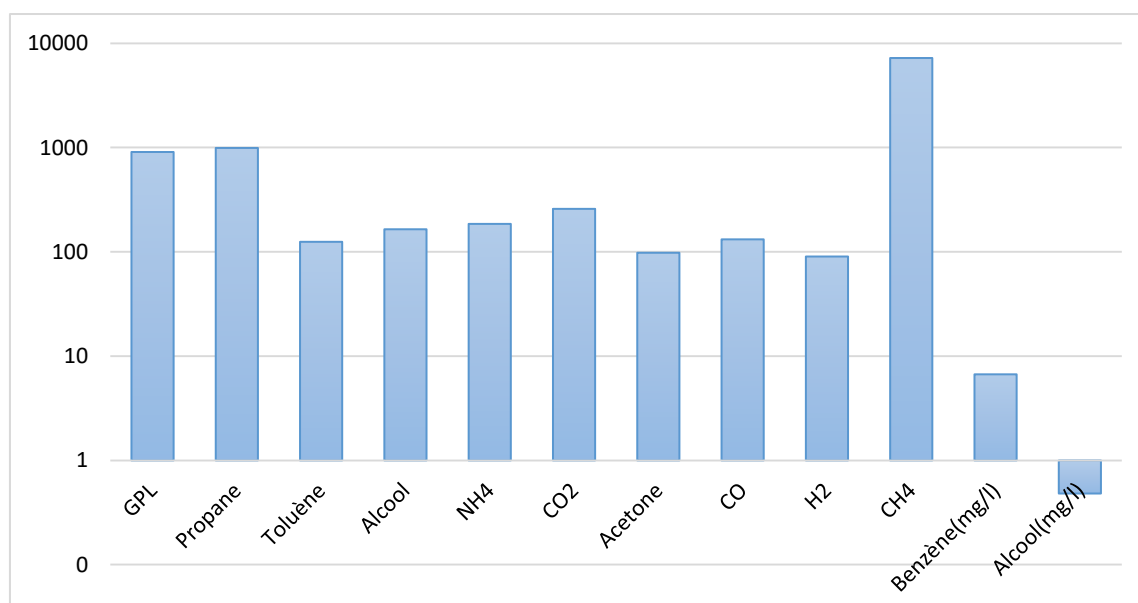


Figure IV.5: Concentration moyenne en ppm des polluants gazeux analysés au niveau de notre site d'étude

On remarque que le GPL est l'élément le plus polluant et le plus présent dans l'air de Draira, cette forte présence est due à son utilisation dans la majorité des voitures essence qui sont en circulation. Les autres gaz analysés sont des gaz qui résultent de la combustion des combustibles fossiles donc leur présence dans l'air de Draria est due principalement au trafic routier.

Tableau IV.2: Teneurs en mg/l du benzène et toluène dans certaines villes

Villes et type de site	Benzène	Toluène
Alger, 2002		
- urbain	9,6	15,2
- périurbain	4,2	23,6
Paris (France), 2001		
- station trafic	9,8	34,9
- station urbaine	4,0	15,0
Sao Paulo (brésil), 2002		
Urbain et station trafic	4,6	44,8
Santiago (chili), 2000		
urbain	19,5	83,5
Karachi (Pakistan), 2002		
urbain	18,2	27,2
Izmir (Turquie), 2001		
Urbain	56,9	106,5

Les teneurs en benzène mesurées à Draria sont par ailleurs justes en dessus à la limite du seuil de $5 \mu\text{g}/\text{m}^3$ prescrit par la directive européenne du 16 novembre 2000.

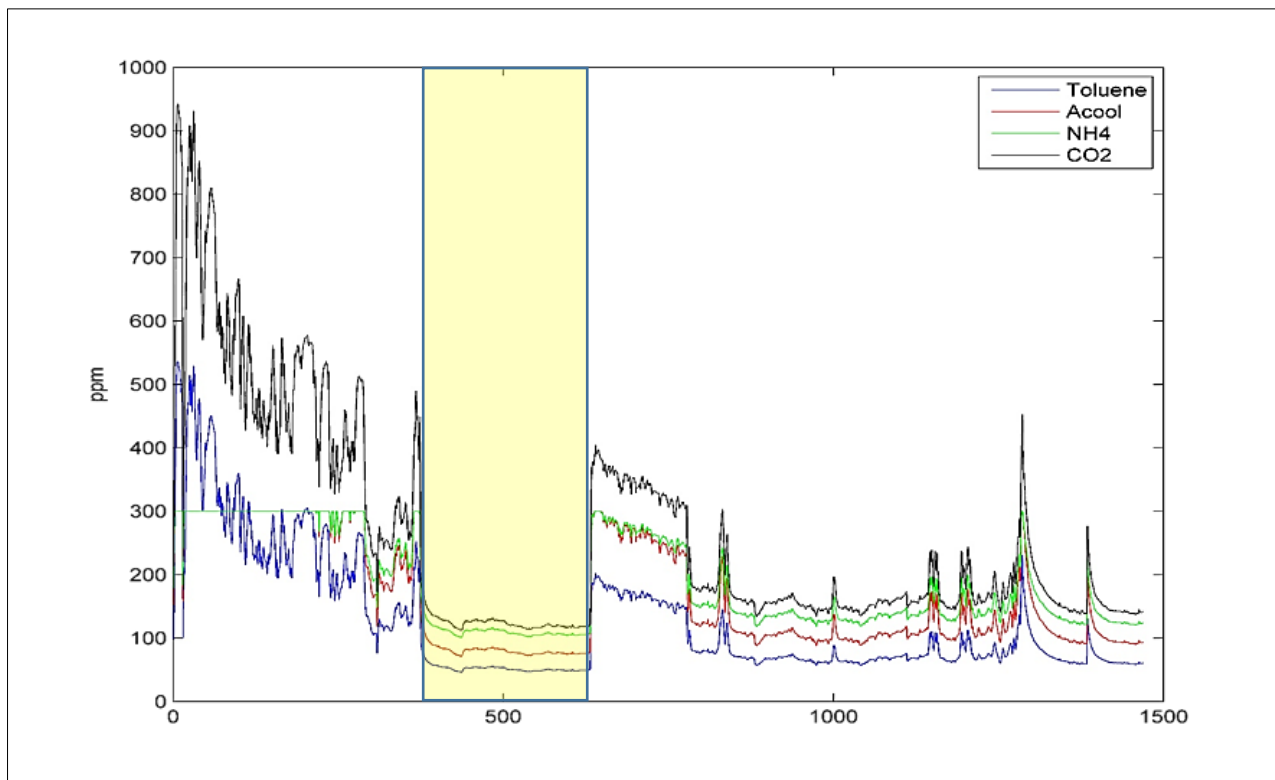


Figure IV.6: Variation de la concentration du toluène, Alcool, NH₄, CO₂ durant la période d'échantillonnage

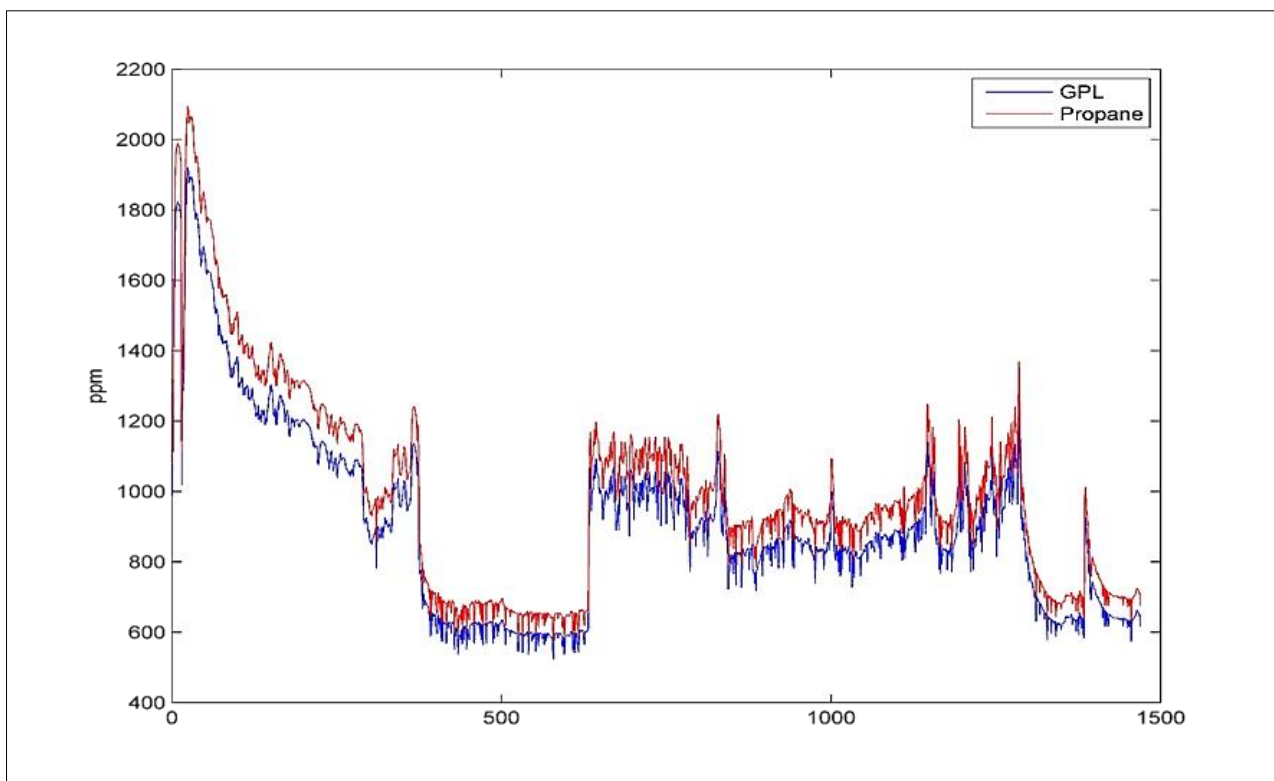


Figure IV.7: Variation de la concentration du propane et GPL durant la période d'échantillonnage

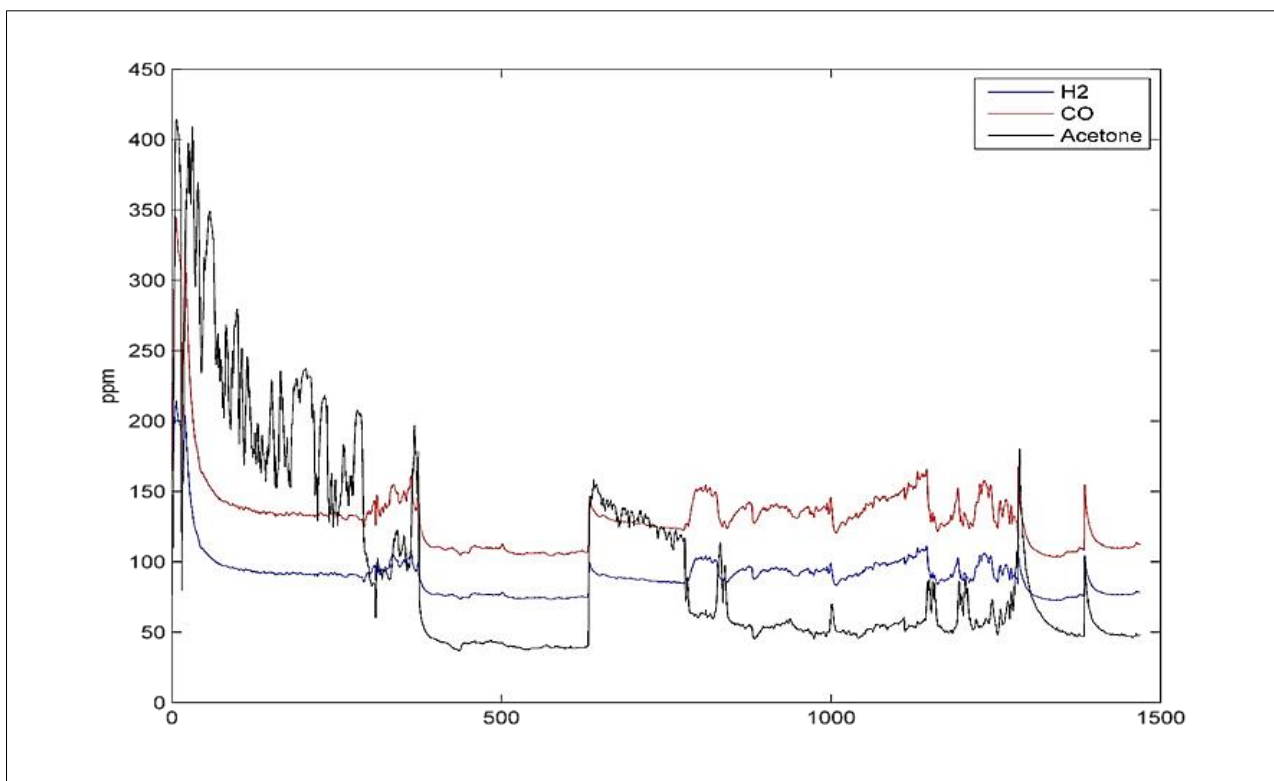


Figure IV.8: Variation de la concentration de l'acétone, CO et H₂ durant la période d'échantillonnage

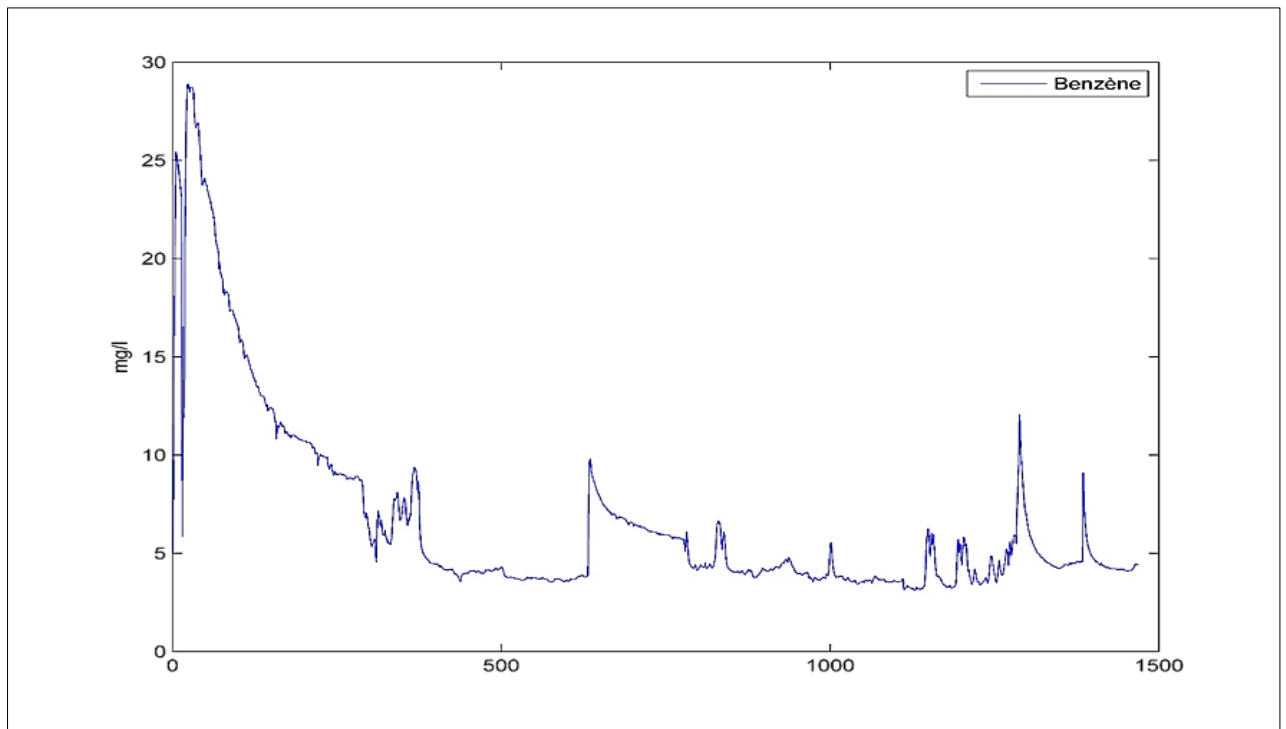


Figure IV.9: Variation de la concentration du benzène durant la période d'échantillonnage

Tableau IV.3: Matrice de corrélation de Pearson des gaz polluants analysés au niveau de notre site d'étude

	GPL	Propane	CH ₄	Toluène	Alcool	NH	CO ₂	Acétone	CO	H ₂	Benzène
GPL	1	1	0,988	0,919	0,830	0,832	0,922	0,920	0,715	0,723	0,889
Propane		1	0,988	0,919	0,830	0,832	0,922	0,920	0,715	0,723	0,889
CH ₄			1	0,943	0,811	0,810	0,941	0,943	0,708	0,713	0,934
Toluène				1	0,877	0,872	0,999	1	0,600	0,602	0,947
Alcool					1	1,000	0,896	0,879	0,410	0,416	0,760
NH						1	0,892	0,874	0,415	0,421	0,754
CO ₂							1	0,999	0,590	0,592	0,940
Acétone								1	0,599	0,601	0,946
CO									1	1	0,599
H ₂										1	0,600
Benzène											1

Il y' a une forte corrélation entre les polluants (la majorité des cas $r > 0.8$) car ils sont tous générés par la combustion des carburants fossiles, à noter qu'il y'a relativement une faible corrélation entre le benzène et l'alcool, NH et CO, rajoutant à ça une faible corrélation aussi entre le couple (alcool-CO), (alcool-H₂), (NH-CO), (NH-H₂).

Pour une meilleure compréhension de ces couples de corrélation une étude détaillée de la pollution de l'air par les gaz d'échappement s'impose.

Afin de mieux interpréter les résultats des mesures, plusieurs paramètres météorologiques ont été pris en considération pendant les campagnes de prélèvement. Ces paramètres météorologiques peuvent avoir une grande influence dans la distribution des gaz polluants. La moyenne quotidienne de l'humidité relative, la vitesse du vent, les précipitations cumulées et les températures recueillies sont présentés dans les figures IV.13-15

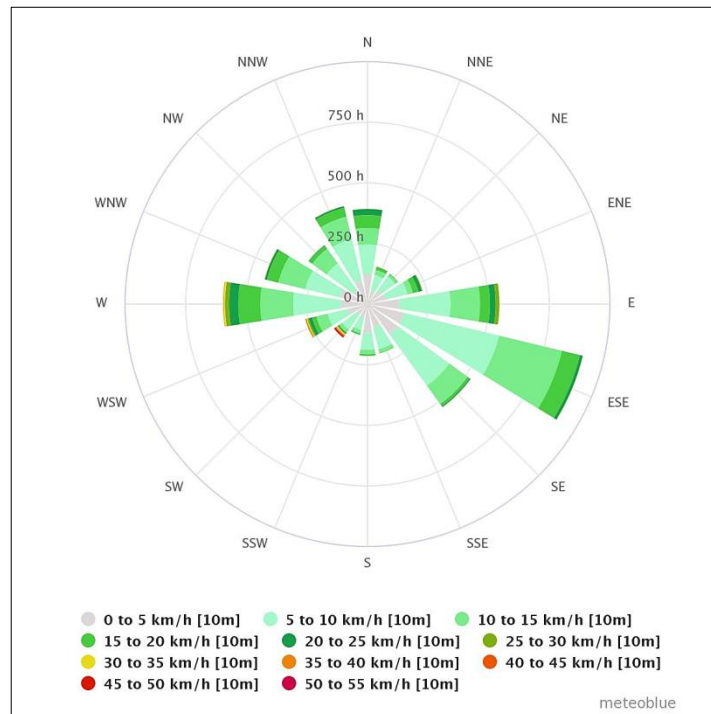


Figure IV.10: Rose des vents enregistrée au niveau de la station de Dar El Beida durant la période d'étude

A partir de ces diagrammes, on peut tirer les informations suivantes.

- Les vents du Sud-Est sont les vents dominants
- Dans la direction Sud-Est, les fréquences de vitesse du vent s'élèvent à environ 10% pour la classe de vitesse 1 à 2 m/s, 19% pour la classe 2 à 3 m/s et environ 20% pour les vitesses allant de 3 à 4 m/s.
- Les vents du Nord sont, avec une fréquence de l'ordre de 5%, moins fréquents que les vents d'Est et sont répartis de manière égale sur les classes de 1 à 2 et 2 à 3 m/s.
- Les vents du secteur Ouest apparaissent à des fréquences de plus de 10% avec des vitesses modérées comprises en 1 et 3 m/s.
- Les vents du Sud sont très rares et constituent moins de 1% de l'ensemble des directions de vent observées.

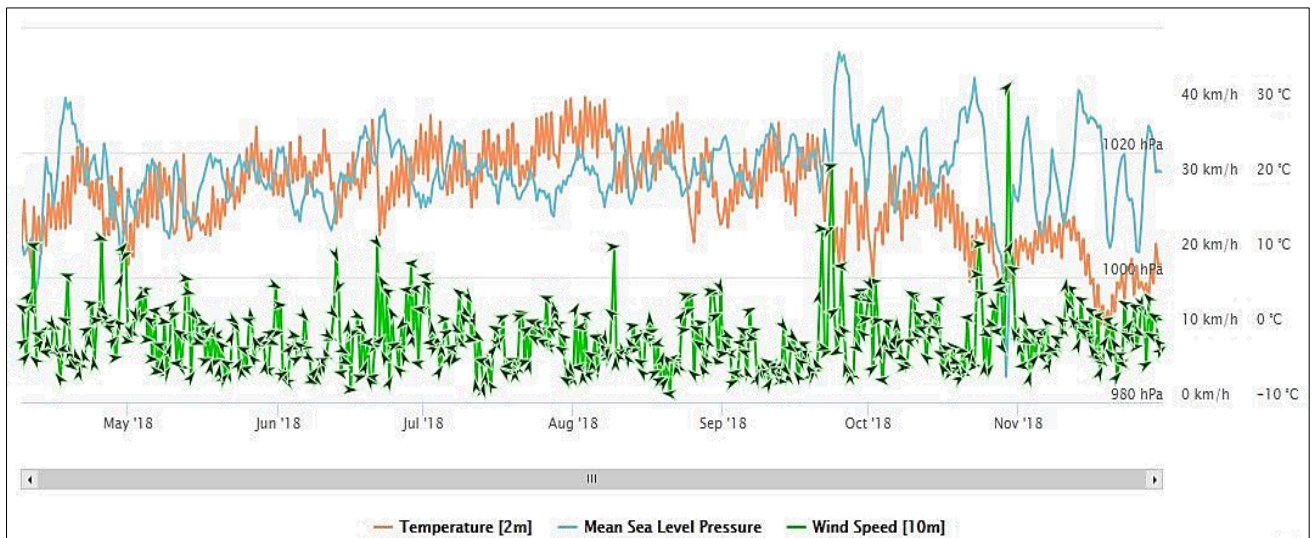


Figure IV.11: Variation de la température, pression atmosphérique et la vitesse du vent durant la période d'étude

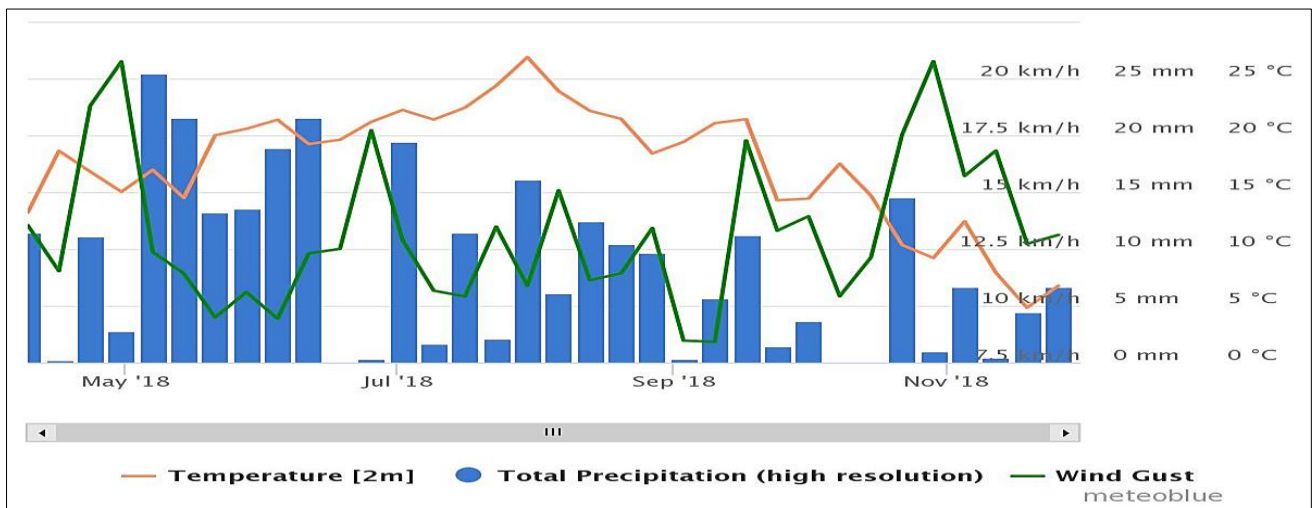


Figure IV.12: Variation de la température, la précipitation et les rafales du vent durant la période d'étude

La perturbation des résultats des analyses représentées par le cadre jaune qui couvre la période allant du 20 Juin au 10 Juillet peut être expliquée par les rafales de vents violant qui ont atteint une vitesse de presque 20km/h et le changement de la direction du vent (Est à l'Ouest), rajoutant à ça l'enregistrement de précipitations de 20 mm ainsi qu'une diminution brutale de la pression atmosphérique et de la densité de l'air.

D'un point de vue global, les fortes concentrations enregistrées au début de la période de la collecte sont dues principalement au trafic routier, on remarque qu'il y'a une décroissance progressive de la concentration ce qui peut être expliqué par le commencement de la période des congés et la fin de l'année universitaire, ces deux paramètres n'ont fait que diminuer le trafic routier contrairement au deux derniers mois de la période de la collecte des échantillons où la pollution par les gaz a connu une augmentation progressive due à la fin de la période estivale, la fin des congés et les rentrées universitaire et scolaire.

IV.1.3. Répartition journalière

Nous présentons sur la figure IV.13 l'évolution des teneurs horaires des gaz polluants détectés au niveau du site d'étude durant la période allant du 03 Juin 2018 au 28 Novembre 2018.

Tableau IV.4: Répartition journalière de la concentration en ppm des polluants gazeux analysés au niveau du site d'étude

	06h-10h	10h-14h	14h-18h	18h-22h	22h-02h	02h-06h
GPL	1528,168	1309,095	1324,553	950,563	460,765	482,506
Propane	1818,625	1490,688	1520,389	983,167	456,217	731,986
Toluène	463,561	495,066	467,777	264,020	153,556	142,168
Alcool	327,586	176,288	260,290	263,534	160,000	80,000
NH₄	252,794	299,776	270,050	251,883	148,529	108,423
CO₂	549,655	561,552	538,213	501,129	204,149	129,236
Acétone	228,743	240,172	196,877	224,252	53,819	50,774
CO	242,223	259,596	255,204	220,066	135,837	111,595
H₂	170,187	192,297	180,352	128,606	35,895	75,618
CH₄	15002,128	14893,608	14835,608	7317,345	6396,569	6384,551

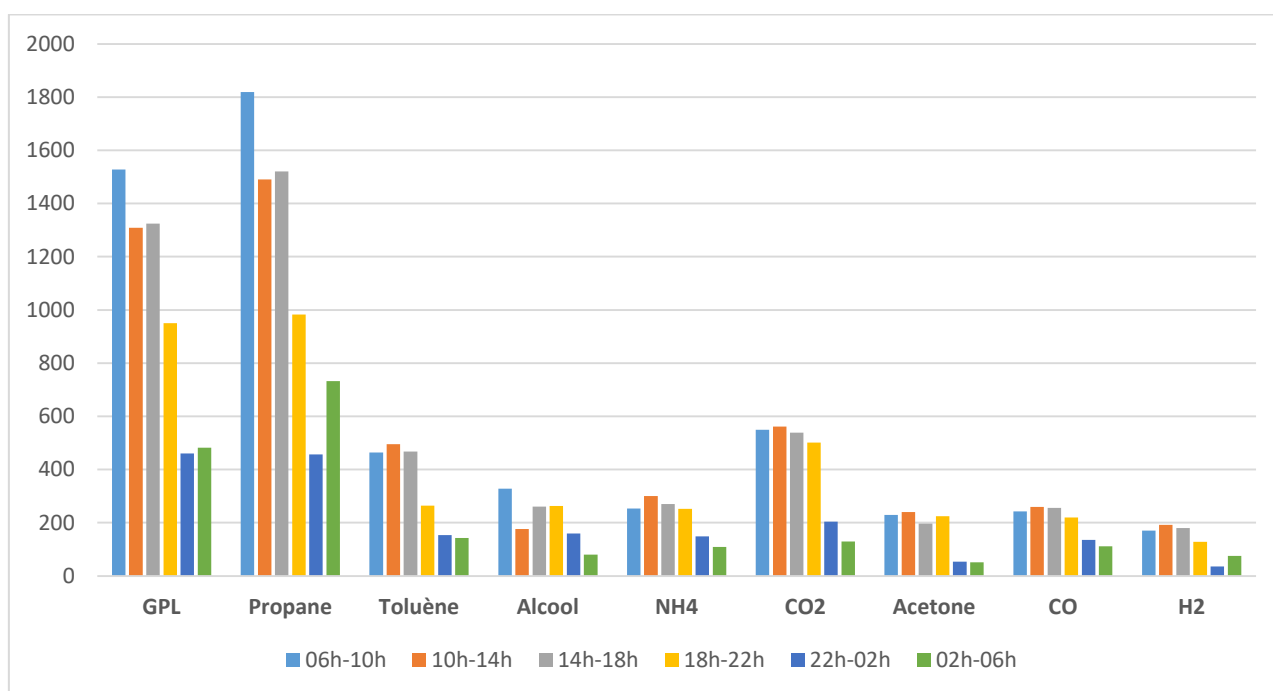


Figure IV.13: Répartition journalière de la concentration en ppm des polluants gazeux analysés au niveau du site d'étude

L'examen du profil journalier moyen (Figure IV.13) montre qu'à Draria l'ensemble des gaz détectés accusent deux pics importants, l'un en début de matinée vers 8 heures et l'autre en fin de journée vers 18-19heures. On observe parfois aussi un léger pic en milieu de journée. Ces pics peuvent être expliqués par l'influence des émissions du trafic routier avoisinant.

L'intense trafic routier qu'on observe aux heures de pointe engendre au site de mesure des teneurs très élevées qui avoisinent le double des teneurs journalières moyennes.

Durant l'après-midi, l'intensification du trafic routier débute des 16 heures et persiste, surtout en période estivale, au-delà de 16 heures (retour des estivants des plages de l'Ouest d'Alger) ce qui explique le large pic observé dans l'évolution journalière.

On note par ailleurs que l'ensemble des polluants évoluent au cours de la journée dans le même sens ; les maxima et minima apparaissent au même horaire. Ce résultat suggère que ces composés ont une même source polluante, en l'occurrence le trafic routier, et subissent les mêmes effets de dispersion.

En termes d'ampleur de pollution, les niveaux relevés en ce site périurbain, qui subit l'influence des émissions du trafic routier, sont très élevés mais nettement inférieurs à ceux observés en plein centre-ville d'Alger (*Kerbachi et al., 1999*).

Une comparaison avec les données recueillies en bibliographie sur des villes de pays développés et pays émergents montre aussi que les taux mesurés à Draria sont relativement alarmants. La situation serait toutefois différente si l'étude était menée en site purement urbain où les démarrages à froid et la circulation au ralenti sont plus fréquents, de même que dans les rues de type canyon où la ventilation naturelle est plus réduite. De telles conditions engendrent en effet une pollution plus intense en polluants gazeux.

Rappelons dans ce contexte que pour l'OMS, le benzène et cancérigène et qu'il n'existe pas un seuil limite au-delà duquel il n'aurait pas de risques pour la santé.

IV.1.4. Répartition hebdomadaire

Afin de vérifier si la réduction de l'intensité du trafic routier durant les jours de fin de semaine a un effet sur les niveaux de pollution de l'air par les gaz, nous avons exploité les données recueillies en fonction du jour de la semaine. Le tableau IV.5 résume les teneurs journalières moyennes de cette pollution en fonction du jour de la semaine.

Tableau IV.5: Répartition hebdomadaire de la concentration en ppm des polluants gazeux analysés au niveau du site d'étude

	Dimanche	Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi
GPL	1391,417	1072,911	1151,790	1231,490	1596,986	774,623	928,026
Propane	1531,603	1829,599	871,968	1056,045	1767,086	914,500	954,367
Toluène	534,453	189,923	261,508	359,387	284,527	204,583	263,708
Alcool	337,028	165,364	179,193	325,511	208,582	161,482	185,328
NH₄	294,190	193,454	252,110	354,754	263,819	172,705	238,528
CO₂	848,770	305,711	466,689	794,725	916,937	217,408	297,346
Acétone	457,581	142,885	110,734	446,757	226,241	124,042	92,963
CO	312,254	222,498	161,019	314,036	380,660	160,135	144,960
H₂	194,753	101,028	154,001	265,670	191,476	66,356	218,023
CH₄	12020,961	9063,741	15029,030	10568,656	13016,502	6592,207	7829,842

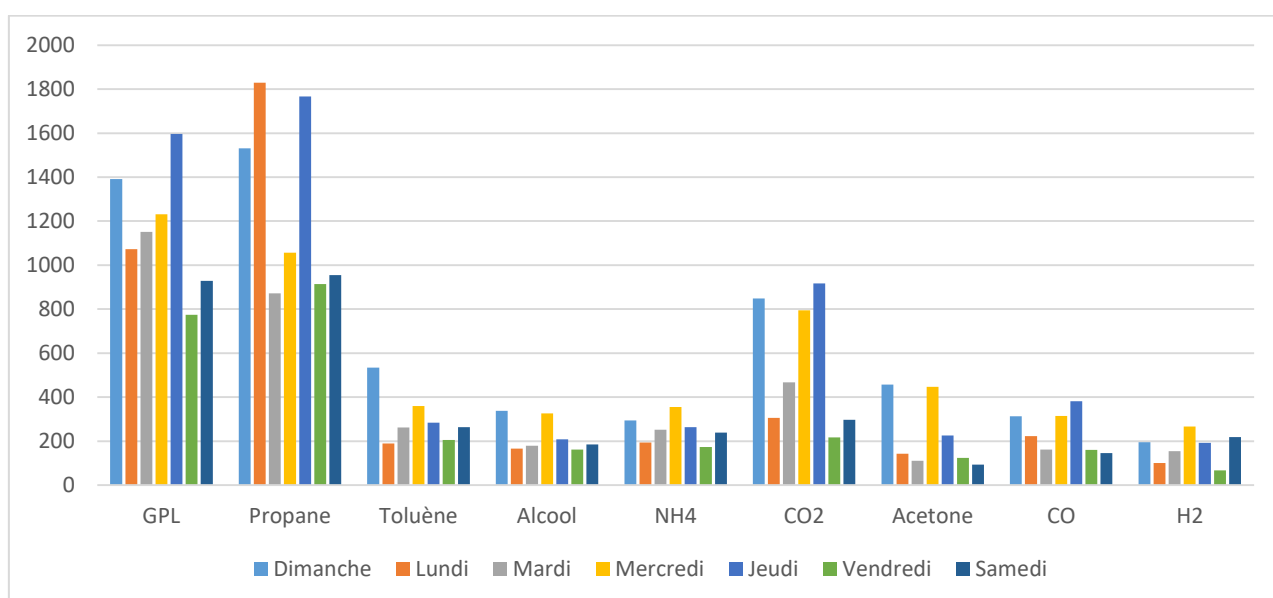


Figure IV.14: répartition hebdomadaire de la concentration en PPM des polluants gazeux analysés au niveau du site d'étude

Ces figures montrent que quel que soit le polluant considéré, le vendredi correspond au plus faible taux de pollution. En ce jour de repos, la pollution est réduite par rapport à la journée la plus polluée d'environ 15 à 20%. Comme plusieurs administrations et services (banque etc.) sont fermés où ne reçoivent pas de public le samedi, ce jour de début de semaine vient, en terme de pollution, juste après le jour férié du vendredi. Le jeudi compte paradoxalement parmi les journées de la semaine qui accusent un fort taux de pollution. En effet, ce n'est pas un vrai jour férié car certaines administrations publiques sont ouvertes, le secteur privé et informel active également ce jour et les secteurs de l'éducation nationale et de l'enseignement supérieur sont aussi en activité. Il en résulte une certaine effervescence de la vie socioéconomique qui se répercute alors sur les niveaux de pollution.

IV.1.5. Répartition mensuelle

Plusieurs études effectuées en Europe, en Asie et aux USA ont montré qu'il existe une différence plus ou moins importante entre les teneurs hivernales et les teneurs estivales des gaz polluants dans l'air. La vitesse du vent et surtout la précipitation sont à l'origine de cette variation.

Nous avons représenté les concentrations enregistrées au niveau de notre site d'étude dans le tableau IV.6.

Tableau IV.6: Répartition mensuelle de la concentration en ppm des polluants gazeux analysés au niveau du site d'étude

	Jun	Juillet	Août	Septembre	Octobre	Novembre
GPL	1359,866	821,079	758,596	884,109	870,370	755,436
Propane	1485,393	898,744	830,609	967,533	952,566	827,214
Toluène	301,509	112,714	100,559	88,140	70,626	76,851
Alcool	297,987	167,611	157,315	137,566	109,702	119,283
NH₄	298,350	187,900	177,029	163,869	139,381	147,653
CO₂	568,554	238,854	215,609	195,095	161,839	173,510
Acétone	234,681	88,729	79,249	69,691	56,037	60,884
CO	157,776	125,513	116,140	135,141	138,531	118,645
H₂	105,856	86,441	80,637	92,425	94,498	82,176

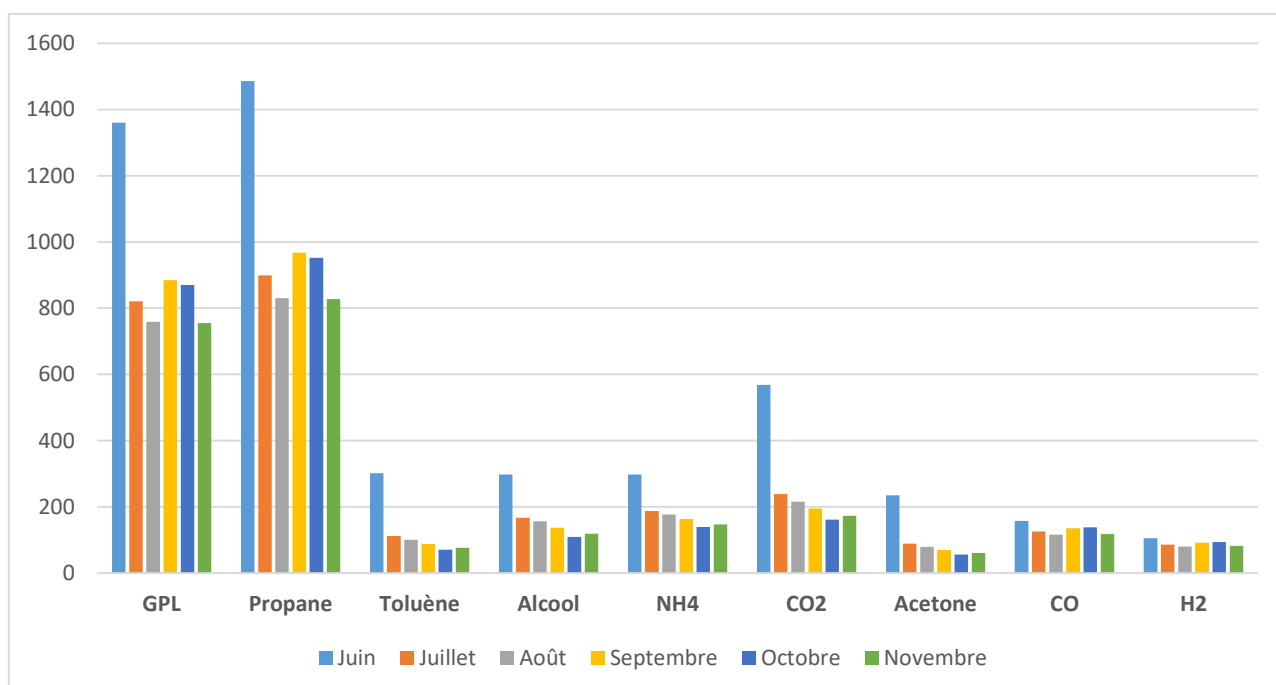


Figure IV.15: Répartition mensuelle de la concentration en ppm des polluants gazeux analysés au niveau du site d'étude

Bien qu'au cours de notre étude qui s'est étalée de mois de Juin jusqu'au mois de Novembre, la saison d'hiver n'a pas été couverte, nous avons jugé utile de vérifier s'il existait une différence entre les deux périodes suivantes :

- Période à température moyenne : ($T_{\text{moyenne}} = 19^{\circ}\text{C}$) correspondant à la période du mois d'Octobre et Novembre.
- Période à fort ensoleillement et forte température : ($T_{\text{moyenne}} = 25,2^{\circ}\text{C}$) et correspondant aux mois de Juin, juillet, août et septembre

L'examen de ces évolutions montre que le mois d'Août accuse, quel que soit le polluant considéré, les teneurs les plus faibles. Les mois de septembre et octobre présentent des évolutions et des niveaux de pollution comparables. Il existe donc un effet de saison.

Ces résultats montrent qu'il existe pour l'ensemble des polluants une différence significative entre les deux saisons (été/automne). Cette variation saisonnière qui se caractérise par un écart de température de l'ordre de 7°C est d'environ 36% pour le benzène.

IV.2. Etude de la pollution par les TSP et ET associés

Une étude antérieure de la pollution atmosphérique par les TSP et les ET associés au niveau de la zone d'étude a été accomplie, approuvée et publiée par **Bouhila et al., 2015**, nous avons fait appel à sa base de donnée afin de modéliser la pollution par les TSP et dix (10) ET associés, la concentration de ces derniers est illustrée dans le tableau IV.7.

Tableau IV.7: La concentration de polluants à modéliser détectés au niveau de notre zone d'étude, TSP en mg/m^3 et les ET en ng/m^3 (**Bouhila et al., 2015**)

Elément	Moyenne	Min	Max	Ecart type
TSP	32,56	3,8	81,1	15,18
Sm	0,07	0,01	0,09	0,033
Br	2,39	0,86	4,05	1,69
Yb	0,04	0,005	0,128	0,024
Sb	0,71	0,36	1,09	0,514
Sc	0,15	0,06	0,29	0,1915
Hf	0,04	0,002	0,08	0,03
Na	106	4	331	172,91
Ce	1,3	0,64	2,28	0,856
Cr	1,43	0,19	3,29	1,638
La	0,66	0,14	2	0,551

On remarque que le chlore, le fer, le calcium et le sodium sont les éléments les plus concentrés dans l'atmosphère de Draria.

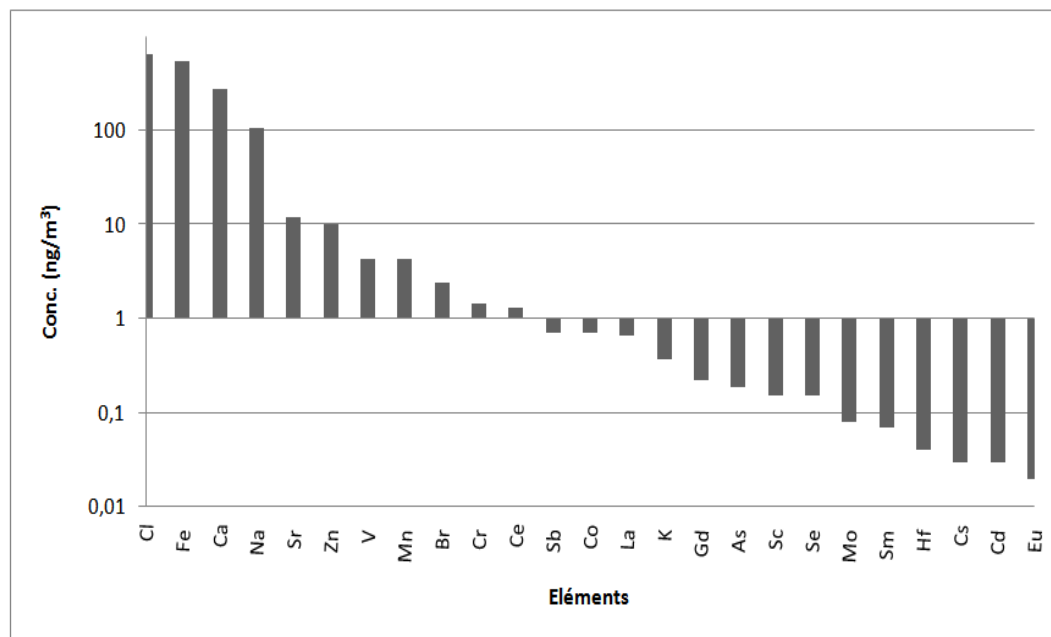


Figure IV.16: *Éléments trace associés aux TSP (ng/m³), déterminés par INAA dans les échantillons d'air de la zone d'étude (Bouhila et al., 2015)*

Le secteur "transport" ne concerne pas seulement le transport routier, qui contribue en grande partie à la pollution particulaire ambiante (Aleman et al., 2010), mais comprend également le transport ferroviaire, aérien et maritime. Nous remarquons aussi des valeurs de Sb assez faibles, en particulier pour le Sc, Cr, Gd, La, et Zn indiquant ainsi l'uniformité et la stabilité de leurs concentrations autour de la station d'échantillonnage pendant la durée des prélèvements.

Comme les données sur les concentrations élémentaires ont suggéré que la relation entre les éléments peut exister, ils ont été testés pour les corrélations de Pearson. La matrice de corrélation effectuée pour les échantillons de l'atmosphère de la ville de Draria est présentée dans le tableau IV.8, où les corrélations possibles sont indiquées en gras. Les coefficients de corrélation linéaire de Spearman (r) avec un test bilatéral de signification (p) ont été calculés pour montrer la relation linéaire entre les différents éléments étudiés. La signification du coefficient de corrélation augmente avec la taille de l'échantillon. Dans notre cas, la taille de l'échantillon est suffisamment grande pour donner une importance pour les corrélations de Pearson.

Tableau IV.8: Matrice de corrélation de Spearman des concentrations des éléments trace associés aux TSP, déterminées par INAA

	TSP	Br	Ce	Cr	Hf	La	Na	Sb	Sc	Sm	Yb
TSP	1	0,142	0,285	0,034	0,314	0,211	0,114	0,302	0,232	0,303	0,332
Br		1	0,462	-0,065	0,515	0,427	0,644	0,489	0,226	0,537	0,198
Ce			1	0,205	0,606	0,641	0,136	0,361	0,322	0,422	0,208
Cr				1	0,325	0,031	-0,077	-0,03	0,091	0,100	0,141
Hf					1	0,335	0,124	0,513	0,463	0,460	0,248
La						1	0,466	0,149	0,026	0,342	0,110
Na							1	0,097	0,067	0,407	0,104
Sb								1	0,383	0,429	0,140
Sc									1	0,234	0,008
Sm										1	0,243
Yb											1

Un examen des résultats donnés dans le tableau IV.8 montre que des corrélations positives significatives entre les éléments analysés ont été trouvées pour plusieurs couples d'éléments : Br-Hf ($r = 0,515$), Br-Na ($r = 0,0644$) et Br-Sm ($r = 0,537$), ce qui prouve que ces éléments peuvent être rejetés par les mêmes sources. Le Ce se trouve à être en corrélation positive avec Hf ($r = 0,513$) et La ($r = 0,641$), à nouveau le cas d'une source commune, ceci est également remarqué pour le couple Hf-Sb ($r = 0,513$). En plus des sources anthropiques communes, l'émission de ces polluants est aussi tributaire des conditions météorologiques, comme les températures et la vitesse du vent.

Il est intéressant de noter que (Sb et Cd), généralement considérés comme de bons traceurs de la circulation (*Wang et al. 2003*) provenant généralement de la re-suspension des poudres de frein (*Iijima et al, 2009; Belzile et al, 2011*) montrent un peu un comportement différent avec une corrélation plus faible avec le groupe ci-dessus (les coefficients de Spearman < 0,5). En outre, la très faible corrélation entre Zn qui a de multiples sources anthropiques (comme une composante de débris de caoutchouc du pneu) (*Almeida et al., 2011*), avec d'autres éléments comme le Gd, Mn, Cl, Sm, Br, La, Sb et Ca suggère encore d'autres sources pour eux, remarqué précédemment par *Sternbeck et al. (2002)*.

La fiabilité de ces interactions inter-éléments et leur signification exacte nécessite une évaluation plus poussée.

IV.3. Modélisation de la pollution par la méthode ARIMA

L'algorithme élaboré et présenté dans le chapitre précédent a été traduit en langage machine par le programme Matlab, l'exécution de ce programme nous a donné les résultats suivant :

Tableau IV.9: Tableau récapitulatif de la modélisation par le modèle ARIMA

	Paramètres du modèle					Résultat modèle				Simulation
	Cste	D	Saisonnalité	MAIgs	SMAIgs	P	Q	Vari.	Distrib.	r^2 sim.
TSP	0	1	1	1	1	2	2	8,9319	Gausse	98,0904
Ce	0	1	1	1	1	2	2	0,0842	Gausse	94,0518
Cr	0	1	1	1	1	2	2	0,1151	Gausse	97,8258
Hf	0	1	1	1	1	2	2	0	Gausse	97,7448
La	0	1	1	1	1	2	2	0,0685	Gausse	88,1275
Na	0	1	1	1	1	2	2	914,184	Gausse	98,4664
Sb	0	1	1	1	1	2	2	0,0159	Gausse	97,0432
Sc	0	1	1	1	1	2	2	0,0002	Gausse	99,6524
Sm	0	1	1	1	1	2	2	0	Gausse	96,9214
Yb	0	1	1	1	1	2	2	0	Gausse	98,2665
Br	0	1	1	1	1	2	2	0,0718	Gausse	98,7487

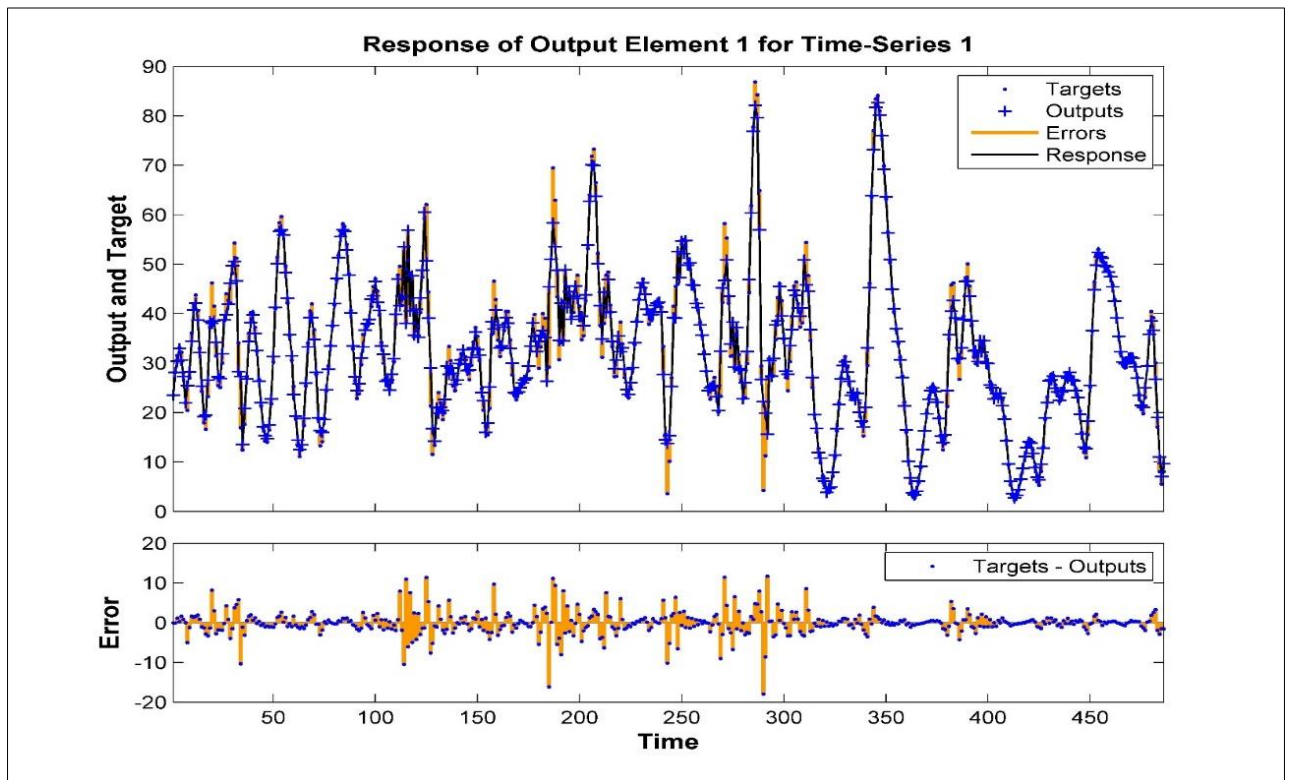


Figure IV.17: Modélisation des TSP par le modèle ARIMA

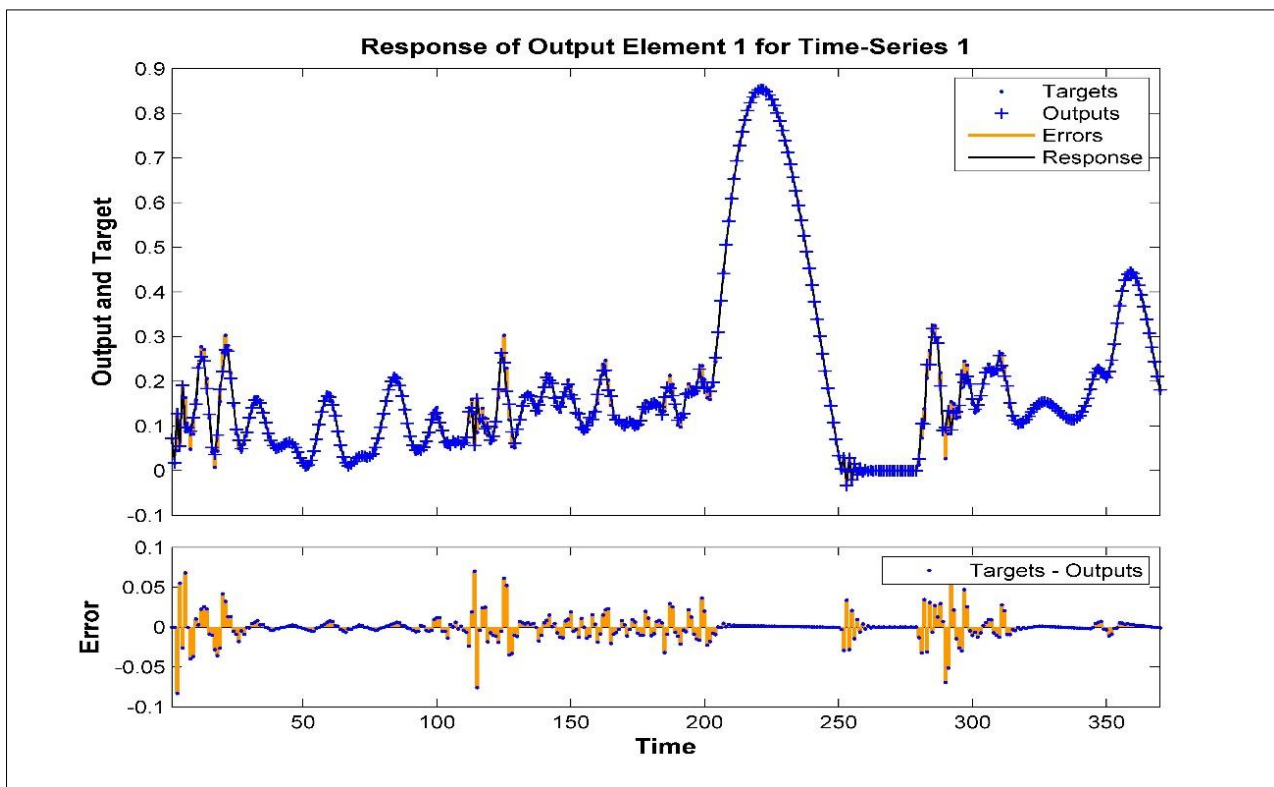


Figure IV.18: Modélisation du Scandium (Sc) par le modèle ARIMA

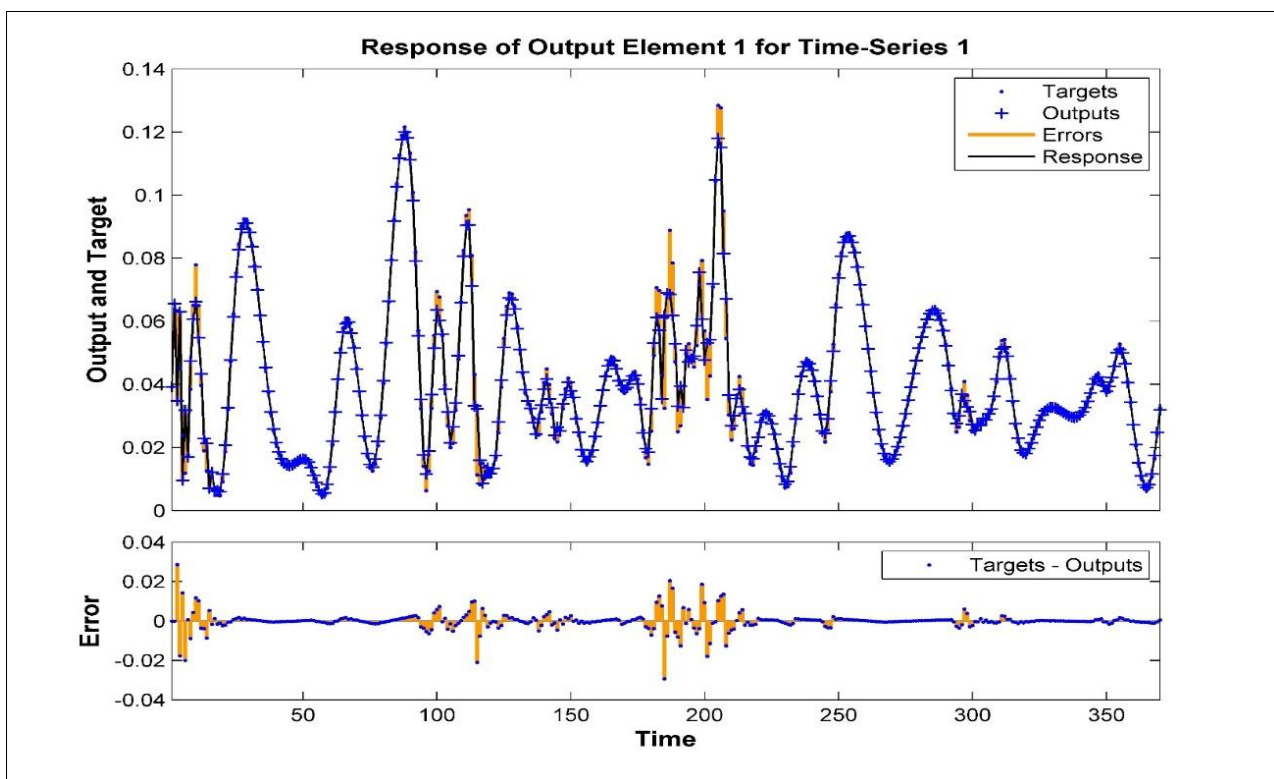


Figure IV.19: Modélisation d'Ytterbium (Yb) par le modèle ARIMA

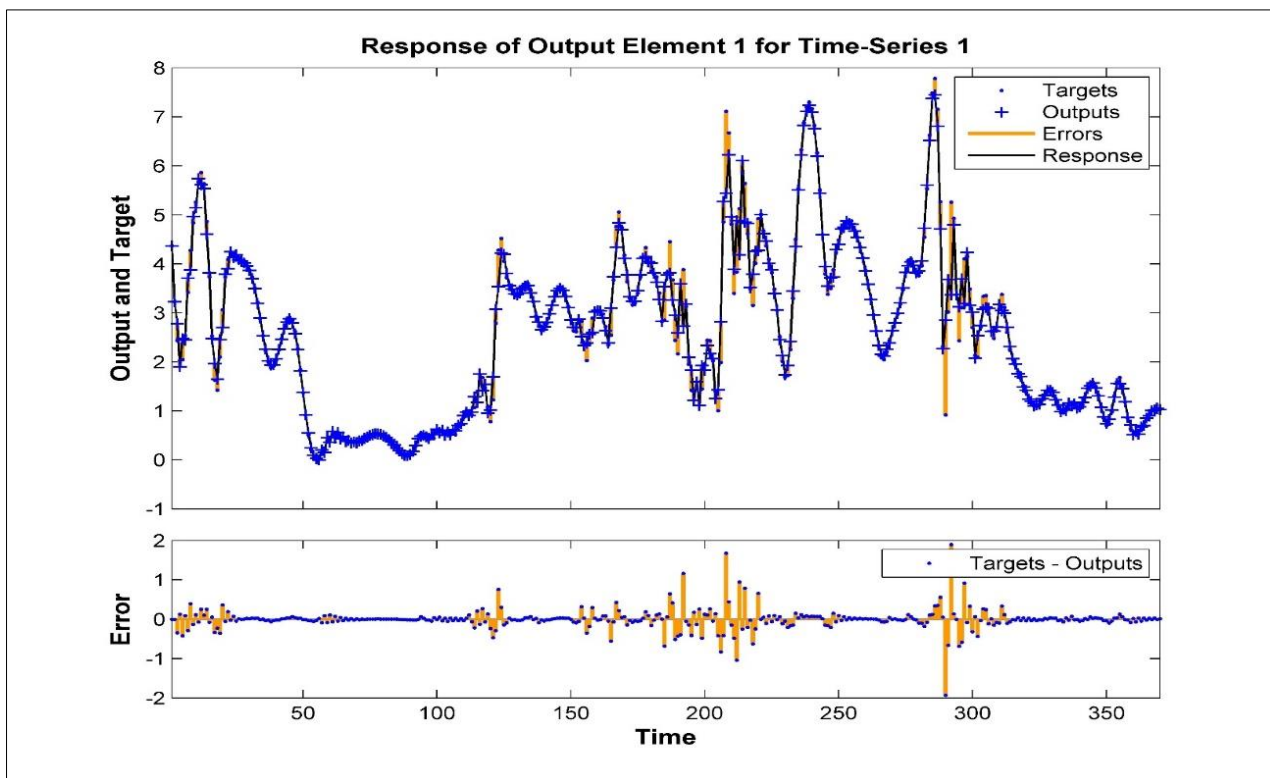


Figure IV.20: Modélisation du Brome (Br) par le modèle ARIMA

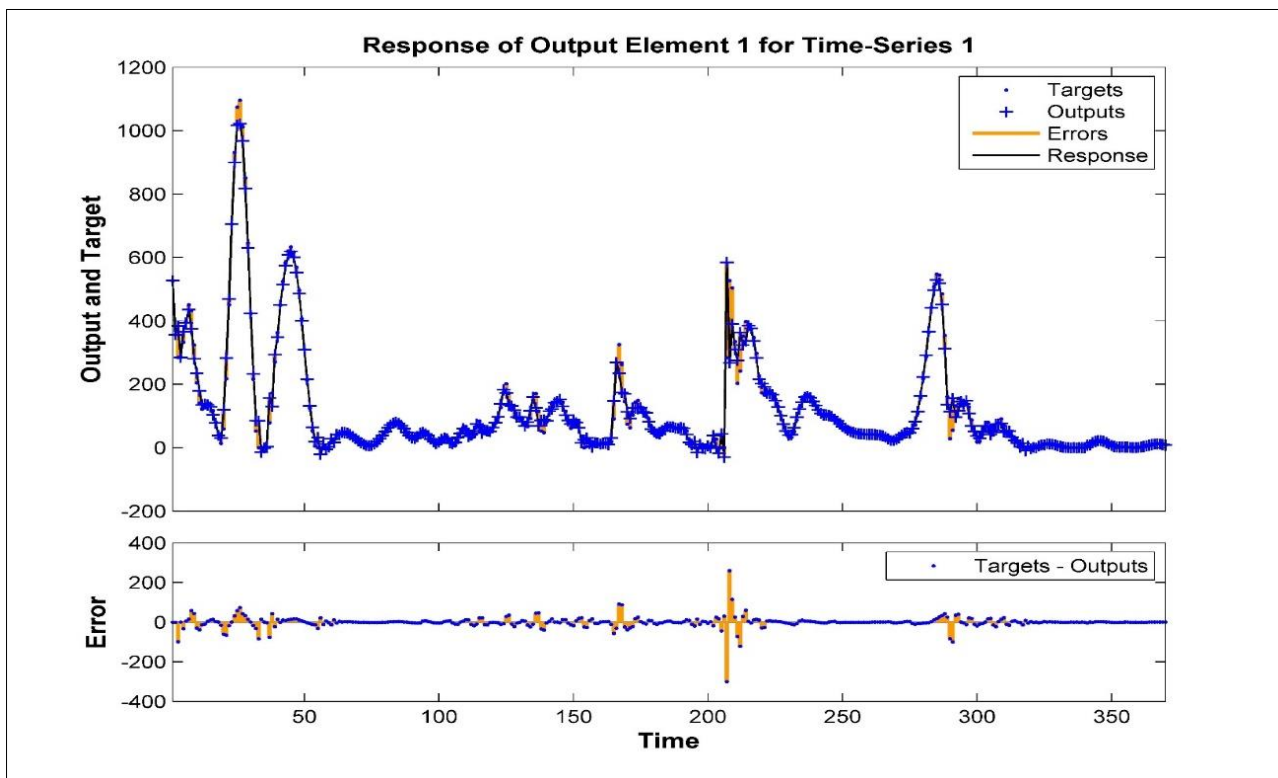


Figure IV.21: Modélisation du Sodium (Na) par le modèle ARIMA

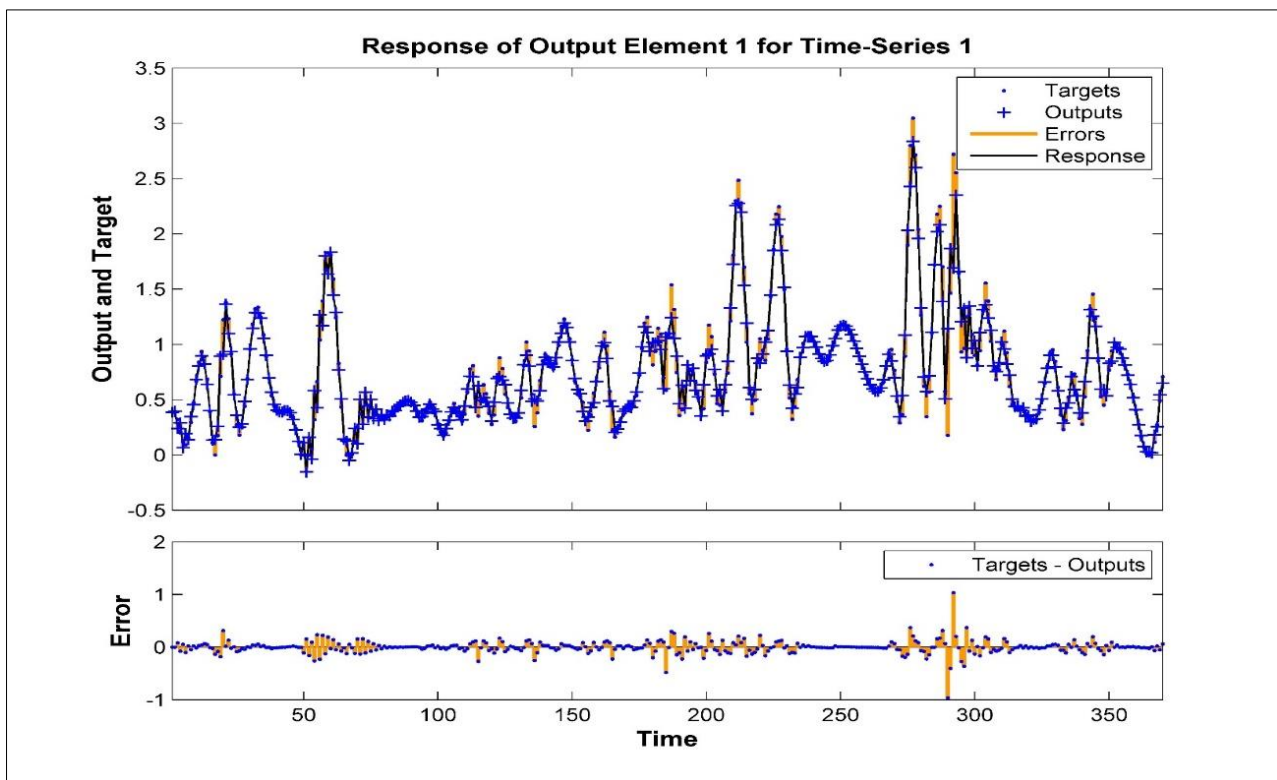


Figure IV.22: Modélisation de l'Antimoine (Sb) par le modèle ARIMA

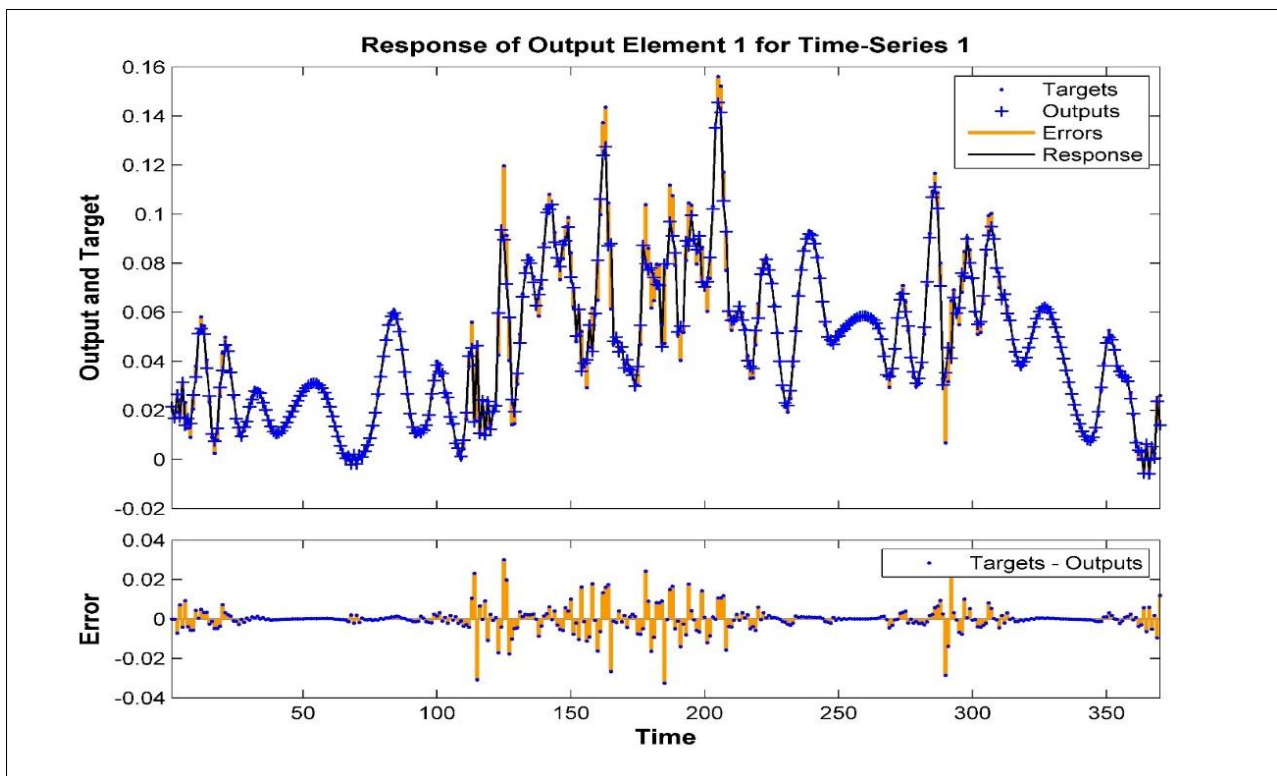


Figure IV.23: Modélisation du Hafnium (Hf) par le modèle ARIMA

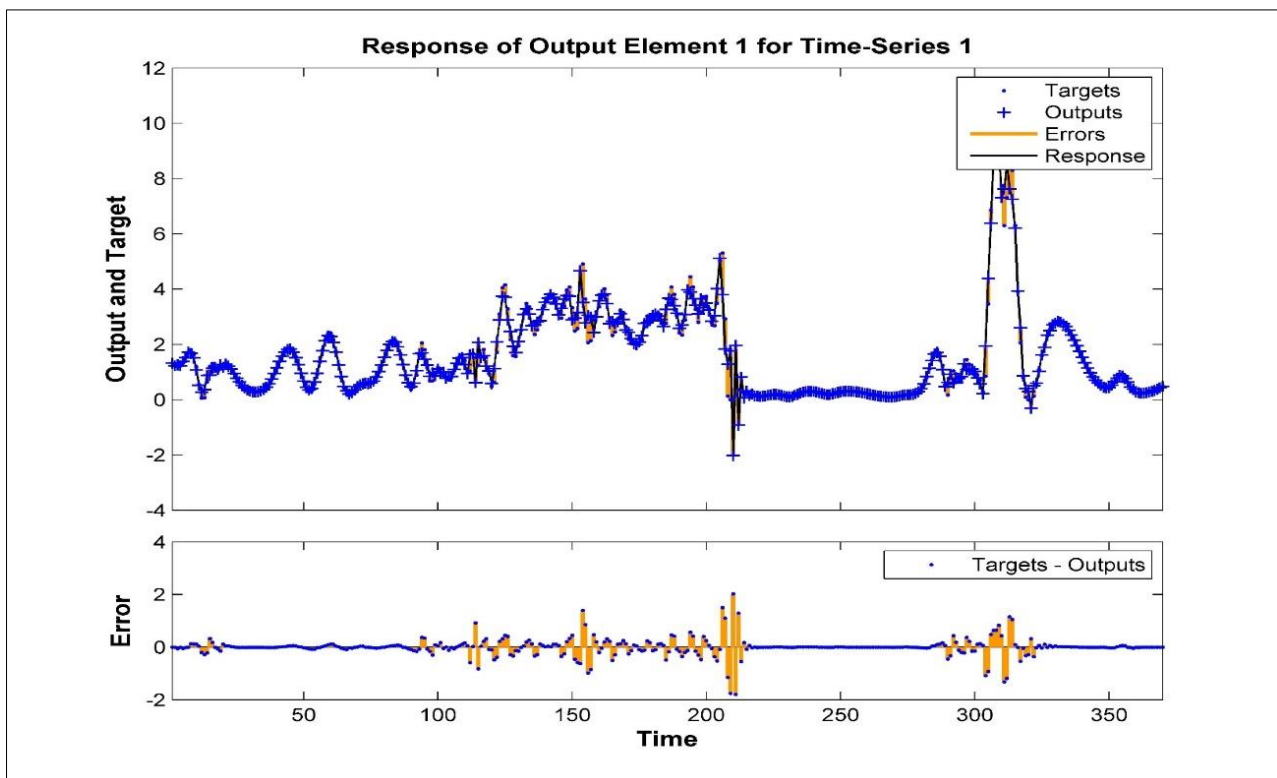


Figure IV.24: Modélisation du Chrome (Cr) par le modèle ARIMA

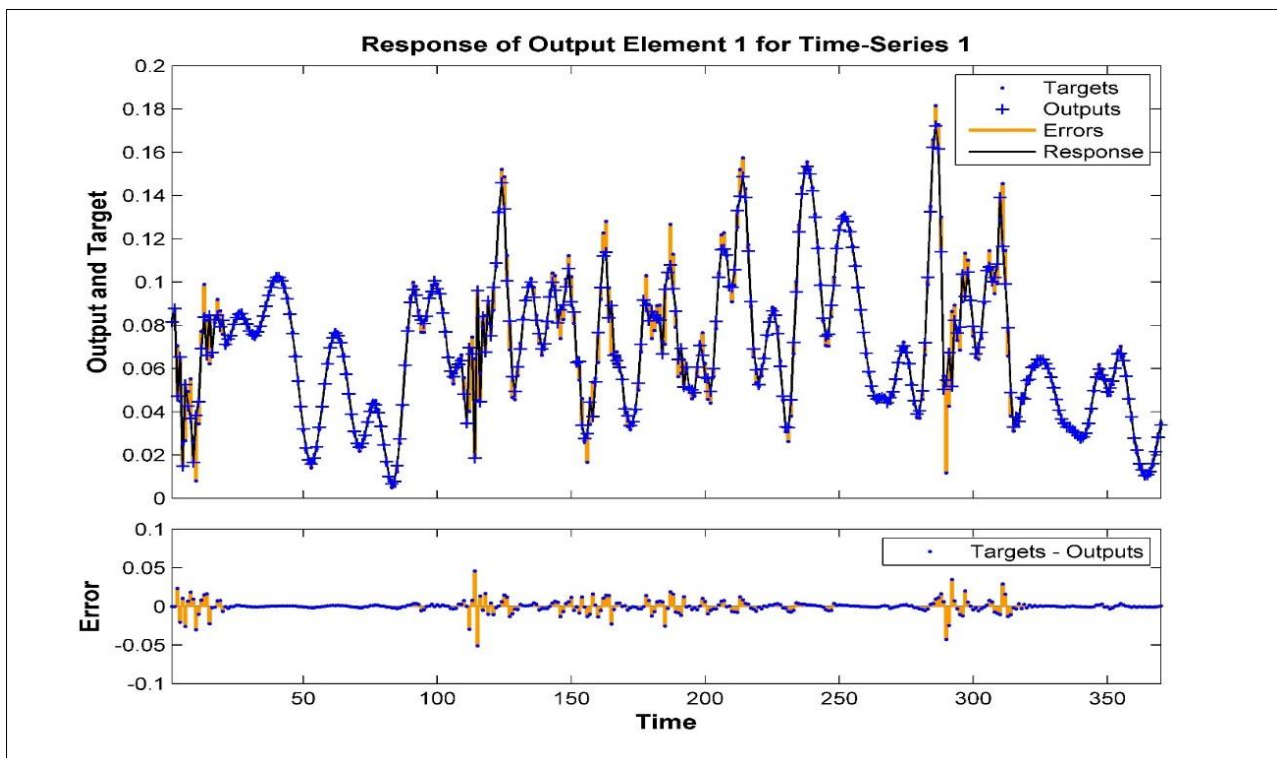


Figure IV.25: Modélisation du Samarium (Sm) par le modèle ARIMA

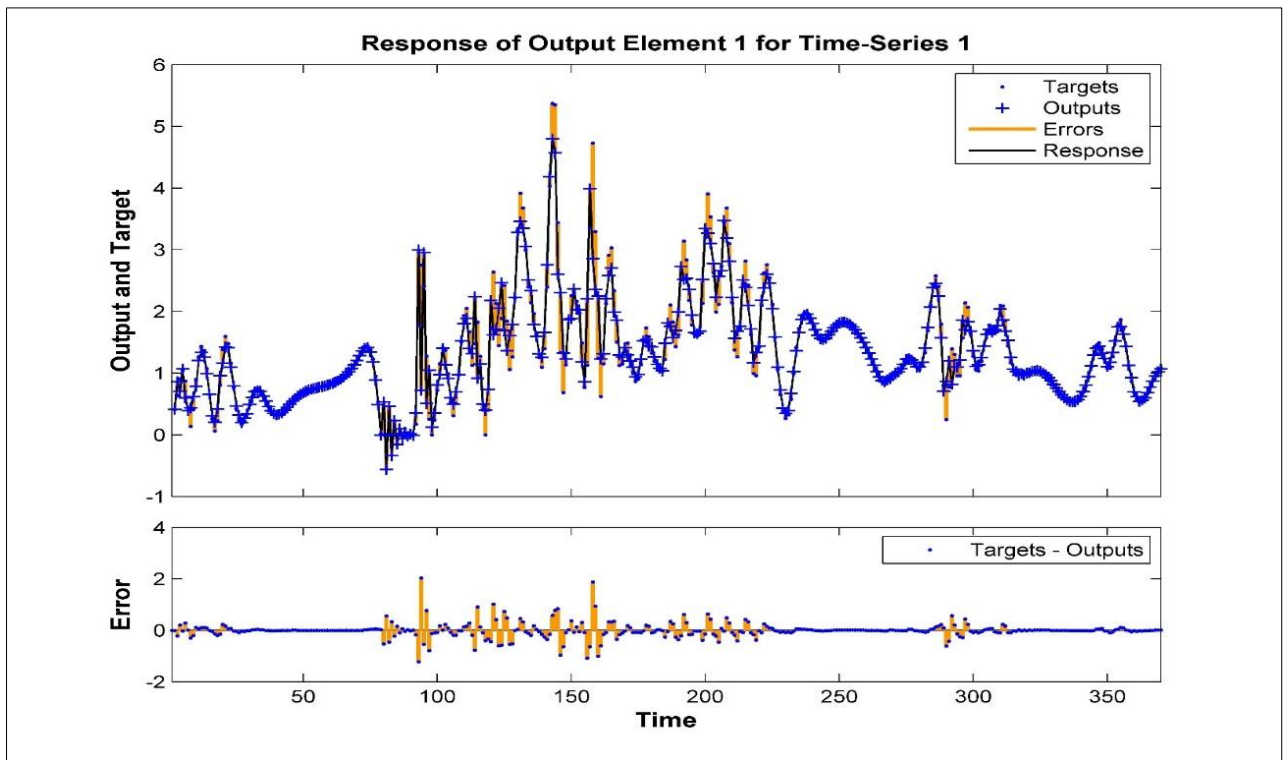


Figure IV.26: Modélisation du Césium (Ce) par le modèle ARIMA

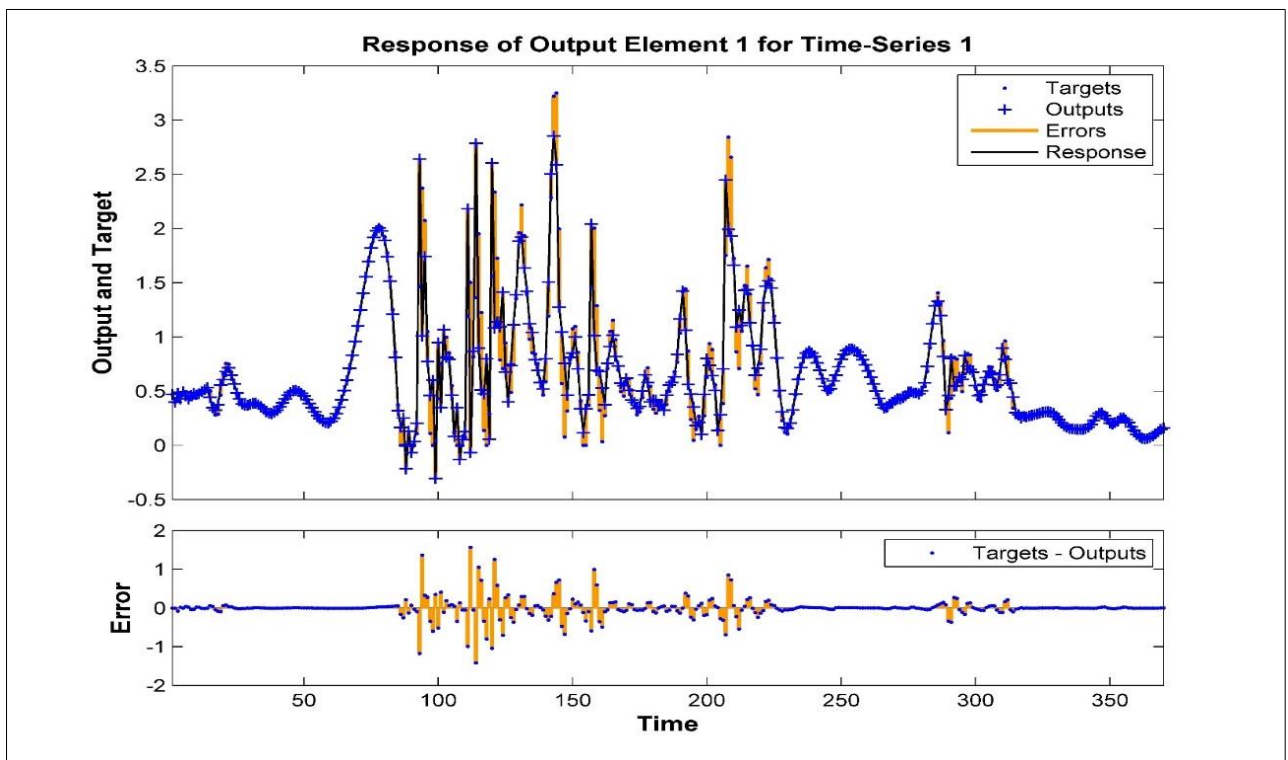


Figure IV.27: Modélisation du Lanthane (La) par le modèle ARIMA

La modélisation par le modèle ARIMA a donné des modèles qui se reprochent de la réalité avec plus au mois des déflexion remarquables durant les jours à très grande

concentration, la marge d'erreur des modèles obtenus est acceptable et il n'y a aucune anomalie à reporter.

IV.4. Modélisation de la pollution par les réseaux de neurones artificiels

Comme nous avons expliqué dans le chapitre précédent, nous avons cherché en première étape à identifier les entrées des modèles à élaborer, la méthode de l'analyse factorielle exploratoire et l'analyse en composantes principales ont fournis le résultat résumé dans les tableaux IV.10 et IV.11

Tableau IV.10: Variance totales expliquée

	Valeurs propres initiales			Extraction somme des carrées		
	Total	% variance	% cumulés	Total	% variance	% cumulés
1	3,491	31,733	31,733	3,491	31,733	31,733
2	1,550	14,088	45,820	1,550	14,088	45,820
3	1,368	12,440	58,260	1,368	12,440	58,260
4	1,179	10,720	68,981	1,179	10,720	68,981
5	0,943	8,571	77,552			
6	0,696	6,328	83,879			
7	0,578	5,252	89,131			
8	0,430	3,908	93,039			
9	0,359	3,267	96,306			
10	0,222	2,016	98,322			
11	0,185	1,678	100,000			

Tableau IV.11: Matrice des composantes principales

	Composantes			
	1	2	3	4
TSP	,506	,092	,444	-,028
Br	,764	-,400	-,007	,058
Ce	,709	,348	-,414	,228
Cr	,257	,715	,090	-,008
Hf	,773	,366	,062	-,231
La	,499	,111	-,499	,544
Na	,263	-,632	,190	,502
Sb	,560	-,231	-,007	-,548
Sc	,417	-,332	-,331	-,425
Sm	,775	-,146	,176	,015
Yb	,277	,137	,750	,202

A partir de ces résultats nous avons pu identifier quatre entrées pour chaque élément à modéliser, comme illustré dans le chapitre précédent, l'historique (J , $J-1$, $J-2$, $J-3$ et $J-4$) des concentrations représente l'unique entrée pour modéliser l'élément en question.

La conversion de l'algorithme représenté dans le chapitre précédent sous langage Matlab et son exécution ainsi que notre méthode de modélisation par les réseaux de neurones artificiels sont résumés dans le tableau IV.12.

Tableau IV.12: Tableau récapitulatif de la modélisation des TSP et les ET associés par les réseaux de neurones artificiels

Élément	Network architecture		Performance					Simulation				
			MSE	Itérations	MSE			r^2 (%)	RMSE	Error (%)		
	Architecture	Fonction			Train	Validation	Test			Classe 1	classe 2	classe 3
TSP	[40-1]	L-P	3,385	215	3,308	3,387	7,277	98,61	3,460	75,00	91,00	0,23
Sm	[10-2-1]	L-L-P	0,037	2172	0,022	0,037	0,017	98,97	0,005	78,14	94,54	1,09
Br	[2-2-1]	L-L-P	2,711	161	1,796	2,757	2,429	99,36	1,407	79,51	92,90	1,09
Yb	[10-2-1]	L-L-P	0,017	126	0,005	0,017	0,015	99,08	0,003	86,89	96,17	1,09
Sb	[10-2-1]	L-L-P	0,093	423	0,664	0,930	0,754	98,42	0,084	75,41	94,81	0,82
Sc	[5-10-2-1]	L-L-L-P	0,168	554	0,139	0,168	0,178	99,80	0,012	80,05	94,81	1,09
Hf	[5-10-2-1]	L-L-P	0,258	1132	0,024	0,027	0,029	98,52	0,050	78,42	93,72	1,64
Na	[5-1]	T-P	464,327	146	658,14	607,97	318,1	99,16	22,470	87,98	96,99	1,37
Ce	[10-2-1]	L-L-P	0,305	1265	0,384	0,363	0,171	97,61	0,179	84,70	94,26	1,37
Cr	[5-10-2-1]	L-L-L-P	0,355	1132	0,485	0,380	0,257	96,29	0,185	82,79	94,81	1,37
La	[5-10-2-1]	L-L-L-P	0,263	321	0,150	0,264	0,306	95,13	0,137	87,16	95,63	1,91

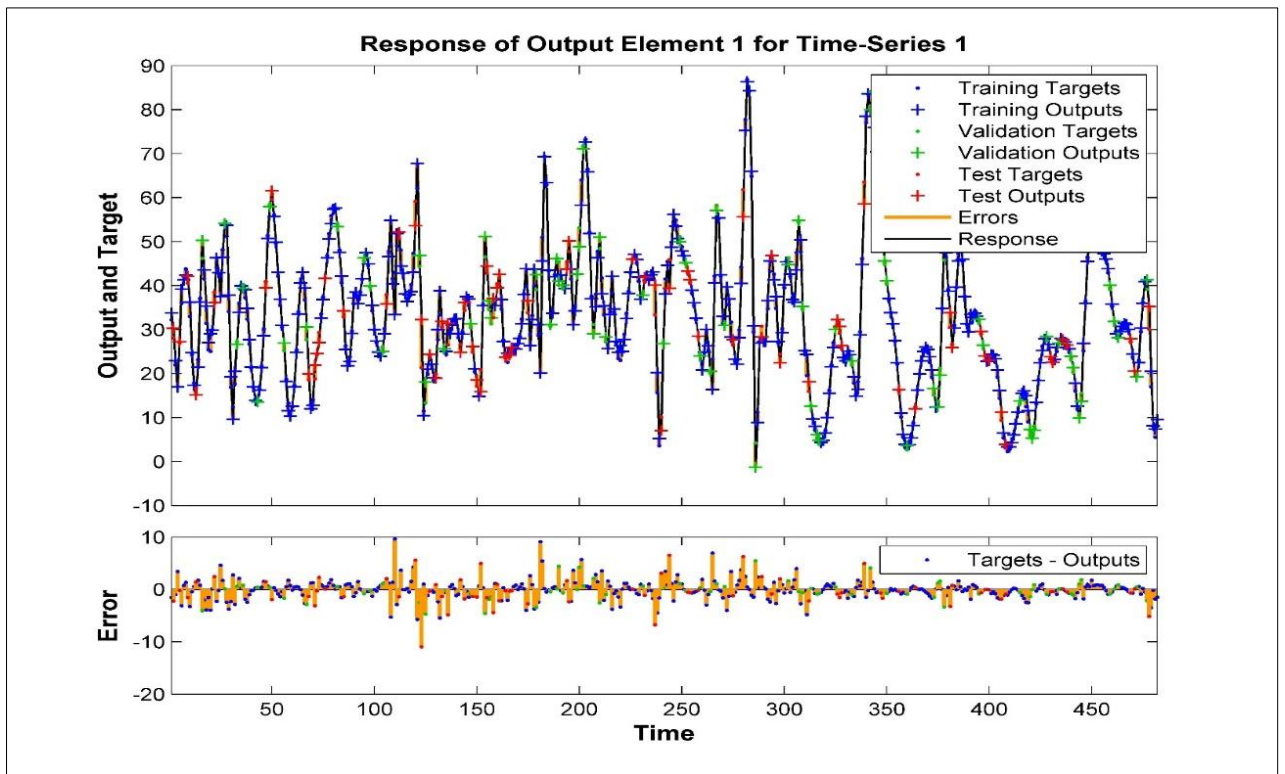


Figure IV.28: Modélisation des TSP par les RNA

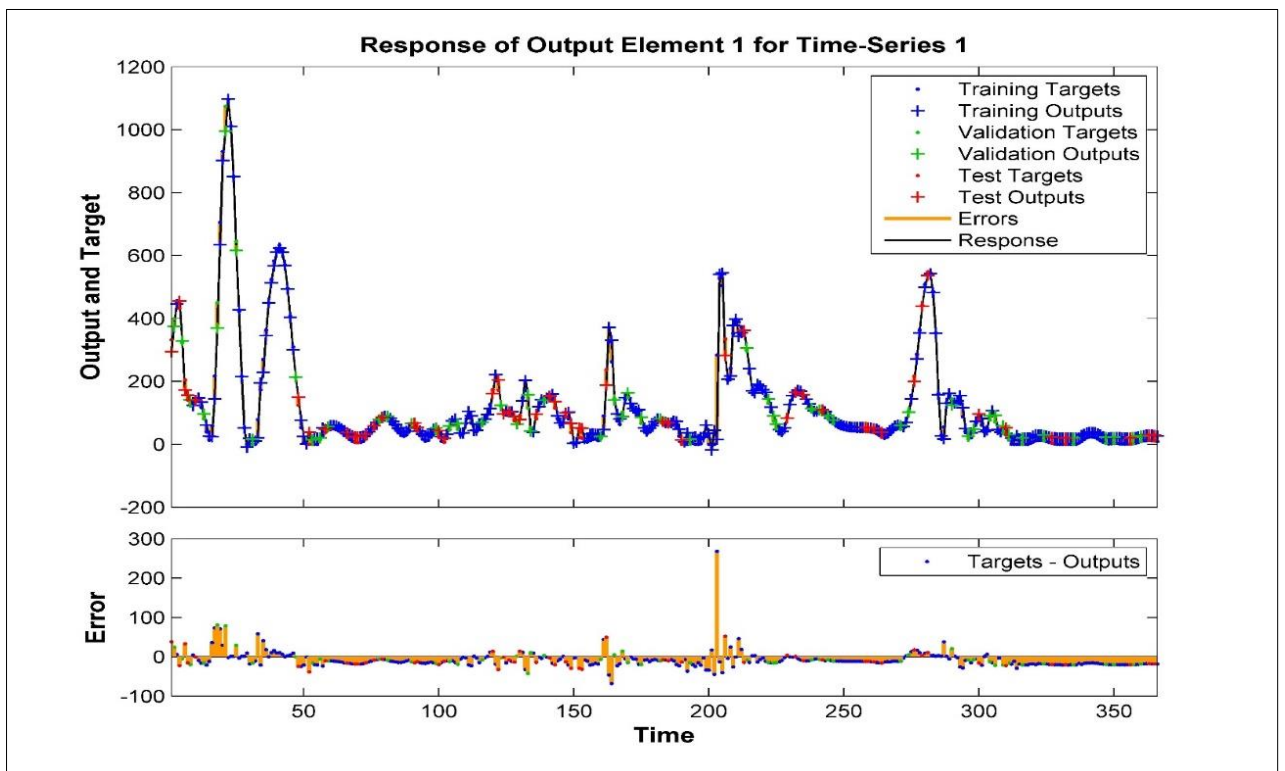


Figure IV.29: Modélisation du Sodium (Na) par les réseaux de neurones artificiels

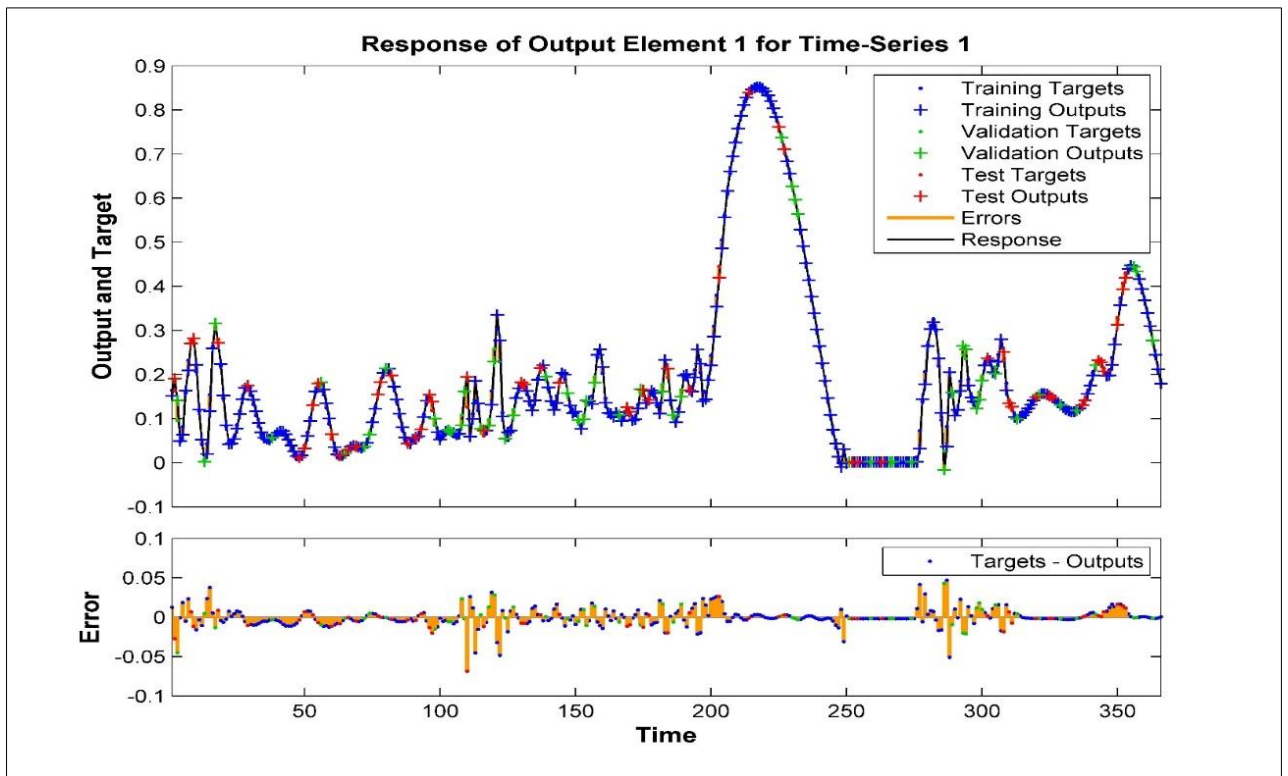


Figure IV.30: Modélisation du Scandium (Sc) par les réseaux de neurones artificiels

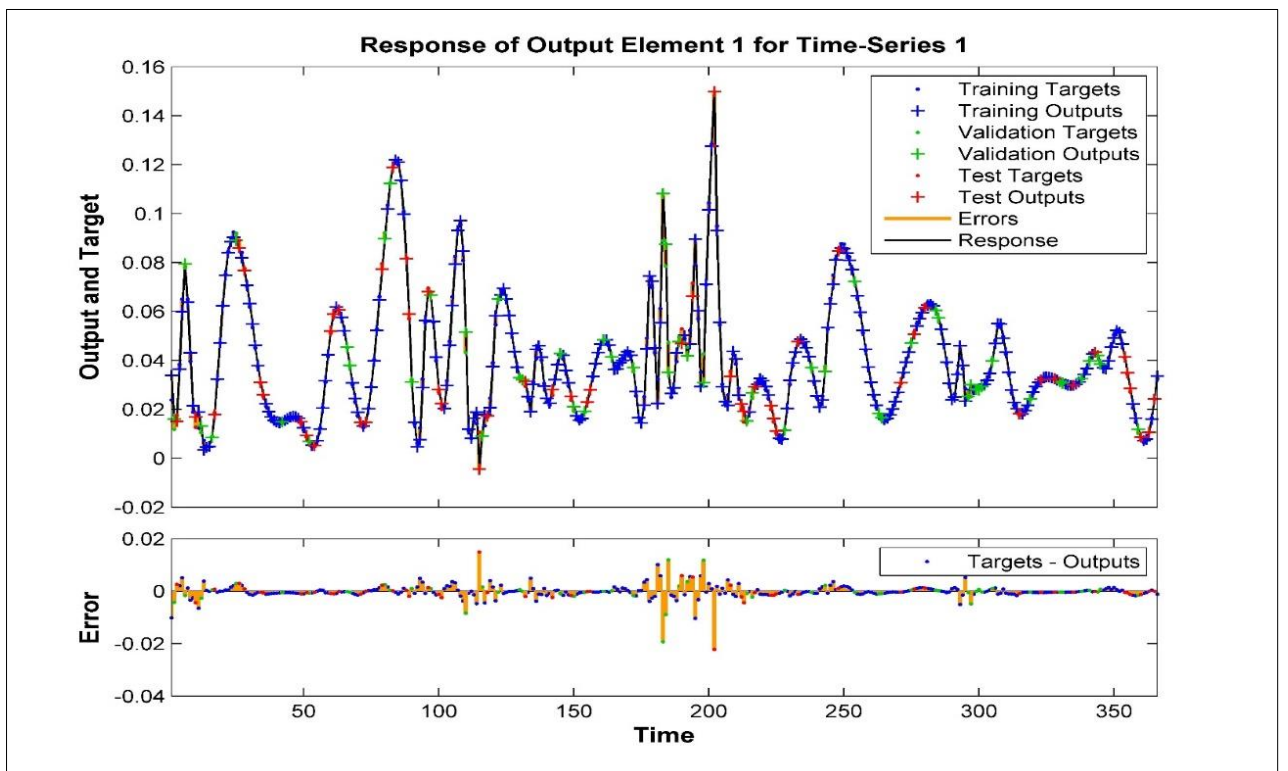


Figure IV.31: Modélisation d'Ytterbium (Yb) par les réseaux de neurones artificiels

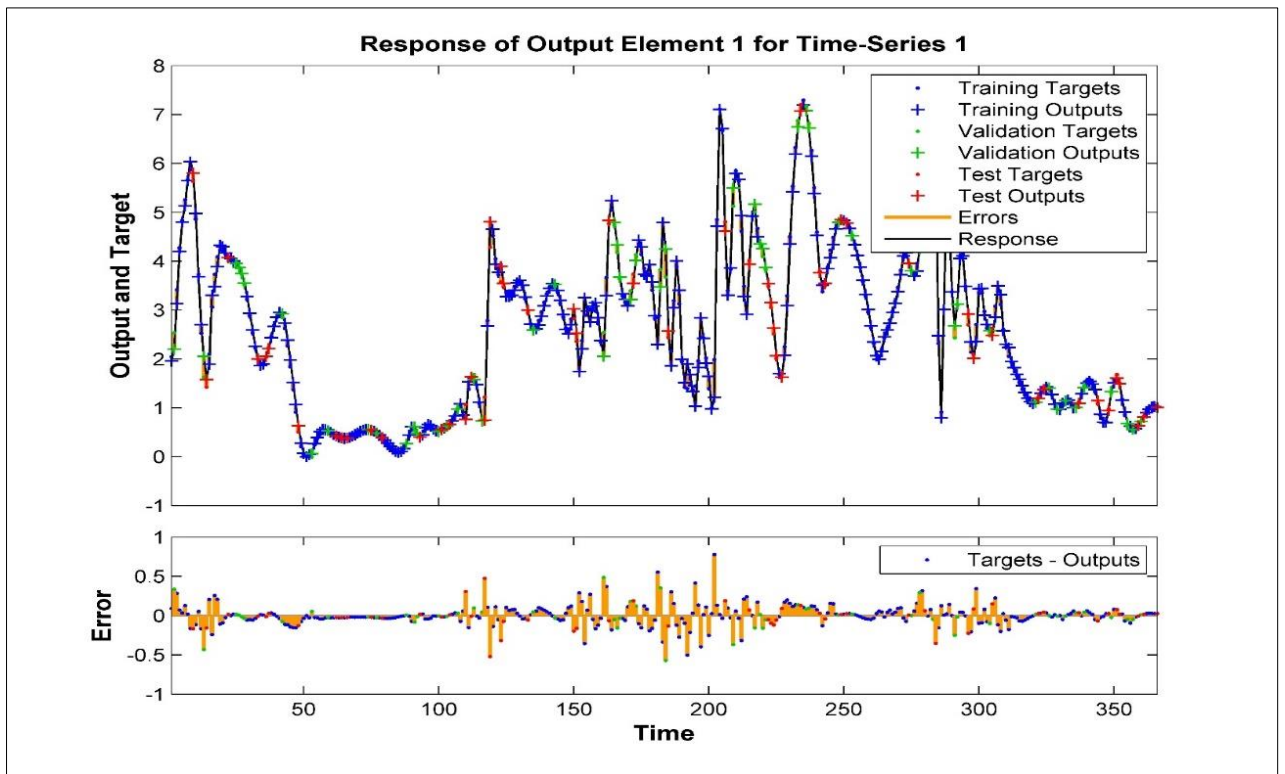


Figure IV.32: Modélisation du Brome (Br) par les réseaux de neurones artificiels

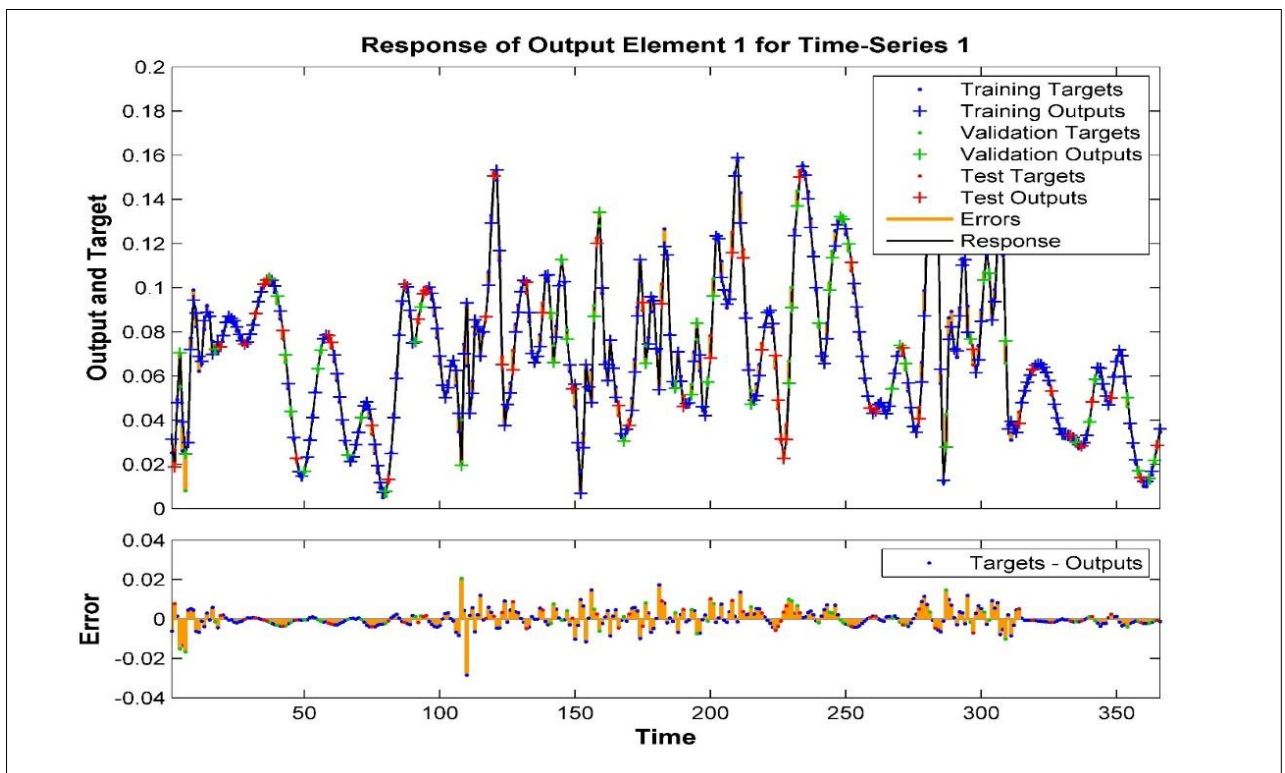


Figure IV.33: Modélisation du Samarium (Sm) par les réseaux de neurones artificiels

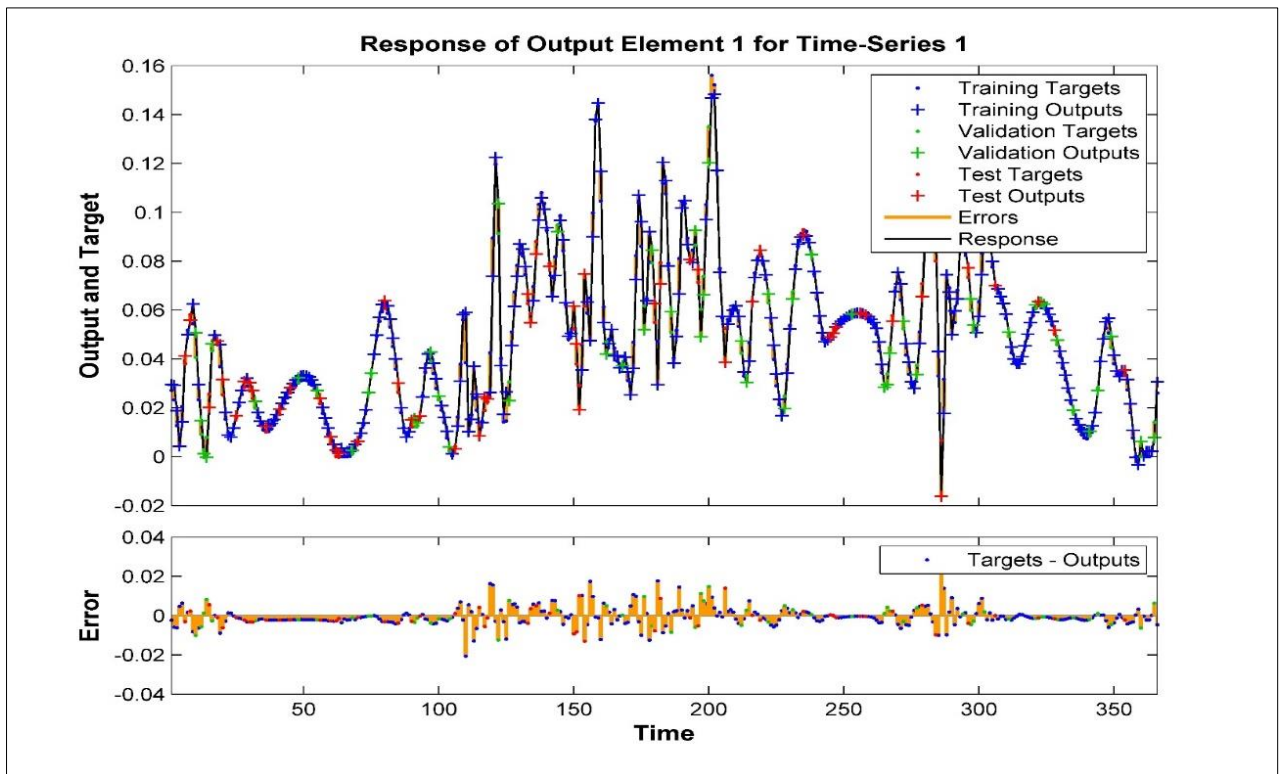


Figure IV.34: Modélisation du Hafnium (Hf) par les réseaux de neurones artificiels

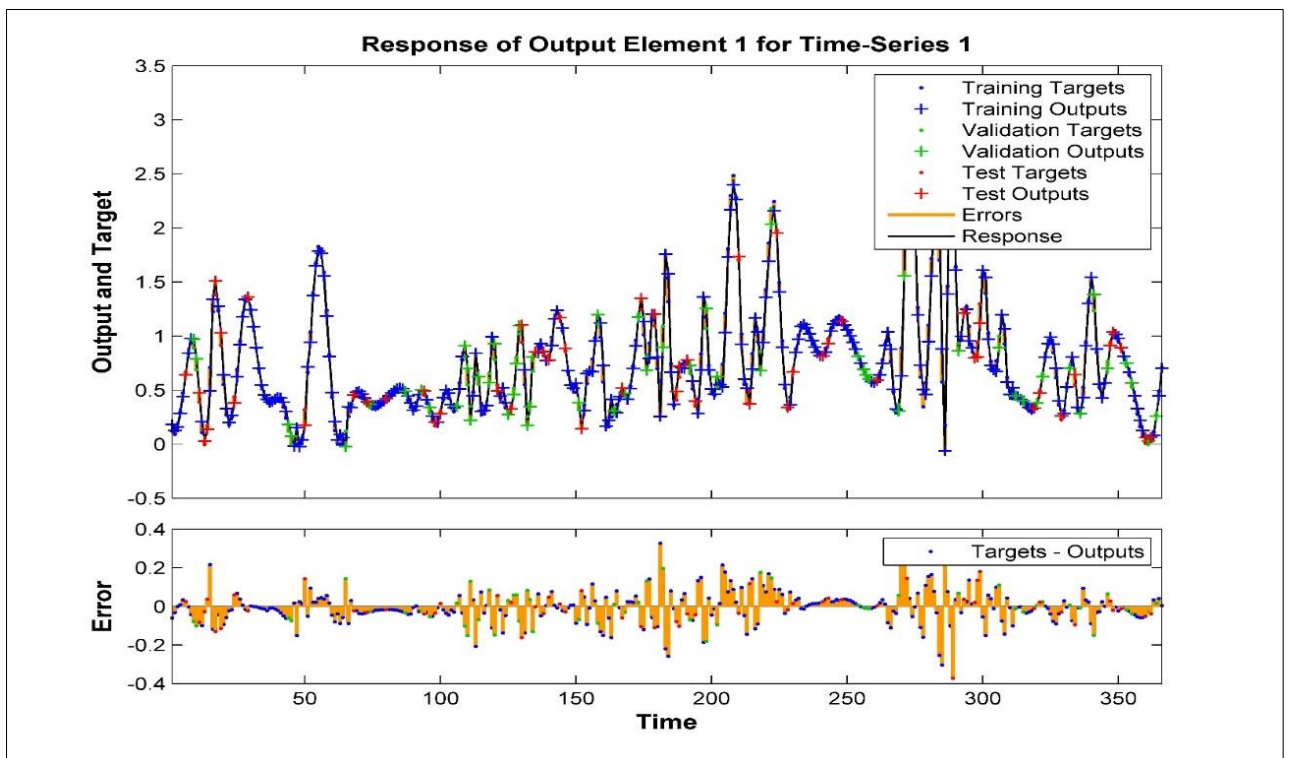


Figure IV.35: Modélisation de l'Antimoine (Sb) par les réseaux de neurones artificiels

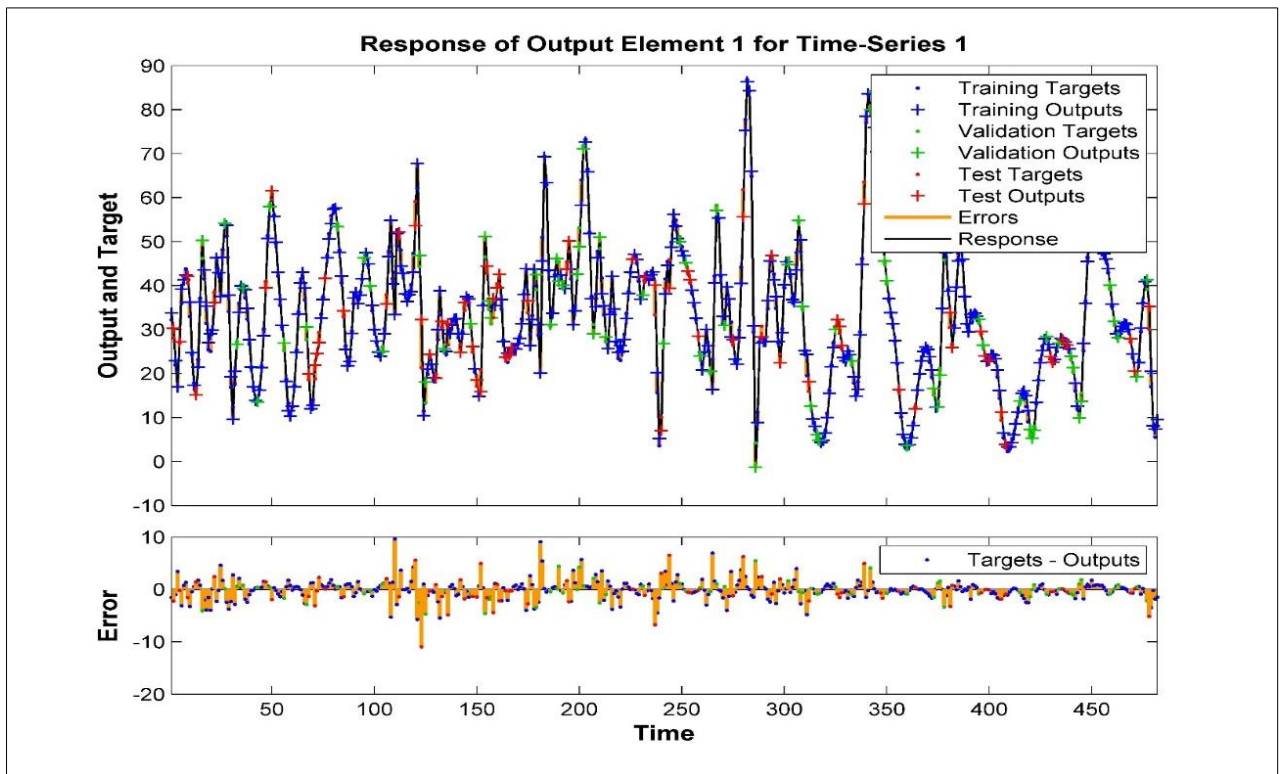


Figure IV.36: Modélisation du Cérium (Ce) par les réseaux de neurones artificiels

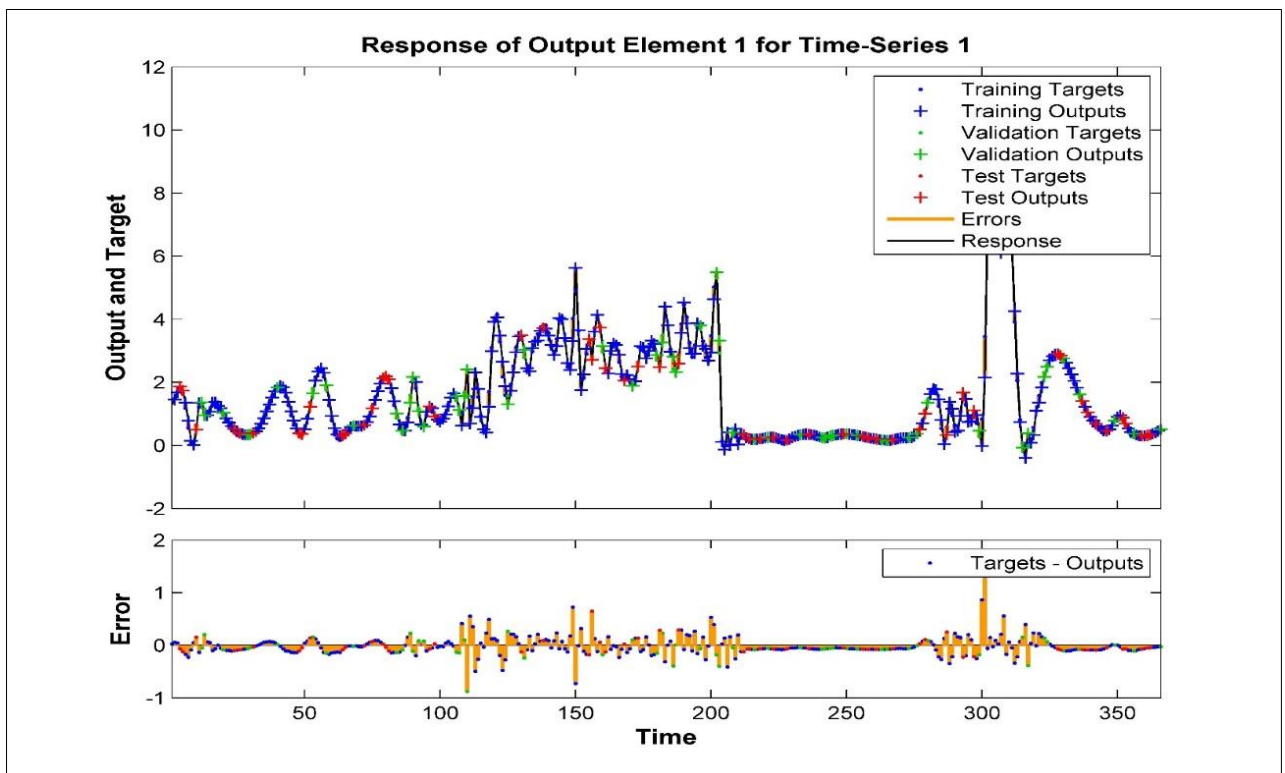


Figure IV.37: Modélisation du Chrome (Cr) par les réseaux de neurones artificiels

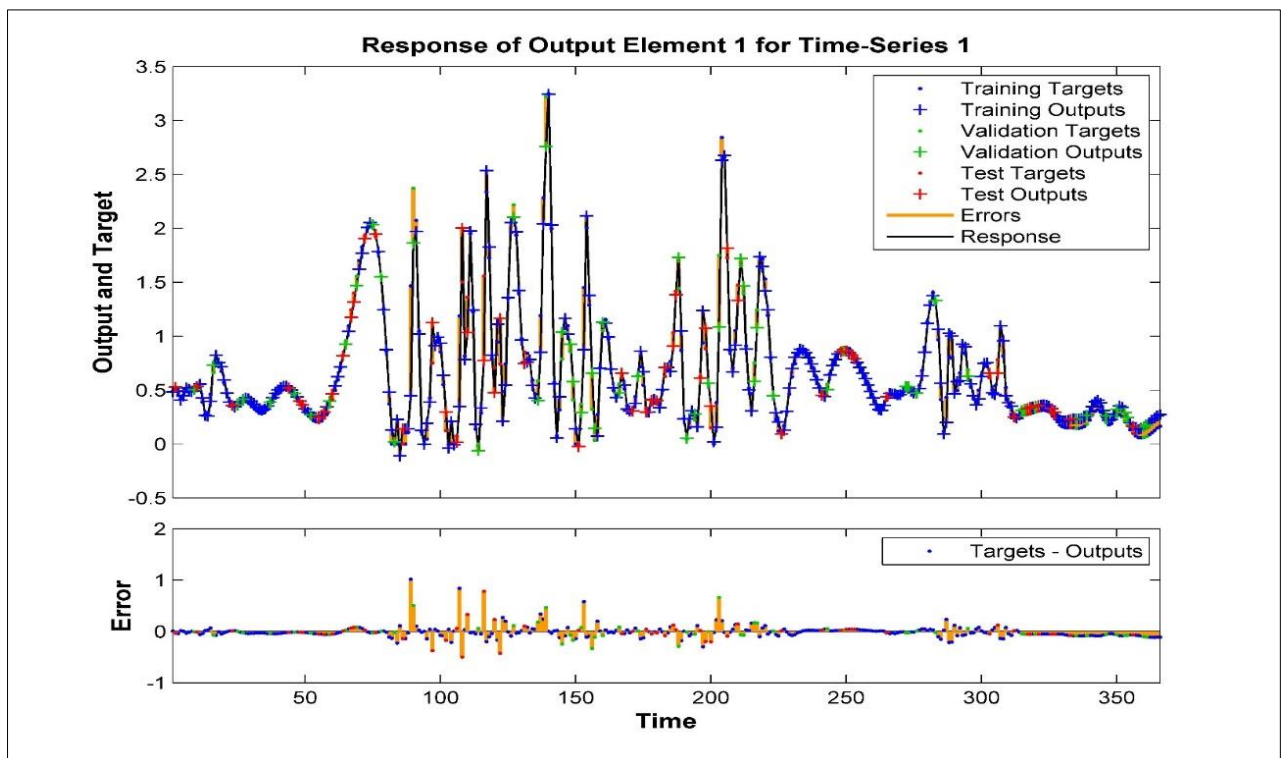


Figure IV.38: Modélisation du Lanthane (La) par les réseaux de neurones artificiels

A la lumière de ces résultats et ces représentations on peut déduire clairement que la modélisation par les réseaux de neurones artificiels a donné des résultats très intéressants en termes de précision et justesse, cette méthode de modélisation a pu couvrir les lacunes trouvées lors de la modélisation par le modèle ARIMA (cas des fortes concentrations et des singularités).

IV.5. Comparaison entre les deux méthodes de modélisation

Tableau IV.13: Comparaison de la modélisation par les deux méthodes

Elément	RNA		ARIMA	
	r^2 (%)	RMSE	r^2 (%)	RMSE
TSP	98,61	3,46	98,0904	3,7093
Ce	98,97	0,005	94,0518	0,1490
Cr	99,36	1,407	97,8258	2,0870
Hf	99,08	0,003	97,7448	0,2481
La	98,42	0,084	88,1275	0,1312
Na	99,8	0,012	98,4664	0,0891
Sb	98,52	0,05	97,0432	0,1013
Sc	99,16	22,47	99,6524	23,275
Sm	97,61	0,179	96,9214	0,7514
Yb	96,29	0,185	98,2665	0,6085
Br	95,13	0,137	98,7487	0,4753

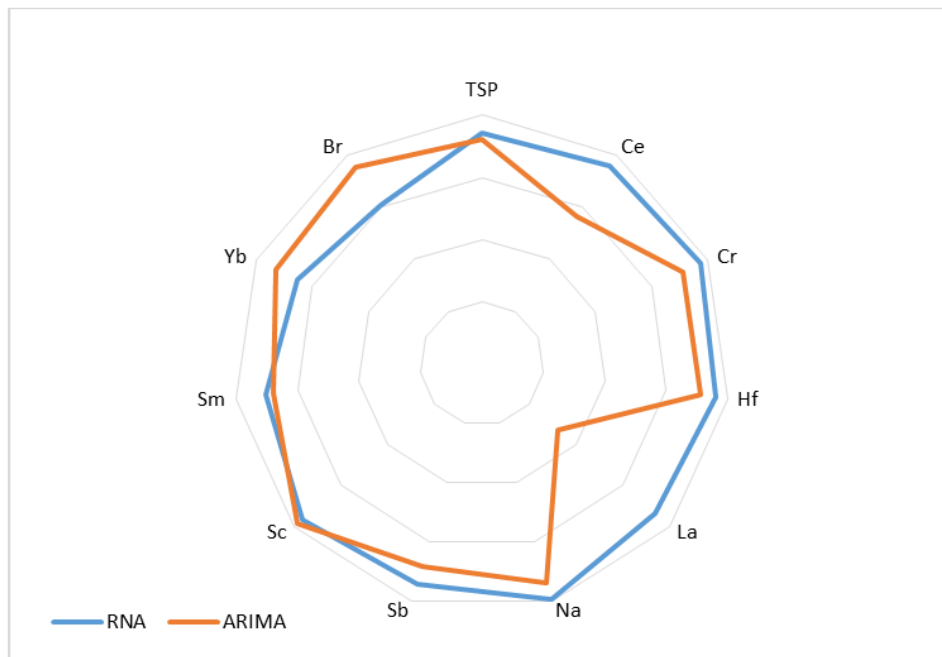


Figure IV.39: Comparaison du coefficient de la régression r^2 des deux méthodes de modélisations

Il est clairement visible sur les figure IV.40 et 41 que la méthode de modélisation par les réseaux de neurones a donné les meilleurs résultat et la meilleure qualité de précision pour l'ensemble des TSP et les ET associés à l'exception du Ytterbium (Yb) et le Brome (Br) qu'est due probablement au sur-apprentissage du réseau soit par le mauvais choix de l'architecture, mais en faisant une combinaison entre les deux critères de validation (r^2 et RMSE) on déduit que la méthode des réseaux de neurones est largement efficace et plus précise.

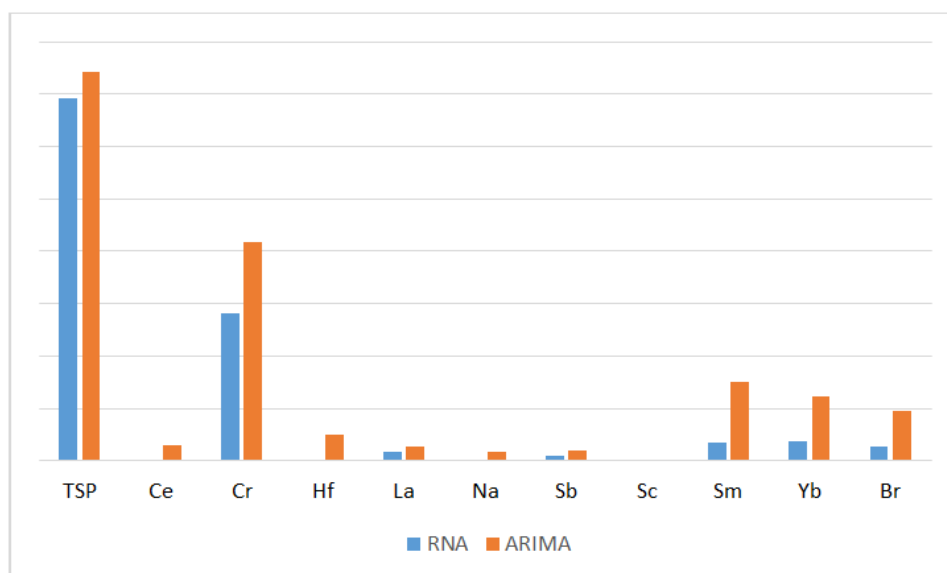


Figure IV.40: Comparaison du RMSE des deux méthodes de modélisations

IV.6. Prédiction de la pollution atmosphérique par la méthode ARIMA

Après l'élaboration et la validation de la modélisation des TSP et des ET associés, nous avons procédé à la prédiction de ces derniers, l'intérêt pratique de cette prédiction c'est d'estimer la pollution *quotidienne* en TSP et en ET associé au niveau de notre site d'étude (Draria) en effectuant que deux prélèvements par semaine et sans analyser le filtre de prélèvement.

L'exécution de l'algorithme présenté dans le chapitre précédent sous langage Matlab nous a donné les résultats présentés dans le tableau ci-dessous :

Tableau IV.14: Tableau récapitulatif de la prédiction des TSP et les ET associés par le modèle ARIMA

	Paramètres du modèle					Résultat modèle				Simulation
	Cste	D	Seasonality	MAIgs	SMAIgs	P	Q	Vari.	Distrib.	r^2 prév.
TSP	0	1	1	1	1	2	2	8,9319	Gauss	87,7963
Ce	0	1	1	1	1	2	2	0,0842	Gauss	66,3676
Cr	0	1	1	1	1	2	2	0,1151	Gauss	63,9621
Hf	0	1	1	1	1	2	2	0	Gauss	63,7931
La	0	1	1	1	1	2	2	0,0685	Gauss	58,1619
Na	0	1	1	1	1	2	2	914,184	Gauss	89,8812
Sb	0	1	1	1	1	2	2	0,0159	Gauss	47,4393
Sc	0	1	1	1	1	2	2	0,0002	Gauss	76,6352
Sm	0	1	1	1	1	2	2	0	Gauss	79,5812
Yb	0	1	1	1	1	2	2	0	Gauss	70,9566
Br	0	1	1	1	1	2	2	0,0718	Gauss	56,5937

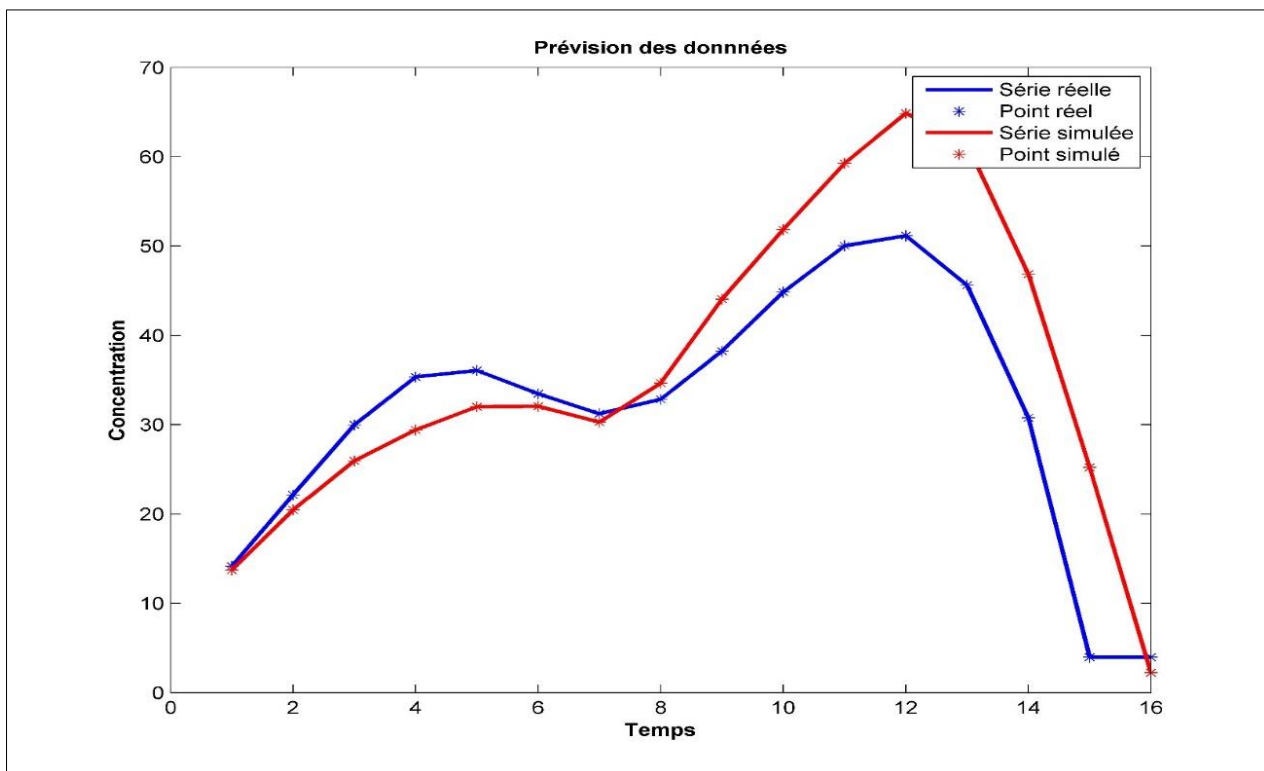


Figure IV.41:Prévision des TSP par le modèle ARIMA

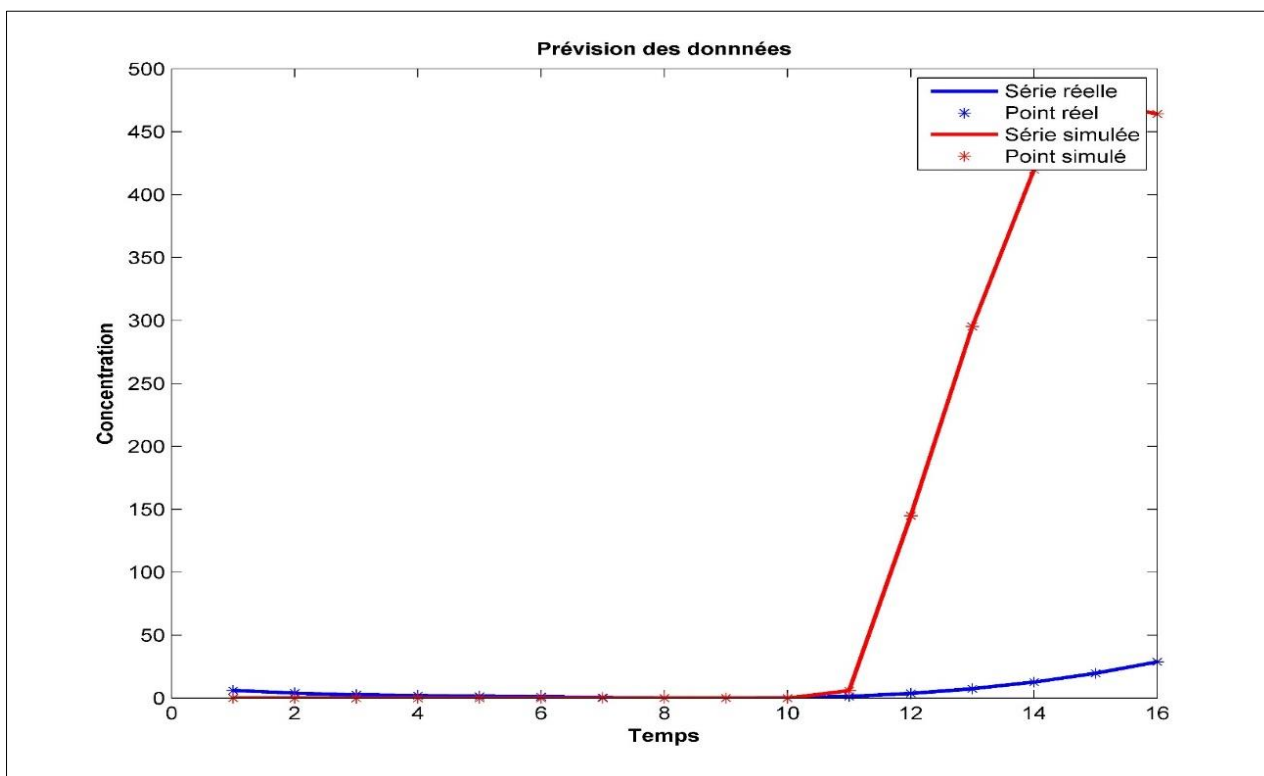


Figure IV.42: Prévision du Sodium(Na) par le modèle ARIMA

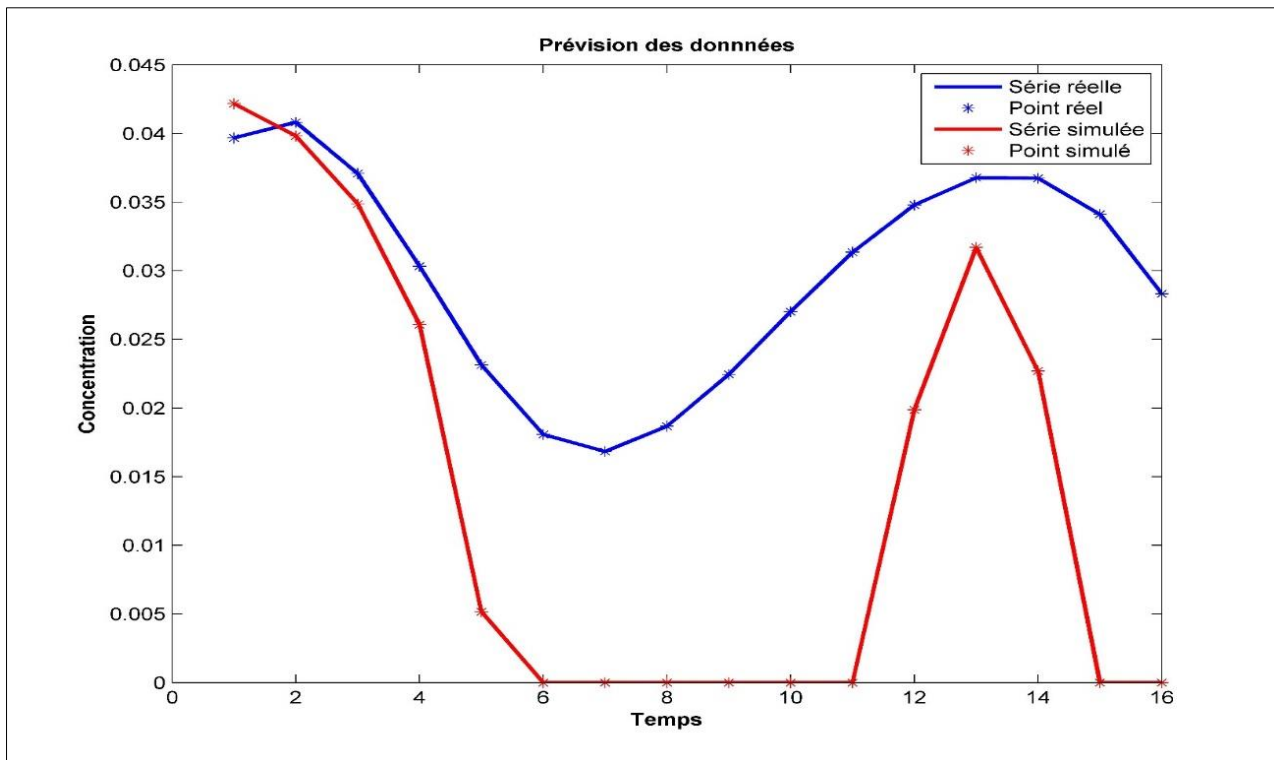


Figure IV.43: Prévision du Samarium (Sm) par le modèle ARIMA

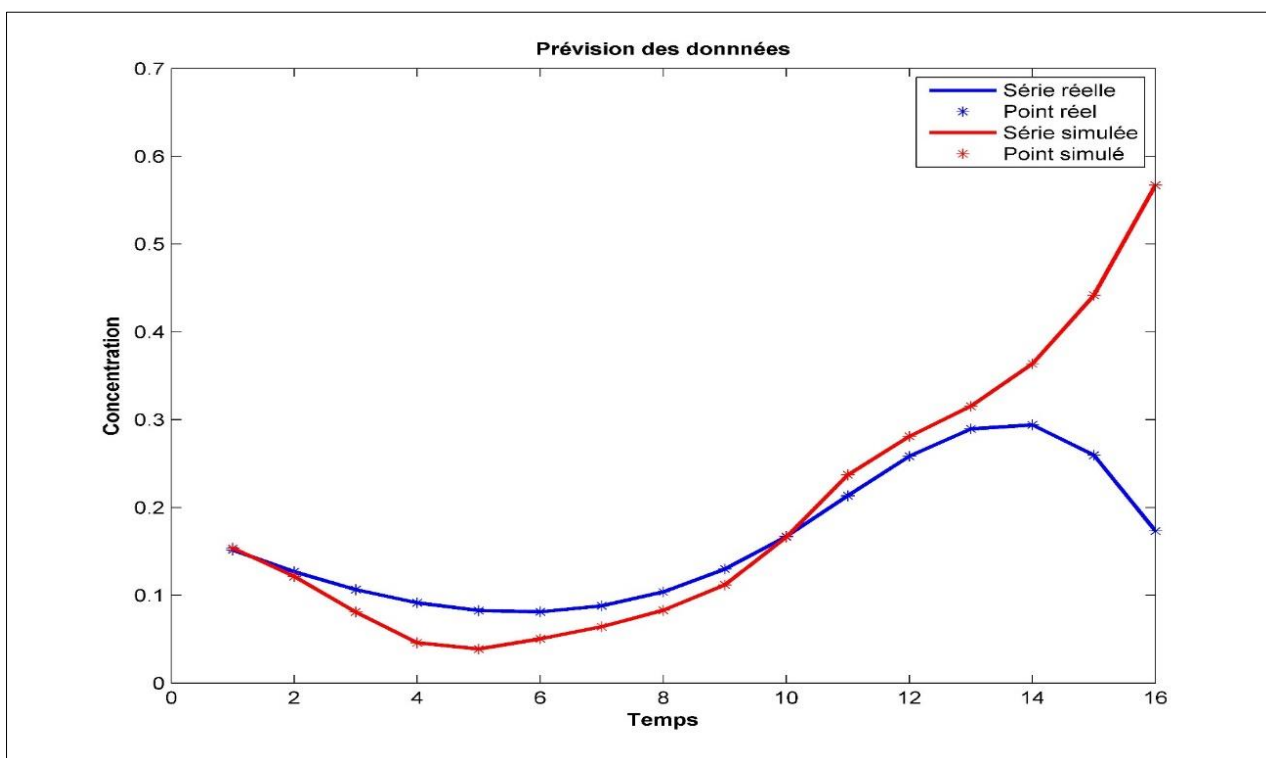


Figure IV.44: Prévision du Scandium (Sc) par le modèle ARIMA

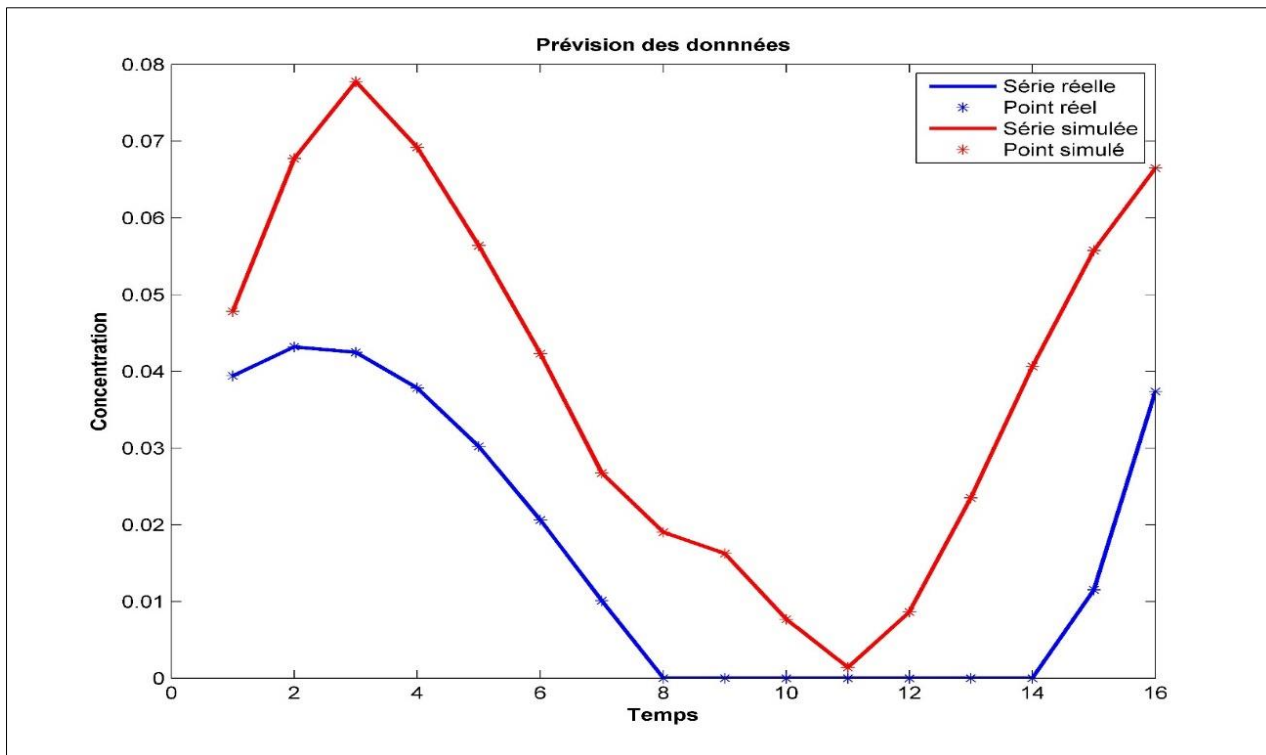


Figure IV.45: Prévision d'Ytterbium (Yb) par le modèle ARIMA

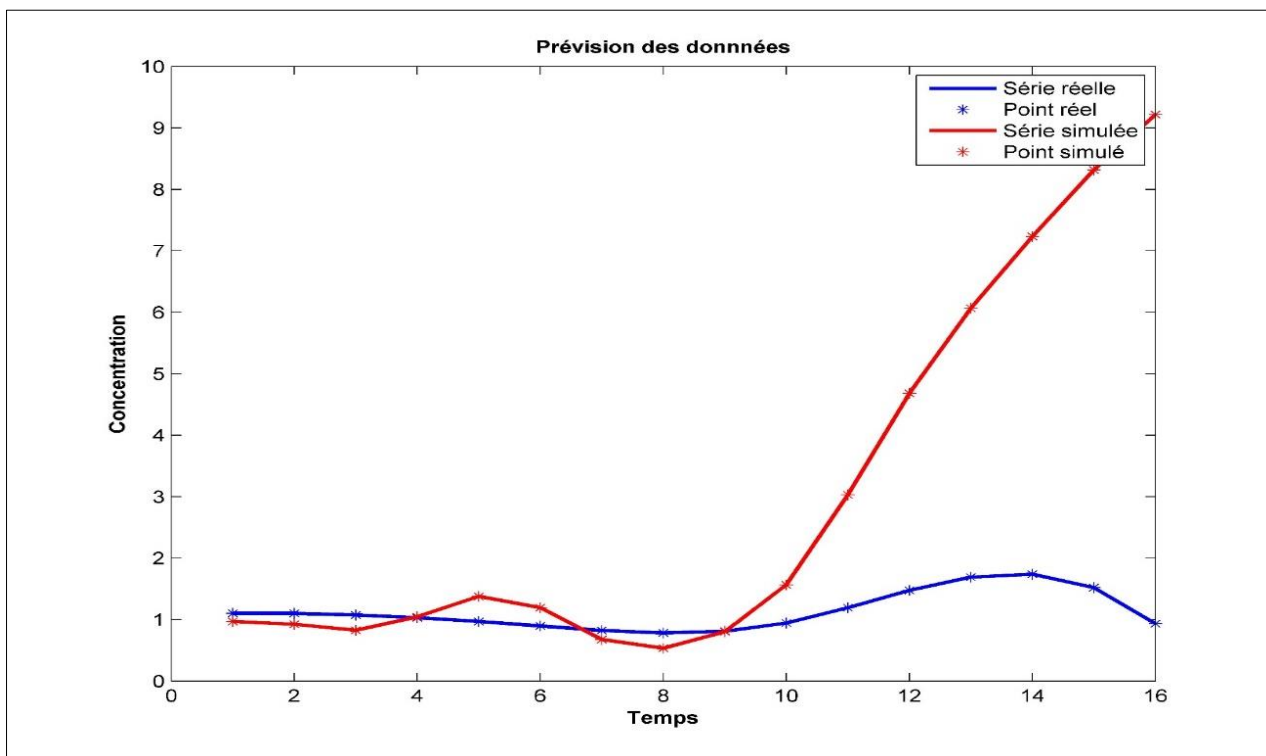


Figure IV.46: Prévision du Césium (Ce) par le modèle ARIMA

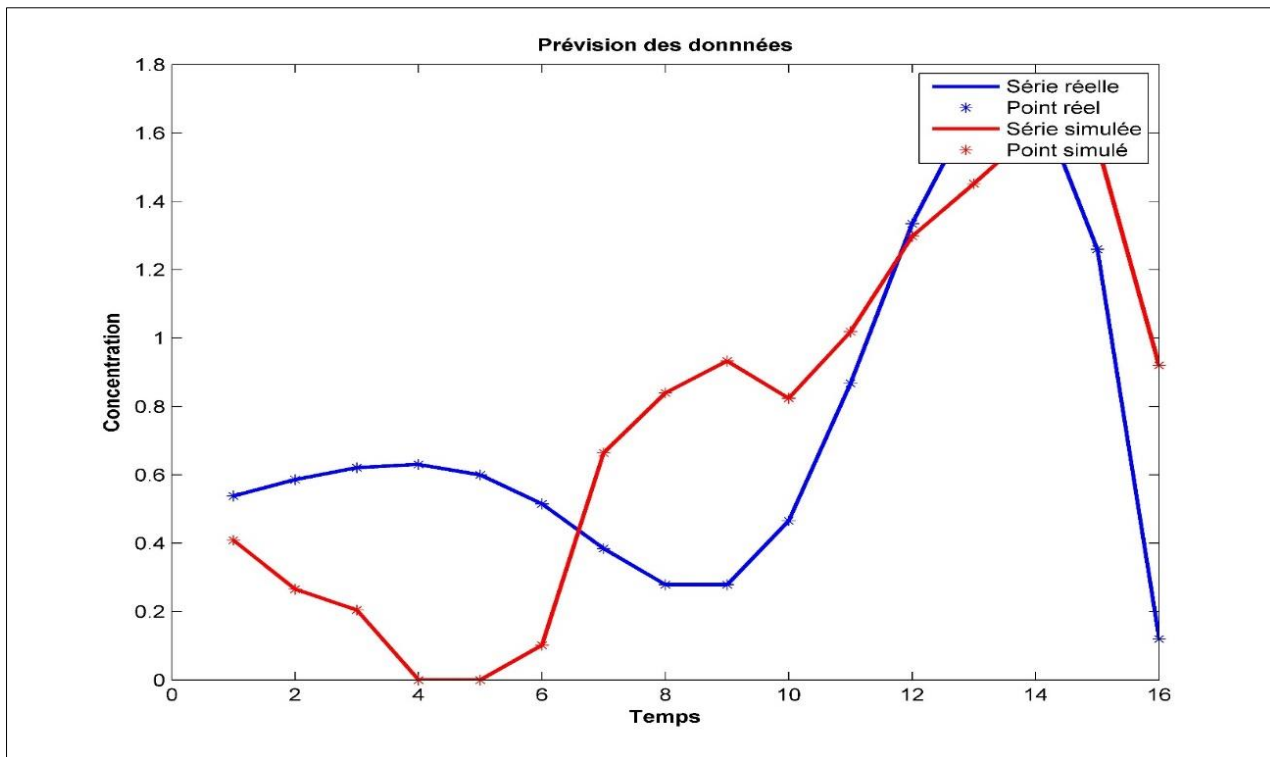


Figure IV.47: Prévision du Chrome (Cr) par le modèle ARIMA

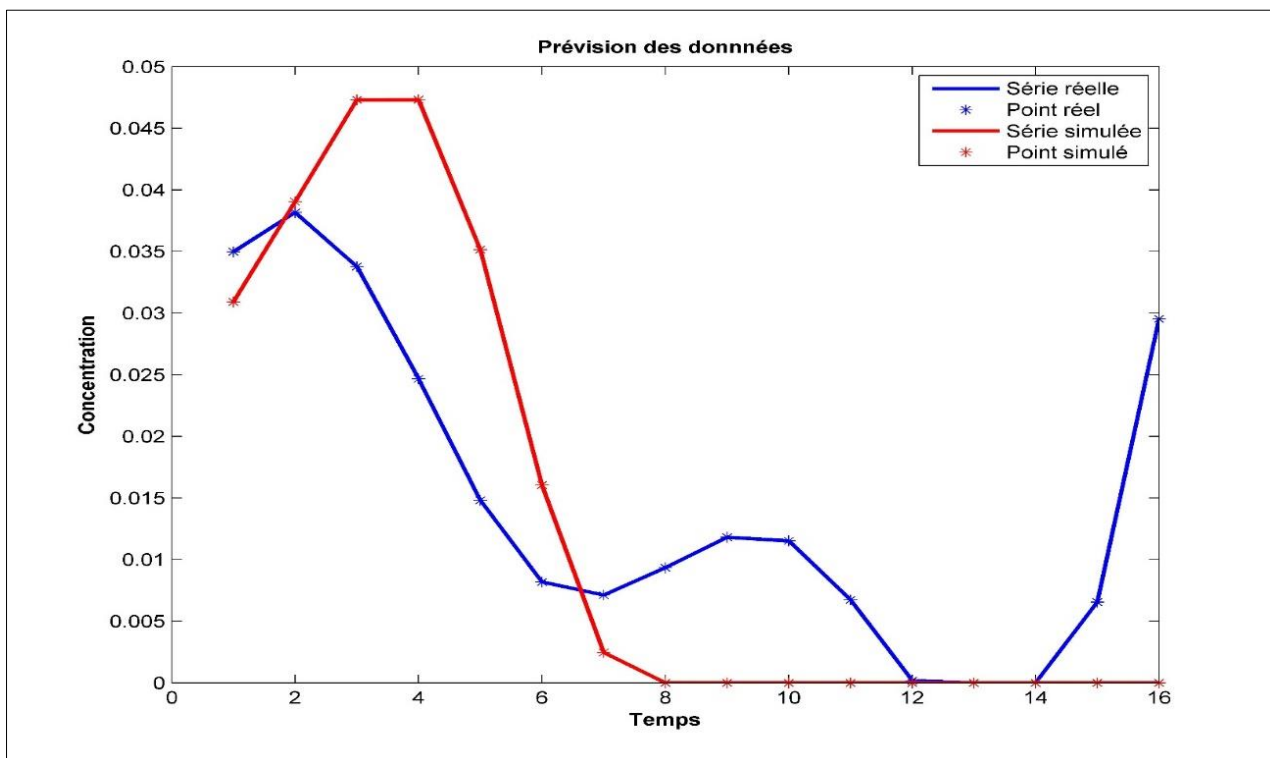


Figure IV.48: Prévision du Hafnium (Hf) par le modèle ARIMA

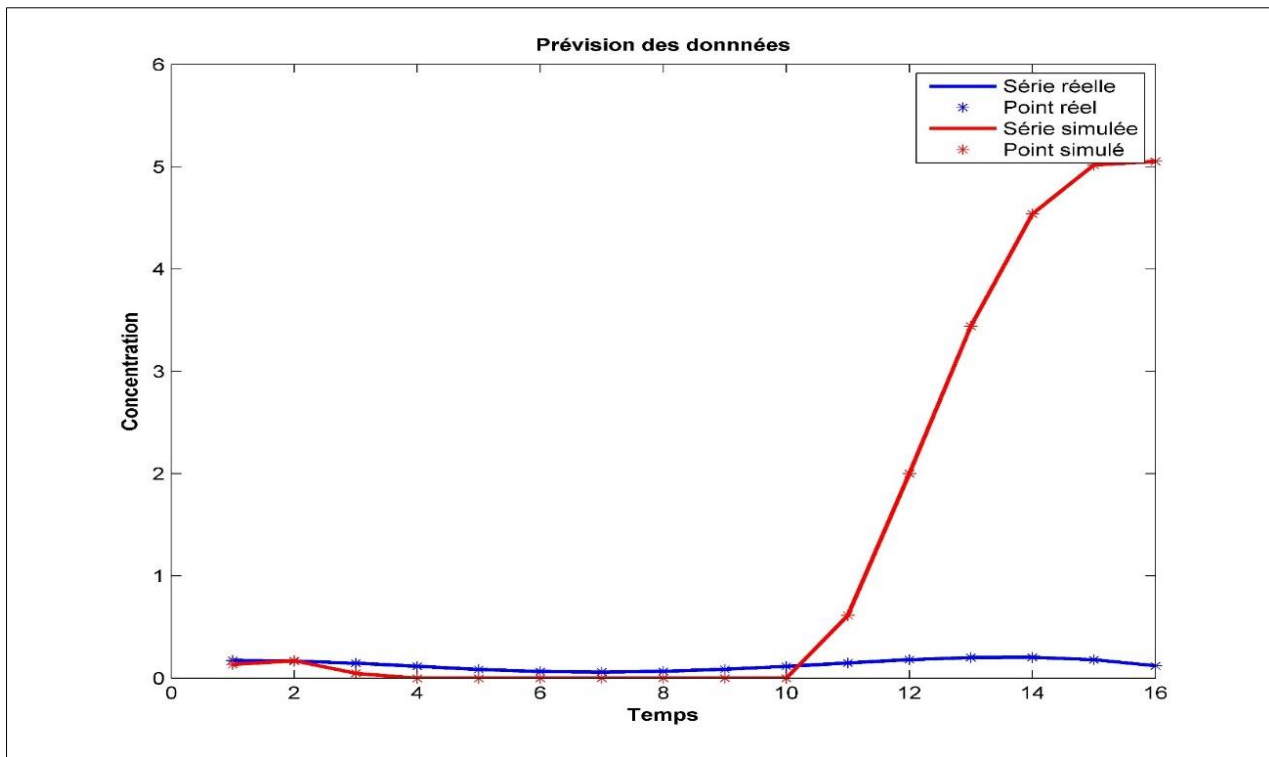


Figure IV.49: Prévision du Lanthane (La) par le modèle ARIMA

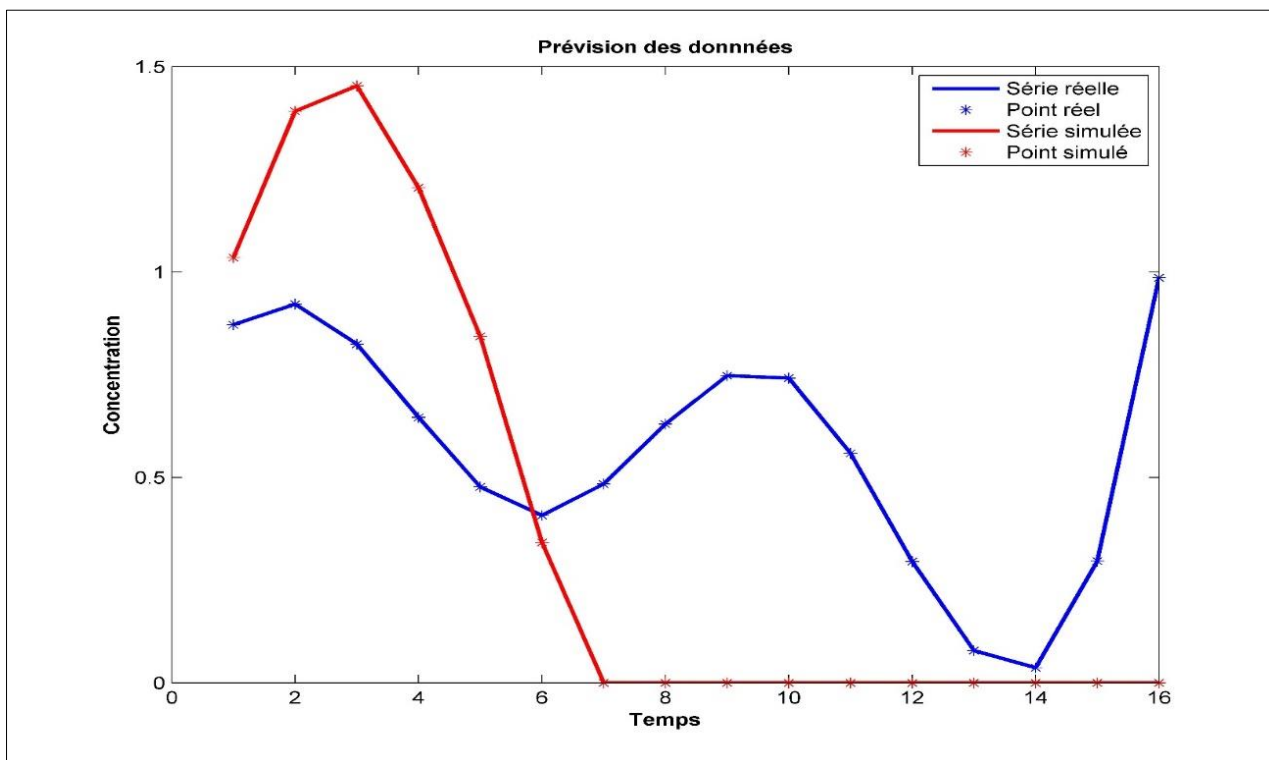


Figure IV.50: Prévision de l'Antimoine (Sb) par le modèle ARIMA

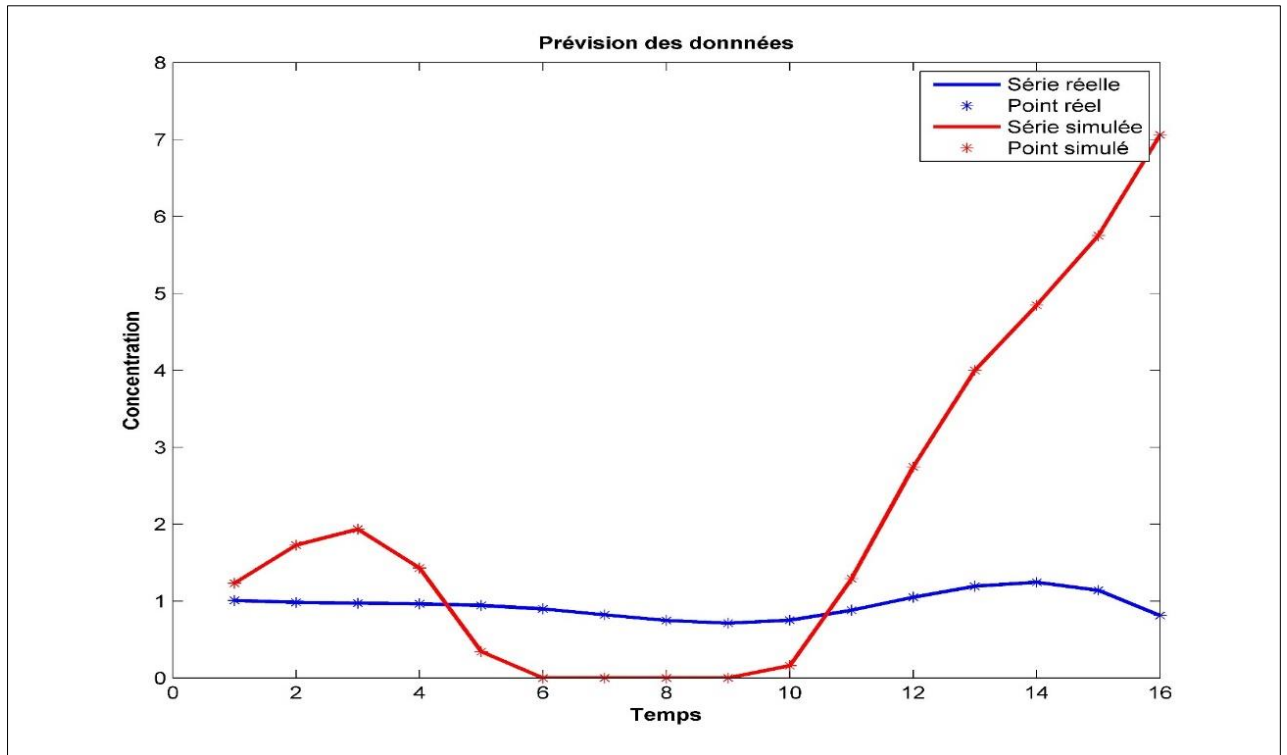


Figure IV.51: Prédiction du Brome (Br) par le modèle ARIMA

A partir de ces résultats on remarque bien que les résultats de la prédiction ne correspondent pas totalement aux concentrations des ET dans l'air, on note une divergence proportionnelle à l'axe temps et une divergence quand il s'agit de jours à forte concentration.

A noter aussi que l'erreur sur les résultats de ces prédictions est considérablement alarmante, une erreur qu'a atteint parfois les 300% (exemple du Lanthane).

IV.7. Prédiction de la pollution atmosphérique par les réseaux de neurones artificiels

Le tableau ci-dessous représente le résultat de la prédiction des TSP et les ET associés par la méthode des réseaux de neurones artificiels, l'algorithme illustré dans le chapitre précédent a été traduit en langage Matlab puis exécuté. La projection de nos modèles élaborés à partir des résultats d'analyse de la pollution atmosphérique au niveau de Draria pendant l'année 2011 sur les quatre premiers mois de l'année 2012 ont donné les résultats représentés dans le tableau ci-dessous:

Tableau IV.15: Tableau récapitulatif de la prévision des TSP et des ET associés par les réseaux de neurones artificiels

Élément	Jour de prévision	Architecture du réseau		Performance				Prévision				
				MSE	MSE			r^2 (%)	RMSE	Erreur (%)		
					Train	Validation	Test			Classe 1	Classe 2	Class 3
TSP	<i>J</i>	[40-2-1]	T-T-P	3,385	3,308	3,387	7,277	97,906	3,460	75,00	91,00	0
	<i>J+1</i>	[40-2-1]	T-T-P	3,453	3,374	3,455	7,423	96,174	3,529	71,250	86,450	0
	<i>J+2</i>	[40-2-1]	T-T-P	3,522	3,442	3,524	7,571	95,365	3,599	67,688	82,128	0
	<i>J+3</i>	[40-2-1]	T-T-P	3,593	3,511	3,595	7,723	93,486	3,671	64,303	78,021	0
	<i>J+4</i>	[35-5-1]	T-T-P	3,664	3,581	3,667	7,877	92,975	3,745	61,088	74,120	0
Sm	<i>J</i>	[10-2-1]	L-L-P	0,037	0,022	0,037	0,017	97,614	0,005	78,14	94,54	1,12
	<i>J+1</i>	[5-2-1]	T-P	0,038	0,022	0,038	0,017	97,228	0,005	74,235	89,809	1,064
	<i>J+2</i>	[5-2-1]	T-P	0,038	0,023	0,038	0,017	96,076	0,005	70,523	85,318	1,011
	<i>J+3</i>	[5-1]	T-P	0,039	0,023	0,039	0,018	94,412	0,005	66,997	81,052	0,960
	<i>J+4</i>	[8-2-1]	T-P	0,040	0,024	0,040	0,018	93,486	0,005	63,647	77,000	0,912
Br	<i>J</i>	[10-2-1]	T-P	2,711	1,796	2,757	2,429	98,331	1,407	79,51	92,90	1,10
	<i>J+1</i>	[8-2-1]	T-P	2,765	1,832	2,812	2,477	97,732	1,435	75,533	88,251	1,045
	<i>J+2</i>	[5-2-1]	T-P	2,821	1,869	2,868	2,527	95,862	1,464	71,756	83,839	0,993
	<i>J+3</i>	[8-2-1]	T-P	2,877	1,906	2,925	2,577	94,605	1,493	68,168	79,647	0,943
	<i>J+4</i>	[2-10-1]	T-P	2,934	1,944	2,984	2,629	94,172	1,523	64,760	75,665	0,896
Yb	<i>J</i>	[2-5-1]	L-L-P	0,017	0,005	0,017	0,015	97,030	0,003	86,89	96,17	1,09
	<i>J+1</i>	[2-10-1]	L-L-P	0,017	0,005	0,017	0,015	97,402	0,003	82,541	91,366	1,038
	<i>J+2</i>	[10-2-1]	L-L-P	0,017	0,005	0,017	0,015	97,302	0,003	78,414	86,798	0,986
	<i>J+3</i>	[10-2-1]	L-L-P	0,018	0,005	0,018	0,015	96,530	0,003	74,493	82,458	0,937
	<i>J+4</i>	[10-2-1]	L-L-P	0,018	0,005	0,018	0,016	96,527	0,003	70,769	78,335	0,890

Sb	<i>J</i>	[5-2-1]	T-P	0,093	0,664	0,930	0,754	98,626	0,084	75,41	94,81	0,82
	<i>J+1</i>	[5-2-1]	T-P	0,095	0,677	0,948	0,769	97,689	0,086	71,639	90,068	0,779
	<i>J+2</i>	[5-2-1]	T-P	0,097	0,690	0,967	0,785	96,076	0,088	68,057	85,565	0,740
	<i>J+3</i>	[8-2-1]	T-P	0,099	0,704	0,987	0,800	95,661	0,089	64,655	81,287	0,703
	<i>J+4</i>	[2-10-1]	T-P	0,101	0,718	1,007	0,816	94,188	0,091	61,422	77,222	0,668
Sc	<i>J</i>	[2-5-1]	L-L-P	0,168	0,139	0,168	0,178	98,154	0,012	80,05	94,81	1,09
	<i>J+1</i>	[2-10-1]	L-L-P	0,171	0,141	0,171	0,182	97,627	0,012	76,052	90,068	1,038
	<i>J+2</i>	[10-2-1]	L-L-P	0,175	0,144	0,175	0,185	97,231	0,013	72,249	85,565	0,986
	<i>J+3</i>	[10-2-1]	L-L-P	0,178	0,147	0,178	0,189	95,942	0,013	68,637	81,287	0,937
	<i>J+4</i>	[10-2-1]	L-L-P	0,182	0,150	0,182	0,193	95,244	0,013	65,205	77,222	0,890
Hf	<i>J</i>	[2-5-1]	L-L-P	0,258	0,024	0,027	0,029	97,383	0,050	78,42	93,72	1,64
	<i>J+1</i>	[2-10-1]	L-L-P	0,264	0,025	0,028	0,030	97,638	0,050	74,495	89,030	1,557
	<i>J+2</i>	[10-2-1]	L-L-P	0,269	0,025	0,028	0,031	96,920	0,051	70,770	84,579	1,480
	<i>J+3</i>	[10-2-1]	L-L-P	0,274	0,026	0,029	0,031	95,391	0,053	67,231	80,350	1,406
	<i>J+4</i>	[10-2-1]	L-L-P	0,280	0,026	0,029	0,032	94,163	0,054	63,870	76,332	1,335
Na	<i>J</i>	[5-2-1]	T-P	464,327	658,140	607,970	318,100	97,992	22,470	87,98	96,99	1,37
	<i>J+1</i>	[5-2-1]	T-P	473,613	671,303	620,129	324,462	96,336	22,919	83,579	92,145	1,298
	<i>J+2</i>	[5-2-1]	T-P	483,086	684,729	632,532	330,951	95,634	23,378	79,400	87,538	1,233
	<i>J+3</i>	[8-2-1]	T-P	492,747	698,424	645,182	337,570	93,844	23,845	75,430	83,161	1,171
	<i>J+4</i>	[10-2-1]	T-P	502,602	712,392	658,086	344,321	92,473	24,322	71,659	79,003	1,113
Ce	<i>J</i>	[2-5-1]	L-L-P	0,305	0,384	0,363	0,171	97,907	0,179	84,70	94,26	1,37
	<i>J+1</i>	[2-10-1]	L-L-P	0,311	0,392	0,370	0,175	96,038	0,183	80,464	89,549	1,298
	<i>J+2</i>	[10-2-1]	L-L-P	0,317	0,400	0,377	0,178	94,147	0,186	76,441	85,072	1,233
	<i>J+3</i>	[10-2-1]	L-L-P	0,323	0,408	0,385	0,182	93,658	0,190	72,619	80,818	1,171
	<i>J+4</i>	[10-2-1]	L-L-P	0,330	0,416	0,393	0,185	92,642	0,194	68,988	76,777	1,113

Cr	<i>J</i>	[2-5-1]	L-L-P	0,355	0,485	0,380	0,257	97,151	0,185	82,79	94,81	1,37
	<i>J+1</i>	[2-10-1]	L-L-P	0,362	0,495	0,388	0,263	97,497	0,189	78,648	90,068	1,298
	<i>J+2</i>	[10-2-1]	L-L-P	0,369	0,505	0,396	0,268	96,173	0,193	74,715	85,565	1,233
	<i>J+3</i>	[10-2-1]	L-L-P	0,376	0,515	0,404	0,273	94,338	0,197	70,979	81,287	1,171
	<i>J+4</i>	[10-2-1]	L-L-P	0,384	0,526	0,412	0,279	93,240	0,201	67,430	77,222	1,113
La	<i>J</i>	[8-2-1]	T-P	0,263	0,150	0,264	0,306	54,929	0,137	87,16	95,63	1,91
	<i>J+1</i>	[5-2-1]	T-P	0,268	0,153	0,269	0,312	54,240	0,140	82,801	90,847	1,817
	<i>J+2</i>	[5-1]	T-P	0,273	0,156	0,275	0,319	97,350	0,142	78,661	86,305	1,726
	<i>J+3</i>	[8-2-1]	T-P	0,279	0,159	0,280	0,325	41,860	0,145	74,727	81,989	1,640
	<i>J+4</i>	[2-10-1]	T-P	0,284	0,162	0,286	0,331	42,700	0,148	70,991	77,890	1,558

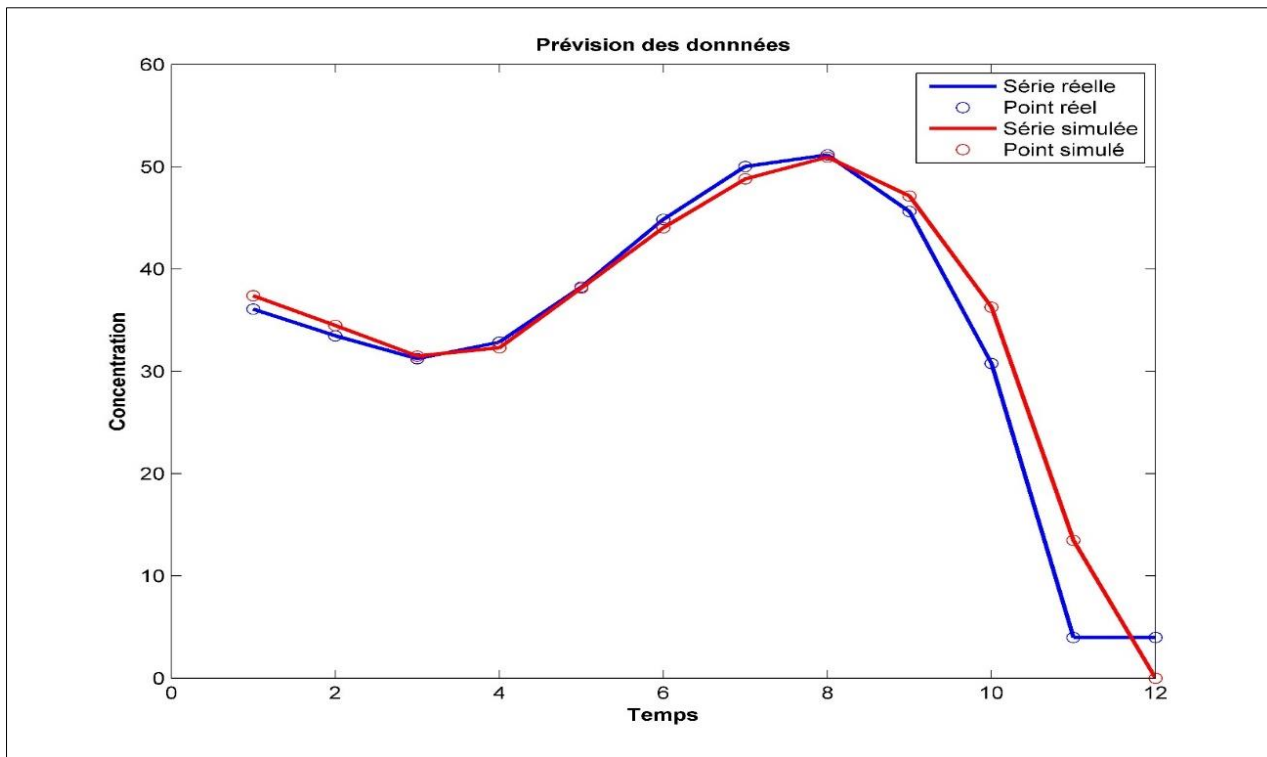


Figure IV.52:Prévision des TSP par les réseaux de neurones artificiels

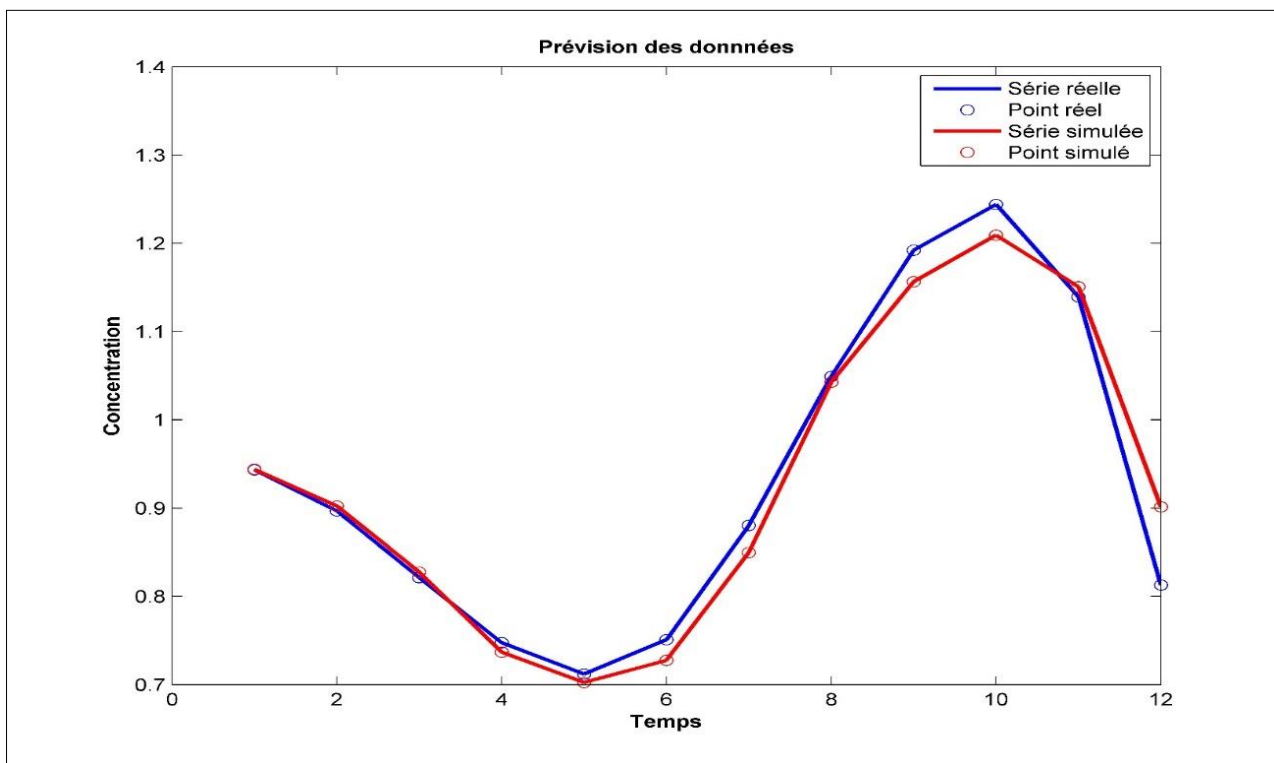


Figure IV.53: Prévision du Brome (Br) par les réseaux de neurones artificiels

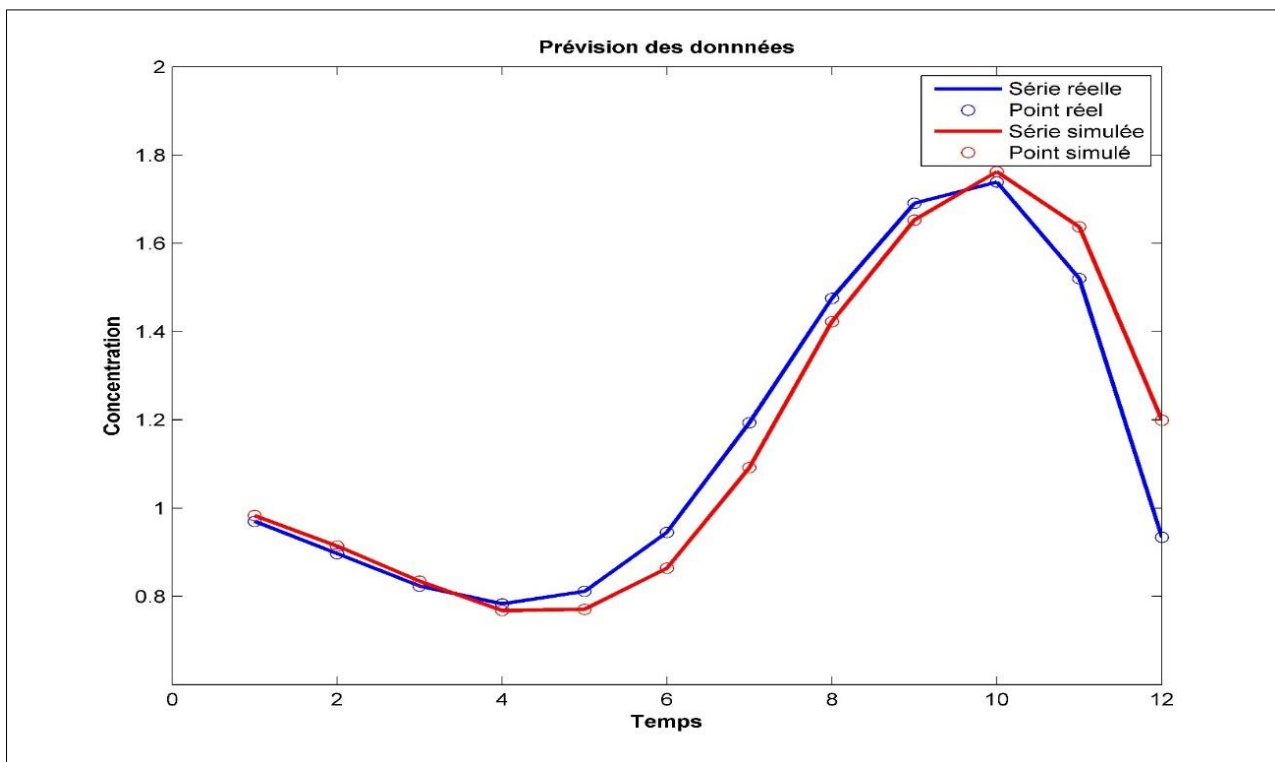


Figure IV.54: Prédiction du Césium (Ce) par les réseaux de neurones artificiels

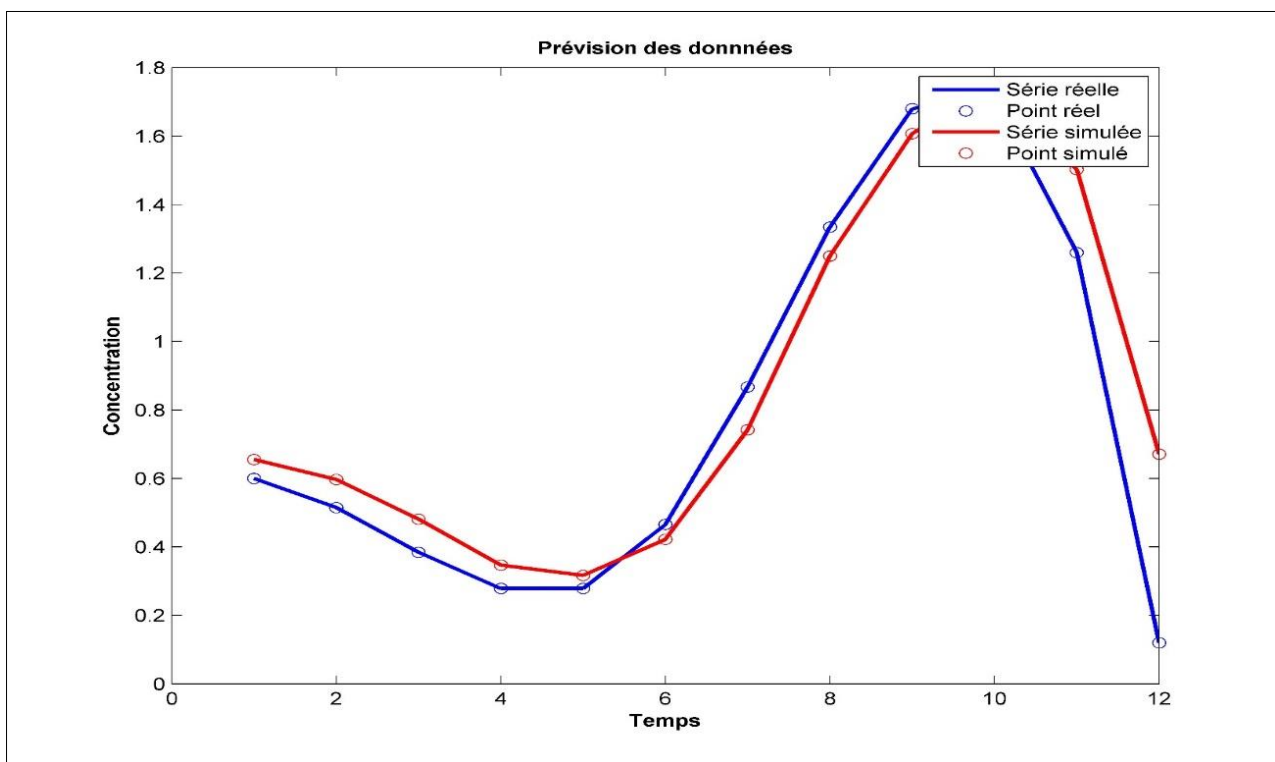


Figure IV.55: Prédiction du Chrome (Cr) par les réseaux de neurones artificiels

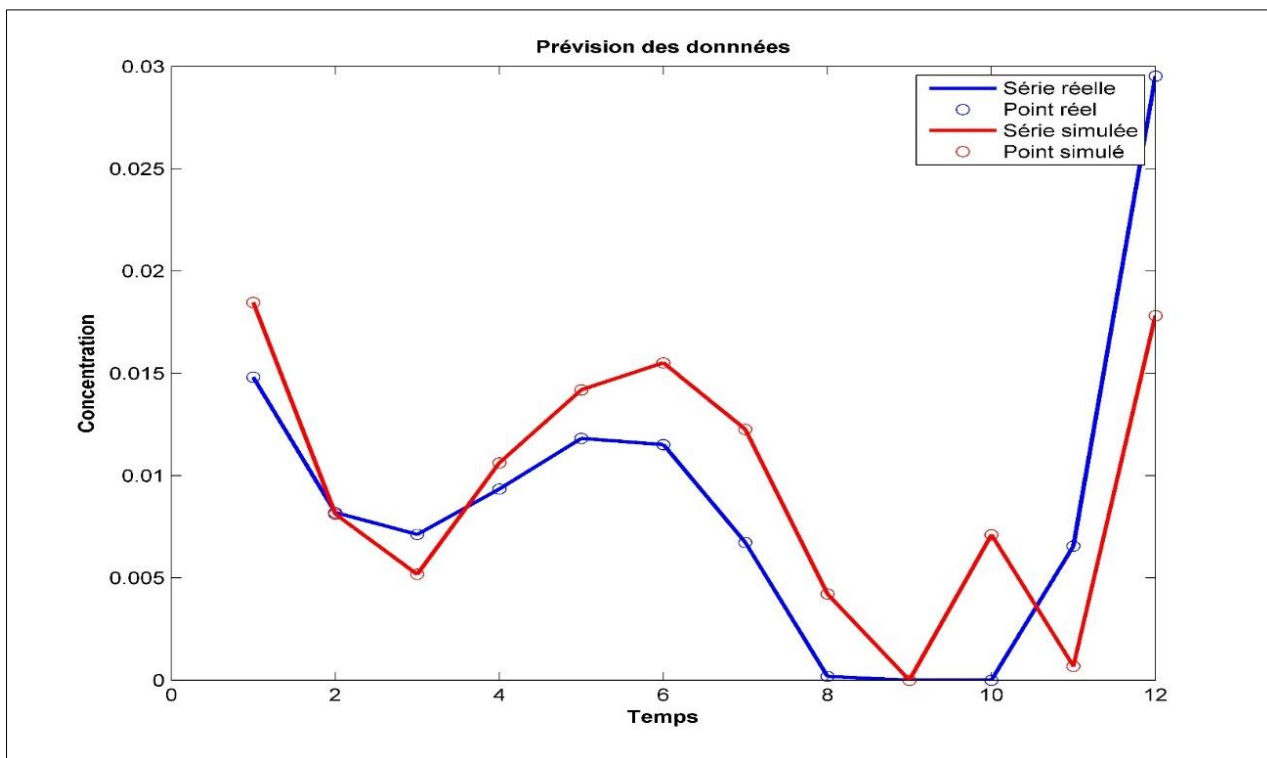


Figure IV.56: Prédiction du Hafnium (Hf) par les réseaux de neurones artificiels

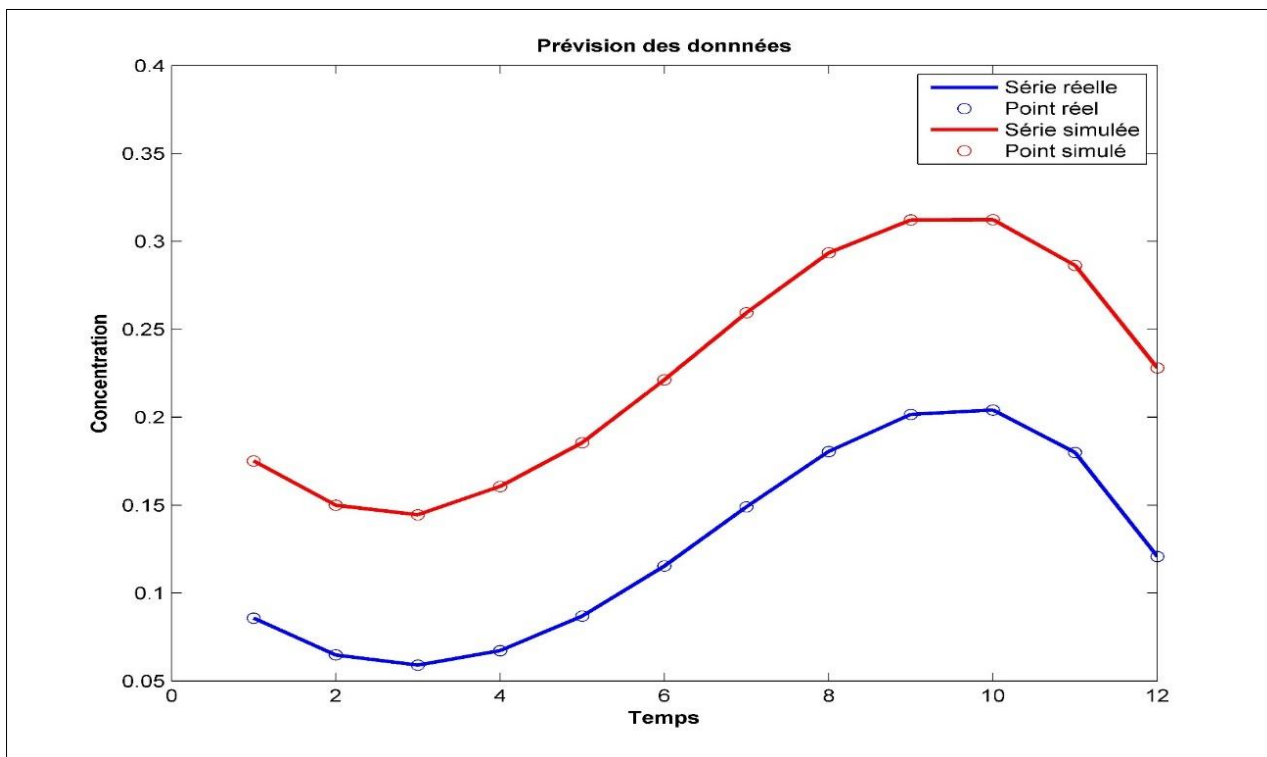


Figure IV. 57: Prédiction du Lanthane (La) par les réseaux de neurones artificiels

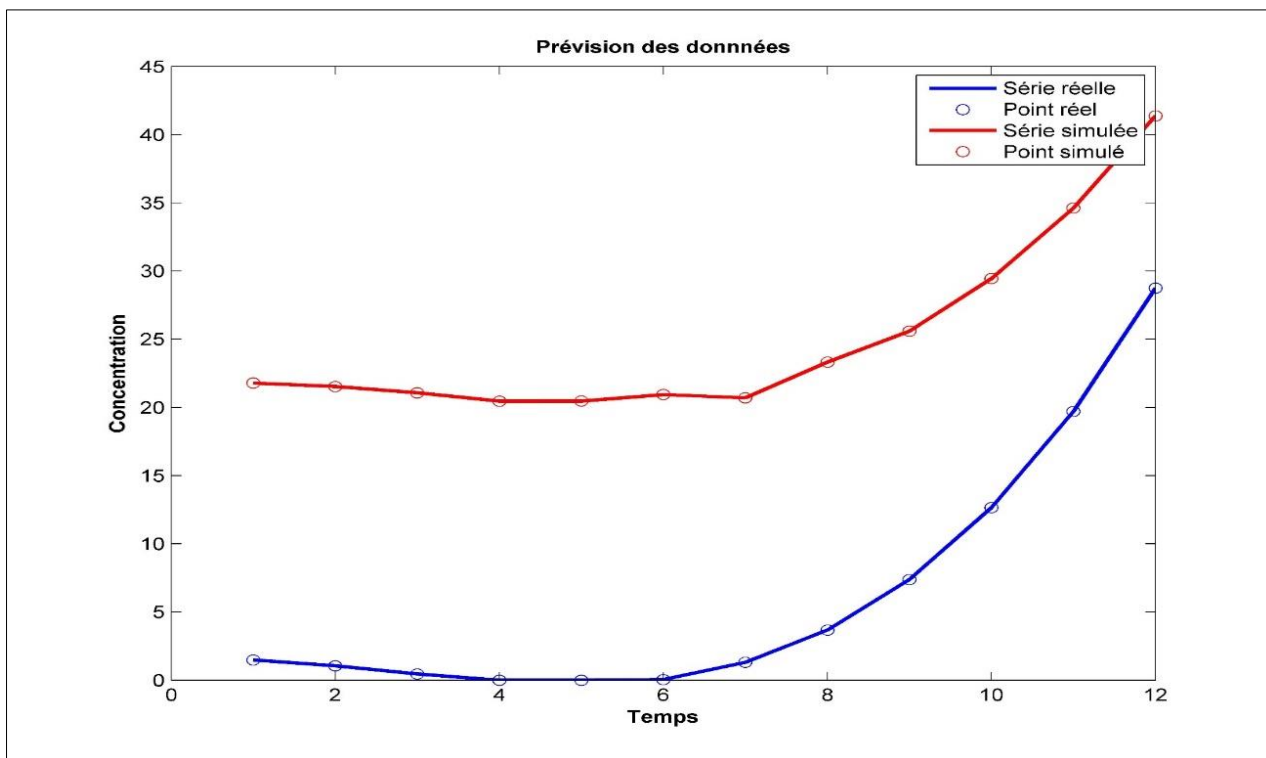


Figure IV.58: Prévion du Sodium (Na) par les réseaux de neurones artificiels

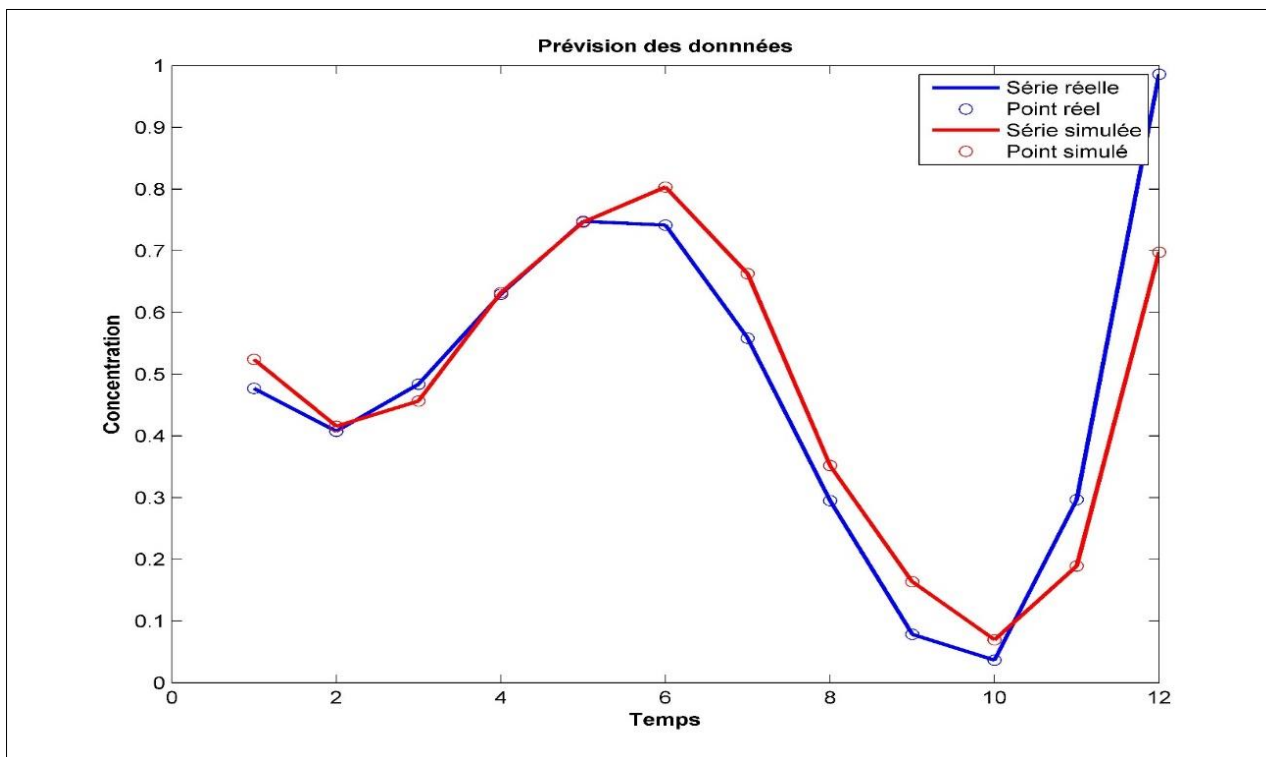


Figure IV.59: Prévion de l'Antimoine (Sb) par les réseaux de neurones artificiels

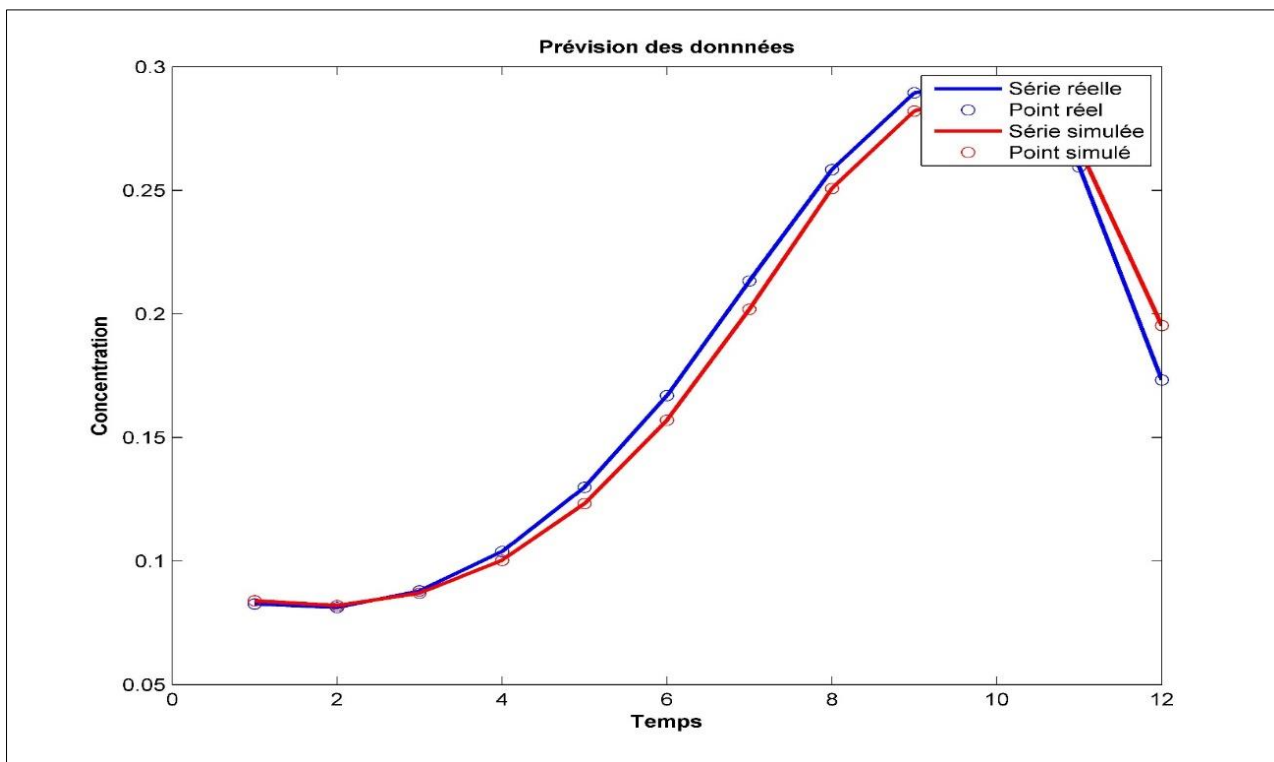


FIGURE IV.60: Prévion du (Sc) par les réseaux de neurones artificiels

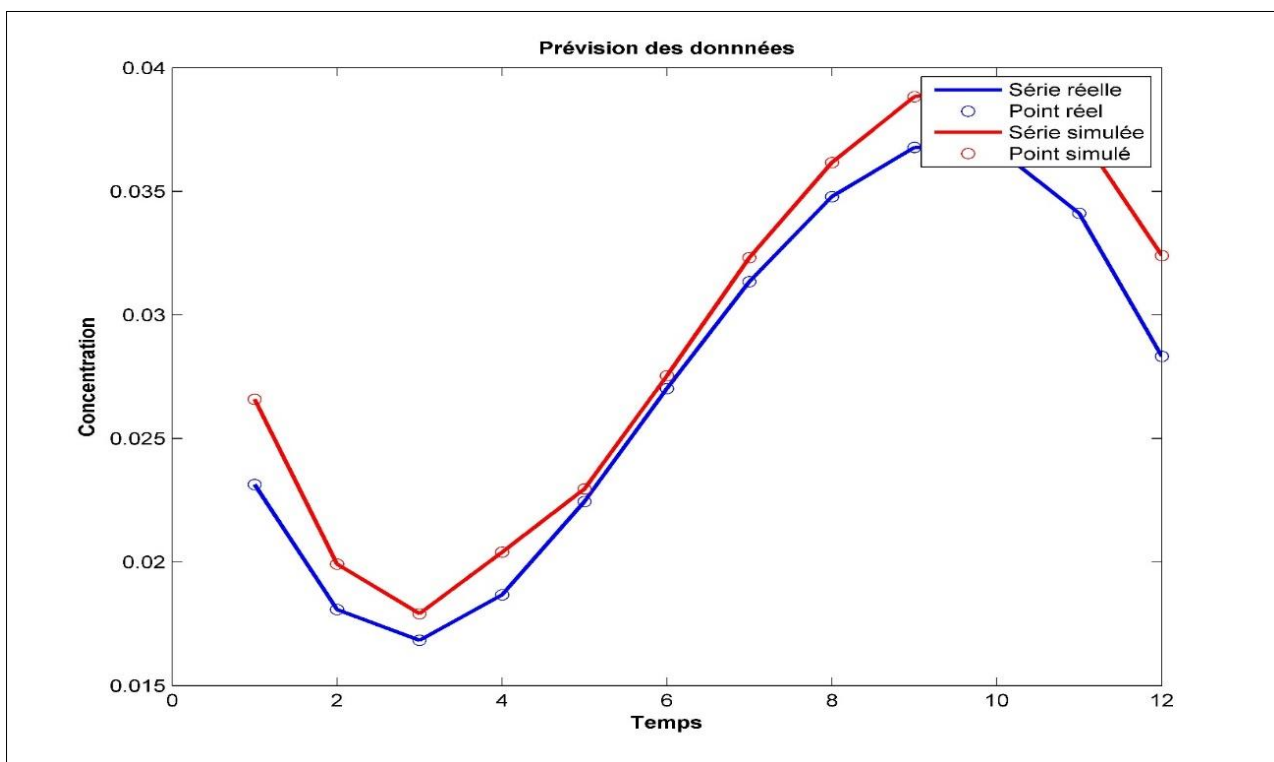


Figure IV.61: Prévion du Samarium (Sm) par les réseaux de neurones artificiels

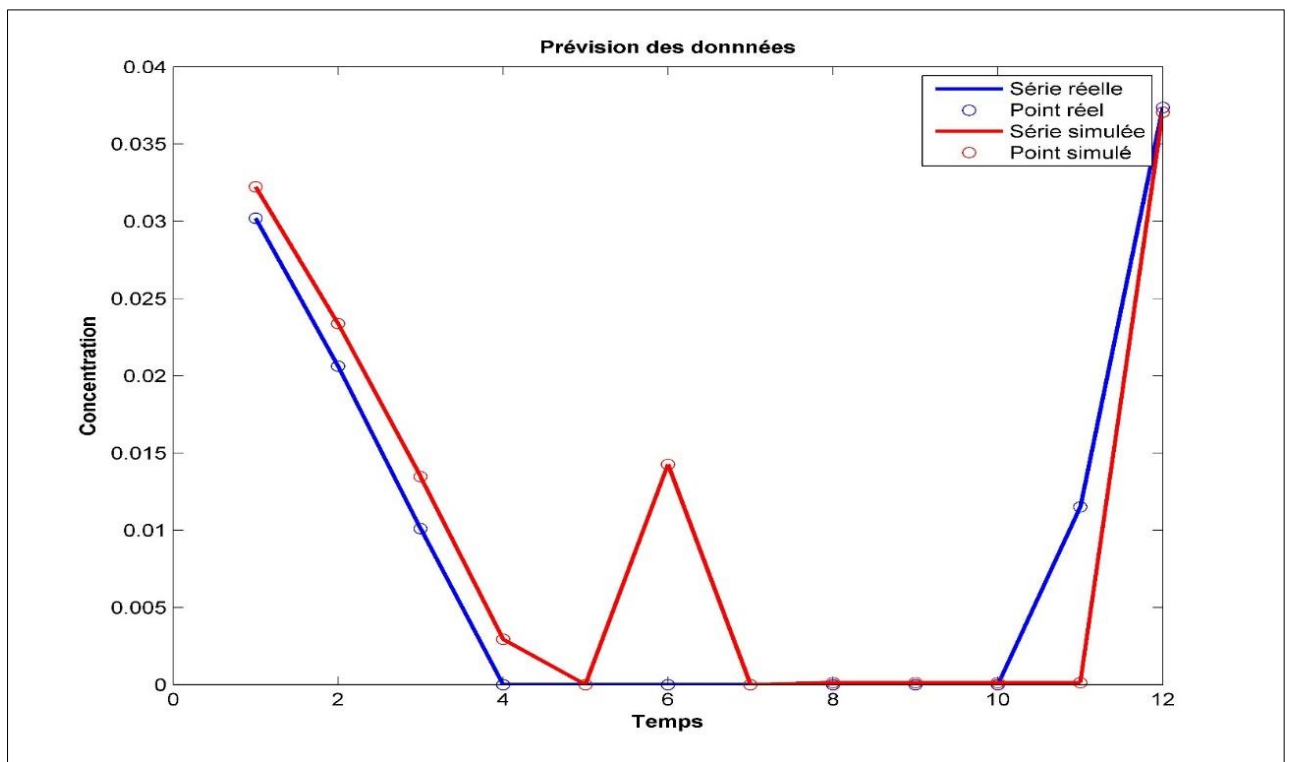


Figure IV.62: *Prévision d'Ytterbium (Yb) par les réseaux de neurones artificiels*

A la lumière de ces résultats et ces graphes on remarque que la méthode de la prévision par les réseaux de neurones artificiels a donné les meilleurs résultats en termes de précision et qualité de prédiction. Les lacunes trouvés en utilisant le modèle ARIMA ont été surmontées et comblées par la méthode des RNA.

Les résultats de la prévision des TSP et des ET associés à ces TSP ont été très cohérents à comparer avec les concentrations réelles, ce qui confirme que nos modèles élaborés ainsi que nos techniques de prévision ont été correctes.

Une seule prévision était hors nos attente, la prévision de l'Ytterbium (Yb) a connu quelques divergences ce qui peut être expliqué par le sur-apprentissage du réseau de neurones de cet élément ou bien par le mauvais choix de l'architecture du réseau.

IV.8. Comparaison entre les deux méthodes de prévision

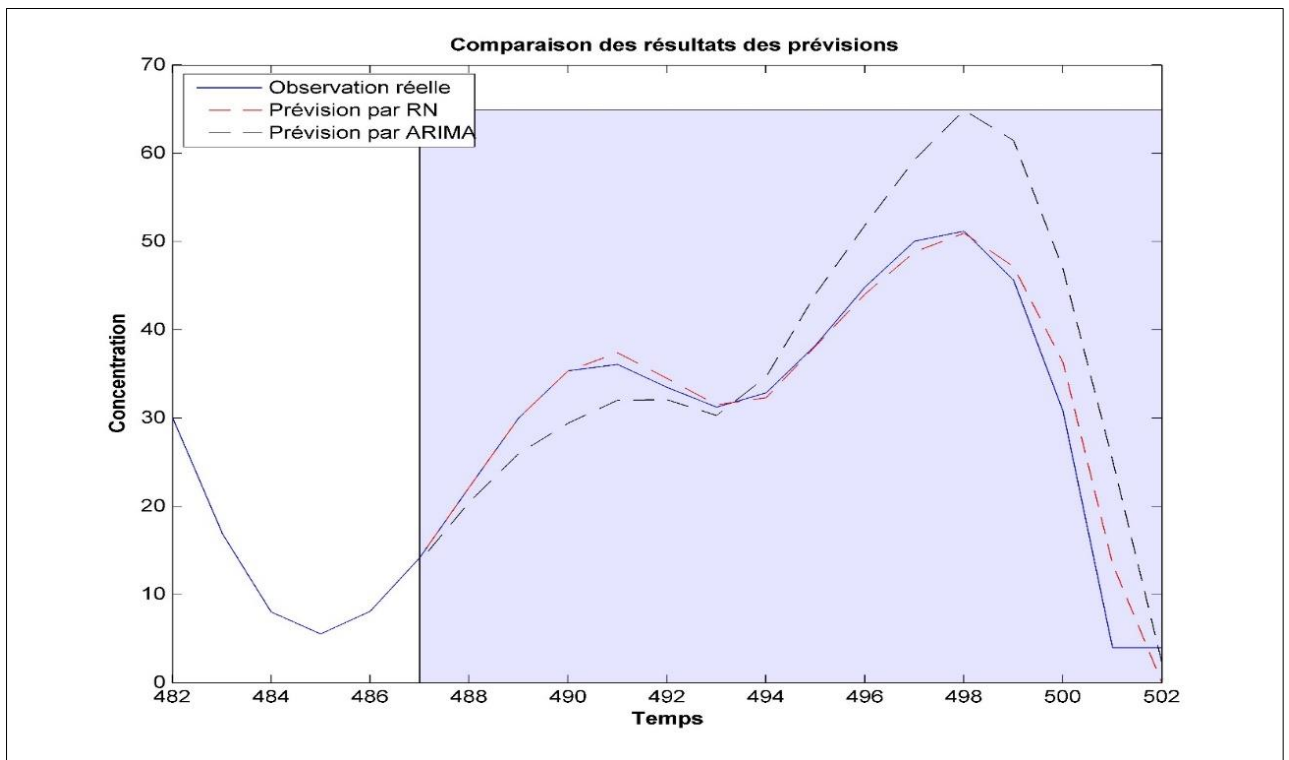


Figure IV.63: Comparaison de la prévision des TSP par les deux méthodes de prévision

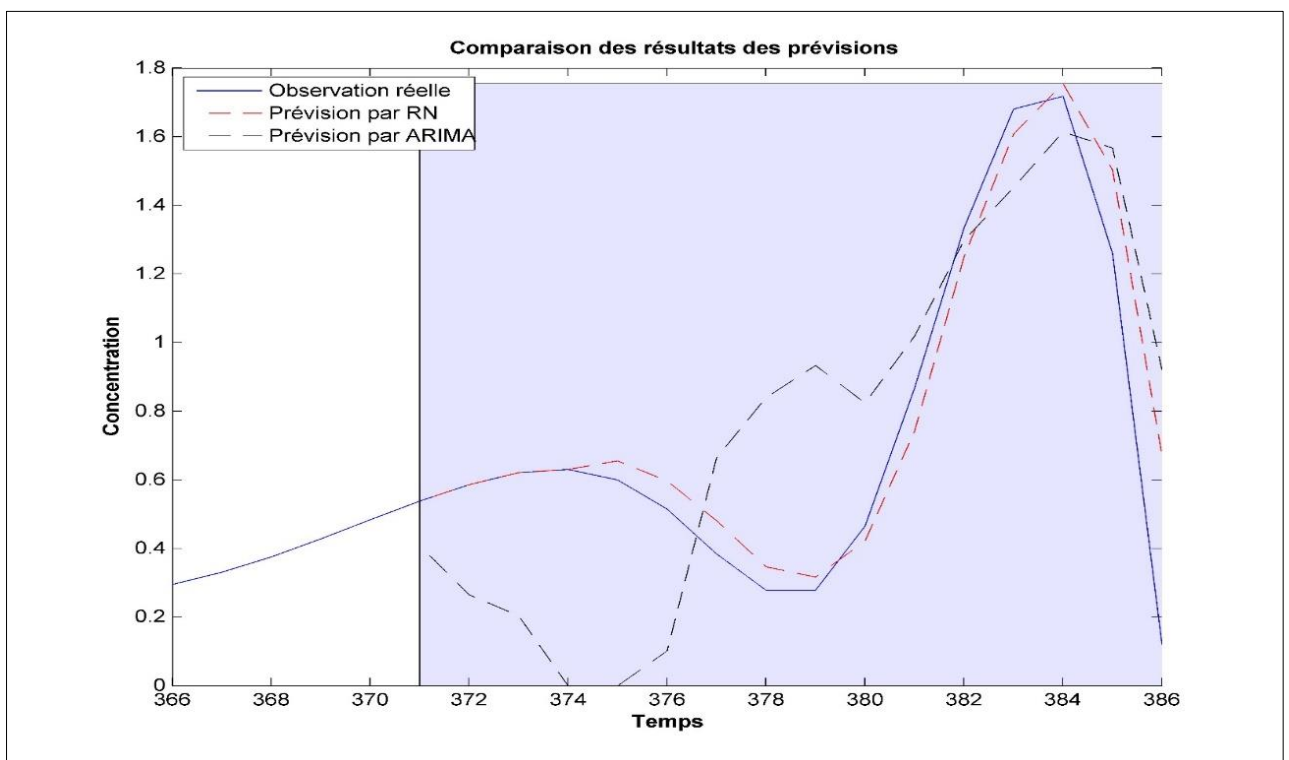


Figure IV.64: Comparaison de la prévision du Chrome (Cr) par les deux méthodes de prévision

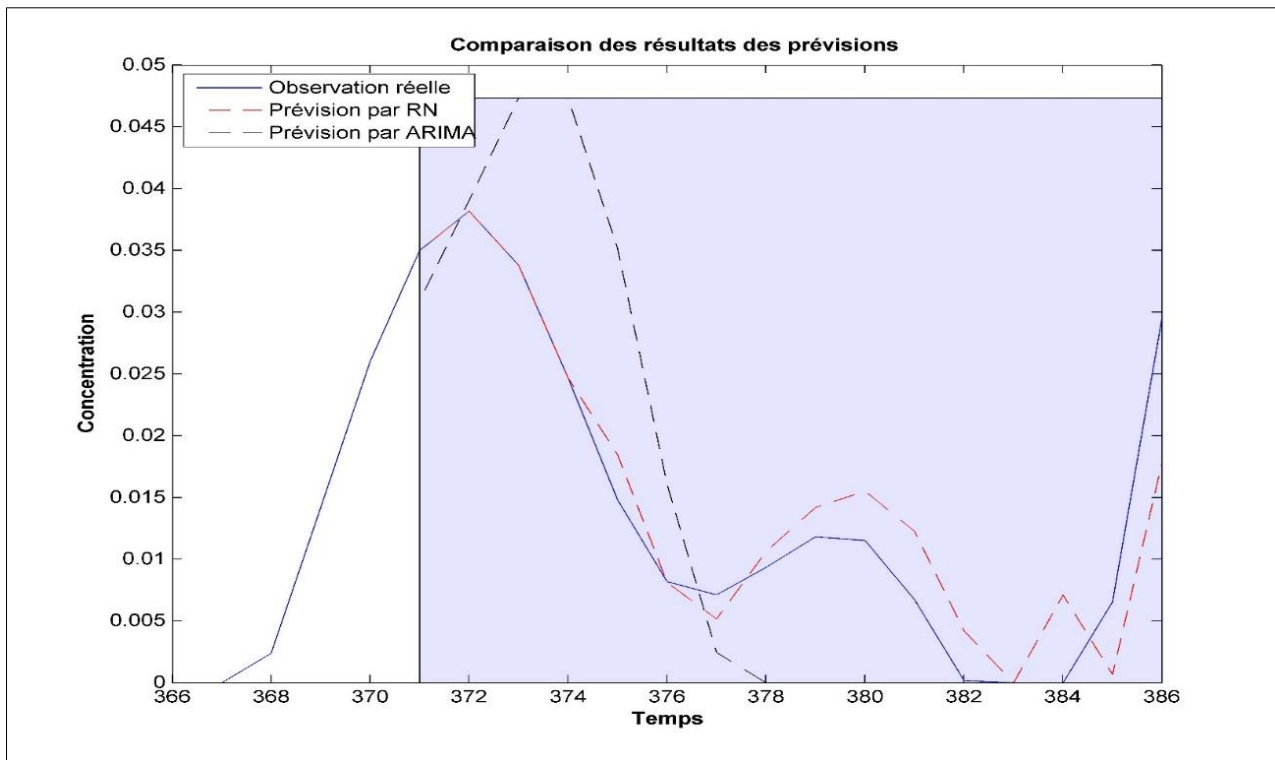


Figure IV.65: Comparaison de la prévision du Hafnium (Hf) par les deux méthodes de prévision

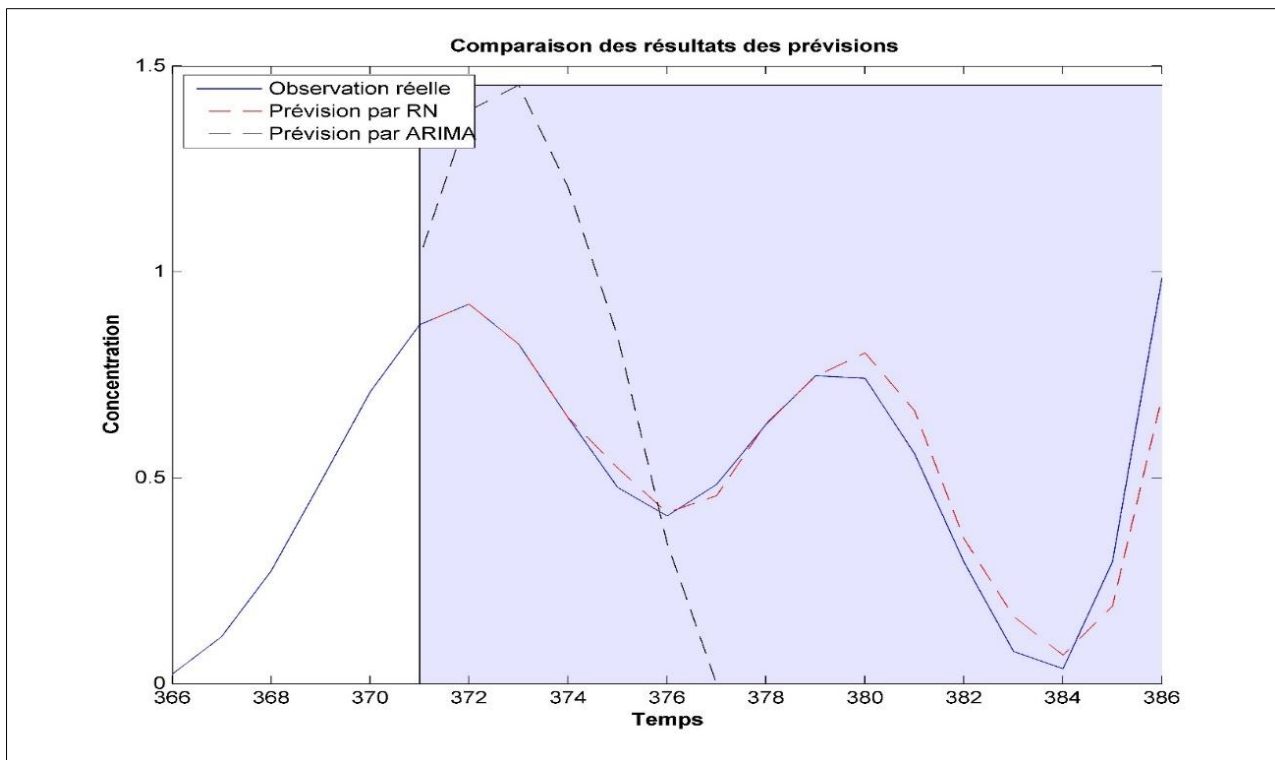


Figure IV.66: Comparaison de la prévision de l'Antimoine (Sb) par les deux méthodes de prévision

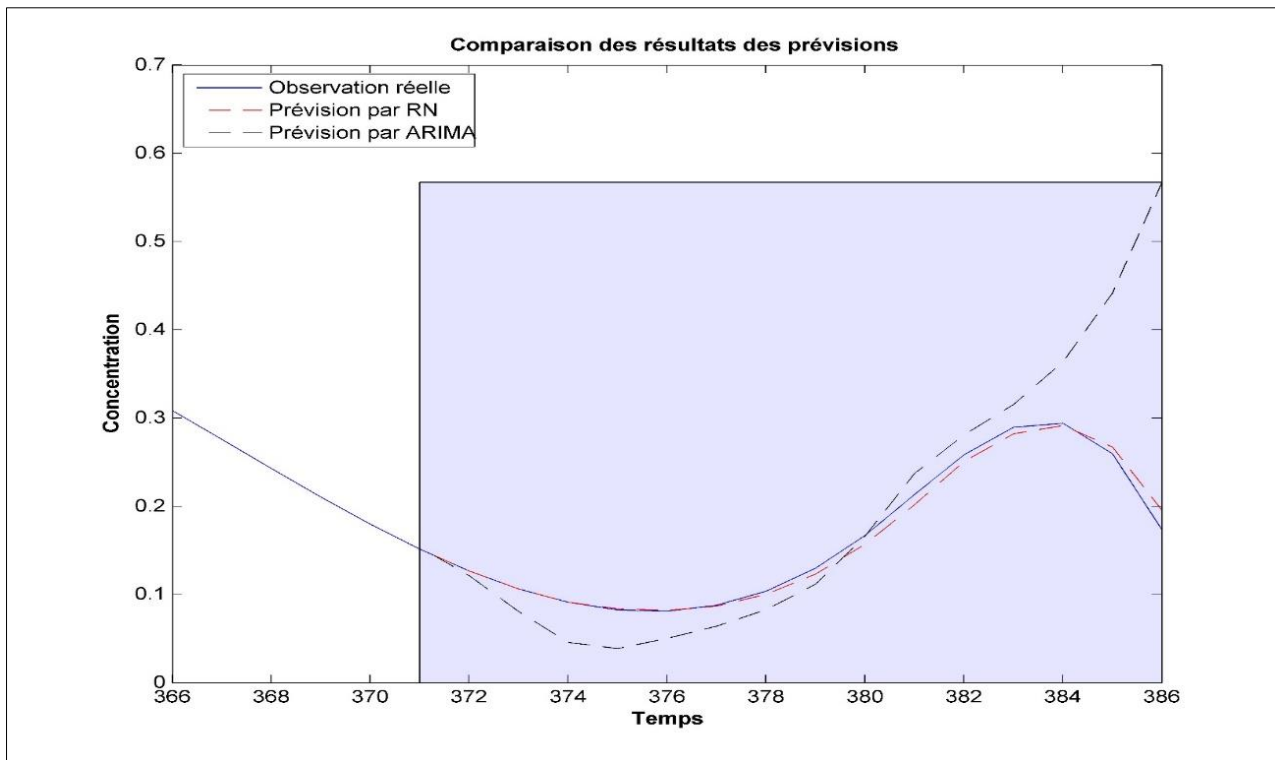


Figure IV.67: Comparaison de la prévision du Scandium (Sc) par les deux méthodes de prévision

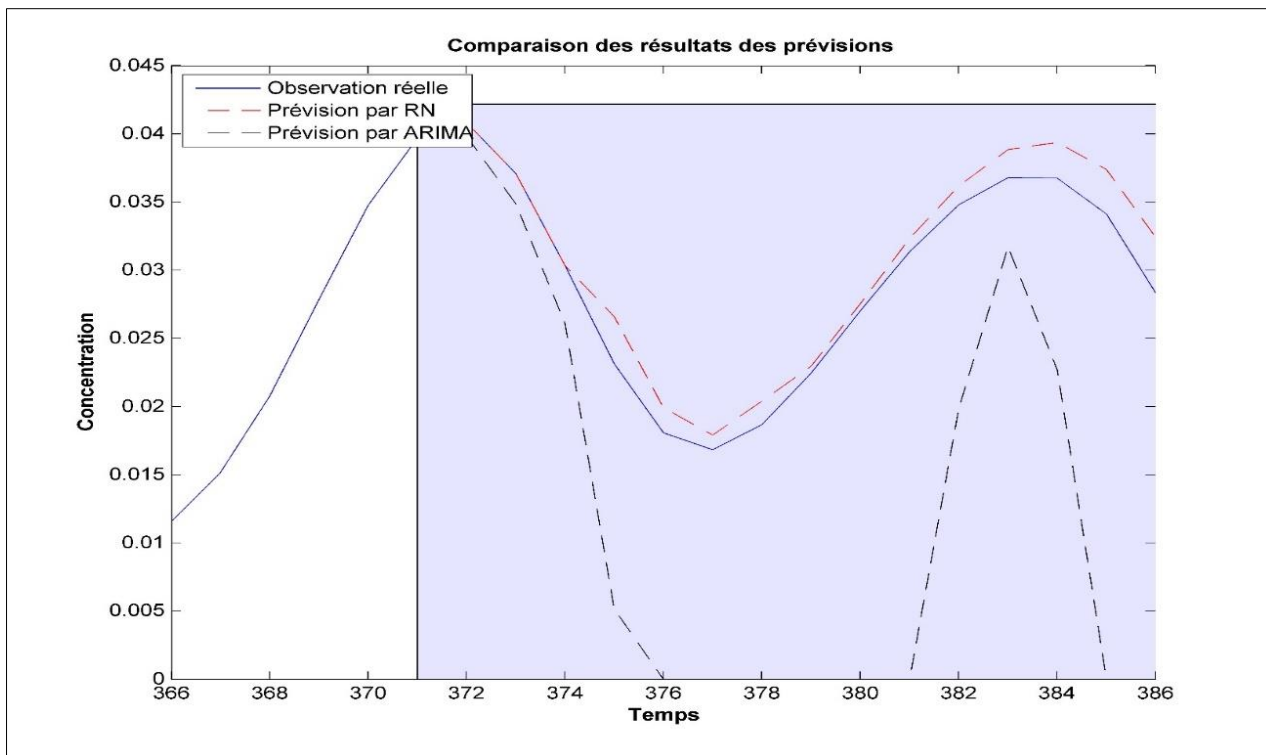


Figure IV.68: Comparaison de la prévision du Samarium (Sm) par les deux méthodes de prévision

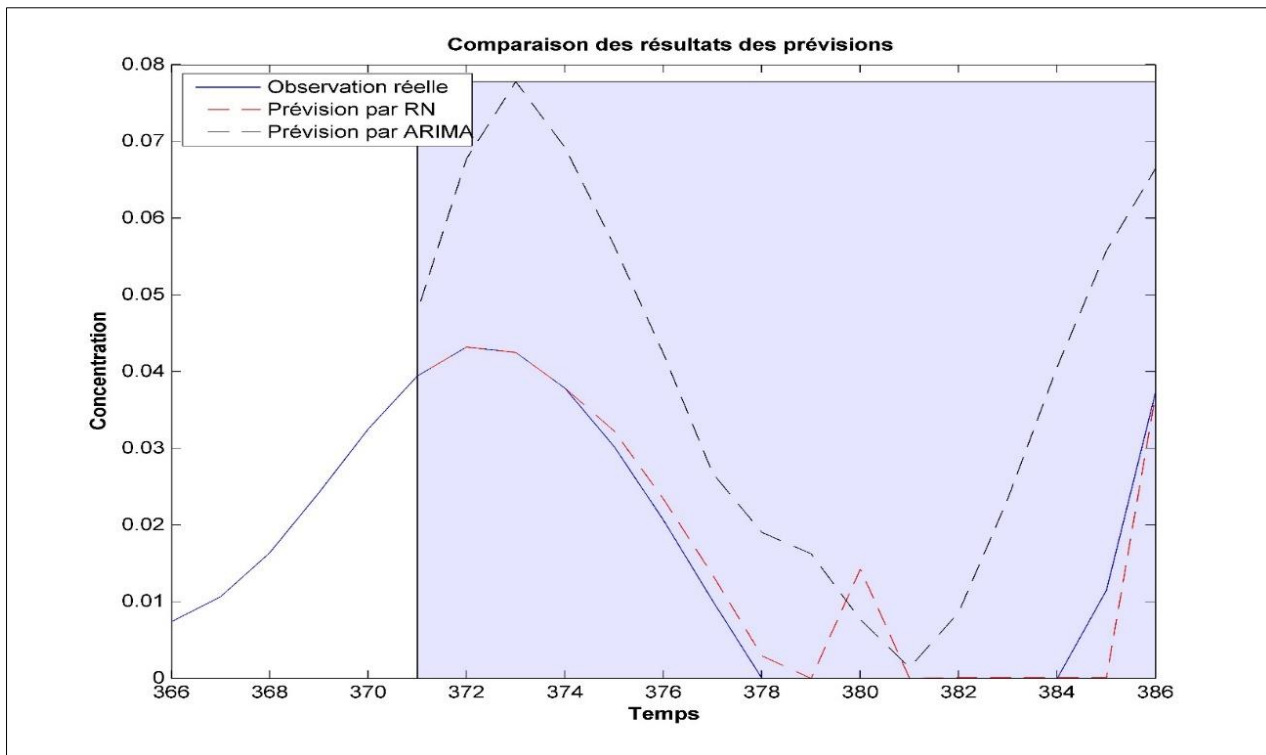


Figure IV.69: Comparaison de la prévision d'Ytterbium (Yb) par les deux méthodes de prévision

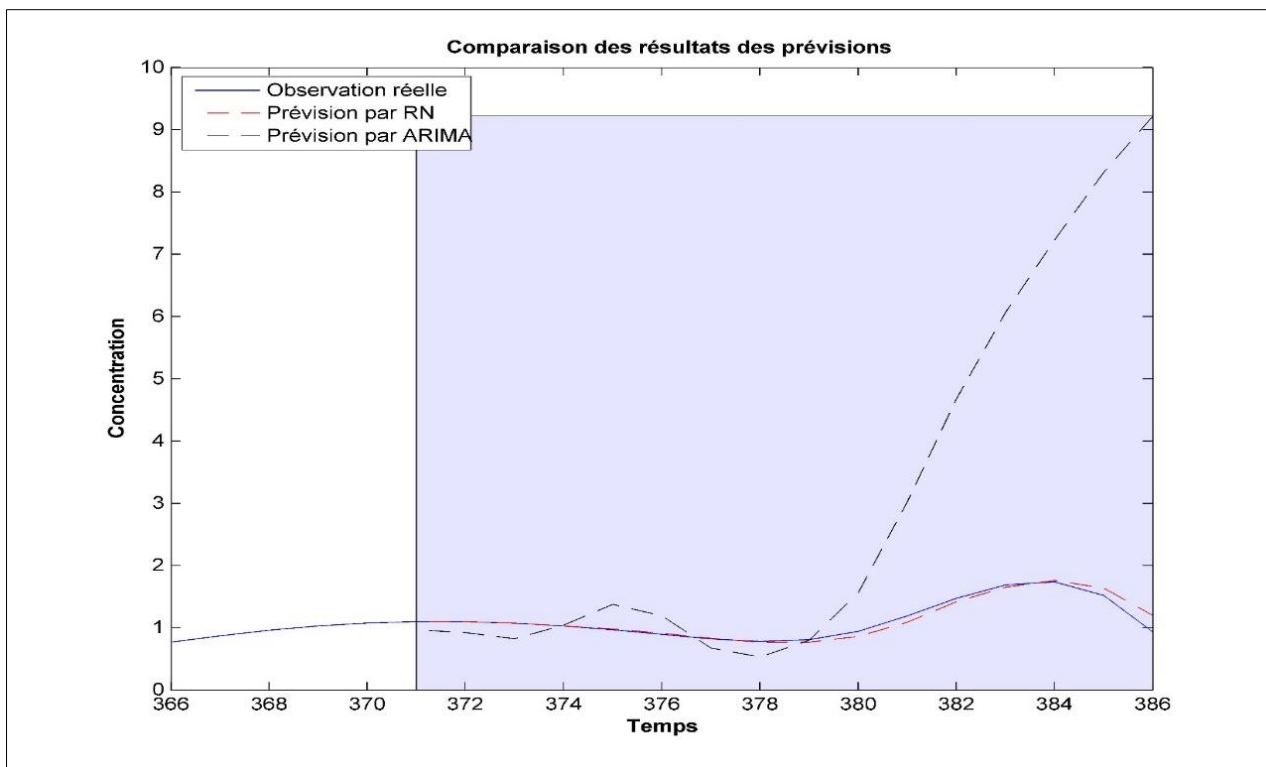


Figure IV.70: Comparaison de la prévision d Césium (Ce) par les deux méthodes de prévision

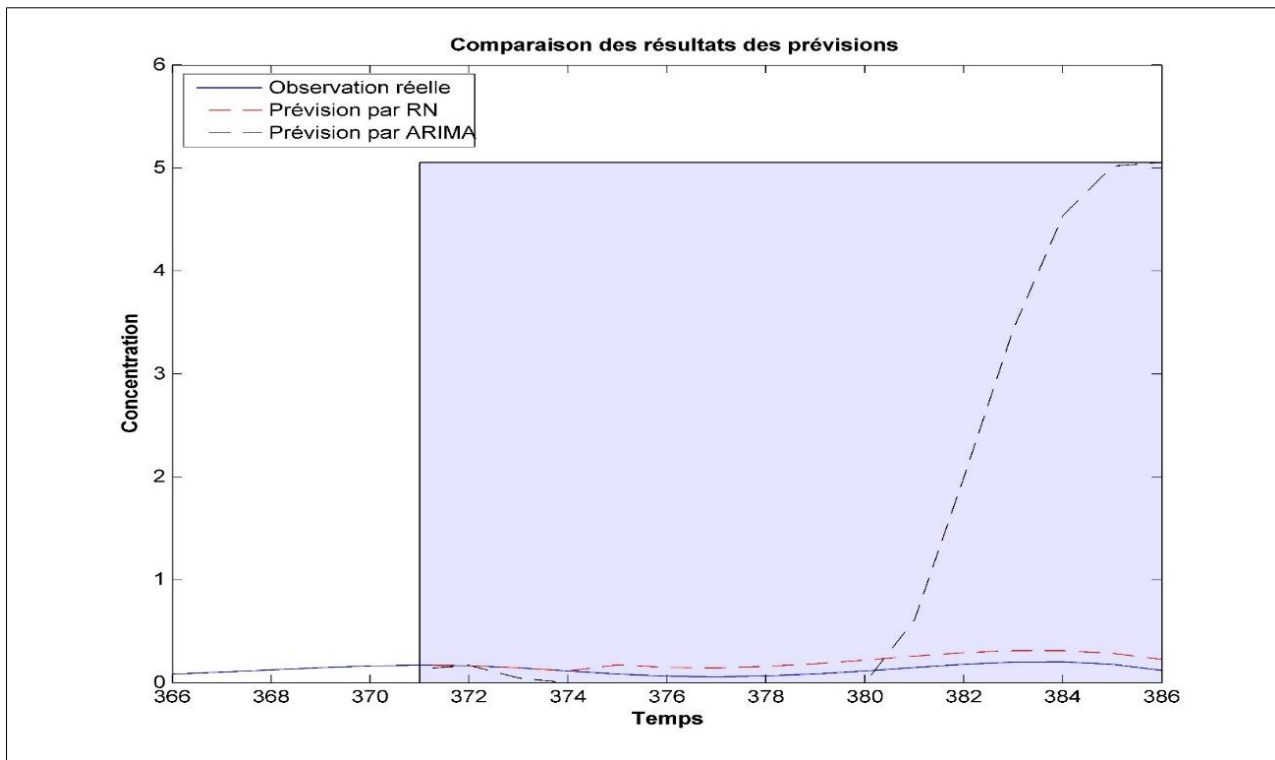


Figure IV.71: Comparaison de la prévision du Lanthane (La) par les deux méthodes de prévision

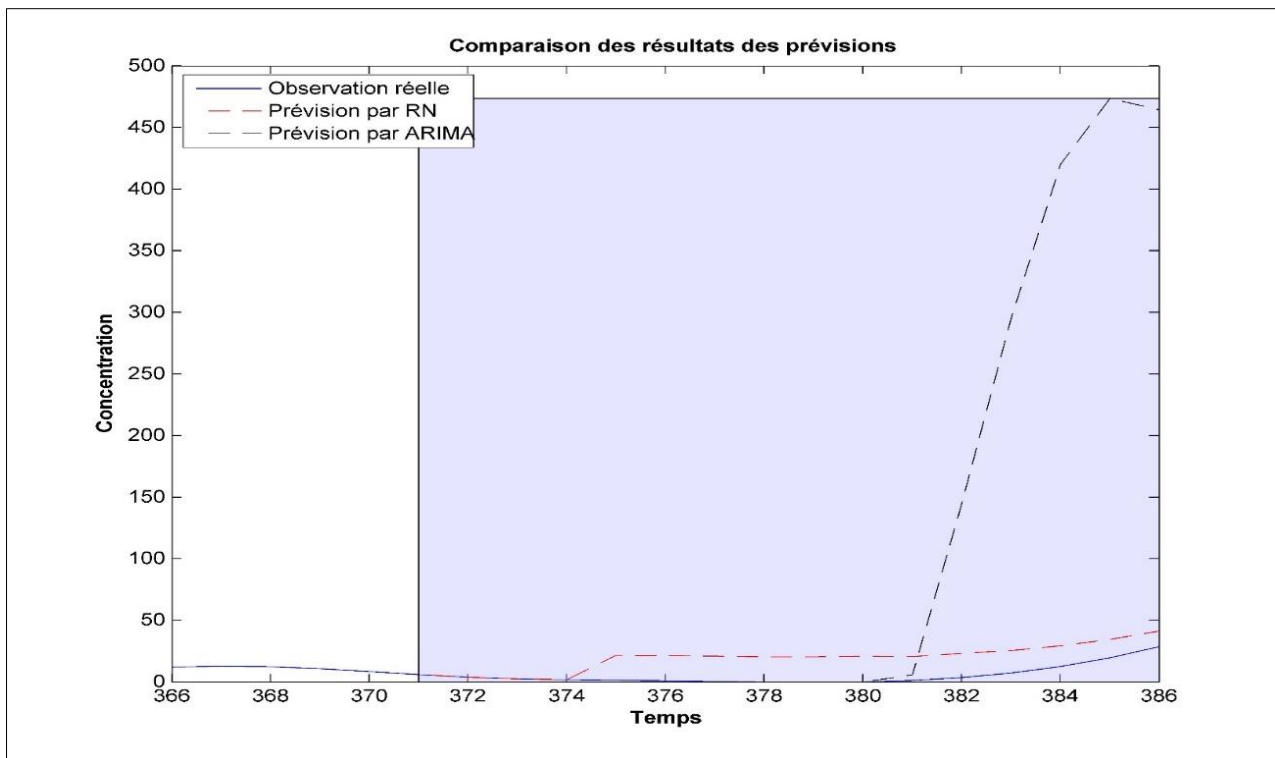


Figure IV.72: Comparaison de la prévision du Sodium (Na) par les deux méthodes de prévision

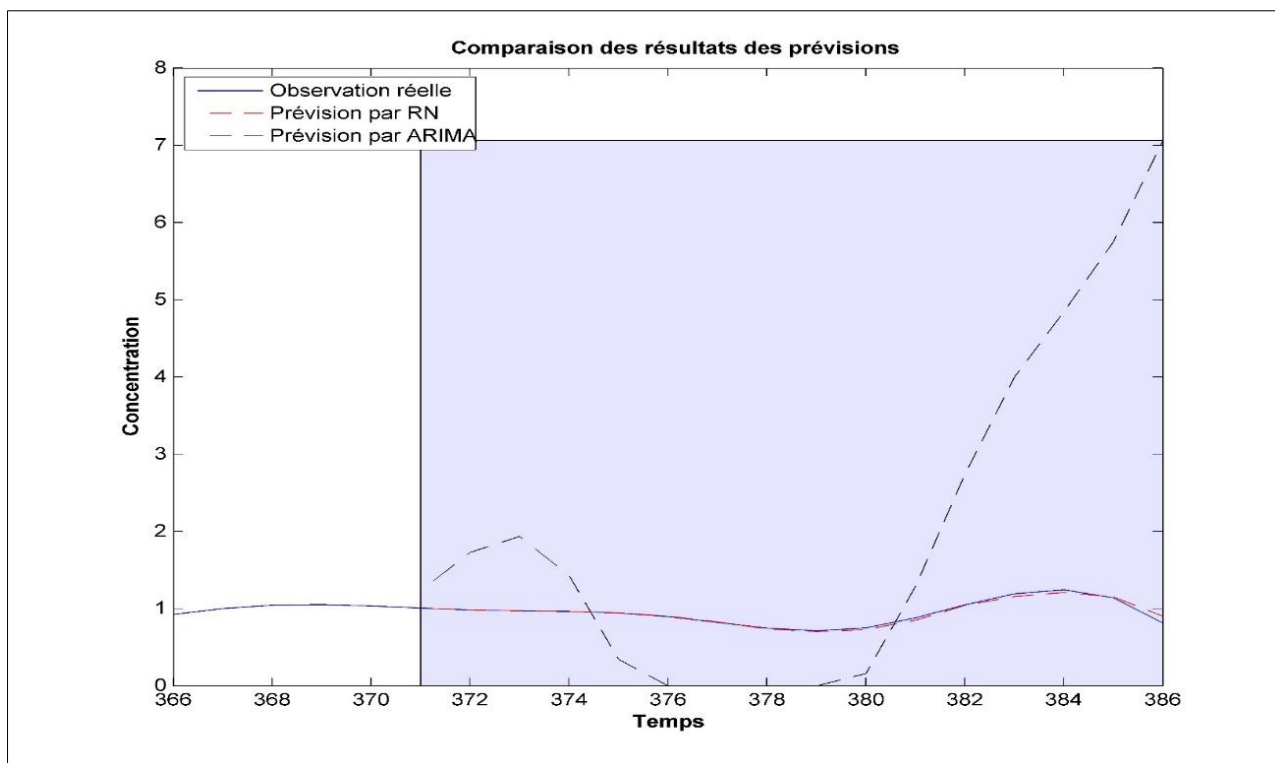


FIGURE IV.73: Comparaison de la prévision du Brome (Br) par les deux méthodes de prévision

Les calculs statistiques de la comparaison entre nos deux méthodes de prévision (ARIMA et réseaux de neurones artificiels) sont présentés dans le tableau ci-dessous :

Tableau IV.16: Comparaison du résultat de la prévision par les deux méthodes de prévisions

Élément	RNA		ARIMA	
	r^2 (%)	RMSE	r^2 (%)	RMSE
TSP	97,5084	3,53	87,7963	17,61
Ce	96,3651	0,03	66,3676	7,59
Cr	94,8080	0,244	63,9621	0,058
Hf	76,2325	0,002	63,7931	0,004
La	99,6825	0,35	58,1619	4,72
Na	99,8948	64,88	89,8812	428,12
Sb	92,6103	0,0075	47,4393	0,6835
Sc	99,3787	0,0053	76,6352	0,1263
Sm	98,9284	0,0067	79,5812	0,063
Yb	91,4177	0,004	70,9566	0,088
Br	98,3633	0,011	56,5937	4,35

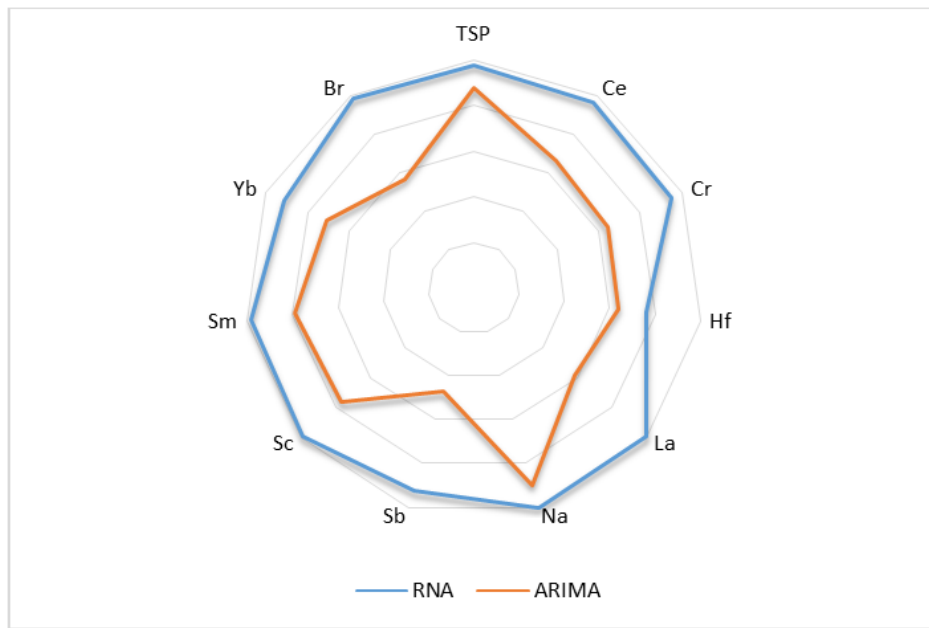


Figure IV.74: Comparaison du r^2 de la prévision des TSP et ET associés par les deux méthodes de prévision

Il est clairement visible dans la figure IV.75 que la méthode de prévision par les réseaux de neurones a donné les meilleurs résultats et la meilleure qualité de précision pour l'ensemble des TSP et ET associés, la combinaison entre les deux critères de validation (r^2 et RMSE) confirme que les RNA sont l'outil le plus performant dans la prévision des TSP et les ET associés pour un horizon de prévision de quatre jours.

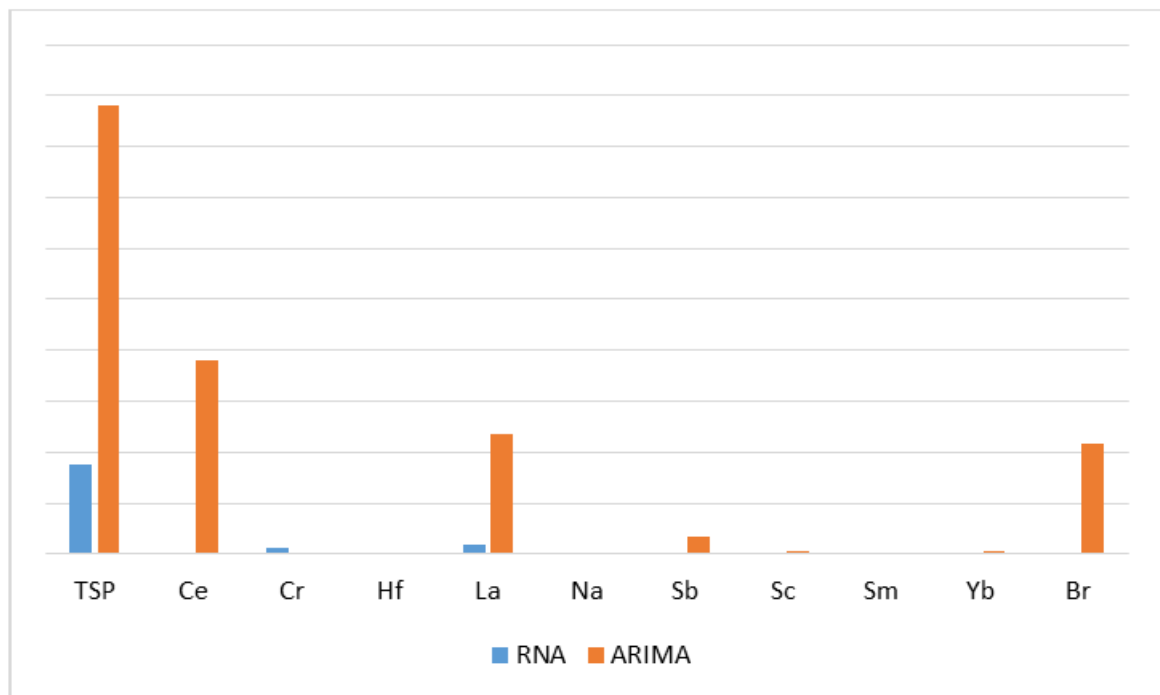


Figure IV.75: Comparaison du RMSE de la prévision des TSP et ET associés par les deux méthodes de prévision

IV.9. La cartographie de la pollution de l'air par les TSP

Nous avons appliqué le modèle Gaussien de la dispersion des polluants dans l'atmosphère pour dresser une cartographie de la pollution par les TSP au niveau de notre zone d'étude, une carte qui s'étend sur un périmètre de 500 mètres dont le centre c'est le CRNA. Pour ce faire nous avons injecté dans un programme informatique rédigé en langage Matlab les valeurs de la vitesse et la direction du vent (figure IV.76) ainsi que la concentration des TSP enregistrée ce jour pour appliquer le modèle plume de la dispersion de la pollution, la sortie du programme nous donne le profil de dispersion de cette pollution (profil de la concentration verticale et profil de la concentration horizontale) en chaque point de ces axes (x, y, z) (figure IV.77).

Nous nous sommes limité à une distance de 500 mètres seulement car nous avons remarqué qu'en dépassant cette distance la valeur de la concentration devient trop faible et non significative

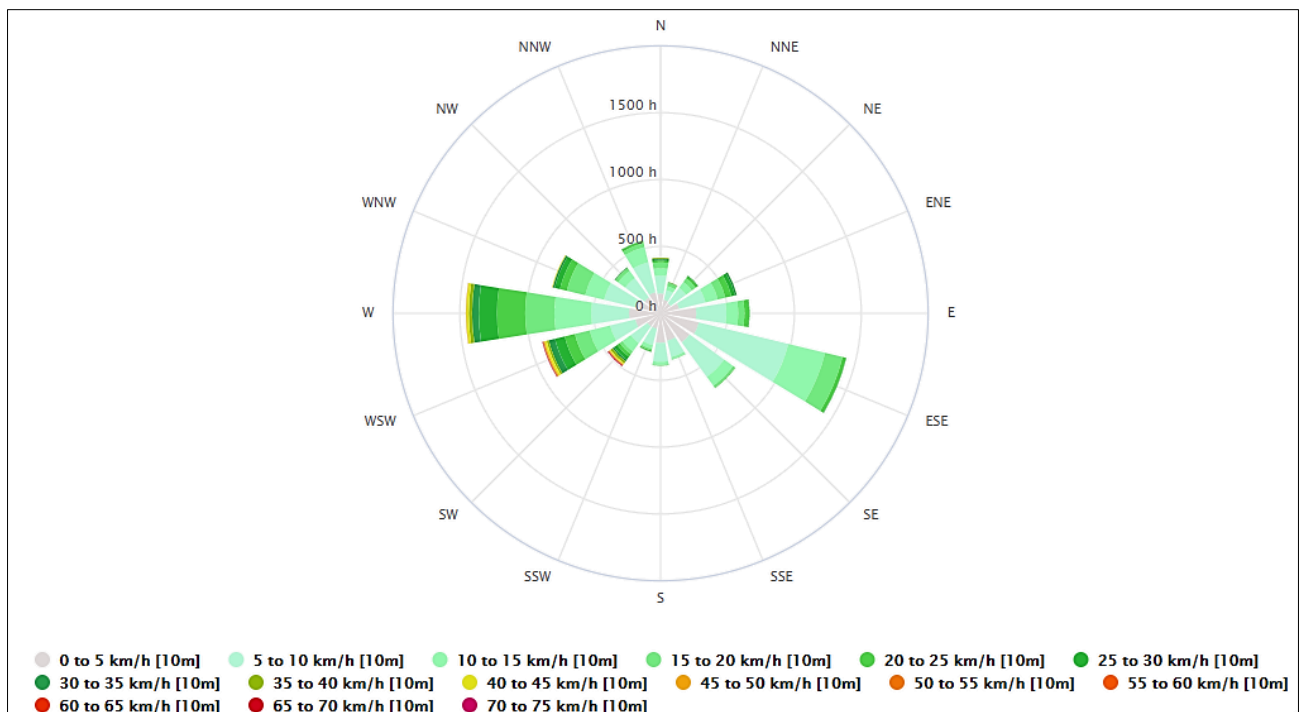


Figure IV.76: Rose des vents enregistrée au niveau de la station de Dar El Beida durant la période d'étude

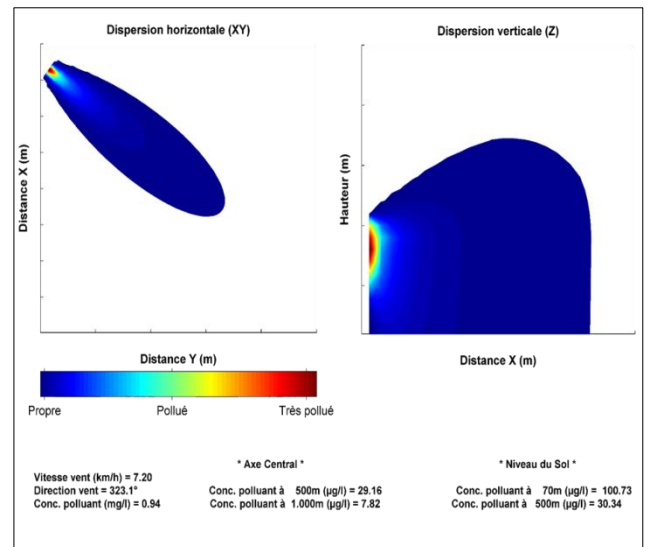
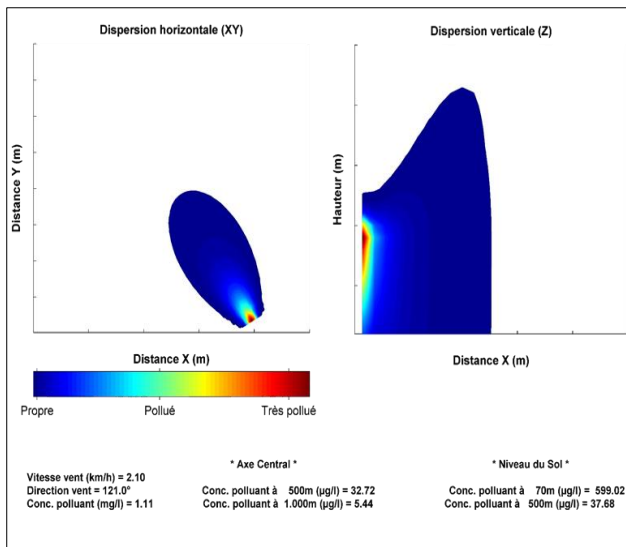
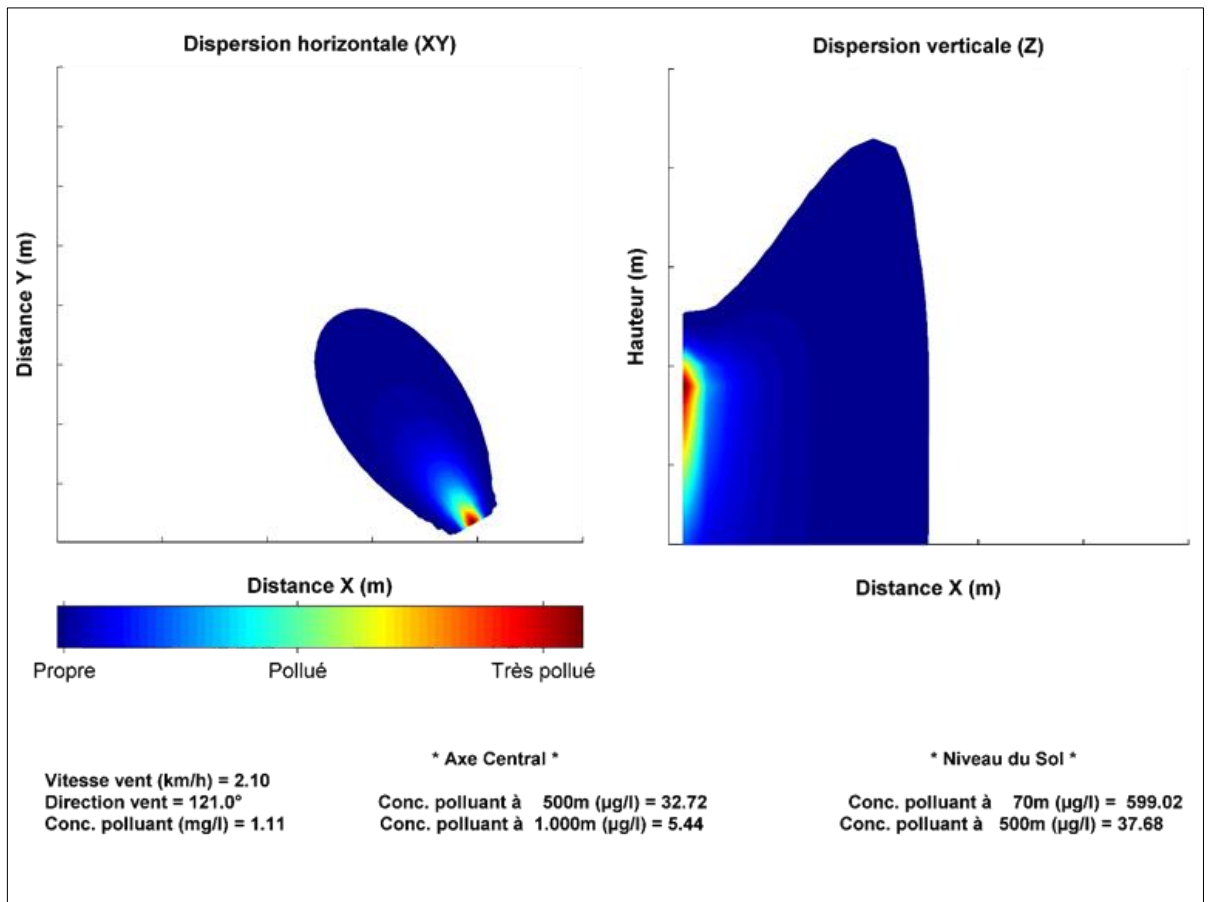


Figure IV.77: Exemple du profil vertical et horizontal de la dispersion TSP par le modèle plume

Pour dresser la cartographie nous avons utilisé une simple stratégie, le programme informatique élaboré nous donne les deux profils de la concentration des TSP dans l'air, nous avons divisé la parcelle où la carte prendra place en une matrice composée de plusieurs carreaux de un mètre de côté, une fois que nous avons eu toutes les concentrations en tout point de la surface nous avons calculé la somme de la concentration en ces points (figure IV.78)

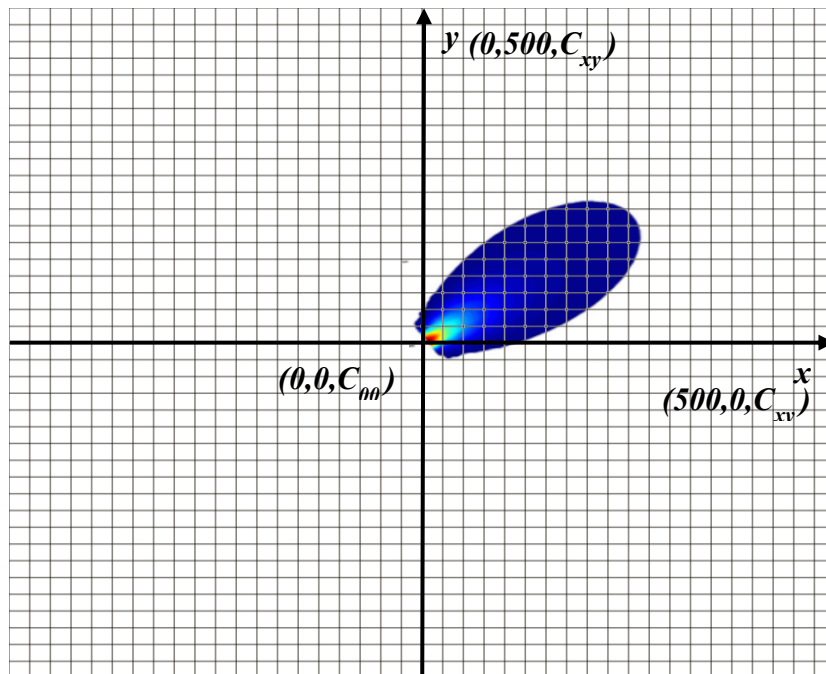


Figure IV.78: Schéma représentatif de la méthode de l'élaboration de la cartographie de la pollution par les TSP

En dernière étape pour couvrir les points manquants nous avons fait une interpolation entre les concentrations avoisinantes. La cartographie finale de la pollution de l'air par les TSP est la suivante :

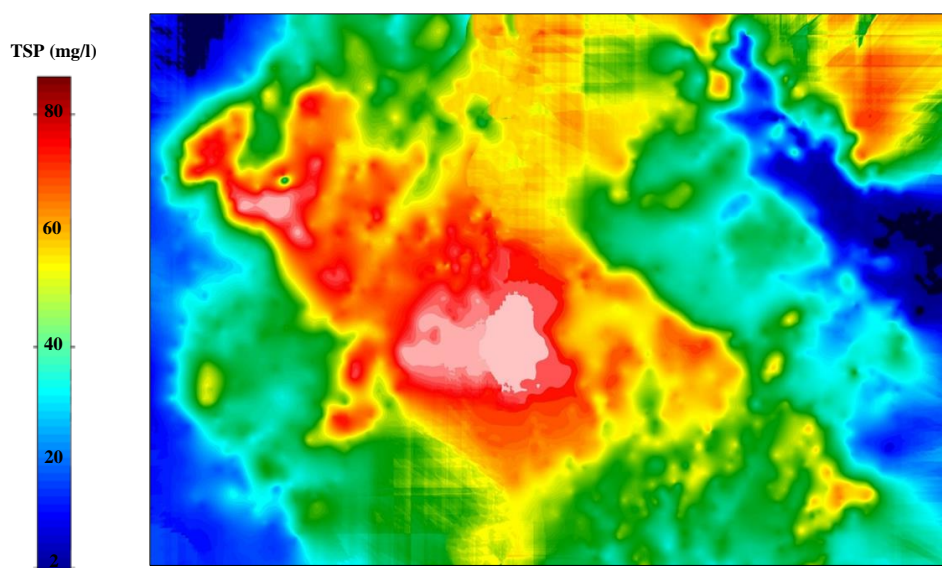


Figure IV.79: Cartographie de la pollution de l'air par les TSP au niveau de notre zone d'étude

La superposition de la cartographie de la pollution de l'air au niveau de notre site d'étude (Draria) sur l'image satellite obtenue par Google Earth Pro est la suivante :

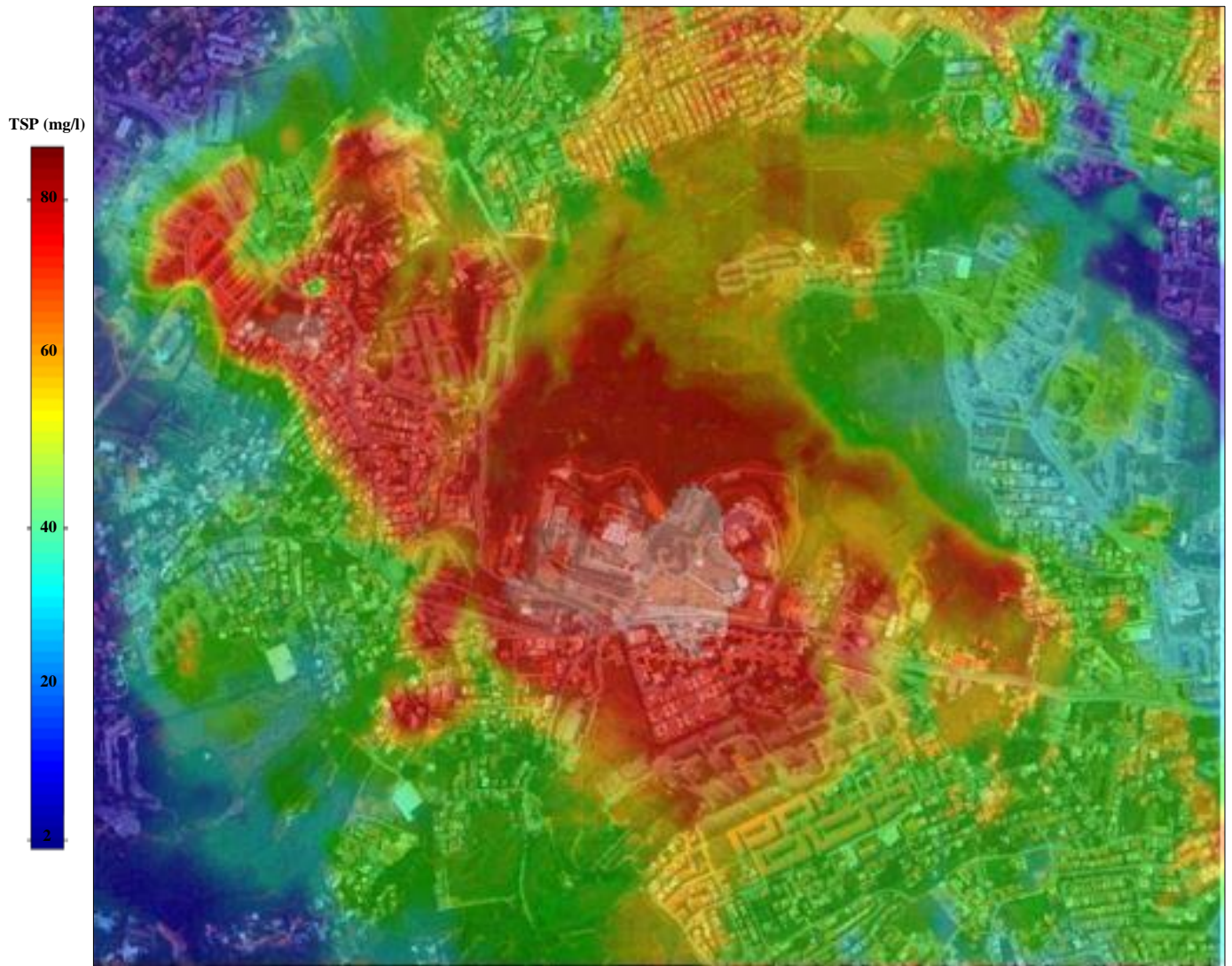


Figure IV.80: Cartographie de la pollution de l'air par les TSP au niveau de notre zone d'étude avec géolocalisation

Conclusion générale

Conclusion générale

Au cours de cette étude, nous avons présenté les niveaux de la pollution de l'air ambiant par les Totales Particules en Suspension (TSP) et les Eléments Traces (ET) associés au niveau de la banlieue de Draria, commune de la wilaya d'Alger, grâce à sa localisation, ce site n'est pas adapté pour le suivi de la pollution urbaine et l'étude de l'exposition des populations. En deuxième phase nous avons présenté les résultats obtenus à partir de notre propre station de prélèvement des gaz polluants, cette station nous a permis d'analyser onze (11) gaz polluants qui sont issus principalement du trafic routier suite à la combustion des carburants fossiles.

La présente étude montre qu'en ce site, la pollution de l'air par les TSP n'est pas inquiétante et relativement faible vu que 85,5% des valeurs enregistrées ne dépassaient pas la valeur cible Algérienne ($50\mu\text{g}/\text{m}^3$) à l'exception de quelques jours où une teneur atmosphérique en TSP supérieure à la norme fixée par l'OMS ($80\mu\text{g}/\text{m}^3$).

Le suivi des teneurs des onze gaz polluants sur la période allant du 03 Juin 2018 au 28 Novembre 2018 nous a permis de tracer les évolutions journalières, hebdomadaire et mensuelles et de mettre en évidence des pics de pollution qui sont liés à l'intensification du trafic routier qui passe à une centaine de mètres du site de mesure. L'étude de l'évolution hebdomadaire a montré que le vendredi (jour férié) accuse le taux de pollution le plus faible. L'évolution mensuelle de cette pollution a montré que le mois d'Août est le mois le moins pollué. Une variation saisonnière est ainsi mise en évidence et on relève en saison estivale une diminution du taux de pollution de l'ordre de 15 à 25 %. Cette diminution est en relation avec l'intensification de la photochimie atmosphérique qui a lieu en été.

En termes de niveaux de pollution, les teneurs mesurées pour le benzène sont relativement élevées mais restent en général en dessous des normes en vigueur dans les pays occidentaux. De même, nous avons montré que, tout comme le CO, le benzène, GPL, Toluène, CO_2 constituent de bon indicateurs de la pollution automobile, afin de mieux comprendre ces corrélations une étude sur les gaz d'échappement des véhicules s'impose.

Nous avons ensuite modélisé cette pollution par les TSP et les ET associés en faisant appel à deux techniques de modélisation différentes, l'objectif n'était pas d'expliquer les phénomènes physiques mis en jeu mais d'élaborer un modèle pour chaque élément qui permettra la simulation de ses concentrations dans le futur proche. Cette étude a permis de démontrer l'intérêt de l'intelligence artificielle par les réseaux de neurones artificiels comme outil de modélisation de la pollution atmosphérique en TSP et en ET associés en modélisant les TSP ainsi que dix (10) Elément Traces associés.

Une fois les modèles obtenus ont été validés nous avons procédé à la prévision de cette pollution en utilisant uniquement les TSP comme entrée. Nous avons proposé un réseau combiné capable de modéliser, détecter, estimer et de prévoir les ET à partir des TSP seulement.

En dernière étape nous avons appliqué sur les TSP le modèle Gaussien de la dispersion des polluants dans le but d'établir une cartographie de la pollution de l'air par les TSP au niveau de Draria, une carte qui s'étale sur un rayon de 500 mètres par rapport au CRND. L'étude de la rose des vents, la rose de pollution et la cartographie nous a permis d'identifier les zones qui sont le plus affectées par la pollution par les TSP.

Globalement, les différentes études que nous avons menées donnent des résultats intéressants et prometteurs qui témoignent de la faisabilité et la validité de notre démarche :

- Les résultats de la pollution par les gaz obtenus à partir de notre propre station de prélèvement présentent une forte **ressemblance** en termes de grandeur et profil avec d'autres études approuvées à Alger et à travers d'autres pays du monde ;
- Les résultats de la prévision des TSP et des ET associés que nous avons obtenu via nos modèles de prévision ont été comparées avec les concentrations réelles, une parfaite **cohérence** en terme de précision a été constatée ;
- Nous avons **créé des scénarios prévisionnels** de la pollution atmosphérique relatifs à notre site d'étude dans un but préventif, que la méthode classique ne le permet pas ;
- Avec nos techniques et nos modèles de prévision, l'obtention de ces résultats **n'a pris que quelques minutes**, alors que l'échantillonnage et la quantification réelle des TSP ainsi que les ET associés par la méthode classique dure pendant des jours.

A la lumière de ces résultats, pour une extension et un enrichissement de ce travail nous proposons :

- L'installation d'autres stations de prélèvement interconnectées au niveau des sites avoisinants afin d'établir une cartographie exacte et en temps réel de la pollution atmosphérique de la région (toute la commune de Draria) puis de la ville d'Alger ;
- L'installation des préleveurs au niveau des axes routiers les plus actifs afin de diminuer les risques de l'exposition à cette pollution en gérant la circulation routière ;
- L'élaboration et la mise en œuvre d'une réglementation en matière de qualité de l'air ambiant qui comporte des prescriptions techniques et des normes d'émissions des rejets par les sources émettrices et la contribution à la normalisation des Eléments Traces.

Références bibliographiques

*R*éférences bibliographiques

- A.Bitouche, « Evaluation de la pollution atmosphérique par les particules fines en site urbain à Alger », Mémoire de Magister, ENP, (2008).
- A.Bitouche, 2008, Evaluation de la pollution atmosphérique par les particules fines en site urbain à Alger, Mémoire de Magister, ENP.
- Adams, F., Dams, R., 1970. Appendix III - A Compilation Of Gamma-Ray Spectra (Germanium Detector). Applied Gamma-Ray Spectrometry (Second Edition Completely Revised and Enlarged), 1970, Pages 469, 471–651.
- Adams, F., Dams, R., 1970. Appendix III - A Compilation Of Gamma-Ray Spectra (Germanium Detector). Applied Gamma-Ray Spectrometry (Second Edition Completely Revised and Enlarged), 1970, Pages 469, 471–651.
- Antoni, J., 2012. A bayesian approach to sound source reconstruction : Optimal basis, regularization, and focusing. The Journal of the Acoustical Society of America , 131, 2873–2890.
- Arbilla G., Bauerfeldt G.F. and De Paula M., Atmospheric levels of aldehydes and BTEX and their relationship with vehicular fleet changes in Rio de Janeiro urban area. Chemosphere, 67, 2007, pp 2096-2103 3- Skov H., Hansen A.B., Lorenzen G., Andersen H.V., Løfstrøm P. and Christensen
- Atkinson, R., Lloyd, A.C., 1984. Evaluation of kinetic and mechanistic data for modeling of photochemical smog. J. Phys. Chem. Ref. Data 13 (2), 315.
- Avino, P., Capannesi, G., Renzi, L., Rosada, A., 2013. Instrumental neutron activation and statistical approach for determining baseline values of essential and toxic elements in hairs of high school students. Ecotox. Environ. Safety 92, 206–214.
- Avino, P., Capannesi, G., Rosada, A., 2008. Heavy metal determination in atmospheric matter by Instrumental Neutron activation Analysis. Microchemical Journal 88, 97–106.
- Baillargeon, J., 2003. L'analyse en composantes principales. UQTR.
- Bell, M.L., Davis, D.L., 2001. Reassessment of the lethal London fog of 1952: novel indicators of acute and chronic consequences of acute exposure to air pollution. Environmental Health Perspectives 109 (Supplement 3), 389–394.
- Berkouki, N., 2007. Caractérisation de la pollution de l'air par les particules fines et des métaux associés, en milieu urbain à Alger centre, magister, ENP.
- Bert, V et deram, 2010. A. Guide des phytotechnologies : utilisation des plantes dans la dépollution et la réhabilitation des sites contaminés par les métaux lourds. France : Environnement et Développement alternatif, 1999. Cité par Fanny VANOBBERGHEN, la phytoremédiation en Wallonie: Evaluation du potentiel d'assainissement des sols contaminés en métaux lourds. Bruxelles : Université libre de Bruxelles.
- Bi, C., Chen, Y., Zhao, Z., et al., 2020. Characteristics, sources and health risks of toxic species (PCDD/fs, PAHs and heavy metals) in PM_{2.5} during fall and winter in an industrial area. Chemosphere 238, 124620-124620.
- Blin, Cathrine, 2004. Les micropolluants dans l'air. Wallonne : s.n.
- Blin, Cathrine. les micropolluants dans l'air. Wallonne : s.n., 2004.

- Boldo E., Medina S., Le Tertre A., Hurley F., Mücke H-G., Ballester F., Aguilera I. & Eilstein D on behalf of the Aphis group. 2006. Aphis: Health impact assessment of long-term exposure to PM2.5 in 23 European cities», *European Journal of Epidemiology*
- Boughedaoui M., R. Kerbachi et R. Jourmard : Mesure de la pollution plombifère dans l'air ambiant d'Alger, *Poll. Atm.*, N°121, 2004, pp.105-111
- Bouhila Z, Mouzai M, Azli T, Nedjar A, Mazouzi C, Zergoug Z, Boukhadra D, Chegrouche S, Lounici H (2015) Investigation of aerosol trace element concentrations nearby Algiers for environmental monitoring using instrumental neutron activation analysis. *Atmos Res* 166:49–59
- Bounouira, H., Embarch, K., Amsil, H., Bounakhla, M., Blaauw, M., 2014. Neutron flux characterization of the Moroccan Triga Mark II research reactor and validation of the ko standardization method of NAA using ko-IAEA program. *J. Radioanal Nucl.Chem.* 300, 465–471.
- Box, G.E.P. & Jenkins, G.M. (1976). *Time series analysis : Forecasting and control*. Oakland, CA : Holden-Day.
- Box, George E. P. (2015) : *Time Series Analysis: Forecasting and Control*. WILEY. ISBN 978-1-118-67502-1.
- Brauer, M., Freedman, G., Frostad, J., et al., 2016. Ambient air pollution exposure estimation for the Global Burden of Disease 2013. *Environ. Sci. Technol.* 50,79e88.
- Brigitte Escofier et Jérôme Pagès, *Analyses factorielles simples et multiples : objectifs, méthodes et interprétation*, Paris, Dunod, Paris, 2008, 318 p.
- Brumer P.h : Approche Globale des problèmes d'environnement liés à l'incinération d'ordures ménagères. *Poll. Atm.*, Juillet-Septembre, 1998, pp. 301-308.
- Bur, Thomas. *IMPACT ANTHROPIQUE SUR LES ELEMENTS TRACES METALLIQUES DANS LES SOLS AGRICOLES DE MIDI-PYRENEES IMPLICATIONS EN TERMES DE LIMITES ET DE CHARGES CRITIQUES*. Toulouse : l'Institut National Polytechnique de Toulouse (INPT), 2008.
- C.S. Benzene exposure and the effect of traffic pollution in Copenhagen, Denmark. *Atmospheric Environment*, 35, 2001, pp 2463-2471
- Calvo, A.I., Alves, C., Castro, A. Pont, V., Vicente, A.M., Fraile, R., 2013. Research on aerosol sources and chemical composition: Past, current and emerging issues. *Atmos. Res.* 120–121, 1–28.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 245-276.
- Chan, K.H., Kurmi, O.P., Bennett, D.A., Yang, L., Chen, Y., Tan, Y., et al., 2019. Solid fuel use and risks of respiratory diseases. A cohort study of 280,000 Chinese never-smokers. *Am. J. Respir. Crit. Care Med.* 199, 352e361.
- Chen, H., Li, L., Lei, Y., Wu, S., Yan, D., Dong, Z., 2020. Public health effect and its economic loss of PM2.5 pollution from coal consumption in China. *Science of the Total Environment* 732, 138973.
- CITEPA (Centre Inter professionnel technique d'Etudes de la Pollution Atmosphérique), 2006. *Emission dans l'air en France, Métaux lourds*, Ministère de l'Ecologie et du Développement Durable. France.
- Cohen, A.J., Brauer, M., Burnett, R., et al., 2017. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the global burden of diseases study 2015. *Lancet* 389, 1907e1918.

- Commenges, D., Jacqmin-Gadda, H., Proust, C. and Guedj, J. (2006). A newton-like algorithm for likelihood maximization: The robust-variance scoring algorithm. Arxiv preprint math/0610402
- Contamine, Jérôme. Le magazine de la chronique scientifique. La modélisation, un outil au service de la Recherche. Novembre 2005, 6.
- Contini, D., Cesari, D., Genga, A., Siciliano, M., Ielpo, P., Guascito, M.R., Conte, M., 2014. Source apportionment of size-segregated atmospheric particles based on the major water-soluble components in Lecce (Italy). *Sci. Total Environ.* 472, 248–261.
- Cornuéjols A. & Miclet L. : Apprentissage artificiel : Concepts et algorithmes. De Bayes et Hume au Deep learning. Eyrolles, 2021
- Coulibaly, P., Anctil, F., Bobee, B., (1999) : Préviation hydrologique par réseaux de neurones artificiels. Etat de l'art. *Journal of Civil Engineering Canada* n° 26, pp. 293-301
- Crabol (B.). – La dispersion des polluants dans l'atmosphère. Clefs CEA no 49, pages 1-13 (1999).
- Currie, L., A., 1968. Limits for qualitative detection and quantitative determination. Application to radiochemistry. *Anal. Chem.* 40, 586–593.
- Daillant, Olivier, 2003. lichens et accumulation des métaux lourds. Association Française de Lichénologie AFL.
- Dan, A., Oosterbaan, J., Jamet, P., 2002. Artificial Neural Networks (ANNs) characterisation of soil pollution : the Polycyclic Hydrocarbons (PAHs) case study. *C. R. Geoscience* 334, 957–965.
- Dang C. T. P. Fontelle et R. Leygonie : Les émissions de poussières et généralités sur leur mesure à l'émission. *Poll. Atm.*, Janvier-mars, 1990, pp. 38-46
- Davalo, E., Naïm, P., (1991) : Des réseaux de Neurones, Edition Eyrolles, Paris.
- De Marco, S.G., Botté, S.E., Marcovecchio, J.E., 2006. Mercury distribution in abiotic and biological compartments within several estuarine systems from Argentina : 1980-2005 period. *Chemosphere* 65, 213-223.
- Detrie J.P. : La pollution atmosphérique, Edition Dunod, Paris, 1969.
- "Diggle P.J. (1977) : The detection of random heterogeneity in plant populations. *Biometrics* 33, 390-394. 1977.
- Ref Type: Journal"
- Djakir Zergoug, Z., Azli, T., Bouhila, Z., Mouzai, M., Nedjar, A., Ferkha, M.K., Lakehal, C., 2013. Biosurveillance de la « Qualité de l'Air : Dosage des Éléments Traces Métalliques dans les Lichens par la technique d'Analyse par Activation Neutronique. Journées Internationales de Biotechnologie 2013 (JIB2013), 21-24 Décembre 2013, Tunisie.
- Doury (A.). – Une méthode de calcul pratique et générale pour la prévision numérique des pollutions véhiculées par l'atmosphère. Rapport CEA-R-4280, Rév. 1 (septembre 1976).
- El Khoukhi, T., Cherkaoui, R.M., Gaudry, A., Ayrault, S., Senhou, A., Chouak, A., Moutia, Z., E., Chakir, 2004. Air pollution biomonitoring survey in Morocco using k0-INAA. *Nucl. Instruments and Methods in Phys. Res. B* 213, 770–774.
- FABLEC, Y. (1999) : Préviation de trajectoires d'avions par réseaux de neurones. Thèse Informatique Fondamentale et Parallélisme. Ecole Nationale de l'Aviation Civile, Toulouse.
- Feng, L., Xiang-yun, Z., Chang-hua, W., Zhi-peng, D., Yan-mao, W., Guo-ru, H., Xiao-lin, L., Min-jian, L., Man-jie, L., and Jiang-Yu, X., 2013. Ecological Risks Assessment and Pollution Source Identification of Trace Elements in Contaminated Sediments from the Pearl River Delta, China. *Biol. Trace Element Res.* 155, 301–313.

- Fontan, J., 2003. Les Pollutions de l'Air, les Connaître pour les Combattre. Paris : Vuibert.
- Framling, K., (1992) : Les réseaux de neurones comme outils d'aide à la décision floue. Rapport de D.E.A. Ecole Nationale Supérieure des Mines, Saint-Etienne
- French, M.N., Krajewski, W.F., Cuykendall, R.R., (1992) : Rainfall Forecasting in Space and Time using a Neural Network, *Journal of Hydrology*, Vol. 137, p. 1-31.
- Fuertes, E., Sunyer, J., Gehring, U., et al., 2020. Associations between air pollution and pediatric eczema, rhinoconjunctivitis and asthma: a meta-analysis of European birth cohorts. *Environ. Int.* 136, 105474
- Gifford, F. (1961) : Use of Routine Meteorological Observations for Estimating Atmospheric Dispersion. *Nuclear Safety*, 2, 44-57.
- Goodwin, P., & Wright, G. (2009) : *Decision analysis for management judgment* (4th ed). Chichester: John Wiley & Sons
- Goyal, N., Jain, S.C., Banerjee, U.C., 2003. Comparative studies on the microbial adsorption of heavy metals. *Advances in Environmental Research* 7, 311-319.
- Goyer, R. A., Clarkson, T. W., 2001. Toxic effects of metals. In : Casarett and Doull's Toxicology : The basic science of poisons. Klaasen, C.D., edit., and Mc Graw-Hill. Medical Division, pub., New York, 811–867.
- Grandy, J., Asl-Hariri, S., Pawliszyn, J., 2015. Chapter 7 - Novel and Emerging Air-Sampling Devices. *Comprehensive Analytical Chemistry* 70, 209-235.
- Green, K. C., & Armstrong, J. S. (2007) : Structured analogies for forecasting. *International Journal of Forecasting*, 23(3), 365–376.
- Greenberg, R. R., Bode, P., De Nadal Fernandes, E. A., 2011. Neutron activation analysis: A primary method of measurement. *Spectrochimica Acta Part B: Atomic Spectroscopy* 66, 193–241.
- Grossberg, S., & Stone, G. (1986). Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance. *Psychological Review*, 93(1), 46–74
- Grubbs, F.E. and Beck G. (1972) : Extension of sample sizes and percentage points for significance tests of outlying observations. *Technometrics*, 14, 847-854.
- Gu, X., Liu, Q., Deng, F., et al., 2019. Association between particulate matter air pollution and risk of depression and suicide: systematic review and meta-analysis. *Br. J. Psychiatr.* 215, 456e467
- Guo, H., Sahu, S.K., Kota, S.H., et al., 2019. Characterization and health risks of criteria air pollutants in Delhi. *Chemosphere* 225, 27e34, 2017.
- Harris, R., & Sollis, R. (2003) : *Applied time series modelling and forecasting*. Chichester, UK: John Wiley & Sons
- He, T., Zhu, J., Wang, J., et al., 2018. Ambient air pollution, h19/DMR methylation in cord blood and newborn size: a pilot study in Zhengzhou city, China. *Chemosphere* 212, 863e871.
- Hebb, D. O. (1949) : *The organization of behavior; a neuropsychological theory*. Wiley.
- Herrebaut L. et J. Beguin : *Mesure des pollutions particulaire et gazeuse, Mesure – Analyse – Contrôle, Nuisance et Environnement*, France, Mai 1976, pp. 24-26.
- Hinds C. W. *Aerosol Technology*. Ed. WJ. USA : s.n, 2012.
- Hoek, G., Pattenden, S., Willers, S., et al., 2012. PM10, and children's respiratory symptoms and lung function in the PATY study. *Eur. Respir. J.* 40 (3), 538e547
- Hsu, K., Gupta, H., Sorooshian, S., (1995) : Artificial neural network modeling of rainfall-runoff process. *Water Resources Research*, 31(10), pp. 2517-2530.

- Hyndman, R. J., & Koehler, A. B. (2006) : Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688.
- Idier, J., 2001. Approche bayésienne pour les problèmes inverses, *Traité IC2, Serie traitement du signal et de l'image*
- J.Priest, « Problems of our Physical Environment», Addison-Wesley Publishing Company ,pp 59-70, (1973).
- Jauffret, A., Grand, C., Cuperlier, N., Gaussier, P., and Tarroux P. (2013). “How can a robot evaluate its own behaviour? A generic model for self-assessment,” in *International Joint Conference on Neural Networks (IJCNN)* (Dallas, TX).
- Jing, C., Ang, Z., Jinzhuo, Z., Renjie, C., Weibing W., Sandie, H., Xiaohui, X., Haidong, K., 2014. Acute effects of air pollution on asthma hospitalization in Shanghai, China. *Environmental Pollution* 191, 139-144.
- Jodouin, J. F., (1994) : *Les réseaux de neurones*, Edition Hermès, Paris .
- Johnson, T.M., Guttikunda, S., Wells, G.J., Artaxo, P., Bond, T.C., Russell, A.G., Watson, J.G., 2011. *West Tools for Improving Air Quality Management: A Review of Top-down Source Apportionment Techniques and Their Application in Developing Countries*, World Bank Energy Sector Management Assistance Program, Formal Report 339/11. The World Bank Group, Washington, DC.
- Jong-Myoung, L., Jin-Hong, L., Jong-Hwa, M., Yong-Sam, C., Ki-Hyun, K., 2010. Airborne PM10 and metals from multifarious sources in an industrial complex area. *Atmos. Res.* 96, 53–64.
- JORA N°1 Décret exécutif n°06-02 du 7 Dhou El Hidja 1426 correspondant au 7 janvier 2006 définissant les valeurs limites, les seuils d’alerte et les objectifs de qualité de l’air en cas de pollution atmosphérique. Art.1 à 10. Pages 3–4.
- JORF n°0001 du 1 janvier 1997 page 11.
- Jose R. Peralta-V.M., Laura L., Mahesh N., (2009). The biochemistry of environmental heavy metal uptake by plants: Implications for the food chain. *The international journal of biochemistry and Cell Biology* 41: 1665-1677.
- Jourdain, F., 2007. *Modélisation des transferts atmosphériques de polluants. s.l. : Technique de l'ingénieur IN65, 3.*
- Julien Bayle, ed. Packt Publishing Limited, (2013) : *C Programming for Arduino*, ISBN 978-1849517591
- Jurado, J.M., Martín, M.J., Pablos, F., Moreda-Piñeiro, A., Bermejo-Barrera, P., 2007. Direct determination of copper, lead and cadmium in aniseed spirits by electrothermal atomic absorption spectrometry. *Food chemistry* 101, 1296-1304.
- K, Oikawa. *Analyse des Echantillons Atmosphériques*, Centre de Recherche. Alger : s.n., 1999.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141-151.
- Kan, H.D., Shi, X.M., 2019. Research progress of ambient air pollution and human health in China. *Zhonghua Yu Fang Yi Xue Za Zhi* [Chinese journal of preventive medicine] 53, 4e9.
- Katharina, J., 2014. Air pollution, avoidance behaviour and children's respiratory health: Evidence from England. *Journal of Health Economics* 38, 23-42.
- Kaygusuz, K., 2012. Energy for sustainable development: A case of developing countries. *Renewable & Sustainable Energy Reviews* 16, 1116–1126.
- Kerbach R., Boughedaoui M., Bounoua L. and Keddou M., Ambient air pollution by aromatic hydrocarbons in Algiers. *Atmospheric Environment*, 40, 2006, pp 3995-4003 2- Martins E.M.,

- Kerbachi R., M. Keddou, M. Boughedaoui et R. Joumard, Aromatic hydrocarbons in the ambient air of Algiers, Proc. Int. symp. "transport and Air pollution", Graz, Austria, 31 May-2. June 1999.
- Kerbachi.R, Cours de pollution atmosphérique 5ème année, ENP, Alger, 2005.
- Kim, H.J., Seo, Y.S., Sung, J., et al., 2019. Interactions of cdh13 gene polymorphisms and ambient PM10 air pollution exposure with blood pressure and hypertension in Korean men. *Chemosphere* 218, 292-298
- Kleimola, L.B., Patel, A.B., Borkar, J.A., Hibberd, P.L., 2015. Consequences of household air pollution on child survival: evidence from demographic and health surveys in 47 countries. *Int. J. Occup. Environ* 21, 294–302.
- Kohonen, T. (1982) Self-Organized Formation of Topologically Correct Feature Maps. *Biological Cybernetics*, 43, 59-69
- Kowalski, Z., Kulczycka, J., Wzorek, Z., 2007. Life cycle assessment of different variants of sodium chromate production in Poland. *Journal of Cleaner Production* 15, 28–37.
- Krauss, G., Depecker, P., Kindagen, J.I., 1997. Méthode des réseaux de neurones pour le calcul des coefficients de vitesse. Application aux écoulements de l'air dans les bâtiments. *Revue Génie thermique* 36, 180–191.
- Krzyzanowski M., 2005. La pollution atmosphérique par les particules en suspension : ses effets nuisibles sur la santé. EURO/04/05, Centre européen de l'environnement et de la santé de l'OMS : 1.
- Kucuksezgin, F., A. Kontas, O. Altay, E. Uluturhan, E. Darilmaz. 2006. Assessment of marine pollution in Izmir Bay: Nutrient, heavy metal and total hydrocarbon concentrations. *Environment International*, 32, Issue 1, 41-5.
- Lane, E.A., Canty, M.J. 2015 Cadmium exposure and consequence for the health and productivity of farmed ruminants. *Research in Veterinary Science*, 101, 132-139.
- Larssen, T., Seip, H. M., Semb, A., Mulder, J., Muniz, I.P., Vogt, R.D., Lydersen, E., Angell, V., Dagang, T., Eilertsen O., 1999. Acid deposition and its effects in China: an overview. *Environmental Science & Policy* 2, 9-24.
- Le Bouffant, L., Addison, J., Bolton, R.E., Bruch, J., Bruyet, B., Daniel, H., Davis, J.M.G., Degueldre, G., Demarez, J., Dodgson, J., Gormley, I.P., Hadden, G.G., Kovacs, M.P., Martin, J.C., Reisner, M.T.R., Robertson, A., Rosmanith, J., 1988. Compared in vitro and in vivo toxicity of coalmine dusts. Relationship with mineralogical composition. *Inhaled Particles VI*, 611-620.
- LEK, S., DIMOPOULOS, I., DERRAZ, M., ELGHACHTOUL, Y., (1996) : Modélisation de la Relation Pluie - Débit à l'aide des Réseaux de Neurones Artificiels, *Revue des Sciences de l'eau*, Vol. 3, p. 319-331.
- Li, J., Liu, C., Cheng, Y., et al., 2019. Association between ambient particulate matter air pollution and ST-elevation myocardial infarction: a case-crossover study in a Chinese city. *Chemosphere* 219, 724-729.
- Li, S., Zhou, G., Liu, Z., et al., 2020. Synthesis and performance characteristics of a new ecofriendly crust-dust suppressant extracted from waste paper for surface mine. *J. Clean. Prod.* 258, 120620.
- Li, T., Yan, M., Sun, Q., et al., 2018. Mortality risks from a spectrum of causes associated with wide-ranging exposure to fine particulate matter: a case-crossover study in Beijing, China. *Environ. Int.* 111, 52-59

- Lin, Z., Meng, X., Chen, R., et al., 2017. Ambient air pollution, temperature and Ka-wasaki disease in Shanghai, China. *Chemosphere* 186, 817e822
- Liu, C., Chen, R., Sera, F., et al., 2019. Ambient particulate air pollution and daily mortality in 652 cities. *N. Engl. J. Med.* 381, 705e715.
- Lu, C., Zhang, W., Zheng, X., et al., 2020. Combined effects of ambient air pollution and home environmental factors on low birth weight. *Chemosphere* 240, 124836.
- Magdelaine C. : Les particules inférieures à 2.5 microns (PM_{2,5}), Les particules atmosphériques, 2001-2002
- Magdelaine C. : Les PM₁₀ : Des particules de diamètre > 2,5 microns, Les particules atmosphériques, 2001-2002.
- Maigeng, Z., Yunning, L., Lijun W., Xingya, K., Xiaohui, X., Haidong, K., 2014. Particulate air pollution and mortality in a cohort of Chinese men. *Environmental Pollution* 186, 1-6.
- Marcel, GIROUX, Louise, Deschaines, Raynald et Chassé. les éléments traces métalliques leur accumulation dans les sols agricoles au Québec. Québec : s.n., 2006
- Marquardt, D.W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, 11(2), 431-441
- Masclat, P., 2005. Pollution atmosphérique : causes, conséquences, solutions, perspectives. Ed. Ellipses. 137, 29–30.
- MINNS, A., HALL, J., (1996) : Artificial neural network as rainfall-runoff models. *Journal des sciences hydrologiques*, 41(3), pp. 399-417
- Mohamedou B. et J. Colosio : Qualité de l'air : une surveillance accrue des particules, mais des concentrations à réduire d'ici l'an 2005, Les données de l'environnement, IFEN, N°58, France, Septembre 2000.
- Mokhnache, L., (2004) : Application des réseaux de neurones dans le diagnostic et la prédiction des isolations de haute tension. Thèse de doctorat. Ecole Nationale Polytechnique, Alger.
- N.Oucher, « Contribution à l'étude, par classe granulométrique, des aérosols particuliers et des métaux lourds qui leur sont associées à Alger », Mémoire de Magister, ENP, (2006).
- National Public Radio, 2002. The killer fog of '52. Text and audio files at <http://discover.npr.org/features/feature.jhtml?wfId=873954> (accessed June 2006).
- Nemery, B., Hoet, P.H. M., Nemmar, A, 2001. The Meuse Valley fog of 1930: an air pollution disaster. *THE LANCET* • Vol 357: 704–08
- Newell, K. M. (1986) : Constraints on the Development of Coordination. In M. G. Wade, & H. T. A. Whiting (Eds.), *Motor Development in Children: Aspects of Coordination and Control* (pp. 341-360). The Netherlands: Martinus Nijhoff, Dordrecht
- Nguyen, T. H., Byeong-Kyu, L., 2010. Characteristics of particulate matter and metals in the ambient air from a residential area in the largest industrial city in Korea. *Atmospheric Research* 98, 526–537.
- O.Blanchard, F. Del Gratta, M. Durif, E. Frejafon, O. Le Bihan, « Exposition par inhalation aux aérosols : Principes et méthodes de mesures », INERIS, Paris, (2004).
- Oberdorster G., Oberdorster E. & Oberdorster J. 2005. Nanotoxicology: an emerging discipline evolving from studies of ultrafine particles », *Environ Health Perspect*, 113, 823-39.
- Official journal of the European Communities, Directive 2000/69/EC of 16 November 2000 relating to limit values for benzene and Carbon monoxide in ambient air, L 313, 13 December 2000, pp. 12-21
- Ontario : Particules en suspension, Qualité de l'air, Ministère de l'Environnement et de l'Énergie, Canada, Mai 2000.

- Oregon Revised Statutes. s.l. : Section 449, 1956. 705 et seq
- P.Masclat, « Pollution atmosphérique : causes, conséquences, solutions, perspectives », Edition Ellipse, Paris, (2005).
- Papoulis, Athanasios (2002) : Probability, Random Variables, and Stochastic processes. Tata McGraw-Hill Education.
- Pasquill (F.). – The estimation of the dispersion of winborne material. Met. Mag., vol. 90, p. 33 (1961).
- Patrick Chantereau et Erik Bartmann, éd. Eyrolles, (2014) : Le grand livre d'Arduino, ISBN 978-2212137019
- Piervittori, R et Maffei, S., 2001.The importance of indicator species in the biomonitoring of atmospheric pollution. A case study in the city of Aosta, NWItaly. Cryptogamie Mycologie22, 297–310.
- Preacher, K. J. & MacCallum, R. C. (2003). Repairing Tom Swift’s electric factor analysis machine.
- Raaschou-Nielsen, O., Beelen, R., Wang, M., Hoek, G., Andersen, Z.J., Hoffmann, B., Stafoggia, M., Samoli, E., Weinmayr, G., Dimakopoulou, K., Nieuwenhuijsen, M., Xun, W.W., Fischer, P., Eriksen, K.T., Sørensen, M., Tjønneland, A., Ricceri, F., de Hoogh, K., Key, T., Eeftens, M., Peeters, P.H., 2016.Particulate matter air pollution components and risk for lung cancer.Environment International, 87, 66-73.
- Reimann, C., De Caritat, P., 1998. Chemical Elements in the Environment Factsheets for the Geochemist and Environmental Scientist.Springer-Verlag, Berlin,Germany.
- Renoux A. : Ce qu’il faut savoir sur l’aérosol atmosphérique, Poll. Atm., Janvier – Mars, 1990, pp. 10-16.
- Revel, G., Ayrault, S., 2000.Comparative use of INAA and ICP-MS methods for environmental studies.Radioanal.Nucl. Chem. 244,73–80.
- Rherib , N., El Amili , O. and EL KHOURCHI , B. (2021) : Impact des caractéristiques des tableaux de bord sur la performance organisationnelle des entreprises marocaines : Analyse factorielle exploratoire. Revue Internationale des Sciences de Gestion. 4, 4 (Oct. 2021).
- Rosenblatt, F.,1959. A probabilistic model for visual perception.Acta physiologica 16, 296–297.
- Rosenlund, M., Picciotto, S., Forastiere, F., Et al., 2008.Traffic-related Air pollution in relation to incidence and prognosis of coronary heart disease.Epidemiology 19,121–128.
- Rosbach, M., Jayasekera, R., Kniewald, G., (2000).Large scale air monitoring biological indicators versus air particulate matter.IAEA-TecDoc- 1152, 71–80.
- Rumelhart, D., Hinton, G. & Williams, R. Learning representations by back-propagating errors. Nature 323, 533–536 (1986). <https://doi.org/10.1038/323533a0>
- Sajikumar, N., Thandaveswara, B.S., (1999) : A Non-Linear Rainfall-Runoff model using an Artificial Neural Network, Journal of Hydrology, Vol. 216, p. 32-55.
- Seow,W.J., Hu,W., Vermeulen, R., Hosgood Iii, H.D., Downward, G.S., Chapman, R.S., et al., 2014. Household air pollution and lung cancer in China: a review of studies in xuanwei. Chin. J. Canc. 33, 471-475.
- Seraghni Nassira. Mesure de la pollution particulaire et metalliquedans l’air au niveau de trois sites urbains de la ville de constantine. Canstantine : Université de Canstantine, 2007.
- Shamseldin, A. Y., (1997) : Application of a Neural Network Technique to Rainfall-Runoff Modelling, Journal of Hydrology, Vol. 199, p. 272-294.
- Simon Monk, (2013) : Programming Arduino Next Steps: Going Further with Sketches, ISBN 978-0071830256

- Spray, J.A. & Newell, K.M. (1986). Time series analysis of motor learning : KR versus no- KR. *Human Movement Science*, 5, 59-74.
- Stern A. C. and Boubel R. W.: *Fundamentals of air pollution*, Academic Press, 2ème édition, London, 1984.
- Sui, Z., Zhang, Y., Wang, T., et al., 2018. Full-scale demonstration of enzyme-treated coal combustion for improved energy efficiency and reduced air pollution. *Energy Fuels* 32 (6), 6584e6594.
- Sun, J., Liang, M., Shi, Z., et al., 2019. Investigating the PM2.5 mass concentration growth processes during 2013-2016 in Beijing and Shanghai. *Chemosphere* 221,452e463.
- TOUZET, C., (1992) : *Les réseaux de neurones artificiels*.
- Vanobberghen Fanny. *La phytoremédiation en Wallonie : Evaluation du potentiel d'assainissement des sols contaminés en métaux lourds*. Bruxelles : université libre de Bruxelles, 2010
- Vogt, W. P. (1993). *Dictionary of statistics and methodology : A nontechnical guide for the social sciences*. Newbury Park, CA : Sage.
- Wang, N., Wu, D., Yang, Y., et al., 2014. Exergy evaluation of a 600MWe supercritical coal-fired power plant considering pollution emissions. *Energy Procedia* 61, 1860- 1863.
- Weber, E., "Air pollution assessment modeling methodology", NATO, challenges of modern society, vol.2, Plenum press, (1982) : Chastain, J.P. 1999. *Air Quality and Odor Control from Swine Production Facilities*. chapter 9 in *Confined Animal Manure Managers Certification Program Manual*, Clemson University, Clemson SC, pp 9-1 to 9-11
- Wilson R, Spengler JD [Eds]. *Particles in our air. Concentrations and health effects*. Cambridge: Harvard University Press, 1996.
- Wilson, W.S.T., Tze, W.W., Andromeda, H.S.W., 2015. Association between air pollution and daily mortality and hospital admission due to ischaemic heart diseases in Hong Kong. *Atmospheric Environment* 120, 360-368.
- Xiu, Z., Nie, W., Yan, J., et al., 2020. Numerical simulation study on dust pollution characteristics and optimal dust control air flow rates during coal mine production. *J. Clean. Prod.* 248, 119197.
- Yahiaoui, G., Ariat, J.L., (2002) *Techniques avancées pour le traitement de l'information*. Edition Cépaduès, Paris.
- Yamani, Mounia El. *Environnement et milieu: pollution atmosphérique urbaine*. s.l. : Afsset, juin 2006
- Yan, B., Liu, S., Zhao, B., et al., (2018) : China's fight for clean air and human health. *Environ. Sci. Technol.* 52, 8063e8064.
- Yin, J., Harrison, R.M., 2008. Pragmatic mass closure study for PM1.0, PM2.5 and PM10 at roadside, urban background and rural sites. *Atmospheric Environment* 42, 980-988
- Zealand, C., Burn, H., Slobodan, Simonovic, P., (1999) : Short term streamflow forecasting using artificial neural network . *J. of hydrology* 1999 N° 214. Pp 32 - 48.
- Zhang, G., Sun, B., Song, S., Wang, H., Zhou, G., 2021. CFD comparative analysis on the pollution characteristics of coal dust under turbulent airflow from coal cutting in the fully mechanized mining face. *Process Safety and Environmental Protection*. 146 515-530.

Annexes

Résumé :

La première partie de la présente étude porte sur la modélisation et la prévision de la qualité de l'air de la région de Draria par les Particules en Suspension Totales (TSP) et les Eléments Traces associés (ET). La modélisation par intelligence artificielle par les réseaux de neurone artificiels a donné de bons modèles en termes d'exactitude et validation. Le réseau de neurones artificiels combiné proposé était capable de détecter, estimer et prévoir en un seul coup les TSP ainsi que les ET associés à partir des TSP prélevés seulement.

La deuxième partie porte sur l'étude de la pollution de l'air du même site d'étude par les gaz polluants, nous avons conçu à base d'Arduino une station intelligente d'auto-échantillonnage et de prélèvement en temps réel de onze (11) gaz polluants. L'étude journalière, hebdomadaire et mensuelle étalées sur six (06) mois ont révélées que la pollution de l'air au niveau de Draria n'est pas inquiétante. En dernière étape nous avons tracé sur un rayon d'un (01) kilomètre la cartographie de la pollution de l'air de la région de Draria.

Mot clés : Pollution atmosphérique, TSP, ET, Modélisation, réseaux de neurones artificiels, Arduino, Station de prélèvement intelligente.

Abstract :

The first part of this study focuses on the modeling and forecasting of air quality in the Draria region by Total Suspended Particles (TSP) and associated Trace Elements (ET). Artificial intelligence modeling by artificial neural networks has given good models in terms of accuracy and validation. The proposed combined artificial neural network was able to detect, estimate and forecast the TSPs as well as the associated ETs from the sampled TSPs only in one go.

The second part is about the study of the air pollution by the polluting gases of the same study site, we have designed, based on Arduino, an intelligent station of self-sampling and real-time sampling of eleven (11) polluting gases. The daily, weekly and monthly study spread over six (06) months revealed that air pollution in Draria is not worrying. In the last step we traced over a radius of one (01) kilometer the mapping of air pollution in the Draria region.

Keywords: Atmospheric pollution, TSP, ET, Modeling and forecasting, artificial neural networks, Arduino, Smart sampling station.

ملخص :

يركز الجزء الأول من هذه الدراسة على النمذجة والتنبؤ بجودة الهواء في منطقة درارية الناجمة عن الجسيمات المعلقة (TSP) والعناصر المعدنية المرتبطة بها (ET). أعطت النمذجة بالذكاء الاصطناعي بواسطة الشبكات العصبية الاصطناعية نماذج جيدة من حيث الدقة وصحة النتائج. كانت الشبكة العصبية الاصطناعية المشتركة المقترحة قادرة على اكتشاف وتقدير والتنبؤ بـ TSPs بالإضافة إلى ETs المرتبطة بـ TSPs التي اعتمدا فقط على TSP.

يتعلق الجزء الثاني بدراسة تلوث الهواء بالغازات الملوثة في نفس موقع الدراسة ، في هذا الصدد صممنا بـ Arduino محطة ذكية لأخذ عينات أحد عشر (11) غازًا ملوثًا ذاتيًا و في الوقت الفعلي. كشفت الدراسة اليومية والأسبوعية والشهرية الممتدة على مدى ستة (06) أشهر أن نسبة تلوث الهواء في درارية لا يدعو للقلق. في الخطوة الأخيرة ، قمنا برسم خريطة تلوث الهواء في منطقة درارية على امتداد كيلومتر واحد (01) كيلومترًا.

الكلمات المفتاحية: تلوث الجوي ، TSP ، ET ، النمذجة والتنبؤ ، الشبكات العصبية الاصطناعية ، Arduino ، محطة أخذ العينات الذكية.