



UNIVERSITE
JEAN LOROUGNON GUEDE

UFR ENVIRONNEMENT

REPUBLIQUE DE CÔTE D'IVOIRE

Union-Discipline-Travail

Ministère de l'Enseignement Supérieur et
de la Recherche Scientifique

ANNEE ACADEMIQUE :

2020-2021

N° D'ORDRE : 0411/2021

N°CARTE D'ETUDIANTE :

CI 0416008441

MASTER

Physique-Chimie

Option : Physique-Chimie appliquée à l'environnement

Spécialité : Modélisation et analyse statistique

THEME :

**MODELISATION DES CONCENTRATIONS DES OXYDES
D'AZOTES (NO_x) PAR LES MODELES AUTOREGRESSIFS A
CHANGEMENT DE REGIMES MARKOVIENS : CAS DE LA
STATION DE TOULOUSE-JACQUIER URBAIN**

LABORATOIRE DES
SCIENCES ET
TECHNOLOGIES DE
L'ENVIRONNEMENT

Présenté par :

BEH Mondjou Georges Constant

JURY

**Président : M. KOUASSI Kouakou Lazare, Professeur Titulaire,
Université Jean LOROUGNON GUEDE**

**Directeur : M. YODE Armel Fabrice Evrard, Maître de Conférences,
Université Félix Houphouët-Boigny**

**Encadreur : M. OKOU Gueï Cyrille, Maître-Assistant,
Université Jean LOROUGNON GUEDE**

Examineur : M. N'DRI Yao Aubin, Maitre-Assistant,

Université Jean LOROUGNON GUEDE

Soutenu publiquement
Le : 06/10/2021



UNIVERSITE
JEAN LOROUGNON GUEDE

UFR ENVIRONNEMENT

REPUBLIQUE DE CÔTE D'IVOIRE

Union-Discipline-Travail

Ministère de l'Enseignement Supérieur et
de la Recherche Scientifique

ANNEE ACADEMIQUE :

2020-2021

N° D'ORDRE : 0411/2021

N°CARTE D'ETUDIANTE :

CI0416008441

LABORATOIRE DES
SCIENCES ET
TECHNOLOGIES DE
L'ENVIRONNEMENT

MASTER

Physique-Chimie

Option : Physique-Chimie appliquée à l'environnement

Spécialité : Modélisation et analyse statistique

THEME :

**MODELISATION DES CONCENTRATIONS DES OXYDES
D'AZOTES (NO_x) PAR LES MODELES AUTOREGRESSIFS A
CHANGEMENT DE REGIMES MARKOVIENS : CAS DE LA
STATION DE TOULOUSE-JACQUIER URBAIN**

Présenté par :

BEH Mondjou Georges Constant

JURY

**Président : M. KOUASSI Kouakou Lazare, Professeur Titulaire,
Université Jean LOROUGNON GUEDE**

**Directeur : M. YODE Armel Fabrice Evrard, Maître de Conférences,
Université Félix Houphouët-Boigny**

**Encadreur : M. OKOU Gueï Cyrille, Maître-Assistant,
Université Jean LOROUGNON GUEDE**

**Examineur : M. N'DRI Yao Aubin, Maître-Assistant,
Université Jean LOROUGNON GUEDE**

Soutenu publiquement
Le : 06/10/2021

DEDICACES

Je dédie cette œuvre

À ma grande famille Spirituelle,

A mon père FEU BEH NIOULE GASTON,

Ma mère GBEHE DEOUE,

Mes frères et mes Sœurs pour leur soutien sans faille,

Ma grande famille BEH,

Mes enseignants du département Physique-Chimie,

À mon tuteur Daplé et sa famille,

Et à mes chers ami(e)s !

REMERCIEMENTS

Ce travail de mémoire de Master en Physique-Chimie, Option Physique-Chimie Appliquée à l'Environnement est le fruit de la contribution de plusieurs personnes à qui je dois toute ma reconnaissance.

Je voudrais exprimer ma considération distinguée à la Professeure **TIDOU Abiba Sanogo Epouse KONE**, Présidente de l'Université Jean Lorougnon GUEDE de Daloa. Merci pour la bonne gestion de notre institution.

Ma reconnaissance va à l'endroit de Monsieur **KOUASSI Kouakou Lazare**, Professeur Titulaire à l'Université Jean Lorougnon GUEDE de Daloa, Directeur de l'UFR Environnement. C'est grâce à vous que l'opportunité de réaliser cette œuvre s'est offerte à nous. Votre passion pour la rigueur, la discipline et l'abnégation dans le travail est pour nous un modèle pour la réussite.

Je remercie plus particulièrement **M. DIBI Brou**, Maître de Conférences à l'Université Jean Lorougnon GUEDE, Directeur du Laboratoire Science et Technologie, pour avoir accepté de nous accueillir dans son laboratoire.

Je voudrais exprimer ma très haute reconnaissance à mon encadreur, Docteur **OKOU Gueï Cyrille**, Maître-Assistant à l'Université Jean Lorougnon GUEDE de Daloa. Je vous remercie pour votre disponibilité, pour votre enthousiasme et pour l'attention particulière avec laquelle vous avez suivi et encadré ce travail. Auprès de vous, j'ai appris des vertus telles que la rigueur, la discipline, l'endurance et le travail bien fait.

Merci à **M. N'DRI Yao Aubin**, Maître-Assistant à l'Université Jean Lorougnon GUEDE, pour avoir accepté de diriger l'équipe d'encadrement de ce travail. Merci pour vos conseils et vos différentes critiques qui ont permis l'amélioration de ce travail.

Je remercie particulièrement le Docteur **KEITA Kolé**, Maître-Assistant à l'Université Jean Lorougnon GUEDE, pour son amour pour le travail et son affection pour moi. Merci pour avoir fait preuve de disponibilité quand besoin y était. Avec vous, je confirme encore que là où il y a la volonté, il y a aussi la possibilité.

À **M. YODE Armel Fabrice Evrard**, Maître de Conférences à l'Université Félix Houphouët-Boigny, un sincère remerciement, Directeur scientifique de ce mémoire. Votre générosité et votre personnalité forcent un grand respect et une admiration considérable de ma part.

Je remercie encore **M. KOUASSI Kouakou Lazare**, Professeur Titulaire et Directeur de l'UFR Environnement pour avoir accepté de présider le jury de soutenance.

Je veux remercier tous les enseignants du parcours physique-chimie de l'Université Jean Lorougnon GUEDE. Merci pour vos efforts et compétences qui m'ont permis d'acquérir de riches connaissances durant mon parcours dans cette Université.

Merci à mes parents, qui m'ont toujours soutenu dans la réalisation de mes études. Merci à ma mère **GBEHE Déoue**, à **Maman Roseline** pour tout, à mes grand-frères **BEH Michel et BEH Sepan Casimir** pour leur amour et fraternité sincère ; merci à ma grande sœur **Mme. Gollys Hilaire Epouse Ella** et sa fille **Florence**, pour son aide et sa disponibilité à toujours répondre à mes besoins. Grâce à vous j'ai toujours eu le courage de continuer mes études. Merci à **Maman MOHIRO Odette**, d'être toujours présente quand besoin y est.

Merci particulier à l'endroit de mes compagnons de tous les jours, il s'agit de mes amis **GNOMBLEI Jean-Luc, ZAMBLE Bi Ziblaho Franck Delore, BROU Konan Florent, ALLAH Kouadio Marc Junior...**

Merci à toute la quatrième promotion de la filière Physique-Chimie, à tous mes amis et connaissances qui m'ont soutenu ou pas.

TABLE DES MATIERES

	<u>Pages</u>
LISTE DES SIGLES ET ABREVIATIONS	iii
LISTE DES TABLEAUX.....	iv
LISTE DES FIGURES	v
INTRODUCTION	1
PREMIERE PARTIE : GENERALITES.....	4
1.1. Présentation de la zone d'étude.....	4
1.2. Généralités sur les NO_x.....	4
1.2.1. Définitions.....	4
1.2.2. Sources des oxydes d'azotes NO_x	5
1.2.3. Mécanismes de formation des oxydes d'azote	6
1.2.4. Les impacts des NO_x	7
1.3. Généralités sur les séries chronologiques.....	8
1.3.1. Définitions.....	8
1.3.2. Quelques propriétés des séries chronologiques.....	9
1.3.3. Présentation des modèles linéaires de séries chronologiques.....	11
1.3.3.1. Processus autorégressif AR(p)	11
1.3.3.2. Processus moyenne-mobile MA(q).....	12
1.3.3.3. Processus autorégressif à moyenne-mobile ARMA (p, q).....	12
1.3.4. Les modèles à changement de régimes	13
1.3.4.1. Les modèles autorégressifs à seuil à transition brute (TAR)	15
1.3.4.2. Les modèles autorégressifs à seuil avec transition retardé (SETAR)	16
DEUXIEME PARTIE : MATERIEL ET METHODES	18
2.1. Matériel	18
2.1.1. Données de l'étude.....	18
2.1.2. Matériels pour le traitement des données.....	18
2.2. Méthodes	20
2.2.1. Les Modèles à variable de transition cachée : Modèles autorégressifs à changement de régimes markovien (MSAR)	20
2.2.2. Les chaînes de Markov : Définitions, propriétés et exemples.....	20

2.2.3. Les modèles MS.....	25
2.2.4. Les modèles autorégressifs à changement de régimes markoviens (MS-AR) ..	25
2.2.5. Application aux données des concentrations des NO _x	27
2.2.5.1. Spécification du modèle.....	28
2.2.5.2. Méthodologie théorique : estimation des paramètres des modèles MSAR.	28
2.2.5.2.1. Fonction de vraisemblance	29
2.2.5.2.2. Algorithme EM (Espérance-maximisation)	29
2.2.5.2.3. Application de l'algorithme EM sur le modèle MS(2)-AR(1)	31
2.2.5.3. Validation des modèles MSAR.....	33
TROISIEME PARTIE : RESULTATS ET DISCUSSION.....	34
3.1. Résultats	34
3.1.1. Résultats des simulations.	34
3.1.2. Résultats sur l'application aux données de NO _x ($\mu\text{g}/\text{m}^3$).....	37
3.1.3. Résultats des estimations par l'algorithme EM	39
3.1.4. Formulation des modèles	42
3.1.5. Validation du modèle	44
3.2. Discussion	45
CONCLUSION... ..	46
REFERENCES... ..	47
ANNEXES.....

LISTE DES SIGLES ET ABREVIATIONS

E :	Ensemble des états
i :	Un état particulier
S_t :	Variable cachée de la chaîne de Markov
Y_t :	Variable observée
ACF :	AutoCorrelation Fonction
PACF :	Partial AutoCorrelation Fonction
AR :	AutoRégressif
MA :	Mobile-Average
ARMA :	AutoRégressif à Moyenne-Mobile
MS :	Markov Switching
MSAR :	Markov Switching Autoregressif
NO :	Monoxyde d'azote
NO ₂ :	Doxyde d'azote
TAR :	Threshold Autoregressifs
SETAR :	Self Excited Autoregressif
EMV :	Estimateur de Maximum de Vraisemblance
MV :	Maximum de Vraisemblance
EM :	Expectation-Maximisation
INDAF :	International Network to study Deposition and Atmospheric chemistry in Africa
MCMC :	Modèles à Chaîne de Markov Cachée
OMS :	Organisation Mondiale de la Santé
GES :	Gaz à Effet de Serre
N ₂ O ₄ :	Peroxyde d'azote
SO ₂ :	Dioxyde de soufre

LISTE DES TABLEAUX

Tableau I : Les principales caractéristiques des NO _x -----	7
Tableau II : Paramètres des simulations de modèles 1 et 2-----	34
Tableau III : Résultats de l'analyse descriptive -----	37
Tableau IV : Résultats de l'estimation des paramètres du modèle MS(2)-AR(1)-----	39
Tableau V : Résultats de validation du modèle.-----	44
Tableau VI : L'analyse statistique par régime (NO ₂)-----	44
Tableau VII : L'analyse statistique par régime (NO) -----	45

LISTE DES FIGURES

Figure 1 : Localisation de la zone d'étude.....	4
Figure 2 : Evolution schématique des NO _x dans l'atmosphère.....	5
Figure 3 : Principales voies de formation du monoxyde d'azote.....	6
Figure 4 : Modèle TAR (2,1) simulé	16
Figure 5 : Modèle SETAR (2,1) simulé.....	17
Figure 6 : L'interface de base de R sous Windows	19
Figure 7 : L'interface de base de RStudio sous Windows.....	19
Figure 8 : Représentation des probabilités d'une chaîne de Markov à deux états	23
Figure 9 : Représentation des probabilités d'une chaîne de Markov à trois états.....	24
Figure 10 : Graphe d'indépendance conditionnelle du modèle MSAR	26
Figure 11 : Récapitulatif des modèles à changement de régime selon l'observabilité de la variable de transition	27
Figure 12 : Graphe du Modèle 1 simulé	34
Figure 13 : Modèle 2 : MS(2)AR(1)	35
Figure 14 : Graphique du modèle 3.....	36
Figure 15 : Représentation du Modèle 4 : MS(3)AR(1).....	36
Figure 16 : Représentation graphique des concentrations de NO et NO ₂	37
Figure 17 : Histogrammes des concentrations de NO et NO ₂	38
Figure 18: PACF de NO et NO ₂	39
Figure 19 : Représentation des différents régimes dans la série de NO	41
Figure 20 : Représentation des différents régimes dans la série de NO ₂	41
Figure 21 : Représentation du Modèle et des données réelles du NO.....	43
Figure 22 : Représentation du Modèle et des données réelles du NO ₂	43

INTRODUCTION

INTRODUCTION

En raison de leurs conséquences immédiates et durables sur le milieu naturel et sur l'homme, les questions de changement et de variabilités climatiques sont placées depuis quelques années au centre des préoccupations des scientifiques et des décideurs politiques dans le monde. Les changements climatiques sont causés par les modifications de l'atmosphère qui résultent de sa transformation chimique par les gaz à effet de serre (GES). Cette perturbation de l'équilibre atmosphérique s'exprime par certains phénomènes extrêmes comme la sécheresse, l'inondation, les cyclones de plus en plus fréquents, une augmentation des températures moyennes etc., sur Terre. Ces phénomènes modifient ainsi ses caractéristiques physiques, chimiques et biologiques (Camirand et *al.*, 2011). En effet, compte tenu de certaines activités humaines ou de certains phénomènes, l'atmosphère se retrouve comme étant le lieu d'introduction de polluants gazeux et particulaires. En d'autres termes, la pollution atmosphérique se traduit par l'émission des gaz à effet de serre en concentration anormale, et cette pollution impacte négativement le climat et l'environnement. Ces émissions peuvent être d'origines naturelle (feux de forêt, éclairs, émissions biogéniques) ou anthropiques (combustion de charbon, de gaz, de mazout et d'essence, transformation du NO, etc.). Il faut aussi souligner que, l'urbanisation, l'exploitation accélérée des ressources minières et pétrolières, le caractère de plus en plus marqué des pratiques agricoles sont autant des facteurs contribuant à l'introduction intensive de divers composés chimiques dans l'atmosphère (Assamoi, 2011).

Les NO_x, qui font partie de ces polluants atmosphériques seront les plus abordés dans notre étude. Ils sont principalement formés à partir des deux gaz prépondérants de l'atmosphère, l'azote et l'oxygène, lorsque ceux-ci sont soumis à la chaleur intense caractérisant les processus de combustion (transports, combustions industrielles et domestiques, etc.). Ils peuvent être également émis par l'industrie chimique (fabrication d'acide nitrique par exemple), lors des opérations de soudages à l'arc, de la déflagration d'explosifs ou par la fabrication et l'utilisation d'engrais azotés. Ils constituent des polluants atmosphériques dans la mesure où ils transportent des espèces chimiques qui, de par leur composition, représentent un véritable danger pour l'homme et son environnement. Par exemple ils sont les constituants les plus importants qui interviennent dans le cycle de l'ozone troposphérique.

Au regard de ce grand défi qui a pour but de maintenir la vie sur Terre, se traduisant par une diminution considérable du taux d'émission de ces polluants, le suivi du niveau de leurs

concentrations dans l'atmosphère doit rester une priorité pour tout citoyen en général et pour les communautés scientifiques en particulier. Ainsi, en 1994, il fut mis sur pied, le réseau international d'observation INDAAF (International Network to study Deposition and Atmospheric chemistry in AFrica). Le réseau INDAAF est un réseau d'observation de la composition chimique de l'atmosphère sur le long terme. Le principal objectif de ce réseau est de suivre et de surveiller la composition chimique atmosphérique et les flux de dépôts secs et humides des polluants atmosphériques, tout en analysant la variation de leurs concentrations dans le temps (hebdomadaire, mensuelle, saisonnière et annuelle) en milieu rural. C'est un projet basé essentiellement sur l'atmosphère du continent africain.

Cependant, l'un des meilleurs moyens de suivre un phénomène physique, c'est sa prévision dans le temps. Ainsi des études sur la modélisation des concentrations des polluants ont été menées. Nous pouvons citer les travaux de Pélé (2020), dans « Modélisation et simulation aux grandes échelles des émissions de NO_x, dans un brûleur semi-industriel », et de Kljatic (2012) sur la « Modélisation des oxydes d'azote et des suies dans les moteurs Diesel ». Tous les deux traitent l'évolution des NO_x de sources anthropiques (industrie et transport) par les séries temporelles.

A notre connaissance cette manière de prédire l'atmosphère est peu développée. En effet, il s'agit d'une étude mathématique sur des observations passées ou présentes d'un processus afin de donner une information prédictive sur son comportement futur (on parle d'étude chronologique). Selon Girard (2011), une méthode de prévision très populaire est basée sur l'étude rigoureuse de séries chronologiques (séries temporelles), qui permet de prédire, par exemple de nombreux phénomènes naturels et financiers. Pour comprendre, l'étude formelle des séries chronologiques consiste à trouver, parmi plusieurs modèles mathématiques, celui qui explique le mieux possible les données observées. À partir de ce modèle, il est possible de faire de la prévision.

Cependant les séries des concentrations des NO_x ne sont pas linéaires. C'est-à-dire que l'on peut observer de fortes variabilités dans l'évolution des concentrations de ces polluants. La question que l'on se pose alors, est la suivante : quelle modélisation pour des variables de concentrations des NO_x, qui tient compte de ses changements structurels ?

Pour répondre à cette question, de nombreuses études ont fait appel à l'utilisation des modèles à changement de régimes, qui visent à comprendre le caractère asymétrique du changement dans la dynamique du phénomène. Les premiers auteurs à mettre sous une équation les modèles

à changement de régimes (Switching Models) sont Quandt & Goldfeld (1973) en économétrie. Ils ont constaté que les modèles traditionnels linéaires ne pouvaient pas être utilisés dans la modélisation des phénomènes aléatoires. Dans leur article, ils ont traité principalement deux types de changement de régimes, à savoir :

- Changement de régime déterministe (basé sur un indicateur temporel).
- Changement de régime stochastique (fait à travers une séquence aléatoire mais indépendante).

Par la suite, la découverte faite par Hamilton (1989), permet au modèle à changement de régimes de connaître un fort développement dans l'application des études économiques et financières. En effet, Hamilton (1989) a présenté dans son article, une extension du modèle développée par ces deux premiers auteurs selon deux voies :

- Il a considéré l'effet autorégressif dans le modèle.
- Le processus inobservable générateur du régime suit une chaîne de Markov.

Aujourd'hui, ces modèles sont bien développés dans la littérature.

C'est dans ce cadre que s'inscrit la présente étude. L'objectif de notre étude consiste à identifier les changements de régimes et modéliser l'évolution des concentrations NO_x par les modèles autorégressifs à changement de régimes Markoviens (MSAR).

Plus spécifiquement, il s'agira de montrer premièrement l'existence de plusieurs régimes dans les séries des concentrations des NO_x , puis construire un modèle qui prend en compte cette non-linéarité.

Pour atteindre notre objectif, notre mémoire sera subdivisé en trois parties structurées comme suit :

La première partie sera consacrée à la généralité sur l'ensemble du travail. C'est-à-dire une étude bibliographique sur la pollution atmosphérique premièrement, puis sur les modèles de séries chronologiques.

La seconde partie intitulée « matériel et méthodes », est la partie pratique du mémoire. Il présentera dans un premier temps le matériel de travail, pour donner ensuite les méthodes d'analyse des données.

Enfin, dans **la troisième partie**, les résultats issus de l'analyse précédente seront présentés, et discutés. Le mémoire se terminera par une conclusion et les perspectives.

PREMIERE PARTIE :
GENERALITES

PREMIERE PARTIE : GENERALITES

1.1. Présentation de la zone d'étude

Les réseaux de mesure assurent le suivi des mesures physiques (vent, température...) et chimiques (concentrations de polluants). Généralement, chaque région française en possède au moins un. Ils gèrent d'importantes bases de données à la disposition de la population et des pouvoirs publics. Pour mener notre étude, nous avons utilisé les données de la région d'Occitanie à Toulouse (France). Nous avons choisi alors la station Toulouse-Jacquier Urbain.

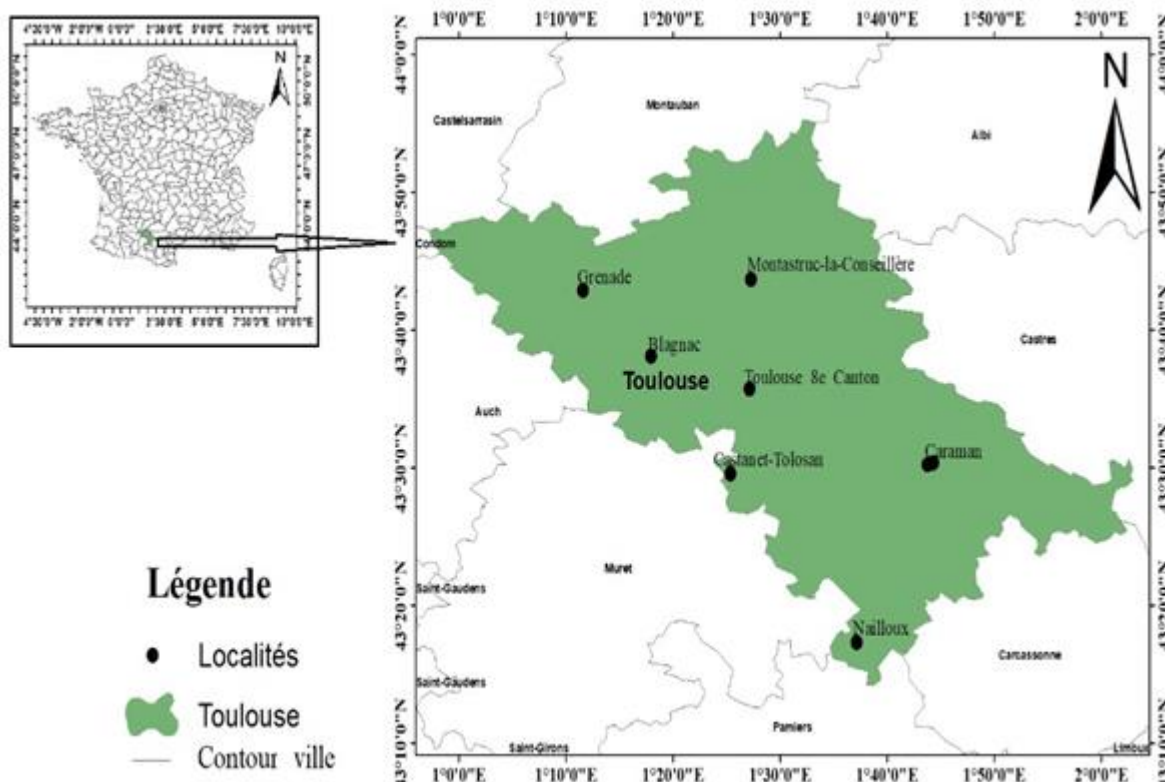


Figure 1 : Localisation de la zone d'étude

1.2. Généralités sur les NO_x

1.2.1. Définitions

Le monoxyde d'azote et le dioxyde d'azote sont les constituants les plus importants qui interviennent dans le cycle de l'ozone troposphérique. Ils sont rassemblés sous la dénomination NO_x. Le dioxyde d'azote (NO₂) est un gaz brun rougeâtre, plus lourd que l'air. Il a une odeur âcre et déplaisante. A température ordinaire, il s'agit plutôt d'un mélange de NO₂ et de peroxyde d'azote (N₂O₄) en équilibre. Les vapeurs de peroxyde d'azote ont un caractère oxydant qui

explique leur toxicité. Le NO_2 , est relativement peu soluble dans l'eau où il réagit pour donner de l'acide nitrique (HNO_3) et des nitrates organiques dont l'effet est toxique.

Le monoxyde d'azote (NO) est le plus abondant des oxydes d'azote. Il a une durée de vie de quelques heures dans l'atmosphère. Il est rapidement oxydé par l'ozone, les radicaux HO_2 et RO_2 pour donner du dioxyde d'azote. La figure 2 ci-dessous représente l'évolution des NO_x dans l'atmosphère.

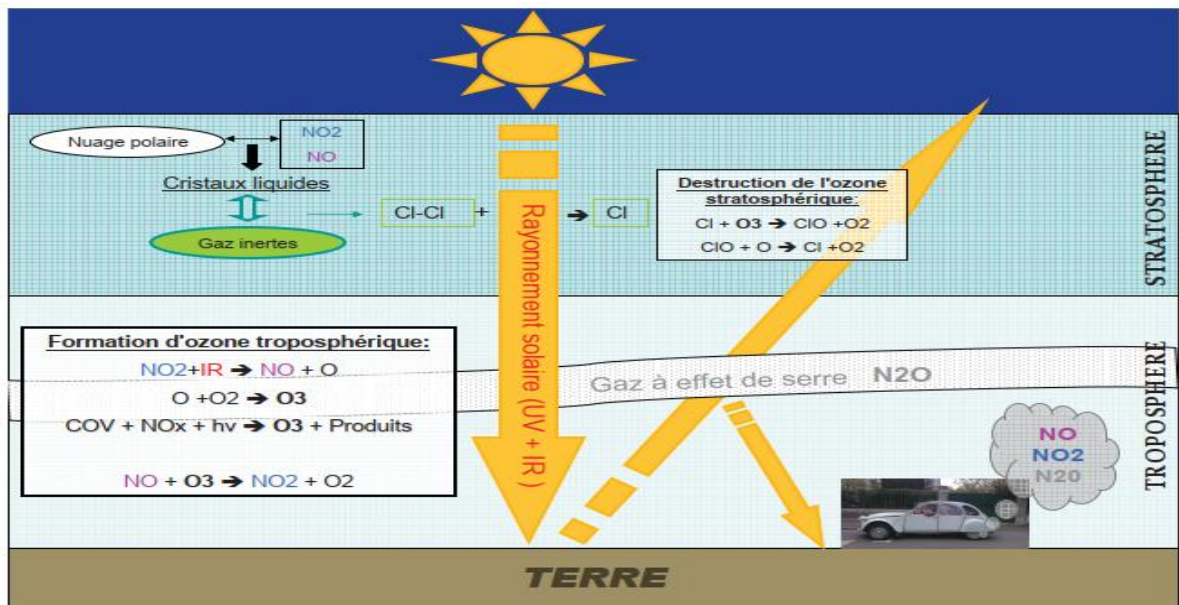


Figure 2 : Evolution schématique des NO_x dans l'atmosphère (Kljahic, 2012)

La troposphère est le siège de transferts d'énergie entre les masses d'air essentiellement par convection thermique générée par les échanges de chaleur entre la surface et l'atmosphère. Dans la troposphère, les NO_x favorisent la formation de l'Ozone.

La chimie troposphérique concerne l'étude des processus qui se produisent dans la troposphère et par lesquels des composés à l'état de traces, provenant de sources naturelles et anthropiques, sont notamment oxydés.

1.2.2. Sources des oxydes d'azotes NO_x

Les oxydes d'azote ont des sources variées, dont plus de 75% sont d'origine anthropique. Au niveau mondial, les émissions d'oxydes d'azote dues à la combustion des combustibles fossiles ont augmenté d'un facteur 10 entre 1900 et 1990. En Europe (CEE), ces émissions proviennent pour 36% des transports, 27% des centrales électriques et 31% de l'industrie. En France, les émissions proviennent pour 76% des transports. Une autre source anthropique, particulièrement importante à l'échelle globale, est liée à la combustion de la biomasse dans les régions tropicales

et subtropicales. Enfin, le développement de l'aviation subsonique entraîne une augmentation des rejets de monoxyde de carbone, d'hydrocarbures et d'oxydes d'azote.

Une classification supplémentaire est généralement utilisée pour différencier l'importance des flux d'émissions. Les sources sont réparties en trois types :

- Les sources ponctuelles qui comprennent, d'après les critères des Directions Régionales de l'Industrie, de la Recherche et de l'Environnement (DRIRE), les émetteurs de type industriel quantitativement les plus importants.
- Les sources linéaires constituées par les principaux axes routiers, ferroviaires, fluviaux et les voies maritimes le cas échéant.
- Les sources surfaciques qui regroupent l'ensemble des sources qui ne sont ni ponctuelles ni linéaires. Les flux d'émissions de ce type de sources sont plus diffus.

Ainsi, le trafic routier est traité à la fois comme une source linéaire, pour les principaux axes de circulation, et surfacique en ce qui concerne le réseau urbain des rues et des voies secondaires en agglomération.

Avant de se prononcer sur les impacts des NO_x , dans la section suivante nous présentons les différents mécanismes de formation de ces polluants.

1.2.3. Mécanismes de formation des oxydes d'azote

Les mécanismes de formation des NO_x durant la décomposition d'hydrocarbures sont complexes et certaines voies de formation sont encore mal comprises. Ces NO_x proviennent de réactions entre l'azote (de l'air ou du combustible) et l'oxygène de l'air.

➤ Formation du NO

On distingue quatre grandes voies de formation du NO, représentées sur la figure suivante :

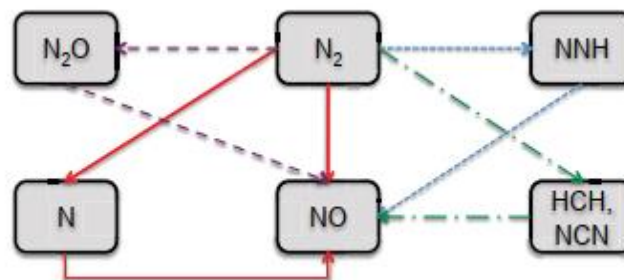
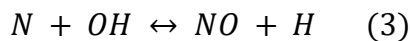


Figure 3 : Principales voies de formation du monoxyde d'azote (Kljahic, 2012)

La principale voie de formation des oxydes d'azote dans les moteurs à combustion interne est celle présentée en rouge, qualifiée de voie thermique. C'est le mécanisme de formation de NO à partir de l'azote moléculaire. La vitesse de production de NO par cette voie devient significative à haute température et en présence d'oxygène (donc pour des mélanges pauvres ou stœchiométriques).

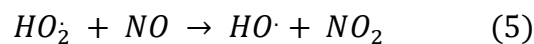
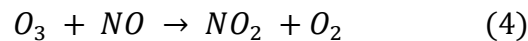
Le mécanisme considéré est le suivant :



La réaction 1 constitue l'étape limitante du mécanisme car elle présente une énergie d'activation élevée (environ 320 KJ/mol) et n'est donc rapide qu'à haute température, ce qui justifie la dénomination de "mécanisme thermique". Pour plus de détails sur les autres voies de formation de la figure 3, voir Kljakic (2012).

➤ Formation du NO₂

La réaction qui contrôle la formation du dioxyde d'azote (NO₂) est la réaction entre le NO formé et soit l'ozone (réaction rapide) ou des radicaux hydro peroxyde HO₂ et méthylperoxyde CH₃O₂· par les équations suivantes :



1.2.4. Les impacts des NO_x

En dehors de leurs effets nocifs sur la santé, les composés azotés ont des répercussions sur l'environnement. Le tableau ci-dessous donne un résumé sur les NO_x (sources et impacts).

Tableau I : Les principales caractéristiques des NO_x (Ouaret, 2016 ; Pélé, 2020)

Polluant	Sources	Impacts sur l'environnement	Impacts sur la santé
Oxydes d'azote (NO _x)	-Anthropiques : Combustion de charbon, de gaz, de fioul et d'essence, Combustion de charbon, de mazout et d'essence, transformation du NO. -Naturelles : Feux de forêt, processus anaérobies dans les sols, éclairs.	-Rôle de précurseur dans la formation d'ozone dans la basse atmosphère (troposphérique) -Contribue à la formation des pluies acides -Contribuent à la concentration de nitrates dans les sols.	-NO : non toxique pour l'homme aux concentrations environnementales -NO ₂ : gaz irritant pour les bronches (augmentation de l'asthme et des infections pulmonaires)

Le monoxyde d'azote NO en lui est peu toxique, il est cependant connu pour avoir un impact sur la fixation de l'oxygène par l'hémoglobine à partir de 15 ppm. En revanche, le NO est un précurseur du dioxyde d'azote qui est lui un gaz toxique. Le NO₂ inhalé peut être absorbé dans les alvéoles pulmonaires, il pénètre profondément dans le système respiratoire entraînant des altérations des structures alvéolaires. Les symptômes de l'action toxique du NO₂ sont l'insomnie, la toux, et une altération des muqueuses.

Après une connaissance générale sur les oxydes d'azote, nous allons présenter les outils mathématiques qui permettront la mise en œuvre de leur modélisation dans la section suivante.

1.3. Généralités sur les séries chronologiques

Dans cette section, nous introduisons d'abord quelques définitions et propriétés des séries temporelles, ainsi que leurs applications aux données des NO_x. Notre objectif est de donner un aperçu sur chacune des principales caractéristiques de l'analyse des séries temporelles et de les appliquer sur les données.

1.3.1. Définitions

Une série chronologique est constituée de valeurs observées à des intervalles de temps réguliers. Par exemple, les débits annuels sur un cours d'eau ou encore les valeurs mensuelles de titres

boursiers, les données météorologiques tels que la température, l'intensité du vent... sont des séries chronologiques (Girard, 2011).

On peut aussi définir une série chronologique comme une suite de nombres réels, indexés par les entiers relatifs tels que le temps. Pour chaque instant du temps, la valeur de la quantité étudiée Y_t est appelée variable aléatoire. L'ensemble des valeurs Y_t quand t varie, est appelé processus aléatoire :

$$\{Y_t, t \in T\}$$

1.3.2. Quelques propriétés des séries chronologiques

Dans cette partie, plusieurs concepts importants liés à l'analyse de séries chronologiques seront abordés. Parmi ceux-ci, on retrouve les notions de stationnarité, d'autocorrélation et de bruit blanc.

❖ La stationnarité

La stationnarité est une caractéristique d'une série chronologique qui implique que le comportement de la série est indépendant du temps. En particulier, on dit qu'une série est stable si elle ne comporte pas de tendance ni effet de saison. Plus formellement, on distingue deux types de stationnarité, à savoir stationnarité forte (stationnarité au sens strict) et stationnarité faible (stationnarité au sens large).

- Un processus aléatoire est dit stationnaire au sens strict si la structure de probabilité est invariante par translation dans le temps :

$$\forall t \in T, f(Y_1, Y_2, \dots, Y_t) = f(Y_{t+h}, Y_{t+h}, \dots, Y_{t+h}), \forall h \in \mathbb{Z}.$$

- Un processus aléatoire est dit faiblement stationnaire si :

$$E(Y_t) = \mu, \forall t \in T$$

$$V(Y_t) = \sigma^2, \forall t \in T$$

$$Cov(Y_{t+h}, Y_t) = \gamma_h, \forall t, h \in \mathbb{Z}$$

Habituellement, la notion de stationnarité au sens strict est beaucoup trop forte pour les applications et il est suffisant de considérer des modèles faiblement stationnaires.

❖ Fonctions d'autocovariance et fonction d'autocorrélation

Dans le processus aléatoire ci-dessus $\{Y_t\}$, la variable aléatoire au temps t dépend généralement, d'une certaine façon, des observations précédentes $\{Y_{t-1}, Y_{t-2}, \dots, Y_1\}$. Ainsi, lorsqu'on travaille

sur un nombre fini de variables aléatoires, nous cherchons souvent les dépendances ou des structures de corrélations entre les différents retards des observations (Oualet, 2016).

Une manière de comprendre les liens d'interdépendance entre les termes d'une telle série consiste à considérer les fonctions d'autocovariance et d'autocorrélation. Ces fonctions mesurent, respectivement, la covariance et la corrélation entre les termes d'une série chronologique.

➤ Fonction d'autocovariance $\gamma(k)$

Soit $\{Y_t\}$ un processus aléatoire tel que la variance $V(Y_t) \leq \infty$;

La fonction d'autocovariance au retard k , notée $\gamma(k)$, du processus $\{Y_t\}$ est définie de la manière suivante, pour tout $t \in \mathbb{Z}$ et $k \in \mathbb{Z}$:

$$\begin{aligned}\gamma(k) &= \text{cov}(Y_t, Y_{t+k}) \\ &= E[(Y_t - E(Y_t))(Y_{t+k} - E(Y_{t+k}))]\end{aligned}$$

➤ Fonction d'autocorrélation $\rho(k)$

La fonction d'autocorrélation ACF en anglais (AutoCorrelation Function) au retard k , notée $\rho(k)$, du processus $\{Y_t\}$ est définie de la manière suivante, pour tout $t \in \mathbb{Z}$ et $k \in \mathbb{Z}$:

$$\rho(k) = \frac{\gamma(k)}{\sigma_{Y_t} \sigma_{Y_{t+k}}} \quad (7)$$

Où σ_{Y_t} est l'écart type du processus au temps t , tel que $\sigma_{Y_t} = \sqrt{\gamma(0)}$

Ainsi, pour un retard k fixé, le nombre $\rho(k) \in [-1, 1]$ mesure la corrélation linéaire entre les variables Y_t et Y_{t+k} .

En particulier, on remarque que $\rho(0) = 1$. Dans une optique prévisionnelle, on s'attachera à mettre en évidence les retards k pour lesquels l'ACF est la plus élevée.

❖ Fonction d'autocorrélation partielle

D'autres mesures de dépendance entre variables existent en statistique. Un outil de diagnostic intéressant est la fonction d'autocorrélation partielle, que l'on note PACF (Partial AutoCorrelation Function). La PACF au retard k , notée $r(k)$, est définie pour tout $k \in \mathbb{Z}$, de la manière suivante :

$$r(k) = \frac{\text{cov}(Y_t - Y_t^*, Y_{t+k} - Y_{t+k}^*)}{\sqrt{\text{var}(Y_t - Y_t^*) \text{var}(Y_{t+k} - Y_{t+k}^*)}} \quad (8)$$

Avec Y_t^* est la régression affine de Y_t sur $Y_{t+1}, Y_{t+2}, \dots, Y_{t+k-1}$ et Y_{t+k}^* est la régression affine de Y_{t+k} sur $Y_{t+k-1}, Y_{t+k-2}, \dots, Y_{t+1}$.

Alors, pour un retard k fixé, le nombre $r(k)$ est le coefficient de corrélation linéaire entre la variable $Y_t - E(Y_t | Y_{t+1}, Y_{t+2}, \dots, Y_{t+k-1})$ et la variable $Y_{t+k} - E(Y_{t+k} | Y_{t+1}, \dots, Y_{t+k-2}, Y_{t+k-1})$. En effet, la fonction d'autocorrélation partielle $r(k)$ permet de mesurer la corrélation entre les variables Y_t et Y_{t+k} en ôtant toutes les observations intermédiaires. Cette mesure de corrélation est délicate à interpréter.

❖ Bruit blanc

Une séquence ε_t est un bruit blanc de moyenne nulle et de variance σ_ε^2 si :

$$\begin{aligned} E(\varepsilon_t) &= 0, \\ E(\varepsilon_t^2) &= \gamma_\varepsilon(0) = \sigma_\varepsilon^2 < +\infty, \\ \gamma(h) &= 0, \forall h \in \mathbb{Z} \text{ pour } h \neq 0. \end{aligned}$$

Par conséquent, le comportement d'un bruit blanc au temps t n'a aucune incidence sur celui-ci au temps $t + h$. On parle de bruit blanc gaussien lorsque les variables aléatoires $\{\varepsilon_t\}$ sont centrées, indépendantes et identiquement distribuées (iid) de variance finie. On le notera $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$.

Il existe plusieurs caractéristiques qui permettent d'appréhender la notion de série chronologique. En dépit de celles décrites dans cette section nous en découvrirons d'autres dans la suite de ce mémoire. Dans ce qui suit, nous présentons quelques modèles linéaires généralement utilisés dans l'analyse des séries chronologiques.

1.3.3. Présentation des modèles linéaires de séries chronologiques

Dans cette partie, on s'intéresse à une brève présentation des processus linéaires des séries chronologiques. Nous effectuons quelques rappels sur les processus de type autorégressif d'ordre p notés AR(p), moyenne-mobile d'ordre q ou MA(q) et les modèles autorégressifs à moyenne-mobile ou ARMA (p, q) qui est la combinaison des deux premiers.

1.3.3.1. Processus autorégressif AR(p)

On appelle processus autorégressif d'ordre p , usuellement noté AR(p), un processus stationnaire dont la réalisation des observations au temps t (Y_t), dépend des réalisations passées.

Il vérifie une relation du type :

$$\begin{aligned} Y_t &= \sum_{i=1}^p \varphi_i Y_{t-i} + \varepsilon_t, \forall t \in \mathbb{Z} & (9) \\ Y_t &= \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \varepsilon_t \end{aligned}$$

Le processus peut se mettre aussi sous la forme :

$$\Phi(B)Y_t = \varepsilon_t \quad (10)$$

Avec $\varphi_i \in \mathbb{R}$, ε_t est un bruit blanc tel que défini plus haut, de variance σ^2 et $\Phi(B)$ est un polynôme caractéristique du degré p définie par :

$$\Phi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p$$

$$\Phi(B) = 1 - \sum_{i=1}^p \varphi_i B^i, \text{ on appelle } B \text{ l'opérateur retard.}$$

1.3.3.2. Processus moyenne-mobile MA(q)

On appelle processus moyenne-mobile d'ordre q , usuellement noté MA(q) (Moving average), un processus (Y_t) défini par :

$$Y_t = \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t, \forall t \in \mathbb{Z} \quad (11)$$

$$Y_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t, \forall t \in \mathbb{Z}$$

Où $\theta_i \in \mathbb{R}$ et ε_t est un bruit blanc de variance σ^2 . Comme dans les processus AR(p), on peut également obtenir une autre formulation du processus MA(q) en appliquant l'opérateur retard.

1.3.3.3. Processus autorégressif à moyenne-mobile ARMA (p, q)

Les processus ARMA (p, q) généralisent les modèles autorégressifs et moyennes mobiles. Ces modèles sont très utiles en pratique pour modéliser des séries réelles en nécessitant moins de paramètres que les modèles AR(p) ou MA(q) simples. Alors un processus stationnaire Y_t admet une représentation ARMA (p, q) s'il satisfait l'équation suivante :

$$Y_t + \sum_{i=1}^p \varphi_i Y_{t-i} = \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i}, \forall t \in \mathbb{Z} \quad (12)$$

Avec $\varphi_p \neq 0, \theta_q \neq 0$ et ε_t est un bruit blanc de variance σ^2 .

Les modèles AR, MA et ARMA aussi appelés modèles de Box-Jenkins, peu développés ci-dessus permettent de modéliser des processus linéaires et stationnaires, c'est-à-dire les processus dont les caractéristiques (moyenne, variance, etc.) ne varient pas avec le temps.

Selon Grégoire (2019), les modèles linéaires classiques sont appropriés et suffisants dans plusieurs applications puisqu'ils parviennent généralement à adéquatement représenter les données macroéconomiques ou encore observées à des espaces temporels relativement grands (mois, années, etc.).

Cependant, il existe certains cas où ces modèles ne peuvent pas bien représenter le vrai modèle sous-jacent aux données, particulièrement dans le cas où il y a plusieurs régimes.

Un régime est par définition la structure dans le temps et l'espace d'un système économique ou physique. Par exemple lorsque l'on veut prévoir l'évolution de certains phénomènes, tel que la propagation d'une nappe de polluant, il est nécessaire de connaître les paramètres météorologiques qui influencent sur sa structure.

Ainsi, la présence de plusieurs régimes sous-entend que les données comportent plusieurs structures sous-jacentes dans le temps et l'espace. Cela se voit notamment en économie, où plusieurs types de jeux de données comportent deux régimes ou plus, avec une transition entre les deux. Il convient alors d'utiliser un modèle plus approprié que les modèles linéaires classiques pour ces données, ce qui nous mène à l'étude des modèles non-linéaires pouvant modéliser les séries temporelles. Les modèles permettant de modéliser ces types de séries sont reconnu sous le nom de « modèles à changement de régimes ».

1.3.4. Les modèles à changement de régimes

Comme nous venons de le voir à la section précédente, les modèles linéaires comportent des limites pouvant restreindre leur pertinence d'application. Il est ainsi justifié d'étendre l'étude aux modèles non-linéaires. Cette section concerne la présentation des modèles non-linéaires des séries temporelles, particulièrement dans notre application (par l'évolution des polluant au cours d'une journée peut être influencée par la température du milieu considéré), dont la modélisation par les équations linéaires est insuffisante. Ceci peut en effet survenir lorsque, pour des raisons structurelles, les valeurs de certains paramètres se modifient au cours du temps de manière continue ou discrète. Le modèle linéaire devenant ainsi inadapté.

Le vrai modèle peut alors contenir des facteurs exogènes dont l'influence change d'une période à l'autre, suggérant ainsi les modèles non linéaires. Les modèles non linéaires permettent de définir différents états de la nature ou régimes. La dynamique de la variable étudiée (moyenne, variance, autocorrélation) dépend du régime dans lequel elle se trouve.

Les modèles à changement de régime décrivent un ensemble d'états dans lesquels chaque portion de variabilité se révèle explicative du phénomène environnemental étudié. Si la succession de ces phases de variabilité se manifeste de façon régulière, celle-ci peut être attribuée à une forme de saisonnalité. Dans la littérature, il existe plusieurs modèles à changement de régimes. Ces derniers regroupent deux types : Les modèles à changement de régimes déterministes introduits par Tong, tels que les modèles TAR (Threshold Autoregressive), et les modèles SETAR (Self Excited Threshold Autoregressive), qui reposent tous les deux sur la définition d'une variable de seuil connue, et les modèles à changement de

régimes markovien (aléatoires ou stochastiques) caractérisés par la méconnaissance de la variable du changement de régime.

Les modèles à changement de régimes aléatoires regroupent plusieurs variantes : (i) les Modèles de Markov Cachés (Hidden Markov Models) qui supposent que la séquence d'observations dépend d'une séquence d'état non observée cachée tirée d'une chaîne de Markov, (ii) les modèles de mélange de distributions (Mixtures Models) qui permettent de prendre en compte l'effet de l'hétérogénéité dans la structure des observations en la départageant en sous-groupes indépendants entre eux du fait que le processus générateur du changement de régime est un processus indépendant, et enfin, (iii) les modèles autorégressifs à changement de régimes Markoviens (MSAR) pour « Markov Switching Autoregressive » qui permettent de capter la majorité des effets non linéaires ainsi que la structure auto-corrélée qui existe dans la plupart des séries des NO_x. Dans notre étude, nous traitons ces modèles non linéaires, mais une particularité sera portée sur les modèles markovien en raison de leur degré de représentativité des faits stylisés non linéaires qui existent dans les domaines de la pollution (Ilham et Nadjet, 2020).

➤ Mise en œuvre du modèle à changement de régime

Considérons l'évolution de la concentration des NO_x, soit le processus aléatoire Y_t , son modèle on a :

$$Y_t = m_t + \varepsilon_t, \forall t \in [1, T] \quad (13)$$

Où m_t représente sa moyenne dans le temps et ε_t est un bruit blanc de variance σ^2 .

Supposons que la variable m_t suit plusieurs comportements sur la période analysée $[1, T]$, on obtient alors un changement d'état dans l'évolution de la concentration de NO_x. Supposons qu'il existe deux régimes gouvernés par une variable (concentration), par exemple un régime où on observe les fortes valeurs de concentration et un régime de basses concentrations. Donc m_t dépend du régime sur lequel le processus se trouve.

Soit $S_t \in \{1, 2\}$ la variable atmosphérique, représentant les régimes au temps t . Dans ce cas la moyenne sera m_{s_t} . L'équation (13) devient :

$$Y_t = m_{s_t} + \varepsilon_t, \quad \forall t \in [1, T] \quad (14)$$

Avec $\varepsilon_t \sim N(0 ; \sigma^2)$

L'équation (14) représente le modèle à changement de régime le plus simple où le changement est observé dans la moyenne. Même si Y_t suit une tendance linéaire dans chaque régime, le modèle (14) est considéré non linéaire. La variable S_t peut être observable comme ne pas l'être.

En fonction du type de comportement observé sur le modèle (14), plusieurs modèles à changement de régime existent dans la littérature. Dans cette étude, nous présentons plusieurs d'entre eux par simulation, un seul sera appliqué à nos données.

1.3.4.1. Les modèles autorégressifs à seuil à transition brute (TAR)

Les modèles autorégressifs à seuil à transition brutale TAR (Threshold Auto-Regressive) supposent une transition entre les régimes par une fonction de transition indicatrice signalée par une variable de transition. Ce type de modèle arbore un comportement incorporant des changements de régime liés au dépassement d'un seuil c par une variable de transition exogène observée que nous représentons ici par $S_t = Y_{t-1}$ (Fouquau, 2008).

➤ Présentation du modèle

Soit le processus $(Y_t)_{t \in \mathbb{Z}}$; l'écriture générale d'un processus TAR pour $M=2$ régimes (pour la simplicité), notée TAR $(M=2, p)$ est la suivante :

$$Y_t = \begin{cases} m_1 + \sum_{i=1}^p \varphi_{i,1} Y_{t-i} + \varepsilon_t & \text{si } Y_{t-1} \leq c \\ m_2 + \sum_{i=1}^p \varphi_{i,2} Y_{t-i} + \varepsilon_t & \text{si } Y_{t-1} > c \end{cases} \quad (15)$$

On peut réécrire ce modèle sous sa forme alternative comme suit :

$$Y_t = \left(m_1 + \sum_{i=1}^p \varphi_{i,1} Y_{t-i} \right) (1 - I(Y_{t-1}, c)) + \left(m_2 + \sum_{i=1}^p \varphi_{i,2} Y_{t-i} \right) I(Y_{t-1}, c) + \varepsilon_t$$

où la fonction indicatrice

$$I(Y_{t-1}, c) = \begin{cases} 0 & \text{si } Y_{t-1} \leq c \\ 1 & \text{si } Y_{t-1} > c \end{cases} \text{ et } \varepsilon_t \text{ est un bruit blanc de variance } \sigma^2.$$

Tant que la concentration du polluant ne dépasse pas la valeur du seuil c , on dira qu'elle se trouve dans le régime $S_t = 1$. Au-delà de ce seuil, nous observons les valeurs qui appartiennent au régime $S_t = 2$, des concentrations du polluant.

Sur la figure ci-dessous est représenté un exemple du processus TAR (2,1) simulé avec les paramètres $\varphi_{1,1} = 0,21$, $\varphi_{1,2} = -0,20$, $m_1 = 21$ et $m_2 = 39$, $c = 500$ la taille de l'échantillon $N=1000$.

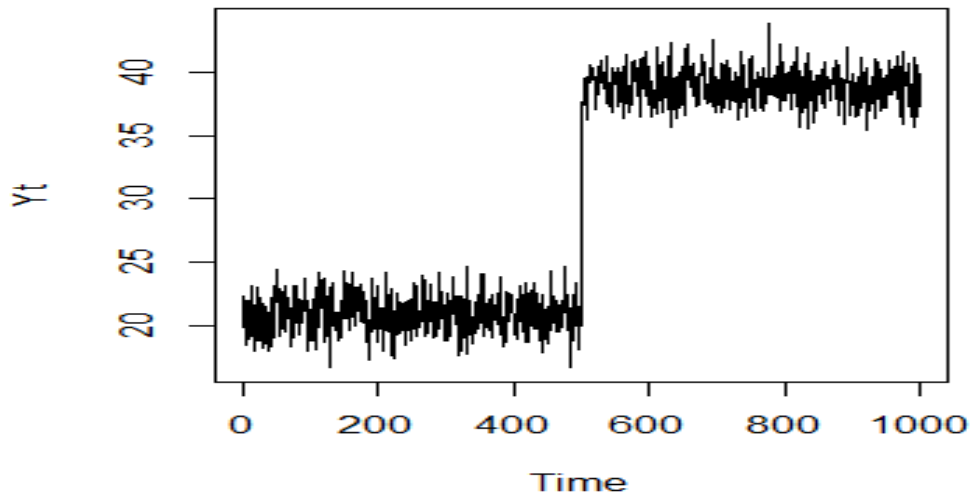


Figure 4 : Modèle TAR (2,1) simulé (sous Rstudio)

L'analyse visuelle du graphique montre que l'évolution des données présente deux régimes. Et le passage d'un régime à l'autre se fait de manière brusque selon le signe de Y_{t-1} par rapport au seuil c .

1.3.4.2. Les modèles autorégressifs à seuil avec transition retardé (SETAR)

Dans le cas où la variable observée Y_t est retardée de $h > 0$ et alors $S_t = Y_{t-h}$, on obtient le modèle SETAR (Self-Exciting Threshold AutoRegressive Model). Le modèle SETAR est un prolongement du modèle de classe TAR (Fofana & Diongue, 2013). L'idée sous-jacente de ce type d'approche est qu'une relation peut être non linéaire sur la période globale considérée mais linéaire par sous périodes. La forme du modèle d'ordre p à deux régimes, SETAR (2, p) d'un processus Y_t , s'écrit :

$$Y_t = \left(m_1 + \sum_{i=1}^p \varphi_{i,1} Y_{t-i} \right) (1 - I(Y_{t-h}, c)) + \left(m_2 + \sum_{i=1}^p \varphi_{i,2} Y_{t-i} \right) I(Y_{t-h}, c) + \varepsilon_t \quad (16)$$

Où $I(Y_{t-h}, c) = \begin{cases} 0 & \text{si } Y_{t-h} \leq c \\ 1 & \text{si } Y_{t-h} > c \end{cases}$, et ε_t est un bruit blanc de variance σ^2 .

Une caractéristique importante du modèle SETAR est que le changement de régime est déterminé par des valeurs retardées de la série. La figure ci-dessous représente un processus SETAR (2,1) simulé avec les paramètres $\varphi_{1,1} = 0,27, \varphi_{1,2} = -0,18, m_1 = -1$ et $m_2 = 1$, $c=500$ et $N=1000$.

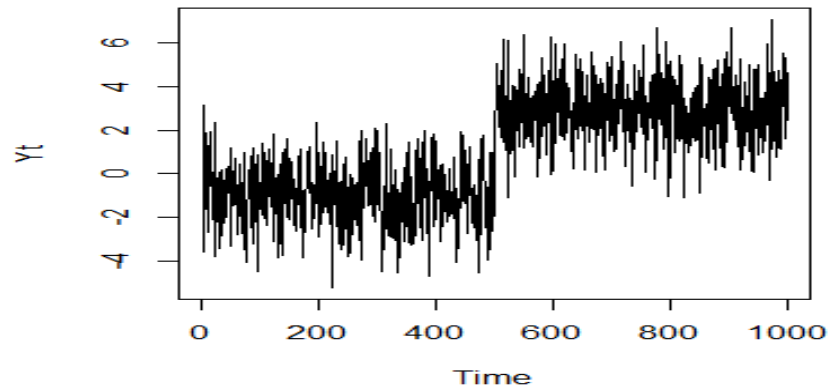


Figure 5 : Modèle SETAR (2,1) simulé (sous Rstudio)

La présente partie, nous a permis d'avoir un aperçu général sur notre travail, c'est-à-dire les notions de pollution par les oxydes d'azote (NO_x) et des séries chronologiques. Dans ce qui suit, nous abordons le matériel de travail et les méthodes utilisées pour l'atteinte de l'objectif de cette étude.

DEUXIEME PARTIE :
MATERIEL ET METHODES

DEUXIEME PARTIE : MATERIEL ET METHODES

2.1. Matériel

2.1.1. Données de l'étude

Les réseaux de surveillance de la qualité de l'air en France sont très majoritairement équipés pour effectuer des mesures au sol. La mesure des polluants est effectuée au moyen d'analyseurs chimiques automatisés. Ces analyseurs délivrent généralement des données qui sont validées par les postes centraux des réseaux. Ainsi, pour notre étude, une base de données journalières dans la région d'Occitanie en France, des concentrations ($\mu g. m^{-3}$) des NO_x (NO et NO_2) à la station de Toulouse-Jacquier Urbain sur une période d'un an (mesure prise du 19/07/2020 à 23h00 min au 11/07/2021 à 23h00 min) a été utilisée. En effet dans cette étude, nous avons utilisé 720 mesures des concentrations des NO_x , avec 355 données pour la concentration NO et 355 données pour le NO_2 .

La section suivante concerne le matériel de traitement de ces données.

2.1.2. Matériels pour le traitement des données

Dans ce mémoire, trois principaux logiciels sont utilisés pour le traitement des données. Il s'agit des logiciels Excel version 2016, R version 3.5.1 et de RStudio.

➤ Logiciel R

R est un langage orienté vers le traitement de données et l'analyse statistique dérivé du langage **S**. Il est développé depuis une vingtaine d'années par un groupe de volontaires de différents pays. C'est un logiciel libre, publié sous licence GNU GPL. Son utilisation présente plusieurs avantages, par exemple :

- C'est un logiciel multiplateforme, qui fonctionne aussi bien sur des systèmes Linux, Mac OS X ou Windows ;
- C'est un logiciel libre, développé par ses utilisateurs et modifiable par tout un chacun ;
- C'est un logiciel gratuit ;
- C'est un logiciel très puissant, dont les fonctionnalités de base peuvent être étendues à l'aide de plusieurs milliers d'extensions
- Etc.

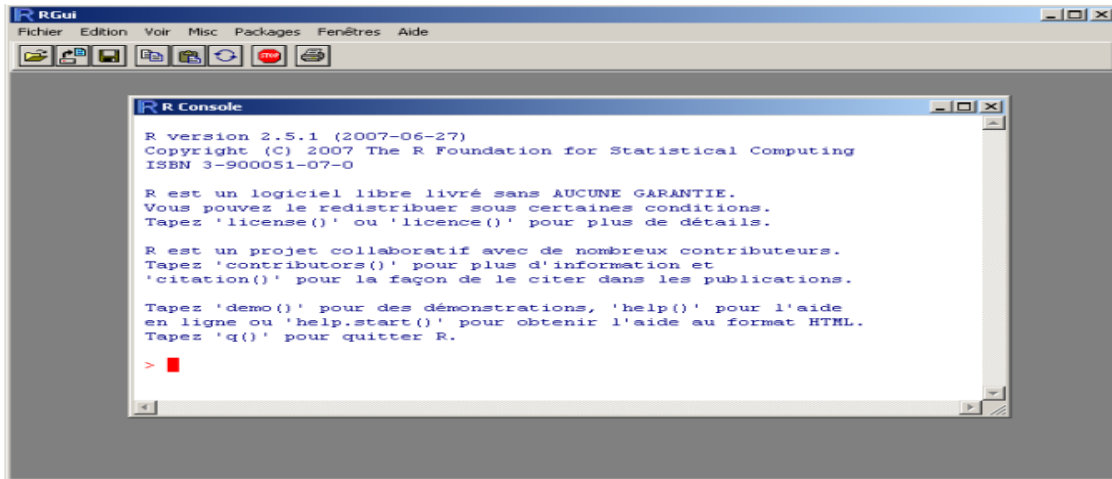


Figure 6 : L'interface de base de R sous Windows

➤ Logiciel RStudio

RStudio est un environnement de développement intégré libre, gratuit, et qui fonctionne sous Windows, Mac OSX et Linux. Il complète R et fournit un éditeur de script avec coloration syntaxique, des fonctionnalités pratiques d'édition et d'exécution du code (comme l'autocomplétions), un affichage simultané du code, de la console R, des fichiers, graphiques et pages d'aide, une gestion des extensions, une intégration avec des systèmes de contrôle de versions comme git, etc. Il intègre de base divers outils comme par exemple la production de rapports au format Rmarkdown. Il est en développement actif et de nouvelles fonctionnalités sont ajoutées régulièrement. Son principal défaut est d'avoir une interface uniquement anglophone.

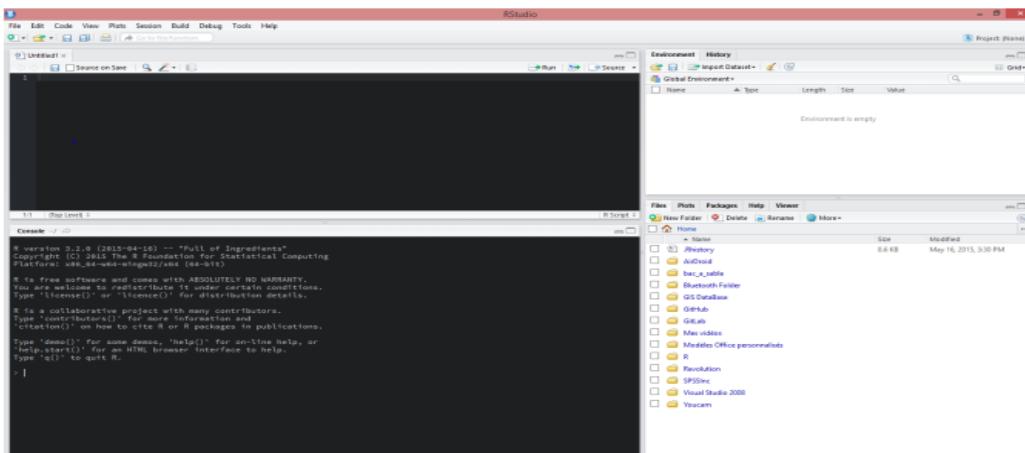


Figure 7 : L'interface de base de RStudio sous Windows

Sous **R**, en général, on ne voit pas directement les données sur lesquelles on travaille ; on ne dispose pas en permanence d'une vue des données sous forme de tableau, comme sous **RStudio**. Pour plus d'information sur **RStudio**, on pourra se référer au site <http://www.rstudio.com/>.

➤ Logiciel Excel

Excel est un logiciel de type « tableur » de la suite bureautique Office de Microsoft. Il permet de visualiser, d'explorer et partager facilement des idées à partir des données.

2.2. Méthodes

Afin d'atteindre nos objectifs, nous avons utilisé les modèles autorégressifs à changement de régimes markoviens. Dans la section suivante, nous donnons alors les détails nécessaires sur ces modèles, puis leur application à nos données.

2.2.1. Les Modèles à variable de transition cachée : Modèles autorégressifs à changement de régimes markovien (MSAR)

Dans le premier chapitre, nous avons vu les modèles à changement de régimes dont les transitions sont observées par une valeur seuil. Cependant, les transitions entre les régimes peuvent être gouvernées par un processus aléatoire inobservé. Le modèle le plus connu pour refléter cette caractéristique de transition probabiliste est celui défini dans le cadre markovien. Avant leur large développement dans le traitement automatique de la parole (Douc et *al.*, 2004), les modèles à variables cachées ont été d'abord introduits dans la littérature économétrique par Goldfeld & Quandt (1973) ; Hamilton (1989,1990). Dans la partie qui suit, nous présentons ce type de modèles qui seront par la suite utilisés pour nos données. Pour la compréhension de ces modèles, nous donnons ici les propriétés importantes pour leur utilisation, c'est-à-dire, nous introduisons dans un premier temps des notions sur les chaînes de Markov à temps discret qui sont nécessaires pour la suite du mémoire. Ensuite, nous présentons les modèles de Markov simples (MS), pour arriver aux modèles autorégressifs à changement de régimes markoviens (MSAR) proprement dit. Dans ce même chapitre, nous simulons également ces modèles pour une vue d'ensemble.

2.2.2. Les chaînes de Markov : Définitions, propriétés et exemples

Les chaînes de Markov sont des suites des variables aléatoires caractérisées par un aspect particulier de dépendance, qui leur attribue des propriétés singulières et un rôle important en modélisation par exemple en biologie, en physique, en économie ou en recherche opérationnelle. Elles ont été introduites par Andrey Markov en 1906. L'hypothèse de départ

c'est comme si plusieurs états de la nature de la variable observée existaient et une variable cachée déterminait l'état du processus observé Y_t . Donc la distribution de Y_t dépend d'un processus sous-jacent (inobservé) S_t , aléatoire et qui représente l'état ou le régime de la nature. Dans notre cas, S_t est caractérisé par un nombre fini d'état ; on note E , l'ensemble de états et i un état particulier de la nature.

Définition 1 :

On dit qu'un processus est une chaîne de Markov d'ordre k si la probabilité d'être dans un régime i à la période suivante dépend du régime présent et des $k-1$ régimes précédents. Alors une chaîne de Markov d'ordre k est une variable aléatoire notée $\{S_t\}$ à valeur dans E telle que $\forall i_t \in E$, et $P(S_t = i_t, \dots, S_0 = i_0) > 0$;

$$P(S_{t+1} = i_{t+1} | S_t = i_t, \dots, S_0 = i_0) = P(S_{t+1} = i_{t+1} | S_t = i_t, \dots, S_{t-k+1} = i_{t-k+1})$$

Pour la suite, nous considérons une chaîne de Markov d'ordre $k = 1$ pour les concentrations des NO_x . Dans ce cas, pour chaque période de temps, il existe donc une certaine probabilité d'appartenir à un régime donné. La probabilité que le système soit dans l'état j à $t + 1$ dépend uniquement de l'état dans lequel il se trouvait à l'instant t . Autrement dit, dans l'évolution au cours du temps, l'état du processus à l'instant $t + 1$ ne dépend que de celui à l'état précédent, mais pas de ses états antérieurs. Ce processus est défini de la manière suivante :

$\forall i_t \in E$, et $P(S_t = i_t, \dots, S_0 = i_0) > 0$;

$$P(S_{t+1} = i_{t+1} | S_t = i_t, \dots, S_0 = i_0) = P(S_{t+1} = i_{t+1} | S_t = i_t) \quad (17)$$

Dans la section suivante, nous verrons les propriétés ou hypothèses des chaînes de Markov qui seront nécessaires pour la suite.

➤ **Propriétés**

Soit $\{S_t\}_t$ est une chaîne de Markov d'ordre 1 à valeur dans E , l'ensemble dénombrable des régimes. Comme on vient de le voir ci-dessus, deux ou plusieurs régimes ou états sont liés par des probabilités qu'on appelle propriétés de transition d'un régime à l'autre.

Définition 2 :

Soient deux états i et j dans E ($i \neq j$) ; la probabilité de transition de l'état i vers j est notée :

$$p_{ij} = P(S_{t+1} = j | S_t = i) \quad (18)$$

La probabilité de ne pas transitité de i est donnée comme suit

$$p_{ii} = P(S_{t+1} = i | S_t = i) \quad (19)$$

Proposition 1 : l'ensemble de toutes les probabilités de transition vérifie les propriétés suivantes :

- $\forall (i, j) \in E, 0 \leq p_{ij} \leq 1$
- $\forall i \in E, \text{on a } \sum_{j \in E} p_{ij} = 1$

De ces probabilités de transition, on définit la matrice de transition entre des états comme suit : $\forall (i, j) \in E = \{1 \dots M\}$,

$$P = (p_{ij})$$

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1M} \\ \vdots & & \ddots & \vdots \\ p_{M1} & p_{M2} & \cdots & p_{MM} \end{pmatrix}$$

En plus de l'hypothèse $k=1$, il faut que S_t vérifie des propriétés supplémentaires suivantes :

- *Homogénéité*

Une chaîne de Markov est dite homogène lorsque la probabilité de transition de i à j ne dépend pas du temps t . c'est-à-dire :

$$p_{ij} = P(S_{t+1} = j | S_t = i) = P(S_1 = j | S_0 = i) \quad (20)$$

- *Irréductibilité*

$\{S_t\}$ est irréductible si tout état est accessible à partir de n'importe quel autre état, c'est-à-dire lorsque tous les états de la chaîne communiquent entre eux.

- *Apériodicité*

$\{S_t\}$ est apériodique si un état i peut survenir de manière irrégulière sans périodicité dans son occurrence.

Exemple 1 : Considérons un exemple sur l'évolution de la température, d'un jour à l'autre, décrite par les chaînes de Markov à deux régimes :

- Régime 1 représente les jours de hausses températures,
- Régime 2 représente les jours de basses températures.

Sous les deux hypothèses suivantes :

$$\begin{cases} p_{11} = P(\text{htemp demain} | \text{htemp aujourd'hui}) = 0.97 \\ p_{21} = P(\text{htemp demain} | \text{btemp aujourd'hui}) = 0.65 \end{cases}$$

On définit la variable aléatoire S_t par :

$$S_t = \begin{cases} 1 & \text{hausse temp le jour } t \\ 2 & \text{basse temp le jour } t \end{cases}$$

La suite $(S_t)_{t \in \mathbb{N}}$ est une chaîne de Markov à temps discret sur les états 1 et 2 dont la matrice de transition P est donnée par :

$$P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}$$

Où :

- p_{11} : représente la probabilité d'avoir de hausse température demain sachant que la température d'aujourd'hui est hausse également ;
- p_{21} : est la probabilité que la température de demain soit hausse sachant que celle d'aujourd'hui est basse ; Maximisation
- p_{12} : la probabilité d'être à basse température demain alors que la température d'aujourd'hui est hausse et
- p_{22} : représente la probabilité d'avoir une basse température demain alors que la température d'aujourd'hui est aussi basse.

Selon la propriété $\sum_{j \in E} p_{ij} = 1$,

On a alors :

$$\begin{cases} p_{11} + p_{12} = 1 \\ p_{21} + p_{22} = 1 \end{cases} \\ \Rightarrow \begin{cases} p_{12} = 1 - 0,97 = 0,03 \\ p_{22} = 1 - 0,65 = 0,35 \end{cases}$$

$$\Rightarrow P = \begin{pmatrix} 0,97 & 0,03 \\ 0,65 & 0,35 \end{pmatrix}$$

La figure (8) ci-dessous représente un schéma d'une chaîne de Markov à deux régimes comme de l'exemple :

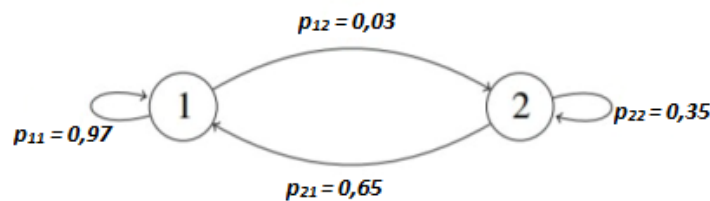


Figure 8 : Représentation des probabilités d'une chaîne de Markov à deux états (Olteneanu, 2006)

Les régimes sont représentés par des cercles numérotés et les probabilités de transitions positives par des flèches.

On définit également les probabilités non conditionnelles π_1 et π_2 telles que :

$$\pi_1 = P(S_t = 1) = \frac{1 - p_{22}}{2 - p_{11} - p_{22}}$$

$$\pi_2 = P(S_t = 2) = \frac{1 - p_{11}}{2 - p_{11} - p_{22}}$$

Cet exemple est le plus simple pour comprendre le comportement des chaînes de Markov.

La figure 9 ci-dessous représente un schéma d'une chaîne de Markov à trois régimes :

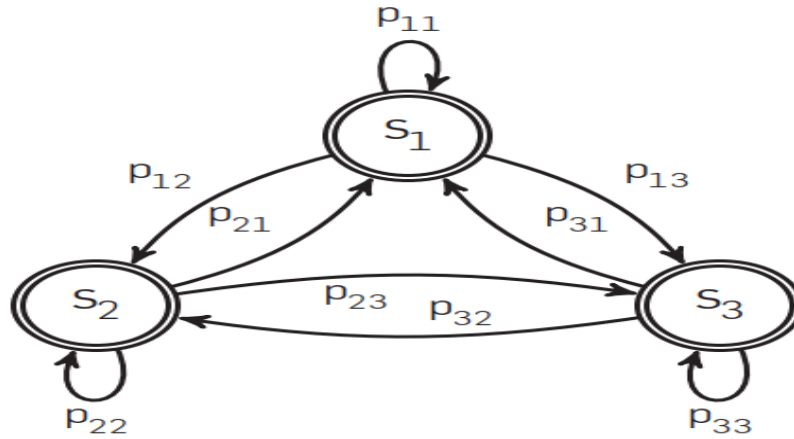


Figure 9 : Représentation des probabilités d'une chaîne de Markov à trois états (Ilham & Nadjat, 2020)

Les régimes sont représentés par des cercles numérotés et les probabilités de transitions positives par des flèches.

Proposition 2 : La relation de Chapman-Kolmogorov

C'est une propriété qui relie, pour une chaîne de Markov homogène, les probabilités de transition entre deux instants t et $(t + m)$ en incluant n transitions :

$$p_{ij}^{(m)} = P(S_{t+m} = j | S_t = i) = P(S_m = j | S_0 = i)$$

En d'autres termes, elle permet de déterminer les probabilités de transition $p_{ij}^{(m)}$ à partir des probabilités de transition entre deux instants consécutifs :

$$p_{ij}^{(m+t)} = \sum_{k \in E} p_{ik}^m p_{kj}^t$$

Cette identité résulte de l'associativité du produit matricielle

$$p^{(m+t)} = p^{(t+m)} = p^t p^m = p^{(t)} p^{(m)}$$

Exemple 2 : Considérons l'exemple de la météo précédent où la matrice de transition est donnée par :

$$P = \begin{pmatrix} 0,97 & 0,03 \\ 0,65 & 0,35 \end{pmatrix}, \text{ la matrice avec deux transitions est donnée par } P^{(2)}.$$

$$P^{(2)} = PP$$

$$P^{(2)} = \begin{pmatrix} 0,97 & 0,03 \\ 0,65 & 0,35 \end{pmatrix} \begin{pmatrix} 0,97 & 0,03 \\ 0,65 & 0,35 \end{pmatrix}$$

$$P^{(2)} = \begin{pmatrix} 0,96 & 0,04 \\ 0,86 & 0,14 \end{pmatrix}$$

2.2.3. Les modèles MS

Un processus à changement de régimes markoviens (MS) est un processus bivarié $\{S_t, Y_t\}$, où :

- $\{Y_t\}$: est un processus observé ;
- $\{S_t\}$: désigne une chaîne de Markov cachée du premier ordre à espace d'état fini $E = \{1, \dots, M\}$.

Dans la partie suivante, on se concentre sur un cas simple des modèles MS populaire, à savoir les modèles MS en moyenne-variance.

Soit le processus (Y_t) suit un modèle à changement de régime markoviens en moyenne-variance, qu'est déterminé par :

$$Y_t = m_t(S_t) + \sigma_t(S_t)\varepsilon_t \quad (21)$$

Où $\varepsilon_t \sim N(0, 1)$, $m_t(S_t)$ et $\sigma_t(S_t)$ dénotent des fonctions mesurables par rapport à une tribu engendrée par la variable aléatoire S_t , où les processus ε_t et (S_t) sont indépendants et S_t une chaîne de Markov homogène, irréductible, apériodique.

Ce modèle signifie que la réalisation de Y_t conditionnellement à S_t suit la normale de moyenne $m_t(S_t)$ et d'écart type $\sigma_t(S_t)$ ($Y_t|S_t \sim N(m(S_t), \sigma(S_t))$) et la densité de probabilité conditionnelle de Y_t est :

$$f(Y_t|S_t = i, \theta) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(\frac{-(y_t - m_i)^2}{2\sigma_i^2}\right) \quad (22)$$

Dans ce cas, le vecteur des paramètres du modèle est $\theta = (\sigma_i^2, m_i)$ et $i = \{1, \dots, M\}$.

2.2.4. Les modèles autorégressifs à changement de régimes markoviens (MS-AR)

On note par (Y_t) la série observable qui représente la variable les concentrations des polluants. Le modèle autorégressif à changement de régimes Markovien MS-AR considère que le comportement de la série Y_t dépend d'un processus inobservable (S_t) exogène générateur du régime à la date t qui est représentée par une chaîne de Markov cachée (Olteneanu, 2006). Comme dans le cas précédent, un MS(M)-AR(p) est un processus bivarié $\{S_t, Y_t\}$, où :

- $\{S_t\}$: est un processus inobservable défini comme une chaîne de Markov homogène, elle-même définie dans un espace d'état fini et discret $E = \{1, \dots, M\}$, dont est M le nombre de régimes de matrice de transition $P = (p_{ij})_{i,j \in E}$ telle que :

$$p_{ij} = P(S_t = j | S_{t-1} = i)$$

- $\{Y_t\}$: est un processus autorégressif linéaire défini par :

$$Y_t = m_t(S_t) + \sum_{i=1}^p \varphi_i(S_t) Y_{t-1} + \sigma_t(S_t) \varepsilon_t \quad (23)$$

Où ε_t est un processus strictement stationnaire, les fonctions $\varphi_i(S_t)$ ou φ_{i,S_t} dépendent d'une chaîne (S_t) de Markov homogène, irréductible, apériodique à espace d'état fini $E = \{1, \dots, M\}$ et indépendante de ε_t pour tout $t \in T$. La chaîne des états cachés $\{S_t\}$ non observable contrôle le processus observé $\{Y_t\}$, illustre le graphe d'indépendance conditionnelle (figure 10).

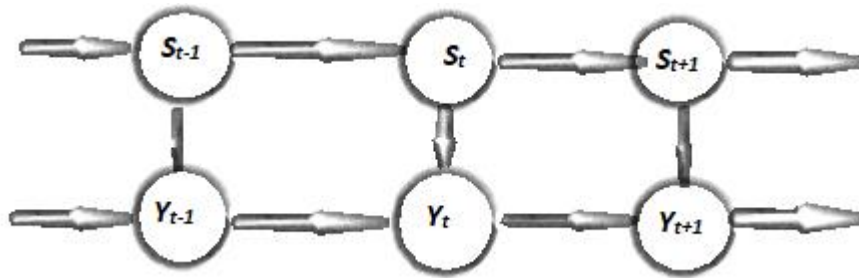


Figure 10 : Graphe d'indépendance conditionnelle du modèle MSAR (Ailliot, 2004)

Donc le changement de régimes s'observe également dans les composantes autorégressives du modèle. Ainsi un modèle autorégressif d'ordre 1 ($p = 1$) à changement de régimes markovien pour deux régimes ($M = 2$) sera noté par exemple par :

$$Y_t = \begin{cases} m_1 + \varphi_{1,1} Y_{t-1} + \sigma_1 \varepsilon_t, & \text{si } S_t = 1 \\ m_2 + \varphi_{1,2} Y_{t-1} + \sigma_2 \varepsilon_t, & \text{si } S_t = 2 \end{cases} \quad (24)$$

Dans ce modèle, on a considéré que les ordres d'auto-régression est le même dans chaque régime pour simplifier le travail. Dans certains cas, ces ordres peuvent être différents. Toutefois, le choix des ordres se fait par des méthodes d'estimation des modèles que nous verrons plus tard dans ce travail.

Ainsi lorsqu'on a par exemple $p_1 = 1, p_2 = 2$; $p_1=2$ et $p_2=4\dots$, les modèles sont représentés respectivement de manière suivante :

$$M_1(p_1 = 1, p_2 = 2) : Y_t = \begin{cases} m_1 + \varphi_{1,1} Y_{t-1} + \sigma_1 \varepsilon_t, & \text{si } S_t = 1 \\ m_2 + \varphi_{1,2} Y_{t-1} + \varphi_{2,2} Y_{t-2} + \sigma_2 \varepsilon_t, & \text{si } S_t = 2 \end{cases}$$

$$M_2(p_1 = 2, p_2 = 4) : Y_t = \begin{cases} m_1 + \varphi_{1,1}Y_{t-1} + \varphi_{2,1}Y_{t-2} + \sigma_1\varepsilon_t, & \text{si } S_t = 1 \\ m_2 + \varphi_{1,2}Y_{t-1} + \varphi_{2,2}Y_{t-2} + \varphi_{3,2}Y_{t-3} + \varphi_{4,2}Y_{t-4} + \sigma_2\varepsilon_t, & \text{si } S_t = 2 \end{cases}$$

Nous terminons cette section par la présentation des modèles à changement de régimes évoqués dans ce travail sous forme d'un diagramme simplifié à la figure (11) suivante :

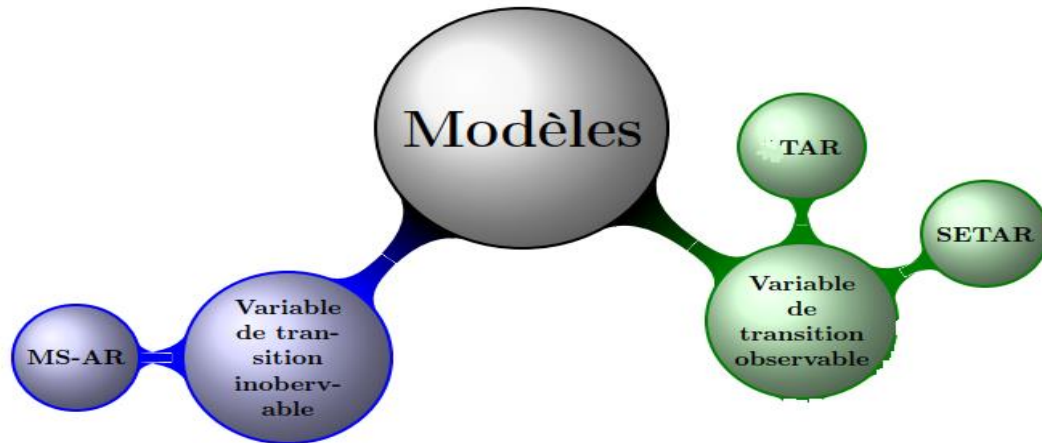


Figure 11 : Récapitulatif des modèles à changement de régime selon l'observabilité de la variable de transition (Fouquau, 2008)

Lorsque la transition est induite par une variable inobservable générée par un processus de Markov, alors le passage d'un régime à un autre est signalé par une matrice de probabilités (MS-AR). Au contraire, si la variable est observée, on distingue deux classes de modèles, selon la forme de la fonction de transition : soit la transition est brutale, soit elle est retardée.

Pour plus de détail sur ces modèles cités et d'autres modèles à changement de régimes, nous proposons les travaux de Dijk *et al.* (2002) ; Chow & Zhang (2013) ; Kock & Teräsvirta. (2011). Cependant la modélisation des phénomènes physiques ne s'arrête pas qu'à la présentation des modèles, dans la section suivante, nous passons à d'autres notions qui permettent d'atteindre notre objectif. Il s'agit alors de passer à l'application sur des données.

2.2.5. Application aux données des concentrations des NO_x

Pour trouver le modèle qui permettra de mieux décrire l'évolution des concentrations des polluants, nous allons analyser les séries temporelles par l'estimation des paramètres des modèles et la simulation des données. Cette partie se termine par la présentation d'une méthode de prévision pour les données à notre disposition. Nous choisissons d'utiliser les modèles MSAR. Néanmoins, ce choix sera justifié plus loin dans ce travail. Ainsi, nous essayons de procéder comme suit :

- Spécification du processus,
- Estimation des paramètres du processus.

- Validation du modèle par test

Nous allons expliquer ces différentes étapes dans ce qui suit.

2.2.5.1. Spécification du modèle

L'objectif de notre étude consiste à identifier les régimes dans l'évolution des concentrations des polluants atmosphériques et les modéliser en utilisant les modèles autorégressifs à changement de régimes markoviens (MS(M)-AR(p)). Pour atteindre cet objectif, nous avons utilisé les données des concentrations des NO_x (NO et NO₂).

Dans cette section, nous essayons de justifier notre choix du modèle à développer dans la suite de notre travail.

- Choix de l'ordre d'autorégression p .

Pour le choix de l'ordre p , nous avons observé la fonction d'autocorrélation partielle (PACF) des séries des NO_x.

Mais avant d'étudier l'évolution proprement dite de nos données, nous allons d'abord simuler plusieurs modèles à changement de régimes markoviens pour comprendre leur comportement.

2.2.5.2. Méthodologie théorique : estimation des paramètres des modèles MSAR.

L'estimation des processus à changement de régime requiert l'estimation de tous ses paramètres.

Afin d'estimer les paramètres des modèles MSAR présentés un peu plus haut, on peut procéder par deux approches :

- Une approche qu'est l'estimation Bayésienne des paramètres du modèle. Etant face à un problème d'estimation sous la présence des données inobservables représentées par la séquence cachée $\{S_t\}_{t \in T}$, qui permet d'estimer les paramètres sous la présence des données manquantes ou inobservables.
- La seconde approche est celle basée sur un maximum de vraisemblance où on estime les paramètres du modèle par la méthode du Maximum de Vraisemblances (MV) ou par l'une de ses dérivées. Pour cette approche, on utilise généralement l'algorithme EM comme une technique d'optimisation itérative qui donne des estimations des paramètres qui converge localement vers l'estimateur de MV (Hamilton, 1990).

Dans ce mémoire, nous allons utiliser l'approche sur la vraisemblance pour la calibration des modèles MSAR, c'est à-dire celle utilisant l'algorithme EM. Nous décrivons dans un premier temps la démarche de manière méthodique, puis l'appliquer en exemple à un modèle particulier MS(2)-AR(1).

2.2.5.2.1. Fonction de vraisemblance

Soit (y_1, \dots, y_T) une observation du processus $\{Y_t\}$, tel que θ , l'ensemble des paramètres inconnus (les probabilités constantes π_i , la matrice de transition P et les coefficients φ_i , $i = 1, \dots, p$ du modèle) (Surprenant, 2018). On cherche à estimer θ à partir des observations disponibles (y_1, \dots, y_T) pour avoir la valeur des paramètres qui rend la fonction de vraisemblance $L(\theta) = P(Y_t; \theta)$ maximale. On a :

$$P(Y_t; \theta) = \sum_{s_1, \dots, s_T \in \{1, \dots, M\}} P(Y_t, \underline{S}_t; \theta) \quad (28)$$

Avec $P(Y_t, \underline{S}_t; \theta)$, la vraisemblance de toutes les données notée $\mathcal{L}(\theta)$ qui vérifie, par application des propriétés d'indépendance conditionnelles du modèle et la formule de Bayes :

$$\begin{aligned} \mathcal{L}(\theta) &= P(Y_t, \underline{S}_t; \theta) \\ \mathcal{L}(\theta) &= f(y_1, S_1 | \theta) \prod_{t=2}^T f(y_t, S_t | Y_{t-1}, S_{t-1}; \theta) \\ \mathcal{L}(\theta) &= f(y_1, S_1 | \theta) P(S_1) \prod_{t=2}^T f(y_t | S_t, Y_{t-1}, S_{t-1}; \theta) P(S_t | Y_{t-1}, S_{t-1}; \theta) \\ \mathcal{L}(\theta) &= f(y_1 | S_1; \theta) P(S_1) \prod_{t=2}^T f(y_t | S_t; \theta) P(S_t | S_{t-1}; \theta) \end{aligned}$$

Si $f(y_1 | S_1; \theta) > 0$

C'est alors cette vraisemblance qu'on cherchera à maximiser, et la méthode usuelle consiste à utiliser l'algorithme EM.

2.2.5.2.2. Algorithme EM (Espérance-maximisation)

On définit l'algorithme EM comme une méthode qui permet d'approcher l'estimateur du maximum de vraisemblance (EMV) quand les données sont incomplètes (cas d'une variable cachée par exemple) ou quand une partie des données est manquante (cas d'une censure par exemple). Il a été introduit d'abord dans le cadre des modèles CMC (Chaînes de Markov Cachées), puis étendu par la suite aux modèles à variables cachées plus généraux par Dempster et al. (1977) et aux modèles MS-AR par Hamilton (1989). Il s'agit d'un algorithme itératif pour trouver des estimateurs du maximum de vraisemblance locaux de la fonction de vraisemblance des observations, lorsque chaque observation contient une partie cachée.

Alors, on considère que chaque donnée est un couple de type (S, Y) où Y est la partie observable et S sa partie cachée, nous supposons connue d'une manière explicite la forme de la fonction densité jointe $P(S, Y; \theta)$ où θ est l'ensemble des paramètres du modèle à estimer.

➤ **Principe générale de l'algorithme E**

Soit $Y = (y_1, \dots, y_T)$ un ensemble d'observations. Faisons l'hypothèse que les observations sont i.i.d, et admettent $L(\theta) = P(Y; \theta)$ comme fonction de vraisemblance que l'on ne peut pas maximiser. Soient les données $S = (s_1, \dots, s_M)$ qui pourraient rendre la maximisation possible et simple si elles étaient connues. Naturellement, l'estimateur est noté $E_{S|Y, \theta^{(n)}}[\log(\mathcal{L}(\theta))]$. La maximisation de la vraisemblance nous donnera une nouvelle valeur de de l'estimateur. En effet c'est une méthode itérative et chaque itération de l'algorithme fait intervenir ses deux étapes citées plus haut. $\theta^{(n)}$ désigne la valeur des paramètres après n itérations. Dans cette partie, nous expliquons plus précisément ces deux étapes dans le cas des modèles $MS(M)$ - $AR(p)$.

✓ Étape E (Expectation) :

On évalue l'espérance selon les données observées et la valeur des paramètres déterminés à l'itération précédente ce qui se calcule comme suit :

$$Y(\theta|\theta^{(0)}) = E_{S|Y, \theta^{(0)}}[\log(\mathcal{L}(\theta))] \quad (29)$$

✓ Etape M (Maximisation) : Cette étape procède à la maximisation de la vraisemblance en utilisant l'estimation des données inconnues effectuées à l'étape précédente, et met à jour la valeur des paramètres pour la prochaine itération. Ainsi, le passage de l'itération (n) à l'itération $(n + 1)$ de l'algorithme consiste déterminer :

$$\hat{\theta}^{(n+1)} = \operatorname{argmax}\{E_{S|Y, \hat{\theta}^{(n)}}[\log(\mathcal{L}(\theta))]\} \quad (30)$$

On répète ces deux étapes jusqu'à obtenir une égalité entre la vraisemblance de l'itération $(n+1)$ et celle de l'itération n . Nous pouvons résumer cet algorithme comme suit :

1. Choisir le vecteur initial de $\theta^{(0)}$ d'une manière arbitraire
2. $n = 0$, on calcule l'espérance :
 - Etape E : $Y(\theta|\theta^{(0)})$
 - Etape M : $\hat{\theta}^{(1)} = \operatorname{argmax}\{E_{S|Y, \hat{\theta}^{(0)}}[\log(\mathcal{L}(\theta))]\}$
3. Si la valeur de $\hat{\theta}^{(n+1)}$ ne converge pas vers la valeur de $\hat{\theta}^{(n)}$ faire :

$$Y(\theta|\theta^n) = \operatorname{argmax}\{E_{S|Y, \hat{\theta}^{(n)}}[\log(\mathcal{L}(\theta))]\}$$

- Etape M : l'étape de maximisation par rapport à $\theta \in \Theta$

$$\hat{\theta}^{(n)} = \operatorname{argmax}\{Y(\theta|\theta^{(n)})\}$$

4. L'algorithme s'arrête si $\hat{\theta}^{(n)} = \hat{\theta}^{(n+1)}$

L'algorithme a été bien développé dans les travaux de Ilham et Nadjat (2020).

2.2.5.2.3. Application de l'algorithme EM sur le modèle MS(2)-AR(1)

Supposons qu'on a $\{y_1, \dots, y_T\}$ une observation du processus aléatoire $(Y_t)_{t \in T}$ (concentration des NO_x) qui suit un modèle MS (2)-AR(1). Soit $S = (S_1, \dots, S_T)$ la réalisation du régime inobservée de tout l'échantillon. On rappelle le modèle (pour les processus stationnaires par régime) :

$$\begin{aligned} Y_t &= m_{S_t} + \varphi_{1,S_t} Y_{t-1} + \varepsilon_t \text{ ou } \varepsilon_t \sim N(0, \sigma_{S_t}^2) \\ \Rightarrow Y_t &= \begin{cases} m_1 + \varphi_{1,1} Y_{t-1} + \varepsilon_t, & \text{si } S_t = 1, \text{ pour } t = 1, 2, \dots, t_0 \\ m_2 + \varphi_{1,2} Y_{t-1} + \varepsilon_t, & \text{si } S_t = 2, \text{ pour } t = t_0 + 1, t_0 + 2, \dots \end{cases} \end{aligned} \quad (31)$$

Selon l'équation (31), les paramètres inconnus à estimer sont :

- Les coefficients du modèle AR(1) dans chaque régime ($\varphi_{1,S_t} = (\varphi_{1,1} \text{ et } \varphi_{1,2})$);
- Les probabilités de transition du régime 1 au régime 2 ($p_{11}, p_{12}, p_{21}, p_{22}$);
- Les probabilités non conditionnelles ($\pi_1 \text{ et } \pi_2$);
- La variance de l'innovation ($\sigma_{S_t}^2$);
- Les deux interceptions ou moyenne m_1 et m_2 .

Donc θ le paramètre inconnu appartient à un espace de $\Theta \in \mathbb{R}^9$, puisque $p_{11} + p_{12} = 1, p_{21} + p_{22} = 1$ et $\pi_1 + \pi_2 = 1$. La vraie valeur de θ sera notée θ^0 et son choix doit être bien fait.

On rappelle alors que la distribution conditionnelle jointe de $(Y|S)$ est donnée par :

$$\begin{aligned} f(Y_t, S_t | \theta) &= f(y_1, S_1 | \theta) \prod_{t=2}^T f(y_t, S_t | Y_{t-1}, S_{t-1}; \theta) \\ &= f(y_1, S_1 | \theta) P(S_1) \prod_{t=2}^T f(y_t | S_t, Y_{t-1}, S_{t-1}; \theta) P(S_t | Y_{t-1}, S_{t-1}; \theta) \\ &= f(y_1 | S_1; \theta) P(S_1) \prod_{t=2}^T f(y_t | S_t; \theta) P(S_t | S_{t-1}; \theta) \end{aligned}$$

Avec $(\varepsilon_{\varphi_1(S_t)}(t))_{t \in T}$ est le processus déterminé par $\varepsilon_{\varphi_1(S_t)}(t) = Y_t - m_{S_t} - \varphi_{1,S_t} Y_{t-1}$ et $f(y_1 | S_1; \theta) > 0$.

Si on considère la loi normale comme loi de distribution de (ε_t) , nous obtenons :

$$f(\varepsilon_{\varphi_1(S_t)}(t)) = \frac{1}{\sqrt{2\pi\sigma_{S_t}^2}} \exp\left(-\frac{(Y_t - m_{S_t} - \varphi_{1,S_t} Y_{t-1})^2}{2\sigma_{S_t}^2}\right)$$

Par la suite, la vraisemblance marginale de Y , $\mathcal{L}_T(Y, \theta)$ sera simplement la somme de $P(Y, S; \theta)$ sur toutes les valeurs possibles de (S_1, \dots, S_T) :

Démonstration :

On a :

$$\begin{aligned} f(Y_t, S_t | \theta) &= [P(S_1 = 2)f(y_1 | S_1 = 2; \theta)\pi_1 + P(S_1 = 1)f(y_1 | S_1 = 1; \theta)(1 - \pi_1)] * \\ &\prod_{t=2}^T P(S_t = 2, S_{t-1} = 2)f(y_t | S_t = 2; \theta)p_{22} \\ &+ P(S_t = 1, S_{t-1} = 2)f(y_t | S_t = 1; \theta)(1 - p_{22}) \\ &+ P(S_t = 2, S_{t-1} = 1)f(y_t | S_t = 2; \theta)(1 - p_{11}) \\ &+ P(S_t = 1, S_{t-1} = 1)f(y_t | S_t = 1; \theta)p_{11} \end{aligned}$$

En appliquant la fonction log sur cette expression, on obtient :

$$\begin{aligned} \log P(y_t, S_t | \theta) &= P(S_1 = 2)[\log f(y_1 | S_1 = 2; \theta) + \log \pi_1] \\ &+ P(S_1 = 1)[\log f(y_1 | S_1 = 1; \theta) + \log(1 - \pi_1)] \\ &+ \sum_{t=2}^T P(S_t = 2) \log f(y_t | S_t = 2; \theta) + P(S_t = 1) \log f(y_t | S_t = 1; \theta) \\ &+ P(S_t = 2, S_{t-1} = 2) \log(p_{22}) + P(S_t = 1, S_{t-1} = 2) \log(1 - p_{22}) \\ &+ P(S_t = 2, S_{t-1} = 1) \log(1 - p_{11}) + P(S_t = 1, S_{t-1} = 1) \log(p_{11}) \end{aligned}$$

Etant donné que les données inobservées S sont présentes dans la vraisemblance, l'estimation paramétrique du modèle MS (2)-AR (1) ne peut pas être réalisée par la méthode MV directement mais plutôt à l'aide de l'algorithme EM dans lequel on détermine l'expression exacte de $Y_t(\theta^{(n+1)} | \theta^{(n)})$ qui est donnée par :

$$\begin{aligned} Y_t(\theta^{(n+1)} | \theta^{(n)}) &= E[\log f(Y_t, S_t | \theta^{(n)})] \\ &= \pi_1^{(n)} [\log f(y_t | S_1 = 2; \theta) + \log \pi_1^{(n)}] \\ &+ (1 - \pi_1^{(n)}) [\log f(y_t | S_1 = 1; \theta^{(n)}) + \log((1 - \pi_1^{(n)}))] \\ &+ \sum_{t=2}^T \{P(S_t = 2 | Y_t; \theta^{(n)}) \log f(y_t | S_t = 2; \theta^{(n)})\} \\ &+ P(S_t = 1 | Y_t; \theta^{(n)}) \log f(y_t | S_t = 1; \theta^{(n)}) \\ &+ P(S_t = 2, S_{t-1} = 2 | Y_t; \theta^{(n)}) \log(p_{22}) \\ &+ P(S_t = 1, S_{t-1} = 2 | Y_t; \theta^{(n)}) \log(1 - p_{22}) \\ &+ P(S_t = 2, S_{t-1} = 1 | Y_t; \theta^{(n)}) \log(1 - p_{11}) \\ &+ P(S_t = 1, S_{t-1} = 1 | Y_t; \theta^{(n)}) \log(p_{11}) \end{aligned}$$

Dans la seconde étape (étape M) de nouveaux estimateurs MV du paramètre, basés sur les inférences retenues dans l'étape E, sont calculés. Les deux étapes sont répétées jusqu'à ce que le maximum de vraisemblance soit atteint.

Dans le but d'évaluer les performances de la procédure d'estimation du modèle MS(2)–AR(1), basée sur l'algorithme EM, étudiée dans le chapitre précédent où :

$$Y_t = m_{S_t} + \varphi_{1,S_t} Y_{t-1} + \varepsilon_t$$

Avec $\varepsilon_t \sim N(0, \sigma_{S_t}^2)$ est un bruit blanc gaussien et S_t est une chaîne de Markov du premier ordre stationnaire, irréductible et homogène et Y_t est stationnaire. Les simulations ont été réalisées pour différentes ensembles des valeurs initiales $\theta^{(0)}$. Les résultats de cette estimation sont donnés dans le Tableau IV du chapitre 3.

2.2.5.3. Validation des modèles MSAR

Pour la validation du modèle estimé, nous étudierons dans un premier temps les erreurs quadratiques des différents modèles estimés pour voir quel modèle minimise ces erreurs. Ensuite nous ferons le test de bruit blanc sur des résidus. Le modèle qui vérifiera ces deux conditions sera le vrai modèle, et sera retenu pour des concentrations des NO_x .

TROISIEME PARTIE : RESULTATS ET DISCUSSION

TROISIEME PARTIE : RESULTATS ET DISCUSSION

3.1. Résultats

3.1.1. Résultats des simulations.

Avant de traiter nos données d'étude, nous avons simulé plusieurs modèles pour la compréhension générale des modèles MSAR. Le tableau ci-dessous renseigne les résultats issus des différentes simulations des modèles MSAR.

Tableau II : Paramètres des simulations de modèles 1 et 2

Paramètres	Modèle 1	Modèle 2
$\varphi_{1,1}$	0,12	0,63
$\varphi_{1,2}$	0,14	0,93
m_1	1,31	0,83
m_2	2,70	1,88
p_{11}	0,98	0,35
p_{22}	0,99	0,20
π_1	0,33	0,55
π_2	0,67	0,45

➤ **Modèle 1 : MS(2)AR(1)**

Nous constatons que les moyennes sont différentes avec les probabilités $p_{11} = P(S_t = 1 | S_{t-1} = 1)$ et $p_{22} = P(S_t = 2 | S_{t-1} = 2)$ très élevées. A partir de ces probabilités conditionnelles, nous déterminons les probabilités d'appartenance aux régimes $\pi_1 = 0,33$ et $\pi_2 = 0,67$. Ces probabilités d'appartenance impliquent que le régime 2 est plus persistant que le régime 1.

La figure (12) suivante donne la représentation du modèle (1).

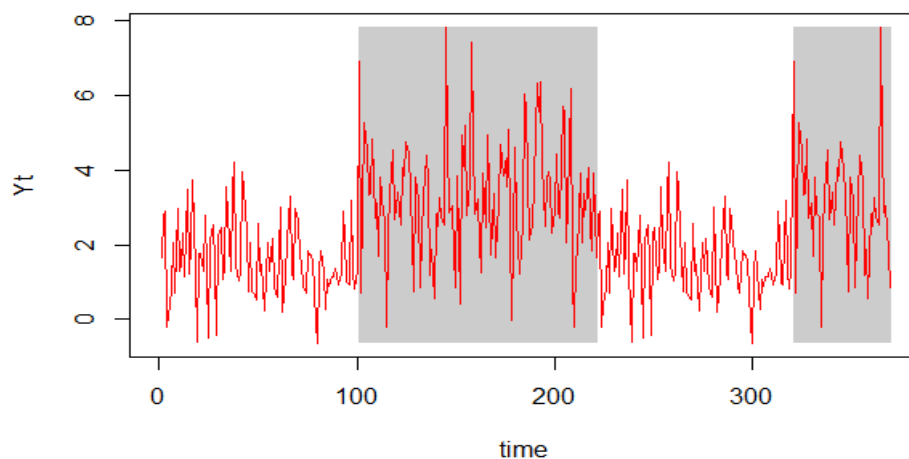


Figure 12 : Graphe du Modèle 1 simulé

Le processus général du modèle 1 est donné par :

$$Y_t = \begin{cases} 1,31 + 0,12Y_{t-1} + \varepsilon_t, & \text{si } S_t = 1 \\ 2,70 + 0,14Y_{t-1} + \varepsilon_t, & \text{si } S_t = 2 \end{cases}$$

➤ **Modèle 2 : MS(2)AR(1)**

Pour les résultats du modèle 2, nous constatons que les probabilités de transition p_{12} et p_{21} sont élevées ($p_{12} = 0,65, p_{21} = 0,80$) et les probabilités d'appartenance à un régime $\pi_1 = 0.55$ et $\pi_2 = 0.45$ sont presque les mêmes. Ce qui signifie qu'il est facile de quitter un régime à un autre.

Nous présentons ce modèle sur la figure (13).

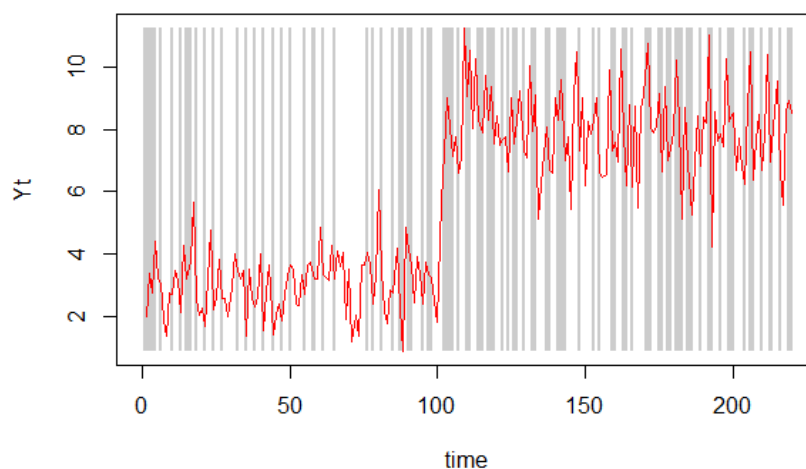


Figure 13 : Modèle 2 : MS(2)AR(1)

L'écriture du modèle 2 est donnée comme suit :

$$Y_t = \begin{cases} 0,83 + 0,63Y_{t-1} + \varepsilon_t, & \text{si } S_t = 1 \\ 1,88 + 0,93Y_{t-1} + \varepsilon_t, & \text{si } S_t = 2 \end{cases}$$

Dans les présentations des modèles suivants, nous donnons une comparaison entre un modèle à changement de régime à deux régimes et à trois régimes. Cette comparaison a pour but de justifier le choix du nombre de régime dans certaines séries.

Soit le modèle suivant simulé avec deux régimes :

$$Y_t = \begin{cases} 1,99 + 0,091Y_{t-1} + \varepsilon_t, & \text{si } S_t = 1 \\ 2,93 + 0,02Y_{t-1} + \varepsilon_t, & \text{si } S_t = 2 \end{cases}$$

Avec la matrice de transition suivante :

$$P_1 = \begin{pmatrix} 0,98 & 0,01 \\ 1,63 \cdot 10^{-7} & 0,99 \end{pmatrix}$$

La représentation de ce modèle est donnée ci-dessous :

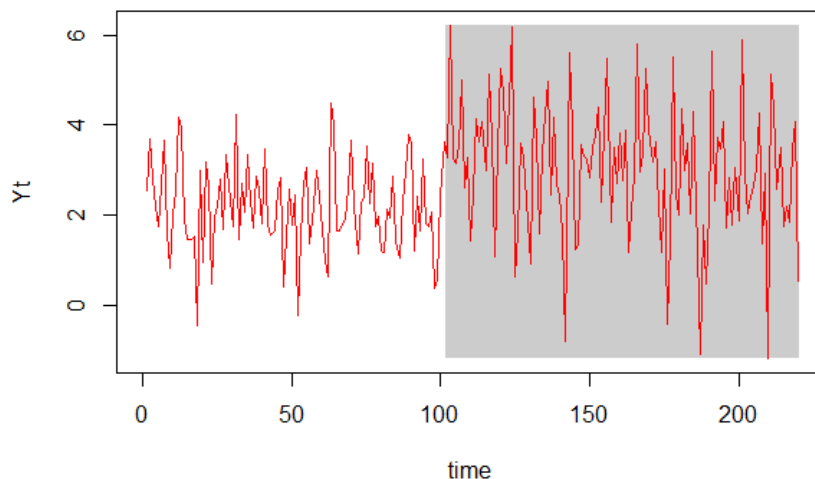


Figure 14 : Graphique du modèle 3

Pour ce même processus, nous avons essayé l'hypothèse de la présence d'un troisième régime, et nous avons obtenu la matrice de transition suivante :

$$P_2 = \begin{pmatrix} 0,98 & 4,5 \cdot 10^{-8} & 0,01 \\ 1,3 \cdot 10^{-33} & 0,97 & 0,2 \\ 2,99 \cdot 10^{-33} & 0,503 & 0,49 \end{pmatrix}$$

Les probabilités de transition sont très faibles (presque nulles) dans cette matrice. Ce qui signifie qu'il est difficile de quitter un régime à un autre. En plus, la probabilité de ne pas transiter du troisième (p_{33}) régime est faible par rapport aux autres régimes. La figure suivante donne la représentation du modèle à trois régimes.

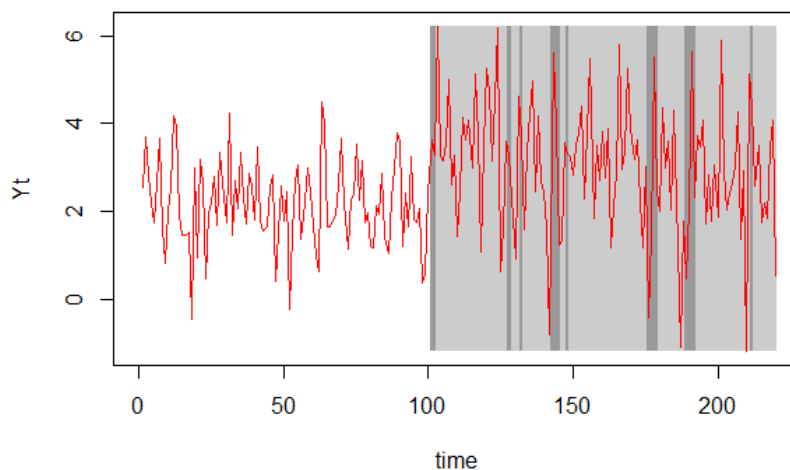


Figure 15 : Représentation du Modèle 4 : MS(3)AR(1)

Dans la suite de notre travail, nous faisons l'application sur les données des NO_x.

3.1.2. Résultats sur l'application aux données de NO_x ($\mu\text{g}/\text{m}^3$)

- **Représentation graphique des données**

Dans cette section, nous faisons une analyse graphique et descriptive des données de NO_x.

Les deux graphes de la figure (16) illustrent les évolutions des concentrations de NO (à gauche) et NO₂ (à droite) sur la période d'étude.

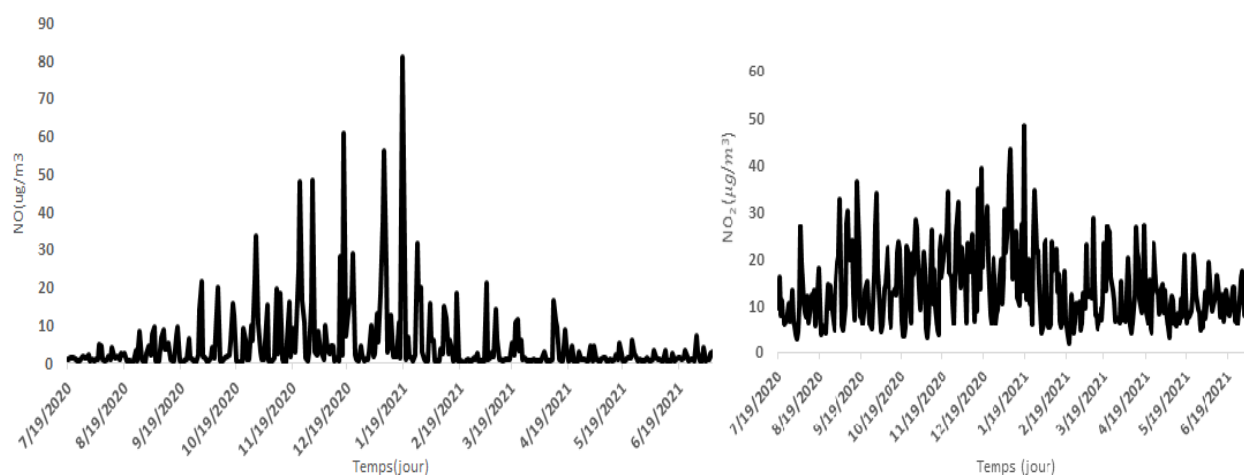


Figure 16 : Représentation graphique des concentrations de NO et NO₂

La figure (16) présente les fortes concentrations de NO sur la période du 23/ 11/2020 au 18/01/2021. Quant au NO₂, la figure présente une forte variabilité des concentrations. Ceci est justifié par la présence des pics de concentrations de NO₂. Dans la section suivante, nous présentons les résultats de l'analyse descriptive des données. Cette étude nous permet de décrire le comportement des concentrations sur la période d'étude

- **Analyse statistique des NO_x**

Le tableau (III) présente les résultats de l'analyse descriptive des concentrations des NO_x

Tableau III : Résultats de l'analyse descriptive

Concentrations	NO	NO ₂
Moyenne	4,8225989	13,5771831
Médiane	1,3500000	11,9000000
Ecart type	8,9684273	7,8498129
Variance	80,43268	61,6195627
Kurtosis	23,89446	1,405174
Skewness	4,260098	1,153118
Minimum	0,1000000	1,9000000
Maximum	80,9000000	48,5000000
Etendue	80,80000	46,6000000
Coefficient de variation	1,8596669	0,5781621
Observations	355	355

Selon l'analyse descriptive des concentrations, nous remarquons que le « skewness » est positif. Ce qui correspond à une distribution asymétrique à droite. L'asymétrie de la série montre qu'il existe de fortes concentrations positives dans les données de NO_x . La valeur du « kurtosis » de la série de NO est supérieure à 3. Ce qui implique que les données appartiennent la famille de distributions leptokurtiques (existence des valeurs extrêmes dans la série).

L'asymétrie observée sur les données de NO indique l'existence d'au moins deux régimes significatifs (régime à fortes concentrations et régime à faibles concentrations). Ce constat est aussi observé sur la série de NO_2 en analysant l'histogramme de la figure (17). La présence d'au moins deux régimes dans les séries montre qu'il y a une non-linéarité dans l'évolution des concentrations des NO_x .

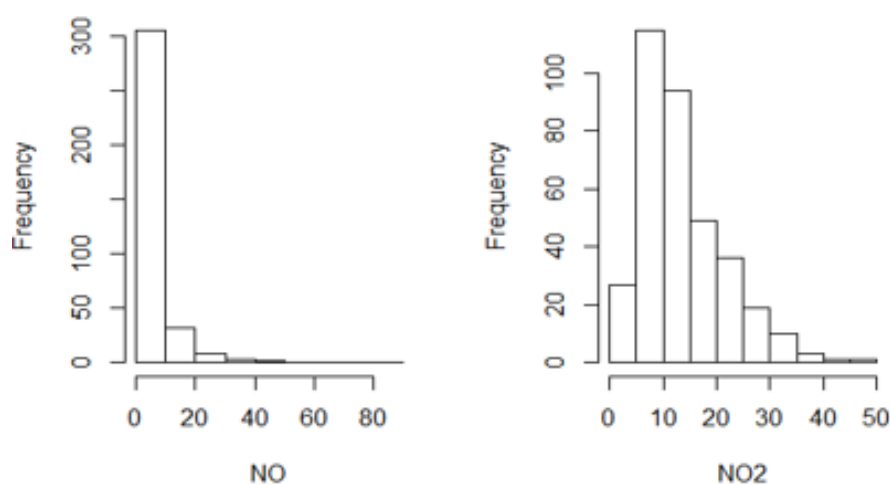


Figure 17 : Histogrammes des concentrations de NO et NO₂

Par conséquent, il est important de prendre en compte ce comportement non-linéaire de ces séries pour la modélisation. Alors, nous avons recours aux modèles non-linéaires de type markovien qui permettent d'identifier les régimes dans la modélisation de ces séries. La section suivante donne les résultats des estimations des modèles MSAR développée par l'algorithme EM.

- Choix de p

Les graphes suivants donnent la représentation des PACF de NO et NO₂

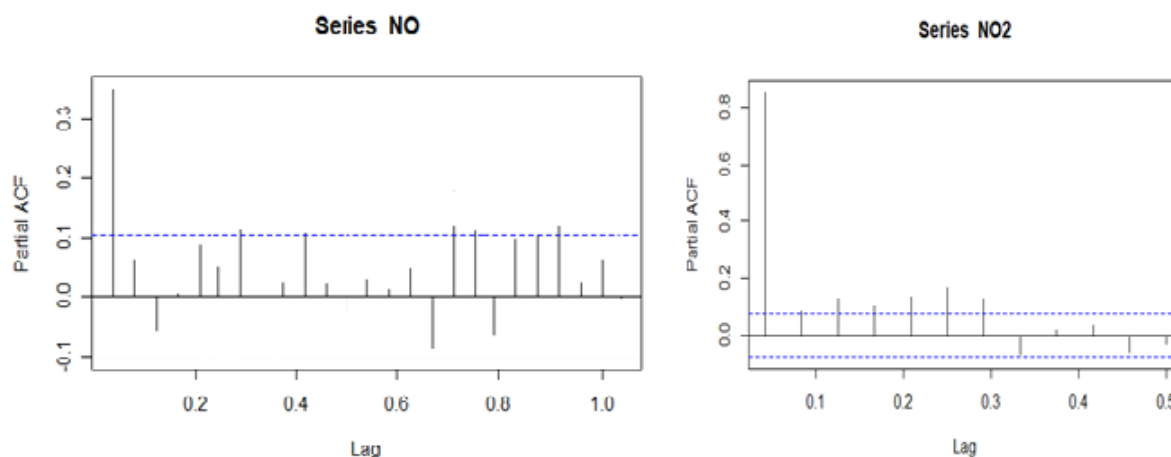


Figure 18: PACF de NO et NO₂

L'analyse de la PACF des NO et NO₂ nous montre qu'un seul pic est significativement hors de l'intervalle de confiance. Ce qui veut dire que l'ordre d'auto-régression $p = 1$.

3.1.3. Résultats des estimations par l'algorithme EM

Le tableau suivant présente les estimateurs des paramètres obtenus à partir des modèles développés par Hamilton (1989).

Tableau IV : Résultats de l'estimation des paramètres du modèle MS(2)-AR(1)

Paramètres ($\hat{\theta}^{(n)}$)	NO	NO ₂
$\hat{\varphi}_{1,1}$	0,039	0,367
$\hat{\varphi}_{1,2}$	0,299	0,441
p_{11}	0,757	0,719
p_{22}	0,549	0,527
π_1	0,649	0,627
\hat{m}_1	0,975	4,893
\hat{m}_2	9,39	13,659
Log-likelihood	-859,72	-1121,4
Observations	355	355

Selon les résultats du tableau, sept paramètres du processus pour chaque composé sont estimés, tels que cités plus haut.

➤ Interprétation des résultats.

Ces résultats ont confirmé l'existence d'au moins deux régimes différents dans les séries de concentrations en NO et en NO₂ durant la période d'étude. Les écarts entre les moyennes des concentrations sont importants.

Au régime 1, on a $\hat{m}_1 = 0,975 \mu g/m^3$ contre $\hat{m}_2 = 9,390 \mu g/m^3$ dans le régime 2 (pour NO). Pour les concentrations du dioxyde d'azote, on obtient $\hat{m}_1 = 4,893 \mu g/m^3$ dans le premier régime et $\hat{m}_2 = 13,659 \mu g/m^3$ dans l'état 2.

Concernant les probabilités de transition du NO et du NO_2 , la probabilité de rester dans le régime 1 est : $p_{11} = P(S_t = 1 | S_{t-1} = 1) = 0,757$ (resp. 0,719) ;

Alors que celle de rester dans le régime 2 est $p_{22} = P(S_t = 2 | S_{t-1} = 2) = 0,549$ (resp. 0,527). Ces probabilités de transition (p_{11}, p_{22}) sont significatives et impliquent l'existence d'au moins deux régimes dans les séries des concentrations de NO_x . Un régime de fortes concentrations des NO_x à moyenne élevée ($9,390 \mu g/m^3$ (NO) et $13,659 \mu g/m^3$ (NO_2), et un autre régime de faibles concentrations où la moyenne est petite ($0,975 \mu g/m^3$ (NO) et $4,893 \mu g/m^3$ (NO_2)).

Nous déterminons les probabilités d'appartenance (non conditionnelles) à chaque régime π_1 et π_2 par la formule suivante :

$$\pi_1 = P(S_t = 1) = \frac{1-p_{22}}{2-p_{11}-p_{22}}, \text{ et } \pi_2 = 1 - \pi_1$$

On obtient alors :

$$\pi_1 = 0,649 \text{ et } \pi_2 = 0,351 \text{ pour le } NO ; \pi_1 = 0,627 \text{ et } \pi_2 = 0,373 \text{ pour } NO_2$$

Ces deux probabilités non-conditionnelles impliquent que dans notre échantillon, 64,9% (62,7%) d'observations appartiennent au régime 1 (faible concentration) et 35,1% (37,3%) appartiennent au régime 2 (forte concentration). Cela veut dire aussi qu'il y a plus de données à faibles concentrations sur la période d'étude.

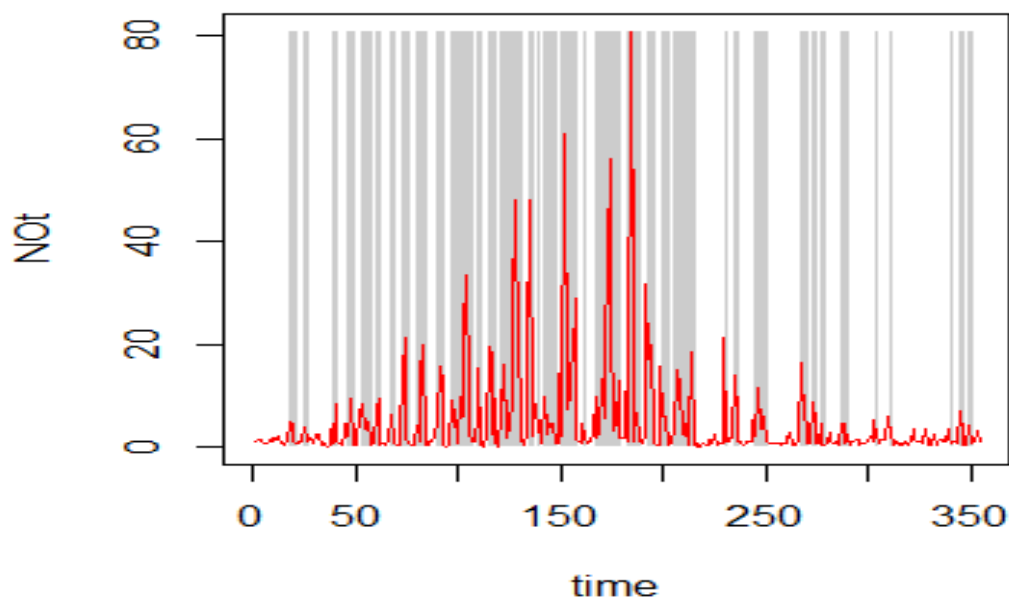


Figure 19 : Représentation des différents régimes dans la série de NO

Sur la figure 18, les zones ombrées représentent le régime 2 où les concentrations sont élevées et les zones claires décrivent le régime 1 : c'est-à-dire les faibles concentrations. On voit que les concentrations du *NO* évolue entre deux régimes alternés.

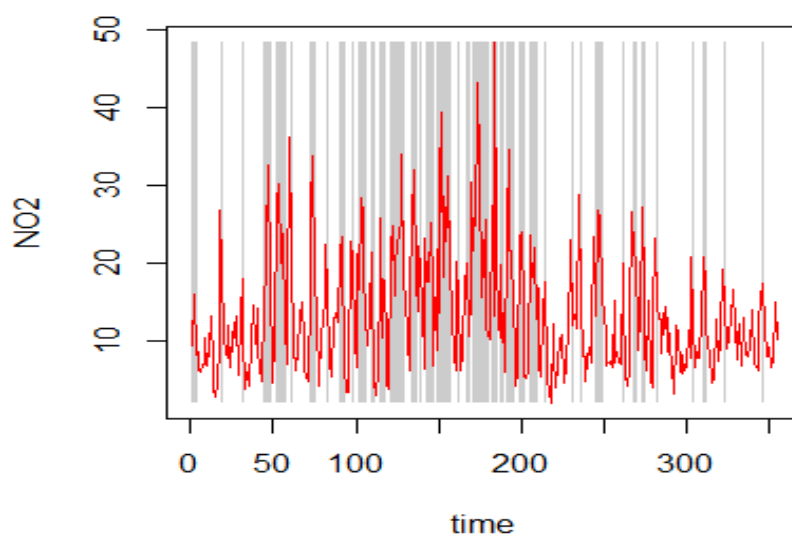


Figure 20 : Représentation des différents régimes dans la série de NO₂

Comme pour les concentrations du *NO*, sur la figure 19, les zones ombrées représentent le régime 2 où les concentrations sont élevées et les zones claires décrivent le régime 1 : c'est-à-dire les faibles concentrations.

➤ **Résultat du modèle MS(3)-AR(1) sur NO_x**

En regardant la matrice de transition du modèle MS(3)AR(1) ci-dessous, nous pouvons suspecter l'existence d'un troisième régime dans la série des concentrations des NO_x.

$$P_{NO} = \begin{pmatrix} 0,709 & 0,235 & 0,055 \\ 0,440 & 0,370 & 0,189 \\ 0,146 & 0,343 & 0,509 \end{pmatrix} \quad \text{et} \quad P_{NO_2} = \begin{pmatrix} 0,586 & 0,252 & 0,161 \\ 0,508 & 0,218 & 0,273 \\ 0,384 & 0,255 & 0,360 \end{pmatrix}$$

Toutes les probabilités de transition et d'appartenance à un régime sont significativement différentes de zéro. Alors les concentrations des NO_x peuvent être suivies pas les modèles à trois régimes.

Dans ce qui suit, nous donnons la formulation des modèles issus de l'estimation pour les concentrations des deux différents composants.

3.1.4. Formulation des modèles

On donne les modèles de chaque polluant qui découlent de l'estimation des paramètres.

De manière générale les modèles MS(2)AR(1) et MS(3)AR(1) sont donnés comme suit :

$$\text{Modèle 1 : } Y_t = \begin{cases} \hat{m}_1 + \hat{\varphi}_{1,1}Y_{t-1} + \varepsilon_t, & \text{si } S_t = 1 \\ \hat{m}_2 + \hat{\varphi}_{1,2}Y_{t-1} + \varepsilon_t, & \text{si } S_t = 2 \end{cases}$$

$$\text{Modèle 2 : } Y_t = \begin{cases} \hat{m}_1 + \hat{\varphi}_{1,1}Y_{t-1} + \varepsilon_t, & \text{si } S_t = 1 \\ \hat{m}_2 + \hat{\varphi}_{1,2}Y_{t-1} + \varepsilon_t, & \text{si } S_t = 2 \\ \hat{m}_3 + \hat{\varphi}_{1,3}Y_{t-1} + \varepsilon_t, & \text{si } S_t = 3 \end{cases}$$

Les séries des concentrations sont stationnaires dans chaque régime.

- Modèle pour le processus de NO :

$$\text{Modèle 1 : } Y_t = \begin{cases} 0,975 + 0,039Y_{t-1} + \varepsilon_t, & \text{si } S_t = 1 \\ 9,390 + 0,299Y_{t-1} + \varepsilon_t, & \text{si } S_t = 2 \end{cases}$$

$$\text{Modèle 2 : } Y_t = \begin{cases} 0,815 + 0,042Y_{t-1} + \varepsilon_t, & \text{si } S_t = 1 \\ 3,53 - 0,05Y_{t-1} + \varepsilon_t, & \text{si } S_t = 2 \\ 15,70 + 0,159Y_{t-1} + \varepsilon_t, & \text{si } S_t = 3 \end{cases}$$

- Modèle pour le processus de NO₂

$$\text{Modèle 1 : } Y_t = \begin{cases} 4,893 + 0,367Y_{t-1} + \varepsilon_t, & \text{si } S_t = 1 \\ 13,659 + 0,441Y_{t-1} + \varepsilon_t, & \text{si } S_t = 2 \end{cases}$$

$$\text{Modèle 2 : } Y_t = \begin{cases} 3,678 + 0,404Y_{t-1} + \varepsilon_t, & \text{si } S_t = 1 \\ 7,99 + 0,598Y_{t-1} + \varepsilon_t, & \text{si } S_t = 2 \\ 17,191 + 0,304Y_{t-1} + \varepsilon_t, & \text{si } S_t = 3 \end{cases}$$

Les figures suivantes donnent l'évolution réelle des données de NO (figure 20) et NO₂ (figure21) et celle des modèles estimés pour le modèle MS(2)AR(1).

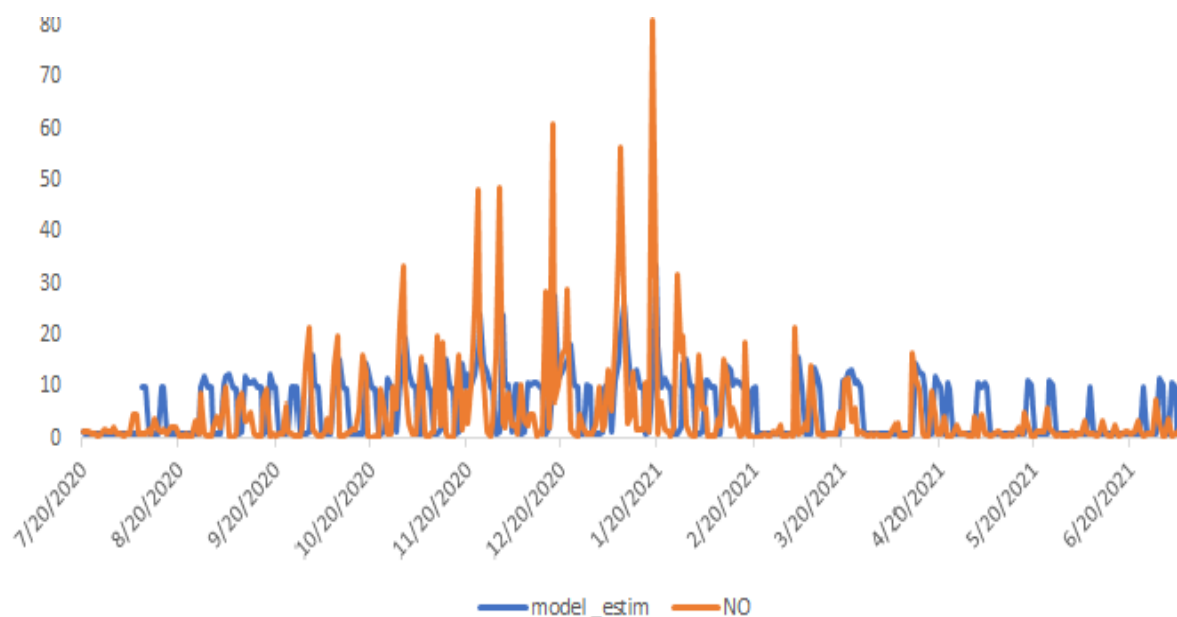


Figure 21 : Représentation du Modèle et des données réelles du NO

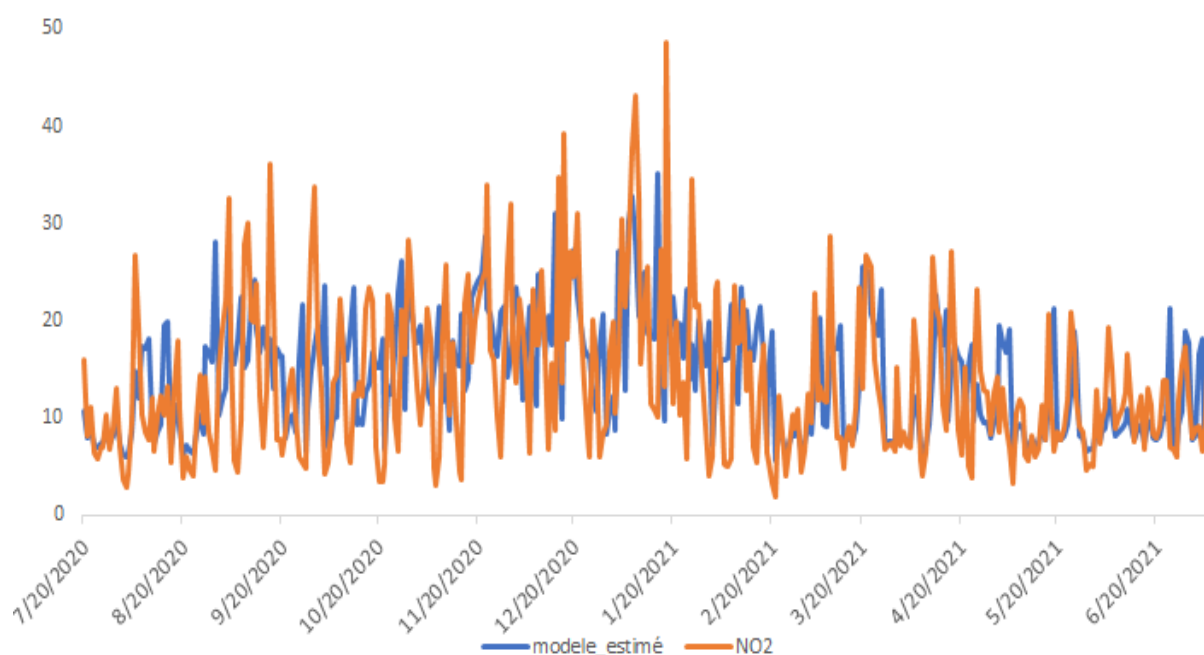


Figure 22 : Représentation du Modèle et des données réelles du NO₂

Les courbes en orange représentent les données réelles et celles en bleu sont issues des modèles. Les modèles 1 et 2 estimés lissent parfaitement les données réelles. Cependant, pour valider les modèles finaux, il est conseillé de tester la robustesse de ces modèles. Pour notre part, nous

avons calculé les erreurs quadratiques de chaque modèle par composantes, puis vérifié quel modèle dont les résidus sont des bruits blancs par le test de Box-Ljung.

3.1.5. Validation du modèle

Le tableau V suivant donne les résultats de validation du modèle.

Tableau V : Résultats de validation du modèle.

Modèles estimés	Erreur quadratique		Test Box-Ljung (p-value)	
	NO	NO ₂	NO	NO ₂
Modèle 1	7,45	8,379	0,75	0,015
Modèle 2	10,05	8,588	$2,156 \cdot 10^{-5}$	0,934

L'étude des erreurs quadratiques pour la série de NO montre que dans le modèle 1, la somme des erreurs quadratiques est plus faible que celle du modèle 2. En plus, le test de bruit blanc montre que les résidus du modèle 1 sont des bruits blancs (iid) car la p-value = 75%.

Pour la série de NO₂, dans le modèle 1, l'erreur quadratique est légèrement inférieure à celle du modèle 2. Cependant, les résidus du modèle 2 sont des bruits blancs (p-value = 93,4%) contrairement au modèle 1 où la p-value est inférieure au seuil de 5%.

De cette analyse, on retient le modèle 1 (deux régimes) pour la série des concentrations de NO et le modèle 2 (trois régimes) pour la série du NO₂.

Les tableaux suivants présentent les résultats de l'analyse statistique par régime pour chaque composante.

Tableau VI : L'analyse statistique par régime (NO₂)

	Régime 1	Régime 2	Régime 3
Moyenne	3,678	7,99	17,191
Médiane	11,600	11,000	13,600
Kurtosis	0,076	0,305	0,724
Skewness	0,842	1,028	1,043
Minimum	1,900	3,000	3,400
Maximum	34,100	36,100	48,500
Etendue	32,600	33,100	45,100
Coefficient de variation	0,529	0,593	0,614

Tableau VII : L'analyse statistique par régime (NO)

	Régime 1	Régime 2
Moyenne	0,975	9,390
Médiane	1,100	2.500
Kurtosis	8,280	11,501
Skewness	2,846	3,076
Minimum	0,2000	0,100
Maximum	21,200	80,900
Etendue	21,000	80,800
Coefficient de variation	1,458	1,613

3.2. Discussion

L'étude des concentrations des NO_x nous a permis de savoir qu'il y a plusieurs régimes dans les séries de ces concentrations. C'est-à-dire deux régimes dans les séries de NO et trois régimes dans celles de NO₂. Les modèles retenus décrivent pratiquement leurs évolutions dans le temps sur la période d'étude. A cet effet, chaque régime correspond à une indicatrice de valeurs de concentrations de ces polluants par rapport à une certaine norme pour la qualité de l'air. On définit alors une valeur limite comme un niveau à atteindre dans un délai donné et à ne pas dépasser, et fixé sur la base des connaissances scientifiques afin d'éviter, de prévenir ou de réduire les effets nocifs sur la santé humaine ou sur l'environnement dans son ensemble.

En situation de santé, pour le dioxyde d'azote, la réglementation a fixé : en moyenne annuelle, la valeur limite est fixée à 40 $\mu\text{g}/\text{m}^3$ selon l'OMS (2006).

Pour notre étude, dans la série des concentrations de NO₂, nous observons les concentrations dans les régimes 1 et 2 qui respectent les normes. C'est-à-dire les concentrations dans ces régimes sont toutes inférieures à la moyenne de 40 $\mu\text{g}/\text{m}^3$. Cependant, dans le troisième régime, où les concentrations sont élevées, on constate deux valeurs au-dessus de cette valeur limite. Il s'agit des concentrations du 1/8/2021 (43,2 $\mu\text{g}/\text{m}^3$) et 1/18/2021 (48,5 $\mu\text{g}/\text{m}^3$). Voir tableaux VI et VII.

Les résultats sur l'étude des concentrations de NO montrent que dans le régime 1, les concentrations en NO sont faibles et donc ne sont pas nocifs pour la santé. Par contre dans le régime 2, où l'étendue est très élevée, les concentrations sont beaucoup dispersées et il y a de fortes valeurs qui seraient nocifs à la santé humaine.

Dans l'ensemble, les concentrations des NO_x, que nous avons étudié dans ce mémoire, respectent la réglementation en ce qui concerne les dangers sur la santé. Même si nous avons

pu identifier les régimes de fortes concentrations dans les différentes séries, il faut noter que ces concentrations restent dans les normes, sauf dans la série de NO₂, où l'on a observé deux dépassements de la valeur limite en 2021. Notre étude est en accord avec nombreuses études faites dans la même région d'Occitanie.

Selon les études de « Atmo Occitanie » (2017) sur l'évaluation de la qualité de l'air sur Toulouse Métropole, en situation de fond urbain, la situation de la qualité de l'air est relativement satisfaisante au regard de la réglementation. Concernant le dioxyde d'azote la situation est globalement stable en 2017, par rapport à 2016.

CONCLUSION GENERALE ET PERSPECTIVES

CONCLUSION

Le travail réalisé dans cette étude s'inscrit dans le cadre général de la modélisation des séries de concentrations des NO_x par des séries chronologiques. La contribution principale de cette étude est d'identifier la non-linéarité dans les séries des concentrations des NO_x troposphériques. Pour cela, l'analyse graphique et la statistique descriptive de ces séries montrent l'existence de deux régimes différents dans leur évolution. Il faut donc faire recours aux modèles qui tiennent compte de cette non-linéarité dans les séries observées.

Les modèles retenus dans la présente étude sont les modèles autorégressifs à changement de régimes markoviens (MSAR). Modèles dans lesquels le changement de régimes est gouverné par une chaîne de Markov cachée.

L'estimation des paramètres des modèles par l'utilisation de l'algorithme EM a permis de retenir le modèle MS(2)AR(1) pour les concentrations du monoxyde d'azote (NO) et le modèle MS(3)AR(1) pour les concentrations du dioxyde d'azote (NO_2), qui expliquent le mieux possible l'évolution des concentrations des NO_x . Ces modèles confirment l'existence de deux (trois) régimes dans les séries de concentrations de NO et NO_2 . Un régime dans lequel on observe les fortes concentrations et un autre régime où les concentrations sont faibles et (les moyennes concentrations). Par une analyse statistique par régime de chaque polluant, nous pouvons dire que les concentrations que nous avons étudiées dans ce travail, respectent globalement les normes en ce qui concernant les dangers sur la santé. Ce résultat est dû à la campagne de réduction des polluants atmosphériques dans la région d'Occitanie.

Dans ce travail, l'étude s'est limitée à l'évolution temporelle des concentrations sans prendre en compte d'autres effets de la nature pouvant influencer cette évolution, alors le modèle utilisé est de type univarié.

Cependant, en perspective nous proposons d'étudier l'évolution de ces polluants par des modèles multivariés qui considéreront par exemple les conditions météorologiques du milieu d'étude. C'est-à-dire étudier l'influence de la température, du vent, de l'humidité relative, de la pluviométrie sur l'évolution des NO_x avec les modèles MSAR.

REFERENCES

REFERENCES

- Ailliot P. (2004). Modèles autorégressifs à changements de régimes markoviens : Application aux séries de vent. Thèse de Doctorat en Mathématiques et Applications, Université de Rennes 1 (Paris, France), 180 p.
- Assamoi A.M. (2011). Emissions anthropiques d'aérosols carbonés en Afrique en 2005 et en 2030 : élaboration d'inventaires et évaluation. Thèse de Doctorat en Physique Chimie de l'Atmosphère, Université de Toulouse (Toulouse, France), 246 p.
- Atmo O. (2017). « Evaluation de la qualité de l'air sur Toulouse Métropole »
Rapport annuel. RA 2017. Consulté le 22 Septembre 2021. <https://www.haute-garonne.gouv.fr/content/download/31792/211352/file/A4%20%20Qualit%C3%A9%20de%20l'air%20Toulouse%20Metropole.pdf>
www.atmo-ociitanie.org
- Camirand J., Gingras C. & Chagnon M.C. (2011). Les changements climatiques : quels en sont les causes et les impacts ?. Nature Québec, Canada, 4 p.
- Chow S.M. & Zhang G. (2013). Nonlinear regime-switching state-space (rsss) models. *Psychometrika*, 78 (4) : 740-768.
- Dempster A.P., Laird N.M. & Rubin D.B. (1977). Maximum Likelihood from Incomplete data via the EM Algorithm. *Journal of the Royal Statistical Society*, 39(1) : 1-38.
- Dijk D., Teräsvirta & Franses P. (2002). Smooth Transition Autoregressive Models-A Survey Of Recent Developments. *Econometric Reviews*, 21(1) : 1-47.
- Douc R., Moulines E. & Rydén T. (2004). Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime. *The Annals of Statistics*, 32(5) : 2254-2304.
- Fofana S. & Diongue A.K. (2013). Comportement de la fonction d'autocorrélation des modèles à changement de régimes : une approche empirique. *Afrika Statistika*, 8 : 515-544.

- Fouquau J. (2008). Modèles à changement de régimes et données de panel : de la non-linéarité à l'hétérogénéité. Thèse de Doctorat en Sciences Economiques, Université d'Orléans (Orléans), 304 p.
- Girard Y. (2011). Series chronologiques à une et plusieurs variables : synthèse des méthodes classiques et modèles à base de copules. Mémoire de Master en Mathématiques et Informatique Appliquées, Université du Québec à Trois-Rivières, Canada, 115 p.
- Goldfeld S. & Quandt R. (1973). The estimation of structural shifts by switching regressions. *Annals of Econometric and Social Measurement*, 2(4) : 475-485.
- Grégoire G. (2019). Sur les modèles non-linéaires autorégressifs à transition lisse et le calcul de leurs prévisions. Mémoire de Master en Faculté des Arts et des Sciences, Université de Montréal, Canada, 103 p.
- Hamilton J.D. (1989). A new approach of economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57 (2) : 357-384.
- Hamilton J.D. (1990). Analysis of times series sujet to changes in regime. *Journal of Econometrics*, 45 : 39-70.
- Ilham Z. & Nadjat S. (2020). Estimation des Modèles Autoregréssifs à Changement de Régime Markovien. Mémoire de master en Probabilités et Statistique, Université Mohammed Seddik Ben Yahia de Jijel, Algérie, 44 p.
- Kljakic P.V. (2012). Modélisation des oxydes d'azote et des suies dans les moteurs Diesel. Thèse de Doctorat en Energetique, Laboratoire Energetique Moléculaire et Macroscopique Combustion (Paris, France), 214 p.
- Kock A.B. & Teräsvirta T. (2011). Forecasting with nonlinear time series models. *The Oxford Handbook of Economic Forecasting*, 34 : 61-87
- Olteneanu M. (2006). Modèles à changement de régime, application aux données financières. Thèse de Doctorat en Mathématiques Appliquées, Université Paris 1-Panthéon-Sorbonne (Paris, France), 131 p.
- Ouaret R. (2016). Déterminants et prévision des fluctuations de la concentration en polluants dans un environnement interieur. Thèse de Doctorat en Sciences et Techniques de l'Environnement, Université Paris-Est (Paris, France), 394 p.

Pélé R. (2020). Modélisation et simulation aux grandes échelles des émissions de NO_x dans un brûleur sémi-industriel. Rapport de Stage Institut National des Sciences Appliquées, Toulouse (France), 61 p.

Surprenant S. (2018). La propagation des récessions au canada. Mémoire de Master en Economique, Université du Québec à Montréal, Canada, 125 p.

ANNEXES

ANNEXES

Annexe : Table de régime de NO

Pour bien observer la dominance du premier régime dans cette zone d'étude, nous pouvons examiner la figure suivante qui donne une indication de chaque observation y_t à son régime d'appartenance. Par exemple pour la série des concentrations du NO.

```

[1] NA 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1
[34] 1 1 1 1 1 1 1 1 2 2 1 1 1 2 2 2 2 2 1 1 2 2 2 2 2 1 1 1 2 2 2 1 1
[67] 1 1 2 2 1 1 1 1 2 2 2 1 1 1 1 2 2 2 2 2 1 1 1 1 1 2 2 2 2 1 1 1 2
[100] 2 1 1 2 2 2 2 2 2 1 1 2 2 2 1 1 1 2 2 2 2 1 1 2 2 2 2 2 2 2 2 2 2
[133] 2 1 1 2 2 2 1 2 2 1 1 2 2 2 2 2 1 1 1 2 2 2 2 2 2 2 2 2 2 1 1 1 2 1
[166] 1 1 1 2 2 1 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2 2 2 2 2 1 1 1 2 2 2 2 1
[199] 1 1 2 2 2 1 1 1 1 2 2 2 2 2 2 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2
[232] 2 1 1 1 2 2 1 1 1 1 1 1 1 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[265] 1 1 1 2 2 2 2 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[298] 1 1 1 1 1 1 2 2 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[331] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 2

```

Figure : Table des régimes pour NO

Cette table confirme la dominance des observations des concentrations du régime 1 qui a été détectée par l'étude des probabilités pour les concentrations de NO. Sa particularité est qu'il attribue chaque donnée à son état d'appartenance instant par instant. Par exemple, parmi les 32 premières observations de notre série, seulement deux sont susceptibles d'appartenir au régime 2.

Résumé

Les variabilités temporelles des GES dans l'atmosphère provoquent des modifications du climat. On en observe déjà les impacts, avec des événements qui se produisent à une vitesse et une ampleur plus élevée que prévues. En effet, dans leur évolution, les concentrations peuvent être fortes ou faibles. Ce comportement ne pouvant pas être suivi par des modèles linéaires, il serait judicieux d'utiliser des modèles qui tiennent compte de ces non-linéarités. Ce mémoire de Master a pour but principal d'identifier des régimes dans les séries des concentrations des NO_x (NO et NO₂) et de modéliser leur évolution temporelle en considérant leur changement structurel par les MSAR. Une analyse de statistique descriptive a été faite pour décrire le comportement des séries de concentrations des NO_x. Les données d'étude sont issues de la station Toulouse-Jaquier Urbain (France). Les résultats de notre étude ont permis d'identifier deux régimes (fortes et faibles) concentrations pour NO et trois régimes pour NO₂. Selon une analyse statistique par régime, nous remarquons que sur la période d'étude, la pollution en NO_x par rapport à la santé humaine est globalement dans les normes. Les paramètres des modèles ont été estimés par la méthode de maximum de vraisemblance développée par l'algorithme EM.

Mots clés : Changement climatique, Concentration des NO_x, Modèles autorégressifs, Toulouse-Jaquier Urbain, Modèles à changement de régime markoviens.

Abstract

Temporal variability of GHGs in the atmosphere is causing climate change. The impacts are already being observed, with events occurring at a higher rate and magnitude than expected. Indeed, in their evolution, concentrations can be high or low. As this behaviour cannot be followed by linear models, it would be wise to use models that take these non-linearities into account. The main goal of this Master thesis is to identify regimes in the NO_x (NO and NO₂) concentration series and to model their temporal evolution by considering their structural change through MSARs. A descriptive statistics analysis was performed to describe the behaviour of the NO_x concentration series. The study data were obtained from the Toulouse-Jaquier Urbain station (France). The results of our study identified two regimes (high and low) for NO and three regimes for NO₂. According to a statistical analysis by regime, we note that over the study period, NO_x pollution in relation to human health is globally within the standards. The parameters of the models were estimated by the maximum likelihood method developed by the EM algorithm.

Keywords: Climate change, NO_x concentration, Autoregressive models, Toulouse-Jaquier Urbain, Markov Regime switching models.