



HAL
open science

Modèles semi-paramétriques appliqués à la prévision des séries temporelles. Cas de la consommation d'électricité.

Vincent Lefieux

► To cite this version:

Vincent Lefieux. Modèles semi-paramétriques appliqués à la prévision des séries temporelles. Cas de la consommation d'électricité.. Mathématiques [math]. Université Rennes 2, 2007. Français. NNT : . tel-00179866

HAL Id: tel-00179866

<https://theses.hal.science/tel-00179866>

Submitted on 16 Oct 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ RENNES 2 – HAUTE BRETAGNE
Unité de Recherche
Ecole Doctorale - Humanités et Sciences de l'Homme

Modèles semi-paramétriques appliqués à la prévision des séries temporelles

Cas de la consommation d'électricité

Thèse de Doctorat

Discipline : Statistiques

Présentée par Vincent LEFIEUX

Directeurs de thèse :

Michel CARBON et Michel DELECROIX

Soutenue le 12 octobre 2007

Jury :

M.	BIAU	Gérard	<i>Rapporteur</i>	Université Pierre et Marie Curie
Mme	BONGRAIN	Marie-Pierre		RTE – DMA
M.	BOSQ	Denis		Université Pierre et Marie Curie
M.	CARBON	Michel		Université Rennes 2
M.	DELECROIX	Michel		Université Pierre et Marie Curie
M.	SIMAR	Léopold	<i>Rapporteur</i>	Université Catholique de Louvain

Table des matières

Remerciements	ix
Introduction générale	xi
1 Prévoir la consommation d'électricité	1
1.1 Problématique	1
1.2 Solution technique adoptée	3
1.2.1 Démarche	3
1.2.2 Analyse de la courbe de charge	6
1.2.3 Principe	11
1.2.4 Modèles utilisés	12
1.2.5 Récapitulatif	17
1.3 Perspectives	17
1.4 Problématique	18
2 Prévisions autorégressives non et semi-paramétriques	21
2.1 Le paradigme non-paramétrique	21
2.2 Prédiction autorégressive à l'aide de la méthode des noyaux	23
2.2.1 Modèle	23
2.2.2 Convergence des méthodes à noyau	25
2.2.3 Mise en œuvre	25
2.3 Le paradigme semi-paramétrique	28
2.4 La méthode MAVE	30
2.4.1 Principe	30
2.4.2 Choix des poids	32
2.4.3 Choix de la réduction de dimension	33
2.4.4 Résultats de convergence	34
2.5 Prédiction autorégressive à l'aide de la méthode MAVE	35
2.5.1 Principe	35

2.5.2	Mise en œuvre	36
2.6	Application à des séries temporelles classiques	38
2.6.1	Démarche	38
2.6.2	Série AR(2)	39
2.6.3	Série ARMA(1,1)	41
2.6.4	Séries "Cigars"	43
2.6.5	Série "Profit margin"	45
2.6.6	Série "Coal"	47
2.7	Application à la consommation d'électricité	49
3	Prévisions semi-paramétriques en présence de variables exogènes	55
3.1	Problématique	55
3.2	Proposition d'un modèle alternatif	58
3.2.1	Introduction	58
3.2.2	Modèle	59
3.3	Estimation du modèle	60
3.3.1	Estimation conditionnelle	60
3.3.2	Estimation globale (PLMAVE)	63
3.4	Problèmes de convergence	65
3.5	Simulations	66
3.5.1	Simulations sur données indépendantes	66
3.5.2	Simulations sur données dépendantes	69
3.6	Prévision à l'aide de la méthode PLMAVE	72
3.7	Application à la consommation d'électricité	73
4	Annexe : De l'ergodicité de modèles autorégressifs	79
4.1	Introduction	79
4.2	Cas d'un modèle non-paramétrique	82
4.3	Cas d'un modèle semi-paramétrique	89
4.4	Cas hétéroscédastique	91
	Conclusion	97

Table des figures

1	Marché de l'électricité	xi
1.1	Réseau de transport 400kV et 225 kV	2
1.2	Prévision journalière de la consommation d'électricité	4
1.3	Consommation journalière d'électricité	7
1.4	Consommation hebdomadaire d'électricité	7
1.5	Consommation annuelle d'électricité	7
1.6	Consommation annuelle d'électricité (haut) et température annuelle (bas)	8
1.7	Impact de la nébulosité sur la consommation d'électricité	9
1.8	Impact du tarif EJP sur la consommation d'électricité	10
1.9	Principe actuel de la prévision de la consommation d'électricité	12
1.10	Dépendance horaire de la consommation d'électricité et de la température	14
1.11	Dépendance horaire de la consommation d'électricité et de la température lissée	15
1.12	Principe de la régression liant la consommation d'électricité à la température	16
2.1	Critère de validation croisée MAVE : AR(2)	39
2.2	Prévision optimale MAVE : AR(2)	40
2.3	Prévision optimale MAVE (détail) : AR(2)	40
2.4	Critère de validation croisée MAVE : ARMA(1,1)	41
2.5	Prévision optimale MAVE : ARMA(1,1)	42
2.6	Prévision optimale MAVE (détail) : ARMA(1,1)	42
2.7	Critère de validation croisée : Cigars	43
2.8	Prévision optimale MAVE : Cigars	44
2.9	Prévision optimale MAVE (détail) : Cigars	44
2.10	Critère de validation croisée MAVE : Profit Margin	45
2.11	Prévision optimale MAVE : Profit margin	46

2.12	Prévision optimale MAVE (détail) : Profit margin	46
2.13	Critère de validation croisée MAVE : Coal	47
2.14	Prévision optimale MAVE : Coal	48
2.15	Prévision optimale MAVE (détail) : Coal	48
2.16	Consommations prévues MAVE (rouge) et réalisées (noir) : Période 1, première semaine tronquée	51
2.17	Consommations prévues MAVE (rouge) et réalisées (noir) : Période 1, seconde semaine tronquée	51
2.18	Consommations prévues MAVE (rouge) et réalisées (noir) : Période 2, première semaine tronquée	52
2.19	Consommations prévues MAVE (rouge) et réalisées (noir) : Période 2, seconde semaine tronquée	53
3.1	Critères de validation croisée : $n = 500$; $d \in \{3, 4, 5\}$	67
3.2	Critères de validation croisée : $n \in \{200, 800\}$; $d \in \{3, 4, 5\}$	68
3.3	Critères de validation croisée : $n = 500$; $d \in \{1, 2, 3\}$	70
3.4	Critères de validation croisée : $n \in \{200, 800\}$; $d \in \{1, 2, 3\}$	71
3.5	Consommations prévues PLMAVE (rouge) et réalisées (noir) : Période 1, première semaine tronquée	75
3.6	Consommations prévues PLMAVE (rouge) et réalisées (noir) : Période 1, seconde semaine tronquée	76
3.7	Consommations prévues PLMAVE (rouge) et réalisées (noir) : Période 2, première semaine tronquée	76
3.8	Consommations prévues PLMAVE (rouge) et réalisées (noir) : Période 2, seconde semaine tronquée	77
4.1	Ecart-type horaire de la consommation d'électricité : 03/04/06-09/07/06	92
4.2	Ecart-type horaire de la consommation d'électricité : 15/01/07-08/04/07	92

Liste des tableaux

2.1	Prévisions optimales "MAVE" et "Noyau" : AR(2)	39
2.2	Prévisions optimales "MAVE" et "Noyau" : ARMA(1,1)	41
2.3	Prévisions optimales "MAVE" et "Noyau" : Cigars	43
2.4	Prévisions optimales "MAVE" et "Noyau" : Profit margin	45
2.5	Prévisions optimales "MAVE" et "Noyau" : Coal	47
2.6	Erreurs de prévision : Période 1	50
2.7	Critères d'erreurs de prévision : Période 1	50
2.8	Erreurs de prévision : Période 2	52
2.9	Critères d'erreurs de prévision : Période 2	52
3.1	Erreurs de prévision : Période 1	74
3.2	Critères d'erreurs de prévision : Période 1	74
3.3	Erreurs de prévision : Période 2	75
3.4	Critères d'erreurs de prévision : Période 2	75

Remerciements

Michel Delecroix m'a permis de découvrir les richesses de la Statistique et a été à l'origine de cette thèse. Je tiens à le remercier chaleureusement pour son soutien et ses conseils avisés durant ces années. Qu'il me soit également permis d'exprimer ma gratitude à Michel Carbon qui a accepté d'encadrer cette thèse.

Je suis sincèrement reconnaissant à Gérard Biau et Léopold Simar d'avoir accepté de consacrer un temps précieux pour rapporter sur ma thèse. Je remercie également Denis Bosq pour sa participation au jury.

Je ne saurais oublier tous les membres, actuels et anciens, du CREST ENSAI et du Laboratoire de Statistiques de Rennes 2, qui ont su m'apporter leur appui et leur amitié, au cours de mes travaux de recherche.

Je tiens également à remercier Anestis Antoniadis pour le temps qu'il m'a accordé lors de discussions fort enrichissantes, et très sympathiques.

J'ai débuté cette thèse au sein d'EDF, dans le département MOSSE, et l'ai poursuivie parmi mes collègues du département DMA de RTE. Je suis très sensible au soutien reçu et je ne saurais l'oublier. Je tiens à remercier tout particulièrement Marie-Pierre Bongrain, Jérôme Ryckbosch et Christophe Crocombette d'avoir permis que cette thèse aboutisse.

Enfin, il me reste à exprimer ma gratitude à ma famille et mes amis, qui ont supporté mes états d'âme avec patience.

Introduction générale

L'ouverture du marché de l'électricité en France a conduit à séparer les activités de production, de transport et de commercialisation. Dans ce cadre, RTE est en charge de la gestion du réseau de Transport et doit à ce titre :

- gérer les infrastructures de réseau
RTE doit entretenir le réseau au coût le plus juste, renforcer sa robustesse et le développer en fonction de la demande
- gérer les flux d'électricité sur le réseau
RTE doit veiller à la sécurité d'approvisionnement et alerter les pouvoirs publics en cas de risque de rupture.

La figure 1 schématise les interactions entre les différents acteurs du marché électrique.

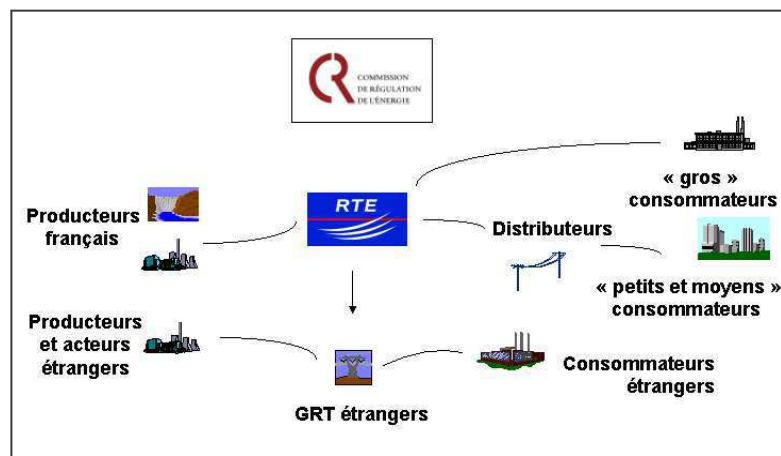


FIG. 1 – Marché de l'électricité

La sûreté de fonctionnement du système électrique, au cœur des responsabilités confiées par la loi à RTE, oblige ce dernier à disposer d'outils statistiques per-

formants. La contrainte majeure est que la production doit, à tout moment, être strictement égale à la consommation, l'électricité ne se stockant pas.

RTE établit donc des prévisions de consommation à différentes échelles de temps, en prenant en compte des aléas pouvant toucher certains producteurs ou certains consommateurs. RTE dispose notamment d'un modèle de prévision à court terme qui utilise les valeurs des températures et nébulosités pour 26 villes représentatives, réparties sur le territoire français. A partir de ces données et également des réalisations du passé, le modèle fournit une prévision pour la journée du lendemain. La prévision définitive est construite par un prévisionniste de RTE qui complète les données par des informations plus qualitatives (hauteur de la couche nuageuse, tarifs. . .).

L'ouverture du marché va conduire le gestionnaire du réseau de transport, à considérer de nouvelles variables explicatives, par exemple les prix de marché (qui peuvent modifier la structure de la consommation). Le modèle actuellement implémenté, de type paramétrique, est peu évolutif et RTE désire accroître ses connaissances afin de disposer de techniques plus souples. L'intérêt s'est porté sur les techniques semi-paramétriques qui ont fait l'objet de développements théoriques assez récents. Ces techniques présentent l'avantage de pouvoir utiliser le cadre général du non-paramétrique tout en palliant son inconvénient majeur : le fléau de la dimension. Une première étude bibliographique nous a permis de constater que ces méthodes ont essentiellement été étudiées dans le cas indépendant. Ce travail étudie l'extension et l'application de quelques modèles semi-paramétriques aux séries temporelles, l'objectif final étant d'obtenir une prévision de la consommation d'électricité.

Le premier chapitre aborde de manière détaillée la prévision de la consommation électrique à court terme à RTE. On présente dans ce cadre les variables d'intérêt : la puissance électrique, ainsi que les variables tarifaires et climatiques ayant un impact sur la puissance. Le besoin de méthodes adaptatives souples est mis en évidence dans ce chapitre.

Le second chapitre considère le problème classique de la prévision de la valeur future d'une série temporelle autorégressive. Longtemps exclusives et encore majoritaires, les modélisations paramétriques de type Box et Jenkins s'avèrent souvent rigides et finalement peu adaptées à la structure de certaines séries observées. La méthodologie non-paramétrique de prévision a permis de considérer un prédicteur efficace dans une classe plus large de processus. On sait cependant que la qualité de l'estimateur résiste mal à un grand nombre de régresseurs, ce qu'on appelle communément le fléau

de la dimension. Nous proposons d'utiliser une méthode semi-paramétrique qui améliore l'approche non-paramétrique pure. La méthode MAVE, basée sur la notion de "directions révélatrices", peut s'appliquer aux séries temporelles. Après avoir défini cette méthode dans le cadre de la régression classique, nous étudions empiriquement son efficacité pour prédire les valeurs futures de quelques séries temporelles.

Le troisième chapitre propose un modèle semi-paramétrique applicable à la consommation d'électricité. La méthodologie MAVE ne permet pas en pratique d'obtenir des résultats satisfaisants pour la consommation d'électricité si on l'utilise en considérant comme régresseurs à la fois le passé de la série et les variables exogènes (climatiques et tarifaires). Pour pallier ce problème, nous proposons un modèle semi-linéaire à directions révélatrices multiples. Après avoir indiqué le mode d'estimation de ce modèle, on étudie son comportement empirique sur des données simulées, ainsi que sur la puissance électrique.

Le quatrième chapitre est une annexe technique qui fournit quelques propriétés de modèles autorégressifs non et semi-paramétriques. Les hypothèses de stationnarité et de mélangeance sont en effet requises pour que les estimateurs soient convergents. Nous utilisons la notion d'ergodicité géométrique pour établir ces résultats.

Chapitre 1

Prévoir la consommation d'électricité

1.1 Problématique

L'électricité ne se stockant pas, la production globale sur le réseau électrique français doit, à tout moment, être strictement égale à la consommation. Toute modification de la demande ou de la production d'électricité en un point du réseau se répercute instantanément sur tout le système électrique. Celui-ci doit donc s'adapter en permanence pour satisfaire l'équilibre offre-demande.

Prendre de "bonnes" décisions en terme de gestion d'un système électrique n'est guère possible si l'on ne dispose pas de techniques efficaces permettant d'en modéliser les aléas, de modéliser l'effet des différentes variables intervenant dans le problème, et d'en faire une prévision aussi précise que possible. En particulier, l'étude et la prévision de la consommation d'électricité est indispensable pour que RTE (Réseau de Transport d'Electricité) puisse garantir à la fois l'équilibre offre/demande à court terme et une exploitation sûre du réseau de transport français (dont la complexité est détaillée dans la figure 1.1).

Les prévisions de consommation doivent être faites à différents horizons de temps : à court terme (du lendemain à la semaine qui suit), à moyen terme (de quelques mois à 1 an), à long terme (pour plusieurs années).

En l'absence de prévision à long terme, RTE ne disposerait pas d'une vision prospective sur la quantité et la localisation de l'énergie à transporter, et les investissements

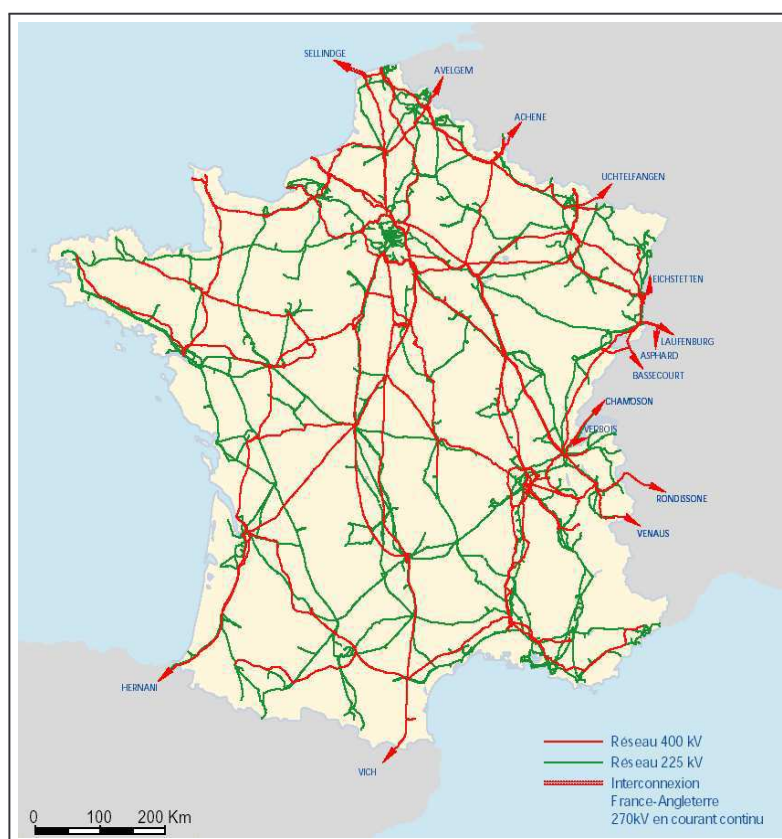


FIG. 1.1 – Réseau de transport 400kV et 225 kV

réseaux nécessaires ne pourraient pas être décidés et construits suffisamment longtemps à l'avance pour pouvoir être mis en exploitation en temps utile.

Les prévisions à moyen terme sont indispensables pour pouvoir trouver des périodes durant lesquelles pourront être réalisées des opérations d'entretien de lignes ou de postes. Ces opérations immobilisent le réseau puisqu'elles conduisent à mettre hors service pour plusieurs semaines l'un de ses éléments, il faut donc vérifier par des calculs de réseau que l'exploitation restera suffisamment sûre pendant les périodes à venir. Les prévisions de consommations permettent de déterminer les injections qui sont utilisées dans ces calculs de réseau.

Enfin, la prévision à court terme est un facteur essentiel pour assurer l'équilibre entre la production et la consommation durant la journée. Chaque jour, la veille pour le lendemain, les producteurs communiquent leurs programmes de marche, c'est-à-dire des courbes indiquant ce qui sera fourni tout au long de la journée. Ces programmes de marche résultent des contrats de fourniture qu'ils ont passés avec leurs clients. On pourrait penser que la somme de ces programmes de marche fournit une image précise de la consommation du lendemain et qu'il est donc inutile que RTE réalise une prévision propre. En réalité ce n'est pas le cas, car il est techniquement plus difficile de faire des prévisions sur un ensemble de clients qu'à l'échelle d'un pays, comme l'illustre la figure 1.2. La prévision faite par RTE permet donc de garantir l'équilibre physique global entre l'offre et la demande à l'échelle du pays. De sa qualité dépendent le niveau des marges d'exploitation nécessaires et plus largement la sûreté du système.

La prévision de consommation est un exercice délicat qui fait appel à une analyse approfondie et à la mise en œuvre de techniques numériques avancées. La problématique la plus difficile techniquement, et également celle pour laquelle l'erreur doit être la plus petite possible est la prévision journalière. Les travaux menés dans le cadre de cette thèse concernent uniquement la prévision à court terme de la consommation d'électricité.

1.2 Solution technique adoptée

1.2.1 Démarche

Les techniques de prévision, si elles sont massivement utilisées par les différents acteurs du marché de l'électricité, le sont également en climatologie, en qualité de l'air,

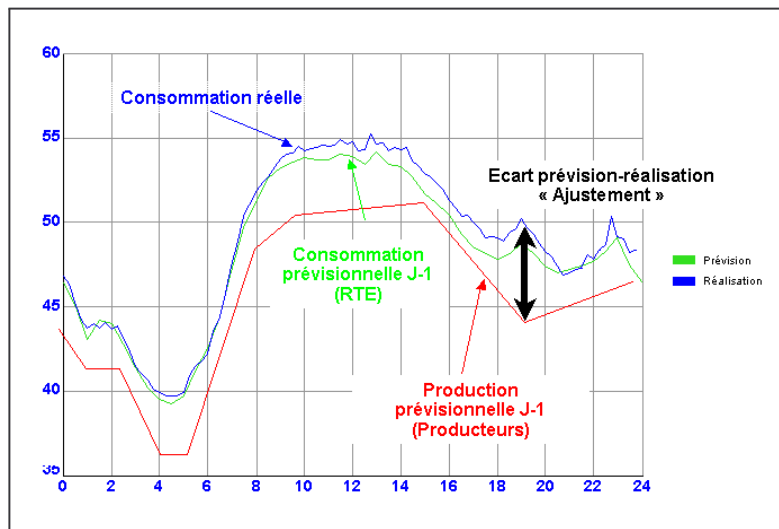


FIG. 1.2 – Prévision journalière de la consommation d'électricité

en finance, dans le domaine commercial... ; il n'en reste pas moins que la problématique de l'électricité présente bel et bien des spécificités, même si sa résolution technique reste classique.

Différents points de vue coexistent afin de prévoir une série temporelle, mais, grosso modo, la démarche est toujours la même : pondérer des motifs observés dans le passé (des formes de courbe réalisées) afin d'en déduire de nouvelles données pour le futur (des courbes prévues).

On peut ainsi distinguer :

- Les techniques statistiques.
- Les techniques issues du domaine de l'intelligence artificielle : on utilise ainsi les réseaux de neurones, la logique floue... Ces techniques élaborent également un système de pondération, non basé sur des résultats théoriques mais sur un ajustement numérique optimum.
- Les heuristiques : ces règles *a priori* sont souvent utilisés par les prévisionnistes ayant une grande expérience de la série temporelle observée.

On constate très souvent la juxtaposition de ces techniques en opérationnel, ce qui s'explique très largement par la non-prédominance d'une technique sur l'autre. Les utiliser conjointement permet parfois de pallier leurs inconvénients respectifs. Par exemple les techniques statistiques sont souvent considérées comme rigoureuses mais exigeantes en hypothèses souvent difficiles à vérifier en pratique. A contrario, le

carcan théorique des réseaux de neurones est moindre, mais ils ont un aspect "boîte noire" indéniable, qui est peu compatible avec le besoin de lisibilité et de traçabilité des décisions qui seront prises sur la base de la prévision.

Dans le domaine électrique, les réseaux de neurones et les méthodes statistiques sont utilisées à proportion égale mais des heuristiques sont toujours employées pour prévoir par exemple les consommations des jours fériés et autres périodes spéciales. RTE a fait le choix d'utiliser majoritairement l'analyse statistique afin d'élaborer ses modèles de prévision. Le modèle ne constitue cependant qu'un outil d'aide à la décision et les répartiteurs du CNES (Centre National d'Exploitation du Système) sont chargés de fournir la prévision finale, en prenant en compte les dernières informations reçues et les modifications inopinées de la demande.

La démarche statistique classique adoptée par EDF (et RTE) consiste à :

- Analyser la courbe de charge réalisée, en identifiant notamment les dépendances temporelles observées dans le passé, le lien avec des variables exogènes (comme la température, la nébulosité, certains tarifs ou prix...).
- Modéliser le phénomène observé : on cale une équation mathématique qu'on valide par la suite à l'aide d'indicateurs statistiques. La forme de l'équation provient de la phase d'analyse précédente. On utilise une équation du type :

$$P_t = f(P_{t-1}, P_{t-2}, P_{t-3}, \dots, Z_t, Z_{t-1}, Z_{t-2}, \dots) + \varepsilon_t$$

où P_t désigne la puissance consommée à l'instant t , Z_t les variables exogènes au problème, ε_t l'erreur traduisant les aléas, et f la "fonction de lien" à estimer.

- "Inférer" le modèle élaboré : le modèle calé dans le passé est "extrapolé" de manière à être utilisable pour le futur, pour générer des prévisions.
- Une fois ces différentes phases réalisées, on doit valider l'outil fourni dans son fonctionnement opérationnel ; l'analyse a posteriori sur les erreurs du modèle permet de le faire évoluer.

En pratique, il existe différentes possibilités pour estimer f .

- Statistique paramétrique : l'estimation de la fonction f se ramène à l'estimation d'un paramètre (vectoriel).
- Statistique non-paramétrique : f appartient à une classe fonctionnelle large sans hypothèses trop contraignantes. Ces méthodes ont comme avantage d'être plus flexibles ; leur grand inconvénient (par exemple pour la méthode des noyaux) est de ne pas converger rapidement dès que le nombre de variables à prendre en compte est grand.

- Statistique semi-paramétrique : les méthodes de réduction de la dimension permettent d'utiliser la souplesse des modèles non-paramétriques tout en éliminant le problème de la dimension. Elles n'ont pas encore été mises en œuvre sur la consommation d'électricité.

On indique par la suite la solution paramétrique adoptée en pratique, ainsi que certaines alternatives déjà envisagées. La phase d'analyse nous sera utile pour élaborer un modèle semi-paramétrique.

1.2.2 Analyse de la courbe de charge

La consommation d'électricité varie en fonction de plusieurs paramètres, dont les principaux sont la saisonnalité, le climats et les tarifs/prix. D'autres variables explicatives existent : des événements exceptionnels peuvent perturber le profil de la consommation (tempêtes du mois de décembre 1999, finale de la coupe du monde de football en 1998), mais leur impact est impossible à prévoir.

Impact de la saisonnalité

On identifie trois cycles temporels :

- un cycle journalier, avec une pointe le matin vers 9h et le soir vers 19h (Cf. Fig. 1.3) ;
- un cycle hebdomadaire, avec 5 jours ouvrables globalement stables et le week-end où la consommation diminue (Cf. Fig. 1.4) ;
- un cycle annuel, avec une pointe de consommation en janvier et un creux vers le 15 août (Cf. Fig. 1.5).

De manière sous-jacente, le cycle annuel dépend de l'activité économique (creux annuel au moment des vacances d'été et hebdomadaire le week-end). Considérer un modèle autorégressif pour les prévisions à court terme permet de prendre en compte les impacts économiques. Remarquons également l'impact du changement d'horaire légal ; par exemple, l'horaire d'été supprime pratiquement la pointe de consommation du soir).

Impact du climat

L'impact du climat doit également être considéré si on veut modéliser correctement la consommation d'électricité.

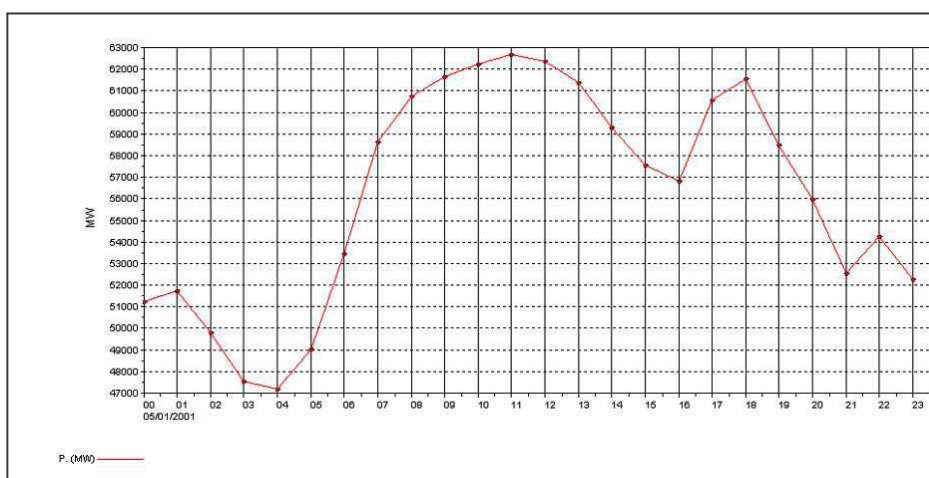


FIG. 1.3 – Consommation journalière d’électricité

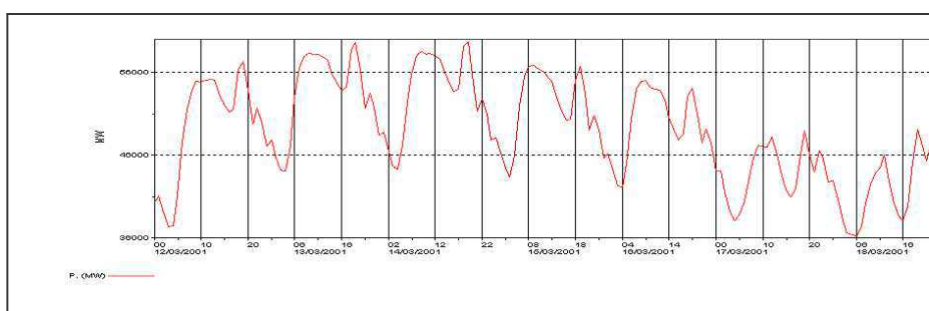


FIG. 1.4 – Consommation hebdomadaire d’électricité

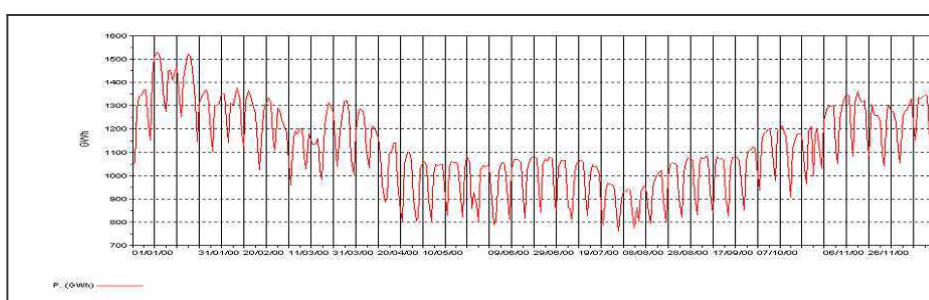


FIG. 1.5 – Consommation annuelle d’électricité

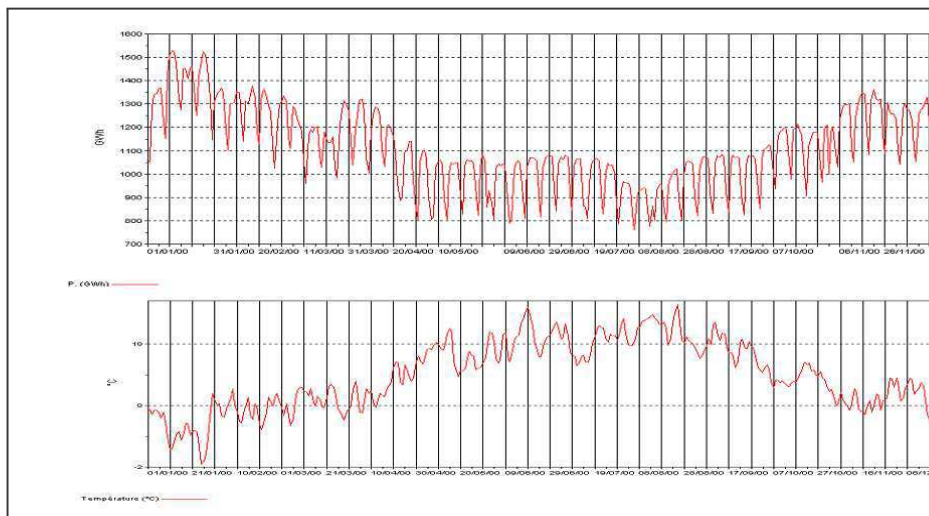


FIG. 1.6 – Consommation annuelle d'électricité (haut) et température annuelle (bas)

Les variations de température se traduisent par des modifications de l'utilisation du chauffage électrique en hiver et de la climatisation en été. Le développement du chauffage électrique dans les années 1970 et plus récemment celui de la climatisation sont à l'origine d'une grande élasticité de la consommation d'électricité à la température.

En hiver par exemple, une variation d'un degré sur l'ensemble du territoire entraîne une augmentation de la consommation équivalente à la consommation de l'agglomération lyonnaise.

La figure 1.6 superpose les variations annuelles de la consommation d'électricité et de la température. On pourrait imaginer que la saisonnalité puisse "modéliser" l'impact des températures, malheureusement, pour des prévisions journalières, on obtiendrait des taux d'erreur inacceptables si on ne modélisait pas précisément les variations des usages sensibles à l'aléa climatique.

Les électriciens savent que la consommation dépend également du passé climatique. Une variation de température agira en effet avec un certain retard sur la demande d'électricité, principalement à cause des temps de réaction humains (inertie de court terme : quelques heures) et à cause de l'isolation thermique des bâtiments (inertie de moyen terme : de quelques heures à une journée).

Une autre variable explicative est la nébulosité (indice de couverture nuageuse¹).

¹La nébulosité est mesurée en octa et est comprise entre 0 (ciel complètement dégagé) et 8 (ciel

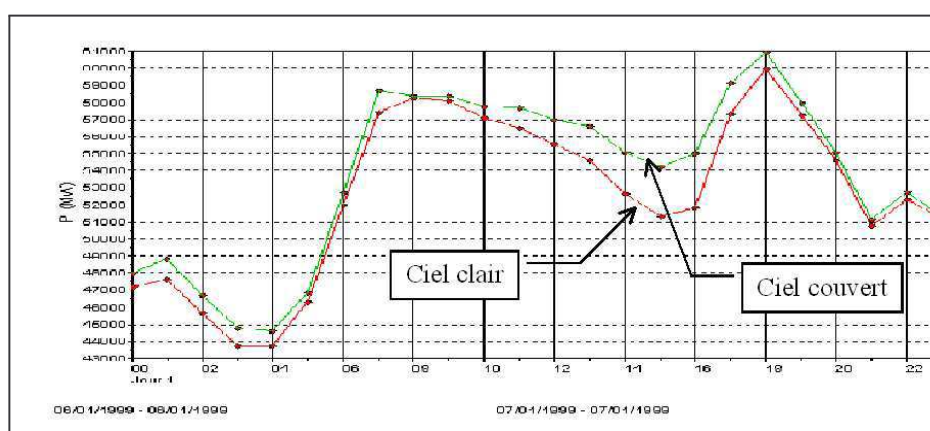


FIG. 1.7 – Impact de la nébulosité sur la consommation d'électricité

La nébulosité est considérée ici simplement comme facteur influençant le volume de chauffage : grâce à l'effet de serre, une nébulosité faible se traduit par un apport calorifique supplémentaire à l'intérieur des bâtiments, et donc par un chauffage moindre. La figure 1.7 met en évidence ce phénomène en juxtaposant deux courbes de charge journalières équivalentes (jour de semaine, températures) pour lesquelles seule la nébulosité diffère de manière importante.

En conclusion, seule la température ressentie nous intéresse et c'est à ce titre qu'on considère la température instantanée et passée, ainsi que la nébulosité.

Impact des tarifs et des prix

Une partie de la demande peut diminuer en fonction des prix et des tarifs de l'électricité. Pour l'instant seul un tarif est significatif d'un point de vue statistique : le tarif EJP (Effacement Jour de Pointe) qui assure un tarif bas toute l'année à l'exception de 22 jours durant lesquels le tarif est dissuasif. EDF utilise ce tarif pour diminuer la consommation de pointe lors de jours de forte consommation (jours froids en général), afin de ne pas utiliser des moyens de production trop coûteux (cela revient à abaisser le coût marginal de production).

La figure 1.8 illustre bien l'effacement : le jour où la température est plus basse, à une consommation moindre car un signal EJP a été envoyé, "effaçant" une partie de la consommation.

complètement couvert). Techniquement, il s'agit d'une variable discrète. Néanmoins la représentation nationale, avec la pondération de 26 stations, nous permet de considérer la nébulosité comme une variable continue.

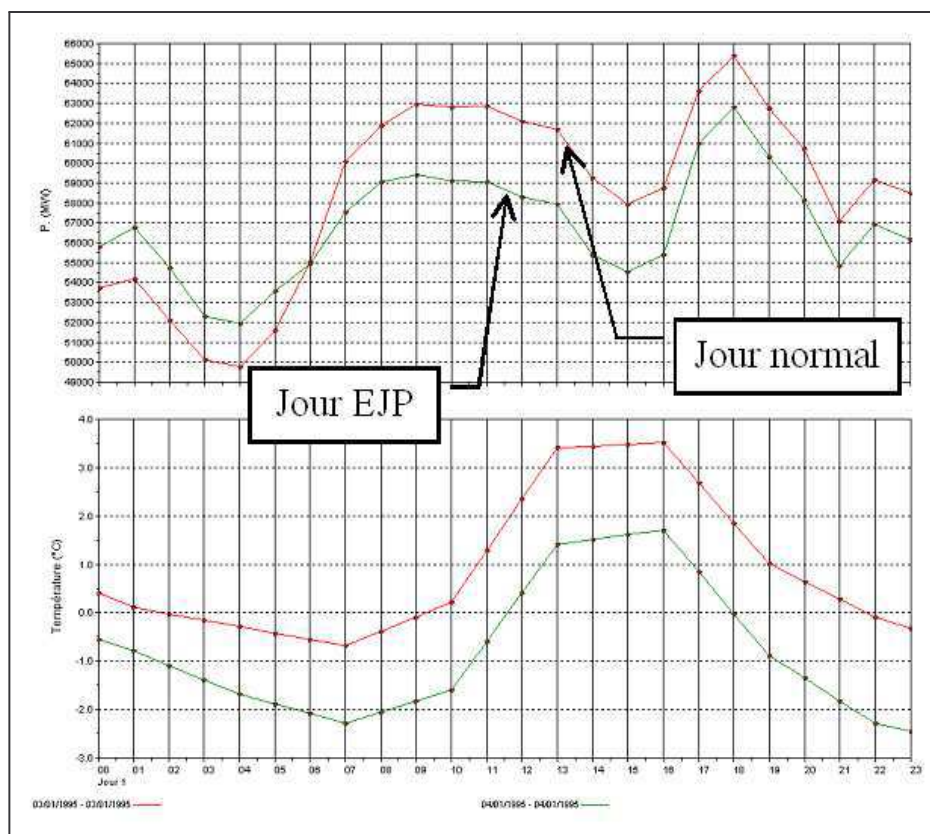


FIG. 1.8 – Impact du tarif EJP sur la consommation d'électricité

Les conclusions de toutes ces analyses sont prises en compte dans l'écriture du modèle retenu, et permettent d'intuire la fonction de dépendance f .

1.2.3 Principe

La solution actuellement en place à RTE s'explique historiquement. Comme indiqué précédemment, les séries temporelles ont connu un fort essor, notamment grâce à Box et Jenkins qui ont développé leurs modèles SARIMA. Leur ouvrage a constitué une référence théorique, mais également un guide pratique assez riche, et ce particulièrement dans le domaine électrique. On peut notamment se référer aux ouvrages de Box *et al.* (1976), ainsi que de Pankratz (1983).

Choisir SARIMA pour effectuer les prévisions était donc a priori un choix techniquement pertinent et quasiment naturel. EDF a ainsi fait le choix d'utiliser un modèle SARIMA dans les années 70. On peut se référer aux travaux de Meslier (1976) et Abadie *et al.* (1979). Malheureusement, ces modèles n'intègrent pas de variables exogènes² et il a donc fallu trouver un biais en leur adjoignant un modèle de régression (liant la puissance consommée avec les variables climatiques et les signaux de prix). On peut également citer des améliorations ou alternatives comme les filtres de Kalman, étudiés notamment par Martin (1999).

Au final, la prévision journalière s'effectue en trois étapes (Cf. Fig. 1.9) :

- On corrige la courbe de charge de l'impact du climat et des prix, afin de se ramener à une série temporelle ne dépendant plus des variables explicatives. On utilise à cet effet un modèle de régression, qu'on surnomme communément dans notre cas modèle de "correction".
- On utilise un modèle SARIMA sur la série ainsi "corrigée" ; la seule dynamique de la série permet de la prévoir.
- On réutilise le modèle de correction afin d'ajouter à la courbe "corrigée" prévue l'impact futur du climat et des prix.

²Les modèles ARIMAX sont trop difficiles à caler dans notre cas.

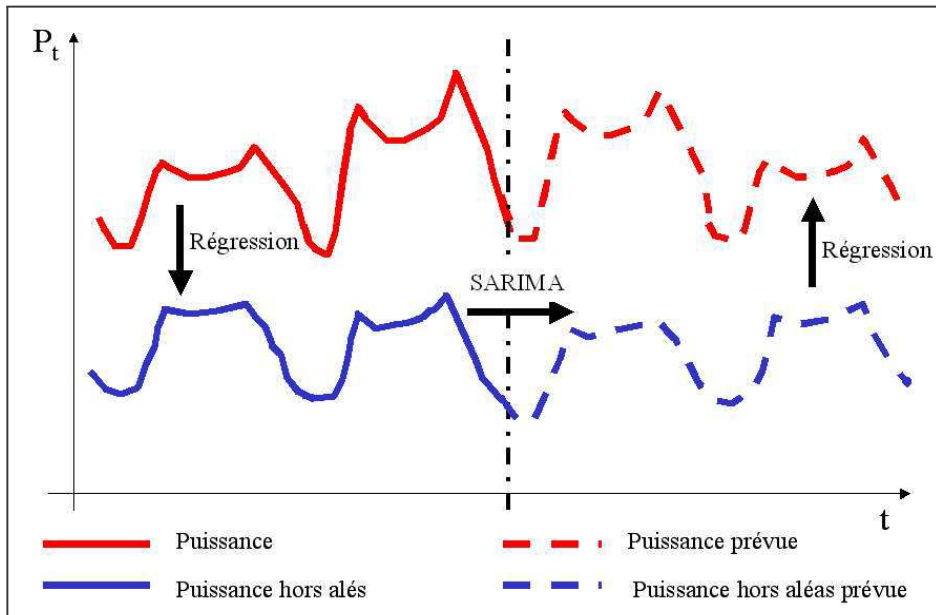


FIG. 1.9 – Principe actuel de la prévision de la consommation d'électricité

1.2.4 Modèles utilisés

Modèle de régression

Le modèle de régression retenu par RTE s'écrit :

$$\begin{aligned}
 P_{h,j,n} = & \left(\sum_{k=1}^7 \Pi_{h,k,n} \cdot \mathbb{I}_{\{\text{jour } j \text{ de type } k\}} \right) \cdot f_h(j) \\
 & + \sum_{k'=1}^5 P_{h,k',n}^{eff} \cdot \mathbb{I}_{\{\text{jour } j \text{ EJP de type } k'\}} \\
 & + g_{h,n}^{\text{hiver}} \cdot F_{\text{hiver}}(TR_{h,j,n}^{\text{hiver}}, T_h^{\text{seuil hiver}}) \\
 & + g_{h,n}^{\text{été}} \cdot F_{\text{été}}(TR_{h,j,n}^{\text{été}}, T_h^{\text{seuil été}}) \\
 & + \varepsilon_{h,j,n}
 \end{aligned} \tag{1.1}$$

Les notations utilisées sont :

- h désigne l'heure, j le jour et n l'année à cheval (du 1^{er} septembre au 31 août suivant).
- k désigne le type de jour (lundi, ..., dimanche)
- k' désigne le type de jour EJP.

- $\Pi_{h,k,n}$: puissance hors aléas (climatiques et tarifaires).
- $P_{h,k',n}^{eff}$: puissance effacée par le tarif EJP.
- $f_h(j)$ désigne la fonction de saisonnalité (j désigne ici la position du jour dans l'année : de 1 à 365) :

$$f_h(j) = 1 + \sum_{m=1}^4 \left[a_h \cos\left(\frac{2\pi m j}{365}\right) + b_h \sin\left(\frac{2\pi m j}{365}\right) \right] \\ + S_h^{\text{octobre}} \cdot \mathbb{I}_{\{\text{jour } j \text{ avant changement d'heure d'octobre}\}} \\ + S_h^{\text{mars}} \cdot \mathbb{I}_{\{\text{jour } j \text{ après changement d'heure de mars}\}}$$

- $g_{h,n}^{\text{hiver}}$: gradient de températures (hiver) ; $g_{h,n}^{\text{été}}$: gradient de températures (été).
- $F_{\text{été}}$ et F_{hiver} : fonctions mesurant les écarts entre les températures ressenties et les températures seuil ; ce sont des fonctions lissées. Par exemple, F_{hiver} est proche de :

$$\tilde{F}_{\text{hiver}}(T, T^{\text{seuil}}) = \max(T - T^{\text{seuil}}, 0)$$

- $T_h^{\text{seuil hiver}}$ et $T_h^{\text{seuil été}}$ sont les températures seuil de démarrage, respectivement du chauffage et de la climatisation.
- $TR_{h,j,n}^{\text{hiver}}$ et $TR_{h,j,n}^{\text{été}}$ sont les températures ressenties modélisées. Par exemple :

$$TR_{h,j,n}^{\text{hiver}} = \frac{1}{1 + \alpha_h + \beta_h} (T_{h,j,n} + \alpha_h T L_{h,j,n}^1 + \beta_h T L_{h,j,n}^2) + \mu_h (8 - N_{h,j,n})$$

où $T_{h,j,n}$ est la température, $N_{h,j,n}$ la nébulosité, $T L_{h,j,n}^1$ et $T L_{h,j,n}^2$ deux températures lissées (à l'aide d'un lissage exponentiel simple) traduisant les inerties de température.

α_h , β_h et μ_h sont les parts respectives des différentes composantes de la température ressentie.

RTE utilise la terminologie suivante :

- Puissance hors aléas :

$$PHA_{h,j,n} = \left(\sum_{k=1}^7 \Pi_{h,k,n} \cdot \mathbb{I}_{\{\text{jour } j \text{ de type } k\}} \right) \cdot f_h(j)$$

- Puissance effacée :

$$PEF_{h,j,n} = \sum_{k'=1}^5 P_{h,k',n}^{eff} \cdot \mathbb{I}_{\{\text{jour } j \text{ EJP de type } k'\}}$$

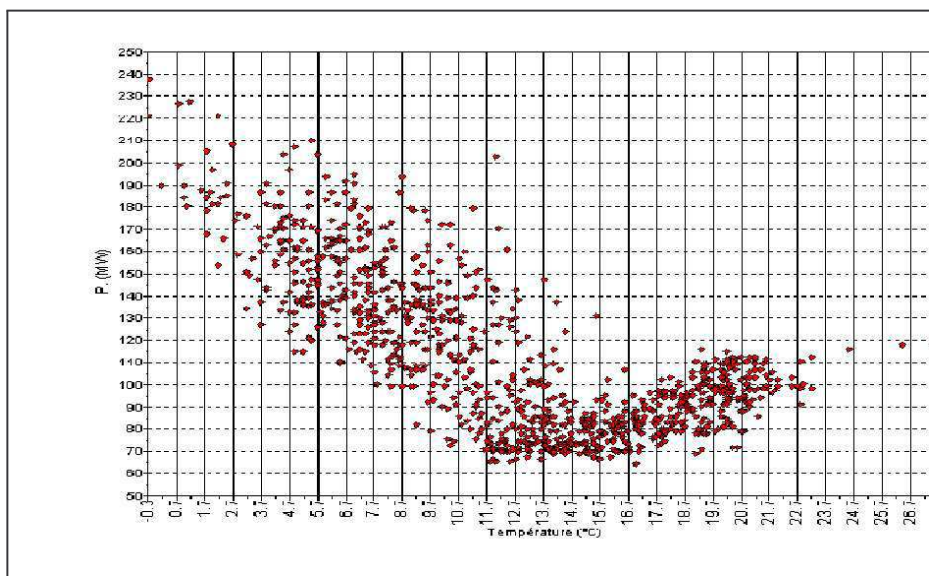


FIG. 1.10 – Dépendance horaire de la consommation d'électricité et de la température

- Puissance sensible à l'aléa climatique hivernal :

$$CH_{h,j,n} = g_{h,n}^{\text{hiver}} \cdot F_{\text{hiver}}(TR_{h,j,n}^{\text{hiver}}, T_h^{\text{seuil hiver}})$$

- Puissance sensible à l'aléa climatique estival :

$$CL_{h,j,n} = g_{h,n}^{\text{été}} \cdot F_{\text{été}}(TR_{h,j,n}^{\text{été}}, T_h^{\text{seuil été}})$$

On peut mettre graphiquement en évidence la modélisation choisie : la figure 1.10 représente les puissances en fonction des températures pour une heure donnée (sur une zone géographique mettant bien en évidence la climatisation).

A titre d'illustration, si on représente ce même nuage de points en fonction non plus de la température brute mais de la température lissée (à l'aide d'un lissage exponentiel simple), on constate que le nuage de point est moins large, rendant la modélisation plus facile (Cf. Fig. 1.11).

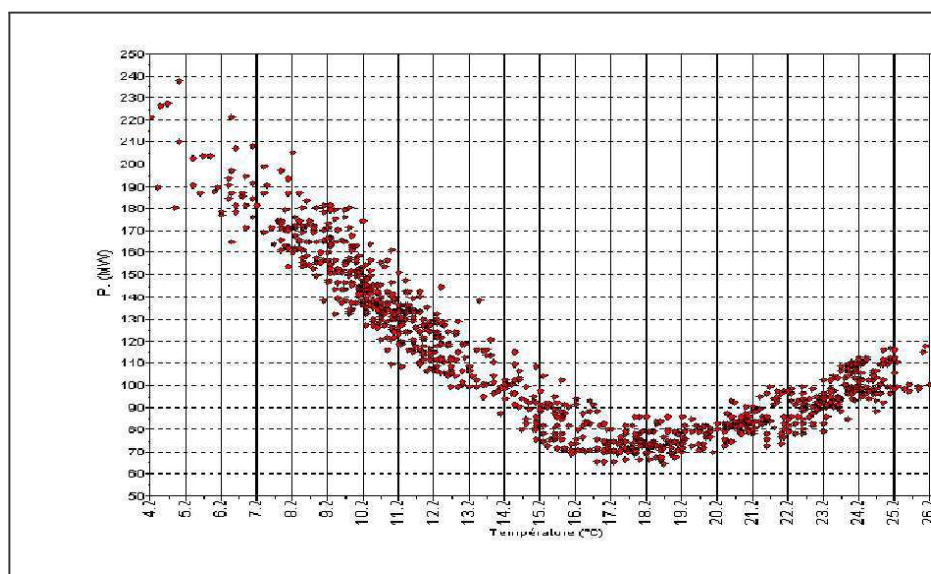


FIG. 1.11 – Dépendance horaire de la consommation d'électricité et de la température lissée

Les caractéristiques de ce modèle sont les suivantes (Cf. Fig 1.12) :

- La relation entre puissance et climat est quasiment linéaire par morceaux.
- Le volume sensible à l'aléa climatique se calcule en multipliant un gradient (estimé par heure et par année) par la différence entre une température ressentie et une température seuil.
- La température ressentie est une combinaison linéaire de la température instantanée, des retards de la température (pour prendre en compte les inerties) et de la nébulosité (pour prendre en compte l'effet de serre).

Cette décomposition est généralement opérée sur cinq années, à l'exception des jours fériés et des périodes atypiques (une année débutant le 1^{er} septembre et se terminant le 31 août, afin d'avoir des années homogènes d'un point de vue climatique). Les différents paramètres sont estimés de manière itérative par la méthode des moindres carrés et un algorithme d'optimisation (Hook et Jeeves).

Les paramètres de la régression sont connus jusqu'au 31 août dernier. Au-delà, nous ne possédons plus de connaissance du lien existant entre la consommation d'électricité et les variables météorologiques et de prix. L'extrapolation des paramètres sur une année permet de déterminer les paramètres du modèle de régression pour l'année en cours.

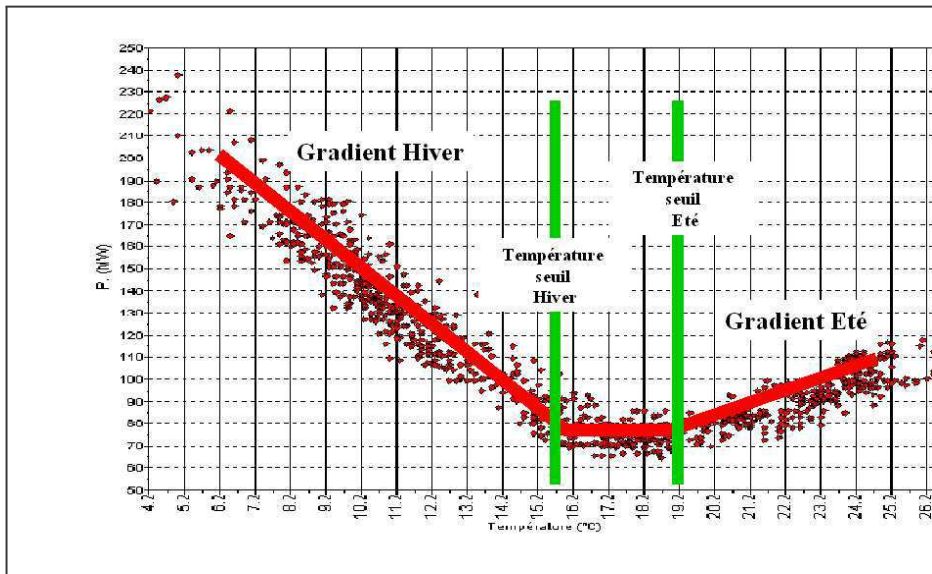


FIG. 1.12 – Principe de la régression liant la consommation d'électricité à la température

Enfin la représentation climatique de la France peut être effectuée au niveau régional³ ou national, en pondérant des stations météorologiques représentatives. La température et la nébulosité qui sont utilisées par la suite sont obtenues en moyennant 26 stations pour lesquelles on dispose de réalisations et de prévisions fournies par Météo-France.

Le modèle SARIMA

Le modèle SARIMA identifié et utilisé à RTE s'écrit de la manière suivante :

$$(I - \varphi B) (I - B^{24}) (I - B^{168}) X_t = (I + \theta_{24} B^{24}) (I + \theta_{168} B^{168}) \varepsilon_t$$

Les paramètres φ , θ_{24} et θ_{168} ont été estimés dans le passé à l'aide des procédures définies par Box et Jenkins. Ces paramètres sont cependant fixés dans l'outil de prévision, ce qui est contraire à l'idée d'estimation dynamique des paramètres (en théorie, on devrait réestimer les paramètres à chaque lancement de prévision). Néanmoins, il a été constaté que les paramètres estimés par le SARIMA étaient stables et les considérer comme constants permet de s'affranchir du problème des données extrêmes qui peut perturber l'estimation, rendant ainsi la prévision plus robuste.

³La France est décomposée en 7 régions électriques, indépendantes des régions administratives.

1.2.5 Récapitulatif

La prévision à l'instant T s'effectue au final en trois étapes :

1. On utilise les paramètres extrapolés de la partie dépendant des aléas climatiques et tarifaires, et les données météorologiques passées afin de déterminer la part dépendante de ces aléas sur le passé ; on soustrait cette part à la courbe de charge passée afin de mettre en évidence la part indépendante des aléas. Cette puissance corrigée vaut :

$$\forall t \leq T : PC_t = P_t - \widehat{PEF}_t - \widehat{CH}_t - \widehat{CL}_t$$

2. On prévoit cette demande d'électricité corrigée pour le lendemain avec le modèle SARIMA. On obtient ainsi \widehat{PC}_t pour $t > T$.
3. On utilise les paramètres extrapolés de la partie dépendante de l'aléa climatique et les données météorologiques prévues (par Météo-France) afin de déterminer la part dépendant des aléas dans le futur. On additionne cette part à la prévision de la demande corrigée des aléas, obtenue à la seconde étape afin d'obtenir une prévision globale de la puissance :

$$\forall t > T : \widehat{P}_t = \widehat{PC}_t + \widehat{PEF}_t + \widehat{CH}_t + \widehat{CL}_t$$

1.3 Perspectives

Si les modèles actuels fournissent à RTE des prévisions de bonne qualité, des travaux en cours explorent certaines voies d'amélioration :

- Incorporer d'autres variables climatiques telles que le vent et l'humidité, et d'autres variables de prix pourrait permettre une meilleure identification des variables explicatives. Il faut noter que cette voie est potentiellement prometteuse mais peut échouer, car on sait bien que même si un phénomène existe physiquement, il n'est pas toujours possible de le mettre en évidence statistiquement. En outre, comme on le réindiquera par la suite, les modèles actuels sont peu évolutifs à cause de leur complexité d'écriture.
- Certains types de jour posent parfois de gros problèmes en terme de prévision, notamment les jours fériés. En l'absence de méthode générale, les heuristiques actuellement adoptées doivent être améliorées. Ceci requiert un effort important de mise au point et de validation.

- Au-delà de ces améliorations, il faudra certainement envisager de prendre en compte les prix pour améliorer la qualité des prévisions. En effet, la demande sera de plus en plus sensible aux prix des marchés qui se sont mis en place en Europe. Pour l'instant seul le tarif EJP d'EDF est pris en compte (et significatif).

1.4 Problématique

Le modèle actuellement en place est complexe et il est très difficile de modifier sa structure. Si on veut tester par exemple l'incorporation d'une nouvelle variable explicative, il est peu probable d'y parvenir facilement, la juxtaposition du modèle de régression et du modèle SARIMA ne permettant pas d'isoler les causes de l'erreur finale de prévision. On connaît l'erreur d'ajustement globale du modèle de régression mais on n'a aucune connaissance de l'erreur de la partie qu'on utilise. Rappelons qu'on utilise seulement les termes \widehat{PEF}_t , \widehat{CH}_t et \widehat{CL}_t , pour lesquels on ne dispose pas de réalisations.

Des tentatives ont déjà été menées pour incorporer le vent et l'humidité et les résultats n'ont pas été probants.

De plus, une prévision même en temps qu'outil d'aide à la décision, devrait fournir un intervalle de confiance sur la prévision, ce qui est impossible pour l'instant ; les méthodes de bootstrap pourraient être envisagées mais les dépendances temporelles complexes et les multiples variables exogènes rendent la tâche complexe.

Le but principal des travaux réalisés consiste à obtenir un modèle unique qui soit contrôlable et pour lequel on dispose de résultats de convergence établis. Les intervalles de confiance, même s'ils ne sont pas l'objet principal de notre travail, devront pouvoir être élaborés de manière assez simple. C'est pourquoi on ne conserve du modèle actuel que les relations explicatives bien mises en évidence au cours de nombreuses études passées.

Le non-paramétrique est une alternative intéressante, en ce sens qu'il ne spécifie pas la fonction de lien. On peut a priori incorporer de nouvelles variables explicatives et simplement réestimer une fonction de lien de dimension plus élevée. Il est également possible d'élaborer des intervalles de confiance en se basant sur l'historique des erreurs de modélisation. L'inconvénient majeur de ces techniques réside dans le fléau de la dimension ; on sait en effet que des problèmes de convergence se posent si la dimension du modèle s'accroît (en pratique, ce problème de convergence va se traduire par des estimateurs de piètre qualité).

Dans notre cas, on dispose de multiples variables explicatives dont le nombre est difficilement réductible. Si on désirait par exemple enlever le passé des températures et ne considérer que la température instantanée, on sait fort bien que notre modèle s'adapterait moins bien, et cela se ressentirait immédiatement sur la prévision. La nécessité d'avoir des erreurs de prévision faibles pour gérer correctement le réseau électrique nous incite malheureusement à conserver de nombreuses variables explicatives, rendant caduques les méthodes non-paramétriques globales.

Afin de conserver l'idée d'une fonction de lien non spécifiée par un modèle paramétrique, on est conduit à envisager des techniques alternatives comme le semi-paramétrique. Si nous parvenons à réduire la dimension des variables explicatives, on pourrait alors envisager d'utiliser les méthodes non-paramétriques.

Nous proposons d'étudier des techniques de réduction de la dimension appliquées à la consommation d'électricité dans les deux chapitres suivants.

Le second chapitre nous permettra notamment de remplacer le modèle SARIMA présenté ici par un modèle semi-paramétrique purement autorégressif.

Le troisième chapitre propose un modèle semi-paramétrique incluant des variables exogènes. On l'appliquera à la consommation d'électricité en considérant les variables météorologiques (températures et nébulosités) comme variables exogènes.

Chapitre 2

Prévisions autorégressives non et semi-paramétriques

On ne s'intéresse dans ce chapitre qu'à des séries autorégressives. Cela peut paraître paradoxal puisqu'on cherche à modéliser la consommation d'électricité avec des variables exogènes. Néanmoins, on cherche à mettre en oeuvre des techniques semi-paramétriques très peu utilisées dans l'exercice de la prévision, et il nous paraît légitime de les tester dans des cas plus simples. Après avoir rappelé l'existence de travaux non-paramétriques sur des séries temporelles, on présente un modèle de régression semi-paramétrique utilisable sur des données dépendantes temporellement. On compare empiriquement les techniques non et semi-paramétriques sur des séries temporelles classiques, et après avoir vérifié l'efficacité de cette nouvelle méthode, on l'applique à la consommation d'électricité "corrigée" des effets des variables exogènes.

2.1 Le paradigme non-paramétrique

Le traitement opérationnel des séries temporelles est jusqu'à présent dominé par les techniques paramétriques, même si des développements théoriques alternatifs, ainsi que des mises en oeuvre, existent.

Considérons un modèle autorégressif classique :

$$X_t = \Phi(X_{t-1}, \dots, X_{t-r}) + \varepsilon_t$$

sans poser pour le moment d'hypothèses particulières sur le résidu.

Les méthodes paramétriques font l'hypothèse que la fonction de lien Φ appartient à une famille paramétrée de fonctions : $\Phi \in \{\Phi_\theta, \theta \in \Theta\}$. Estimer le lien Φ_θ revient donc à estimer le paramètre θ .

Par exemple, les modèles AR d'ordre r précisent que :

$$\Phi(X_{t-1}, \dots, X_{t-r}) = \varphi_1 X_{t-1} + \dots + \varphi_r X_{t-r}$$

On détermine donc la fonction de lien temporel en estimant $(\varphi_1, \dots, \varphi_r)$.

Il est clair que des problèmes de spécification peuvent se poser dans ces modèles, les rendant souvent caducs. Les modèles AR supposent par exemple que le processus à un instant t dépend linéairement de son passé (avec une mémoire donnée par l'ordre du processus AR).

Notons que la prévision en place à RTE utilise deux modèles purement paramétriques :

- Le modèle de régression : il s'agit d'un modèle paramétrique non linéaire. Il n'existe pas de résultats théoriques sur la significativité des différents paramètres. Les méthodes bootstrap pourraient néanmoins apporter quelques réponses sur la variabilité des paramètres estimés.
- Le modèle SARIMA pour établir la fonction de lien temporel.

Afin d'éviter de fixer une structure a priori de la fonction de régression, il est possible de faire appel aux méthodes non-paramétriques. Il existe maintenant de nombreux développements théoriques sur ces méthodes dont on peut citer quelques représentants :

- Les méthodes "sieves" ("tamis") telles que la décomposition en fonctions orthogonales, les séries de Fourier, les ondelettes...
- Les méthodes locales comme la méthode des noyaux, les polynômes locaux, les fonctions splines...

Dans le domaine électrique, des prévisions ont déjà été effectuées à l'aide des méthodes non-paramétriques. On peut notamment citer Harvey *et al.* (1993) qui ont utilisé des fonctions splines variant dans le temps afin de modéliser la consommation d'électricité ainsi que la température en Angleterre. Remarquons cependant que les erreurs obtenues peuvent être de l'ordre de 10% au niveau de la pointe de consommation, ce qui n'est pas acceptable en pratique. Afin d'être juste, il faut noter que la température est elle-même prévue à l'aide des splines, alors qu'aucun météorologue n'utilise de modèle stochastique pour fournir des prévisions pour le lendemain¹.

¹Les prévisions sont déterministes (issues d'équations de thermodynamique) jusqu'à 5 jours et sont stochastiques pour un horizon de 10 jours.

En France, des travaux ont également été menés sur les méthodes non-paramétriques. Ainsi Poggi (1994) s'est basé sur les travaux de Carbon *et al.* (1993) pour prévoir ce que nous avons appelé dans la première partie, la puissance corrigée des aléas PC_t ; il a remplacé le modèle SARIMA par un modèle autorégressif non-paramétrique. Notons qu'il existe d'autres alternatives non-paramétriques basées sur la méthode des noyaux, notamment les méthodes du mode et du mode conditionnel : on pourra notamment se référer à Matzner *et al.* (1998).

2.2 Prédiction autorégressive à l'aide de la méthode des noyaux

On présente une méthode de prédiction non-paramétrique à l'aide de la méthode des noyaux qu'on utilise dans la suite à des fins de comparaisons mais aussi dans les méthodes semi-paramétriques.

2.2.1 Modèle

On peut utiliser un modèle autorégressif non-paramétrique classique, et utiliser la méthode des noyaux. Carbon *et al.* (1993) ont notamment étudié et appliqué ce modèle.

Supposons qu'on dispose d'une série temporelle modélisable par un processus $(X_t)_{t \in \mathbb{N}}$ du second ordre, stationnaire et ergodique.

Si on désire prévoir X_{T+l} avec $l \in \mathbb{N}^*$ à partir de (X_1, \dots, X_T) , on peut utiliser le prédicteur suivant :

$$\widehat{X}_T(l) = \mathbb{E}(X_{T+l} / X_T, \dots, X_1)$$

On suppose ici qu'on dispose d'un processus r -markovien, ce qui est une hypothèse raisonnable pour les séries temporelles classiques :

$$\begin{aligned} \widehat{X}_T(l) &= \mathbb{E}(X_{T+l} / X_T, \dots, X_1) \\ &= \mathbb{E}(X_{T+l} / X_T, \dots, X_{T-r+1}) \\ &= \mathbb{E}\left(X_{T+l} / X_T^{(r)}\right) \end{aligned}$$

où $X_T^{(r)} = (X_T, \dots, X_{T-r+1})^T$ ².

²On désigne par x^T la transposée de x dans tout le document.

Cette hypothèse est étayée par des résultats sur les martingales. On sait que pour r et T "grands", les quantités $\mathbb{E}(X_{T+l}/X_T, \dots, X_1)$ et $\mathbb{E}(X_{T+l}/X_T, \dots, X_{T-r+1})$ sont proches de $\mathbb{E}(X_{T+l}/\mathcal{F}_{-\infty}^T(X))$, où $\mathcal{F}_{-\infty}^T(X)$ est la tribu engendrée par $\{X_t\}_{t \leq T}$:

$$\mathcal{F}_{-\infty}^T(X) = \sigma(X_t, t \leq T)$$

On peut donc estimer le modèle suivant :

$$X_{T+l} = \Phi_l(X_T, \dots, X_{T-r+1}) + \varepsilon_t$$

où :

$$\forall x \in \mathbb{R}^r : \Phi_l(x) = \mathbb{E}(X_{T+l} / X_T^{(r)} = x)$$

On obtient :

$$\forall x \in \mathbb{R}^r : \hat{\Phi}_l(x) = \frac{\sum_{t=r}^{T-l} K_h(X_t^{(r)} - x)}{\sum_{t=r}^{T-l} K_h(X_t^{(r)} - x)} \cdot X_{t+l}$$

On utilise dans tout le document la notation $K_h(\cdot) = \frac{1}{h^r} K\left(\frac{\cdot}{h}\right)$ où $K : \mathbb{R}^r \rightarrow \mathbb{R}$ est un "noyau" que nous supposons vérifier :

- K est une fonction de densité symétrique bornée
- $\lim_{\|x\| \rightarrow +\infty} \|x\|^r K(x) = 0$ ($x \in \mathbb{R}^r$) où $\|x\|$ désigne³ la norme euclidienne sur \mathbb{R}^r
- $\int_{\mathbb{R}^r} \|x\|^2 K(x) dx < +\infty$

Si on utilise cette estimation de la fonction Φ_l , on obtient une expression simple pour la prévision de X_{T+l} :

$$\begin{aligned} \hat{X}_T(l) &= \hat{\Phi}_l(X_T^{(r)}) \\ &= \sum_{t=r}^{T-l} \alpha_{t,T}^{r,l} X_{t+l} \end{aligned}$$

où :

$$\alpha_{t,T}^{r,l} = \frac{K_h(X_t^{(r)} - X_T^{(r)})}{\sum_{t=r}^{T-l} K_h(X_t^{(r)} - X_T^{(r)})}$$

³Sauf mention contraire, on conservera cette notation par la suite.

2.2.2 Convergence des méthodes à noyau

Rhomari (1994) et Bosq (1996) ont étudié les propriétés théoriques des prédicteurs à noyau. Ils ont établi des résultats de convergence pour $\widehat{\Phi}_l$ ainsi que pour l'erreur de prévision quadratique. On peut notamment citer le théorème suivant sur l'erreur de prévision quadratique.

Théorème 2.1 *Bosq (1996)*

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus strictement stationnaire, géométriquement mélangeant, borné et markovien d'ordre r vérifiant :

$$X_{t+l} = \Phi_l(X_t, \dots, X_{t-r+1}) + \varepsilon_t$$

Soit $(Z_t)_{t \in \mathbb{Z}}$ le processus associé défini par $Z_t = (X_{t+l}, X_t^{(r)})$.

Soit $\widehat{\Phi}_l$ l'estimateur de Φ_l à partir de l'échantillon $(Z_t)_{t \in \{r, \dots, T-l\}}$.

Soient les fonctions suivantes :

$$\begin{aligned} \forall x \in \mathbb{R}^r : f_l(x) &= \int_{\mathbb{R}} f_{(X_{t+l}, X_t^{(r)})}(u, x) du \\ \varphi_l(x) &= \int_{\mathbb{R}} u f_{(X_{t+l}, X_t^{(r)})}(u, x) du \\ \Phi_l(x) &= \frac{\varphi_l(x)}{f_l(x)} \end{aligned}$$

Supposons que f_l et φ_l sont continûment différentiables sur $S = \overline{\{x \in \mathbb{R}^r / f_l(x) > 0\}}$ (\overline{E} désigne l'adhérence de E).

Pour un noyau K lipschitzien et pour une fenêtre $h \sim \left(\frac{(\log T)^{2-\varepsilon}}{T}\right)^{\frac{1}{r+2}}$ où $\varepsilon \in]0, 2[$, on a :

$$\mathbb{E} \left[\left(\widehat{\Phi}_l(X_{T+l}) - \Phi_l(X_{T+l}) \right)^2 \right] = O \left(\frac{(\log T)^{2\varepsilon + (2-\varepsilon)\frac{2}{r+2}}}{T^{\frac{2}{r+2}}} \right)$$

Notons qu'il existe également un théorème dans le cas où le processus n'est pas borné.

2.2.3 Mise en œuvre

Il n'y a pas à proprement parler de paramètres à estimer mais certains choix doivent être effectués comme le noyau, la fenêtre et la mémoire.

Choix du noyau

En pratique, on choisit très souvent un noyau produit, c'est-à-dire tel que :

$$\forall (x_1, \dots, x_r) \in \mathbb{R}^r : K(x_1, \dots, x_r) = \prod_{i=1}^r K_0(x_i)$$

où K_0 est un noyau unidimensionnel.

Même si des travaux ont été menés pour optimiser le choix du noyau K , notamment par Epanechnikov (1969), l'impact du noyau reste minime dans les résultats de convergence. Ainsi, les noyaux produits uniforme, gaussien et d'Epanechnikov sont quasiment équivalents.

Choix de la fenêtre

Le choix de la fenêtre h (qui dépend de T) est majeur. En pratique, on a choisi de mettre en oeuvre un critère de *validation croisée*.

On minimise à cet effet le critère suivant :

$$CV_l(h) = \frac{1}{T-l-r+1} \sum_{t=r}^{T-l} \left[X_{t+l} - \widehat{\Phi}_{l,-t} \left(X_t^{(r)} \right) \right]^2$$

où $\widehat{\Phi}_{l,-t}$ est l'estimateur de Φ_l calculé à l'aide de toutes les données sauf la t -ième. On peut décider d'estimer une fenêtre pour tous les horizons de prévision considérés, ou distinguer la fenêtre en fonction de ce même horizon. On choisit cette seconde option dans la suite.

Choix de la mémoire

Le choix de la mémoire peut être effectué de manière analogue à la validation croisée pour h . On peut par exemple minimiser la quantité suivante :

$$M(r) = \frac{1}{T-l-r+1} \sum_{t=r}^{T-l} \left[X_{t+l} - \widehat{\Phi}_l \left(X_t^{(r)} \right) \right]^2$$

Il est à noter que le calcul de $M(r)$ fait intervenir la fenêtre qu'il faut optimiser par le critère de validation croisée énoncé auparavant.

Afin de limiter les calculs, on peut présélectionner certaines tailles de mémoire si on a une connaissance suffisante de la série temporelles à traiter. En pratique, la mémoire pour la consommation d'électricité peut être choisie comme un multiple de 24 (ou 12) si on travaille sur une série au pas horaire.

Qualité de la prédiction

Afin de mesurer la qualité de la prédiction, on effectue une analyse a posteriori : on tronque la série (de l_{\max} observations) et on compare les prévisions obtenues avec les réalisations.

On utilise classiquement les deux indicateurs de qualité suivants :

– L'erreur moyenne observée :

$$MAPE = \frac{1}{l_{\max}} \sum_{i=T-l_{\max}+1}^T \frac{|X_j - \hat{X}_j|}{|X_j|}$$

– L'erreur quadratique moyenne observée :

$$RMSE = \sqrt{\frac{1}{l_{\max}} \sum_{i=T-l_{\max}+1}^T (X_j - \hat{X}_j)^2}$$

On peut remarquer que le MAPE est très souvent utilisé en pratique, alors que la minimisation de notre problème est effectuée à l'aide d'un critère quadratique. Il pourrait donc être plus légitime d'utiliser un critère cohérent tel que le RMSE, mais les praticiens restent très attachés aux erreurs relatives.

Horizon de la prédiction

En pratique, la prédiction peut s'effectuer de deux manières différentes. Si on désire prévoir X_{T+l} pour $l \in \{1, \dots, l_{\max}\}$ à partir de (X_1, \dots, X_T) , on peut :

- Effectuer l_{\max} prévisions "indépendantes" : on calcule $\hat{X}_T(l)$ pour $l \in \{1, \dots, l_{\max}\}$. On estime donc $\Phi_1, \dots, \Phi_{l_{\max}}$.
- Effectuer l_{\max} prévisions de manière itérative : on calcule $\hat{X}_T(1)$, puis $\hat{X}_{T+1}(1)$ en remplaçant X_{T+1} par $\hat{X}_T(1)$, ... On n'estime alors que Φ_1 .

Il n'existe pas de résultat théorique sur la qualité comparée de ces deux procédés. Il ne semble pas nécessaire d'effectuer des tests importants pour déterminer le meilleur choix ; seule une mise en œuvre réellement opérationnelle justifierait cette comparaison.

Stationnarisation des données

Théoriquement, on applique cette méthode sur des processus stationnaires. Si les séries temporelles présentent des saisonnalités et/ou une tendance, on devrait alors

les stationnariser, en utilisant par exemple une décomposition saisonnière ou des différenciations adéquates. En pratique, on constate qu'il est parfois judicieux de conserver les saisonnalités afin de comparer des motifs $X_t^{(r)}$ qui incluent ces saisonnalités, notamment dans le cas de la consommation d'électricité.

2.3 Le paradigme semi-paramétrique

Bien qu'étant plus satisfaisant intellectuellement que les méthodes paramétriques, le non-paramétrique soulève des problèmes :

- La fonction de régression est ininterprétable (la fonction de régression ne se résume pas à l'étude de quelques paramètres dont on peut étudier les variations).
- Contrairement au choix du noyau K , le choix de la fenêtre h peut s'avérer difficile.
- Enfin, le problème majeur est la dimension de l'espace des variables explicatives (communément appelé le fléau de la dimension). Plus il y a de variables explicatives à prendre en compte, moins on est assuré de la convergence des estimateurs. Ce problème théorique de convergence se traduit en pratique par de grosses erreurs d'ajustement.

Afin de pallier certains errements du paramétrique et du non-paramétrique, des modèles de régression semi-paramétriques ont été proposés. Il en existe deux principaux types, dont on donne ici quelques représentants.

Approximation fonctionnelle

- Les *modèles additifs généralisés*, étudiés entre autres par Auestad *et al.* (1990), Härdle *et al.* (1996).

En notant $Y \in \mathbb{R}$ la variable à expliquer et $X = (X^1, \dots, X^p)^T \in \mathbb{R}^p$ les p variables explicatives, on suppose que :

$$\mathbb{E}(Y / X) = \sum_{j=1}^p g_j(X^j)$$

où $g_j : \mathbb{R} \rightarrow \mathbb{R}$ sont p fonctions mesurables inconnues.

- Les *modèles autorégressifs à coefficients fonctionnels*, étudiés entre autres par Chen *et al.* (1993).

On suppose que :

$$X_t = f_1(X_{t-1}^*) X_{t-1} + \dots + f_r(X_{t-1}^*) X_{t-r} + \varepsilon_t$$

où $f_i : \mathbb{R} \rightarrow \mathbb{R}$ sont r fonctions mesurables inconnues et $X_{t-1}^* = (X_{t-i_1}, \dots, X_{t-i_k})^T$ (avec $i_j \in \mathbb{N}^*$ pour $j \in \{1, \dots, k\}$).

Réduction de la dimension

- Les modèles de "Projection Pursuit Regression" (PPR), étudiés entre autres par Huber (1985), Hall (1989), Chen (1991), Xia *et al.* (1997).

On suppose que :

$$\mathbb{E}(Y / X) = \sum_{j=1}^D g_j(\beta_j^T X)$$

où $g_j : \mathbb{R} \rightarrow \mathbb{R}$ sont D fonctions mesurables inconnues et $\beta_j \in \mathbb{R}^p$ sont D paramètres (vectoriels) inconnus.

- Les modèles à direction révélatrice unique (SIM : Single Index Model), étudiés entre autre par Härdle *et al.* (1993), Ichimura (1993), Carroll *et al.* (1997), Delecroix *et al.* (1999), Xia *et al.* (1999), Delecroix *et al.* (2006).

On suppose que :

$$\mathbb{E}(Y / X) = g(\beta^T X)$$

où $g : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction mesurable inconnue et $\beta \in \mathbb{R}^p$ est un paramètre (vectoriel) inconnu.

- Les modèles à directions révélatrices multiples, étudiés par Li *et al.* (2001), Xia *et al.* (2002), Li *et al.* (2004).

On suppose que :

$$\mathbb{E}(Y / X) = g(B^T X)$$

où $g : \mathbb{R}^D \rightarrow \mathbb{R}$ est une fonction mesurable inconnue et $B = (\beta^1, \dots, \beta^D)_{p \times D}$ est un paramètre (matriciel) inconnu.

Ce sont les modèles à direction révélatrices qui occupent notre attention. Ils ont comme intérêt de spécifier au minimum le modèle; on utilise une fonction de lien non-paramétrique en ayant déterminé au préalable des combinaisons linéaires de variables explicatives qui contiennent le maximum d'information, palliant ainsi le fléau de la dimension.

Il existe différentes méthodes d'estimation pour ces méthodes :

- "Average Derivative Estimate" (ADE), étudié notamment par Härdle *et al.* (1989), Härdle *et al.* (1992), Horowitz (1998), Horowitz *et al.* (1996), Hristache *et al.* (2001).
- "Sliced Inverse Regression" (SIR), étudié notamment par Li (1991), Schott (1994), Zhu *et al.* (1996).
- "Moving Average (conditional) Variance Estimation" (MAVE), étudié par Xia *et al.* (2002).

La méthode ADE présente l'inconvénient majeur de ne s'appliquer qu'aux problèmes à direction révélatrice unique. Au vu de la complexité de notre série d'étude (Cf. Chapitre 1), il ne nous apparaît pas adéquat de ne considérer qu'une seule direction révélatrice.

La méthode SIR présente quant à elle l'inconvénient d'imposer une structure probabiliste incompatible avec la plupart des séries temporelles. La famille probabiliste la plus générale concernée par la méthode SIR est la famille des distributions symétriques elliptiques. Si on considère un processus $(X_t)_{t \in \mathbb{Z}}$ répondant à une telle loi, ce processus est alors réversible temporellement, c'est-à-dire vérifie :

$$\forall n \in \mathbb{N}^*, \forall (t_1, \dots, t_n) \in \mathbb{Z}^n : \mathcal{L}(X_{-t_1}, \dots, X_{-t_n}) = \mathcal{L}(X_{t_1}, \dots, X_{t_n})$$

Or il est évident que ce type d'hypothèse est incompatible avec la grande majorité des séries temporelles.

C'est pourquoi on se focalise sur l'étude de la méthode MAVE qui ne présente aucun de ces deux inconvénients, même si sa mise en œuvre est plus complexe.

2.4 La méthode MAVE

2.4.1 Principe

La méthode MAVE est une méthode d'estimation d'un modèle de régression à directions révélatrices multiples. Elle a été introduite par Xia *et al.* (2002).

On considère le modèle de régression suivant :

$$Y = g(B^T X) + \varepsilon \quad (2.1)$$

où :

- $Y \in \mathbb{R}$ est la variable à expliquer
- $X \in \mathbb{R}^p$ sont p variables explicatives
- ε est une v.a.r. dans $\mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ admettant une fonction de densité.
- $\mathbb{E}(\varepsilon/X) = 0$ p.s.
- $g : \mathbb{R}^D \rightarrow \mathbb{R}$ est une fonction mesurable inconnue
- $B = (\beta^1, \dots, \beta^D)_{p \times D}$, avec $D \in \{1, \dots, p\}$, est une matrice inconnue telle que $B^T B = I_D$

Cette régression semi-paramétrique consiste à estimer, de manière non paramétrique, une fonction de lien g inconnue sur D directions révélatrices. On constate aisément que plus D est petit, plus on évite le fléau de la dimension, inconvénient typique des

méthodes non-paramétriques. Les D directions révélatrices sont les combinaisons linéaires de variables explicatives qui contiennent le maximum d'information pour la variable à expliquer Y .

L'hypothèse sous-jacente du modèle (2.1) est :

$$\mathbb{E}(Y / X = x) = \mathbb{E}(Y / B^T X = B^T x)$$

$B^T X$ contient toute l'information nécessaire à expliquer Y ; la projection linéaire de X (de dimension p) sur un espace de dimension D plus faible n'absorbe pas d'information. On appelle $\beta_1^T X, \dots, \beta_D^T X$ directions révélatrices.

On peut remarquer que si $D = 1$, on retrouve alors les modèles SIM.

Pour un nombre de directions révélatrices D donné, la direction B (2.1) est solution de :

$$\min_{B/B^T B=I_D} \mathbb{E} \left[(Y - \mathbb{E}(Y / B^T X))^2 \right] \quad (2.2)$$

où $B = (\beta_1, \dots, \beta_D)_{p \times D}$ avec $D \in \{1, \dots, p\}$.

La méthode MAVE repose sur la variance conditionnelle. On considère la variance de Y conditionnellement à $B^T X$:

$$\sigma_B^2(B^T X) = \mathbb{E} \left[(Y - \mathbb{E}(Y / B^T X))^2 / B^T X \right]$$

On en déduit que :

$$\mathbb{E} [\sigma_B^2(B^T X)] = \mathbb{E} \left[(Y - \mathbb{E}(Y / B^T X))^2 \right]$$

Le problème de minimisation (2.2) est donc équivalent à :

$$\min_{B/B^T B=I_D} \mathbb{E} [\sigma_B^2(B^T X)] \quad (2.3)$$

La résolution du modèle (2.1) se fait selon cette dernière formule et est appelée méthode MAVE : Minimum Average (conditional) Variance Estimation.

Soit la fonction suivante :

$$g_B(u_1, \dots, u_D) = \mathbb{E}(Y / \beta_1^T X = u_1, \dots, \beta_D^T X = u_D)$$

Soit $(Y_i, X_i)_{i \in \{1, \dots, n\}}$ un échantillon de (Y, X) .

En effectuant un développement limité de $\mathbb{E}(Y_i / B^T X_i)$ en X_l pour $l \in \{1, \dots, n\}$, on en déduit que :

$$\mathbb{E}(Y_i / B^T X_i) \simeq a + b^T B^T (X_i - X_l)$$

où

$$\begin{cases} a = g_B(B^T X_l) \\ b = (b_{(1)}, \dots, b_{(D)})^T \end{cases}$$

avec :

$$\forall k \in \{1, \dots, D\} : b_{(k)} = \left. \frac{\partial g_B(u_1, \dots, u_D)}{\partial u_k} \right|_{u_1 = \beta_1^T X_l, \dots, u_D = \beta_D^T X_l}$$

On estime alors $\sigma_B^2(B^T X_l)$ de manière analogue à la méthode des polynôme locaux, étudiés notamment par Fan *et al.* (1996). On obtient ainsi :

$$\hat{\sigma}_B^2(B^T X_l) = \min_{a,b} \sum_{i=1}^n \{Y_i - [a + b^T B^T (X_i - X_l)]\}^2 \omega_{il}$$

où $(\omega_{il})_{i \in \{1, \dots, n\}}$ sont des poids traduisant la proximité entre X_i et X_l , tels que :

- $\forall l \in \{1, \dots, n\} : \omega_{il} \in]0, 1[$
- $\sum_{i=1}^n \omega_{il} = 1$

On estime $\mathbb{E}[\sigma_B^2(B^T X)]$ par $\frac{1}{n} \sum_{j=1}^n \hat{\sigma}_B^2(B^T X_j)$.

On estime donc B , solution de (2.2), en résolvant le problème suivant :

$$\min_{B/B^T B = I_D, (a_j, b_j)_{j \in \{1, \dots, n\}}} \sum_{j=1}^n \sum_{i=1}^n \{Y_i - [a_j + b_j^T B^T (X_i - X_j)]\}^2 \omega_{ij} \quad (2.4)$$

2.4.2 Choix des poids

Il y a deux choix possibles pour déterminer les poids $(\omega_{ij})_{(i,j) \in \{1, \dots, n\}^2}$:

- Choix simplifié

La méthode des polynômes locaux impliquerait le choix suivant :

$$\omega_{ij} = \frac{K_h(B^T (X_i - X_j))}{\sum_{l=1}^n K_h(B^T (X_l - X_j))}$$

On constate que la présence des directions révélatrices accroît la complexité de l'estimation. C'est pourquoi on peut simplifier ces poids en considérant leur version non-paramétrique :

$$\omega_{ij} = \frac{K_h(X_i - X_j)}{\sum_{l=1}^n K_h(X_l - X_j)}$$

– Choix affiné

On procède alors de manière itérative

- On définit un premier jeu de poids afin d'initialiser l'estimation :

$$\omega_{ij} = \frac{K_h(X_i - X_j)}{\sum_{l=1}^n K_h(X_l - X_j)}$$

Pour une dimension D connue, il est alors possible d'obtenir une première estimation \hat{B} de B .

- On utilise ensuite un jeu de poids améliorés :

$$\tilde{\omega}_{ij} = \frac{K_h\left(\hat{B}^T(X_i - X_j)\right)}{\sum_{l=1}^n K_h\left(\hat{B}^T(X_l - X_j)\right)}$$

Une fois qu'on a estimé B par \hat{B} , il est possible de réestimer B en minimisant l'équation (2.4) avec les poids précédents.

Cette étape est réitérée de manière à converger. La méthode est appelée rMAVE ("refined MAVE").

2.4.3 Choix de la réduction de dimension

Le choix de la dimension D s'effectue par un critère de validation croisée noté CV :

$$\hat{D} = \arg \min_{d \in \{1, \dots, p\}} CV(d)$$

où :

$$- \forall d \in \{1, \dots, p\} : CV(d) = \frac{1}{n} \sum_{j=1}^n (Y_j - \hat{a}_{d,j})^2$$

$$- \forall d \in \{1, \dots, p\}, \forall j \in \{1, \dots, n\} : \hat{a}_{d,j} = \frac{\sum_{i=1, i \neq j}^n K_{h_d}^{(i,j)} Y_i}{\sum_{i=1, i \neq j}^n K_{h_d}^{(i,j)}}$$

$$- \forall d \in \{1, \dots, p\}, \forall (i, j) \in \{1, \dots, n\}^2 : K_{h_d}^{(i,j)} = K_{h_d}\left(\hat{\beta}_1^T(X_i - X_j), \dots, \hat{\beta}_d^T(X_i - X_j)\right)$$

En pratique, si p est trop grand, on se fixera un nombre maximum de directions à tester.

2.4.4 Résultats de convergence

Il existe deux théorèmes importants pour la méthode MAVE, l'un sur le choix de la dimension D par le critère de validation croisée, et la qualité de l'estimateur de B .

Soient les hypothèses suivantes :

HM1 $(Y_i, X_i)_{i \in \mathbb{N}}$ est strictement stationnaire et β -mélangeant :

$$\beta(k) = \sup_{i \in \mathbb{N}^*} \left(\mathbb{E} \left[\sup_{A \in \mathcal{F}_{i+k}^{+\infty}} |\mathbb{P}(A / \mathcal{F}_1^i) - \mathbb{P}(A)| \right] \right) \xrightarrow{k \rightarrow +\infty} 0$$

où $\mathcal{F}_i^j = \sigma((X_l, Y_l), l \in \{i, \dots, j\})$.

HM2 $\forall k \in \mathbb{N}^* : \mathbb{E}[|Y|^k] < +\infty$ et $\mathbb{E}[\|X\|^k] < +\infty$

HM3 f_X et f_Y admettent des dérivées bornées, et sont elle-même bornées sur un voisinage de 0.

HM4 Les densités conditionnelles $f_{X/Y}$ et $f_{(X_0, X_l)/(Y_0, Y_l)}$ pour $l \in \mathbb{N}^*$ sont bornées.

HM5 g admet des dérivées (jusqu'à l'ordre 3) continues et bornées.

HM6 K est une fonction de densité symétrique sphérique.

Théorème 2.2 *Xia et al. (2002)*

Sous les hypothèses (HM1-HM6), en supposant que f_X est à support compact, on a :

$$\mathbb{P}(\widehat{D} = D) \xrightarrow{n \rightarrow +\infty} 1$$

Théorème 2.3 *Xia et al. (2002)*

Sous les hypothèses (HM1-HM6), si $h \xrightarrow{n \rightarrow +\infty} 0$ et $\frac{nh^p}{\log(n)} \xrightarrow{n \rightarrow +\infty} 0$, on a :

- Si $d < D$:

$$\left\| (I - BB^T) \widehat{B} \right\| = O_{\mathbb{P}} \left(h^2 + \frac{\log(n)}{nh^{p+1}} \right)$$

- Si $d \geq D$:

$$\left\| (I - \widehat{B}\widehat{B}^T) B \right\| = O_{\mathbb{P}} \left(h^3 + \frac{\log(n)}{nh^{p+1}} \right)$$

Notons que $\left\| (I - BB^T) \widehat{B} \right\|$ et $\left\| (I - \widehat{B}\widehat{B}^T) B \right\|$ sont des mesures de la distance entre \widehat{B} et B .

d désigne le nombre de colonnes de la matrice \widehat{B} et D celui de B .

2.5 Prévision autorégressive à l'aide de la méthode MAVE

2.5.1 Principe

La méthode MAVE a été appliquée à des modèles de régression, sur des données simulées et des données réelles. Aucune des hypothèses de la méthode MAVE n'est contradictoire avec les séries temporelles classiques et il est donc possible d'utiliser cette méthode pour élaborer une prévision. On reprend ici la méthodologie proposée par Carbon *et al.* (1993).

Pour prévoir X_{T+l} avec $l \in \mathbb{N}^*$ à partir de (X_1, \dots, X_T) , on utilise le prédicteur suivant :

$$\widehat{X}_T(l) = \mathbb{E}(X_{T+l} / X_T, \dots, X_1)$$

On suppose encore ici qu'on dispose d'un processus r -markovien :

$$\begin{aligned} \widehat{X}_T(l) &= \mathbb{E}(X_{T+l} / X_T, \dots, X_1) \\ &= \mathbb{E}\left(X_{T+l} / X_T^{(r)}\right) \text{ où } X_T^{(r)} = (X_T, \dots, X_{T-r+1})^T \\ &= g_l\left(B_l^T X_T^{(r)}\right) \end{aligned}$$

où $B_l = (\beta_1, \dots, \beta_D)_{r \times D}$.

On peut noter que la fonction de lien g_l , ainsi que la matrice B_l , dépendent de l'horizon de prévision l .

Remarque 2.1

Considérons le modèle autorégressif semi-paramétrique vérifiant les hypothèses du modèle MAVE :

$$\forall t \in \mathbb{N}^* : X_t = g\left(B^T (X_{t-1}, \dots, X_{t-r})^T\right) + \varepsilon_t$$

Le quatrième chapitre (annexe technique) définit des conditions nécessaires et suffisantes sur la fonction de lien g pour que le processus $(X_t)_{t \in \mathbb{N}^*}$ soit stationnaire et mélangeant. Cette hypothèse est en effet requise pour établir la convergence des estimateurs à l'aide de la méthode MAVE. On pourra constater que les conditions requises sont assez naturelles et faciles à obtenir.

2.5.2 Mise en œuvre

On se propose d'utiliser ici partiellement la méthode MAVE pour prévoir une série temporelle "issue" d'un processus $(X_t)_{t \in \mathbb{N}}$.

Supposons la matrice B_l connue. On peut alors déterminer la fonction de lien g_l par la méthodes des noyaux. Cette méthode est intéressante pour un prévisionniste car elle permet de calculer des similarités (Cf. Poggi (1994)) qui mesurent la distance entre le dernier bloc connu de r données, et les blocs passés de même taille : la pondération peut ainsi être affichée clairement, donnant une indication au prévisionniste (opérationnel), pas toujours au fait des modèles statistiques.

On a :

$$\forall x \in \mathbb{R}^D : g_l(x) = \mathbb{E} \left(X_{T+l} \mid B_l^T X_T^{(r)} = x \right)$$

On obtient :

$$\begin{aligned} \widehat{X}_T(l) &= \widehat{g}_l \left(B_l^T X_T^{(r)} \right) \\ &= \sum_{t=r}^{T-l} \alpha_{t,T}^{r,l} X_{t+l} \end{aligned}$$

où :

$$\alpha_{t,T}^{r,l} = \frac{K_h \left[B_l^T \left(X_t^{(r)} - X_T^{(r)} \right) \right]}{\sum_{t=r}^{T-l} K_h \left[B_l^T \left(X_t^{(r)} - X_T^{(r)} \right) \right]}$$

Afin d'estimer la matrice B_l contenant les coefficients des directions révélatrices, on utilise la méthode MAVE.

On se ramène ainsi au problème d'optimisation suivant (pour D fixée) :

$$\widehat{B}_l = \arg \min_{B/B^T B = I_D, (a_j, b_j)_{j \in \{1, \dots, n\}} j=r+1 i=r+1} \sum_{j=r+1}^{T-l} \sum_{i=r+1}^{T-l} \left\{ X_{i+l} - \left[a_j + b_j^T B^T \left(X_i^{(r)} - X_j^{(r)} \right) \right] \right\}^2 \omega_{ij}$$

Choix du noyau

On choisit pour K le noyau produit gaussien :

$$K(x_1, \dots, x_D) = \frac{1}{(2\pi)^{\frac{D}{2}}} \exp \left(-\frac{1}{2} \sum_{i=1}^D x_i^2 \right)$$

Choix de la fenêtre

On doit en utiliser différentes pour lesquelles on utilise encore un critère de validation croisée :

- Pour l'estimation de B_l : on doit optimiser la fenêtre pour chacune des dimensions d testées pour trouver la dimension optimale \hat{D} .
- Pour l'estimation de g_l par la méthode du noyau.

Choix de la mémoire

On peut d'ores et déjà constater que l'utilisation de la méthode MAVE est complexe numériquement. C'est pourquoi, choisir la taille de la mémoire r optimale par validation croisée peut être téméraire. En pratique, on choisira sur les séries temporelles un nombre limité de valeurs de r potentielles, basées sur notre connaissance a priori des séries.

Horizon de prédiction

On a présenté comment obtenir la prédiction $\hat{X}_T(l)$ de X_{T+l} avec $l \in \mathbb{N}^*$ à partir de (X_1, \dots, X_T) .

En pratique, on souhaite obtenir $\hat{X}_T(l)$ pour $l \in \{1, \dots, l_{\max}\}$. On peut alors opérer de 2 façons différentes

- Effectuer l_{\max} prévisions "indépendantes" : on calcule $\hat{X}_T(l)$ pour $l \in \{1, \dots, l_{\max}\}$. On doit alors estimer l_{\max} matrices B_l . Pour chacune d'entre elles, on doit choisir par validation croisée la dimension \hat{D} optimale, ce qui est très coûteux numériquement. On peut réduire la difficulté en estimant qu'une seule matrice B pour tous les horizons tout en déterminant une fenêtre optimale pour chacun des horizons de prédiction
- Effectuer l_{\max} prévisions de manière itérative : on calcule $\hat{X}_T(1)$, puis $\hat{X}_{T+1}(1)$ en remplaçant X_{T+1} par $\hat{X}_T(1)$, ..., et enfin $\hat{X}_T(l_{\max})$. On peut alors décider d'estimer une seule matrice, celle correspondant à $l = 1$, qu'on réutilise à chaque itération. On n'a plus alors qu'une seule validation croisée à effectuer pour le choix de \hat{D} .

2.6 Application à des séries temporelles classiques

2.6.1 Démarche

On a repris quelques séries utilisés par Carbon *et al.* (1993) afin de comparer les résultats obtenus avec la méthodes des noyaux classiques et la méthode MAVE. Ils ont montré que les prévisions non-paramétriques à l'aide de la méthode des noyaux fournissent des resultats équivalents ou meilleurs que les modèles SARIMA. Nous avons adopté la même démarche sans relancer les prévisions par la méthode de Box et Jenkins. Notons qu'on a choisi des séries de taille suffisamment importante pour que la méthode MAVE puisse présenter un intérêt.

Le but initial de cette étude était de vérifier l'efficacité de la méthode MAVE sur des séries "simples" afin d'envisager de l'appliquer à notre problématique de la consommation d'électricité beaucoup plus délicate.

Pour chacune des séries considérées, on a tronqué les données de $l_{\max} = 6$ observations et prévu les données manquantes :

- à l'aide de la méthode des noyaux ;
- à l'aide de la méthode MAVE : on a estimé qu'une seule matrice B pour tous les horizons de prévision et une fonction de lien g_l différente pour chaque horizon l .

Comme indiqué auparavant, on utilise les RMSE et MAPE comme critères d'erreur. Les résultats suivants montrent que la méthode MAVE est équivalente ou supérieure à la méthode du noyau. Cette comparaison est également effectuée dans la suite du document sur la consommation d'électricité, hors aléas exogènes, et on pourra constater que la réduction de la dimension prend alors tout son sens sur une problématique avec de nombreuses variables explicatives.

2.6.2 Série AR(2)

On a simulé le processus AR(2) suivant :

$$X_t = 1000 + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \varepsilon_t$$

où :

- $t \in \{1, \dots, 100\}$
- $\varphi_1 = 0.7, \varphi_2 = 0.2$
- $\varepsilon_t \rightsquigarrow \mathcal{N}(0, 5^2)$

Méthode	r	D	RMSE	MAPE (%)
Noyau	5	-	24.45	0.185
MAVE	10	2	21.39	0.150

TAB. 2.1 – Prévisions optimales "MAVE" et "Noyau" : AR(2)

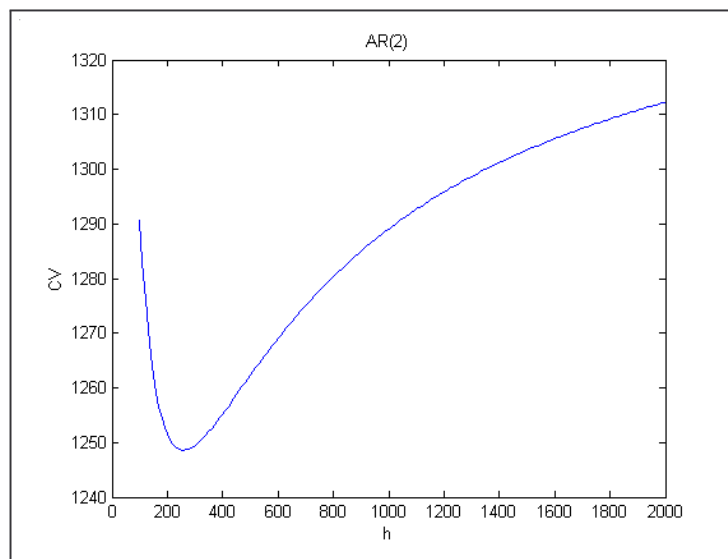


FIG. 2.1 – Critère de validation croisée MAVE : AR(2)

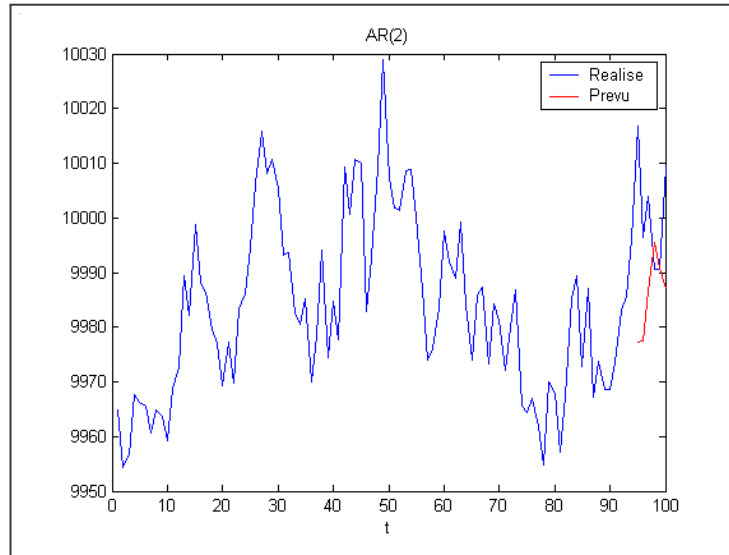


FIG. 2.2 – Prévion optimale MAVE : AR(2)

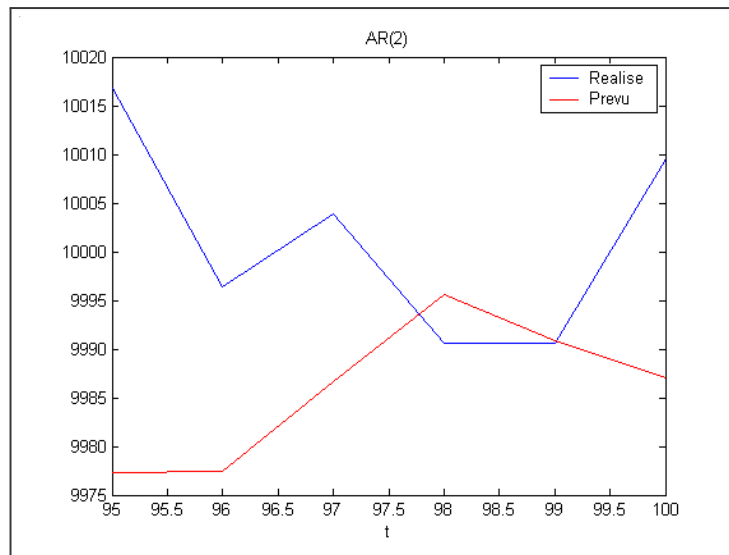


FIG. 2.3 – Prévion optimale MAVE (détail) : AR(2)

2.6.3 Série ARMA(1,1)

On a simulé le processus ARMA(1,1) suivant :

$$X_t = 1000 + \varphi_1 X_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

où :

- $t \in \{1, \dots, 100\}$
- $\varphi_1 = 0.8, \theta_1 = 0.2$
- $\varepsilon_t \rightsquigarrow \mathcal{N}(0, 5)$

Méthode	r	D	RMSE	MAPE (%)
Noyau	20	-	4.81	0.074
MAVE	10	2	5.31	0.091

TAB. 2.2 – Prévisions optimales "MAVE" et "Noyau" : ARMA(1,1)

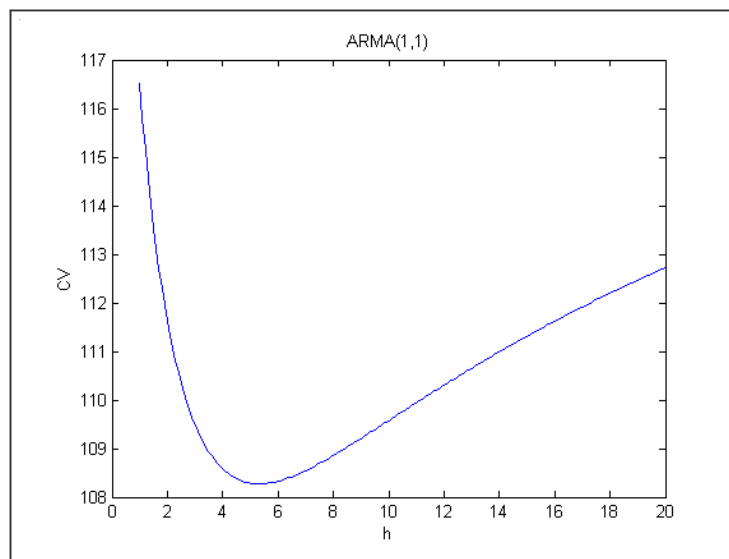


FIG. 2.4 – Critère de validation croisée MAVE : ARMA(1,1)

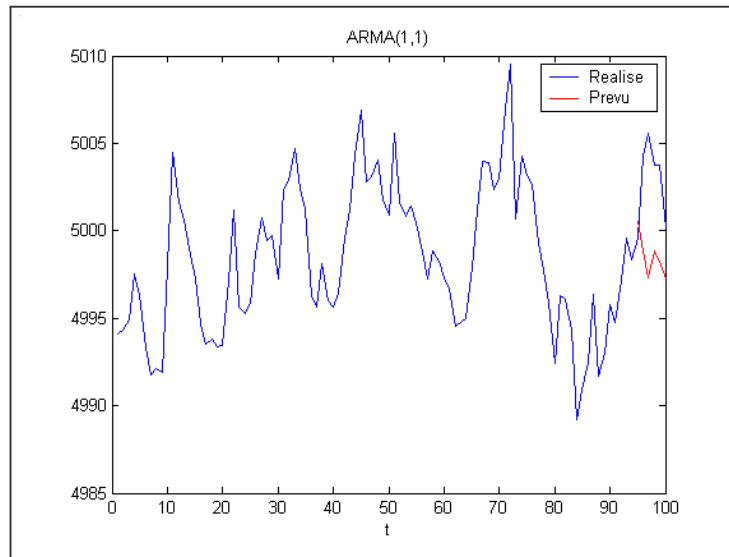


FIG. 2.5 – Prévision optimale MAVE : ARMA(1,1)

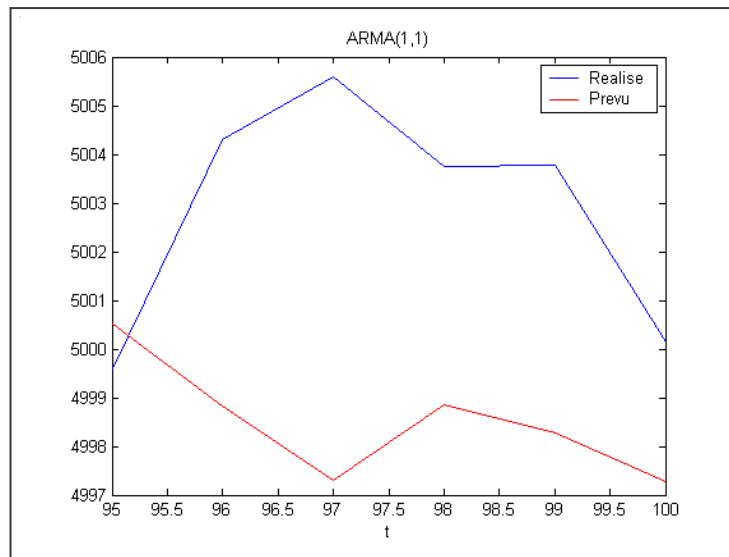


FIG. 2.6 – Prévision optimale MAVE (détail) : ARMA(1,1)

2.6.4 Séries "Cigars"

La série "Cigars" issue de Pankratz (1983) contient la consommation mensuelle américaine de cigares de 1969 à 1976.

Cette série est la seule à avoir subi une différenciation préalable (d'ordre 12) qui a permis d'améliorer de manière significative les résultats.

Méthode	r	D	RMSE	MAPE (%)
Noyau	5	-	23.39	3.12
MAVE	10	2	19.30	2.66

TAB. 2.3 – Prévisions optimales "MAVE" et "Noyau" : Cigars

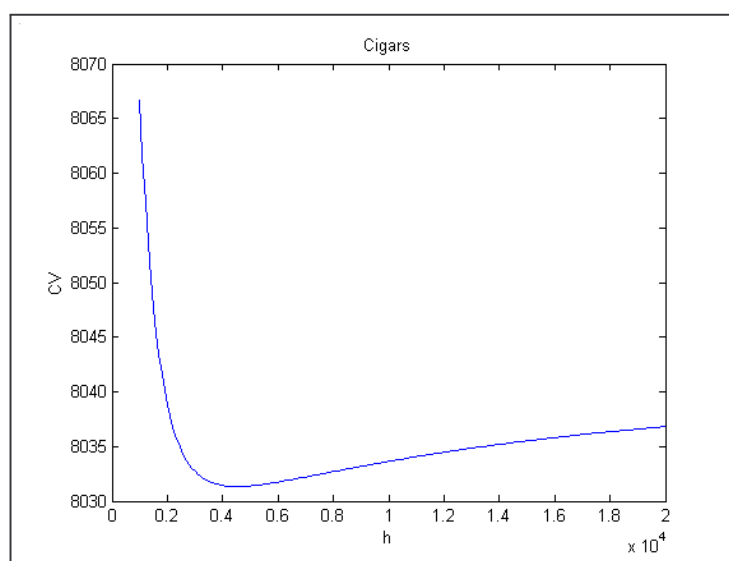


FIG. 2.7 – Critère de validation croisée : Cigars

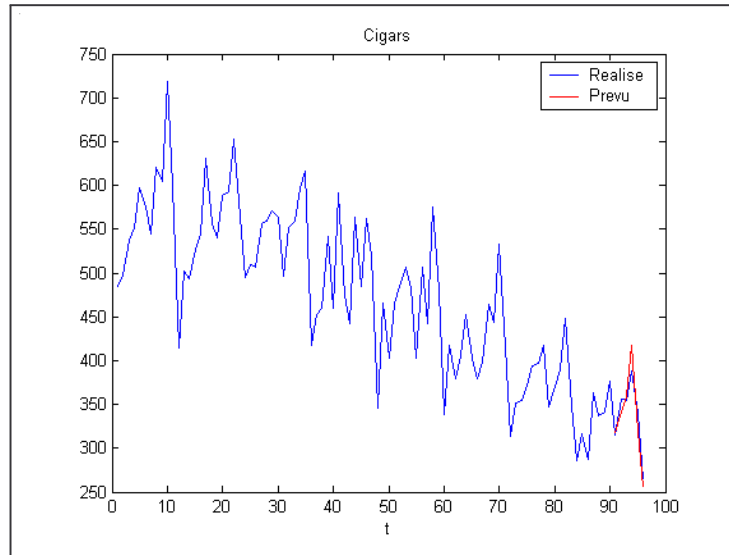


FIG. 2.8 – Prédiction optimale MAVE : Cigars

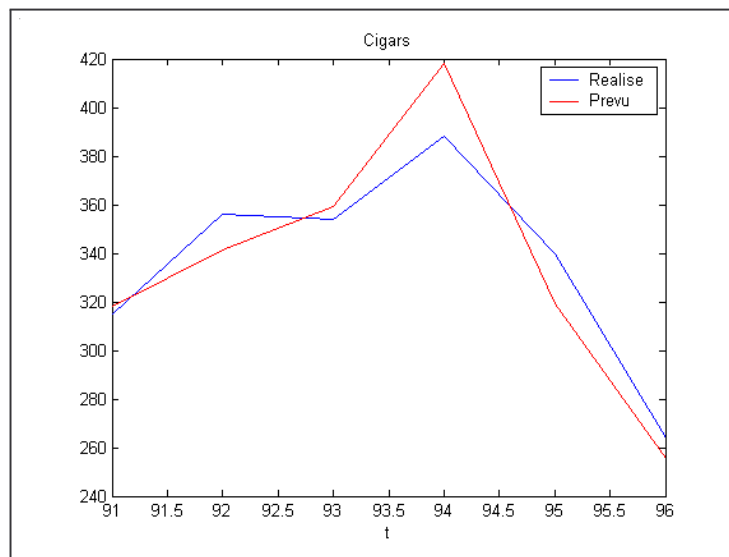


FIG. 2.9 – Prédiction optimale MAVE (détail) : Cigars

2.6.5 Série "Profit margin"

La série "Profit Margin" issue de Pankratz (1983) contient les bénéfices trimestriels des compagnies américaines en % de 1953 à 1972.

Méthode	r	D	RMSE	MAPE (%)
Noyau	24		0.432	9.22
MAVE	12	2	0.389	8.42

TAB. 2.4 – Prévisions optimales "MAVE" et "Noyau" : Profit margin

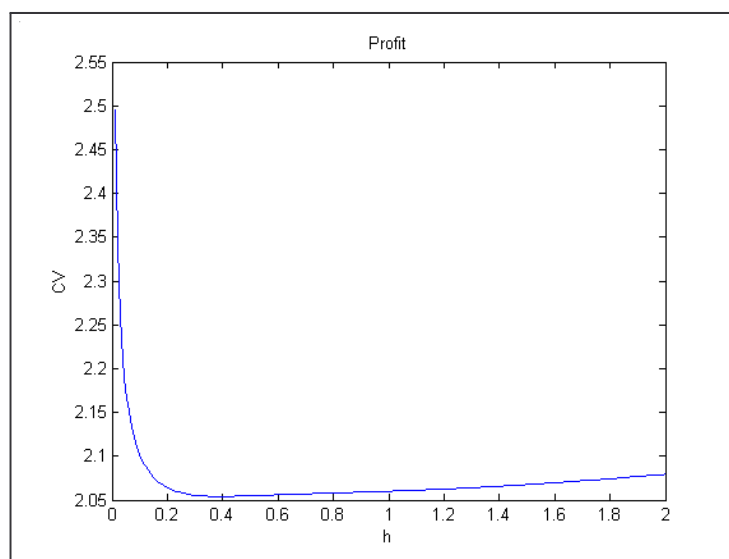


FIG. 2.10 – Critère de validation croisée MAVE : Profit Margin

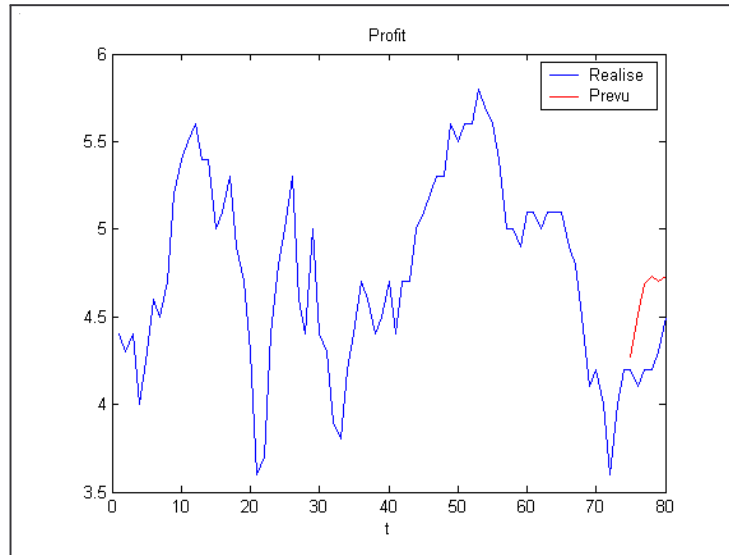


FIG. 2.11 – Prévision optimale MAVE : Profit margin

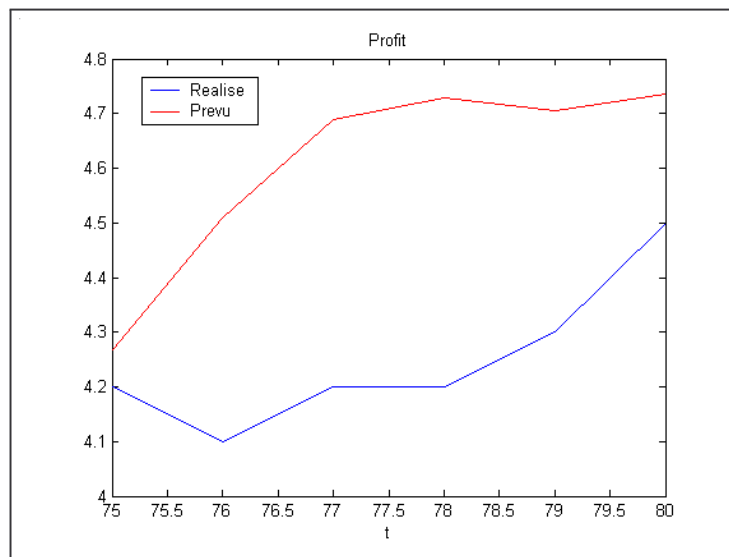


FIG. 2.12 – Prévision optimale MAVE (détail) : Profit margin

2.6.6 Série "Coal"

La série "Coal" issue de Pankratz (1983) contient la production mensuelle de charbon aux USA de 1952 à 1959.

Méthode	r	D	RMSE	MAPE (%)
Noyau	4		1342	3,35
MAVE	10	2	1681	4.92

TAB. 2.5 – Prévisions optimales "MAVE" et "Noyau" : Coal

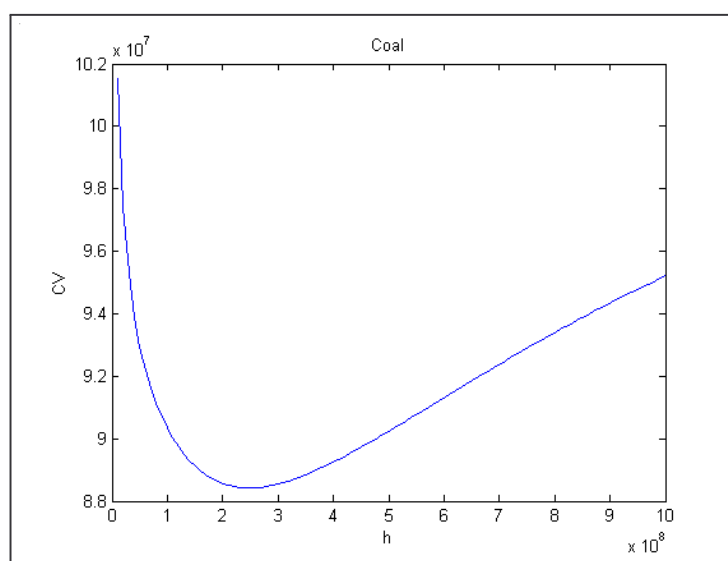


FIG. 2.13 – Critère de validation croisée MAVE : Coal

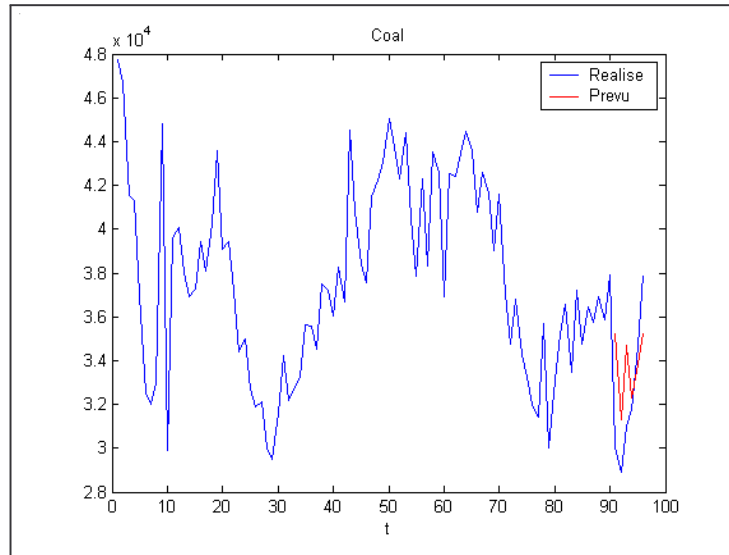


FIG. 2.14 – Prévision optimale MAVE : Coal

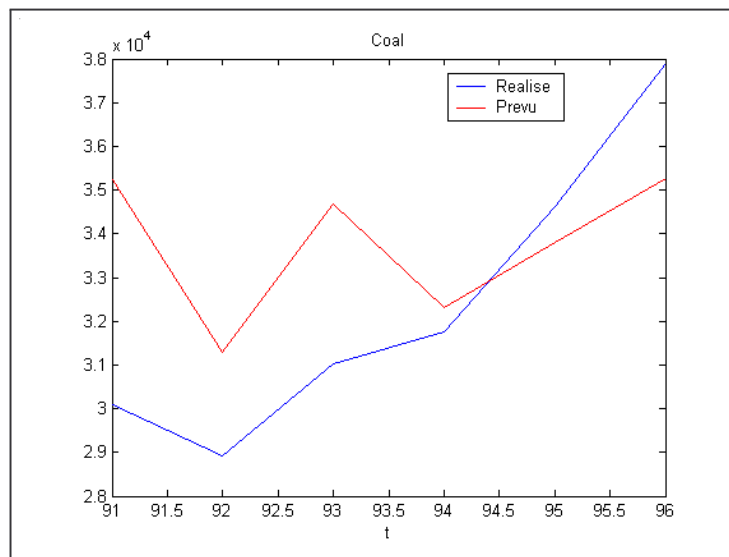


FIG. 2.15 – Prévision optimale MAVE (détail) : Coal

2.7 Application à la consommation d'électricité

On a exposé de manière détaillée le processus opérationnel utilisé à RTE pour effectuer la prévision de consommation d'électricité. On utilise un modèle de régression pour éliminer l'impact des variables explicatives dans le passé et rajouter cet impact dans le futur. Le modèle permettant de traiter la série "corrigée" de manière autorégressive est un modèle SARIMA. A ce stade, on est en mesure de remplacer cette étape par une prévision autorégressive basée sur MAVE. On travaille ainsi sur la série suivante :

$$\forall t \leq T : PC_t = P_t - \widehat{PEF}_t - \widehat{CH}_t - \widehat{CL}_t$$

Pour des commodités d'écriture, on renomme cette série $(X_t)_{t \in \mathbb{N}}$.

A partir de (X_1, \dots, X_T) , on souhaite prévoir X_{T+l} pour $l \in \{1, \dots, 24\}$, c'est-à-dire effectuer une prévision journalière en considérant qu'on se place à un instant T correspondant à 0h d'un jour donné. En pratique, la prévision n'est pas effectuée à minuit par RTE mais on a procédé de cette façon là par commodité de calculs.

On a sélectionné deux périodes durant lesquelles la consommation est stable, c'est-à-dire non perturbée par des "ruptures" (Noël par exemple). Ces périodes sont :

- Période 1 : du 3 avril 2006 au 9 juillet 2006
- Période 2 : du 15 janvier 2007 au 8 avril 2007

Comme indiqué auparavant, le choix de la mémoire est fastidieuse pour les calculs et on a choisi $r = 24$ qui nous semblait adapté à la consommation d'électricité. De même, nous avons estimé une seule matrice B pour chaque horizon de prévision. Nous avons ainsi conservé une estimation dynamique, en évitant de calculer à chaque étape une nouvelle matrice B .

Afin de comparer les résultats entre le modèle SARIMA et la méthode MAVE, nous avons tronqué les deux dernières semaines et effectué une prévision à minuit des 24 puissances horaires suivantes. Nous avons ainsi prévu a posteriori 14 jours (336 points horaires) la veille pour le lendemain.

Le critère de validation croisée nous a conduit à sélectionner 5 directions révélatrices pour la première période, et 7 pour la seconde.

Les résultats obtenus figurent dans les tableaux et graphiques suivants. Pour des raisons de confidentialité, nous ne fournissons que les résultats synthétiques obtenus, cela sera également le cas dans le troisième chapitre.

On constate que la méthode MAVE est légèrement moins bonne que le modèle SARIMA sur la première période, et légèrement meilleure sur la seconde période.

Pour la première période, le MAPE est de 1.8% avec la méthode MAVE alors qu'il ne vaut que 1.2% avec la méthode de Box et Jenkins. Les erreurs maximales et minimales sont également plus importantes avec la méthode MAVE.

Pour la seconde période, le MAPE vaut 1.1% pour la méthode MAVE contre 1.3% pour SARIMA. Les erreurs minimales et maximales sont moins importantes pour MAVE cette fois-ci.

Ces résultats fort encourageants devraient être confirmés sur des périodes plus longues si on désirait remplacer le modèle SARIMA par la méthode MAVE. Par contre, le protocole expérimental deviendra plus complexe puisqu'il faudra traiter les périodes dites spéciales; or le semi-paramétrique ne semble pas être plus robuste à des modifications de lois au cours du temps que ne le sont les techniques paramétriques classiques.

Notons cependant que ce choix n'est absolument pas une priorité pour RTE; ces premiers essais nous ont essentiellement permis de montrer les performances du semi-paramétrique, et d'envisager de les étendre au cas de la consommation d'électricité avec prise en compte des variables exogènes (Cf. Chapitre 3).

	MAVE	SARIMA
Moyenne	-104	-109
Ecart-type	1156	608
Minimum	-2570	-1887
Quantile 5%	-1172	-999
Médiane	-132	-185
Quantile 95%	1678	905
Maximum	2496	2108

TAB. 2.6 – Erreurs de prévision : Période 1

	MAVE	SARIMA
MAPE (%)	1.8	1.2
RMSE	1159	617

TAB. 2.7 – Critères d'erreurs de prévision : Période 1

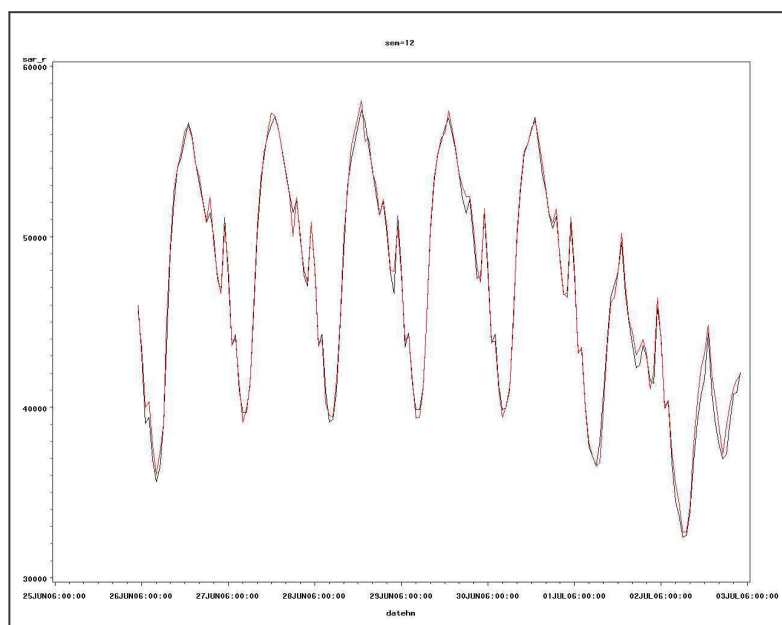


FIG. 2.16 – Consommations prévues MAVE (rouge) et réalisées (noir) : Période 1, première semaine tronquée

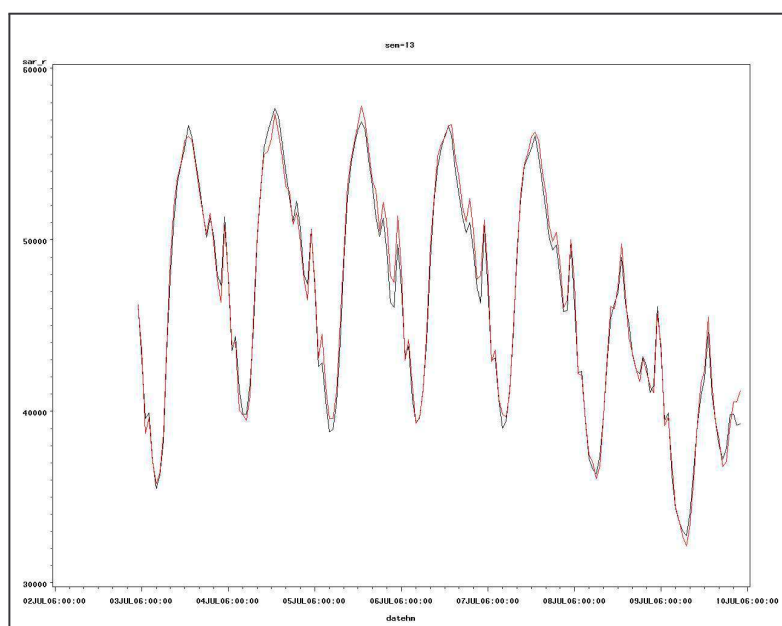


FIG. 2.17 – Consommations prévues MAVE (rouge) et réalisées (noir) : Période 1, seconde semaine tronquée

	MAVE	SARIMA
Moyenne	-109	344
Ecart-type	608	646
Minimum	-1087	-1180
Quantile 5%	-599	-659
Médiane	-185	334
Quantile 95%	905	1418
Maximum	1908	2007

TAB. 2.8 – Erreurs de prévision : Période 2

	MAVE	SARIMA
MAPE (%)	1.1	1.3
RMSE	732	774

TAB. 2.9 – Critères d'erreurs de prévision : Période 2

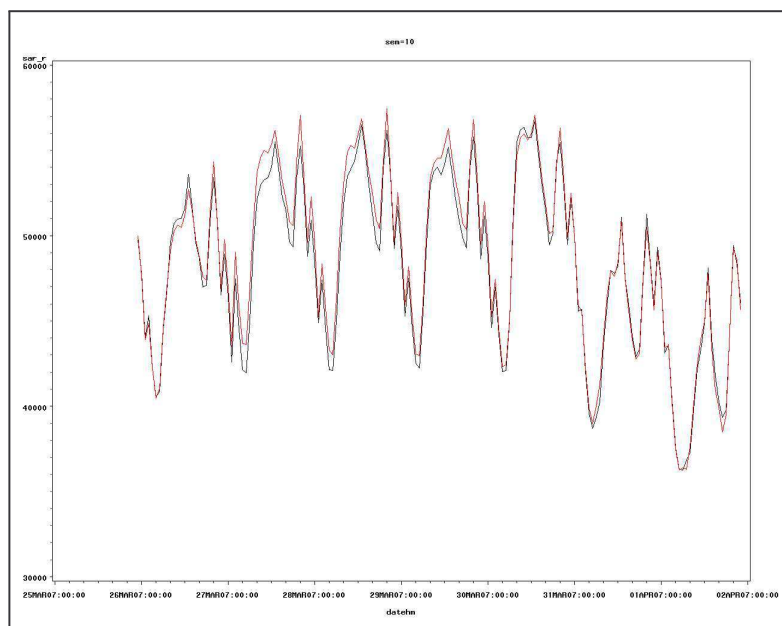


FIG. 2.18 – Consommations prévues MAVE (rouge) et réalisées (noir) : Période 2, première semaine tronquée

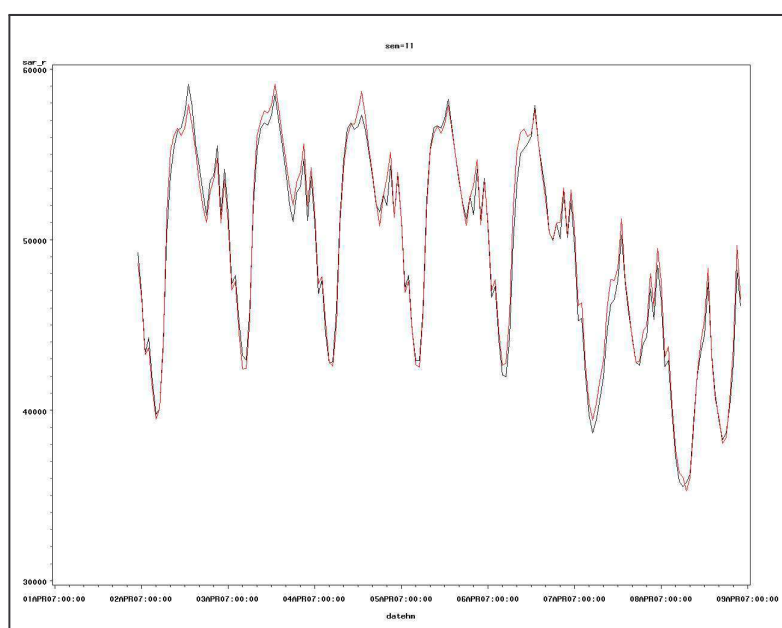


FIG. 2.19 – Consommations prévues MAVE (rouge) et réalisées (noir) : Période 2, seconde semaine tronquée

Chapitre 3

Prévisions semi-paramétriques en présence de variables exogènes

3.1 Problématique

Les résultats satisfaisants de la prévision autorégressive à l'aide de la méthode MAVE, à la fois sur les séries basiques, et sur la consommation d'électricité, nous ont incités à l'étendre en incorporant les variables explicatives telles que la température et la nébulosité. En considérant la consommation d'électricité P_t comme variable à expliquer, on peut considérer comme variables explicatives le passé de la consommation, la température instantanée et son passé, et enfin la nébulosité. On ne prend pas en compte le tarif EJP car son effet est de plus en plus faible et risquerait de perturber l'estimation de notre modèle.

Si on reprend les notations utilisées pour la méthode MAVE, on peut considérer comme variables d'intérêt :

- $Y = P_t$
- $X = (P_{t-1}, \dots, P_{t-r}, T_t, \dots, T_{t-r'+1}, N_t)^T$

Pour expliquer la puissance à l'instant t , on prend ainsi en compte les r puissances précédentes, la température à l'instant t et les r' températures passées, ainsi que la nébulosité à l'instant t . Ce choix de variables n'est pas effectué au hasard mais repose sur notre connaissance de cette série, qu'on a exposée dans le premier chapitre.

On considère donc le modèle suivant :

$$P_t = g\left(B^T (P_{t-1}, \dots, P_{t-r}, T_t, \dots, T_{t-r'+1}, N_t)^T\right) + \varepsilon_t \quad (3.1)$$

avec les hypothèses classiques du modèle MAVE.

La matrice B est ici de dimension $(r + r' + 2) \times D$, qui peut devenir très grande si on prend des valeurs de r et r' importantes.

On a tenté de multiples essais pour estimer ce modèle, malheureusement sans succès. On peut émettre à ce stade différentes hypothèses expliquant cet échec.

Le nombre de variables explicatives

On a essayé d'estimer ce modèle pour $r \in \{24, 48, 168\}$ et $r' \in \{24, 48\}$. On avait en effet mis en évidence des saisonnalités journalières et hebdomadaires pour la consommation d'électricité, ainsi que l'influence des températures des 2 derniers jours. On considérait alors entre 49 et 217 variables explicatives au total.

On a essayé de limiter le nombre de variables en sélectionnant celles qui nous paraissaient les plus censées à partir de notre connaissance a priori des données. Malheureusement, cette sélection manuelle ne nous a pas permis d'aboutir, sachant que les temps de calcul limitent fortement le nombre d'essais réalisables dans un temps décent. L'échec se concrétise de deux manières différentes : le résidu du modèle peut être très important, et si ce résidu ne nous semble pas trop élevé, c'est la qualité de la prévision qui devient alors très mauvaise.

A ce niveau, il faudrait envisager une procédure de sélection de variables. On peut facilement imaginer que cette procédure serait de toute façon pénalisée par les temps de calcul. L'idée qui nous a guidé dans le choix des variables explicatives était de considérer les variables pour lesquelles les coefficients de B étaient plus importants. Une autre idée, peu rigoureuse théoriquement, serait d'effectuer une ACP afin de détecter a priori les combinaisons les plus explicatives. En dehors du manque de rigueur, différents essais d'ACP avaient déjà été effectués dans le passé, sans succès, sans doute à cause de la difficulté à séparer l'impact du passé de la consommation et des variables météorologiques.

La séparabilité des impacts autorégressifs et de l'influence météorologique

On peut indiquer ici qu'on a effectué des tentatives en normalisant les variables pour limiter les effets d'échelle qui existent entre la consommation d'électricité, la température et la nébulosité.

Il est clair que les puissances passées dépendent également des températures passées ce qui pose un problème de spécification de notre modèle. L'estimation des directions révélatrices s'effectue avec des variables explicatives qui présentent de fortes dépendances entre elles ; on se rapproche des problèmes de colinéarité des modèles linéaires généralisés.

La convergence de l'algorithme d'optimisation

Il existe de nombreuses boucles dans l'algorithme utilisé et il est vraisemblable qu'on soit confronté à des problèmes de convergence. Le choix initial de la fenêtre n'est

sans doute pas neutre, même si les auteurs de la méthode MAVE ne mentionnent pas ce problème. On dispose maintenant de moyens de calcul très importants, mais il n'en demeure pas moins que cette méthode est réellement coûteuse et on n'a aucune garantie de l'existence d'un minimum global. Seule l'analyse numérique pourrait répondre à cette question non triviale.

Le choix de la fenêtre

On optimise le modèle à l'aide de la méthode MAVE, en utilisant une seule fenêtre pour toutes les directions. On pourrait ne plus imposer cette unicité afin d'ajuster au mieux le modèle, mais au détriment du temps de calcul. Cette voie n'a donc pas été envisagée.

La juxtaposition d'un modèle de régression et d'un modèle autorégressif pour prévoir la courbe de charge a été envisagé historiquement pour utiliser un modèle SARIMA et prendre en compte les variables exogènes. La séparation des effets autorégressifs et des effets dus aux variables exogènes est assurée actuellement par deux modèles utilisés indépendamment. Si on désire toujours unifier ces deux modèles, il nous semble judicieux de considérer un seul modèle qui sépare lui-même ces deux effets, d'où une équation du type :

$$P_t = f(T_t, \dots, T_{t-r'+1}, N_t) + g(P_{t-1}, \dots, P_{t-r}) + \varepsilon_t$$

où f et g sont des fonctions de lien inconnues.

On peut envisager d'estimer de manière non ou semi-paramétrique les deux fonctions f et g , mais la difficulté technique est très importante. On n'a pas trouvé de modèles de ce type dans la littérature, du moins avec autant de variables explicatives, que ce soit en non ou en semi-paramétrique.

Il existe par contre de nombreuses tentatives de réduction de la dimension, qui incorporent certaines linéarités.

Citons par exemple les travaux de Damon *et al.* (2005) qui ont travaillé sur la prévision des pics d'ozone : ils ont utilisé un modèle autorégressif hilbertien avec prise en compte de variables exogènes. En pratique, ils ont appliqué un modèle ARHX(1), comprenant une partie autorégressive d'ordre 1 et 6 variables exogènes. Cette voie pourrait faire l'objet de recherches ultérieures, même si la mémoire de la consommation d'électricité peut constituer un écueil.

On peut également citer les travaux de Hastie *et al.* (1993) qui ont proposé un modèle de régression additif non-paramétrique à coefficients variables :

$$Y = \Phi_0(Z) + \Phi_1(Z) X_1 + \dots + \Phi_p(Z) X_p + \varepsilon$$

On aurait ainsi pu envisager un modèle autorégressif prenant en compte les variables météorologiques dans les poids. On n'a pas exploré cette voie qui ne nous semble pas correspondre à la dynamique de nos données.

Xia *et al.* (1999) ont quant à eux étendu ce modèle en utilisant une direction révélatrice simple pour réduire la dimension des variables explicatives entrant en jeu dans le calcul des poids :

$$Y = \Phi_0(g_{\theta_0}(Z)) + \Phi_1(g_{\theta_0}(Z))X_1 + \dots + \Phi_p(g_{\theta_0}(Z))X_p + \varepsilon$$

Au vu de nos données, nous avons préféré considérer un modèle semi-paramétrique semi-linéaire.

3.2 Proposition d'un modèle alternatif

3.2.1 Introduction

La littérature est abondante sur les modèles de régression semi-linéaires ("partially linear models"). On en expose quelques uns, qu'on se propose d'étendre au cas semi-paramétrique multidirectionnel.

Robinson (1988) a considéré le modèle de régression semi-linéaire suivant :

$$Y = \Phi(X) + \alpha^T Z + \varepsilon \tag{3.2}$$

où :

- $Y \in \mathbb{R}$ est la variable à expliquer
- $X \in \mathbb{R}^p$ sont p variables explicatives
- $Z \in \mathbb{R}^q$ sont q variables explicatives
- $\Phi : \mathbb{R}^p \rightarrow \mathbb{R}$ est une fonction mesurable inconnue
- $\alpha = (\alpha_1, \dots, \alpha_q)^T$ est un paramètre (vectoriel) inconnu
- $\mathbb{E}(\varepsilon / X, Z) = 0$ p.s.

Afin d'estimer le modèle (3.2), on conditionne par rapport à X :

$$\mathbb{E}(Y / X) = \Phi(X) + \alpha^T \mathbb{E}(Z / X)$$

D'où :

$$Y - \mathbb{E}(Y / X) = \alpha^T [Z - \mathbb{E}(Z / X)] + \varepsilon$$

Après avoir estimé $\mathbb{E}(Y / X)$ et $\mathbb{E}(Z / X)$ de manière non-paramétrique, on peut estimer α par la méthode des moindres carrés, puis estimer Φ de manière non-paramétrique.

L'inconvénient majeur est que les différentes études sur ces modèles semi-linéaires ont été majoritairement effectuées en posant une hypothèse d'indépendances de l'échantillon $(Z_i, X_i, Y_i)_{i \in \{1, \dots, n\}}$, ce qui ne convient évidemment pas au cas des séries temporelles.

Shen (1996), après avoir étudié la notion d'exogénéité, a proposé le modèle de séries temporelles suivant :

$$X_{t+1} = \Phi(X_t) + \Psi(Z_t) + \varepsilon_t \quad (3.3)$$

où :

- $X_t \in \mathbb{R}$ est la variable endogène
- $Z_t \in \mathbb{R}$ est la variable exogène

La méthode d'estimation est similaire à celle de Robinson (1988). Il suppose que $(X_t, Z_t)_{t \in \mathbb{N}}$ est un processus géométriquement mélangeant et que la fonction Ψ est développable en série de Mac-Laurin.

Ce type de modèle, quoiqu'intéressant, ne considère qu'une variable endogène et qu'une variable exogène, essentiellement pour pallier le fléau de la dimension. La consommation d'électricité est trop complexe pour pouvoir être modélisable de cette façon et il faut donc étendre les résultats à des dimensions plus élevées. Le semi-paramétrique apparaît là encore comme une alternative intéressante.

3.2.2 Modèle

On considère le modèle de régression semi-linéaire suivant :

$$Y = \alpha^T Z + g(B^T X) + \varepsilon \quad (3.4)$$

où :

- $Y \in \mathbb{R}$ est la variable à expliquer
- $X \in \mathbb{R}^p$ sont p variables explicatives
- $Z \in \mathbb{R}^q$ sont q variables explicatives
- ε est une v.a.r. dans $\mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ admettant une fonction de densité.
- $\mathbb{E}(\varepsilon / X, Z) = 0$ p.s.
- $g : \mathbb{R}^D \rightarrow \mathbb{R}$ est une fonction mesurable inconnue
- $B = (\beta^1, \dots, \beta^D)_{p \times D}$, avec $D \in \{1, \dots, p\}$, est une matrice inconnue telle que $B^T B = I_D$
- $\alpha = (\alpha_1, \dots, \alpha_q)^T$ est un paramètre inconnu

Dans notre cas d'étude, on considérera par la suite :

- $Y = P_t$

- $X = (P_{t-1}, \dots, P_{t-r})^T$
- $Z = (T_t, \dots, T_{t-r'+1}, N_t)^T$

Ce modèle a été étudié dans le cas d'une seule direction révélatrice par Carroll *et al.* (1997), ainsi que par Xia *et al.* (2006). Ils ont proposé le modèle GPLSIM en considérant $\beta = (\beta_1, \dots, \beta_p)^T$ à la place de la matrice B .

Notre propos peut paraître paradoxal puisqu'on a abandonné la méthode MAVE globale pour notre série d'étude, pour des raisons de complexité de modélisation, et on propose comme alternative un modèle a priori plus général. En effet, le modèle (3.4) inclut les modèles semi-paramétriques de type MAVE : il suffit d'enlever la partie linéaire, c'est-à-dire poser $\alpha = 0$. Néanmoins, notre objectif est de séparer les variables explicatives et de les "répartir" entre X et Z , au lieu de les inclure toutes dans X . C'est pourquoi, on considère que le modèle semi-linéaire est plus spécifié que le modèle général envisagé auparavant.

3.3 Estimation du modèle

On expose deux procédures d'estimation du modèle (3.4), l'une basée sur le conditionnement partiel du modèle (3.4), et l'autre basée sur une méthode dérivée de MAVE.

3.3.1 Estimation conditionnelle

Afin d'estimer le modèle (3.4), on peut procéder en deux temps : conditionner le problème de manière non-paramétrique puis utiliser les moindres carrés. Ainsi, si on conditionne le modèle (3.4) par rapport à X , on obtient :

$$\mathbb{E}(Y/X) = \alpha^T \mathbb{E}(Z/X) + g(B^T X)$$

On en déduit :

$$Y - \mathbb{E}(Y/X) = \alpha^T [Z - \mathbb{E}(Z/X)] + \varepsilon$$

A partir d'un échantillon $(X_i, Y_i, Z_i)_{i \in \{1, \dots, n\}}$, si on dispose d'estimations de $\mathbb{E}(Y_i/X_i)$ et $\mathbb{E}(Z_i/X_i)$, on se ramène alors à un modèle de régression linéaire classique.

On peut estimer $\mathbb{E}(Y_i/X_i)$ et $\mathbb{E}(Z_i/X_i)$ de manière non-paramétrique : on note respectivement \widehat{Y}_i et \widehat{Z}_i ces estimateurs.

L'estimateur de Nadaraya-Watson nous fournit :

$$\widehat{Y}_i = \frac{\sum_{j=1}^n K_h(X_i - X_j) Y_j}{\sum_{j=1}^n K_h(X_i - X_j)} \text{ et } \widehat{Z}_i = \frac{\sum_{j=1}^n K_h(X_i - X_j) Z_j}{\sum_{j=1}^n K_h(X_i - X_j)}$$

On dispose ainsi des valeurs $(Y_i - \widehat{Y}_i, Z_i - \widehat{Z}_i)_{i \in \{1, \dots, n\}}$ pour estimer α par les moindres carrés.

On a :

$$\widehat{\alpha} = S_{Z_i - \widehat{Z}_i, Z_i - \widehat{Z}_i}^{-1} S_{Y_i - \widehat{Y}_i, Z_i - \widehat{Z}_i} \quad (3.5)$$

où :

$$S_{U,V} = (U_1, \dots, U_n) (V_1, \dots, V_n)^T$$

Il est possible d'affiner cet estimateur en filtrant les dénominateurs $\sum_{j=1}^n K_h(X_i - X_j)$ trop faibles qui conduisent numériquement à une division par 0.

De même, si on rencontre des problèmes d'hétéroscédasticité, on peut "pondérer" notre problème en ajoutant une matrice de variance-covariance estimée. Il est également possible de pallier ce problème en normalisant les données en amont.

Une fois $\widehat{\alpha}$ déterminé, on peut estimer g et B à partir de $(X_i, Y_i - \widehat{\alpha}^T Z_i)_{i \in \{1, \dots, n\}}$ par la méthode MAVE. On a en effet :

$$\forall i \in \{1, \dots, n\} : Y_i - \widehat{\alpha}^T Z_i = g(B^T X_i) + \varepsilon'_i$$

On obtient le problème de minimisation suivant, en supposant la dimension D connue :

$$\min_{B/B^T B = I_D, (a_j, b_j)_{j \in \{1, \dots, n\}}} \sum_{j=1}^n \sum_{i=1}^n \{Y_i - \widehat{\alpha}^T Z_i - [a_j + b_j^T B^T (X_i - X_j)]\}^2 \omega_{ij} \quad (3.6)$$

Algorithme

L'algorithme finalement utilisé est le suivant.

1. Estimation non-paramétrique de $\mathbb{E}(Y_i/X_i)$ et $\mathbb{E}(Z_i/X_i)$ à l'aide de la méthode des noyaux.

On utilise les échantillon $(X_i, Y_i)_{i \in \{1, \dots, n\}}$ et $(X_i, Z_i)_{i \in \{1, \dots, n\}}$.

On optimise un critère de validation croisée afin d'obtenir la fenêtre optimale.

2. Estimation par les moindres carrés de $\hat{\alpha}$.

On utilise l'échantillon $(Y_i - \hat{Y}_i, Z_i - \hat{Z}_i)_{i \in \{1, \dots, n\}}$.

3. Estimation semi-paramétrique de \hat{B} .

On utilise l'échantillon $(X_i, Y_i - \hat{\alpha}^T Z_i)_{i \in \{1, \dots, n\}}$

On optimise deux critères de validation croisée afin d'obtenir le nombre optimal de directions révélatrices, et la fenêtre optimale associée.

– Choix du nombre optimal de directions.

On utilise un critère de validation croisée afin de déterminer la dimension optimale :

$$\hat{D} = \underset{d \in \{1, \dots, d_{\max}\}}{\operatorname{arg\,min}} \operatorname{CV}(d)$$

– Pour chaque valeur $d \in \{1, \dots, d_{\max}\}$ testée (parmi lesquelles on recherche la dimension optimale), on utilise un critère de validation croisée afin d'obtenir la fenêtre optimale. On peut également itérer en modifiant les poids au fur et à mesure (méthode rMAVE).

Cette méthode n'est pas celle qu'on favorisera par la suite car les résultats peuvent être mauvais si la dimension de X est trop importante : on est alors de nouveau confronté au fléau de la dimension. En pratique, on peut envisager une alternative :

1. Estimation non-paramétrique de $\mathbb{E}(Y_i/X_i)$ et $\mathbb{E}(Z_i/X_i)$ à l'aide de la méthode des noyaux.
2. Estimation par les moindres carrés de $\hat{\alpha}$.
3. Estimation semi-paramétrique de \hat{B} .
4. Estimation semi-paramétrique de $\mathbb{E}(Y_i/X_i)$ et $\mathbb{E}(Z_i/X_i)$ à l'aide de la méthode MAVE :

$$\hat{Y}_i = \sum_{j=1}^n \frac{K_h \left[\hat{B}^T (X_i - X_j) \right]}{\sum_{j=1}^n K_h \left[\hat{B}^T (X_i - X_j) \right]} Y_j \quad \text{et} \quad \hat{Z}_i = \sum_{j=1}^n \frac{K_h \left[\hat{B}^T (X_i - X_j) \right]}{\sum_{j=1}^n K_h \left[\hat{B}^T (X_i - X_j) \right]} Z_j$$

5. On réitère les étapes 2, 3 et 4 jusqu'à l'obtention de résultats stables.

3.3.2 Estimation globale (PLMAVE)

Au lieu d'estimer dans un premier temps des espérances conditionnelles afin de se ramener à une résolution des moindres carrés, puis une résolution MAVE, on considère le problème dans sa globalité. On évite ainsi les problèmes de convergence des estimateurs non-paramétriques.

Les paramètres α et B solutions de (3.4) vérifient :

$$\min_{\alpha \in \mathbb{R}^q, B/B^T B = I_D} \mathbb{E} \left[(Y - \alpha^T Z - \mathbb{E}(Y / B^T X))^2 \right] \quad (3.7)$$

où $B = (\beta_1, \dots, \beta_D)_{p \times D}$ avec $D \in \{1, \dots, p\}$.

On réutilise l'idée de la méthode MAVE qui consiste à travailler sur la variance conditionnelle. On considère la variance de Y conditionnellement à $B^T X$:

$$\sigma_B^2 (B^T X) = \mathbb{E} \left[(Y - \alpha^T Z - \mathbb{E}(Y / B^T X))^2 / B^T X \right]$$

On en déduit que :

$$\mathbb{E} [\sigma_B^2 (B^T X)] = \mathbb{E} \left[(Y - \alpha^T Z - \mathbb{E}(Y / B^T X))^2 \right]$$

Le problème de minimisation (3.7) est donc équivalent à :

$$\min_{\alpha \in \mathbb{R}^q, B/B^T B = I_D} \mathbb{E} [\sigma_B^2 (B^T X)] \quad (3.8)$$

Soit la fonction suivante :

$$g_B(u_1, \dots, u_D) = \mathbb{E}(Y / \beta_1^T X = u_1, \dots, \beta_D^T X = u_D)$$

Soit $(Y_i, Z_i, X_i)_{i \in \{1, \dots, n\}}$ un échantillon de (Y, Z, X) .

De manière analogue à la méthode MAVE, on effectue un développement limité de $\mathbb{E}(Y_i / B^T X_i)$ en X_l pour $l \in \{1, \dots, n\}$:

$$\mathbb{E}(Y_i / B^T X_i) \simeq a + b^T B^T (X_i - X_l)$$

On estime alors $\sigma_B^2 (B^T X_l)$:

$$\hat{\sigma}_B^2 (B^T X_l) = \min_{a,b} \sum_{i=1}^n \{Y_i - \alpha^T Z_i - [a + b^T B^T (X_i - X_l)]\}^2 \omega_{il}$$

où :

$$\begin{cases} a = g_B(B^T X_l) \\ b = (b_{(1)}, \dots, b_{(D)})^T \end{cases}$$

avec :

$$\forall k \in \{1, \dots, D\} : b_{(k)} = \left. \frac{\partial g_B(u_1, \dots, u_D)}{\partial u_k} \right|_{u_1 = \beta_1^T X_l, \dots, u_D = \beta_D^T X_l}$$

$(\omega_{il})_{i \in \{1, \dots, n\}}$ sont des poids traduisant la proximité entre X_i et X_l , tels que :

- $\forall l \in \{1, \dots, n\} : \omega_{il} \in]0, 1[$
- $\sum_{i=1}^n \omega_{il} = 1$

On estime donc B , solution de (3.7), en résolvant le problème suivant :

$$\min_{\alpha \in \mathbb{R}^q, B/B^T B = I_d, (a_j, b_j)_{j \in \{1, \dots, n\}}} \sum_{j=1}^n \sum_{i=1}^n \{Y_i - \alpha^T Z_i - [a_j + b_j^T B^T (X_i - X_j)]\}^2 \omega_{ij} \quad (3.9)$$

On peut résoudre l'équation (3.9) de manière itérative :

1. On initialise l'algorithme avec $\tilde{\alpha}$ et \tilde{B} . On peut par exemple considérer $\tilde{B} = I_p$ et prendre pour $\tilde{\alpha}$ l'estimation de α dans la régression linéaire suivante :

$$Y = c + \alpha^T Z + \varepsilon$$

2. On estime $(a_j, b_j)_{j \in \{1, \dots, n\}}$ pour (α, B) fixés.
3. On estime (α, B) pour $(a_j, b_j)_{j \in \{1, \dots, n\}}$ fixés.
4. On réitère les étapes 2 et 3 jusqu'à l'obtention de résultats stables.

Notons que le choix des poids est similaire à celui de la méthode MAVE. On peut choisir des poids simples ou des poids affinés comme mentionné dans le second chapitre.

On nomme cette méthode PLMAVE (Partially Linear MAVE) dans la suite du document.

3.4 Problèmes de convergence

Xia *et al.* (2006) ont établi le résultat de convergence suivant dans le cas d'une direction révélatrice unique ($B = \beta \in \mathbb{R}^p$).

Soient les hypothèses suivantes :

HP1 $(Y_i, Z_i, X_i)_{i \in \mathbb{N}}$ est strictement stationnaire et géométriquement mélangeant.

HP2 Il existe un compact D tel que $\mathbb{P}(X \in D) = 1$.

Les fonctions de densité f_X et $f_{\beta^T X}$ de X et $\beta^T X$ (pour $\|\beta\| = 1$) admettent des dérivées bornées.

Il existe $c_0 \in \mathbb{R}$ tel que $\{x / f_X(x) \geq c_0\}$ et $\{x / f_{\beta^T X}(\beta^T x) > c_0\}$ sont non vides.

HP3 $\forall (\beta_1, \beta_2) \in \mathbb{R}^p \times \mathbb{R}^p / \beta_1 \perp \beta_2, \exists c \in \mathbb{R} : f_{(\beta_1^T X, \beta_2^T X)}(x_1, x_2) < c f_{\beta_1^T X}(x_1) f_{\beta_2^T X}(x_2)$

HP4 g admet des dérivées (jusqu'à l'ordre 3) continues et bornées.

$\mathbb{E}(Z / X = x), \mathbb{E}(Z^T Z / X = x), \mathbb{E}(U / \beta^T X = x)$ et $\mathbb{E}(U^T U / \beta^T X = x)$,

où $U = (X^T, Z^T)^T$, admettent des dérivées bornées.

$\mathbb{E}(Y^k / X = x), \mathbb{E}(|Z|^k / X = x), \mathbb{E}(|Z_l Z_1| / X_1 = x_1, X_l = x_l)$

et $\mathbb{E}(|Z_l Z_1| / \beta^T X_1 = \beta^T x_1, \beta^T X_l = \beta^T x_l)$, pour tout $k \in \mathbb{N} / k > 2, (x, x_1, x_l) \in (\mathbb{R}^p)^3$, sont des fonctions bornées.

HP5 H est une fonction de densité admettant une dérivée bornée, et à support compact.

K est une fonction de densité symétrique sphérique admettant une dérivée bornée, à support compact et admettant une transformée de Fourier absolument intégrable.

HP6 $\mathbb{E}[(Z - \mathbb{E}(Z / X = x))(Z - \mathbb{E}(Z / X = x))^T]$ est une matrice définie positive.

Théorème 3.1 Xia *et al.* (2006)

Soit $(\hat{\alpha}, \hat{\beta})$ l'estimateur de (α, β) obtenu à l'aide de GPLSIM en initialisant l'algorithme par $(\tilde{\alpha}, \tilde{\beta})$.

Sous les hypothèses (HP1-HP6), et en supposant que $h \sim n^{-\delta}$ où $\frac{1}{6} < \delta < \frac{1}{4}$, on a :

$$\sqrt{n} \begin{pmatrix} \hat{\alpha} - \alpha \\ \hat{\beta} \pm \beta \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(0, W)$$

On pourra trouver l'expression complète de W dans Xia *et al.* (2006).

Remarquons que le signe indéterminé de l'expression ci-dessus correspond au signe de $\tilde{\beta}^T \beta$.

Nous ne sommes pas encore parvenus à établir de résultats de convergence pour le modèle (3.4), estimé à l'aide de la méthode dérivée de MAVE. La difficulté technique n'est pas anodine mais nous pensons qu'il sera possible de mettre en évidence une convergence en loi dans un futur proche. La démonstration de ce résultat est très technique et l'étendre au cas général n'est pas évident. Nous pensons qu'il est possible d'établir un résultat similaire mais en considérant $\left\| \widehat{B}\widehat{B}^T - BB^T \right\|$ au lieu de $\widehat{\beta} \pm \beta$. Le travail qui suit consiste à analyser les résultats de la méthode appliquée à des séries temporelles simulées, ainsi qu'à la consommation d'électricité en posant la conjecture que les estimateurs convergent.

3.5 Simulations

On a appliqué la méthode PLMAVE sur deux jeux de données, l'un correspondant à un modèle de régression classique, l'autre à des données dépendantes temporellement. On s'est inspiré de jeux de données utilisés par Xia *et al.* (2002) et Xia *et al.* (2006) qu'on a complexifié afin de disposer de modèles à directions révélatrices multiples.

3.5.1 Simulations sur données indépendantes

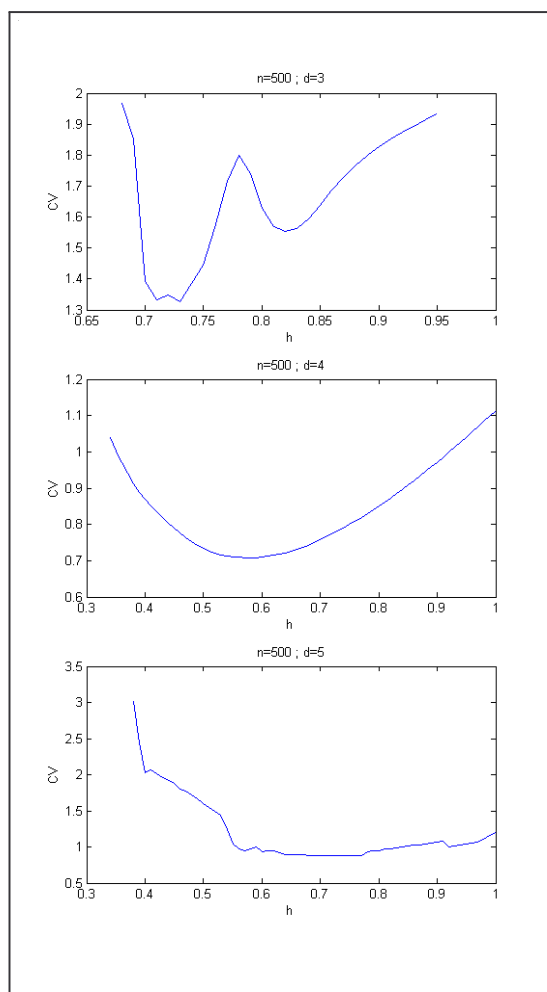
On a simulé $n \in \{200, 500, 800\}$ données indépendantes issues du modèle suivant :

$$Y = \alpha^T Z + X^T \beta_1 (X^T \beta_2)^2 + (X^T \beta_3) (X^T \beta_4) + 0.5\varepsilon$$

où :

- $\alpha = (2, 5)^T$
- $\beta_1 = \frac{1}{\sqrt{30}} (1, 2, 3, 4, 0, 0, 0, 0, 0, 0)^T$
- $\beta_2 = \frac{1}{\sqrt{35}} (-2, 1, -4, 3, 1, 2, 0, 0, 0, 0)^T$
- $\beta_3 = \frac{1}{\sqrt{15}} (0, 0, 0, 0, 2, -1, 2, 1, 2, 1)^T$
- $\beta_4 = \frac{1}{2} (0, 0, 0, 0, 0, 0, -1, -1, 1, 1)^T$
- $Z \rightsquigarrow \mathcal{N}(0, I_2)$, $X \rightsquigarrow \mathcal{N}(0, I_{10})$ et $\varepsilon \rightsquigarrow \mathcal{N}(0, 1)$ sont des v.a. indépendantes.

On a estimé $(\alpha, \beta_1, \beta_2, \beta_3, \beta_4)$ à l'aide de PLMAVE. On présente les résultats obtenus avec $n \in \{200, 500, 800\}$ et $d \in \{3, 4, 5\}$.

FIG. 3.1 – Critères de validation croisée : $n = 500$; $d \in \{3, 4, 5\}$

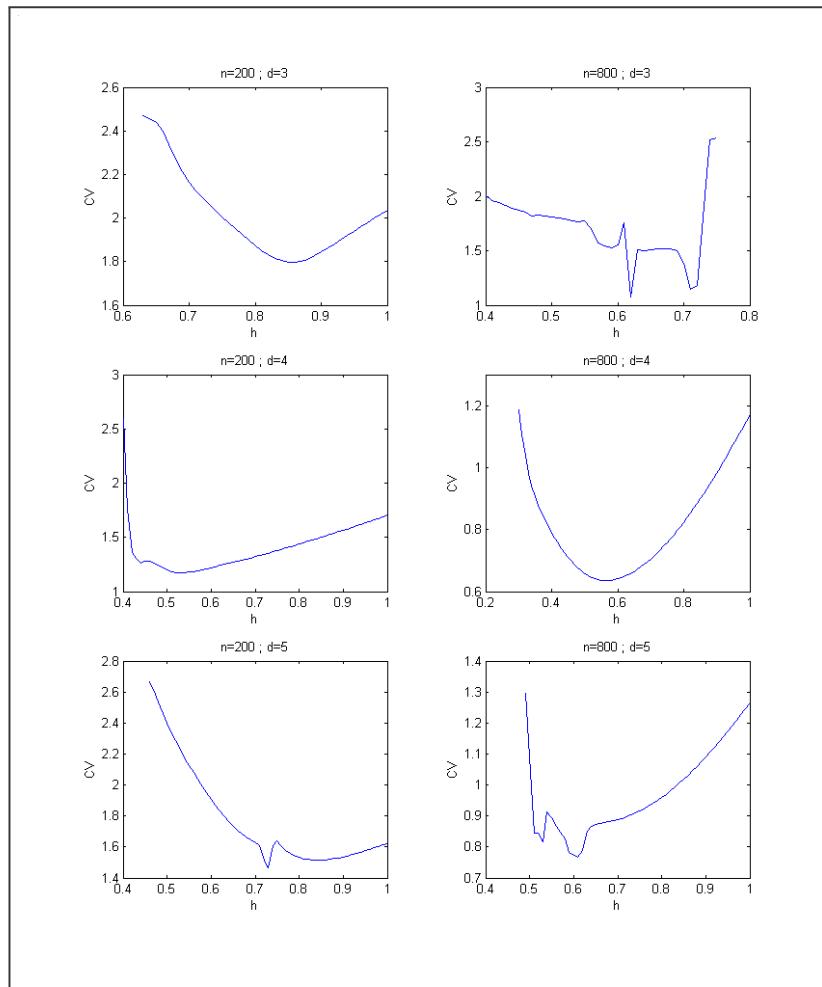


FIG. 3.2 – Critères de validation croisée : $n \in \{200, 800\}$; $d \in \{3, 4, 5\}$

On peut constater sur la figure 3.1 que les valeurs les plus faibles du critère de validation croisée sont bien obtenues pour $d = 4$. De même, la figure 3.2 montre que ce critère d'erreur décroît naturellement avec n . On peut également noter que le critère de validation croisée est assez bruité dans cet exemple, d'autant plus si la dimension testée est différente de la dimension simulée.

Les résultats permettent de vérifier la validité de notre procédure sur un modèle de régression, et on a obtenu pour la fenêtre optimale :

$$\begin{cases} \hat{\alpha} = (2.0297, 5.1209)^T \\ \left\| \left(I_{10} - \hat{B}\hat{B}^T \right) B \right\| = 1.43 \end{cases}$$

3.5.2 Simulations sur données dépendantes

On a simulé $n \in \{200, 500, 800\}$ données issues du modèle suivant :

$$Y_t = \alpha^T Z_t + 2 \exp \left[-3 (X_t^T \beta_1)^2 - 2 (X_t^T \beta_2) \right] + 0.5 \varepsilon_t$$

où :

- $\alpha = (1, 2, 3)^T$
- $\beta_1 = \frac{1}{3} (-2, 1, 2)^T$
- $\beta_2 = \frac{1}{\sqrt{2}} (1, 0, 1)^T$
- $Z_t = (Z_{t,1}, Z_{t,2}, Z_{t,3})^T$ où $(Z_{t,i})_{i \in \{1,2,3\}}$ sont des v.a.r. indépendantes de loi $\mathcal{B} \left(\frac{1}{2} \right)$.
- $X_t = (U_{t-1}, U_{t-2}, U_{t-3})^T$ où :

$$U_t = 0.4U_{t-1} - 0.5U_{t-2} + V_t$$

où V est un bruit blanc fort de loi $\mathcal{U}([-1, 1])$.

- Z_t, X_t et ε_t sont des v.a. indépendantes.

On a estimé $(\alpha, \beta_1, \beta_2)$ à l'aide de PLMAVE. On présente les résultats obtenus avec $n \in \{200, 500, 800\}$ et $d \in \{3, 4, 5\}$.

Les figures 3.3 et 3.4 conduisent aux mêmes conclusions que précédemment. Notons seulement que le critère de validation croisée n'est pas bruité dans cet exemple.

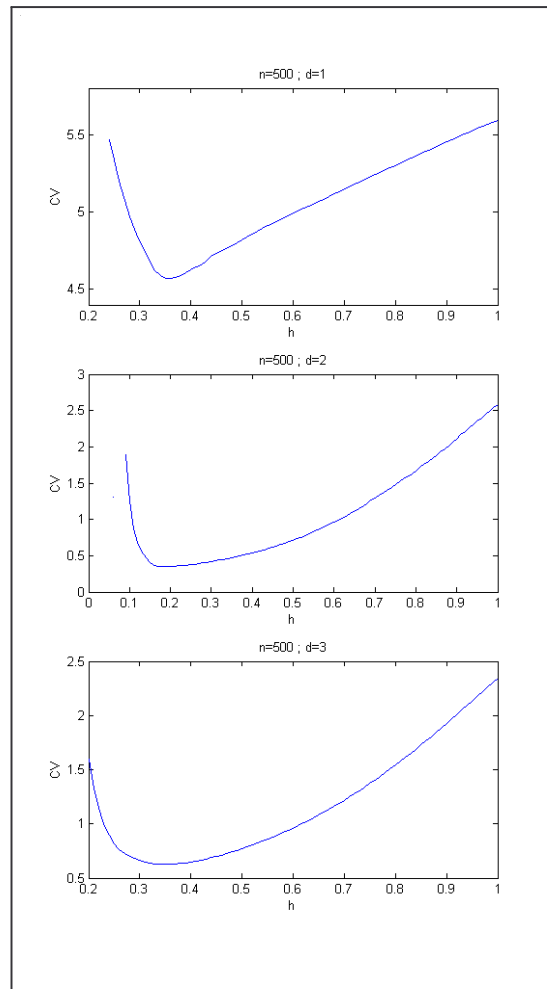
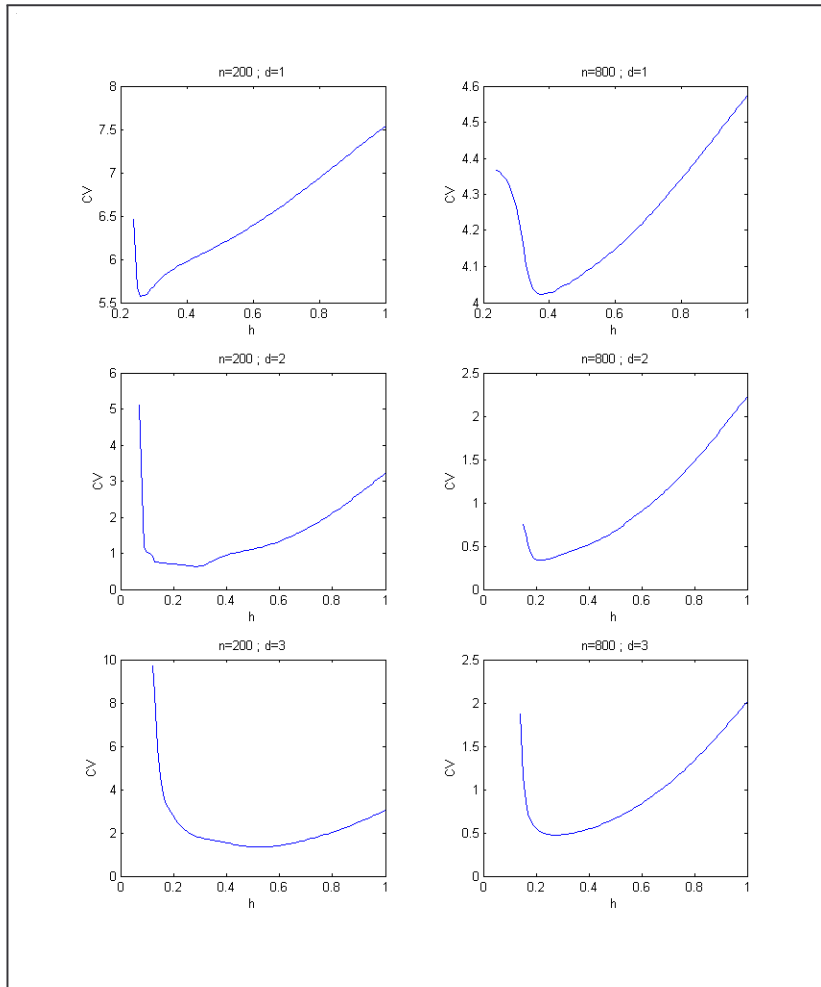


FIG. 3.3 – Critères de validation croisée : $n = 500$; $d \in \{1, 2, 3\}$

FIG. 3.4 – Critères de validation croisée : $n \in \{200, 800\}$; $d \in \{1, 2, 3\}$

Les résultats permettent encore de vérifier la validité de notre procédure sur un modèle de séries temporelles, et on a obtenu pour la fenêtre optimale :

$$\begin{cases} \widehat{\alpha} = (2.03, 5.12)^T \\ \widehat{\beta}_1 - \beta_1 = (0.15, 0.02, 0.11)^T \\ \widehat{\beta}_2 - \beta_2 = (-0.09, 0.13, 0.12)^T \\ \left\| (I_3 - \widehat{B}\widehat{B}^T) B \right\| = 0.074 \end{cases}$$

3.6 Prédiction à l'aide de la méthode PLMAVE

On désire prévoir X_{T+l} avec $l \in \mathbb{N}^*$ à partir de $(Y_t, Z_t)_{t \in \{1, \dots, T\}}$, en supposant le modèle suivant vérifié :

$$X_{t+l} = \alpha_l^T Z_{t+l} + g_l \left(B_l^T X_t^{(r)} \right) + \varepsilon_t \quad (3.10)$$

où Z_{t+l} contient les variables exogènes. On suppose évidemment qu'on est en mesure de disposer de prévisions convenables des variables exogènes (à partir d'avis d'experts ou de modèles).

On est en mesure d'estimer α_l et B_l avec la méthode PLMAVE à partir de $(X_t, Z_t)_{t \in \{1, \dots, T\}}$. Disposant ainsi de $U_t = X_t - \widehat{\alpha}_l^T Z_t$ pour $t \in \{1, \dots, T\}$, ainsi que de \widehat{B}_l , on peut écrire :

$$\forall t \in \{r, \dots, T-l\} : U_{t+l} = g_l \left(B_l^T X_t^{(r)} \right) + \varepsilon'_t$$

On peut donc prévoir U_{T+l} à l'aide de la méthode des noyaux, et obtenir la prévision de X_{T+l} à l'aide de la relation suivante :

$$\widehat{X}_T(l) = \widehat{U}_T(l) + \widehat{\alpha}_l^T \widehat{Z}_{T+l}$$

où \widehat{Z}_{T+l} est connu a priori.

Dans le cas de la consommation d'électricité, on considère notamment comme variables exogènes des variables climatiques pour lesquelles Météo-France fournit des prévisions.

Il est également possible d'élaborer des intervalles de confiance en se basant sur les erreurs de prévision a posteriori qu'on note $(e_{t(l)})_t$. L'intervalle de confiance se détermine ainsi à partir des quantiles empiriques des erreurs passées. On peut se référer à Carbon *et al.* (1993) pour plus de détails.

3.7 Application à la consommation d'électricité

Nous avons appliqué la méthode que nous avons nommé PLMAVE à la consommation d'électricité. On a considéré les variables suivantes :

- $Y = P_t$
- $X = (P_{t-1}, \dots, P_{t-r})^T$

Quant à Z qui contient les variables exogènes, on a considéré les quatre variables météorologiques suivantes :

- la différence entre la température brute et la température seuil utilisée habituellement dans les modèles de RTE (Cf. Chapitre 1) ;
- la différence entre une première température lissée et la température seuil ;
- la différence entre une seconde température lissée et la température seuil ;
- la nébulosité.

Les données météorologiques utilisées sont des pondérations de données mesurées sur 26 stations françaises, et représentatives pour la consommation d'électricité.

Notons que les températures seuil sont horaires et que les lissages utilisés sont ceux utilisés par RTE. Ces variables météorologiques "construites" n'ont pas fait l'objet d'une étude particulière dans cette application. Nous pouvons néanmoins citer l'existence de modèles incluant ces problèmes de seuil ; on pourra notamment se référer à Chen (1993). Au vu de la complexité de notre problème, il n'était pas raisonnable d'ajouter cet aspect dans le modèle.

Le fait d'avoir de forts impacts horaires nous a conduit à considérer une partie linéaire dépendante de l'heure. Nous n'avons pas effectué de différenciation mais avons considéré une mémoire de taille 24 afin de bien prendre en compte l'importance de la saisonnalité journalière.

On a de nouveau sélectionné les deux périodes considérées dans le second chapitre :

- Période 1 : du 3 avril 2006 au 9 juillet 2006
- Période 2 : du 15 janvier 2007 au 8 avril 2007

On constate que la méthode PLMAVE fournit des résultats moins bons que le modèle actuellement en place pour les deux périodes considérées. Le MAPE vaut ainsi 1.2% pour PLMAVE contre 0.9% pour le modèle RTE, sur la première période. Sur la seconde période, ce même MAPE vaut 1.8% pour PLMAVE contre 1.2% pour RTE. Nous pouvons remarquer que la prise en compte de variables exogènes (météorologiques) est totalement justifiée sur le second jeu de données (fin de l'hiver 2006-2007), mais moins sur le premier jeu de données (printemps 2006). Le modèle PLMAVE fournit néanmoins des résultats acceptables sur ces deux périodes.

Il est certain que ces résultats sont loins d'être mauvais, et une étude approfondie des variables exogènes pourrait conduire très rapidement à l'amélioration des résultats. Il est également à noter que le modèle PLMAVE est plus simple que le modèle défini dans le premier chapitre (régression et SARIMA), même si la procédure d'estimation est plus coûteuse en temps de calcul. Le modèle actuellement en place a fait l'objet de très nombreuses études et d'améliorations successives, ce que nous n'avons pas encore pu réaliser en pratique pour ce nouveau modèle. Gageons que l'avenir le rendra compétitif, et permettra de définir de manière plus aisée des intervalles de confiance.

	PLMAVE	RTE
Moyenne	294	273
Ecart-type	686	482
Minimum	-2096	-1043
Quantile 5%	-844	-541
Médiane	300	249
Quantile 95%	1433	1071
Maximum	2033	2231

TAB. 3.1 – Erreurs de prévision : Période 1

	PLMAVE	RTE
MAPE (%)	0.9	1.2
RMSE	554	738

TAB. 3.2 – Critères d'erreurs de prévision : Période 1

	PLMAVE	RTE
Moyenne	142	148
Ecart-type	1351	908
Minimum	-4182	-2657
Quantile 5%	-2028	-1534
Médiane	183	287
Quantile 95%	2281	1379
Maximum	4302	2789

TAB. 3.3 – Erreurs de prévision : Période 2

	PLMAVE	RTE
MAPE (%)	1.8	1.2
RMSE	1356	919

TAB. 3.4 – Critères d'erreurs de prévision : Période 2

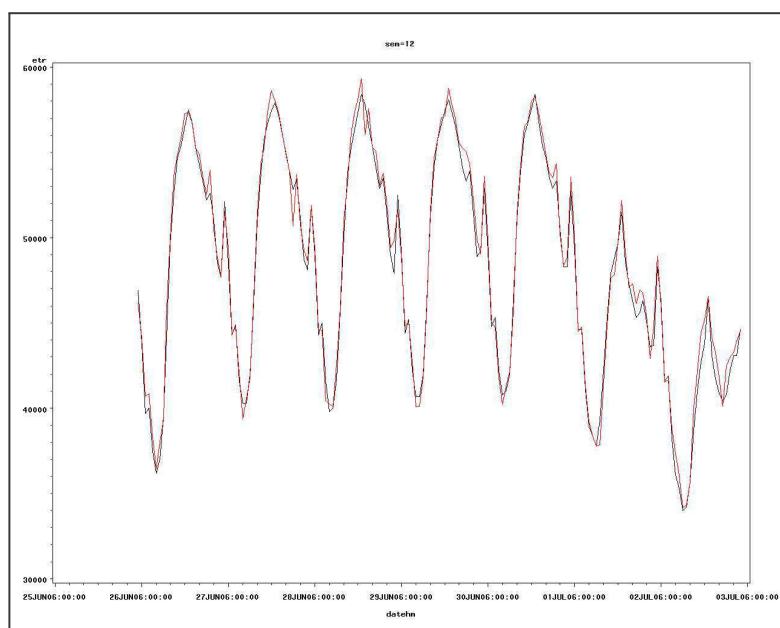


FIG. 3.5 – Consommations prévues PLMAVE (rouge) et réalisées (noir) : Période 1, première semaine tronquée

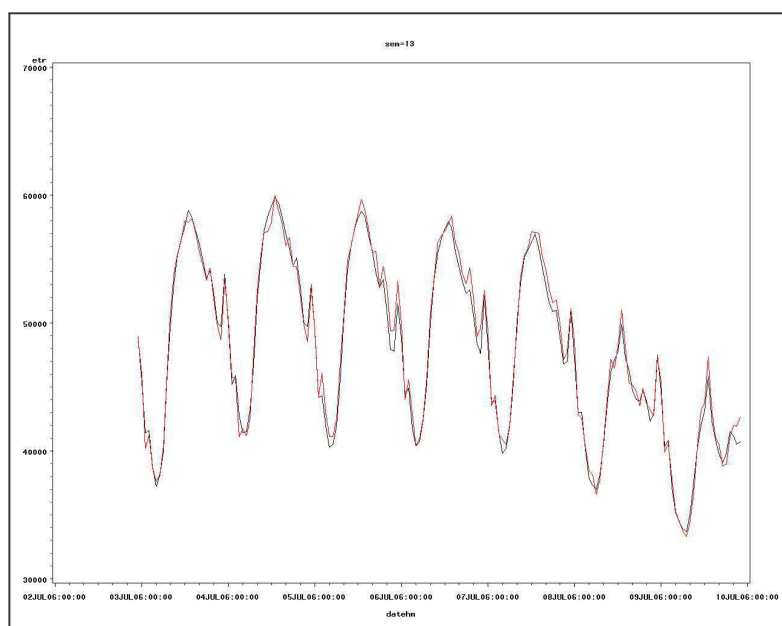


FIG. 3.6 – Consommations prévues PLMAVE (rouge) et réalisées (noir) : Période 1, seconde semaine tronquée

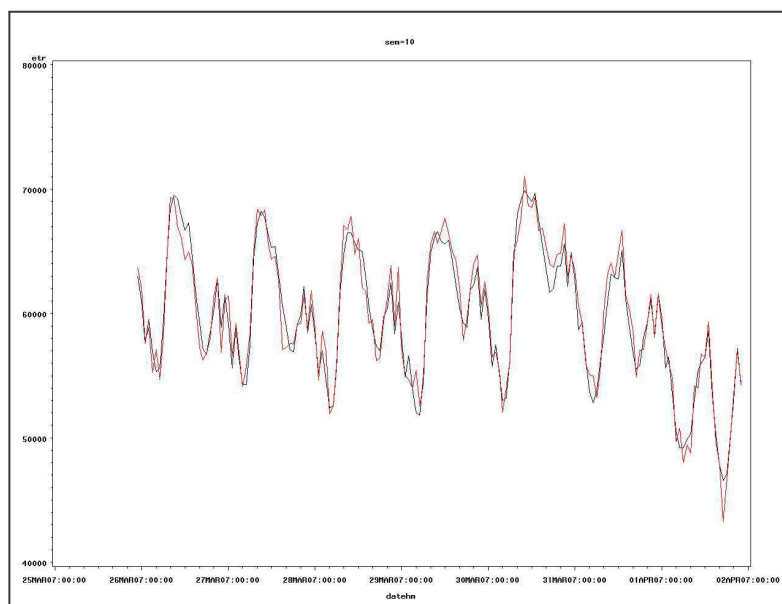


FIG. 3.7 – Consommations prévues PLMAVE (rouge) et réalisées (noir) : Période 2, première semaine tronquée

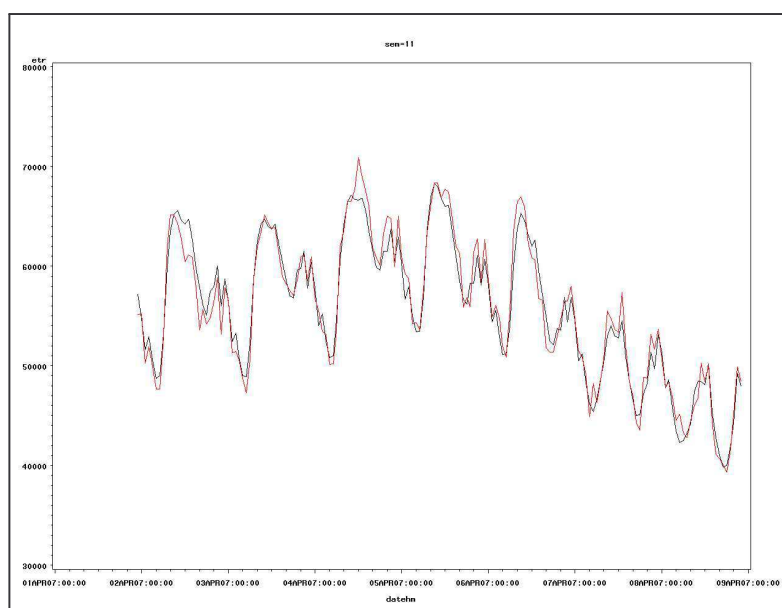


FIG. 3.8 – Consommations prévues PLMAVE (rouge) et réalisées (noir) : Période 2, seconde semaine tronquée

Chapitre 4

Annexe : De l'ergodicité de modèles autorégressifs

4.1 Introduction

On utilise abondamment les modèles ARMA de Box et Jenkins en pratique. Ces modèles nécessitent de travailler sur une série stationnaire afin d'être en mesure de pouvoir estimer correctement les polynômes autorégressif et moyenne mobile. Malheureusement, les séries temporelles réelles sont rarement modélisables par des processus stationnaires et l'alternative proposée par les auteurs consiste en la différenciation de la série, en "tendance" et/ou en "saisonnalité". Notons qu'on n'a aucune certitude de stationnariser une série à l'aide de la différenciation, ni même avec la décomposition saisonnière, même si cela fonctionne dans de nombreux cas.

Si on considère une série "stationnaire", on sait qu'il est toujours possible de la modéliser par un modèle ARMA. Supposons qu'on a un processus $(X_t)_{t \in \mathbb{Z}}$ ARMA (AutoRegressive Moving Average) d'ordre (p, q) , noté $ARMA(p, q)$:

$$\forall t \in \mathbb{Z} : X_t - \sum_{i=1}^p \varphi_i X_{t-i} = \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i}$$

où $(\varphi_1, \dots, \varphi_p) \in \mathbb{R}^p$, $\varphi_p \neq 0$, $(\theta_1, \dots, \theta_q) \in \mathbb{R}^q$, $\theta_q \neq 0$ et $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc faible de variance σ^2 .

Si on réécrit ce processus en posant :

$$AR(B) = I - \sum_{i=1}^p \varphi_i B^i$$

$$MA(B) = I + \sum_{i=1}^q \theta_i B^i$$

où I est l'opérateur identité et B l'opérateur retard, on obtient :

$$\forall t \in \mathbb{Z} : AR(B) X_t = MA(B) \varepsilon_t$$

Notons qu'on a transformé les notations classiques Φ et Θ , respectivement en AR et MA , afin d'éviter des confusions avec ce qui suit.

On sait que le processus $(X_t)_{t \in \mathbb{Z}}$ est stationnaire si et seulement si le polynôme AR a toutes ses racines de module différent de 1 (en dehors du cercle unité).

Les modèles ARMA sont sûrement les modèles autorégressifs les plus utilisés mais des développements méthodologiques assez récents nous incitent à considérer des modèles alternatifs au paramétrique. L'ouvrage de Tong (1990) en dresse une liste non exhaustive. On est ainsi conduit à considérer des modèles autorégressifs :

– *non-paramétrique*, par exemple :

$$\forall t \in \mathbb{Z} : X_t = \Phi(X_{t-1}, \dots, X_{t-r}) + \varepsilon_t$$

où Φ est une fonction inconnue.

– *semi-paramétrique*, par exemple :

$$\forall t \in \mathbb{Z} : X_t = g\left(B^T(X_{t-1}, \dots, X_{t-r})^T\right) + \varepsilon_t$$

où g est une fonction inconnue et B une matrice inconnue (de dimension adéquate).

Il est possible de déterminer des conditions sur Φ et g afin que les modèles ci-dessus soient stationnaires. On va utiliser pour cela la notion d'ergodicité géométrique¹.

Pham (1986) a montré que si un processus est géométriquement ergodique alors il est non seulement strictement stationnaire, mais également géométriquement mélangeant. Notons que ce résultat n'est pas valable dans le cas de l'ergodicité simple, d'où l'intérêt majeur de l'ergodicité géométrique.

¹On redonne les définitions de l'ergodicité et de mélangeance dans la suite.

Lors de l'étude de modèles autorégressifs non linéaires, on pose très souvent des conditions de mélangeance du processus, hypothèse qu'il est difficile de vérifier. La définition même de la fonction de lien peut nous amener à rejeter ou non cette condition.

Les résultats énoncés ici posent des conditions suffisantes pour considérer l'ergodicité géométrique des modèles autorégressifs non linéaires, non et semi-paramétriques. Ces conditions impliquent la stationnarité et la mélangeance géométrique des processus considérés.

Afin d'étudier l'ergodicité géométrique des modèles autorégressifs, il est utile de réécrire des modèles sous la forme de chaînes de Markov qui sont des outils puissants pour établir de telles propriétés. Notons que les travaux de Tweedy (1975) constituent la contribution majeure dans l'étude de l'ergodicité.

4.2 Cas d'un modèle non-paramétrique

On considère tout d'abord le modèle non-paramétrique suivant :

$$\forall t \in \mathbb{N}^* : X_t = \Phi(X_{t-1}, \dots, X_{t-r}) + \varepsilon_t \quad (4.1)$$

où :

- $(\varepsilon_t)_{t \in \mathbb{N}^*}$ est une suite de v.a.r. i.i.d. dans $\mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ admettant une fonction de densité.
- $\forall t \in \mathbb{N}^* : \varepsilon_t$ est indépendant de $\{X_{t-s}, s \in \mathbb{N}^* / t - s \geq -r + 1\}$
- $\Phi : \mathbb{R}^r \rightarrow \mathbb{R}$ est une fonction mesurable

On initialise le processus par (X_0, \dots, X_{-r+1}) .

Le modèle (4.1) est un modèle autorégressif non-paramétrique très classique, notamment étudié par Carbon *et al.* (1993) qui l'ont estimé à l'aide de la méthode des noyaux.

On peut réécrire le modèle (4.1) sous la forme suivante :

$$\begin{cases} \underline{X}_0 = (X_0, \dots, X_{-r+1}) \\ \forall t \in \mathbb{N}^* : \underline{X}_t = \underline{\Phi}(\underline{X}_{t-1}) + \underline{\varepsilon}_t \end{cases} \quad (4.2)$$

où :

- $\underline{X}_t = (X_t, \dots, X_{t-r+1})^T \in \mathbb{R}^r$
- $\underline{\Phi}(\underline{X}_t) = (\Phi(\underline{X}_t), X_t, \dots, X_{t-r+2})^T \in \mathbb{R}^r$
- $\underline{\varepsilon}_t = (\varepsilon_t, 0, \dots, 0)^T \in \mathbb{R}^r$

$(\underline{X}_t)_{t \in \mathbb{N}}$ est une chaîne de Markov homogène dont l'espace d'états est $(\mathbb{R}^r, \mathcal{B}(\mathbb{R}^r))$.

On a en effet :

$$\forall A \in \mathcal{B}(\mathbb{R}^r), \forall x \in \mathbb{R}^r, \forall (t, s) \in \mathbb{N}^2 : \mathbb{P}(\underline{X}_t \in A / \underline{X}_{t-1} = x) = \mathbb{P}(\underline{X}_s \in A / \underline{X}_{s-1} = x)$$

La probabilité de transition π de $(\underline{X}_t)_{t \in \mathbb{N}}$ vaut :

$$\forall A \in \mathcal{B}(\mathbb{R}^r), \forall x \in \mathbb{R}^r : \pi(x, A) = \mathbb{P}(\underline{X}_t \in A / \underline{X}_{t-1} = x)$$

Il existe de nombreux résultats sur l'ergodicité géométrique des processus unidimensionnels non-linéaires autorégressifs ; citons notamment les travaux de An *et al.* (1996), An *et al.* (1997) et Tong (1990).

Si la chaîne de Markov $(\underline{X}_t)_{t \in \mathbb{N}}$ définie par (4.2) est géométriquement ergodique alors le processus $(X_t)_{t \in \mathbb{N}}$ est lui-même géométriquement ergodique.

On donne ici quelques définitions de base utiles à notre propos.

Définition 4.1

On dit que le processus $(X_t)_{t \in \mathbb{N}}$ est fortement mélangeant (α -mélangeant) si :

$$\alpha(k) = \sup_{A,B} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \xrightarrow{k \rightarrow +\infty} 0$$

où $A \in \sigma(X_s, s \leq t)$ et $B \in \sigma(X_s, s \geq t+k)$, $k \in \mathbb{N}$.

On dit que le processus $(X_t)_{t \in \mathbb{N}}$ est géométriquement mélangeant s'il existe $\rho \in]0, 1[$ et $a \in \mathbb{R}^+$ tels que :

$$\alpha(k) = \sup_{A,B} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \leq a\rho^k$$

Les définitions suivantes portent sur les chaînes de Markov.

Définition 4.2

On dit que la chaîne de Markov $(\underline{X}_t)_{t \in \mathbb{N}}$ définie par (4.2) est μ -irréductible, où μ désigne la mesure de Lebesgue, si pour tout $A \in \mathcal{B}(\mathbb{R}^r)$ tel que $\mu(A) > 0$, pour tout $x \in \mathbb{R}^r$, il existe $t_0 \in \mathbb{N}^*$ tel que :

$$\pi^{t_0}(x, A) > 0$$

Définition 4.3

Soit la chaîne de Markov $(\underline{X}_t)_{t \in \mathbb{N}}$ définie par (4.2).

On appelle cycle de longueur k un k -uplet de boréliens de \mathbb{R}^r (A_0, \dots, A_{k-1}) disjoints tels que :

$$\begin{cases} \forall i \in \{0, \dots, k-2\}, \forall x \in A_i : \pi(x, A_{i+1}^C) = 0 \\ \forall x \in A_{k-1} : \pi(x, A_0^C) = 0 \end{cases}$$

Soit d le plus grand commun diviseur des longueurs k .

La chaîne de Markov $(\underline{X}_t)_{t \in \mathbb{N}}$ est alors dite périodique de période d .

Si $d = 1$, la chaîne de Markov $(\underline{X}_t)_{t \in \mathbb{N}}$ est alors dite apériodique.

Définition 4.4

Soit la chaîne de Markov $(\underline{X}_t)_{t \in \mathbb{N}}$ définie par (4.2).

On dit que le borélien B de \mathbb{R}^r est petit pour π s'il existe une mesure ν strictement positive et $t_0 \in \mathbb{N}^*$ tels que :

$$\forall x \in B, \forall A \in \mathcal{B}(\mathbb{R}^r) : \pi^{t_0}(x, A) > \nu(A)$$

L'ergodicité géométrique qui occupe notre attention est définie comme suit.

Définition 4.5

On dit que la chaîne de Markov $(\underline{X}_t)_{t \in \mathbb{N}}$ définie par (4.2) est géométriquement ergodique s'il existe une probabilité π_0 et $\rho \in]0, 1[$ tels que :

$$\forall x \in \mathbb{R}^r : \rho^{-t} \left\| \pi^t(x, \cdot) - \pi_0(\cdot) \right\|_{VT} \xrightarrow{t \rightarrow +\infty} 0 \quad (4.3)$$

où $\|\cdot\|_{VT}$ est la norme de variation totale.

π_0 est appelée probabilité invariante.

Remarque 4.1

Pour une mesure μ définie sur un espace probabilisé (Ω, \mathcal{A}) , on définit pour $A \in \mathcal{A}$: $\mu^+ = \sup_{A \in \mathcal{A}} \mu(A)$ et $\mu^- = -\inf_{A \in \mathcal{A}} \mu(A)$. La norme de variation totale vaut :

$$\|\mu\|_{VT} = \mu^+ + \mu^-$$

Remarque 4.2

Si la formule (4.3) est valable pour $\rho = 1$, la chaîne de Markov $(\underline{X}_t)_{t \in \mathbb{N}}$ définie par (4.2) est dite ergodique (au sens de Harris).

Il existe différents résultats sur l'ergodicité géométrique du modèle (4.1), obtenus à partir des lemmes suivants :

Lemme 4.1 Chan et al. (1985)

Supposons que la fonction Φ du modèle (4.1) est bornée sur tout ensemble borné.

La chaîne de Markov $(\underline{X}_t)_{t \in \mathbb{N}}$ définie par (4.2) est alors apériodique et μ -irréductible.

Lemme 4.2 Tong (1990)

Soit une chaîne de Markov $(\underline{X}_t)_{t \in \mathbb{N}}$ définie par (4.2), apériodique et μ -irréductible.

Supposons que la fonction Φ du modèle (4.1) est bornée sur tout ensemble borné.

Supposons également qu'il existe une fonction f positive, $q \in \mathbb{N}^*$, $(c_1, c_2) \in \mathbb{R}^{+*2}$,

$\rho \in]0, 1[$ et un petit espace C tels que :

$$- \forall x \notin C : \mathbb{E} [f(\underline{X}_{t+1}) / \underline{X}_t = x] \leq \rho f(x) - c_1$$

$$- \forall x \in C : \mathbb{E} [f(\underline{X}_{t+1}) / \underline{X}_t = x] \leq c_2$$

La chaîne de Markov $(\underline{X}_t)_{t \in \mathbb{N}}$ est alors géométriquement ergodique.

Le lemme suivant, connu sous le nom du "h-step criteria", permet d'obtenir des résultats d'ergodicité d'une chaîne de Markov, à partir de ceux établis sur une sous-suite de cette chaîne.

Lemme 4.3 *Tjøstheim (1990)*

Soit $(\underline{X}_t)_{t \in \mathbb{N}}$ une chaîne de Markov.

S'il existe $h \in \mathbb{N}^*$ tel que $(\underline{X}_{ht})_{t \in \mathbb{N}}$ est géométriquement ergodique alors $(\underline{X}_t)_{t \in \mathbb{N}}$ est géométriquement ergodique.

Lemme 4.4

Soit $(\underline{X}_t)_{t \in \mathbb{N}}$ une chaîne de Markov.

S'il existe $h \in \mathbb{N}^*$ tel que $(\underline{X}_{ht})_{t \in \mathbb{N}}$ vérifie les hypothèses du lemme 4.2 alors $(\underline{X}_t)_{t \in \mathbb{N}}$, et donc $(X_t)_{t \in \mathbb{N}}$, est géométriquement ergodique.

Démonstration

Ce résultat provient directement des lemmes 4.2 et 4.3. \diamond

Les deux théorèmes suivants nous permettent d'établir une condition suffisante pour l'ergodicité du modèle (4.1), ainsi que du modèle semi-paramétrique considéré par la suite.

Théorème 4.1 *An et al. (1996)*

Soit une chaîne de Markov $(\underline{X}_t)_{t \in \mathbb{N}}$ définie par (4.2).

Supposons que la fonction Φ du modèle (4.1) vérifie :

$$|\Phi(x_1, \dots, x_r)| \leq \lambda \max\{|x_1|, \dots, |x_r|\} + c$$

où $\lambda \in]0, 1[$ et $c \in \mathbb{R}$.

Le processus $(X_t)_{t \in \mathbb{N}}$ est alors géométriquement ergodique.

On fournit ici la démonstration de An et al. (1996) car ce schéma de preuve est utilisé dans la suite. De plus, la démonstration donnée par les auteurs est incomplète, alors que les arguments finaux sont majeurs.

Démonstration

On utilise la fonction suivante :

$$f(x_1, \dots, x_r) = \max(|x_1|, \dots, |x_r|)$$

On a :

$$\begin{aligned}
|X_1| &= |\Phi(X_0, \dots, X_{-r+1}) + \varepsilon_1| \\
&\leq |\Phi(X_0, \dots, X_{-r+1})| + |\varepsilon_1| \\
&\leq \lambda \max\{|X_0|, \dots, |X_{-r+1}|\} + c + |\varepsilon_1| \\
&= \lambda f(\underline{X}_0) + c + |\varepsilon_1| \\
|X_2| &= |\Phi(X_1, \dots, X_{-r+2}) + \varepsilon_2| \\
&\leq |\Phi(X_1, \dots, X_{-r+2})| + |\varepsilon_2| \\
&\leq \lambda \max\{|X_1|, |X_0|, \dots, |X_{-r+2}|\} + c + |\varepsilon_2| \\
&\leq \lambda \max\{\lambda \max\{|X_0|, \dots, |X_{-r+1}|\} + c + |\varepsilon_1|, |X_0|, \dots, |X_{-r+2}|\} + c + |\varepsilon_2| \\
&= \lambda^2 \max\{\max\{|X_0|, \dots, |X_{-r+1}|\}, |X_0|, \dots, |X_{-r+2}|\} + c + |\varepsilon_1| + c + |\varepsilon_2| \\
&\leq \lambda \max\{|X_0|, \dots, |X_{-r+1}|\} + c + |\varepsilon_1| + c + |\varepsilon_2| \\
&= \lambda f(\underline{X}_0) + 2c + |\varepsilon_1| + |\varepsilon_2|
\end{aligned}$$

On peut montrer par récurrence que :

$$|X_r| \leq \lambda f(\underline{X}_0) + rc + |\varepsilon_1| + \dots + |\varepsilon_r|$$

On en déduit que :

$$\begin{aligned}
f(\underline{X}_r) &= \max\{|X_r|, \dots, |X_1|\} \\
&\leq \lambda f(\underline{X}_0) + rc + |\varepsilon_1| + \dots + |\varepsilon_r|
\end{aligned}$$

D'où :

$$\begin{aligned}
\forall t \in \mathbb{N}^* : \mathbb{E}[f(\underline{X}_{rt}) / \underline{X}_{r(t-1)} = x] &= \mathbb{E}[f(\underline{X}_r) / \underline{X}_0 = x] \\
&\leq \mathbb{E}[\lambda f(\underline{X}_0) + rc + |\varepsilon_1| + \dots + |\varepsilon_r| / \underline{X}_0 = x] \\
&= \lambda \mathbb{E}[f(\underline{X}_0) / \underline{X}_0 = x] + rc \\
&\quad + \mathbb{E}[|\varepsilon_1| / \underline{X}_0 = x] + \dots + \mathbb{E}[|\varepsilon_r| / \underline{X}_0 = x] \\
&\leq \lambda f(x) + rc + \mathbb{E}[|\varepsilon_1|] + \dots + \mathbb{E}[|\varepsilon_r|] \\
&= \lambda f(x) + c'
\end{aligned}$$

où $c' = rc + \mathbb{E}[|\varepsilon_1|] + \dots + \mathbb{E}[|\varepsilon_r|]$.

– Soit ρ tel que $0 < \lambda < \rho < 1$.

On a :

$$\mathbb{E}[f(\underline{X}_r) / \underline{X}_0 = x] \leq \rho f(x) - [(\rho - \lambda) f(x) - c']$$

Etant donné que f est une norme sur \mathbb{R}^r , il existe $\delta \in \mathbb{R}^{+*}$ tel que :

$$\forall x \in \mathbb{R}^r : f(x) \geq \delta \|x\|$$

où $\|x\|$ désigne la norme euclidienne sur \mathbb{R}^r .

Les normes sont en effet équivalentes sur l'espace de dimension finie \mathbb{R}^r .

Soit $M \in \mathbb{R}^{+*}$ tel que $(\rho - \lambda)M - c' \geq c_1 > 0$.

Considérons le petit ensemble $C = \left\{ x \in \mathbb{R}^r \mid \|x\| < \frac{M}{\delta} \right\}$.

On a :

$$\begin{aligned} \forall x \notin C : \mathbb{E}[f(\underline{X}_r) / \underline{X}_0 = x] &\leq \rho f(x) - [(\rho - \lambda)f(x) - c'] \\ &\leq \rho f(x) - [(\rho - \lambda)\delta \|x\| - c'] \\ &\leq \rho f(x) - [(\rho - \lambda)M - c'] \\ &\leq \rho f(x) - c_1 \end{aligned}$$

– f étant bornée sur C , il existe donc $c_2 \in \mathbb{R}^{+*}$ tel que :

$$\begin{aligned} \forall x \in C : \mathbb{E}[f(\underline{X}_r) / \underline{X}_0 = x] &\leq \lambda f(x) + c' \\ &\leq c_2 \end{aligned}$$

L'ergodicité géométrique est ainsi démontrée à l'aide du lemme 4.4 avec $h = r$. \diamond

Théorème 4.2

Soit une chaîne de Markov $(\underline{X}_t)_{t \in \mathbb{N}}$ définie par (4.2).

Supposons que la fonction Φ du modèle (4.1) vérifie :

$$|\Phi(x_1, \dots, x_r)| \leq \varphi_1 |x_1| + \dots + \varphi_r |x_r| + c$$

où $(\varphi_i)_{i \in \{1, \dots, r\}} \in \mathbb{R}^{+r}$ est telle que $\sum_{i=1}^r \varphi_i < 1$, et $c \in \mathbb{R}$.

Le processus $(X_t)_{t \in \mathbb{N}}$ est alors géométriquement ergodique.

Démonstration

Il n'est pas nécessaire d'utiliser ici le critère de Tjøstheim (1990).

On utilise la fonction suivante :

$$f(x_1, \dots, x_r) = \beta_1 |x_1| + \dots + \beta_r |x_r|$$

où $(\beta_i)_{i \in \{1, \dots, r\}} \in \mathbb{R}^{+r}$.

On précise les conditions supplémentaires que doivent vérifier les $(\beta_i)_{i \in \{1, \dots, r\}}$ dans la suite.

On a :

$$\forall t \in \mathbb{N}^* : \mathbb{E}[f(\underline{X}_t) / \underline{X}_{t-1} = x] = \mathbb{E}[f(\underline{X}_1) / \underline{X}_0 = x]$$

$$\begin{aligned}
f(\underline{X}_1) &= \beta_1 |X_1| + \beta_2 |X_0| + \dots + \beta_r |X_{-r+2}| \\
&= \beta_1 |\Phi(X_0, \dots, X_{-r+1}) + \varepsilon_1| + \beta_2 |X_0| + \dots + \beta_r |X_{-r+2}| \\
&\leq \beta_1 |\Phi(X_0, \dots, X_{-r+1})| + |\varepsilon_1| + \beta_2 |X_0| + \dots + \beta_r |X_{-r+2}| \\
&\leq \beta_1 (\varphi_1 |X_0| + \dots + \varphi_r |X_{-r+1}| + c + |\varepsilon_1|) + \beta_2 |X_0| + \dots + \beta_r |X_{-r+2}| \\
&= (\beta_1 \varphi_1 + \beta_2) |X_0| + \dots + (\beta_1 \varphi_{r-1} + \beta_r) |X_{-r+2}| + \beta_1 \varphi_r |X_{-r+1}| + \beta_1 c + \beta_1 |\varepsilon_1| \\
&= \frac{\beta_1 \varphi_1 + \beta_2}{\beta_1} \beta_1 |X_0| + \dots + \frac{\beta_1 \varphi_{r-1} + \beta_r}{\beta_{r-1}} \beta_{r-1} |X_{-r+2}| + \frac{\beta_1 \varphi_r}{\beta_r} \beta_r |X_{-r+1}| + \beta_1 c + \beta_1 |\varepsilon_1|
\end{aligned}$$

Sous la condition $\sum_{i=1}^r \varphi_i < 1$, on peut contraindre les $(\beta_i)_{i \in \{1, \dots, r\}}$ de telle sorte que :

$$\begin{cases} \forall i \in \{1, \dots, r-1\} : 0 < \frac{\beta_1 \varphi_i + \beta_{i+1}}{\beta_i} < 1 \\ 0 < \frac{\beta_1 \varphi_r}{\beta_r} < 1 \end{cases}$$

Sous cette contrainte, on considère :

$$\lambda = \max \left(\frac{\beta_1 \varphi_1 + \beta_2}{\beta_1}, \dots, \frac{\beta_1 \varphi_{r-1} + \beta_r}{\beta_{r-1}}, \frac{\beta_1 \varphi_r}{\beta_r} \right) \in]0, 1[$$

On a donc :

$$\begin{aligned}
f(\underline{X}_1) &\leq \lambda (\beta_1 |X_0| + \dots + \beta_r |X_{-r+1}|) + \beta_1 c + \beta_1 |\varepsilon_1| \\
&= \lambda f(\underline{X}_0) + \beta_1 c + \beta_1 |\varepsilon_1|
\end{aligned}$$

D'où :

$$\begin{aligned}
\mathbb{E}[f(\underline{X}_1) / \underline{X}_0 = x] &\leq \lambda f(x) + \beta_1 c + \beta_1 \mathbb{E}[|\varepsilon_1|] \\
&= \lambda f(x) + c'
\end{aligned}$$

où $c' = \beta_1 c + \beta_1 \mathbb{E}[|\varepsilon_1|]$.

La fin de la démonstration est analogue à celle du théorème 4.1, sachant que f est encore une norme sur \mathbb{R}^r . \diamond

Remarque 4.3

Le théorème 4.2 implique le théorème 4.1.

4.3 Cas d'un modèle semi-paramétrique

On considère maintenant un modèle autorégressif semi-paramétrique (à directions révélatrices multiples) :

$$\forall t \in \mathbb{N}^* : X_t = g \left(B^T (X_{t-1}, \dots, X_{t-r})^T \right) + \varepsilon_t \quad (4.4)$$

où :

- $(\varepsilon_t)_{t \in \mathbb{N}^*}$ est une suite de v.a.r. i.i.d. dans $\mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ admettant une fonction de densité.
- $\forall t \in \mathbb{N}^* : \varepsilon_t$ est indépendant de $\{X_{t-s}, s \in \mathbb{N}^* / t - s \geq -r + 1\}$
- $g : \mathbb{R}^D \rightarrow \mathbb{R}$ est une fonction mesurable
- $B = (\beta^1, \dots, \beta^D)_{r \times D}$, avec $1 \leq D < r$, est une matrice telle que $B^T B = I_D$

Remarque 4.4

La condition de normalisation sur la matrice B est celle choisie dans la méthode MAVE de Xia et al. (2002) que nous avons utilisée. Il n'est toutefois pas nécessaire d'avoir l'orthogonalité ici et on peut restreindre notre hypothèse en posant :

$$\forall j \in \{1, \dots, D\} : \|\beta^j\| = \sqrt{\frac{1}{r} \sum_{i=1}^r (\beta_i^j)^2} = 1$$

On énonce ci-après deux théorèmes qui posent des conditions suffisantes pour l'ergodicité géométrique du modèle (4.4).

Théorème 4.3

Soit $(X_t)_{t \in \mathbb{N}}$ un processus vérifiant le modèle (4.4).

Si la fonction g vérifie la condition suivante :

$$|g(x_1, \dots, x_D)| \leq \frac{\lambda}{\sqrt{r}} \max\{|x_1|, \dots, |x_D|\} + c \quad (4.5)$$

où $\lambda \in]0, 1[$ et $c \in \mathbb{R}$, alors le processus $(X_t)_{t \in \mathbb{N}}$ est géométriquement ergodique.

Démonstration

On a :

$$\begin{aligned}
& \left| g \left(B^T (x_{t-1}, \dots, x_{t-r})^T \right) \right| \\
&= \left| g \left(\beta_1^1 x_{t-1} + \dots + \beta_r^1 x_{t-r}, \dots, \beta_1^D x_{t-1} + \dots + \beta_r^D x_{t-r} \right) \right| \\
&\leq \frac{\lambda}{\sqrt{r}} \max \left(\left| \beta_1^1 x_{t-1} + \dots + \beta_r^1 x_{t-r} \right|, \dots, \left| \beta_1^D x_{t-1} + \dots + \beta_r^D x_{t-r} \right| \right) + c \text{ d'après (4.5)} \\
&\leq \frac{\lambda}{\sqrt{r}} \max \left(\left| \beta_1^1 x_{t-1} \right| + \dots + \left| \beta_r^1 x_{t-r} \right|, \dots, \left| \beta_1^D x_{t-1} \right| + \dots + \left| \beta_r^D x_{t-r} \right| \right) + c \\
&\leq \frac{\lambda}{\sqrt{r}} \max \left(\left| \beta_1^1 \right| + \dots + \left| \beta_r^1 \right|, \dots, \left| \beta_1^D \right| + \dots + \left| \beta_r^D \right| \right) \max \left(\left| x_{t-1} \right|, \dots, \left| x_{t-r} \right| \right) + c \\
&= \lambda \max \left(\frac{\left| \beta_1^1 \right| + \dots + \left| \beta_r^1 \right|}{\sqrt{r}}, \dots, \frac{\left| \beta_1^D \right| + \dots + \left| \beta_r^D \right|}{\sqrt{r}} \right) \max \left(\left| x_{t-1} \right|, \dots, \left| x_{t-r} \right| \right) + c \\
&\leq \lambda \max \left(\left| x_{t-1} \right|, \dots, \left| x_{t-r} \right| \right) + c
\end{aligned}$$

En effet, l'inégalité de Cauchy nous assure que :

$$\forall j \in \{1, \dots, D\} : \sum_{i=1}^r \left| \beta_i^j \right| \leq \sqrt{\sum_{i=1}^r 1^2} \sqrt{\sum_{i=1}^r (\beta_i^j)^2} = \sqrt{r}$$

L'ergodicité géométrique est ainsi démontrée à l'aide du théorème 4.1. \diamond

Théorème 4.4

Soit $(X_t)_{t \in \mathbb{N}}$ un processus vérifiant le modèle (4.4).

Si la fonction g vérifie la condition suivante :

$$\left| g(x_1, \dots, x_D) \right| \leq \frac{1}{\sqrt{r}} (\varphi_1 |x_1| + \dots + \varphi_D |x_D|) + c \quad (4.6)$$

où $(\varphi_i)_{i \in \{1, \dots, D\}} \in \mathbb{R}^{+D}$ est telle que $\sum_{i=1}^D \varphi_i < 1$, et $c \in \mathbb{R}$, alors le processus $(X_t)_{t \in \mathbb{N}}$ est géométriquement ergodique.

Démonstration

On a :

$$\begin{aligned}
& \left| g \left(B^T (x_{t-1}, \dots, x_{t-r})^T \right) \right| \\
&= \left| g \left(\beta_1^1 x_{t-1} + \dots + \beta_r^1 x_{t-r}, \dots, \beta_1^D x_{t-1} + \dots + \beta_r^D x_{t-r} \right) \right| \\
&\leq \frac{1}{\sqrt{r}} \left(\varphi_1 \left| \beta_1^1 x_{t-1} + \dots + \beta_r^1 x_{t-r} \right| + \dots + \varphi_D \left| \beta_1^D x_{t-1} + \dots + \beta_r^D x_{t-r} \right| \right) + c \text{ d'après (4.6)} \\
&= \frac{\varphi_1 |\beta_1^1| + \dots + \varphi_D |\beta_1^D|}{\sqrt{r}} |x_{t-1}| + \dots + \frac{\varphi_1 |\beta_r^1| + \dots + \varphi_D |\beta_r^D|}{\sqrt{r}} |x_{t-r}| + c \\
&= \varphi'_1 |x_{t-1}| + \dots + \varphi'_r |x_{t-r}| + c
\end{aligned}$$

Or :

$$\begin{aligned}
\sum_{i=1}^r \varphi'_i &= \frac{\varphi_1 |\beta_1^1| + \dots + \varphi_D |\beta_1^D|}{\sqrt{r}} + \dots + \frac{\varphi_1 |\beta_r^1| + \dots + \varphi_D |\beta_r^D|}{\sqrt{r}} \\
&= \frac{\varphi_1 (|\beta_1^1| + \dots + |\beta_r^1|) + \dots + \varphi_D (|\beta_1^D| + \dots + |\beta_r^D|)}{\sqrt{r}} \\
&\leq \varphi_1 + \dots + \varphi_D \\
&< 1
\end{aligned}$$

L'ergodicité géométrique est ainsi démontrée à l'aide du théorème 4.2. \diamond

Remarque 4.5

Le théorème 4.4 implique le théorème 4.3.

4.4 Cas hétéroscédastique

On a considéré jusqu'à présent des processus homoscedastiques, hypothèse contenable pour de nombreuses séries temporelles réelles. Ainsi, la consommation d'électricité a une variance qui dépend fortement de l'heure (Cf. Fig. 4.1 & 4.2) et on peut être amené à normaliser cette série afin de se ramener à une variance constante au cours du temps.

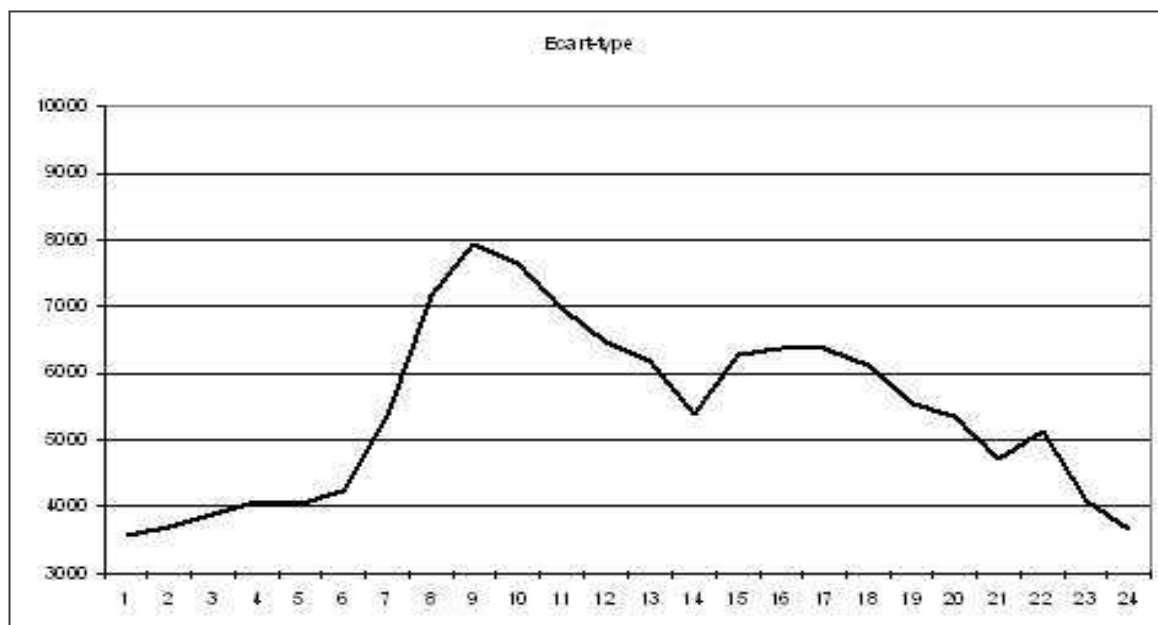


FIG. 4.1 – Ecart-type horaire de la consommation d'électricité : 03/04/06-09/07/06

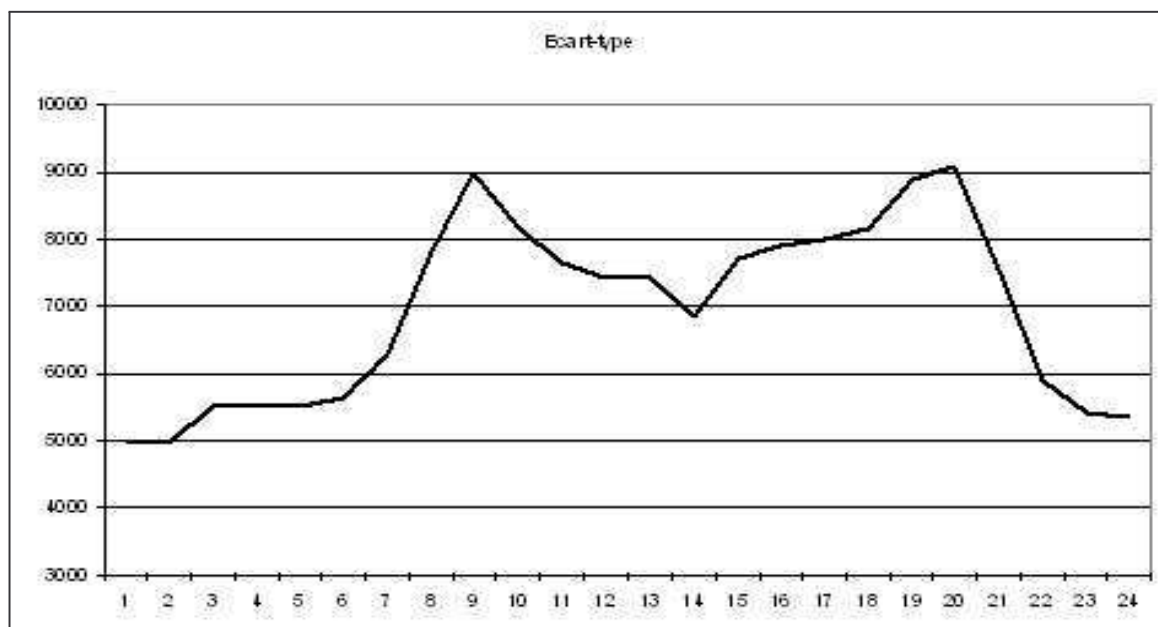


FIG. 4.2 – Ecart-type horaire de la consommation d'électricité : 15/01/07-08/04/07

On considère maintenant le modèle suivant :

$$\forall t \in \mathbb{N}^* : X_t = \Phi(X_{t-1}, \dots, X_{t-r}) + \sigma(X_{t-1}, \dots, X_{t-r}) \varepsilon_t \quad (4.7)$$

où :

- $(\varepsilon_t)_{t \in \mathbb{N}^*}$ est une suite de v.a.r. i.i.d. dans $\mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ admettant une fonction de densité.
- $\forall t \in \mathbb{N}^* : \varepsilon_t$ est indépendant de $\{X_{t-s}, s \in \mathbb{N}^* / t - s \geq -r + 1\}$
- $\Phi : \mathbb{R}^r \rightarrow \mathbb{R}$ est une fonction mesurable
- $\sigma : \mathbb{R}^r \rightarrow \mathbb{R}$ est une fonction mesurable

On peut réécrire le modèle (4.7) sous la forme suivante :

$$\begin{cases} \underline{X}_0 = (X_0, \dots, X_{-r+1}) \\ \forall t \in \mathbb{N}^* : \underline{X}_t = \underline{\Phi}(\underline{X}_{t-1}) + \underline{\sigma}(\underline{X}_{t-1}) \cdot \underline{\varepsilon}_t \end{cases} \quad (4.8)$$

où :

- $\underline{X}_t = (X_{t-1}, \dots, X_{t-r+1})^T \in \mathbb{R}^r$
- $\underline{\Phi}(\underline{X}_t) = (\Phi(\underline{X}_t), X_t, \dots, X_{t-r+2})^T \in \mathbb{R}^r$
- $\underline{\sigma}(\underline{X}_t) = (\sigma(\underline{X}_t), 0, \dots, 0)^T \in \mathbb{R}^r$
- $\underline{\varepsilon}_t = (\varepsilon_t, 0, \dots, 0)^T \in \mathbb{R}^r$
- \cdot désigne le produit scalaire

Chen *et al.* (2000) ont étudié ce problème et ont proposé des conditions pour obtenir l'ergodicité géométrique du modèle (4.7). Ils ont notamment énoncé le théorème suivant.

Théorème 4.5 *Chen et al. (2000)*

Soit une chaîne de Markov $(\underline{X}_t)_{t \in \mathbb{N}}$ définie par (4.8).

Supposons que la fonction Φ du modèle (4.7) vérifie :

$$|\Phi(x_1, \dots, x_r)| \leq \varphi_1 |x_1| + \dots + \varphi_r |x_r| + c_1$$

où $(\varphi_i)_{i \in \{1, \dots, r\}} \in \mathbb{R}^{+r}$ et $c_1 \in \mathbb{R}$.

Supposons que la fonction σ du modèle (4.7) vérifie :

$$|\sigma(x_1, \dots, x_r)| \leq \alpha_1 |x_1| + \dots + \alpha_r |x_r| + c_2$$

où $(\alpha_i)_{i \in \{1, \dots, r\}} \in \mathbb{R}^{+r}$ et $c_2 \in \mathbb{R}$.

Si $\sum_{i=1}^r (\varphi_i + \alpha_i \mathbb{E}(|\varepsilon_1|)) < 1$ alors le processus $(X_t)_{t \in \mathbb{N}}$ est géométriquement ergodique.

On est en mesure d'établir un résultat similaire pour un processus autorégressif semi-paramétrique (à directions révélatrices multiples) hétéroscédastique.

Théorème 4.6

Soit $(X_t)_{t \in \mathbb{N}}$ un processus vérifiant le modèle suivant :

$$\forall t \in \mathbb{N}^* : X_t = g \left(B^T (X_{t-1}, \dots, X_{t-r})^T \right) + \sigma (X_{t-1}, \dots, X_{t-r}) \varepsilon_t \quad (4.9)$$

où :

- $(\varepsilon_t)_{t \in \mathbb{N}^*}$ est une suite de v.a.r. i.i.d. dans $\mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ admettant une fonction de densité.
- $\forall t \in \mathbb{N}^* : \varepsilon_t$ est indépendant de $\{X_{t-s}, s \in \mathbb{N}^* / t - s \geq -r + 1\}$
- $g : \mathbb{R}^D \rightarrow \mathbb{R}$ est une fonction mesurable
- $\sigma : \mathbb{R}^r \rightarrow \mathbb{R}$ est une fonction mesurable
- $B = (\beta^1, \dots, \beta^D)_{r \times D}$, avec $1 \leq D < r$, est une matrice telle que $B^T B = I_D$

Supposons que les fonctions g et σ vérifient les conditions suivantes :

$$|g(x_1, \dots, x_D)| \leq \frac{1}{\sqrt{r}} (\varphi_1 |x_1| + \dots + \varphi_D |x_D|) + c_1 \quad (4.10)$$

$$|\sigma(x_1, \dots, x_r)| \leq \alpha_1 |x_1| + \dots + \alpha_r |x_r| + c_2 \quad (4.11)$$

où $(c_1, c_2) \in \mathbb{R}^2$, $(\varphi_i)_{i \in \{1, \dots, D\}} \in \mathbb{R}^{+D}$ et $(\alpha_i)_{i \in \{1, \dots, r\}}$ sont telles que :

$$\sum_{i=1}^D \varphi_i + \sum_{i=1}^r \alpha_i \mathbb{E}(|\varepsilon_1|) < 1$$

Le processus $(X_t)_{t \in \mathbb{N}}$ est alors géométriquement ergodique.

Démonstration

On a :

$$\begin{aligned} & \left| g \left(B^T (x_{t-1}, \dots, x_{t-r})^T \right) \right| \\ & \leq \frac{\varphi_1 |\beta_1^1| + \dots + \varphi_D |\beta_1^D|}{\sqrt{r}} |x_{t-1}| + \dots + \frac{\varphi_1 |\beta_r^D| + \dots + \varphi_D |\beta_r^D|}{\sqrt{r}} |x_{t-r}| + c_1 \\ & = \varphi'_1 |x_{t-1}| + \dots + \varphi'_r |x_{t-r}| + c \end{aligned}$$

Or :

$$\begin{aligned}
\sum_{i=1}^r (\varphi'_i + \alpha_i \mathbb{E}(|\varepsilon_1|)) &= \sum_{i=1}^r \varphi'_i + \sum_{i=1}^r \alpha_i \mathbb{E}(|\varepsilon_1|) \\
&= \frac{\varphi_1 |\beta_1^1| + \dots + \varphi_D |\beta_1^D|}{\sqrt{r}} + \dots + \frac{\varphi_1 |\beta_r^1| + \dots + \varphi_D |\beta_r^D|}{\sqrt{r}} + \sum_{i=1}^r \alpha_i \mathbb{E}(|\varepsilon_1|) \\
&= \frac{\varphi_1 (|\beta_1^1| + \dots + |\beta_r^1|) + \dots + \varphi_D (|\beta_1^D| + \dots + |\beta_r^D|)}{\sqrt{r}} + \sum_{i=1}^r \alpha_i \mathbb{E}(|\varepsilon_1|) \\
&\leq \sum_{i=1}^D \varphi_i + \sum_{i=1}^r \alpha_i \mathbb{E}(|\varepsilon_1|) \\
&< 1
\end{aligned}$$

L'ergodicité géométrique est ainsi démontrée à l'aide du théorème 4.5. \diamond

Conclusion

Les travaux initiés dans cette thèse avaient pour objectifs d'étudier les techniques semi-paramétriques appliquées à la consommation d'électricité. Les modèles utilisés aujourd'hui, entièrement paramétriques, sont très spécifiés et peu évolutifs ; ajouter de nouvelles variables explicatives peut devenir très laborieux. Or l'ouverture à la concurrence du marché de l'électricité conduit à "perturber" de plus en plus la courbe de charge et nous oblige à envisager de nouvelles variables explicatives (les prix de marché par exemple). Disposer de techniques plus souples que celles en place, présente donc un intérêt majeur pour RTE, en charge de l'équilibre offre/demande sur le réseau de transport.

Le second chapitre a démontré que la prévision d'une série temporelle autorégressive à l'aide d'une méthode de réduction de la dimension (MAVE), est efficace et conduit à des résultats similaires voire meilleurs que la prévision non-paramétrique dans certains cas. Si le modèle semi-paramétrique utilisé est suffisamment souple pour s'appliquer à de nombreux cas pratiques, on a constaté que certains calages revêtent une grande importance : le choix de la fenêtre évidemment, mais également la mémoire du processus sous-jacent et le nombre de directions révélatrices. Le praticien devra consacrer suffisamment de temps pour caler ces différents critères afin de s'assurer de la qualité des résultats obtenus au final.

Nous avons défini un modèle semi-paramétrique pour modéliser la consommation d'électricité dans le troisième chapitre. Le nombre et la diversité des variables explicatives (endogènes et exogènes) nous ont conduit à considérer un modèle semi-paramétrique semi-linéaire. Nous ne sommes pas encore parvenus à démontrer la convergence des estimateurs, et des travaux futurs seront menés dans ce sens. Il existe néanmoins des résultats dans le cas d'une seule direction révélatrice qui nous ont incités à poursuivre et à étudier les résultats obtenus sur des séries simulées et sur la courbe de charge. Les prévisions obtenues pour la consommation électrique sont légèrement moins bonnes que celles issues de la procédure actuelle, mais le modèle est beaucoup plus simple et plus "contrôlable". En outre, les prévisionnistes de RTE

ont besoin d'intervalles de prévision afin de disposer d'un indice de confiance sur les prévisions fournies par le modèle. Le modèle que nous avons défini permettrait de fournir beaucoup plus facilement cet intervalle.

Le seul inconvénient que nous avons mis en évidence est le temps de calcul qui peut devenir important si on veut caler correctement le modèle. Quoi qu'il en soit, nous disposons d'arguments solides pour poursuivre dans le semi-paramétrique et effectuer une mise en oeuvre réellement opérationnelle pour RTE.

Bibliographie

- [1] Abadie J., Meslier F., "Etude de l'utilisation des modèles A.R.I.M.A pour la prévision à court terme de l'énergie journalière produite par Electricité de France", R.A.I.R.O. Recherche opérationnelle/operation Research, Vol. 13, No. 1
- [2] Amato U., Antoniadis A., De Feis I., "Dimension reduction in functional regression with applications", Computational Statistics and Data Analysis, 2006, Vol. 50, No. 9
- [3] An H., Chen M., Huang F., "The geometric ergodicity and existence of moments for a class of non-linear time series model", Statistics & Probability Letters, 1997, Vol. 31
- [4] An H., Huang F., "The geometric ergodicity of non-linear time autoregressive models", Statistica Sinica, 1996, Vol. 6
- [5] Auestad B., Tjøstheim A., "Search for significant variables in nonparametric additive regression", Biometrika, 1996, Vol. 83, No. 3
- [6] Barnett W. A., Powell J., Tauchen G., "Nonparametric and semiparametric methods in econometrics and statistics", Cambridge University Press, 2004
- [7] Boneu M., Delecroix M., Malin M., "Semiparametric versus nonparametric estimation in single index regression model : a computational approach", Computational Statistics, 1993, Vol. 8, No. 3
- [8] Bosq D., "Nonparametric statistics for stochastic processes", Springer, Lecture Notes in Statistics, Vol. 110, 1996
- [9] Box G. E. P., Jenkins G. M., "Time Series Analysis : Forecasting and Control", Holden-Day, 1976
- [10] Cai Z., Fan J., Yao Q., "Functional-coefficient regression models for nonlinear time series", Journal of the American Statistical Association, 2000, Vol. 95, No. 451
- [11] Camlong-Viot C., "Modèle additif de régression sous des conditions de mélange", Thèse Université Toulouse III, 2000

-
- [12] Camlong-Viot C., Sarda P., Vieu P., "Additive time series : the kernel integration method", *Mathematical methods of Statistics*, 2000, Vol. 9, No. 4
- [13] Carroll R. J., Fan J., Gijbels I., Wand P., "Generalized partially linear single-index models", *Journal of the American Statistical Association*, 1997, Vol. 92, No. 438
- [14] Carbon M., Delecroix M., "Non-parametric vs parametric forecasting in time series : a computational point of view", *Applied Stochastic Models and Data Analysis*, 1993, Vol. 9
- [15] Chan K. S., Tong H., "On the use of deterministic Lyapunov function for the ergodicity of stochastic difference equations", *Advances in Applied Probability*, 1985, Vol. 17
- [16] Chan K. S., Tong H., "On likelihood ratio tests for threshold autoregression", *Journal of the Royal Statistical Society*, 1990, Vol. 52, No. 3
- [17] Chen H., "Estimation of a projection-pursuit type regression model", *The Annals of Statistics*, 1991, Vol. 19, No. 1
- [18] Chen M., Chen G., "Geometric ergodicity of nonlinear autoregressive models with changing conditional variances", *The Canadian Journal of Statistics*, 2000, Vol. 28, No. 3
- [19] Chen R., "Threshold variable selection in open-loop threshold autoregressive models", *Journal of Time Series Analysis*, 1995, Vol. 16, No. 5
- [20] Chen R., Tsay R., "Functional-coefficient autoregressive models", *Journal of the American Statistical Association*, 1993, Vol. 88, No. 421
- [21] Damon J., Guillas S., "The inclusion of exogenous variable in functional autoregressive ozone forecasting", *Université Paris VI ISUP-LSTA Rapport technique*, 2001, No. 4
- [22] Delecroix M., Hristache M., "M-estimateurs semi-paramétriques dans les modèles à direction révélatrice unique", *Bulletin of the Belgian Mathematical Society*, 1999, Vol. 6, No. 2
- [23] Delecroix M., Hristache M., Patilea V., "On semiparametric M-estimation in single-index regression", *Journal of Statistical Planning and Inference*, 2006, Vol. 136
- [24] Doukhan P., "Mixing. properties and examples", *Springer, Lecture Notes in Statistics*, No. 85, 1995
- [25] Epanechnikov V. A., "Nonparametric estimation of a multidimensional probability density", *Theory of Probability and Applications*, 1969, Vol. 14

-
- [26] Fan J., Gijbels I., "Local polynomial modelling and its applications", Chapman & Hall, 1996
- [27] Fan J., Zhang W., "Statistical estimation in varying coefficient models", The Annals of Statistics, 1997, Vol. 27, No. 5
- [28] Hall P., "On projection pursuit regression", The Annals of Statistics, 1989, Vol. 17, No. 2
- [29] Härdle W., "Applied nonparametric regression", Econometric Society Monographs, 1992, Vol. 19
- [30] Härdle W., Hall P., Ichimura H., "Optimal smoothing in single-index models", The Annals of statistics, 1993, Vol. 21, No. 1
- [31] Härdle W., Hart J., Marron J. S., Tsybakov A. B., "Bandwidth choice for average derivative estimation", Journal of the American Statistical Association, 1992, Vol. 87, No. 417
- [32] Härdle W., Korostelev A., "Search for significant variables in nonparametric additive regression", Biometrika, 1996, Vol. 83, No. 3
- [33] Harvey A., Koopman S. J., "Forecasting hourly electricity demand using time-varying splines", Journal of the American Statistical Association, 1993, Vol. 88, No. 424
- [34] Härdle W., Liang H., Gao J., "Partially linear models", Physica-Verlag, 2000
- [35] Härdle W., Mammen E., "Comparing nonparametric versus parametric regression fits", The Annals of Statistics, 1993, Vol. 21, No. 4
- [36] Härdle W., Müller M., Sperlich S., Werwatz A., "Nonparametric and semiparametric models", Springer, 2004
- [37] Härdle W., Stocker T. M., "Investigating smooth multiple regression by the method of average derivatives", Journal of the American Statistical Association, 1989, Vol. 84, No. 408
- [38] Härdle W., Vieu P., "Kernel regression smoothing of time series", Journal of Time Series Analysis, 1992, Vol. 13, No. 3
- [39] Hastie T., Tibshirani R., "Varying-coefficient models", Journal of the Royal Statistical Society, 1993, Vol. 55, No. 4
- [40] Horowitz J., "Semiparametric methods in econometrics", Springer, Lecture Notes in Statistics, Vol. 131, 1998
- [41] Horowitz J., Härdle W., "Direct semiparametric estimation of single-index models with discrete covariates", Journal of the American Statistical Association, 1996, Vol. 91, No. 436, Theory and methods

- [42] Hristache M., Juditsky A., Spokoiny V., "Direct estimation of the single index coefficients in single index models", *The Annals of Statistics*, 2001, Vol. 29, No. 3
- [43] Huber P., "Projection Pursuit", *The Annals of Statistics*, 1985, Vol. 13, No. 2
- [44] Ichimura H., "Semiparametric least squares (SLS) and weighted SLS estimation of single-index models", *Journal of Econometrics*, 1993, Vol. 58
- [45] Kim T. Y., Cox D. D., "Asymptotic behaviors of some measures of accuracy in nonparametric curve. Estimation with dependant observations", *Journal of Multivariate Analysis*, 1995, Vol. 53
- [46] Li K. C., "Sliced inverse regression for dimension reduction", *Journal of the American Statistical Association*, 1991, Vol. 86, No. 414
- [47] Li M. C., Chan K. S., "Semiparametric reduced-rank regression", The University of Iowa, Department of Statistics and Actuarial Science, Technical report, 2001
- [48] Li M. C., Chan K. S., Tong H., "Partially linear reduced-rank regression", The University of Iowa, Department of Statistics and Actuarial Science, Technical report, 2004
- [49] Linton O. B., Härdle W., "Estimation of additive regression models with known links", *Biometrika*, 1996, Vol. 83, No. 3
- [50] Martin M. M., "Filtrage de Kalman d'une série temporelle saisonnière. Application à la prévision de consommation d'électricité", *Revue de Statistique Appliquée*, 1999, Vol. 47, No. 4
- [51] Matzner-Løber E., Gannoun A., De Gooijer J. G., "Nonparametric forecasting : comparison of three kernel-based methods", *Communications in Statistics - Theory and Methods*, 1998, Vol. 27
- [52] McCullagh P., Nelder J. A., "Generalized linear models", Chapman & Hall, 1989, Second Edition
- [53] Meslier F., "Contribution à l'analyse des séries chronologiques et application à la mise au point de modèles de prévision à court terme relatifs à la demande journalière relevée à Paris Monsouris", Thèse Université Paris IX, 1976
- [54] Meyn S. P., Tweedie R. L., "Markov chains and stochastic stability", Springer, 1993
- [55] Meyn S. P., Tweedie R. L., "State-dependant criteria for convergence Markov chains", *Annals of Applied Probability*, 1994
- [56] Mokkadem A., "Critères de mélange pour des processus stationnaires. Estimation sous des hypothèses de mélange. Entropie des processus linéaires", Thèse Université Paris Sud, 1987

-
- [57] Opsomer J. D., Ruppert D., "Fitting a bivariate additive model by local polynomial regression", *The Annals of Statistics*, 1997, Vol. 25, No. 1
- [58] Pankratz A., "Forecasting with univariate Box-Jenkins models", Wiley, 1983
- [59] Pham D. T., "The mixing property of bilinear and generalised random coefficient autoregressive models", *Stochastic Processes and their Applications*, No. 23
- [60] Pinkse J., "Nonparametric regression estimation using weak separability", 2001, Preprint
- [61] Poggi J. M., "Prévision non-paramétrique de la consommation électrique", *Revue de Statistique Appliquée*, 1994, Vol. 42, No. 4
- [62] Rhomari N., "Filtrage non-paramétrique pour les processus non-markoviens", Thèse Université Paris 6, 1994
- [63] Schott J. R., "Determining the dimensionality in sliced inverse regression", *Journal of the American Statistical Association*, 1994, Vol. 89, No. 425
- [64] Shen J., "Etude probabiliste et statistique de modèles non-linéaires en présence de variables exogènes", Thèse Université Paris 6, 1995
- [65] Tjøstheim D., "Non-linear time series and Markov chains", *Advances in Applied Probability*, Vol. 22
- [66] Tong H., "Non-linear time series", Oxford University Press, 1990
- [67] Tweedie R. L., "Criteria for classifying general Markov chains", *Advances in Applied Probability*, 1975, Vol. 8
- [68] Xia Y., An H. Z., "Projection pursuit autoregression in time series", *Journal of Time Series Analysis*, 1999, Vol. 20, No. 6
- [69] Xia Y., Härdle W., "Semiparametric estimation of generalized partially linear single-index models", *Journal of Multivariate Analysis*, 2006, Vol. 97
- [70] Xia Y., Li W. K., "On single-index coefficient regression models", *Journal of the American Statistical Association*, 1999, Vol. 94, No. 448
- [71] Xia Y., Tong H., Li W. K., "On extended partially linear single-index models", *Biometrika*, 1999, Vol. 86, No. 4
- [72] Xia Y., Tong H., Li W. K., "An adaptative estimation of dimension reduction space", *Journal of the Royal Statistical Society, Ser. B*, 2002, Vol. 64, No. 3
- [73] Yang L., Tschernig R., "Multivariate bandwidth selection for local linear regression", *Journal of the Royal Statistical Society, Ser. B*, 1999, Vol. 61, No. 4
- [74] Zhu L. X., Fang K. T., "Asymptotics for kernel estimate of sliced inverse regression", *The Annals of Statistics*, 1996, Vol. 24, No. 3

Résumé

Une prévision correcte de la consommation d'électricité est fondamentale pour le bon fonctionnement du réseau électrique français, dont Réseau de Transport d'Electricité a la charge. Les prévisions utilisées quotidiennement par RTE sont issues d'un modèle alliant une régression paramétrique non linéaire et un modèle SARIMA.

Dans l'idée d'obtenir un modèle de prévision adaptatif, des méthodes de prévision non-paramétriques ont déjà été testées sans succès véritable. On sait notamment que la qualité d'un prédicteur non-paramétrique résiste mal à un grand nombre de variables explicatives, ce qu'on appelle communément le fléau de la dimension.

On a proposé récemment des méthodes semi-paramétriques d'estimation d'une régression qui améliorent l'approche non-paramétrique pure. L'une d'elles, basée sur la notion de "directions révélatrices" appelée MAVE (Moving Average -conditional- Variance Estimation), peut s'appliquer aux séries temporelles. Nous étudions empiriquement son efficacité pour prédire les valeurs futures d'une série temporelle autorégressive.

Nous adaptons ensuite cette méthode, d'un point de vue pratique, pour prédire la consommation électrique. Nous proposons un modèle semi-paramétrique semi-linéaire, basé partiellement sur la méthode MAVE, qui permet de prendre en compte simultanément l'aspect autorégressif du problème, et l'introduction de variables exogènes. La procédure d'estimation proposée se révèle efficace en pratique.

Mots-clefs

Prédictions non-paramétriques. Prédictions semi-paramétriques. Modèles à directions révélatrices multiples. Modèles semi-linéaires. Réduction de la dimension. Séries temporelles. Consommation d'électricité.

Abstract

Réseau de Transport d'Electricité (RTE), in charge of operating the French electric transportation grid, needs an accurate forecast of the power consumption in order to operate it correctly. The forecasts used everyday result from a model combining a nonlinear parametric regression and a SARIMA model.

In order to obtain an adaptive forecasting model, nonparametric forecasting methods have already been tested without real success. In particular, it is known that a nonparametric predictor behaves badly with a great number of explanatory variables, what is commonly called the curse of dimensionality.

Recently, semiparametric methods which improve the pure nonparametric approach have been proposed to estimate a regression function. Based on the concept of "dimension reduction", one those methods (called MAVE : Moving Average -conditional- Variance Estimate) can apply to time series. We study empirically its effectiveness to predict the future values of an autoregressive time series.

We then adapt this method, from a practical point of view, to forecast power consumption. We propose a partially linear semiparametric model, based on the MAVE method, which allows to take into account simultaneously the autoregressive aspect of the problem and the exogenous variables. The proposed estimation procedure is practicaly efficient.