



CONTRIBUTION À LA PRÉVISION DES PROCESSUS ALÉATOIRES PAR L'ANALYSE HARMONIQUE

Gabriel Cohen

► To cite this version:

Gabriel Cohen. CONTRIBUTION À LA PRÉVISION DES PROCESSUS ALÉATOIRES PAR L'ANALYSE HARMONIQUE. Statistiques [stat]. Conservatoire National des Arts et Métiers - CNAM Paris, 1999. Français. NNT : . tel-01357158

HAL Id: tel-01357158

<https://theses.hal.science/tel-01357158>

Submitted on 29 Aug 2016

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.

THESE
Présentée en vue de l'obtention du DOCTORAT du
CONSERVATOIRE NATIONAL DES ARTS ET METIERS
PARIS

Spécialité
INFORMATIQUE

par Gabriel COHEN

Sujet de la thèse
CONTRIBUTION A LA PREVISION DES PROCESSUS
ALEATOIRES PAR L'ANALYSE HARMONIQUE

Présentée le 1er décembre 1999 devant le jury composé de:

M. Michel CARBON, Professeur à l'Université de Rennes II, rapporteur
M. Pierre CAZES, Professeur à l'Université Paris-Dauphine, rapporteur
M. Jean Claude DEVILLE, Inspecteur Général de l'INSEE
M. Pierre-Louis GONZALEZ, Maître de Conférences, CNAM
M. Gilbert SAPORTA, Professeur CNAM

TABLE DES MATIERES

INTRODUCTION	4
Chapitre I : LES DIFFERENTS MODELES DE DESCRIPTION D'UNE SERIE CHRONOLOGIQUE	6
I-1) Les méthodes de description classiques	8
I-2) Les méthodes liées à l'interprétation d'une série chronologique comme réalisation d'un processus	9
Adaptation des outils habituels de l'analyse des processus aux études d'une série chronologique	
La classe des processus AutoRégressif Moving Average (ARMA)	
I-3) Les méthodes d'analyse utilisant la connaissance théorique du noyau de covariance	24
Application au processus de Poisson	
Chapitre II : DESCRIPTION DES FAMILLES DE TRAJECTOIRES A L'AIDE DE L'ANALYSE HARMONIQUE	27
II-1) Les notations adaptées à ce type d'analyse	
II-2) La méthode descriptive d'un ensemble de séries chronologiques à travers l'analyse harmonique.	33
II-3) Approximation discrète d'un processus et autres travaux sur l'analyse harmonique	37
Chapitre III : LES METHODES DE PREVISION D'UNE SERIE CHRONOLOGIQUES	41
III-1) Les méthodes classiques de prévision	
III-2) La prévision des processus stationnaires et la méthode de Box et Jenkins	42

Chapitre IV : LA PREVISION ET L'ANALYSE HARMONIQUE	49
IV-1) Adaptation de l'analyse harmonique aux problèmes de la prévision d'un processus	
Théorème Principal (comportement des composantes temporelles) Comportement des valeurs propres en fonction du temps écoulé	
IV-2) Réalité et illusion de la prévision harmonique	57
Processus équivalents Indépendance de l'avenir en fonction du passé Théorème de périodicité parfaite Théorème de répétition parfaite	
IV-3) La place des processus M.A. dans l'analyse harmonique	62
IV-4) Analyse harmonique et prévision des processus stationnaires (AR, Poisson...)	64
IV-5) Stratégie de prévision des familles de trajectoires	65
Chapitre V : VALIDITÉ DE L'ANALYSE HARMONIQUE DANS LE CAS D'UN NOYAU DE COVARIANCE QUELCONQUE	67
V-1) Les différentes procédures possibles pour sélectionner les composantes temporelles significatives .	
V-2) Etude de l'indice fort de perte d'information .	69
Chapitre VI : APPLICATIONS PRATIQUES	74
VI-1) Organisation des études pratiques	
VI-3) Indices trimestriels de la production industrielle en France de 1980 à 1988 (données INSEE)	75
VI-4) La consommation d'électricité en France de 1980 à 1972 (Données E.D.F.)	82
Chapitre VII : CONCLUSIONS ET PROSPECTIVES	85
VI-1) Premières conclusions	
VI-2) Autres perspectives liées à l'analyse harmonique	86

ANNEXES

ORGANISATION DES ANNEXES	88
Annexe I : INDICES TRIMESTRIELS DE LA PRODUCTION INDUSTRIELLE EN FRANCE DE 1980 À 1988	89
Détail des calculs et des études faits sur ces données	
Annexe II : LA CONSOMMATION D'ÉLECTRICITÉ EN FRANCE DE 1980 À 1972	135
Détail des calculs et des études faits sur ces données	
Annexe III : APPLICATION DE L'ANALYSE HARMONIQUE AU CHOIX DU NOMBRE D'AXES EN ANALYSE EN COMPOSANTES PRINCIPALES	147
Essai d'utilisation des vecteurs propres de la matrice de covariance dans le choix du nombre d'axes significatifs	
Annexe IV : DEMONSTRATION DE QUELQUES CALCULS CONNUS ET UTILES	163
Annexe V : LE TRAITEMENT INFORMATIQUE DES DONNEES	168
BIBLIOGRAPHIE	171

INTRODUCTION

La prévision est une préoccupation apparue avec les premiers relevés statistiques. Au début du XX^{ème} siècle les besoins de rationaliser l'économie ont amené les scientifiques à développer des techniques de plus en plus complexes d'analyse d'une série chronologique. Dans les années 70 le développement de l'informatique et des communications ont augmenté les besoins des décideurs en matière de prospectives de toutes sortes. C'est dans ce mouvement général que se placent les progrès effectués par les techniques de prévision statistiques.

A ce jour, un grand nombre de logiciels standards (tableurs,.....) proposent, pour prolonger une série chronologique, des solutions automatiques basées sur la régression linéaire ou polynomiale.

Les séries économiques ou financières demandent des outils plus sophistiqués. Pour répondre à ces exigences de nombreuses méthodes ont été proposées. Une des plus élaborée et efficace fut celle proposée par Box et Jenkins, en 1976.

Il n'est plus rare, pour le statisticien, de posséder un ensemble de séries chronologiques appartenant à un même environnement. Par exemple :

- * Les relevés météorologiques quotidiens à différents endroits d'une région.
- * Le nombre de voitures, par minute, aux différents points du périphérique.
- * Le nombre de bouteilles vendues quotidiennement dans chaque supermarché d'une chaîne commerciale.
- *

Le traitement isolé de chaque série chronologique prive toute analyse et toute prévision de l'information contenue dans l'appartenance à une même famille. Plus précisément l'analyse et la prévision d'une famille Ω de séries chronologiques $\{X_t(\omega)\}_{\omega \in \Omega}$ définie sur un intervalle de temps T doit pouvoir prendre en compte la structure de $\Omega \times T$. Le travail de prévision n'est plus dans ce cas l'analyse d'une série de valeurs et l'estimation de la valeur à prédire mais la description d'un tableau de données et son prolongement.

En 1974, dans son étude sur la constitution des familles, Jean Claude Deville propose une méthode, intitulée «Méthodes statistiques et numériques de l'analyse harmonique» qui intègre l'information contenue dans une famille de séries chronologiques.

Cette méthode, essentiellement descriptive, met en oeuvre les savoir-faire de l'analyse des données et de l'analyse fonctionnelle. Elle a pour origine les travaux de Karhunen et Loeve qui dès 1945 proposaient une décomposition optimale d'un processus en processus quasi déterministe.

Le calcul des processus quasi déterministes est lié fortement à l'intervalle de temps considéré. Dans le cas général il n'est pas évident que les processus quasi déterministes calculés sur un intervalle de temps T aient un lien clair avec ceux calculés sur un intervalle $T+1$. C'est pourquoi, dans un premier temps, la possibilité d'utiliser la décomposition de Karhunen et Loeve pour réaliser une prévision est apparue comme impossible et que cette direction n'a pas été exploitée plus tôt.

Notre contribution essentielle a été de montrer deux résultats :

- Dans les situations pratiques il y a de bonnes raisons théoriques d'espérer que les premiers processus quasi déterministes calculés à l'instant T et ceux calculés à instant $T+1$ soient comparables sur T
- L'information apportés par les autres processus quasi déterministes est négligeable.

Grâce à ces résultats nous avons pu légitimement faire une prévision en prolongeant chaque processus quasi déterministe et en reconstituant le processus. Cette méthode fournit un nouvel outil de prévision que nous avons utilisé pour réaliser deux études et leurs prévisions.

- Une étude sur les indices trimestriels de la production industrielle en France de 1988 publiée par l'INSEE
- Une étude sur la consommation d'électricité par secteurs économiques de Janvier 1968 à Décembre 1992.

Nous avons comparé nos résultats aux résultats obtenus par une méthode classique de prévision sur chaque série chronologique et nous avons constaté que le traitement global de notre méthode apportait des améliorations notables.

Pour réaliser les études pratiques un logiciel sous C++ a été réalisé dont les résultats ont été récupérés sur EXCEL. Ce programme est actuellement dans un environnement Windows. Il est déposé à l'agence de la protection des programmes: 119, avenue de Flandre à Paris, France.

CHAPITRE I

LES DIFFERENTS MODELES DE DESCRIPTION ET DE PREVISION DES SERIES CHRONOLOGIQUES

« Toute prévision scientifique, toute science expérimentale se fonde sur une projection sur l'avenir. L'esprit humain ne saurait prévoir qu'en tirant des leçons du passé et en faisant appel à son imagination » (Extrait de l'Encyclopédie Universalis à la rubrique : prévision économique) .

La prévision comporte, de par sa nature même, deux aspects :

- ⇒ Un travail d'observation et d'analyse des informations livrées par le passé
- ⇒ Un travail d'imagination qui permet de proposer une vision probable du futur.

La prévision des séries chronologiques n'échappe pas à cette démarche. C'est pourquoi nous nous proposons de faire un rapide tour des méthodes descriptives utilisées avant de rappeler les méthodes prévisionnelles qui y ont été attachées.

Nous ne prétendons pas faire une présentation exhaustive des outils d'analyse et de prévision des séries chronologiques proposées en statistique mais simplement nous souhaitons permettre au lecteur de se situer rapidement. Nous recommandons pour plus d'information le livre de Guy Mélard [25] ou Michel Tenenhaus [32].

A ce jour les méthodes de prévision d'une seule série chronologique sont disponibles dans de nombreux logiciels statistiques.

Les méthodes élémentaires basées sur la régression linéaire ou polynomiale sont à la disposition du grand public dans tous les tableurs commercialisés.

Différentes méthodes classiques permettent de décomposer une série en tendance et en saisonnalité. On utilise un modèle multiplicatif ou un modèle additif suivant les situations.

Ces différentes techniques de prévision procèdent de démarches empiriques et sont relativement insatisfaisantes sur le plan scientifique.

Pour sortir du cadre trop étroit qu'impose le modèle classique on est amené à interpréter la série chronologique $\{x_{t_0}, x_{t_0-1}, x_{t_0-2}, \dots, x_{t_0-n}\}$ comme une réalisation d'un processus discret $\{X_t\}_{t \in \mathbb{Z}}$ aux instants t compris entre t_0-n et t_0 ($t_0-n \leq t \leq t_0$). Si on fait l'hypothèse que la variable X_{t_0} peut s'expliquer en fonction du passé alors:

$$X_{t_0} = f(X_{t_0-1}, X_{t_0-2}, \dots, X_{t_0-n}).$$

Cette relation s'écrit pour t_0+1 $X_{t_0+1} = f(X_{t_0}, X_{t_0-1}, X_{t_0-2}, \dots, X_{t_0-n+1})$ et permet de calculer la valeur x_{t_0+1} de la prévision de la série chronologique de départ :

$$X_{t_0+1} = f(X_{t_0}, X_{t_0-1}, X_{t_0-2}, X_{t_0-3}, \dots, X_{t_0-n})$$

Sous ces hypothèses la prévision consiste à trouver la meilleure relation liant le passé au futur c'est à dire optimiser la relation $X_{t_0} = f(X_{t_0-1}, X_{t_0-2}, \dots, X_{t_0-p})$

Box et Jenkins, en 1976, ont proposé une méthode qui sélectionne dans la vaste classe des modèles ARIMA (AutoRegressive Integrated Moving Average) le modèle le mieux adapté à la série chronologique étudiée. C'est une réponse statistique adaptée au problème de la prévision d'une série chronologique.

On doit cependant mentionner, pour les séries trop courtes ou trop irrégulières, que cette méthode est mal adaptée et qu'il est préférable d'utiliser un lissage exponentiel.

Il y a peu de méthode de prévision qui prennent en compte un ensemble de séries chronologiques, c'est à dire un tableau chronologique. Un travail de description a cependant été proposé par J.C. Deville en 1974, dans son étude sur la constitution des familles. Cette méthode de description utilise l'analyse harmonique c'est à dire en pratique l'analyse des vecteurs propres de l'opérateur de covariance. C'est à partir des résultats de cette étude et des travaux de G.Saporta sur les vecteurs propres du noyau de covariance des processus de Poisson que nous avons imaginé la méthode de prévision présentée au chapitre IV.

Dans un premier temps nous avons pensé que notre méthode n'avait aucun lien avec les méthodes traditionnelles de prévision d'une seule série chronologiques si ce n'est dans la comparaison des résultats. Cependant lors de l'étude de la classe des processus ARIMA utilisée par Box et Jenkins on peut remarquer que la matrice de covariance est connue. Les processus quasi déterministes qui permettent en analyse harmonique de reconstituer les données ne peuvent pas être déterminés (on ne connaît qu'une seule série) mais il est possible d'estimer les vecteurs propres qui les engendrent. Des liens plus clairs et plus structurés restent à établir entre la description des processus à l'aide des modèles ARIMA et celle proposée par l'analyse harmonique. Nous avons pensé qu'il était cependant intéressant dans

le premier chapitre de présenter pour les processus ARIMA standards, les vecteurs propres et les valeurs propres de leurs matrice de covariance.

I-1) LES MÉTHODES DE DESCRIPTION CLASSIQUES :

I-1-a) Les méthodes élémentaires

Les premières méthodes mises en place furent des méthodes faisant appel à un modèle déterministe.

La régression linéaire propose de remplacer une série chronologique par une droite la plus proche possible des données. Cette méthode simple s'applique aux données qui sont relativement linéaires (voir Michel Tenenhaus [32] page 57 à 65) .

Dans le même ordre d'idée la régression polynomiale, cherche à approcher les données par un polynôme (spline). (Michel Tenenhaus [32] page 271) .

I-1-b) Les méthodes de décomposition d'une série chronologique en tendances et saisonnalité

Pour étudier des séries économiques liées aux périodes de l'année on est obligé de faire une hypothèse sur le mode de « fabrication » de la série. On suppose que la série est organisée à partir de la somme ou du produit de trois composantes :

- ⇒ La tendance
- ⇒ la saisonnalité
- ⇒ le résidu

La série chronologique X_t s'écrit alors :

$$X_t = T_t \cdot S_t \cdot R_t \quad \text{ou} \quad X_t = T_t + S_t + R_t$$

T_t représente la tendance de la série, S_t la saisonnalité et R_t le résidu. La tendance est déterminée par lissage. Dans la méthode classique ce lissage est proposé en utilisant des moyennes mobiles. Cette technique est trop sensible aux valeurs aberrantes, notamment aux valeurs aberrantes extrêmes. Des méthodes plus robustes ont été mises au point dans les années 70, en particulier par Tukey (Michel Tenenhaus [32] ,page 273).

La saisonnalité S_t est obtenue par soustraction ($x_t - T_t = S_t$) pour le modèle additif ou par division ($x_t / T_t = S_t$) pour le modèle multiplicatif.

Le résidu comme son nom l'indique n'est que le reste de la série lorsque on a enlevé la tendance et la saisonnalité, c'est à dire dans le modèle additif $R_t = x_t - T_t - S_t$ et dans le modèle multiplicatif $R_t = x_t / (T_t \cdot S_t)$ (Michel Tenenhaus [32] page 266 à 284)

Implicitement dans ces méthodes on espère que le résidu ne comporte plus d'information significative mais décrit les fluctuations aléatoires du processus; le résidu est alors appelé bruit blanc. Si on imagine que le résidu renferme encore de l'information on s'efforce à le modéliser par une combinaison linéaire de bruit blancs décalés. Cette méthode sera décrite dans le paragraphe sur les modèles ARIMA.

I-c) La régression linéaire multiple :

Dans cette méthode on suppose que la série chronologique x_t que l'on désire analyser peut se décrire à l'aide d'une combinaison linéaire des k séries $\{x_{t-k}\}$ chronologiques décalées. Le problème consiste à trouver la meilleure combinaison possible

Cela revient à rechercher les meilleurs coefficients $\{a_1, a_1, a_1, a_1, \dots, a_k\}$ tels que la série des combinaisons linéaires de la série décalées plusieurs fois :

$$a_1 x_{t-1} + a_2 x_{t-2} + a_3 x_{t-3} + \dots + a_k x_{t-k}$$

soit la « plus proche » possible de la série x_t pour tout t .

A ce stade du calcul l'estimation des coefficients ne fait pas appel à un modèle probabiliste mais à des techniques de l'analyse des données. On pourra regarder avec profit Guy Mélard [25] pages 171 à 181.

On remarquera que toutes les méthodes énumérées ci-dessus ne font appel à aucune hypothèse sur la structure des données, seule la connaissance des valeurs est prise en compte.

I-2) LES MÉTHODES LIÉES À L'INTERPRÉTATION D'UNE SÉRIE CHRONOLOGIQUE COMME RÉALISATION D'UN PROCESSUS

Si on désire aller plus loin dans la description d'une série chronologique on est amené à interpréter les données comme une famille $\{X_t\}_{t \leq p}$ de variables aléatoires de second ordre dont une réalisation est la série des valeurs observées $\{x_t, x_{t-1}, x_{t-2}, x_{t-3}, \dots, x_{t-n}\}$. Cette interprétation permet d'utiliser tous les outils fournis par les statistiques et les probabilités et tout particulièrement les résultats sur les processus définis sur un intervalle donné T .

Remarque : Nous donnerons plusieurs définitions équivalentes des processus au chapitre II qui justifieront l'existence des différents calculs en particulier de la moyenne et de la variance

I-2-a) Adaptation des outils habituels de l'analyse des processus aux études d'une série chronologique

Lorsque on possède les valeurs d'une série chronologique on est capable dans la pratique de la transformer, à l'aide d'opérations simples, en une série chronologique qui peut être interprétée comme la réalisation d'un processus stationnaire. C'est pourquoi, dans la suite de cet exposé, nous présenterons uniquement la modélisation des processus stationnaires.

I-2-aa) Définition : Espérance mathématique

L'espérance mathématique du processus est l'ensemble des valeurs $\{E(X_t)\}_t$ obtenues en prenant les espérances mathématiques de chacune des variables X_t .

Rappelons que la moyenne d'un processus quelconque à l'instant t se calcule par la formule $E(X_t) = \int_{\Omega} X_t(\omega) dP\omega$; pour estimer cette moyenne il faudrait posséder un ensemble de réalisations $\{\bar{x}_t(\omega_j); \bar{x}_{t-1}(\omega_j); \bar{x}_{t-2}(\omega_j); \dots; \bar{x}_{t-n}(\omega_j)\}_{j \in \{1,2,\dots,m\}}$ et estimer la moyenne par la formule $\bar{x}_t = \frac{1}{m} \sum_{j=1}^m x_{t-i}(\omega_j)$ pour tout t . On obtiendrait ainsi une moyenne qui serait une fonction.

Dans la mesure où on ne connaît qu'une seule réalisation du processus cette démarche n'a pas de sens. En faisant l'hypothèse que le processus dont les données sont une réalisation est stationnaire on suppose implicitement que le processus a une moyenne nulle ou du moins constante. Dans ces conditions on est en droit de supposer que les données fluctuent autour de la valeur moyenne (hypothèse d'ergodicité); la moyenne du processus est alors un nombre que l'on estime à l'aide de la moyenne empirique par la formule : $\frac{1}{n+1} \sum_{i=0}^n x_{t-i}$

I-2-ab) Définition: Processus centré

Le processus centré est le processus défini par $X_t - E(X_t)$

L'espérance d'un processus centré est nulle. La série des valeurs centrées est alors estimée en soustrayant aux valeurs brute la moyenne empirique.

Pour des raisons de commodité nous supposons à partir de maintenant que tous les processus sont centrés et de second ordre.

I-2-ac) Définition: Noyau de covariance

On est amené à définir les liens entre les différentes variables $\{X_t\}_t$ en calculant leurs covariances deux à deux. On crée ainsi une fonction réelle à deux variables appelé noyau de covariance:

$$N(t,t') = E(X_t \cdot X_{t'})$$

I-2-ad) Définition: Processus stationnaires

Un processus stationnaire est un processus ayant une structure probabiliste stable au cours du temps. (Tennenhaus [30] page 286)

On en déduit qu'un processus stationnaire a une moyenne et une variance constante et que son noyau de covariance dépend uniquement du décalage :

$$N(t, t+h) = \gamma(h)$$

Cette hypothèse permet d'estimer la variance $V(t)$ par la formule $\frac{1}{n} \sum_{i=0}^{n-1} (x_{t-i})^2$ et les covariances $N(t, t+1) = \gamma(1)$; $N(t, t+2) = \gamma(2)$; par les formules $\frac{1}{n-1} \sum_{i=0}^{n-2} x_{t-i} x_{t-i-1}$; $\frac{1}{n-2} \sum_{i=0}^{n-3} x_{t-i} x_{t-i-2}$;

Remarque : Les processus aléatoires stationnaires centrés et sans composante déterministes X_t se décomposent suivant une combinaison linéaire de bruits blancs (Tenenhaus [30] page 288).

$$X_t = \sum_{i=0}^{\infty} \alpha_i e_{t-i}$$

I-2-ae) Définition : Opérateur de covariance

On associe au noyau de covariance l'opérateur de covariance:

$$U(f)(t) = \int_T N(t, t') f(t') dt'$$

où $f(t)$ est une fonction définie sur l'intervalle T d'observation du processus.

Remarque : Pour la justification de l'existence de l'intégrale et les propriétés de $U(f)(t)$ voir [15] J.C. Deville

Désignons par $f_1, f_2, f_3, \dots, f_n$ les vecteurs propres du noyau de covariance dont l'image par U n'est pas nulle et considérons un ensemble de vecteurs propres du noyau $g_1, g_2, g_3, \dots, g_k$ orthogonaux entre eux formant avec les vecteurs $f_1, f_2, f_3, \dots, f_n$ une base de l'espace des fonctions de carré intégrable. Si $X(\omega, t)$ est une trajectoire du processus

$$X_t(\omega) = X(\omega, t) = [X(\omega)](t) = \sum_1^n a_i(\omega) f_i(t) + \sum_1^k b_j(\omega) g_j(t)$$

De plus la projection de toute réalisation sur le noyau est presque sûrement nulle (Annexe 4_3). Cela permet lors de la décomposition d'un processus suivant les vecteurs propres associés à l'opérateur de covariance de ne pas prendre en compte la base du noyau de covariance. Nous pouvons ainsi écrire presque pour tout ω :

$$X_t(\omega) = X(\omega, t) = [X(\omega)](t) = \sum_1^n a_i(\omega) f_i(t) \quad (1)$$

L'égalité ci dessus permet de décrire une trajectoire en utilisant uniquement les fonctions propres de l'opérateur de covariance.

Dans les études suivantes on est capable d'estimer les matrices de covariances des processus associés à une série chronologique donnée. Nous serons de ce fait à même de décomposer le processus à l'aide des fonctions propres de son noyau de covariance.

I-2-b) La classe des processus ARMA

Comme nous l'avons dit au début de ce chapitre pour étudier une série chronologique on recherche un processus dont la série peut être considérée comme une réalisation. On est amené à rechercher une classe de processus suffisamment large pour répondre à cette demande et suffisamment simples pour être facilement étudiée. Une fois la tendance et la saisonnalité « retirée » de la série statistique donnée le bruit restant peut souvent être supposé stationnaire. (Tenenhaus [30] page 286 et suivantes)

Pour décrire l'information restante dans ce bruit, on suppose qu'il est engendré par l'information issue du passé et d'une combinaison de chocs aléatoires.

Le processus le plus élémentaire qui représente les chocs aléatoires est le bruit blanc. La combinaison de bruit blancs décalés dans le temps engendre des processus appelés Moyennes Mobile, ou MA (Moving Average).

Les processus obtenus à l'aide de combinaison linéaire du processus avec décalage arrière (X_{t-h}) permettent d'engendrer une autre classe de processus appelés processus Auto Régressifs (AR)

La combinaison des MA et des AR engendre une classe assez vaste de processus capables de générer les processus correspondants à notre attente. Cette classe s'appelle la classe des processus ARMA (Auto Regressive Moving Average)

Nous présenterons en premier le bruit blanc et la promenade aléatoire. Ensuite nous ferons un rapide rappel des processus MA (Moving Average) et des processus AR (Auto Régressif) pour terminer par les processus ARMA.

I-2-ba) Deux exemples « naturels »: le bruit blanc et la promenade aléatoire;

I-2-baa) Le bruit blanc

Définition

Le bruit blanc est un processus de moyenne nulle constitué de variables non corrélées entres elles ayant la même variance.

Plus précisément si e_t est un bruit blanc :

$$E(e_t) = 0 \quad , \quad N(t,t') = E(e_t \cdot e_{t'}) = 0 \quad \text{si } t \neq t' \quad \text{et} \quad V(e_t) = \sigma^2$$

Remarque : Dans la suite du texte les bruits blancs seront toujours notés e_t

Si on désire estimer les erreurs de prévision on ajoute souvent implicitement l'hypothèse de normalité.

Sa matrice de covariance est alors la matrice identité multipliée par la variance σ^2 . Elle comporte une seule valeur propre σ^2 et tout vecteur est vecteur propre. On ne peut rien tirer de cette information. L'absence de sens des vecteurs propres est une particularité des bruits blancs liée à l'égalité des valeurs propres. Nous montrerons que la possibilité de remplacer un processus par sa décomposition en processus quasi déterministes (engendrés par les vecteurs propres de l'opérateur de covariance) est d'autant plus légitime que les valeurs propres sont séparées.

I-2-bab) Le processus de promenade aléatoire (random walk process):

Lorsque on observe une série statistique on peut imaginer que le passage de la donnée x_{t-1} à x_t se fait d'une façon aléatoire. C'est à dire que la différence $x_{t-1} - x_t$ est une réalisation d'un bruit blanc. La définition du processus qui est à l'origine de cette série statistique est appelée « promenade aléatoire » (Guy Méléard [25] page 333).

Définition :

Un processus discret X_t , stationnaire et centré, est appelé promenade aléatoire si et seulement si

$$X_t - X_{t-1} = e_t \quad (1) \quad \text{pour tout } t$$

La définition fait apparaître le processus décalé X_{t-1} . Si on note B l'opérateur linéaire de décalage défini par $B(X_t) = X_{t-1}$ on peut écrire l'égalité (1) sous la forme :

$$(1 - B)X_t = e_t \quad (2)$$

A partir de ces définitions on obtient les propriétés suivantes:

De la relation (1) on déduit que :

$$E(X_t) = E(X_{t-1}) = E(X_{t-k}) = 0$$

En multipliant la relation (1) par le processus X_{t-k} on établit la relation :

$$X_{t-k} X_t - X_{t-k} X_{t-1} = X_{t-k} e_t$$

Le processus étant stationnaire et centré il se décompose suivant la série $X_t = \sum_{i=0}^{\infty} \alpha_i e_{t-i}$ (voir ci dessus la remarque de I-2-ac). De ce fait $\text{cov}(X_{t-k}, e_t) = 0$ (Guy Mélard [25] page 313)

En prenant l'espérance des deux membres on obtient :

$$E(X_{t-k} X_t) - E(X_{t-k} X_{t-1}) = E(X_{t-k} e_t) = 0$$

d'où on déduit pour tout $k > 0$:

$$\text{Cov}(X_{t-k}, X_t) = \text{Cov}(X_{t-k}, X_{t-1}) \text{ donc}$$

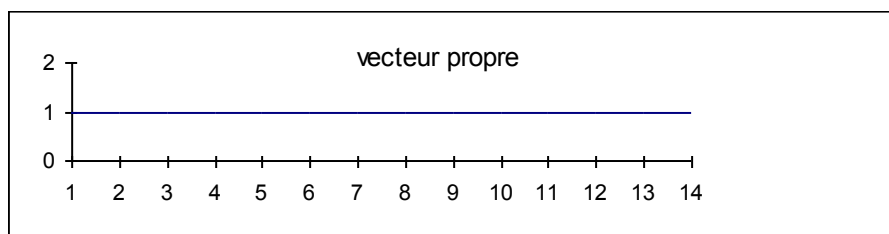
$$\text{pour } k=1 : \text{Cov}(X_{t-1}, X_t) = \text{Cov}(X_{t-1}, X_{t-1}) = V(X_{t-1}) = \sigma^2$$

Le processus étant stationnaire le noyau de covariance est une fonction $\gamma(n)$ de l'écart n entre les rangs

$$\text{pour } k=2 : \text{Cov}(X_{t-2}, X_t) = \text{Cov}(X_{t-2}, X_{t-1}) = \gamma(1) = \text{Cov}(X_{t-1}, X_t) = \sigma^2$$

$$\text{pour } k=n : \text{Cov}(X_{t-n}, X_t) = \text{Cov}(X_{t-n}, X_{t-1}) = \gamma(n-1) = \gamma(n-2) = \dots = \gamma(1) = \sigma^2$$

La matrice de covariance est composée de valeurs toutes égales à σ^2 et comportant qu'un seul vecteur propre à un coefficient près par exemple le vecteur $(1, 1, 1, 1, 1, \dots)$



I-2-bb) Le modèle MA (Moving Average) :

Les deux exemples que nous avons traités ci-dessus mettent en valeur le rôle du bruit blanc dans la construction des modèles. L'ensemble des processus que l'on peut engendrer à l'aide de combinaisons linéaires de bruit blanc auquel on a appliqué l'opérateur B de décalage arrière s'appelle le modèle moyenne mobile ou M.A.

Définition des processus MA

Un processus $MA(q)$ est défini par l'une des équations suivantes:

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \theta_3 e_{t-3} \dots - \theta_q e_{t-q}$$

$$Y_t = (1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 \dots - \theta_q B^q) e_t$$

Nous présenterons d'une façon détaillée uniquement les processus $MA(1)$ comme le fait Guy Mélard [25] page 305 en y ajoutant la description de la matrice de covariance et des vecteurs propres associés.

Les processus $MA(1)$ sont de la forme

$$Y_t = e_t - \theta e_{t-1} \quad (1)$$

En multipliant chaque membre de l'égalité par le processus Y_{t-q} et en y appliquant l'opérateur espérance mathématiques on a la relation :

$$\text{cov}(Y_t, Y_{t-q}) = E((e_t - \theta e_{t-1})(e_{t-q} - \theta e_{t-q-1}))$$

d'où on déduit si $q=0$ on a $V(Y_t) = (1+\theta^2) \sigma^2$; si $q=1$ on a $\text{cov}(Y_t, Y_{t-1}) = -\theta \sigma^2$ et si $q > 1$
 $\text{cov}(Y_t, Y_{t-1}) = \text{cov}(Y_t, Y_{t-2}) = \text{cov}(Y_t, Y_{t-3}) = \dots = 0$

On peut supposer que le bruit blanc a une variance égale à 1, la matrice de covariance à la forme suivante répétée à l'infini:

$$\begin{array}{cccccccc} 1+\theta^2 & -\theta & 0 & 0 & 0 & 0 & 0 & 0 \\ -\theta & 1+\theta^2 & -\theta & 0 & 0 & 0 & 0 & 0 \\ 0 & -\theta & 1+\theta^2 & -\theta & 0 & 0 & 0 & 0 \\ 0 & 0 & -\theta & 1+\theta^2 & -\theta & 0 & 0 & 0 \\ 0 & 0 & 0 & -\theta & 1+\theta^2 & -\theta & 0 & 0 \\ 0 & 0 & 0 & 0 & -\theta & 1+\theta^2 & -\theta & 0 \\ 0 & 0 & 0 & 0 & 0 & -\theta & 1+\theta^2 & -\theta \\ 0 & 0 & 0 & 0 & 0 & 0 & -\theta & 1+\theta^2 \end{array}$$

La valeur de θ est stictement comprise entre -1 et +1 (Guy Mélard [25] page 308)

I-2-bba) Etude approfondie d'un cas particulier: le modèle $MA(1)$ avec $\theta = -0.5$:

La matrice de covariance de ce processus est une matrice M_n , comportant n lignes et n colonnes de la forme :

$$\begin{array}{cccccccc} 1,25 & 0,50 & 0,00 & 0,00 & 0,00 & . & . & . & 0,00 & 0,00 \\ 0,50 & 1,25 & 0,50 & 0,00 & 0,00 & . & . & . & 0,00 & 0,00 \\ 0,00 & 0,50 & 1,25 & 0,50 & 0,00 & . & . & . & 0,00 & 0,00 \\ 0,00 & 0,00 & 0,50 & 1,25 & 0,50 & . & . & . & 0,00 & 0,00 \\ 0,00 & 0,00 & 0,00 & 0,50 & 1,25 & . & . & . & 0,00 & 0,00 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0,00 & 0,00 & 0,00 & 0,00 & 0,00 & . & . & . & 0,50 & 0,00 \\ 0,00 & 0,00 & 0,00 & 0,00 & 0,00 & . & . & . & 1,25 & 0,50 \\ 0,00 & 0,00 & 0,00 & 0,00 & 0,00 & . & . & . & 0,50 & 1,25 \end{array}$$

n est aussi le nombre d'instants considérés.

Nous allons montrer que cette matrice M_n comporte n valeurs propres distinctes comprises entre 2,25 et 0,25.

Considérons la matrice $B_n = -2 M_n + 4,5 I$, B_n est la matrice à n lignes et n colonnes :

$$\begin{array}{cccccccc} 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & -1 & 2 \end{array}$$

Remarquons que M_n et B_n ont les mêmes vecteurs propres et que les valeurs propres sont liées par la relation : $\mu_p = -0.5 \lambda_p + 2.25$.

Il nous suffit de trouver les valeurs propres et les vecteurs propres de la matrice B_n pour connaître les valeurs propres et les vecteurs propres de la matrice M_n .

Lemme : La matrice B_n possède n valeurs propres distinctes tel que si $p \in [1, 2, 3, \dots, n]$ alors la valeur propre de rang p est $\mu_p = 2 \left(1 - \cos \left(\frac{p\pi}{n+1} \right) \right)$ et le vecteur propre de rang p est $f_p = \left(\sin \left(\frac{p\pi}{n+1} \right); \sin \left(\frac{2p\pi}{n+1} \right); \sin \left(\frac{3p\pi}{n+1} \right); \dots; \sin \left(\frac{np\pi}{n+1} \right) \right)$

Démonstration :

Nous allons montrer que les valeurs propres de la matrice B_n sont $\mu_p = 2 \left(1 - \cos \left[\frac{p\pi}{n+1} \right] \right)$ et que les vecteurs propres sont $f_p = \left\{ \sin \left(\frac{kp\pi}{n+1} \right) \right\}_{k \in [1, 2, \dots, n]}$ pour p prenant les valeurs de 1 à n .

Soit $f = (x_1, x_2, x_3, \dots, x_{n-1}, x_n)$ un vecteur propre associé à une valeur propre μ . La relation $(B_n - \mu I_n) f = 0$ se traduit par le système d'équations suivant :

$$\begin{array}{rcl} (2 - \mu) x_1 & - & x_2 & = & 0 \\ -x_1 & (2 - \mu) x_2 & - & x_3 & = & 0 \\ & -x_2 & (2 - \mu) x_3 & - & x_4 & = & 0 \\ & \dots & \dots & \dots & \dots & \dots & \dots \\ & & & & -x_{n-1} & + (2 - \mu) x_n & = & 0 \end{array}$$

Posons $x_0 = 0$ et $x_{n+1} = 0$ la famille $\{x_k\}_{k \in [0, 1, \dots, n+1]}$ vérifie la relation $-x_k + (2 - \mu)x_{k+1} - x_{k+2} = 0$ pour tout $k \in [0, 1, \dots, n+1]$. Recherchons les solutions de la forme $e^{i\theta k}$, la relation ci dessus s'écrit :

$$-e^{i\theta k} + (2 - \mu)e^{i\theta(k+1)} - e^{i\theta(k+2)} = e^{i\theta k} [1 + (2 - \mu)e^{i\theta} - (e^{i\theta})^2] = 0$$

On en déduit que $e^{i2\theta} - (2 - \mu)e^{i\theta} + 1 = 0$ car $e^{i\theta k}$ n'est jamais nul. Posons $X = e^{i\theta}$, on doit trouver des valeurs de θ telles que l'équation caractéristique du système $X^2 - (2 - \mu)X + 1 = 0$ soit satisfaite.

Nous devons en premier résoudre une équation du second degré dont le discriminant est :

$$\Delta = (2 - \mu)^2 - 4.$$

Remarquons que $e^{i2\theta} - (2 - \mu)e^{i\theta} = -1$ entraîne :

$\cos 2\theta - (2 - \mu) \cos \theta = -1$ et $\sin 2\theta - (2 - \mu) \sin \theta = 0$. Sachant que $\cos 2\theta = \cos^2 \theta - \sin^2 \theta$ on déduit de la première relation: $\cos 2\theta + 1 = 2 \cos^2 \theta = (2 - \mu) \cos \theta$ ou encore $2 - \mu = 2 \cos \theta$ d'où $\Delta = -4 \sin^2 \theta$. Le discriminant étant négatif l'équation caractéristique admet deux solutions $X_1 = \frac{2 \cos \theta + 2i \sin \theta}{2} = e^{i\theta}$ et $X_2 = \frac{2 \cos \theta - 2i \sin \theta}{2} = e^{-i\theta}$

Revenons à notre système d'équation; toute combinaison linéaire de la puissance k ème $(\alpha e^{i\theta k} + \beta e^{-i\theta k}, \alpha$ et β complexes), des solutions de l'équation caractéristique sont solution du système d'équation initial de plus $x_0 = 0$ et $x_{n+1} = 0$ donc :

$$\alpha + \beta = 0 \quad \alpha e^{i\theta(n+1)} + \beta e^{-i\theta(n+1)} = 0$$

1. De la première relation on en déduit que $x_k = \gamma \sin k\theta$. Comme on désire garder uniquement les solutions réelles il faut que γ soit réel et non nul.
2. La seconde égalité s'écrit $x_{n+1} = \gamma \sin [(n+1)\theta] = 0$ et si γ n'est pas nul alors $\theta = \frac{p\pi}{n+1}$ pour $p \in [1, 2, 3, \dots, n]$ (la valeur 0 et $n+1$ ne sont pas admises car alors $\mu = 0$).

Pour tout $p \in [1, 2, 3, \dots, n]$ nous pouvons donner les éléments caractéristiques de la matrice B_n

- les valeurs propres sont $\mu_p = 2 \left(1 - \cos \left(\frac{p\pi}{n+1} \right) \right)$
- une des vecteurs propres correspondant est $f_p = \left\{ \sin \left(\frac{kp\pi}{n+1} \right) \right\}_{k \in [1, 2, \dots, n]}$

Pour terminer la démonstration il suffit de constater que toutes les valeurs propres λ_p trouvées sont différentes et au nombre de n .

Les valeurs propres et les vecteurs propres de M_n s'en déduisent directement :

- les valeurs propres sont $\lambda_p = \cos \left(\frac{p\pi}{n+1} \right) + 1.25$
- les vecteurs propres correspondants sont $f_p = \left\{ \sin \left(\frac{kp\pi}{n+1} \right) \right\}_{k \in [1, 2, \dots, n]}$

Difficultés d'interprétation de ce résultat :

Ce résultat très simple permet de représenter les différents vecteurs propres normés en fonction de la valeur propre choisie et du nombre d'observations n .

Faisons deux remarques :

Remarque N°1 :

Le nombre de valeurs propres augmentent en fonction du nombre d'instants considérés et elles sont comprises entre 2.25 et 0.25, elle sont donc de plus en plus proche si le nombre d'instants considérés croît. Nous verrons que cette situation est très défavorable lorsque on désire faire de la prévision.

Remarque N°2 :

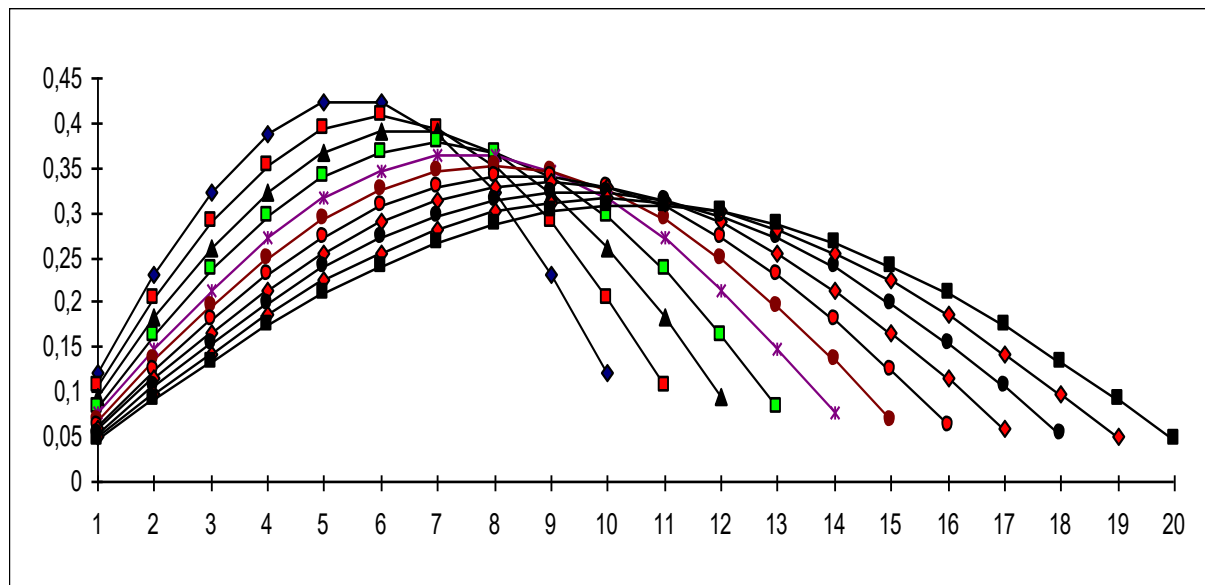
Si on désire comparer le vecteur propre de rang k d'un processus en fonction de la variation du nombre d'instants n où il est connu, on peut avoir deux représentations possible des résultats:

- Les processus sont connus en $t_0, t_0 + 1, t_0 + 1, t_0 + 2, t_0 + 3, \dots, t_0 + k, \dots$, c'est à dire à partir d'une date et au fur et à mesure du temps.
- Les processus sont connus à l'instant présent t_0 et dans les instants passés $t_0 - 1, t_0 - 2, t_0 - 3, t_0 - 4, \dots, t_0 -$

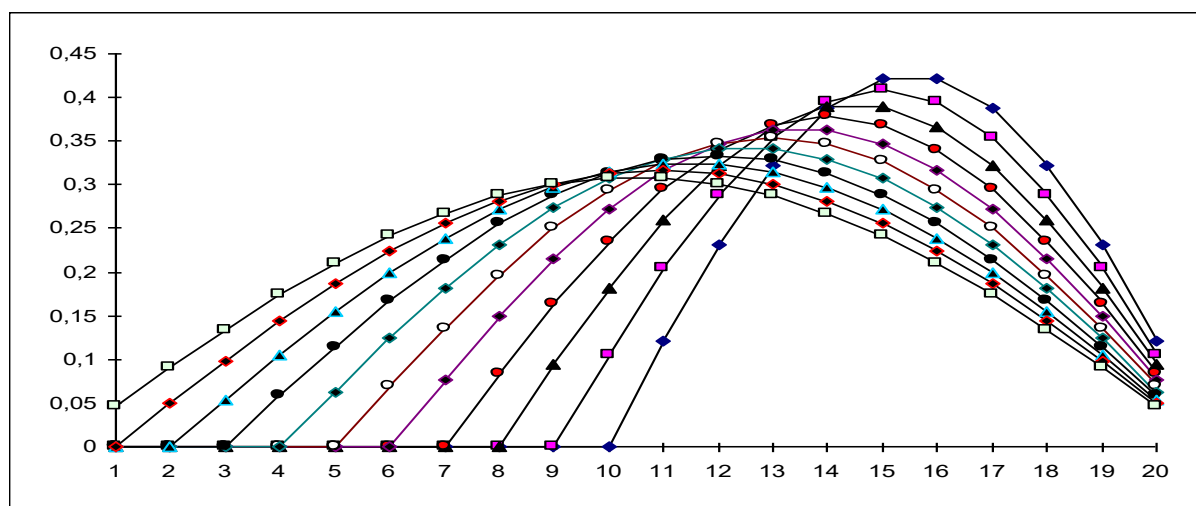
Dans ces deux situations les matrices considérées sont les mêmes mais la représentation graphique n'est pas la même. Pour illustrer ces remarques nous allons

représenter la première fonction propre de 10 à 20 instants dans les deux cas évoqués précédemment.

Premier cas le processus est connu à partir d'une date donnée et on représente le premier vecteur propre en fonction du nombre d'instants écoulés



Second cas: le processus est connu dans le passé et on « remonte » plus ou moins dans le temps

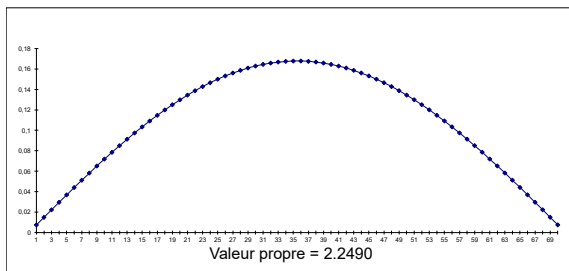


Ces deux exemples montrent les difficultés de tout vocabulaire imprécis et la nécessité de définir correctement les périodes ou les dates mises en jeux.

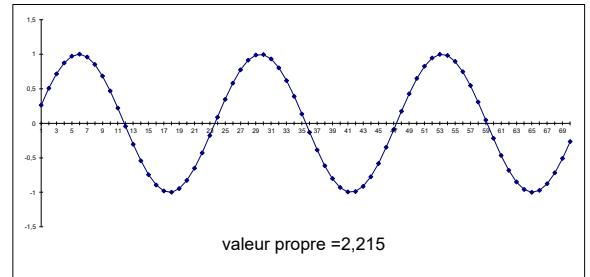
Examinons dans la page suivante les 10 premiers vecteurs propres pour un MA(1) étudié sur 70 instants.

Les 10 premiers vecteurs propres d'un MA(1) de coefficient 0.5 calculés sur 70 instants

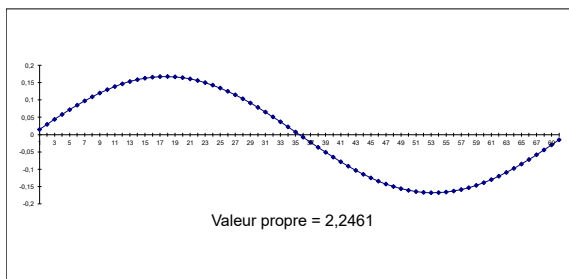
Vecteur propre N°1



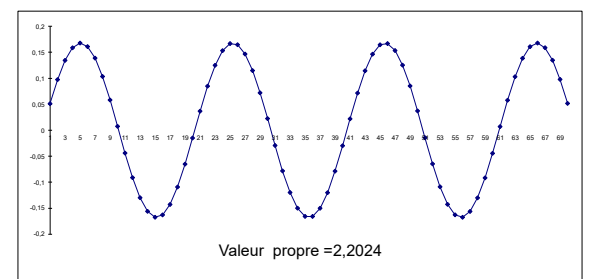
Vecteur propre N°6



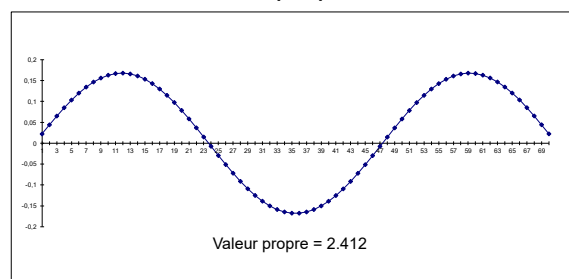
Vecteur propre N°2



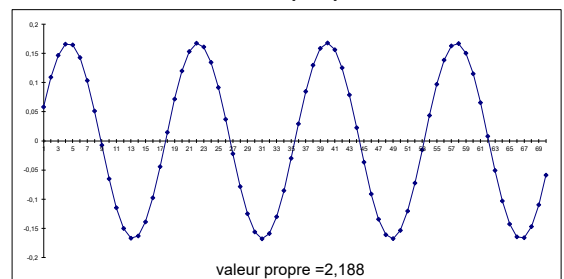
Vecteur propre N°7



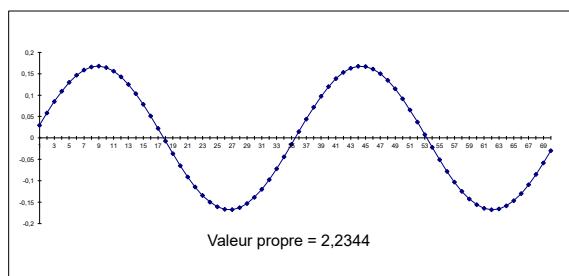
Vecteur propre N°3



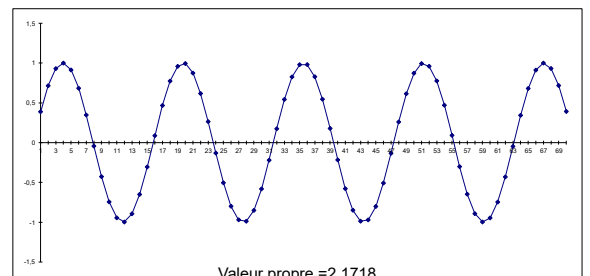
Vecteur propre N°8



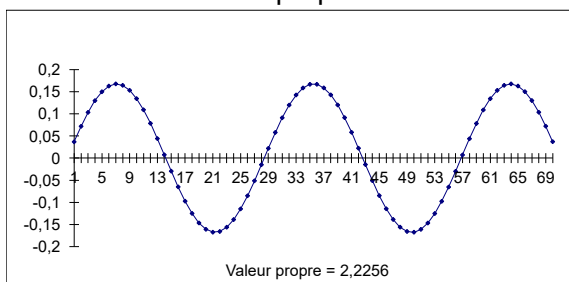
Vecteur propre N°4



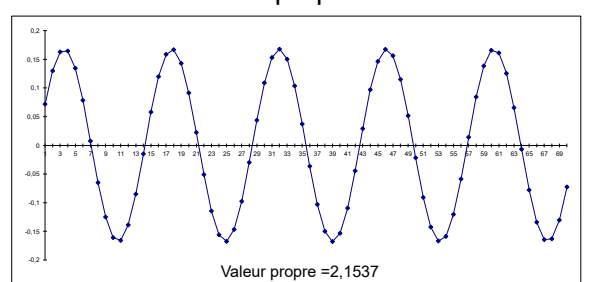
Vecteur propre N°9



Vecteur propre N°5



Vecteur propre N°10



I-2-bba) le modèle MA(1) avec θ strictement compris entre 1 et -1

Notons $M_n(\theta)$ la matrice carrée à n lignes et n colonnes définie par :

$$\begin{array}{cccccc} 1+\theta^2 & -\theta & 0 & 0 & 0 & \dots \\ -\theta & 1+\theta^2 & -\theta & 0 & 0 & \dots \\ 0 & -\theta & 1+\theta^2 & -\theta & 0 & \dots \\ 0 & 0 & -\theta & 1+\theta^2 & -\theta & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & -\theta & 1+\theta^2 & \dots \end{array}$$

Nous pouvons écrire que $M_n(\theta) = \theta B_n + [(1 + \theta^2) - 2\theta] I$ alors si μ_p est la valeur propre associée au vecteur propre x de B_n on a :

$$M_n(\theta)(x) = \theta B_n(x) + [(1 + \theta^2) - 2\theta] x = [\theta \mu_p + (1 + \theta^2) - 2\theta] x$$

Ce qui démontre que x est un vecteur propre de $M_n(\theta)$ de valeur propre :

$$\begin{aligned} \lambda_p &= \theta \mu_p + (1 + \theta^2) - 2\theta = 2\theta \left(1 - \cos\left(\frac{p\pi}{n+1}\right) \right) + (1 + \theta^2) - 2\theta \\ &= (1 + \theta^2) - 2\theta \cos\left(\frac{p\pi}{n+1}\right) \end{aligned}$$

En résumé les éléments caractéristiques de la matrice $M_n(\theta)$ sont :

- les valeurs propres $\lambda_p = 1 + \theta^2 - 2\theta \cos\left(\frac{p\pi}{n+1}\right)$
- les vecteurs propres correspondants $f_p = \left\{ \sin\left(\frac{kp\pi}{n+1}\right) \right\}_{k \in [1, 2, \dots, n]}$

Nous remarquerons que les vecteurs propres d'un MA(1) sont indépendants de la valeur de θ mais que les valeurs propres en dépendent.

Ces valeurs propres sont cependant positives et comprises entre :

$$\theta^2 + 2\theta + 1 \text{ et } \theta^2 - 2\theta + 1$$

La longueur de l'intervalle qui les contient est donc de 4 quelque soit la valeur de n .

1-2-b-c) Le modèle AR (Auto Régressif) :

Les processus auto régressifs sont une généralisation de la promenade aléatoire et sont défini par:

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \dots - \phi_p B^p) Y_t = e_t$$

que l'on note AR(p) ou ARMA(p,0)

Comme pour le modèle MA(q), où nous avons présenté le modèle MA(1) en détail, nous présenterons le modèle AR(1); c'est à dire le modèle $(1 - \phi B) Y_t = e_t$ qui s'écrit sous forme explicite : $Y_t - \phi Y_{t-1} = e_t$.

On montre que $-1 < \phi < 1$ et que les coefficients de la matrice de covariance $a_{ij} = \phi^{(|i-j|)}$ (voir Guy Mélard [25] pages 314).

Nous étudierons plus particulièrement le cas $\phi = 0,9$; la matrice de covariance qui est théoriquement infinie à gauche s'écrit:

$$\begin{array}{cccccc} 0,90 & 0,81 & 0,73 & 0,66 & 0,59 & 0,53 & 0,48 \\ 0,81 & 0,90 & 0,81 & 0,73 & 0,66 & 0,59 & 0,53 \\ 0,73 & 0,81 & 0,90 & 0,81 & 0,73 & 0,66 & 0,59 \\ 0,66 & 0,73 & 0,81 & 0,90 & 0,81 & 0,73 & 0,66 \\ 0,59 & 0,66 & 0,73 & 0,81 & 0,90 & 0,81 & 0,73 \\ 0,53 & 0,59 & 0,66 & 0,73 & 0,81 & 0,90 & 0,81 \\ 0,48 & 0,53 & 0,59 & 0,66 & 0,73 & 0,81 & 0,90 \end{array}$$

La recherche des vecteurs propres et des valeurs propres de ce type de matrice ne nous a pas semblé évidente. Nous avons donc traité le problème grâce au calculateur. Le théorème de Gerschgorin permet cependant de démontrer que les valeurs propres sont bornées par $2\phi / 1 - \phi$.

Rappel du théorème de Gerschgorin:

Si $A = (a_{ij})_{(i,j) \in [1,n]^2}$ est une matrice et si λ est un vecteur propre de A alors

$$\lambda \in \bigcup_{1 \leq i \leq n} \left\{ x \in \mathbb{R} / |x - a_{ii}| \leq \sum_{i \neq j} |a_{ij}| \right\}$$

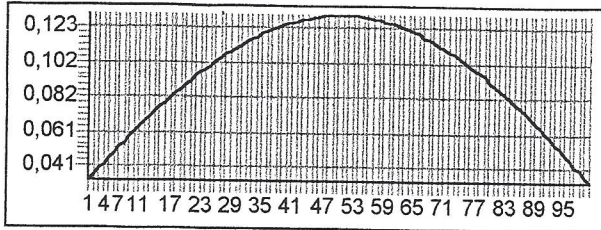
si on applique ce théorème à la matrice d'un AR(1) on a :

$$|\lambda - 1| \leq 2 \sum_{i=1}^{n-1} \phi^i \leq 2 \phi \sum_{i=0}^{\infty} \phi^i = 2 \frac{\phi}{1 - \phi}$$

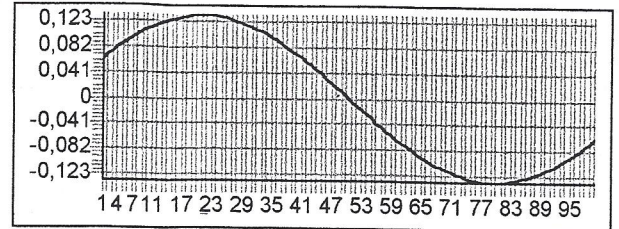
Nous présentons dans les pages suivantes les graphes des vecteurs propres classés par rang de valeurs propres décroissantes pour 100 instants

⇒ Graphe des premiers vecteurs d'un AR(1) avec $\phi = 0.9$ sur 100 instants

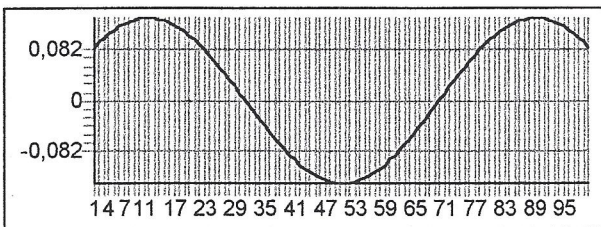
Vecteur propre n°1
valeur propre $\lambda_1 = 17.8$



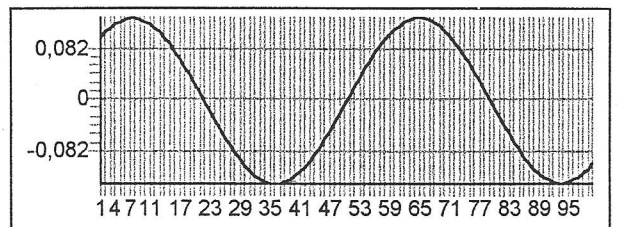
Vecteur propre n°2
valeur propre $\lambda_2 = 15.1$



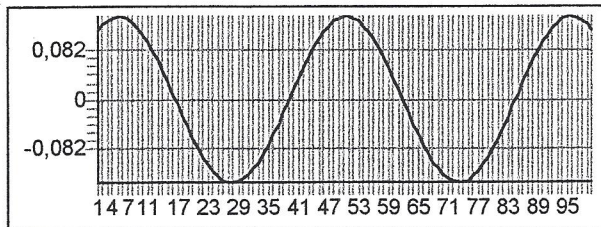
Vecteur propre n°3
valeur propre $\lambda_3 = 11.9$



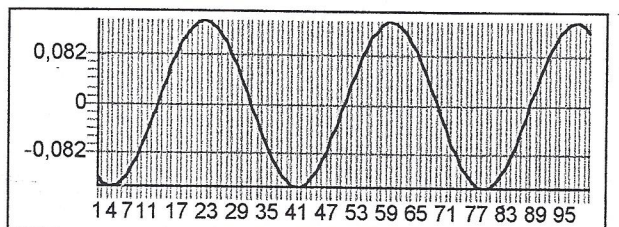
Vecteur propre n°4
valeur propre $\lambda_4 = 9.1$



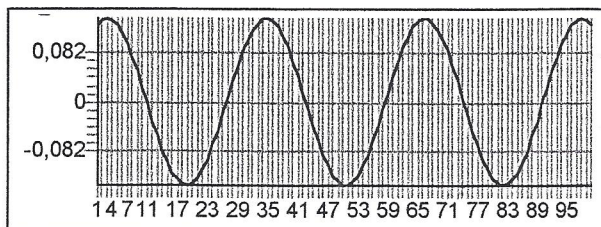
Vecteur propre n°5
valeur propre $\lambda_5 = 7.0$



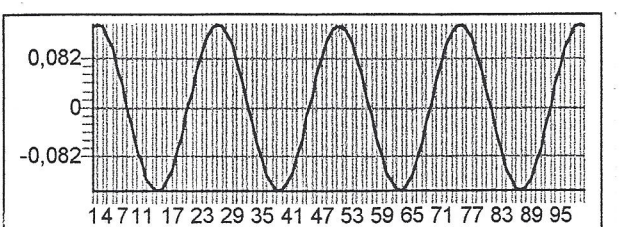
Vecteur propre n°6
valeur propre $\lambda_6 = 5.4$



Vecteur propre n°7
valeur propre $\lambda_7 = 4.2$



Vecteur propre n°8
valeur propre $\lambda_8 = 3.3$



On remarquera que toutes les valeurs propres sont inférieures à $2 \times 0.9 / (1-0.9) + 1 = 19$

I-3) LES MÉTHODES D'ANALYSE UTILISANT LA CONNAISSANCE THÉORIQUE DU NOYAU DE COVARIANCE

Dans le paragraphe précédent nous avons recherché les valeurs propres et les vecteurs propres des processus discrets du type ARIMA à l'aide du calcul ou à défaut à l'aide du calculateur. La recherche théorique des vecteurs propres et des valeurs propres associées est forcément plus satisfaisante mais pas toujours facile.

Dans les situations concrètes la connaissance du processus se fait à l'aide de relevés statistiques. La recherche d'un modèle sensé représenter ces données n'est pas toujours une démarche qui se justifie. Cela nous encline à penser que dans la majorité des situations concrètes il faut recourir à un calculateur.

Si le processus étudié est un processus continu la situation est comparable. Dans le cas général on utilisera un calculateur pour déterminer les valeurs propres et les vecteurs propres du processus. Dans ce cas un problème supplémentaire se posera car les données sont discrètes et qu'il faudra justifier l'approximation du processus par ces données.

Certains cas particuliers de processus continus permettent de trouver analytiquement l'expression des vecteurs propres et des valeurs propres. C'est le cas des processus de Poisson. Les processus de Poisson continus sont des processus définis sur un intervalle T réel qui satisfont à l'hypothèse d'accroissement stationnaire. C'est à dire que pour tout h , $X_{t+h} - X_t$ est stationnaire.

Dans ce cas le noyau de covariance prend une forme très simple:

$$N(t,t') = \min(t,t')$$

Nous allons présenter ce cas classique en reprenant les résultats de l'article de 1985 de Gilbert Saporta cité en bibliographie [27], puis nous comparerons les vecteurs propres et les valeurs propres d'un processus de Poisson défini sur l'intervalle $T = [0,T]$ puis $T' = [0,T+h]$. L'intérêt de cette comparaison trouvera son expression lors de la prévision des processus à l'aide de la décomposition harmonique. Comme nous l'avons vu les vecteurs propres d'un processus (par exemple d'un MA) ne sont pas stables lorsque l'intervalle change. Cette constatation nous amènerait à penser que la connaissance des vecteurs propres d'un processus n'apporte aucun élément exploitable dans le cadre d'une prévision. La réalité est plus complexe et les cas que nous avons rencontrés font plutôt partie des exceptions que des généralités. Nous allons vérifier que dans le cas d'un processus de Poisson les vecteurs propres associés à l'intervalle T et ceux associés à l'intervalle T' sont proches sur l'intervalle T . En d'autres termes la connaissance des vecteurs propres sur T permettra la prévision des vecteurs propres associés à T' . Or connaître les vecteurs propres sur T' c'est connaître le processus sur T' . Ainsi nous serons capables de prévoir le processus à l'horizon h .

I-3-a) Spectre d'un processus de Poisson :

Le noyau de covariance d'un processus de Poisson U défini sur $[0,T]$ est :
(G.Saporta [26] pages 46 et 47)

$$N(t, t') = \min(t, t')$$

Si f est un vecteur propre de U de valeur propre λ on a:

$$U(f)(t) = \int_0^t s f(s) ds + \int_t^T t f(s) ds = \lambda f(t)$$

Après deux dérivations on a l'équation différentielle:

$$\lambda f''(t) + f(t) = 0$$

qui admet pour solutions :

$$\lambda_i(T) = \frac{4T^2}{(2i-1)\pi^2} \quad \text{et} \quad f_{i,T}(t) = \sqrt{\frac{2}{T}} \sin\left[\left(i - \frac{1}{2}\right)\pi \frac{t}{T}\right]$$

I-3-b) Comportement spectral d'un processus de Poisson

Nous allons montrer que l'écart entre deux fonctions propres calculées T et $T+h$ est uniformément négligeable.

Dans un premier temps posons :

$$\sqrt{\frac{2}{T+h}} = \sqrt{\frac{2}{T}} + \varepsilon(h) \quad \text{alors} \quad \Delta(h) = \left| f_{i,T}(t) - f_{i,T+h}(t) \right|$$

peut s'écrire:

$$\Delta(h) = \left| \sqrt{\frac{2}{T}} \sin\left[\frac{1}{2}\left(i - \frac{1}{2}\right)\pi \frac{t}{T}\right] - \left(\sqrt{\frac{2}{T}} + \varepsilon(h)\right) \sin\left[\frac{1}{2}\left(i - \frac{1}{2}\right)\pi \frac{t}{T+h}\right] \right|$$

ou encore

$$\Delta(h) = \left| \sqrt{\frac{2}{T}} \left(\sin\left[\frac{1}{2}\left(i - \frac{1}{2}\right)\pi \frac{t}{T}\right] \sin\left[\frac{1}{2}\left(i - \frac{1}{2}\right)\pi \frac{t}{T+h}\right] \right) + \varepsilon(h) \sin\left[\frac{1}{2}\left(i - \frac{1}{2}\right)\pi \frac{t}{T+h}\right] \right|$$

utilisons la formule élémentaire de trigonométrie

$$\sin a - \sin b = 2 \cos \frac{a+b}{2} \sin \frac{a-b}{2}$$

Alors :

$$\Delta(h) = \left| \sqrt{\frac{2}{T}} \left(\cos\left[\frac{1}{2}\left(i - \frac{1}{2}\right)\pi t \frac{2T+h}{(T+h)T}\right] - \sin\left[\frac{1}{2}\left(i - \frac{1}{2}\right)\pi t \frac{h}{(T+h)T}\right] \right) + \varepsilon(h) \sin\left[\frac{1}{2}\left(i - \frac{1}{2}\right)\pi \frac{t}{T+h}\right] \right|$$

$$\Delta(h) \leq \left| \sqrt{\frac{2}{T}} \sin \left[\frac{1}{2} \left(i - \frac{1}{2} \right) \pi t \frac{h}{T(T+h)} \right] \right| + |\varepsilon(h)|$$

Par exemple pour $i=1$, $T = 18$, $h=1$.

$$\Delta(h) \leq 0.01$$

quelque soit la valeur de t

Nous généraliserons ce résultat dans les chapitres ultérieurs pour les processus qui ne sont pas soumis à des contraintes d'accroissement stationnaire et nous vérifions sur des exemples que la majorité des processus concrets satisfont à ces conditions.

I-3-c) Vecteur propres associés à un processus de Poisson :

Nous représentons les différents éléments caractéristiques d'un processus de Poisson développés sur un intervalle $[0,20]$. Nous représentons dans un premier temps les valeurs propres. En observant les résultats il est raisonnable de garder entre 3 et 5 valeurs propres.

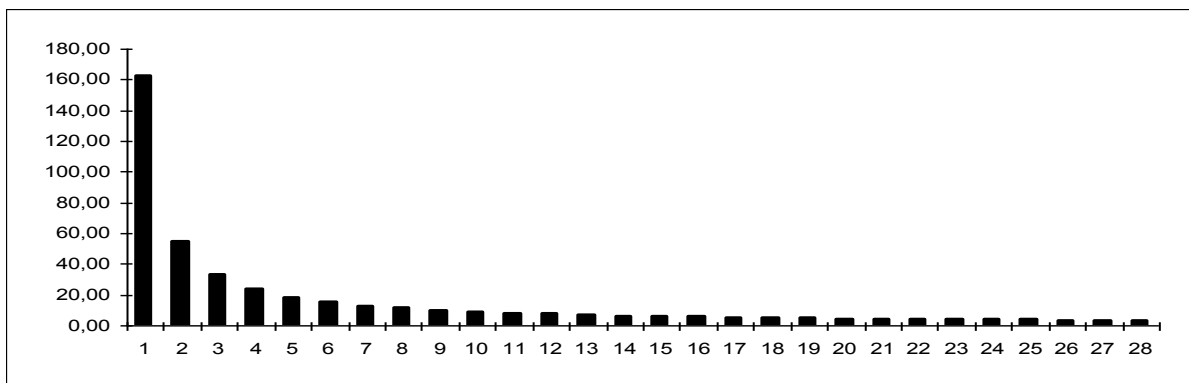
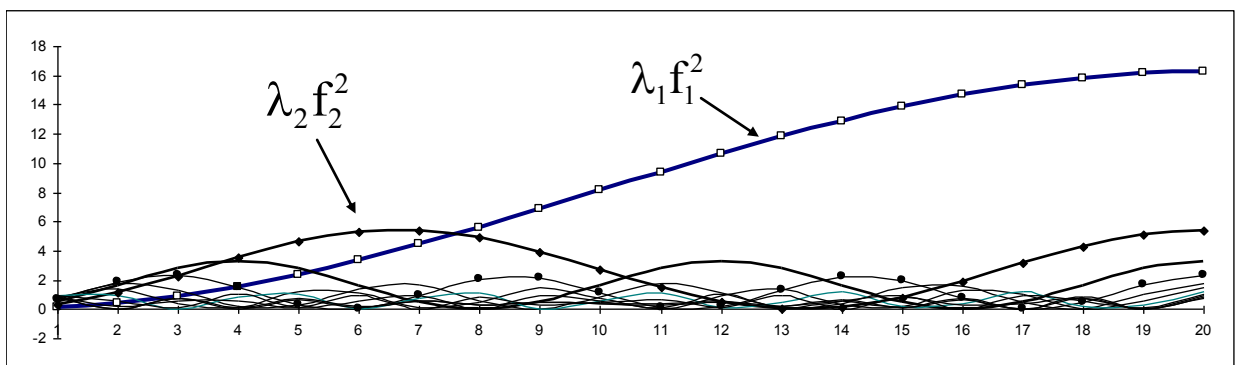


Diagramme des premières valeurs propres

Nous avons représenté ci dessous les variances de chaque composante temporelle c'est à dire les fonctions $\lambda_1 f_1^2$; $\lambda_2 f_2^2$; $\lambda_3 f_3^2$;



On observe la convergence uniforme vers « zéro » de la variance des composantes temporelles.

CHAPITRE II

DESCRIPTION DES FAMILLES DE TRAJECTOIRES A L'AIDE DE L'ANALYSE HARMONIQUE

Dans ce chapitre nous étudierons les outils adaptés à l'étude d'un ensemble de trajectoires considérées comme les réalisations d'un processus. Cette situation diffère fondamentalement du cas où on ne possède qu'une trajectoire. La moyenne du processus et le noyau de covariance ne nécessite pas pour être estimé une hypothèse de stationnarité. La recherche de modèles de processus du type ARIMA n'est plus nécessaire car on peut directement travailler dans l'ensemble de tous les processus, à condition de prendre les précautions d'usage pour que les différentes définitions puissent prendre sens.

II-1) LES NOTATIONS ADAPTÉES À CE TYPE D'ANALYSE

Toute étude de processus fait intervenir deux ensembles : « l'univers des possibles » ou « espace des individus » et celui du temps. Lorsqu'on étudie une seule série chronologique on est amené à considérer le processus défini par la famille des variables aléatoires $\{X_t\}_{t \in T}$. Comme on s'intéresse aux lois de chaque variable et à leurs liens, l'ensemble des événements Ω est sous-jacent mais n'est pas explicitement nommé en général. La définition complète doit être dans ce cas :

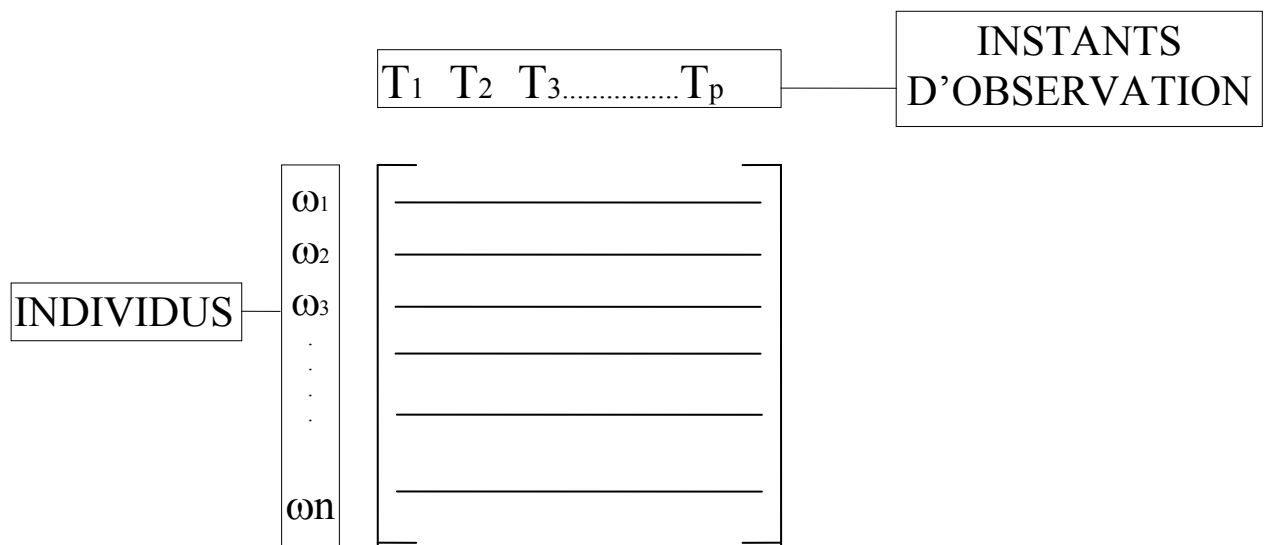
II-1-a) Définition adapté à l'études d'une série chronologique :

$(\Omega, \mathcal{A}, P)$ étant un espace probabilisé muni d'une tribu $\mathcal{A}$ de parties de Ω et d'une probabilité P , le processus X est une application mesurable vérifiant :

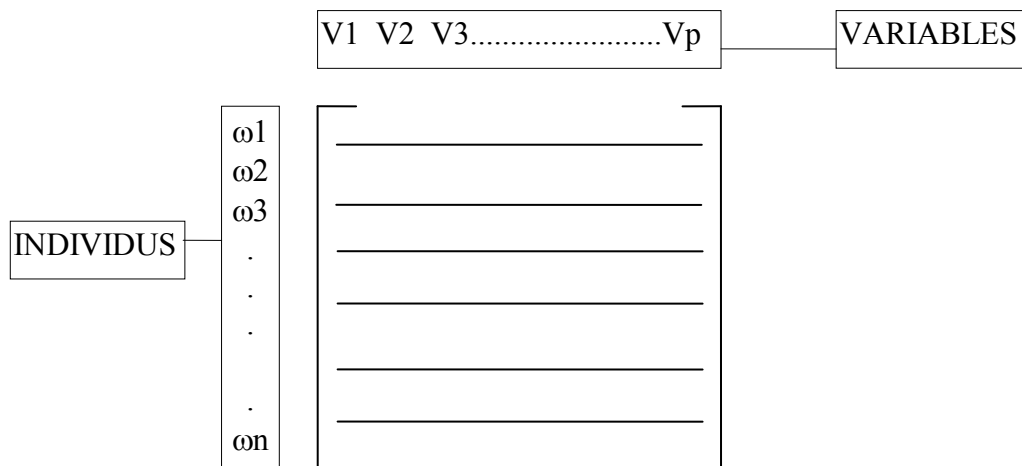
$$X: (\Omega, \mathcal{A}, P) \longrightarrow (\Psi, \mathcal{B})^{\otimes T} = (\Psi^T, \mathcal{B}^{\otimes T})$$

où Ψ est un ensemble muni de la tribu $\mathcal{B}$; dans notre cas $\Psi = \mathbf{R}$ (ou un intervalle T de $\mathbf{R}$) muni de la tribu habituelle

Lors de l'étude d'un ensemble de trajectoires dont on connaît expérimentalement un certain nombre de valeurs la situation n'est pas comparable car on dispose d'un tableau ayant la forme suivante :



Cette présentation est identique à celle utilisée en analyse des données où on range les informations dans un tableau du type suivant :



Un rôle d'égale importance est joué par l'espace des individus et l'espace des variables ou des instants. C'est pourquoi nous préférons, dans ce cas, utiliser la définition proposée par Gilbert Saporta ([28] page 123) :

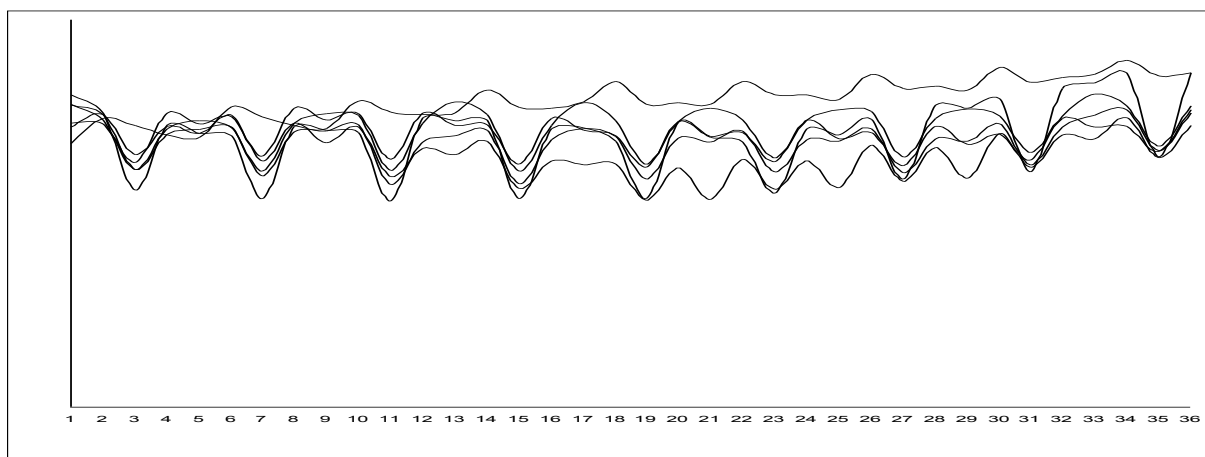
II-1-b) Définition adaptée à l'études des données engendrées par des séries chronologiques:

Un processus aléatoire est une application X mesurable et de carré intégrable définie sur $\Omega \times T$ par:

$$X: (\omega, t) \in \Omega \times T \longrightarrow X(\omega, t) \in \mathbb{R}$$

où $\Omega \times T$ est munie de la tribu produit de $(\Omega, \mathcal{A}, P)$ et de (T, μ) muni de la mesure de Lebesgue μ .

Cette présentation met en valeur les relevés d'un processus présentés sous forme d'un tableau mais ne met pas en valeur l'ensemble des trajectoires. Si le processus est présenté sous forme de trajectoires comme c'est le cas ci dessous il est naturel de définir un processus comme une application de l'ensemble des individus dans un ensemble de fonctions. C'est pourquoi nous proposons une nouvelle définition;



Ensemble des trajectoires d'un processus économique (données INSEE)

II-1-c) Définition adaptée à l'études d'un ensemble de trajectoires

Un processus X est une application de l'ensemble des individus $(\Omega, \mathcal{A}, P)$ dans l'ensemble des classes de fonctions de carré intégrable $L^2(T)$ vérifiant :

a) X est défini de la façon suivante :

$$X: \omega \in \Omega \longrightarrow X(\omega) \in L^2(T)$$

avec $[X(\omega)](t) = X(\omega, t)$ presque partout.

b) X , est intégrable au sens de l'intégrale de Bochner.

Remarque 1 :

L'intégrale de Bochner est une intégrale définie directement sur des fonctionnelles mesurables à valeur dans l'espace $L^2(T)$ (plus généralement dans un espace vectoriel topologique _E.V.T._).

De telles conditions, qui sont de fait peu restrictives, permettent sans précaution supplémentaire d'intégrer X , $X(\omega)$ par rapport à t , $X(t)$ par rapport à ω , ou $X(\omega, t)$ par rapport à ω et à t . Elle permettent aussi d'avoir tous les résultats dont on a besoin sur l'existence de la moyenne, de la variance, du noyau de covariance, et de l'opérateur associé (voir annexe IV)

Remarque 2

Dans la pratique ces trois définitions sont équivalentes c'est pourquoi nous ne prendrons pas de précautions particulières pour expliciter les outils que nous manipulerons. Nous ferons sans cesse des abus de langages en confondant les éléments des fonctions de carré intégrables $L^2(T)$ et ceux des classes correspondantes de l'ensemble $L^2(T)$.

La définition II-1-c) nous amène à définir la moyenne d'un processus, la variance, et la covariance de deux processus comme des classes de fonctions; c'est à dire des éléments de $L^2(T)$.

A nouveau, ces définitions sont équivalentes à des ensembles de mesure nulle près sur T aux définitions déjà données dans le chapitre précédent, mais elles apportent un nouveau regard sur les processus qui trouvera tout au long des démonstrations et des présentations graphiques sa justification.

Donnons ces définitions :

II-1-d) Moyenne d'un processus:

Définition :

La moyenne d'un processus X est la fonction notée $E(X)$ définie par:

$$E(X)(t) = \int_{\Omega} X(\omega, t) dP\omega$$

où $\int_{\Omega} X(\omega) dP\omega$ est une fonction de $L^2(T)$ qui prend pour presque tout t la valeur

$$\int_{\Omega} X(\omega, t) dP\omega$$

Remarque :

Si le cardinal de Ω est fini l'intégrale est une somme finie. Dans ce cas, qui est celui rencontré dans beaucoup de situations financières ou économiques on peut espérer que la fonction moyenne sera d'autant plus régulière que le nombre d'individus sera important; cette remarque aura un intérêt pratique lors de la prévision du processus.

II-1-e) Processus centré :

Définition :

On appelle processus centré associé à X le processus défini par.

$$\omega \longrightarrow X(\omega) - E(X)$$

Remarques :

- Dans la suite du texte nous supposons toujours que les processus considérés sont centrés.
- L'espérance mathématique d'un processus centré est la fonction nulle.

II-1-f) Variance d'un processus:

Définition :

La variance d'un processus X centré est la fonction $V(X)$ définie par :

$$V(X) = \int_{\Omega} X^2(\omega) dP\omega$$

$$\text{où} \quad [X^2(\omega)](t) = X^2(\omega, t) = (X(\omega, t))^2$$

Remarque :

La notation fonctionnelle permet de donner une définition simple de la variance d'un processus. Cette définition est une extension du mot variance car le résultat n'est pas un réel mais une fonction (ou une classe de fonction).

II-1-g) Noyau de Covariance:

Définition :

Si X est un processus centré le noyau de covariance est la quantité :

$$N(t,t') = \int_{\Omega} X(\omega, t) X(\omega, t') dP_{\omega}$$

Remarques:

- Un processus définit à chaque instant une variable aléatoire. Le noyau de covariance mesure le lien entre toutes ces variables. C'est lui qui décrit la déformation du processus au fur et à mesure de l'écoulement du temps. Il est la « mémoire » du processus. Si X est un processus défini sur un intervalle T divisé en deux intervalles consécutifs $T = T' + T''$. L'analyse du processus ne peut être considérée comme le résultat de l'analyse sur T' puis sur T'' sauf dans des cas très particuliers que nous étudierons. C'est cette observation qui fait toute la difficulté de l'étude des processus.
- Si on discrétise le temps ce noyau de covariance va devenir une matrice de covariance qui pourra être analysée à l'aide des méthodes de l'analyse des données.

II-1_h) Opérateur de covariance:

Définition :

L'opérateur de covariance est l'application linéaire U définie par :

$$U: f \in L^2(T) \longrightarrow U(f)$$

$$\text{avec} \quad U(f) = \int_T N(t,t') f(t') dt';$$

Remarque 1:

On remarquera que $U(f)$ est un élément de $L^2(T)$ (J.C Deville [15] page 19)

Remarque 2:

Lors d'une analyse de données la matrice de covariance est interprétée comme l'application linéaire dont les vecteurs propres sont les axes principaux. Il est naturel d'associer, par analogie, au noyau de covariance l'opérateur de covariance défini ci dessus

II-2) LA MÉTHODE DESCRIPTIVE D'UN ENSEMBLE DE SÉRIES CHRONOLOGIQUES À TRAVERS L'ANALYSE HARMONIQUE.

L'analyse harmonique se propose de décrire un processus en le décomposant suivant « la meilleur » base de l'espace Hilbertien $L^2(T)$.

Les valeurs propres et les vecteurs propres de l'opérateur de covariance ont des propriétés remarquables bien connus. Ces propriétés sont liées à trois types de raisons :

1. Des résultats analytiques très généraux liés aux propriétés communes à tous les opérateurs de Hilbert-Schmidt.
2. Un résultat sur les trajectoires, liés aux propriétés probabilistes de l'ensemble des individus et au rôle du noyau de covariance
3. Une propriété de convergence uniforme de la variance cumulée des processus quasi déterministe vers la fonction variance, liée à la continuité du noyau de covariance

Rappelons ces résultats.

1) Résultats analytiques très généraux :

L'opérateur de covariance est défini à l'aide du Noyau de covariance $N(t,t')$ qui est un élément de $\mathcal{L}^2(\Omega \times T, P \otimes dt)$ est un opérateur de Hilbert-Schmidt, d'où les résultats

- Le spectre de l'opérateur de covariance est réduit à une famille $\{\lambda_i\}$ dénombrable de valeurs propres.
- Les valeurs propres, non nulles sont séparées, réelles, bornées et classées par ordre décroissant.
- A chaque valeur propre non nulle est associé un sous espace propre de dimension 1.
- Le noyau de l'opérateur de covariance associé à la valeur propre nulle est de dimension finie.
- Si à chaque valeur propre non nulle λ_i (les valeurs propres multiples étant comptées avec leur ordre de multiplicité) on associe un vecteur propre f_i . Il est possible à l'aide de ces vecteurs propres de définir une famille orthogonale. Complétés par une base orthogonale du noyau, ils forment une base orthogonale de l'espace.
- Presque à tout instant on a les égalités: $V(X)(t) = N(t,t) = \sum_{i \in p} \lambda_i f_i^2(t)$

2) Résultat lié aux propriétés probabilistes :

Presque pour tout $\omega \in \Omega$, la projection de toute trajectoire $X(\omega)$ sur le noyau $\text{Ker}(U)$ de l'opérateur de covariance est nulle.

3) Propriété de convergence uniforme

Le théorème de Mercer permet d'approcher uniformément, dans le cas d'un noyau de covariance continu, la fonction variance par la série $\sum_{i>N} \lambda_i f_i^2$.

II-2-a) Rappel des principaux résultats sur les processus quasi déterministes en analyse harmonique :

II-2-aa) Définition d'un processus quasi déterministe:

Soit f une fonction et une variable aléatoire γ . Le processus X défini par :

$$X : (\omega, t) \longrightarrow \gamma(\omega) f(t)$$

est appelé processus quasi déterministe .

Réciproquement tout processus pouvant se décomposer en un produit d'une variable aléatoire et d'une fonction sera appelé processus quasi déterministe.

II-2-ab) Définition d'un processus élémentaire:

Soit X un processus et un vecteur propre de l'opérateur de covariance n'appartenant pas à son noyau. Nous appellerons processus élémentaire X_i associé à f_i le processus quasi déterministe défini par :

$$X_i : \omega \longrightarrow \langle X(\omega), f_i \rangle f_i$$

La décomposition optimale d'un processus à l'aide de processus quasi déterministe est basée sur le théorème fondamental suivant :

II-2-ac) Théorème fondamental

La norme de la projection des trajectoires de X sur N vecteurs orthogonaux est de variance maximale si les N vecteurs choisis sont les N premiers vecteurs propres de l'opérateur de covariance.

Plus explicitement :

Désignons P le projecteur sur un sous-espace engendré par N vecteurs.

$$V(\|P(X(\omega))\|) \leq \sum_{i=1}^N \lambda_i$$

où $\lambda_1 \geq \lambda_2 \geq \lambda_3 \dots \geq \lambda_N$ sont les valeurs propres associées aux vecteurs propres de l'opérateur de covariance. L'égalité ayant lieu lorsque le projecteur P est défini par $P(X(\omega)) = \sum_{i=1}^N \langle X(\omega), f_i \rangle f_i$

Ce résultat n'est que la généralisation du résultat fondamental de l'analyse des données. Il est dans le droit fil de la préoccupation de l'analyse spectrale qui se proposait à l'origine de généraliser les résultats sur les espaces vectoriels de dimension fini aux espaces vectoriels de fonctions.

II-2-b) Autres propriétés des vecteurs propres et des valeurs propres de l'opérateur associé au noyau de covariance :

Dans un premier temps donnons la définition d'un processus élémentaire et quelques propriétés.

Un processus élémentaire est un cas particulier d'un processus quasi déterministe; c'est à dire :

II-2-bb) Quelques propriétés utiles :

⇒ Si le processus donné est centré alors chaque processus élémentaire est centré.

La variance totale d'un processus élémentaire est la fonction : $\lambda_i f_i^2$

⇒ La covariance de deux processus élémentaires différents est nulle

$$\text{COV}(X_i, X_j) = 0$$

⇒ Presque pour tout $\omega \in \Omega$, la projection de toute trajectoire $X(\omega)$ sur le noyau $\text{Ker}(U)$ de l'opérateur de covariance est nulle.

⇒ Pour tout processus X est la somme de ses processus élémentaires :

$$X = \sum_{i \in N} X_i$$

Plus précisément, presque pour tout individu ω de Ω on a la décomposition suivante:

$$X(\omega) = \sum_{i \in N} \langle X(\omega), f_i \rangle f_i$$

où f_i parcourt l'ensemble des vecteur propres de l'opérateur de covariance excepté ceux appartenant à son noyau.

II-2-c) Convergence uniforme de la série $\sum_{i \in \mathbb{N}} \lambda_i f_i^2$: Rappel du théorème de Mercer

Si le noyau de covariance est continu sur $T \times T$ la série $\sum_{i \in \mathbb{N}} \lambda_i f_i^2$ converge uniformément vers la fonction variance

On peut utiliser ce théorème pour définir un indice de perte d'information lorsque on remplace le processus par une somme des premiers processus élémentaires.

Il suffit de poser :

$$Y(\omega) = \sum_{i \leq p} \langle X(\omega), f_i \rangle f_i$$

et

$$Z(\omega) = \sum_{i > p} \langle X(\omega), f_i \rangle f_i$$

Alors $X = Y + Z$ et $V(X) = V(Y) + V(Z)$.

Si le théorème de Mercer peut s'appliquer alors la quantité $V(Z)(t) = \sum_{i > p} \lambda_i f_i^2(t)$ peut être uniformément rendue aussi petit que l'on désire à condition que p soit assez grand.

Ce résultat permet dans la pratique de remplacer X par un nombre fini de processus quasi déterministes sous l'hypothèse de continuité du noyau de covariance. La continuité du noyau de covariance n'entraîne pas celle des trajectoires, mais celle des fonctions propres. Dans le cas de processus économiques cette continuité n'est cependant pas évidente. Si l'hypothèse de continuité du noyau de covariance était indispensable, l'analyse harmonique ne pourrait être utilisée qu'avec un nombre limité de processus. Heureusement, comme nous le montrerons, la convergence uniforme de la série $\sum_{i \in \mathbb{N}} \lambda_i f_i^2$ reste valable, même si le noyau de covariance n'est pas continu. Cela justifie le caractère très général des études faites à l'aide de ces techniques.

II-3) APPROXIMATION DISCRÈTE D'UN PROCESSUS ET AUTRES TRAVAUX SUR L'ANALYSE HARMONIQUE

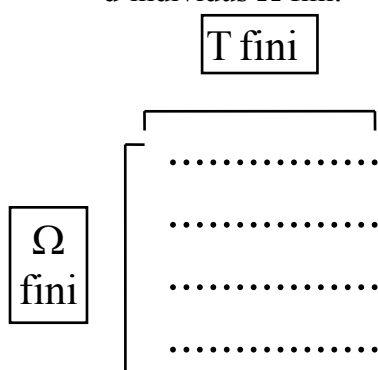
L'analyse harmonique des processus est né de l'enrichissement mutuel des savoir-faire de l'analyse des données, de l'algèbre linéaire, de l'analyse fonctionnelle et de l'analyse numérique. L'étude des processus, à l'aide des outils de l'analyse harmonique peut être regardée comme le prolongement, de l'ACP, du discret au continu.

L'idée d'utiliser l'analyse harmonique pour étudier les processus est relativement ancienne (cf Loeve 1945 [18]). La quantité de calculs nécessaires à été, comme en analyse des données, un frein aux recherches. Dès que les calculs ont été possibles des études concrètes ont été faites et les théories se sont développées .

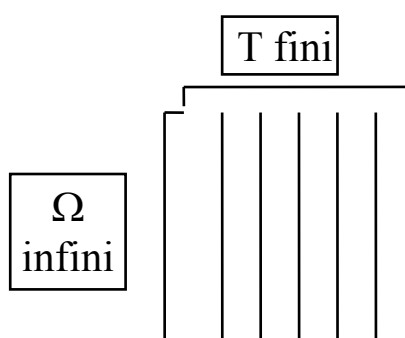
Ces études amenaient naturellement les chercheurs vers des questions de légitimité des calculs mais surtout des approximations . Dans quels cas peut on dire que la connaissance des processus sur un nombre discret d'instantns permet de connaître les éléments caractéristique du processus ? Après discrétisation du processus la matrice de covariance et ses éléments caractéristiques sont-ils une bonne approximation des éléments caractéristiques de l'opérateur de covariance?

Dès 1982, lors d'un congrès commun de la « Classification Society » et de la « Psychometric Society » au Canada Ramsay J.O., présentait les différentes analyses factorielles sous la forme de quatre diagrammes,.

Cas totalement discret:
nombre de variables T fini et nombre d'individus Ω fini.

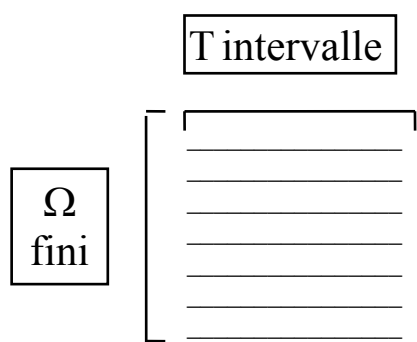


Cas mixte de type 1:
Nombre de variables T finies
et ensemble Ω infini.



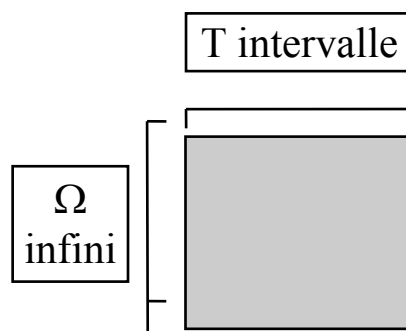
Cas mixte de type 2:

T intervalle et nombre d'individus Ω fini.



Cas général:

T intervalle et Ω infini.



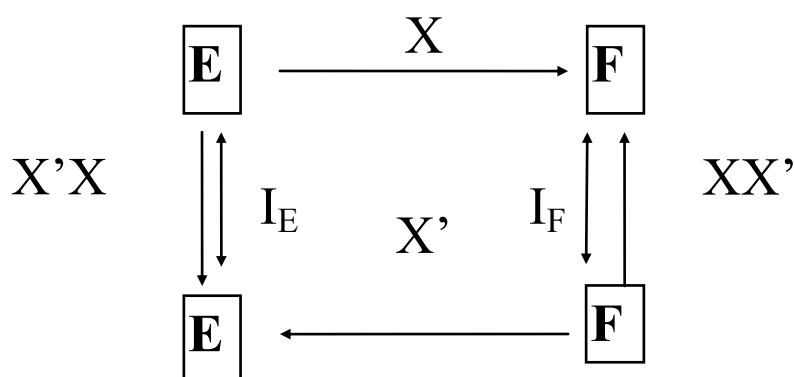
Les différentes études faites sur l'analyse harmonique d'un processus se sont orientées vers l'approximation des ensembles infinis par des ensembles discrets, seuls susceptibles d'être traités par informatique.

Le cas mixte de type 1 a été proposé par Y.Escoufier dans son étude sur l'hydrométrie (cf [17] 1970) Ω étant le bassin de l'Allier et T la hauteur de pluie

Le cas mixte de type 2 a été abordé par J.C. Deville (cf [15] 1974) dans son étude sur la constitution des familles. Ω étant l'ensemble des familles et T le temps du mariage.

Le cas général a été étudié à Toulouse par J. Dauxois et A. Pousse (cf [11] 1976) avec leur thèse qui annonce clairement, dans son titre, les objectifs : « Les analyses factorielles en calcul des probabilités et en statistique : essais de synthèse ». Puis dans la thèse de P.Besse (cf [5] 1979)

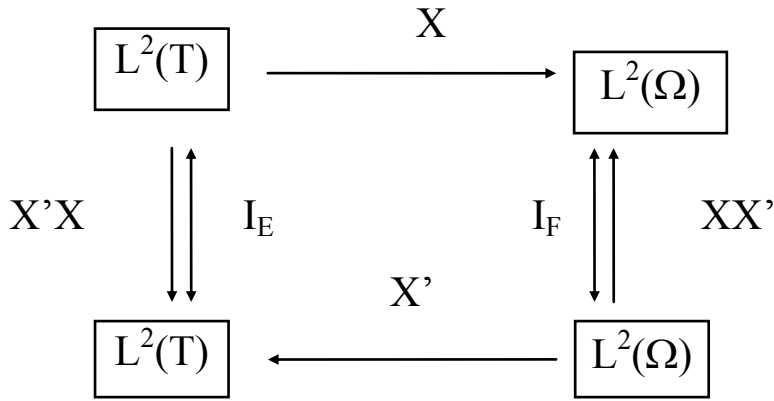
Dans le cas discret l'analyse des données, pour prendre en compte simultanément l'espace des individus et des variables présente le diagramme de dualité suivant.



Dans cette présentation E est l'espace des coefficients de combinaison des variables de dimension p s'il y a p variables , F est l'ensemble des combinaisons des individus n s'il y a n individus, X le tableau des données avec en ligne les individus et en colonne les variables.

Les travaux sur la recherche d'un cadre général d'études des processus à l'aide des outils de l'analyse des données ont naturellement pris en compte ce cadre de présentation en interprétant l'espace E (respectivement F) comme représentant l'espaces $L^2(T)$ (respectivement $L^2(\Omega)$) des fonctions de carré intégrable défini sur T (respectivement Ω)

Ce qui donne le diagramme suivant :



où X est l'application de $L^2(T)$ dans $L^2(\Omega)$ définie par :

$$X(f)(\omega) = \int_T X(\omega, t') f(t') dt'$$

et X' l'application de $L^2(\Omega)$ dans $L^2(T)$ définie par :

$$X(Y)(t) = \int_{\Omega} X(\omega, t) Y(\omega) dP\omega$$

$X'X$ est alors l'opérateur de covariance. Remarquons au passage que XX' est l'opérateur d'Escoufier.

L'introduction de métriques sur l'espace des individus et le désir d'approcher le processus par des processus discrets complexifie le diagramme de base en faisant appel à différentes théories d'analyse fonctionnelle ou numérique (dual topologiques, espaces de Sobolev, distributions).

Nous retiendrons deux résultats qui nous semblent significatifs :

Dauxois et Pousse (cf [11] page 266) ont montré que l'ACP approchée d'un processus X , dont on connaît une réalisation sur un échantillon de taille n converge uniformément vers l'ACP de X lorsque la taille de l'échantillon croît indéfiniment.

Besse démontre ([cf [5] page 61) que si la fonction

$$t \longrightarrow V(t) = N(t, t)$$

est continue alors l'ACP d'un processus X défini sur un intervalle T est approchée uniformément par sa restriction discrète sur T , à condition que la discrétisation de T soit convenablement choisie

De toutes ces études, il en résulte, dans des conditions très générales, que le praticien ayant choisi convenablement une discrétisation des ensembles Ω et T , l'ACP de X peut être sans trop de danger remplacée par celle de sa discrétisation. Le calcul matriciel fait habituellement pour définir les valeurs propres et les vecteurs propres associés au processus est donc légitime.

Notre travail ne s'est pas situé exactement dans le prolongement des études que nous venons de présenter. Nous avons admis les résultats d'approximation des processus par des données discrètes. Nous avons plutôt essayé de préciser les relations entre les différents éléments de l'analyse harmonique des processus lorsque l'intervalle de temps varie.

CHAPITRE III

LES METHODES DE PREVISION D'UNE SERIE CHRONOLOGIQUE

III-1) LES METHODE CLASSIQUES DE PREVISION

III-1-a) Les méthodes prévision usuelles:

III-1-aa) Les méthodes élémentaires

Ces méthodes utilisent des modèles simples et déterministes. La prévision espérée est obtenue en prolongeant le modèle d'une façon mécanique.

La prévision à l'aide de la régression linéaire ou polynomiale consiste à prolonger la droite ou le polynôme en utilisant son équation. Cette méthode, par sa simplicité, reste souvent une bonne solution lorsque les données sont celles d'une seule série chronologique sans saisonnalité ou stationnarité évidente.

III-1-ab) Les méthodes de décomposition d'une série chronologique en tendances et saisonnalité

Si on a une série décomposée en tendance et saisonnalité on les prolonge pour réaliser la prévision.

⇒ La tendance se prolonge à l'aide d'une régression linéaire ou polynomiale.

⇒ On effectue la prévision définitive en corrigeant la tendance à l'aide du coefficient saisonnier

Dans ces méthodes on abandonne l'information contenue dans le résidu, c'est à dire la différence entre la série reconstituée grâce à la tendance et la saisonnalité et les données d'origine. Cette attitude se justifie si le résidu ne comporte plus d'information, c'est à dire s'il peut être considéré comme la réalisation d'un bruit blanc.

III-1-b) Méthode de lissage exponentiel simple

Cette méthode de prévision est très adaptée aux séries chronologiques courtes qui ne contiennent pas de tendance ou de saisonnalité.

L'idée est simple. On se propose de prédire la série de proche en proche en utilisant pour prédire la valeur de la série au temps $t+1$ une moyenne pondérée entre la dernière valeur connue x_t et la dernière prévision y_t :

$$y_{t+1} = \alpha x_t + (1-\alpha) y_t$$

ou α est un nombre compris entre 0 et 1 et $y_0 = 0$

On peut ainsi définir la prévision au rang t par la formule :

$$y_t = \alpha ((1-\alpha)^0 x_t + (1-\alpha)^1 x_{t-1} + (1-\alpha)^2 x_{t-2} + (1-\alpha)^3 x_{t-3} + \dots + (1-\alpha)^{n-1} x_1 + (1-\alpha)^n x_0)$$

Pour déterminer la constante de lissage on minimise la quantité :

$$\sum_{t=1}^n (y_t - x_t)^2$$

qui est un polynôme en α de degré pair.

On peut utiliser d'autres critères (Guy Mélard [25] page 144)

Cette méthode de lissage exponentiel est à rattacher au processus de type ARIMA (0,1,1) pour la prévision (Michel Tenenhaus [30] page 336).

III-1-b) Autres méthodes de lissage exponentiel:

Les autres méthodes de lissages utilisent les mêmes idées de base que le lissage exponentiel simple mais en appliquant ces idées à des séries ayant une tendance et/ou une saisonnalité. Deux méthodes sont à citer :

- ◇ Le lissage de Holt qui accepte les séries chronologiques contenant une tendance (Michel Tenenhaus [30] page 337).
- ◇ Le lissage de Winter qui généralise la notion de lissage pour toutes les séries chronologiques comportant tendance, saisonnalité et résidu (Michel Tenenhaus [30] page 340). C'est cette méthode que nous utilisons lors de prolongement automatiques de processus élémentaires dans le logiciel professionnel que nous proposons.

III-2) LA PREVISION DES PROCESSUS STATIONNAIRES ET LA MÉTHODE DE BOX-JENKINS

La méthode de prévision proposée par Box et Jenkins est basée sur la modélisation de processus stationnaires à l'aide des processus de la famille ARMA.

Tout processus du type ARMA(p,q) s'écrit de la façon suivante :

$$X_t = \sum_{i=1}^p \varphi_i X_{t-i} + \sum_{i=1}^q \psi_i e_{t-i} + e_t$$

Si on désire prévoir à l'instant t+1 la valeur prise par une série de données $\{x_t, x_{t-1}, x_{t-2}, x_{t-3}, \dots, x_{t-p}\}$ on est amené à rechercher le modèle ARMA(p,q) dont une réalisation sera la dite série :

$$X_{t+1} = \sum_{i=1}^p \varphi_i X_{t-i+1} + \sum_{i=1}^q \psi_i e_{t-i+1} + e_t$$

puis écrire que :

$$x_{t+1} = \sum_{i=1}^p \varphi_i x_{t-i+1} + \sum_{i=1}^q \psi_i \tilde{e}_{t-i+1} \quad (1)$$

Pour que le calcul (1) soit possible il faut connaître la valeur exacte de :

1. La série $\{\varphi_i\}_{i \in [1,2,\dots,p]}$
2. La série $\{\psi_i\}_{i \in [1,2,\dots,q]}$
3. La valeur $\{\tilde{e}_i\}_{i \in [1,2,\dots,q]}$ des réalisations du bruit blanc

La prévision d'une série chronologique de variance constante et de moyenne nulle comporte trois phases :

- ⇒ La détermination du type de processus ARIMA à sélectionner
- ⇒ L'estimation des valeurs des paramètres du processus
- ⇒ Les valeurs de la réalisation du bruit blanc associé à ces données

III-2-a) La détermination du type de processus ARMA à sélectionner :

Dans ce type de prévision on ne dispose que d'une série chronologique comme données. On suppose que ces données sont une réalisation d'une partie d'un processus stationnaire discret. Cette hypothèse permet d'estimer les autocorrélations. Pour estimer l'autocorrélation $\rho_1, \rho_2, \dots, \rho_k$ on utilise les valeurs de $r_1, r_2, \dots, r_k$ définies par :

$$r_1 = \frac{\sum_{t=2}^n x_t x_{t-1}}{\sum_{t=1}^n x_t^2} \quad ; \quad r_2 = \frac{\sum_{t=3}^n x_t x_{t-2}}{\sum_{t=1}^n x_t^2} \quad ; \dots\dots\dots; \quad r_k = \frac{\sum_{t=k+1}^n x_t x_k}{\sum_{t=1}^n x_t^2}$$

La série des données doit être « assez longue » pour que les estimations des autocorrélations soient possibles en fonction du décalage obligatoire dans les calculs.

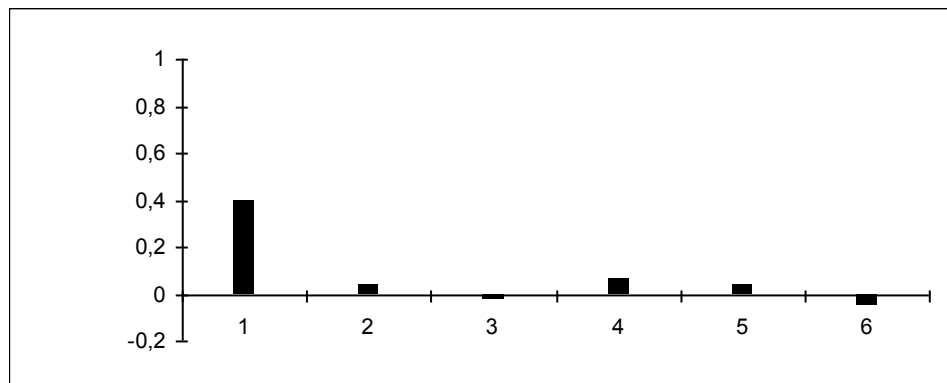
III-2-aa) Identification de processus MA

Si le processus est du type ARMA(0,q) (moyenne mobile « pur ») il s'écrit:

$$X_t = \sum_{i=1}^p \theta_i X_{t-i} + \sum_{i=1}^q \psi_i e_{t-i} + e_t = \sum_{i=1}^q \psi_i e_{t-i} + e_t$$

La covariance de X_t et de X_{t+k} est nulle dès que $k > q$. On en déduit que les autocorrélations sont nulles. En conséquence si les estimations des autocorrélations ne sont significativement différentes de zéro à partir d'un rang q , on en déduira qu'il existe un MA(q) dont une des réalisation est l'ensemble des données.

Ainsi il apparaît nécessaire d'estimer l'ensemble des autocorrélations $\rho_1, \rho_2, \dots, \rho_k$ et de les représenter par un diagramme en bâton. Ce diagramme est appelé autocorrélogramme simple



corrélogramme d'un MA(1) de la forme $X_t = 0.5 e_{t-1} + e_t$

Pour plus de précision on pourra consulter Guy Mélard [25] page 305 et suivantes ou Tenenhaus [30] page 293 et 294 .

Dans le cas où le corrélogramme comporte, quelque soit le rang, des valeurs nettement différentes de zéro on est amené à chercher une autre façon d'identifier le processus.

III-2-ab) Identification de processus AR

Supposons que le processus cherché soit un autorégressif (ARMA (p,0)) son écriture est de la forme $X_t = \sum_{i=1}^p \varphi_i X_{t-i} + e_t$ Le bruit blanc peut être considéré comme non corrélé avec les processus X_{t-h} .

Calculons le coefficient de corrélation de X_t avec X_{t-h} , $\rho_h = \text{cov}(X_t, X_{t-h}) / \sigma^2 = \text{cov}(\sum_{i=1}^p \varphi_i X_{t-i} + e_t, X_{t-h}) / \sigma^2$ où σ^2 est la variance estimée. On déduit les relation suivantes.

$$\begin{aligned} \rho_1 &= \varphi_1 1 + \varphi_2 \rho_1 + \varphi_3 \rho_2 + \dots + \varphi_p \rho_{p-1} \\ \rho_2 &= \varphi_1 \rho_1 + \varphi_2 1 + \varphi_3 \rho_2 + \varphi_4 \rho_3 + \dots + \varphi_p \rho_{p-2} \\ \rho_3 &= \varphi_1 \rho_2 + \varphi_2 \rho_1 + \varphi_3 1 + \varphi_4 \rho_1 + \dots + \varphi_p \rho_{p-1} \\ &\dots \\ &\dots \\ \rho_p &= \varphi_1 \rho_{p-1} + \varphi_2 \rho_{p-2} + \varphi_3 \rho_{p-3} + \dots + \varphi_p 1 \end{aligned}$$

Sur ces relations on peut faire deux remarques :

1. Les coefficients φ_i pour i compris entre 1 et p sont les coefficients de régression multiple du processus X_t sur $X_{t-1}, X_{t-2}, X_{t-3}, \dots, X_{t-p}$
2. Ces coefficients peuvent être déterminés en résolvant le système d'équation ci dessus.

Le dernier coefficient φ_n s'appelle coefficient de corrélation partiel.

En pratique on ignore si le processus est autorégressif ainsi que la valeur de p . C'est pour quoi on procède de la façon suivante :

On vérifie d'abord que les données peuvent être considérées comme la réalisation d'un processus stationnaire puis pour k prenant les valeurs successives 1,2,3,... on considère la régression linéaire de X_t sur $X_{t-1}, X_{t-2}, X_{t-3}, \dots, X_{t-k}$.

La processus X_t s'écrit alors $X_t = \sum_{i=1}^k \varphi_i X_{t-i} + R_{k,t}$ où $R_{k,t}$ est un processus non corrélé avec la famille des processus décalés $X_{t-1}, X_{t-2}, X_{t-3}, \dots, X_{t-k}$.

En utilisant un raisonnement identique à celui fait ci-dessus on obtient la famille d'équation suivante appelées équations de Yule-Walker :

$$\begin{aligned}
\rho_1 &= \varphi_{1,k} 1 + \varphi_{2,k} \rho_1 + \varphi_{3,k} \rho_1 + \dots + \varphi_{k,k} \rho_{p-1} \\
\rho_2 &= \varphi_{1,k} \rho_1 + \varphi_{2,k} 1 + \varphi_{3,k} \rho_1 + \varphi_{4,k} \rho_2 + \dots + \varphi_{k,k} \rho_{p-2} \\
\rho_3 &= \varphi_{1,k} \rho_2 + \varphi_{2,k} \rho_1 + \varphi_{3,k} 1 + \varphi_{4,k} \rho_1 + \dots + \varphi_{k,k} \rho_{p-1} \\
&\dots \\
&\dots \\
&\dots \\
&\dots \\
&\dots \\
\rho_p &= \varphi_{1,k} \rho_{p-1} + \varphi_{2,k} \rho_{p-2} + \varphi_{3,k} \rho_{p-3} + \dots + \varphi_{k,k} 1
\end{aligned}$$

Le terme $\varphi_{k,k}$ s'appelle aussi coefficient de corrélation partielle. Si le processus est un processus autorégressif AR(p) le coefficient de corrélation partielle $\varphi_{k,k}$ sera nul si $k > p$. Dans la pratique il y aura une certaine valeur de k à partir de laquelle on pourra considérer que toutes les corrélations partielles sont nulles.

La représentation graphique de la suite $\varphi_{k,k}$ est appelée corrélogramme partiel du processus. Pour plus de précision voir Tenenhaus [30], page 287 à 290.

III-2-ac) Cas général :

III-3-aca) Processus stationnaires

Pour étudier un processus stationnaire on regarde le corrélogramme simple et le corrélogramme partiel. Si l'un des deux corrélogrammes s'annule on est dans le cas d'un MA ou d'un AR. Dans le cas contraire on est en présence d'un ARMA(p,q) ou p et q sont non nuls. Il n'y a pas dans ce cas de solution automatique.

On est amené à transformer le processus d'origine pour se ramener à un cas précédent (type AR ou MA) pour cela on étudie :

- la série des différences premières $\{Z_t, Z_{t-1}, Z_{t-2}, Z_{t-3}, \dots, Z_{t-n-2}\} = \{X_t - X_{t-1}, X_{t-1} - X_{t-2}, X_{t-2} - X_{t-3}, X_{t-3} - X_{t-4}, \dots, X_{t-n-2} - X_{t-n-1}\}$
- Dans le cas où cette nouvelle étude ne donne pas de résultats satisfaisants on essaie des différences premières pondérées $Z_t = X_t - \varphi_1 X_{t-1}$ avec des valeurs particulières jusqu'à ce qu'on trouve un ARMA.
- Si cette transformation ne donne pas de résultats satisfaisants on essaie alors une différence d'ordre deux ou une combinaison de différences d'ordre deux et d'ordre un, du type :

$$Z_t = X_t - \varphi_1 X_{t-1} - \varphi_2 X_{t-2}$$

En général on trouve ainsi le type de modélisation du processus cherché.

Si les données comportent une périodicité évidente de longueur « s » on définit la nouvelle série des données par $z_t = x_t - x_{t-s}$ et on étudie la série des données z_t . On obtient ainsi une série stationnaire puis on recommence la même méthode que ci dessus:

1. spécification ou identification
2. estimation
3. adéquation

(voir G.Mélard [25] page 348)

III-2-ac) Processus non stationnaires

Dans le cas où la série des données n'est pas stationnaire on recherche des transformations qui la rende stationnaire c'est à dire de moyenne nulle, de variance contante et de coefficient de corrélation stables par translation. Puis on procède de la même façon que ci dessus.

III-2-b) L'estimation des valeurs des paramètres du processus

Deux solutions peuvent être employées :

III-2-ba) Méthode par approche successive

A l'aide de plusieurs essais on est arrivé à connaître la partie autorégressive du processus on peut donc écrire que $Z_t = \sum_{i=1}^q \psi_i e_{t-i} + e_t = X_t - \sum_{i=1}^p \phi_i X_{t-i}$ Dans un premier temps on suppose que Z_t est un MA(1). On réalise autant de fois que nécessaire le test du bruit blanc sur le processus $Z_t - \psi_1 e_{t-1}$ en faisant varier la valeur ψ_1 ; si on a des résultats satisfaisants on a ainsi déterminé la valeur de tous les coefficients. Sinon on fait l'hypothèse que le processus Z_t est un MA(2) et on recommence les tests.

III-2-bb) Méthode du maximum de vraisemblance

La méthode consiste à remplacer les essais successifs par l'optimisation d'une fonction de vraisemblance.

Le processus $X_t = \sum_{i=1}^p \phi_i X_{t-i} + \sum_{i=1}^q \psi_i e_{t-i} + e_t$ vérifie la relation

$$Z_t = \phi(B)X_t = \psi(B)e_t \text{ où } B \text{ désigne l'opérateur de décalage arrière et } \phi = \sum_{i=1}^p \phi_i B_i$$

$$\text{et } \psi = \sum_{i=1}^q \psi_i B_i$$

On cherche à estimer les coefficients φ_i et ψ_i . Pour cela on va utiliser la méthode du maximum de vraisemblance

On fait l'hypothèse que les bruits blancs suivent une loi normale $N(0, \sigma)$. On peut écrire X_t comme étant somme infinie de bruit blancs $X_t = \varphi(B)^{-1} \psi(B) e_t = \sum_{i=1}^{\infty} \beta_i e_{t-i} + e_t$ la loi de X_t est de ce fait multi-normale.

Les paramètres β_i peuvent s'écrire en fonction de φ et ψ c'est à dire des paramètres φ_i et ψ_i .

En utilisant cette décomposition et en appliquant la méthode du maximum de vraisemblance on peut obtenir la meilleure estimation des paramètres φ et ψ ; cette méthode est proposée par différents logiciels dont le logiciel SAS. Pour plus de prévisions se reporter à Thenenhaus [30] page 305 et 306

III-2-c) La dernière phase de la prévision :

Il ne nous reste qu'à estimer les valeurs $\{\tilde{e}_t\}_{i \in [1, 2, \dots, q]}$ des réalisations du bruit blanc et de finir le calcul.

III-2-ca) Estimation des valeurs $\{\tilde{e}_{t-i}\}_{i \in [1, 2, \dots, q]}$

La série des données transformée $(x_t, x_{t-1}, x_{t-2}, \dots, x_{t-n+1})$ est considérée comme une réalisation du processus $X_t = \sum_{i=1}^p \varphi_i X_{t-i} + \sum_{i=1}^q \psi_i e_{t-i} + e_t$

qui peut s'écrire à l'instant $t+1$ de prévision $X_{t+1} = \sum_{i=1}^p \varphi_i X_{t-i+1} + \sum_{i=1}^q \psi_i e_{t-i+1} + e_t$

La réalisation des données nous permet d'écrire que l'estimation de la prévision $\tilde{x}_{t+1} = \sum_{i=1}^p \varphi_i x_{t-i+1} + \sum_{i=1}^q \psi_i \tilde{e}_{t-i+1} + \tilde{e}_t$. Pour effectuer le calcul il ne nous manque qu'à estimer les différentes valeurs des $\tilde{e}_{t-i+1}$ pour i prenant la valeur 1 à q , la valeur de $\tilde{e}_t$ étant estimée à zéro.

Pour effectuer ces derniers calculs on utilise la méthode dite de prévision arrière (backforecasting) Tenenhaus [30], page 329 et 330

III-2-cb) Estimation de la valeur prévue :

La détermination de la prévision se fait en utilisant la formule

$$\tilde{x}_{t+1-h} = \sum_{i=1}^p \varphi_i x_{t-i+1} + \sum_{i=1}^q \psi_i \tilde{e}_{t-i+1}$$

On estime l'erreur faite en calculant les résidus par la

formule : $\bar{e}_{t-h} = x_t - \tilde{x}_t$. Si l'estimation du modèle a été correctement fait on est en présence de la réalisation d'un bruit blanc.

CHAPITRE IV

LA PREVISION ET L'ANALYSE HARMONIQUE

IV-1) ADAPTATION DE L'ANALYSE HARMONIQUE AUX PROBLÈMES DE LA PRÉVISION D'UN PROCESSUS:

Nous avons vu que l'analyse harmonique permet de décrire un ensemble de trajectoires interprétées comme la réalisation d'un processus. Le processus est estimé par la matrices des données $\{ X(\omega, t) \}$ où ω parcourt l'ensemble des individus et t l'ensemble des instants observés. Les fonctions propres et les valeurs de l'opérateur de covariance sont les éléments principaux de cette analyse. Les valeurs propres indiquent le « poids » du vecteur propre correspondant dans l'analyse. La variation des vecteurs propres (croissance, périodicité,), apportent l'information sur l'évolution du processus en fonction du temps. Les coefficients de projection de chaque trajectoire du processus marquent l'influence de chaque composante temporelle sur la réalisation étudiée. Dans la mesure où on travaille sur des processus centrés il ne faut pas oublier l'information apportée par la fonction moyenne.

Précisons le vocabulaire de l'analyse harmonique en fonction des remarques ci-dessus et de l'analogie avec l'Analyse en Composantes Principales. Par définition, le couple (fonction propre / valeur propre) s'appellera composante temporelle du processus . Nous parlerons de la composante temporelle de rang 1 (respectivement de rang 2,3,4,.....,p) pour désigner une des deux fonctions propres normées associées à la première valeur propre (respectivement 2,3,4,.....,p valeur propre).

Pour que la description du processus, proposé par l'analyse harmonique soit meilleure que l'observation directe des trajectoires, il faut que le nombre de composantes temporelles soit nettement moins important que le nombre total de trajectoires. Remarquons en premier que l'analyse harmonique est liée à la diagonalisation de la matrice de covariance qui dépend, dans sa dimension, du nombre d'instant p pris en compte et non du nombre n des individus (nombre de trajectoires).

Malgré cette observation, si la prise en compte de l'ensemble des composantes temporelles étaient obligatoires, l'analyse harmonique perdrait de son intérêt !!!

Au paragraphe **II_2_c)** nous avons rappelé qu'il est possible de remplacer le processus X par la somme Y des p processus élémentaires de telle façon que le résidu Z , différence entre X et Y , soit uniformément négligeable. Cette nouvelle observation permet de réduire ainsi le nombre de composantes temporaires à observer. Dans les exemples concrets ce nombre se réduit à 4 à 7 suivant les cas. Il ne faut pas oublier que ce dernier résultat est une conséquence du théorème de Mercer et suppose la continuité du noyau de covariance. Nous reviendrons plus en détail sur cette observation en montrant dans le cas de l'analyse harmonique, que les résultats sur l'approximation uniforme du processus restent valables sans hypothèse de continuité.

En conclusion :

Le traitement d'un processus par l'analyse harmonique simplifie notablement le nombre d'informations à analyser au regard du nombre de données brutes; pour nous en persuader il suffit de compter sur un exemple:

Si les données d'un processus comportent 30 périodes d'enregistrement et 100 réalisations; le nombre de données est de $30 \times 100 = 3\,000$ (dans l'étude de J.C. Deville un échantillon de 98 795 femmes avait été traité sur une période de 20 ans). Si l'information principale peut être décrite par 7 fonctions propres et la projection de chaque réalisation sur ces vecteurs propres, le nombre de données est alors de $7 \times 100 = 700$. Un gain important dans la simplification de l'étude.

Le passage de la description à la prévision d'un processus à l'aide de l'analyse harmonique se fait apparemment d'une façon naturelle en prolongeant les composantes temporelles (c'est à dire dans notre exemple 7 composantes temporelles) puis en reconstituant chaque trajectoire. L'apparente simplicité de la méthode cache une difficulté théorique importantes. Pour que la technique de prolongement proposée ait un sens il faudrait que les composantes temporelles calculées sur l'intervalle $[a,b+h]$ aient leurs restrictions sur l'intervalle $[a,b]$ identiques aux composantes temporelles calculées sur ce dernier intervalle. Cela est malheureusement connu comme faux car les noyaux de covariance calculés sur les intervalles $[a,b]$ $[a,b+h]$ n'ont pas de relations simples. On pourrait imaginer que les composantes temporelles calculées sur les intervalles $[a,b]$ et $[a,b+h]$ n'aient aucun lien entre elles. Nous allons montrer que si cette situation peut se produire théoriquement, dans la pratique, on peut affirmer que les premières composantes temporelles ont la « même forme ». Cette situation favorable se produira si les valeurs propres sont « bien séparées » entre elles.

En résumé l'analyse harmonique permet de faire la prévision d'un processus en prolongeant un nombre réduit de ses composantes temporelles

Nous allons préciser et démontrer les affirmations ci-dessus dans les paragraphes suivants.

IV-1-a) Notations

Soient $a < b < c$ trois réels on note $T=[a,c]$, $\tilde{T}=[a,b]$, $\bar{T}=[b,c]$.

Si f est une fonction définie sur T on note $\tilde{f}$ (respectivement $\bar{f}$) la fonction égale à f sur $\tilde{T}$ (respectivement $\bar{T}$) et nulle ailleurs.

On remarque alors que $f = \tilde{f} + \bar{f}$.

Soit X un processus aléatoire défini sur $\Omega \times T$, $N(t,t')$ son noyau de covariance et U l'opérateur de covariance associé .

On définit $\tilde{X}$ (respectivement $\bar{X}$) sur T par $\tilde{X}(\omega, t) = X(\omega, t)$
(respectivement $\bar{X}(\omega, t) = X(\omega, t)$) pour tout $t \in \tilde{T}$ (respectivement $t \in \bar{T}$) et zéro ailleurs.

On définit alors comme d'habitude : $\tilde{N}(t, t'), \bar{N}(t, t'), \tilde{U}, \bar{U}$, les noyaux de covariance et les opérateurs associés à $\tilde{X}$ et à $\bar{X}$.

Le premier théorème que nous allons démontrer indique que la restriction de toute fonction nulle sur l'intervalle $\bar{T}$ a la même image par l'opérateur U ou par l'opérateur $\tilde{U}$.

IV-1-b) Théorème n°1

Pour tout $g \in L^2(T)$ on a :

$$(\forall t, t \in T)(U(\tilde{g})(t) = \tilde{U}(\tilde{g})(t))$$

Démonstration :

En effet,

$$\begin{aligned} U(\tilde{g})(t) &= \int_T N(t, t') \tilde{g}(t') dt' = \\ &= \int_{\tilde{T}} N(t, t') \tilde{g}(t') dt' + \int_{\bar{T}} N(t, t') \tilde{g}(t') dt' = \\ &= \int_T \tilde{N}(t, t') \tilde{g}(t') dt' = \tilde{U}(\tilde{g})(t) \end{aligned}$$

car $\tilde{g}(t)$ est nulle sur $\bar{T}$ et, $N(t, t') = \tilde{N}(t, t')$ sur $\tilde{T} \times \tilde{T}$; ce qui termine la démonstration.

Le théorème n° 2 évalue l'écart entre l'image d'une fonction par U et l'image de sa restriction par l'opérateur $\tilde{U}$;

IV-1-c) Théorème n°2

Pour tout $g \in L^2(T)$ et pour tout $t \in \tilde{T}$ on a :

$$U(g)(t) = \tilde{U}(\tilde{g})(t) + v(t)$$

où

$$v(t) = \int_{\bar{T}} N(t, t') \bar{g}(t') dt'$$

Démonstration :

Calculons :

$$\begin{aligned} U(g)(t) &= \int_T N(t, t') g(t') dt' = \int_{\tilde{T}} N(t, t') \tilde{g}(t') dt' + \\ &\int_{\bar{T}} N(t, t') \bar{g}(t') dt' = \tilde{U}(\tilde{g})(t) + v(t) \end{aligned}$$

ce qui termine la démonstration

Les résultats du théorème suivant étant acquis nous serons en mesure de démontrer le théorème de Principal

IV-1-d) Théorème n°3

Si f est un vecteur propre de U , presque pour tout $t \in \tilde{T}$ on a :

$$U(f)(t) = \lambda \sum_{i \in N} \langle \tilde{f}, g_i \rangle g_i(t)$$

où g_i est une base de vecteurs propres de $L^2(\tilde{T})$

Démonstration:

Si f est un élément de $L^2(T)$ et si $t \in \tilde{T}$ alors $\tilde{f}$ est un élément de $L^2(\tilde{T})$ et $f(t) = \tilde{f}(t)$; de ce fait :

$$f(t) = \tilde{f}(t) = \sum_{i \in N} \langle \tilde{f}, g_i \rangle g_i(t) = \sum_{i \in N} \langle f, g_i \rangle g_i(t)$$

$$\text{car } \langle \tilde{f}, g_i \rangle = \langle f, g_i \rangle$$

$$\text{donc : } U(f)(t) = \lambda f(t) = \lambda \tilde{f}(t) = \sum_{i \in N} \lambda \langle \tilde{f}, g_i \rangle g_i(t)$$

Le théorème principal qui suit donne une idée du comportement des vecteurs propres de l'opérateur U en fonction des vecteurs propres de l'opérateur $\tilde{U}$; dit autrement, ce théorème situe les vecteurs propres du futur en fonction de ceux du passé. Si f est un vecteur propre de U et λ la valeur propre associée à f , plus λ sera éloignée d'une valeur propre λ_i de $\tilde{U}$ associée à g_i plus les fonctions propres f et g_i seront dans des espaces orthogonaux.

Nous verrons dans les théorèmes suivants que cette situation est plus fréquente que cela ne pourrait paraître surtout sur les premiers vecteurs propres de U .

IV-1-e) Théorème principal :

Soit f un vecteur propre de U et $\{g_i\}$ une base de vecteurs propres de $\tilde{U}$; alors pour tout $i \in N$ on a si λ est différent de λ_i :

$$\left| \langle \tilde{f}, g_i \rangle \right| \leq \frac{\|v\|}{|\lambda - \lambda_i|}$$

où λ est la valeur propre associée à f , $\{\lambda_i\}$ la famille de valeurs propres associées à la famille $\{g_i\}$ différentes de λ et v défini par

$$v(t) = \int_{\tilde{T}} N(t, t') \tilde{f}(t') dt'$$

Démonstration :

Si $t \in \tilde{T}$ calculons

$$U(f)(t) = \tilde{U}(\tilde{f})(t) + v(t) = \tilde{U} \left[\sum_{i \in N} \langle \tilde{f}, g_i \rangle g_i \right](t) + v(t)$$

$$= \sum_{i \in N} \lambda_i \langle \tilde{f}, g_i \rangle g_i + v(t) = \sum_{i \in N} \langle \lambda_i \tilde{f} + v, g_i \rangle g_i(t)$$

d'après le théorème N° 3

$$U(f)(t) = \sum_{i \in N} \lambda_i \langle \tilde{f}, g_i \rangle g_i(t) = \sum_{i \in N} \langle \lambda_i \tilde{f}, g_i \rangle g_i(t)$$

De ce fait on a l'égalité

$$\sum_{i \in N} \langle \lambda_i \tilde{f}, g_i \rangle g_i(t) = \sum_{i \in N} \langle \lambda_i \tilde{f} + v, g_i \rangle g_i(t)$$

comme $\{g_i\}$ est une base, pour tout i on a :

$$\langle \lambda_i \tilde{f}, g_i \rangle = \langle \lambda_i \tilde{f} + v, g_i \rangle$$

ou encore

$$(\lambda - \lambda_i) \langle \tilde{f}, g_i \rangle = \langle v, g_i \rangle$$

d'après l'inégalité de Cauchy-Schwarz :

$$|\langle v, g_i \rangle| \leq \|v\| \|g_i\| = \|v\|$$

d'où la relation :

$$|\lambda - \lambda_i| |\langle \tilde{f}, g_i \rangle| \leq \|v\|$$

La démonstration est terminée.

Le Théorème suivant précise les conditions du théorème principal. La majoration fait intervenir le maximum du noyau de covariance ainsi que la quantité. $\|\tilde{f}\|_1 \mu(\tilde{T})^{1/2}$ où μ représente la mesure de Lebesgue sur l'intervalle T .

On pourrait s'attendre à ce que cette quantité fasse intervenir le temps écoulé lors de la première analyse par rapport à celui écoulé globalement. En d'autres termes il est « naturel » d'espérer que plus le processus est connu sur une période longue, plus la prévision est sûre. Cette affirmation est à la fois juste et fausse. Elle est fausse car rien n'interdit de penser que le processus, à partir d'une certaine date, fait intervenir des composantes temporelles, qui ne s'étaient pas « exprimées », prenant des valeurs faibles en début de processus et des valeurs fortes en fin. Elle est vraie car toutes les composantes temporelles étant normées il est raisonnable de penser que :

$$\|\tilde{f}\|_1 \approx \mu(\bar{T}) / \mu(T) \quad \text{et que} \quad \mu(\tilde{T}) \approx \mu(T) \quad \text{donc que} \quad \|\tilde{f}\|_1 \mu(\tilde{T})^{1/2} \approx \mu(\bar{T}) / \mu(\tilde{T})^{1/2}$$

Cette dernière remarque montre, s'il était encore nécessaire, que les résultats énoncés dépendent intimement des données.

IV-1-f) Théorème

Avec les notations précédentes si le noyau de covariance est borné avec

$M = \sup_{(t,t') \in T \times T} [N(t,t')] ;$ si de plus on suppose que pour tout $i \in N$ on a $\lambda \neq \lambda_i$ alors

$$\left| \langle f, g_i \rangle \right| \leq \frac{M \|f\|_1 \mu(\tilde{T})^{1/2}}{|\lambda - \lambda_i|}$$

Démonstration:

Il suffit de démontrer que $\|v\| \leq M (1 - \|\sqrt{f}\|^2)^{1/2} \mu(\tilde{T})^{1/2}$

Calculons :

$$\begin{aligned} \|v\|_{\tilde{T}}^2 &= \int_{\tilde{T}} \left[\int_{\tilde{T}} N(t,t') f(t') dt' \right]^2 dt \leq \int_{\tilde{T}} \left[\int_{\tilde{T}} |N(t,t') f(t')| dt' \right]^2 dt \\ &\leq M^2 \int_{\tilde{T}} \left[\int_{\tilde{T}} |f(t')| dt' \right]^2 dt \leq M^2 \int_{\tilde{T}} \|f\|_1^2 dt \leq M^2 \|f\|_1^2 \left[\int_{\tilde{T}} 1 dt \right] \leq M^2 \|f\|_1^2 \mu(\tilde{T}) \end{aligned}$$

D'où en utilisant le théorème principal :

$$\left| \langle \tilde{f}, g_i \rangle \right| \leq \frac{\|v\|}{|\lambda - \lambda_i|} \leq \frac{M \|f\|_1 \mu(\tilde{T})^{1/2}}{|\lambda - \lambda_i|}$$

L'application du théorème principal suppose que les valeurs propres sont espacées. Nous allons voir que c'est une hypothèse raisonnable.

IV-1-g) Théorème :

La somme des premières valeurs propres de l'opérateur de covariance croît avec l'intervalle T

Démonstration :

Avec les notations précédentes soient $\{\lambda_i\}$ les valeurs propres associées à l'opérateur U et $\{\lambda'_i\}$ les valeurs propres associées à l'opérateur $\tilde{U}$ montrons que pour tout $N \in \mathbb{N}$ on a:

$$\sum_{i \leq p} \lambda'_i \leq \sum_{i \leq p} \lambda_i$$

Soit $\{g_i\}_{i \in N}$ l'ensemble des vecteurs propres de $\tilde{U}$; notons g_i^0 la fonction égale à g_i sur $\tilde{T}$ et nulle sur $\bar{\tilde{T}}$; Désignons par P la projection orthogonale sur l'espace engendré par la famille $\{f_i\}_{i \leq p}$ des vecteurs propres de U et Q la projection orthogonale sur l'espace engendré par la famille $\{g_i^0\}_{i \leq p}$ d'après le théorème de J.C Deville

$$V(\|Q(X(\omega))\|) \leq V(\|P(X(\omega))\|) = \sum_{i < p} \lambda_i$$

Dans l'espace $L^2(\tilde{T})$ notons K la projection orthogonale sur l'espace engendré par la famille $\{g_i\}_{i \leq p}$ on a:

$$V(\|Q(X(\omega))\|_T) = V(\|K(X(\omega))\|_T) = \sum_{i \leq p} \lambda'_i$$

en effet :

$$V(\|Q(X(\omega))\|_T) = V\left(\sum_{i \leq p} \langle X(\omega), g_i^0 \rangle\right) =$$

$$V\left(\sum_{i < p} \langle X(\omega), g_i^0 \rangle^2\right) = V\left(\sum_{i \leq p} \langle X(\omega), g_i \rangle^2\right) =$$

$$V(\|K(\tilde{X}(\omega))\|_{\tilde{T}}) = \sum_{i \leq p} \lambda'_i$$

Remarque:

Nous aurions souhaité que le théorème soit démontré pour la relation $\lambda'_i \leq \lambda_i$. Nous avons cherché des contre-exemples sans succès. De toute façon on a pour $N = 1$ la relation $\lambda'_1 \leq \lambda_1$.

Conclusion générale:

Il est raisonnable de penser que lorsque l'intervalle de temps T est « grand », les valeurs propres sont espacées. De ce fait la projection d'une fonction propre de T (calculée sur T) sur toute fonction propre de $\tilde{T}$ est très faible sauf pour une valeur. Cela engendre la proximité des fonctions propres convenablement choisies.

IV-2) RÉALITÉ ET ILLUSION DE LA PRÉVISION HARMONIQUE :

La méthode de prévision que nous proposons se base sur le prolongement des composantes temporelles. Dans ce paragraphe nous nous sommes posés différentes questions.

1. Les composantes temporelles sont elles à elle seules capables de décrire un processus ?
2. Si les données d'un processus défini sur $[a,b]$ et celles définies sur $[b,c]$ sont indépendantes qu'en est-il des composantes temporelles ?
3. Lorsque on génère un processus en dupliquant exactement les données qu'en est-il des composantes temporelles ?
4. Lorsqu'un processus comporte des intervalles sur lesquels les données se répètent exactement, comment se comportent les composantes temporelles ?

Les théorèmes suivants répondent à ces questions en confirmant nos intuitions.

1. A une famille de composantes temporelles correspond une infinité de processus qui forme une classe (théorème VI-2-a)
2. et 3. sont des situations opposées. Pour que les composantes temporelles puissent se calculer intervalle par intervalle il faut que les processus partiels soient indépendants (théorème VI-2-b) ou identiques (théorème VI-2-c), le théorème principal (IV-1e) pouvant être considéré comme une situation intermédiaire entre ces deux situations extrêmes.
4. Si les données se répètent sur certains intervalles toutes les composantes temporelles se répètent aussi.

IV-2-a) Théorème

Soit, $\{f_i\}_{i \in \mathbb{N}}$ une famille orthogonale de fonctions de $L^2(T)$ et $\{\lambda_i\}_{i \in \mathbb{N}}$ une famille de réels positifs et $\{Y_i\}_{i \in \mathbb{N}}$ une famille de variable aléatoires réelles définies sur Ω telles que :

$$E(Y_i) = \int_{\Omega} Y_i(\omega) dP\omega = 0$$

et

$$\int_{\Omega} Y_i(\omega) Y_j(\omega) dP\omega = \delta_{ij} \lambda_i \lambda_j$$

de plus supposons que $\sum_{i \in \mathbb{N}} \lambda_i$ converge ainsi que les séries $\sum_{i \in \mathbb{N}} (Y_i(\omega))^2$

Alors le processus défini par $X(\omega, t) = \sum_{i \in \mathbb{N}} Y_i(\omega) f_i(t)$ admet pour composantes temporelles la famille $\{f_i, \lambda_i\}_{i \in \mathbb{N}}$

Démonstration:

Remarquons avant la démonstration que la série $\sum_{i \in N} Y_i(\omega) f_i(t)$ converge pour tout ω dans $L^2(T)$ ce qui définit $X(\omega, t)$.

Désignons par N le noyau de covariance du processus X alors :

$$\begin{aligned} N(t, t') &= \int_{\Omega} X(\omega, t) X(\omega, t') dP\omega = \\ &= \int_{\Omega} \left[\sum_{i \in N} Y_i(\omega) f_i(t) \right] \left[\sum_{j \in N} Y_j(\omega) f_j(t') \right] dP\omega \\ &= \sum_{i, j \in N \times N} \left(\int_{\Omega} Y_i(\omega) Y_j(\omega) dP\omega \right) f_i(t) f_j(t') = \sum_{i \in N} \lambda_i f_i(t) f_i(t') \end{aligned}$$

de ce fait :

$$\begin{aligned} U(f_k)(t) &= \int_T N(t, t') f_k(t') dt' = \int_T \left(\sum_{i \in N} \lambda_i f_i(t) f_i(t') \right) f_k(t') dt' \\ &= \sum_{i \in N} \int_T \lambda_i f_i(t) f_i(t') f_k(t') dt' = \lambda_k f_k \end{aligned}$$

Ce théorème démontre qu'à partir d'une famille de composantes temporelles il est possible de créer une infinité de processus. Pour cela il suffit de choisir des familles différentes de $\{Y_i\}_{i \in N}$; ces dernières pouvant être engendrées par des processus différents donnés à l'avance. Cette remarque amène à considérer une partition de l'ensemble des processus avec la relation d'équivalence :

IV-2-b) Définition

Deux processus sont équivalents si et seulement si ils admettent la même famille de composantes temporelles.

IV-2-c) Théorème : (Indépendance de l'avenir en fonction du passé).

Si X est un processus défini sur $[a, c]$ et si b est un point à l'intérieur de $[a, c]$ alors les restrictions $\tilde{X}$ et $\bar{X}$ de X respectivement sur $[a, b]$ et $[b, c]$ admettent pour composantes temporelles les restrictions des composantes temporelles de X si pour tout t élément de $[a, b]$ et t' élément de $[b, c]$ on a :

$$\int_{\Omega} X(\omega, t) X(\omega, t') dP\omega = 0 \quad (I)$$

De plus si f est un vecteur propre de U de valeur propre λ , $\tilde{f}$ (respectivement $\bar{f}$) est un vecteur propre de $\tilde{U}$ (respectivement $\bar{U}$) de valeur propre λ

Démonstration :

Nous allons dans un premier temps montrer que le noyau de covariance de X est la somme des noyaux de covariance de $\tilde{X}$ et de $\bar{X}$.

Calculons :

$$\begin{aligned} N(t, t') &= \int_{\Omega} X(\omega, t) X(\omega, t') dP\omega \\ &= \int_{\Omega} (\tilde{X}(\omega, t) + \bar{X}(\omega, t)) (\tilde{X}(\omega, t') + \bar{X}(\omega, t')) dP\omega \\ &= \tilde{N}(t, t') + 2 \int_{\Omega} \tilde{X}(\omega, t) \bar{X}(\omega, t') dP\omega + \bar{N}(t, t') \\ &= \tilde{N}(t, t') + \bar{N}(t, t') \end{aligned}$$

Remarquons au passage que :

$\tilde{N}(t, t') = 0$ (respectivement $\bar{N}(t, t') = 0$) si $t \notin \bar{T}$ ou $t' \notin \bar{T}$ (respectivement $t \notin \tilde{T}$ ou $t' \notin \tilde{T}$)

Soit f un vecteur propre de U de valeur propre λ alors si t est un élément de $\tilde{T}$:

$$\begin{aligned} U(f)(t) &= \lambda f(t) = \int_T N(t, t') f(t') dt' = \int_T (\tilde{N}(t, t') + \bar{N}(t, t')) f(t') dt' \\ &= \int_T \tilde{N}(t, t') \tilde{f}(t') dt' = \tilde{U}(\tilde{f})(t) = \lambda \tilde{f}(t) \end{aligned}$$

ce qui prouve que $\tilde{f}$ est un vecteur propre de $\tilde{U}$. La démonstration est terminée.

Remarque :

La réciproque du théorème ci-dessus est fausse comme le montre le théorème de périodicité parfaite qui suit.

IV-2-d) Théorème de périodicité parfaite

Soit $\tilde{T}$ un intervalle de longueur a et $\bar{T}$ l'intervalle obtenu en décalant $\tilde{T}$ de a que nous noterons $\bar{T} = \tilde{T} + a$. Soient deux processus $\tilde{X}$ et $\bar{X}$ respectivement nuls sur $\bar{T}$ et $\tilde{T}$ et tels que pour tout $t \in \bar{T}$ et $\omega \in \Omega$, $\bar{X}(\omega, t) = \tilde{X}(\omega, t-a)$ (périodicité parfaite). Alors si $(\tilde{f}, \lambda)$ est une composante temporelle de $\tilde{X}$

$$a) \tilde{f}(t) = 0 \text{ pour tout } t \in \bar{T}$$

$$b) (\bar{f}, \lambda) \text{ est une composante temporelle de } \bar{X} \text{ si } \bar{f}(t) = \tilde{f}(t-a) \text{ pour tout } t \in \bar{T} \text{ et } \tilde{f}(t) = 0 \text{ pour tout } t \in \tilde{T}$$

c) De plus si on pose $X = \tilde{X} + \bar{X}$ le couple $(f, 2\lambda)$, tel que $f = \tilde{f} + \bar{f}$ est une composante temporelle de X

d) Réciproquement si (f, λ) est une composante temporelle de X Le couple $(\tilde{f}, \frac{1}{2}\lambda)$ est une composante temporelle de $\tilde{X}$ où $\tilde{f}(t) = f(t)$ si $t \in \tilde{T}$ et $\tilde{f}(t) = 0$ si $t \in \bar{T}$.

Démonstration:

a) Le noyau de $\tilde{X}$ est nul si un des instant est sur $\bar{T}$.

Soit $t \in \bar{T}$ $U_{\tilde{X}}(\tilde{f})(t) = \lambda \tilde{f}(t) = \int_{\tilde{T}} \tilde{N}(t, t') \tilde{f}(t') dt' = 0$; ce qui démontre le a) du théorème

b) Calculons pour $t \in \bar{T}$:

$$U_{\bar{X}}(\bar{f})(t) = \int_{\bar{T}} \bar{N}(t, t') \bar{f}(t') dt' = \int_{\tilde{T}} \bar{N}(t, t') \bar{f}(t') dt' = \int_{\tilde{T}} \tilde{N}(t-a, t'-a) \tilde{f}(t'-a) dt' \\ = \lambda \tilde{f}(t-a) = \lambda \bar{f}(t), \text{ ce qui démontre le b) car dans le cas où } t \in \tilde{T} \text{ la valeur est nulle.}$$

c) Examinons la forme particulière du noyau de covariance de X .

$$N(t, t') = \int_{\Omega} (\tilde{X}(t, \omega) + \bar{X}(t, \omega)) (\tilde{X}(t', \omega) + \bar{X}(t', \omega)) dP\omega \\ = \tilde{N}(t, t') + \int_{\Omega} \tilde{X}(t, \omega) \bar{X}(t', \omega) dP\omega + \int_{\Omega} \bar{X}(t, \omega) \tilde{X}(t', \omega) dP\omega + \bar{N}(t, t')$$

$$\text{Si } (t, t') \in \tilde{T} \times \tilde{T} \text{ alors } N(t, t') = \tilde{N}(t, t')$$

$$\text{Si } (t, t') \in \tilde{T} \times \bar{T} \text{ alors } N(t, t') = \int_{\Omega} \tilde{X}(t, \omega) \bar{X}(t', \omega) dP\omega$$

Ces résultats étant acquis nous nous proposons d'évaluer, quant $t \in \tilde{T}$, $U(f)(t)$:

$$U(f)(t) = \int_{\tilde{T}} N(t, t') f(t') dt' = \int_{\tilde{T}} N(t, t') \tilde{f}(t') dt' + \int_{\bar{T}} N(t, t') \bar{f}(t') dt' = \\ \int_{\tilde{T}} \tilde{N}(t, t') \tilde{f}(t') dt' + \int_{\bar{T}} \int_{\Omega} \tilde{X}(t, \omega) \bar{X}(t', \omega) dP\omega \bar{f}(t') dt' =$$

$$\int_T \tilde{N}(t, t') \tilde{f}(t') dt' + \int_T \int_{\Omega} \tilde{X}(t, \omega) \tilde{X}(t'-a, \omega) dP \omega \tilde{f}(t'-a) dt'$$

Posons $u = t' - a$ alors :

$$\begin{aligned} U(f)(t) &= \int_T \tilde{N}(t, t') \tilde{f}(t') dt' + \int_T \int_{\Omega} \tilde{X}(t, \omega) \tilde{X}(u, \omega) dP \omega \tilde{f}(u) dt' = \\ &2 \int_T \tilde{N}(t, t') \tilde{f}(t') dt' = \tilde{U}(\tilde{f})(t) = 2 \lambda \tilde{f}(t) \end{aligned}$$

La démonstration du c) est terminée.

d) La démonstration de la réciproque se fait en utilisant le même calcul que dans le point c) ci dessus et permet de démontrer que $\tilde{U}(\tilde{f})(t) = 1/2 \lambda f(t)$

IV-2-e) Théorème de répétition parfaite

Soit I et J deux sous intervalles disjoints de l'intervalle T tels $J = I + p$ où p est un nombre positif supérieur à la longueur de I . Soit (f, λ) une composante temporelle d'un processus X défini sur T tel que pour tout $\omega \in \Omega$ et $t \in I$ on ait $X(\omega, t) = X(\omega, t+p)$ alors

$$f(t) = f(t+p)$$

Démonstration

Remarquons que $N(t, t') = N(t+p, t')$ si $t \in I$.

En effet :

$$N(t, t') = \int_{\Omega} X(\omega, t) X(\omega, t') dP \omega = \int_{\Omega} X(\omega, t+p) X(\omega, t') dP \omega = N(t+p, t').$$

De ce fait :

$$U(f)(t) = \int_T N(t, t') f(t') dt' = \int_T N(t+p, t') f(t') dt' = U(f)(t+p)$$

La démonstration est terminée

Remarque :

Le théorème de répétition parfaite qui semble être un cas général du précédent, ne fait que généraliser une partie du résultat. Dans le cas de la périodicité parfaite les composantes temporelles du processus global sont des « copies » des composantes temporelles du processus de base et de ce fait se ressemblent. Dans le cas du théorème de répétition parfaite il n'y a pas de rapport évident entre le processus global et les processus qui se répètent. Par contre les composantes temporelles globales du processus sont identiques sur les intervalles où les données se répètent.

IV-2-f) Conclusion

La connaissance des composantes temporelles d'un processus (fonctions et valeurs propres de l'opérateur de covariance) ne donne qu'un aperçu du processus car à une famille de composantes temporelles correspond une infinité de processus. Trois cas particuliers éclairent la complexité de la situations générale:

- Si deux processus sont mis bout à bout sans intervention du passé sur l'avenir ($\int_{\Omega} X(\omega, t) X(\omega, t') dP\omega = 0$ pour tout $(t, t') \in \tilde{T} \times \bar{T}$) le processus X ainsi formé admet pour composantes temporelles la simple juxtaposition des composantes temporelles des processus originels. La réciproque étant fausse.
- Dans des « circonstances opposées » si un processus est formé de la juxtaposition d'un processus donné et de sa réplique (périodicité parfaite du processus), les composantes temporelles du processus global sont aussi (au coefficient 2 près) la réplique des composantes temporelles du processus dupliqué.
- Plus généralement si un processus se répète exactement sur deux sous intervalles du temps, les composantes temporelles du processus global se répètent. Attention il ne faut pas en déduire que les composantes temporelles globales sont une réplique des composantes temporelles partielles comme c'était le cas précédemment.

La situation générale est encore plus complexe car le futur est lié au passé sans se répliquer. Dans cette situation le théorème principal indique que plus les premières valeurs propres du « passé » seront écartées les une des autres plus il sera raisonnable de prévoir l'avenir en prolongeant les composantes temporelles connues.

IV-2) LA PLACE DES PROCESSUS M.A. EN ANALYSE HARMONIQUE :

IV-2-a) Analyse harmonique d'un processus MA défini sur un intervalle T :

Nous nous proposons d'effectuer une analyse harmonique d'un processus à moyenne mobile défini sur un intervalle de temps T. L'étude de ces processus est plus facile à réaliser lorsqu'ils sont définis sur un intervalle $[a, b]$ que lorsque ils le sont sur un ensemble discret.

Théorème :

L'analyse harmonique d'un processus de type M.A. est réduite au noyau de son opérateur de covariance.

Démonstration:

Rappelons qu'un bruit blanc défini sur $[a,b]$ est un processus e défini par

$$e : (\omega, t) \in \Omega \times [a,b] \longrightarrow e(\omega, t) \in \mathbb{R}$$

tel que pour tout t et t' différents de $[a,b]$ $E(e(t), e(t')) = 0$ et $E(e(t), e(t)) = 1$

Par définition un processus à Moyenne Mobile (Mobil Average) est défini par la combinaison linéaire $\sum_{i=1}^k B^i e$ ou B^i est l'opérateur de décalage de retard i . Pour fixer les idées nous prendrons un processus X à moyenne mobile d'ordre 2 défini par :

$$X(\omega, t) = e(\omega, t) + \alpha e(\omega, t-1) + \beta e(\omega, t-2)$$

Posons $\delta(t, t') = 0$ si t est différent de t' et $\delta(t, t) = 1$ le noyau de covariance de X s'écrit :

$$\begin{aligned} N(t, t') &= \int_{\Omega} [e(\omega, t) + \alpha e(\omega, t-1) + \beta e(\omega, t-2)][e(\omega, t') + \alpha e(\omega, t'-1) + \beta e(\omega, t'-2)] dP\omega = \\ &= \int_{\Omega} e(\omega, t)e(\omega, t') dP\omega + \int_{\Omega} \alpha e(\omega, t-1)e(\omega, t') dP\omega + \int_{\Omega} \beta e(\omega, t-2)e(\omega, t') dP\omega \\ &+ \int_{\Omega} \alpha e(\omega, t)e(\omega, t'-1) dP\omega + \int_{\Omega} \alpha^2 e(\omega, t-1)e(\omega, t'-1) dP\omega + \int_{\Omega} \alpha \beta e(\omega, t-2)e(\omega, t'-1) dP\omega \\ &+ \int_{\Omega} \beta e(\omega, t)e(\omega, t'-2) dP\omega + \int_{\Omega} \alpha \beta e(\omega, t-1)e(\omega, t'-2) dP\omega + \int_{\Omega} \beta^2 e(\omega, t-2)e(\omega, t'-2) dP\omega \\ &= \delta(t, t') + \alpha \delta(t-1, t') + \alpha \delta(t, t'-1) + \alpha^2 \delta(t-1, t'-1) + \alpha \beta \delta(t-1, t'-2) + \alpha \beta \delta(t-2, t'-1) \\ &+ \beta \delta(t-2, t') + \beta \delta(t, t'-2) + \beta^2 \delta(t-2, t'-2) \end{aligned}$$

Soit (f, λ) une composante temporelle de X par définition :

$U(f)(t') = \int_T N(t, t') f(t') dt'$ or $\int_T \delta(t, t') f(t') dt' = 0$ pour tout t et t' du segment $[a,b]$. En explicitant l'écriture du noyau on obtient:

$$U(f)(t') = 0$$

Ce qui prouve que les composantes temporelles d'un processus à moyenne mobile sont des vecteurs appartenant au noyau.

IV-2-b) Modification des processus à l'aide des processus de type MA

Nous nous proposons de voir si l'analyse harmonique d'un processus est perturbé par l'addition d'un processus de type MA. Le théorème suivant montre que ce n'est pas le cas. Cette remarque pourra justifier une modification des données d'un processus avant l'analyse harmonique.

Théorème

Soit Y un processus et X un processus de type MA, supposons que le processus Y et le processus bruit blanc $\mathcal{E}$ ne soient pas corrélés :

$$\text{Cov}(Y(t), \mathcal{E}(t')) = 0 \text{ pour tout } t \text{ et } t' \text{ (même } t=t') \quad (1).$$

Alors les processus Y et $Y+X$ sont équivalents au sens de l'analyse harmonique.

L'hypothèse $\text{Cov}(Y(t), \mathcal{E}(t')) = 0$ peut sembler forte et non justifiée . Elle est cependant habituellement faite car $\mathcal{E}$ est considéré comme engendré par des chocs aléatoires qui ne sauraient avoir des liens avec le processus Y . Voir par exemple Guy Mélard [25] deux dernières lignes de la page 313.

Démonstration:

L'hypothèse (1) permet d'écrire que le noyau du processus $S = X + Y$ est la somme des noyaux de X et de Y : $N_S = N_X + N_Y$

Soit U_S l'opérateur de covariance de S et (f, λ) une composante temporelle alors :

$$\begin{aligned} U_S(f)(t) &= \int_T N_S(t, t') f(t') dt' = \int_T N_X(t, t') f(t') dt' + \int_T N_Y(t, t') f(t') dt' \\ &= U_X(f)(t') + U_Y(f)(t') = U_Y(f)(t') = \lambda f(t) \end{aligned}$$

ce qui termine la démonstration

IV-4) ANALYSE HARMONIQUE ET PRÉVISION DES PROCESSUS STATIONNAIRES (AR, POISSON ...)

IV-4-a) Processus du type M.A. :

Nous avons déjà vu que les processus M.A. définis sur un intervalle $[a, b]$ n'interviennent pas dans une analyse harmonique.

IV-4-b) Processus du type A.R. :

Nous n'avons pas de résultats généraux sur les processus auto-régressifs (A.R.) mais nous avons remarqué, dans le cas discret, que les composantes temporaires semblent globalement sinusoïdales et nous avons démontré que les valeurs propres sont bornées.

Pour faire une prévision d'un auto-régressif dont on a estimé le modèle il est toujours possible de calculer la matrice de covariance et par la même les composantes temporelles.

Si on possède une seule série chronologique la prévision de cette série peut se faire en décomposant à l'instant t la série sur ses composantes temporelles puis en recomposant la série à l'instant $t+1$. Cependant la proximité des valeurs propres rend impossible le choix d'un nombre fini et significatif de composantes temporelles. Il n'y a donc aucun intérêt pratique qui justifie l'utilisation de l'analyse harmonique lors de l'étude d'un processus auto-régressif.

IV-4-c) Processus de Poisson :

Les processus de Poisson permettent de calculer les composante temporelles d'une façon théoriques. Les valeurs propres sont très séparées (Saporta [27] page 113) . Il est tout à fait légitime d'utiliser la méthodes de prévision proposée par l'analyse harmonique. Si on possède une seule réalisation il suffit de calculer ses coordonnées sur chacune des composantes temporelles significatives et de recomposer les données à l'instant suivant.

IV-4-a) Conclusion :

La prévision des processus stationnaire n'est pas différente de celle des autres processus. On doit cependant remarquer que la matrice est plus facile à estimer quand on a peu de données (une seule série chronologique, assez longue, peut suffire). L'analyse harmonique est toujours possible, mais les calculs et la prévision ne le sont pas toujours. Cela dépend du comportement des valeurs propres de la matrice de covariance. Si les valeurs propres sont trop proches de zéro ou trop proches entre elles les calculateurs peuvent donner des valeurs fausses. De plus le théorème principal ne s'applique plus et la prévision est fantaisiste.

Dans le cas où la matrice de covariance possède des vecteurs propres séparés on peut légitimement utiliser l'analyse harmonique pour réaliser une prévision.

Dans ce dernier cas il est possible de proposer un intervalle de confiance à la prévision. On procède de la même façon que lors d'un lissage exponentiel ou d'une prévision à l'aide de la méthode de Box et Jenkins. On réalise les prévisions pas à pas du temps t au temps $t+1$ jusqu'à l'avant dernière valeur connue. Ensuite on calcule les erreurs entre les prévisions et les valeurs réelles. Cette série des erreurs sera supposée gaussienne et permettra d'estimer un intervalle de confiance autour de la prévision.

IV-5) STRATÉGIE DE PRÉVISION DES FAMILLES DE TRAJECTOIRES

IV-5-a) Le principe et ses limites :

La prévision d'une famille de trajectoires est basée sur l'idée que le futur est construit sur les informations apportées simultanément par toutes les trajectoires .

L'analyse harmonique répond à cette exigence car le calcul des composantes temporelles fait intervenir toutes les données disponibles.

La technique que nous proposons est basée sur le prolongement des composantes temporelles que nous supposons être moins nombreuses et plus régulières que les trajectoires. L'analyse harmonique s'adresse de ce fait à des ensembles importants de trajectoires sauf dans le cas où on est capable de déterminer théoriquement les composantes principales (processus de Poisson,.....).

Dans la méthode de prévision à l'aide des techniques de l'analyse harmonique le prolongement des tendances reste une question délicate.

Si le prolongement de la moyenne et des tendances sont exactes la prévision devrait être très satisfaisante, mais dans le cas contraire, toutes les prévisions des trajectoires seront mauvaises. Il est donc conseillé, comme dans toute méthode de prévision d'utiliser d'autres méthodes pour conforter ou infirmer les prévisions faites.

Nous espérons cependant, comme nous l'avons testé dans nos exemples avoir une meilleure prévision à l'aide de l'analyse harmonique d'un ensemble de trajectoires que lorsque on fait une prévision trajectoire par trajectoire.

IV-5-b) La procédure proposée

Dans toutes les prévisions que nous proposerons nous utiliserons la procédure suivante:

- 1.Centrer le processus
- 2.Calculer la matrice de covariance
- 3.Rechercher les valeurs propres et les vecteurs propres
- 4.Vérifier l'écart entre les premières valeurs propres
- 5.Sélectionner les vecteurs propres significatifs
- 6.Prolonger ces vecteurs propres
- 7.Prolonger la moyenne
- 8.Reconstituer les données

CHAPITRE V

VALIDITE DE L'ANALYSE HARMONIQUE DANS LE CAS D'UN NOYAU DE COVARIANCE QUELCONQUE

V-1) LES DIFFÉRENTES PROCÉDURES POSSIBLES POUR SÉLECTIONNER LES COMPOSANTES TEMPORELLES SIGNIFICATIVES

.

Les besoins de la prévision, à l'aide de l'analyse harmonique, nous ont amené à sélectionner un nombre minimum de composantes temporelles afin de réduire les erreurs lors du prolongement. Rappelons que si X est un processus on a

$$X(\omega) = \sum_{i \in N} \langle X(\omega), f_i \rangle f_i$$

et si on pose

$$Y(\omega) = \sum_{i \leq p} \langle X(\omega), f_i \rangle f_i$$

l'erreur commise est le processus $Z = X - Y$ qui s'écrit :

$$Z(\omega) = \sum_{i > p} \langle X(\omega), f_i \rangle f_i$$

Le premier critère de choix proposé par J.C Deville [15] correspond à la démarche de l'analyse en composante principale. Nous savons que :

$$V(\|X\|) = \sum_{i \in N} \lambda_i \quad \text{et} \quad V(\|Y\|) = \sum_{i < p} \lambda_i \quad \text{et} \quad V(\|Z\|) = \sum_{i \geq p} \lambda_i$$

On peut de ce fait définir un indice de perte d'information par exemple à 5% en choisissant p de telle façon que

$$V(\|Z\|) / V(\|X\|) = \frac{\sum_{i \geq p} \lambda_i}{\sum_{i \in N} \lambda_i} = \leq 5 \%$$

Cet indice comporte deux risques d'erreur.

1. Le traditionnel problème du « coude » que l'on rencontre en analyse en composantes principales.
2. Le trop grand effet de « lissage de la norme ». En effet le fait de calculer la norme de chacune des trajectoires avant de mesurer la perte d'information par la variance risque de faire croire que la perte d'information est négligeable, alors qu'elle ne l'est qu'en moyenne. On peut redouter en particulier les effets de pics.

Le théorème de Mercer permet d'imaginer un autre indice qui se basera sur la convergence uniforme de la série de fonctions $V(Z) = \sum_{i \geq p} \lambda_i f_i^2$

On définit ainsi une nouvelle procédure pour déterminer le rang p_0 à partir duquel la perte d'information sera uniformément négligeable en fixant une valeur de ε à l'avance et en choisissant p_0 tel que :

$$V(Z)(t) = \int_{\Omega} Z(\omega, t)^2 dP\omega = \sum_{i \geq p_0} \lambda_i f_i^2(t) < \varepsilon$$

Nous proposons de choisir la valeur de ε graphiquement à l'aide du logiciel que nous avons conçu à cet effet.

Nous dirons que Z est un « bruit parasite » ou tout simplement un bruit. Nous ne devons pas confondre ce bruit avec un bruit blanc. Nous remarquerons cependant, dans ce cas, que le noyau de covariance N_Z de ce processus vérifie les relations suivantes:

$$|N_Z(t, t)| = \left| \int_{\Omega} Z(\omega, t) Z(\omega, t') dP\omega \right| \leq \left(\int_{\Omega} Z(\omega, t)^2 dP\omega \right)^{\frac{1}{2}} \left(\int_{\Omega} Z(\omega, t')^2 dP\omega \right)^{\frac{1}{2}} \leq \varepsilon$$

ce qui montre que les corrélations sont très faibles et de ce fait que Z peut être considéré comme comportant peu d'information. Il n'est cependant pas à confondre ou à rapprocher d'un bruit blanc dont la variance est constante et généralement égale à 1.

Rappelons que :

$$E(Z(\omega, t)) = 0 \quad \text{et} \quad V(Z(\omega, t)) < \varepsilon$$

V-1-a) Définition :

Ce nouvel indice sera appelé indice fort de perte d'information et il s'exprime par la fonction I_p par :

$$I_p : t \longrightarrow V(Z)(t)$$

Remarque :

Cet indice n'est pas un nombre mais une fonction. Pour tout t la valeur de la fonction variance est la variance de la variable aléatoire obtenue en observant le processus à l'instant t . C'est la variance de chaque variable aléatoire de la famille $\{Z_t\}_{t \in [a, b]}$

La difficulté principale de cet indice est liée aux résultats sur la convergence uniforme de la série. Les processus stationnaires et en particulier les processus de Poisson ont un noyau

de covariance continu et le théorème de Mercer permet de justifier l'utilisation de l'indice fort, mais les processus en général et les processus économiques en particulier ont des noyaux de covariance qui échappent à ce cadre trop étroit. Seule l'hypothèse que chaque trajectoire du processus est de carré intégrable est acceptable. Cette hypothèse entraîne que le noyau de covariance est un élément de $L^2(T \times T)$ mais n'implique pas qu'il soit continu. Pour justifier l'utilisation de l'indice fort dans tous les cas il nous faut rétablir le théorème de Mercer dans le cas particulier d'un noyau de covariance sous la seule hypothèse d'appartenance à l'ensemble $L^2(T \times T)$. C'est ce que nous allons faire dans le paragraphe suivant.

V-2) ETUDE DE L'INDICE FORT DE PERTE D'INFORMATION

Notons 1_T la fonction égale à 1 sur l'intervalle T . Nous cherchons à démontrer que l'indicateur fort vérifie pour tout $\varepsilon > 0$ donné et pour n choisi assez grand : $V(Z) < \varepsilon 1_T$

Formellement nous avons vu que nous souhaitons que cet indicateur s'écrive :

$$V(Z) = \sum_{i > p} \lambda_i f_i^2$$

L'étude de l'indicateur fort se ramènera à l'étude de cette dernière série.

Afin d'obtenir l'ensemble des résultats théoriques souhaitables sur cet indicateur, nous allons oeuvrer en deux temps.

- Dans un premier temps nous allons démontrer que l'égalité ci-dessus est vraie presque pour tout t , à condition que chaque réalisation du processus soit dans l'espace des fonctions de carré intégrable.

Dans un second temps nous généraliserons le théorème de Mercer en remplaçant l'hypothèse de continuité du noyau de covariance par sa seule appartenance à l'espace $L^2(T \times T)$ des fonctions de carré intégrable définies sur $T \times T$. Plus précisément nous montrerons que la série $\sum_{i \in p} \lambda_i f_i^2$ converge (quasiment) uniformément vers $V(X)$ dans le cas général où chaque réalisation du processus est une fonction de carré intégrable.

V-2-a) Théorème :

Si X est un processus presque à tout instant on a les égalités:

$$V(X)(t) = N(t, t) = \sum_{i \in N} \lambda_i f_i^2(t)$$

Démonstration :

La démonstration va être faite en trois étapes.

1. Une étude du noyau de covariance de $Z = X_p = X - \sum_{i \leq p} X_i$
2. Une démonstration de l'égalité $\int_T V(t) dt = \sum_{i \in N} \lambda_i$
3. L'utilisation simultanée de la norme de $L^1(T)$ de la positivité de la variance pour finir la démonstration (Rappelons que $L^1(T)$ est l'espace des classes de fonctions dont la valeur absolue est intégrable. C'est un espace de Banach dont la norme est désignée par $\| \cdot \|_{L^1}$. La convergence pour la norme de $L^2(T)$ étant plus forte que celle de $L^1(T)$, il ne faut pas espérer que la série $\sum_{i \in p} \lambda_i f_i^2(t)$ converge dans $L^2(T)$.)

Etape n°1 de la démonstration :

Nous allons étudier le processus $Z = X_p = X - \sum_{i \leq p} X_i$

Nous avons par définition :

$$X_p(\omega, t) = X(\omega, t) - \sum_{i \leq p} \langle X, f_i \rangle f_i(t)$$

Montrons dans un premier temps que le noyau N_p de X_p n'est autre que:

$$N_p(t, t') = N(t, t') - \sum_{i \leq p} \lambda_i f_i(t) f_i(t')$$

Calculons :

$$N_p(t, t') = \int_{\Omega} \left(X(\omega, t) - \sum_{i \leq p} \langle X(\omega), f_i \rangle f_i \right) \left(X(\omega, t') - \sum_{i \leq p} \langle X(\omega), f_i \rangle f_i \right) dP$$

Développons cette expression en utilisant la linéarité de l'intégrale

$$\begin{aligned} N_p(t, t') &= \\ N(t, t') &- \left[\sum_{i < p} \left(\int_{\Omega} X(\omega, t') \langle X(\omega), f_i \rangle dP \omega \right) f_i(t) \right] \\ &- \left[\sum_{i < p} \left(\int_{\Omega} X(\omega, t) \langle X(\omega), f_i \rangle dP \omega \right) f_i(t') \right] \end{aligned}$$

$$+ \left[\sum_{i < p, j < n} \left(\int_{\Omega} \langle X(\omega), f_i \rangle \langle X(\omega), f_j \rangle dP\omega \right) f_i(t) f_j(t') \right]$$

d'où (voir annexe IV-1-c) :

$$N_p(t, t') = N(t, t') - 2 \sum_{i \leq p} \lambda_i f_i(t) f_i(t') + \sum_{i \leq p} \lambda_i f_i(t) f_i(t') =$$

$$N(t, t') - \sum_{i \leq p} \lambda_i f_i(t) f_i(t')$$

remarquons au passage que pour $t = t'$ l'expression ci dessus est positive car il s'agit de la variance du processus X_p d'où l'inégalité :

$$N(t, t) \geq \sum_{i < p} \lambda_i f_i(t)^2$$

Ce qui prouve que la série à terme positif $\sum_{i \in N} \lambda_i f_i^2(t)$ est convergente pour tout t .

Pour démontrer le théorème il suffit de montrer que :

$$\left\| \sum_{i \in N} \lambda_i f_i^2 - V \right\|_{L^1} = 0$$

Etape n°2 de la démonstration :

Démontrons d'abord que :

$$\int_T V(t) dt = \sum_{i \in N} \lambda_i$$

Nous avons les égalités suivantes :

$$\int_T V(t) dt = \int_T \left[\int_{\Omega} X(\omega, t)^2 dP\omega \right] dt = \int_{\Omega} \left[\int_T X(\omega, t)^2 dt \right] dP\omega = \int_{\Omega} \|X(\omega)\|_{L^2}^2 dP\omega$$

$$= \sum_{i \in N} \int_{\Omega} \langle X(\omega), f_i \rangle^2 dP\omega + \sum_{j \in N} \int_{\Omega} \langle X(\omega), \varphi_j \rangle^2 dP\omega$$

où $\{\varphi_i\}$ est une base du noyau de U .

D'autre part nous savons que :

$$\int_{\Omega} \langle X(\omega), f_i \rangle^2 dP\omega = \lambda_i \quad \text{et que} \quad \int_{\Omega} \langle X(\omega), \varphi_j \rangle^2 dP\omega = 0$$

d'où le résultat :

$$\int_T V(t) dt = \sum_{i \in N} \lambda_i$$

Etape n°3 de la démonstration :

Finissons le calcul :

$$\left\| \sum_{i \in N} \lambda_i f_i^2 - V \right\|_{L^1} = \int_T \left| V(t) - \sum_{i \in N} \lambda_i f_i^2(t) \right| dt \stackrel{(1)}{=} \\ \int_T V(t) dt - \int_T \left[\sum_{i \in N} \lambda_i f_i^2(t) \right] dt = \sum_{i \in N} \lambda_i - \sum_{i \in N} \lambda_i = 0$$

L'égalité (1) ayant lieu car $V(t) \geq \sum_{i < p} \lambda_i f_i^2(t)$

La démonstration est terminée.

V-2-b) Théorème :

Si le noyau de covariance est un élément de $L^2(T) \times L^2(T)$ alors la série $\sum_{i > N} \lambda_i f_i^2$ converge « quasiment » uniformément vers la fonction variance V .

Démonstration

La démonstration est une application directe du théorème suivant V-2-c) et du théorème V-2-a)

Remarque

Contrairement à la situation exploitée par le théorème de Mercer la convergence uniforme de la série $\sum_{i > N} \lambda_i f_i^2$ n'est pas liée à la structure de l'opérateur de covariance du processus mais au fait qu'il s'agit de séries à termes positifs. Nous allons montrer qu'une série à termes positifs, qui converge, presque pour tout t , vers une fonction (son enveloppe supérieure) est convergente quasi uniformément. Cette notion de quasi uniformité est une notion de convergence uniforme sauf sur une petite partie de T ; partie « qui peut se déplacer » mais dont

la dimension est laissée au libre choix de l'utilisateur; de ce fait cette restriction de l'uniformité presque partout n'est pas un handicap dans les applications pratiques.

V-2-c) Théorème

Toute suite $\{f_n\}$ de fonctions croissantes sur T convergentes presque partout vers une fonction f converge quasi uniformément vers cette fonction au sens suivant:

Pour tout $\varepsilon > 0$ et tout $\delta > 0$, il existe un sous ensemble mesurable $\mathcal{A}$ de l'intervalle T et il existe un entier N , tel que pour tout $n > N$ on ait : $\mu(\mathcal{A}) < \delta$ et $\sup_{t \in T \setminus \mathcal{A}} |f_n(t) - f(t)| < \varepsilon$

Démonstration :

Posons :

$$A_{\varepsilon, n} = \{t \in T, f(t) - f_n(t) \geq \varepsilon\}$$

et $\bigcap_{n \in \mathbb{N}} A_{\varepsilon, n} = A_{\varepsilon, \infty}$

Alors, la suite $\{f_n\}$ étant croissante on a : $A_{\varepsilon, \infty} \subset A_{\varepsilon, n+1} \subset A_{\varepsilon, n}$

Montons que $\mu(A_{\varepsilon, \infty}) = 0$

Raisonnons par l'absurde : supposons que $\mu(A_{\varepsilon, \infty}) > 0$ alors:

La suite $\{f_n\}$ serait divergente sur $A_{\varepsilon, \infty}$, ensemble de mesure non nulle, ce qui est contraire à l'hypothèse. Comme $\mu(A_{\varepsilon, \infty}) = 0$ pour tout $\delta > 0$ donné il existe N_0 tel que dès que $n > N_0$ on a :

$$\mu(A_{\varepsilon, n}) \leq \mu(A_{\varepsilon, N_0}) \leq \delta$$

c'est à dire : $\sup_{t \in T \setminus \mathcal{A}} (f(t) - f_n(t)) < \varepsilon$ pour $\mathcal{A} = A_{\varepsilon, N_0}$

ce qui termine la démonstration.

Remarques:

- L'application de ce lemme à la série $\sum_{i \in \mathbb{N}} \lambda_i f_i^2$ assure sa quasi convergence presque partout.
- On remarquera au passage que le Théorème V-2-a) démontre que la série $\sum_{i \in \mathbb{N}} \lambda_i$ est convergente. C'est à dire que l'opérateur de covariance est nucléaire.

CHAPITRE VI

APPLICATIONS PRATIQUES

VI-1) ORGANISATION DES ÉTUDES PRATIQUES

Notre travail avait pour but de définir des outils d'analyse et de prévision des processus ne suivant pas de loi connues à l'avance. Il apparaît naturel de réaliser une application sur des données réelles. Nous avons porté notre choix sur deux jeux de données:

- La production industrielle en France de 1980 à 1988 (données INSEE)
- La consommation d'électricité en France, par trimestre, dans les différents secteurs économiques de 1972 à 1980 (données E.D.F.)

Sur ces deux jeux de données nous réaliserons une vérification des différents théorèmes démontrés :

1. La vérification du théorème lié à l'indice fort.
2. La vérification des proximités des fonctions propres (théorème de principal)
3. Le comportement des composantes temporelles en fonction de la répétition des données.
4. La vérification de la croissance des valeurs propres en fonction de la longueur des intervalles.

Puis nous réaliserons des prévisions pratiques que nous comparerons à des prévisions faites sur les mêmes données d'une façon plus traditionnelle: c'est-à-dire courbe par courbe.

La seconde étude comporte une originalité qui a attiré notre attention. La quatrième composante temporelle a un prolongement inattendu qui induit une erreur de prévision grave. Cette situation est à redouter car elle entraîne des erreurs sur toutes les prévisions de toutes les données. Nous avons trouvé une méthode pour éviter ce genre d'erreur. Cela nous engage vers deux réflexions.

1. La prévision systématique voire automatique est possible mais dangereuse car elle peut engendrer des erreurs très grossières.
2. Les techniques de prévision devront être utilisées par des opérateurs avertis sous peine d'engendrer les plus grandes fantaisies.

Pour valider nos résultats nous avons utilisé des données sur des périodes légèrement plus courtes que les périodes réelles pour pouvoir comparer les différentes prévisions et la réalité.

Tous ces résultats ont été réalisés à l'aide du logiciel que nous avons développé. Dans ce chapitre nous ne présenterons que les résultats et les tableaux significatifs. L'ensemble des résultats et des tableaux étant reportés en annexe.

VI-2) INDICES TRIMESTRIELS DE LA PRODUCTION INDUSTRIELLE EN FRANCE DE 1980 À 1988

VI-2-a) Présentation des données.

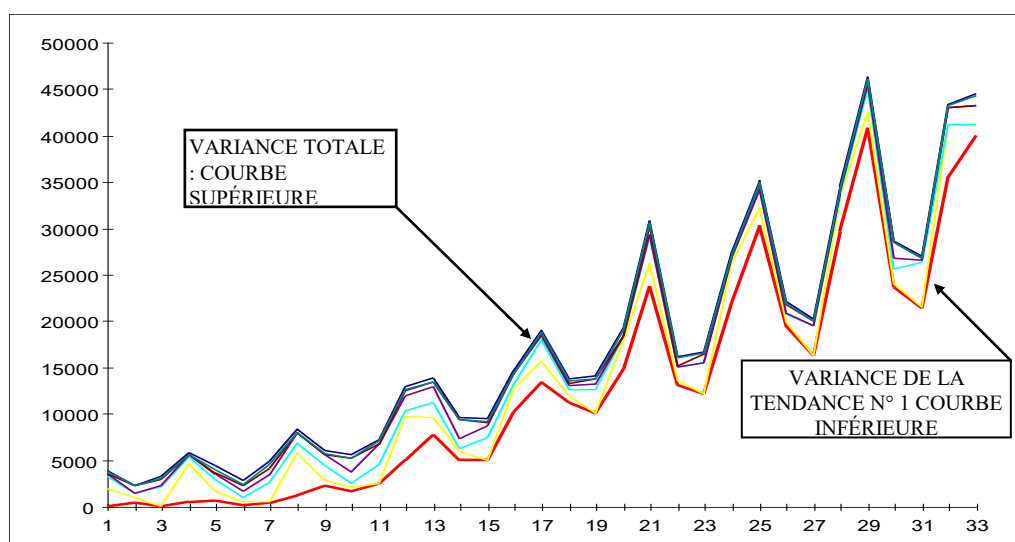
Les données sont extraites de l'annuaire rétrospectif de la France de 1948 à 1988 de l'INSEE. Elles concernent un ensemble de 28 séries chronologiques décrivant la production trimestrielle des différents secteurs de l'économie française de 1980 à 1988. Ces données peuvent être rangées dans un tableau de 28 lignes et 36 colonnes.

Dans la majorité des cas nous avons gardé un tableau de 28 lignes et 33 colonnes afin de pouvoir comparer les prévisions avec la réalité.

VI-2-b) La vérification des théorèmes liés à l'indice fort.

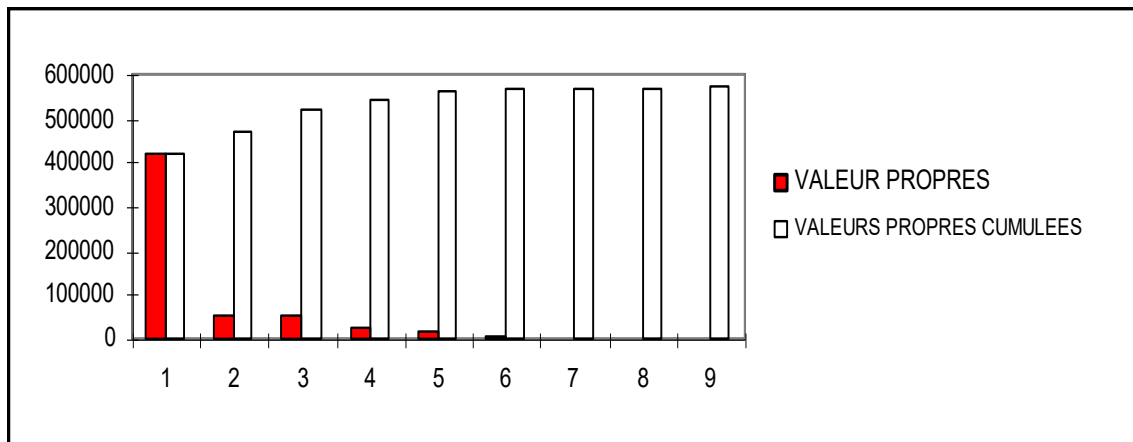
Le logiciel que nous avons élaboré calcule les 33 valeurs propres et les 33 vecteurs propres du noyau de covariance qui n'est ici qu'une matrice. En réalité dès le rang numéro 15 les valeurs propres sont nulles.

Sur les 15 valeurs propres et vecteurs propres restantes l'indice fort (V-2-b) nous permet de sélectionner 6 composantes temporelles comme le montre le graphique ci dessous :



Le choix de la sixième composante temporelle se fait par la proximité des variances cumulées des six premières composantes temporelles et de la variance globale du processus.

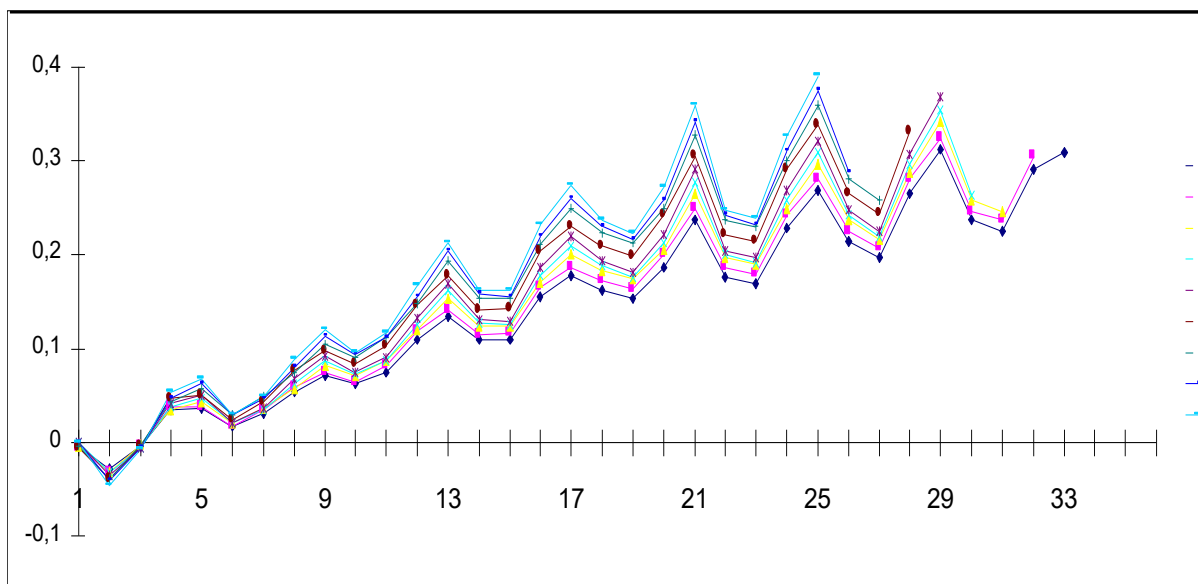
L'information portée par les composantes temporelles 7 à 15 peut être considérée comme un bruit. Les six premières composantes temporelles apportent 98,56 % de l'information mais les cinq premières 97,52% et les 7 premières 99,02. Sans l'indicateur fort nous aurions eu des difficultés à prendre une décision comme le montre le graphique des 9 premières valeurs propres ci-dessous.



VI-2-c) Représentation des composantes temporelles

Nous allons vérifier la proximité des fonctions propres (théorème de Principal) et de la croissance des valeurs propres en fonction de la longueur des intervalles.

La prévision des processus à l'aide du prolongement des composantes temporelles suppose que l'information de la composante temporelle de rang n sur l'intervalle $[1, T]$ contienne une grande partie de l'information sur la composante temporelle de même rang calculée sur l'intervalle $[1, T+1]$. Plus simplement dit, les composantes temporelles de même rang ont « les mêmes formes » sur des intervalles communs. Nous avons représenté sur les graphiques ci-dessous les composantes temporelles de rang 1 pour des intervalles de 25 trimestres à 33.



Nous pouvons remarquer que les vecteurs propres se « collent » les uns aux autres. Cette observation permettra de faire une correction lors de la prévision. Pour ne pas trop charger le texte nous avons représenté l'ensemble des composantes temporelles en annexe.

Dans le tableau ci-dessous nous reportons les valeurs propres des six premières composantes temporelles calculées sur les intervalles de 25 à 33 trimestres.

	Trimestre N°25	Trimestre N°26	Trimestre N°27	Trimestre N°28	Trimestre N°29	Trimestre N°30	Trimestre N°31	Trimestre N°32	Trimestre N°33
Composante temporelle N°1	201020	218841	234219	262319	302349	324490	345195	379772	419288
Composante temporelle N°2	38722	39434	40265	43574	45061	46059	48070	50590	51766
Composante temporelle N°3	34666	35318	38154	38426	41257	42710	44871	48043	48164
Composante temporelle N°4	15353	15405	16430	16799	17821	18466	19034	21069	23188
Composante temporelle N°5	9121	11247	11247	12357	12358	15289	15372	15654	15736
Composante temporelle N°6	2271	2663	2663	3083	3888	4258	4387	4806	5986

Tableau des valeurs propres

Sur ce tableau nous pouvons observer que les valeurs propres croissent en fonction de la longueur de l'intervalle. Cela correspond à l'intuition : « plus le temps s'écoule plus l'information que l'on a du processus est importante ». Il ne faut cependant pas oublier le sens du théorème de principal et de son lemme. Les vecteurs propres sont d'autant plus proches que les valeurs propres de rangs successifs sont écartées.

Il est à remarquer que les valeurs propres 2 et 3, quelque soit le rang sont relativement peu écartées (environ 7% alors que les autres écarts sont de 20% à 80%). On peut interpréter ce résultat en disant que les composantes temporelles 2 et 3 engendrent un seul espace de dimension 2. Cet espace contient deux composantes admettant une période de 4 mois. Cet espace peut être identifié comme l'espace des saisons .

VI-2-d) Comportement des composantes temporelles en fonction de la répétition:

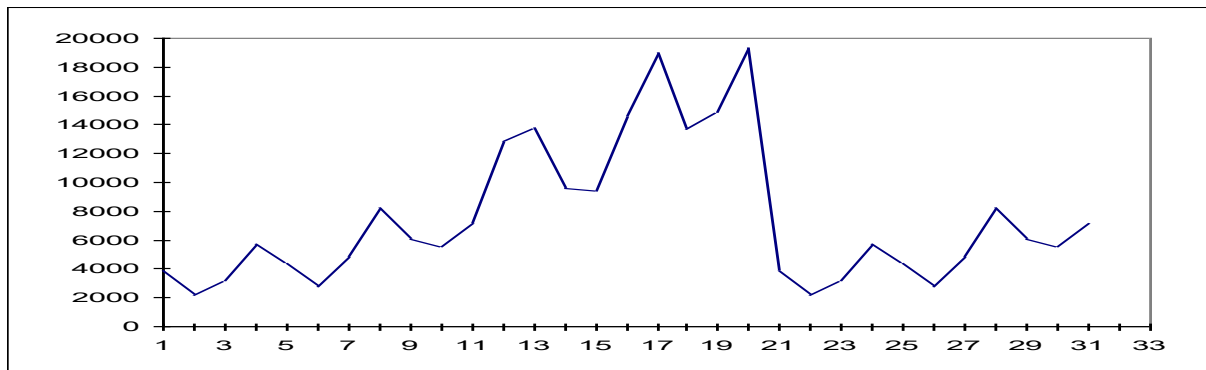
Si les données se répètent sans que le processus soit obligatoirement périodique les composantes temporelles du processus global comportent des répétitions sur les mêmes intervalles que celles du processus (IV-2-d). Pour illustrer ce résultat nous avons reconstitué un processus artificiellement en prenant 20 trimestres des indices de production industrielle que nous avons complété par les 10 premières valeurs de ce processus pour obtenir un processus défini sur 30 périodes.

Nous avons analysé les nouvelles données et recherché les composantes temporelles.

Les tableaux ci-dessous illustrent les résultats attendus :

Représentation de la fonction variance :

Cette fonction comporte les répétitions sur les mêmes intervalles que celles du processus.



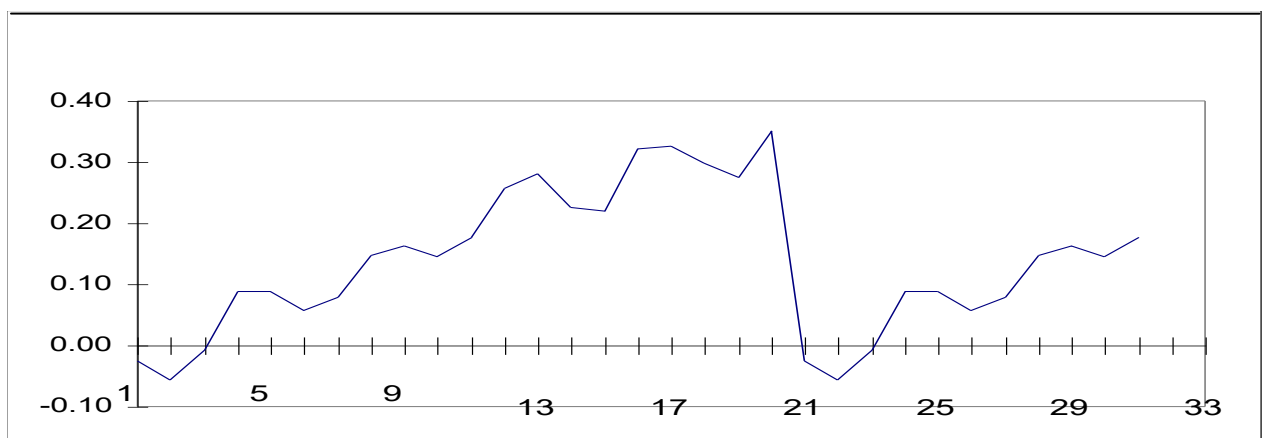
Dans les deux tableaux ci-après nous avons représenté les deux premières composantes temporelles et les valeurs propres correspondantes, sous forme de tableaux et de graphiques

COMPOSANTE TEMPORELLE N°1

Valeur propre 118034

Les 31 valeurs du vecteur propre correspondant avec les répétitions à partir du rang 21

1	2	3	4	5	6	7	8	9	10	11
-0.03	-0.06	-0.01	0.09	0.09	0.06	0.08	0.15	0.16	0.15	0.18
12	13	14	15	16	17	18	19	20		
0.26	0.28	0.23	0.22	0.32	0.33	0.30	0.28	0.35		
21	22	23	24	25	26	27	28	29	30	31
-0.03	-0.06	-0.01	0.09	0.09	0.06	0.08	0.15	0.16	0.15	0.18



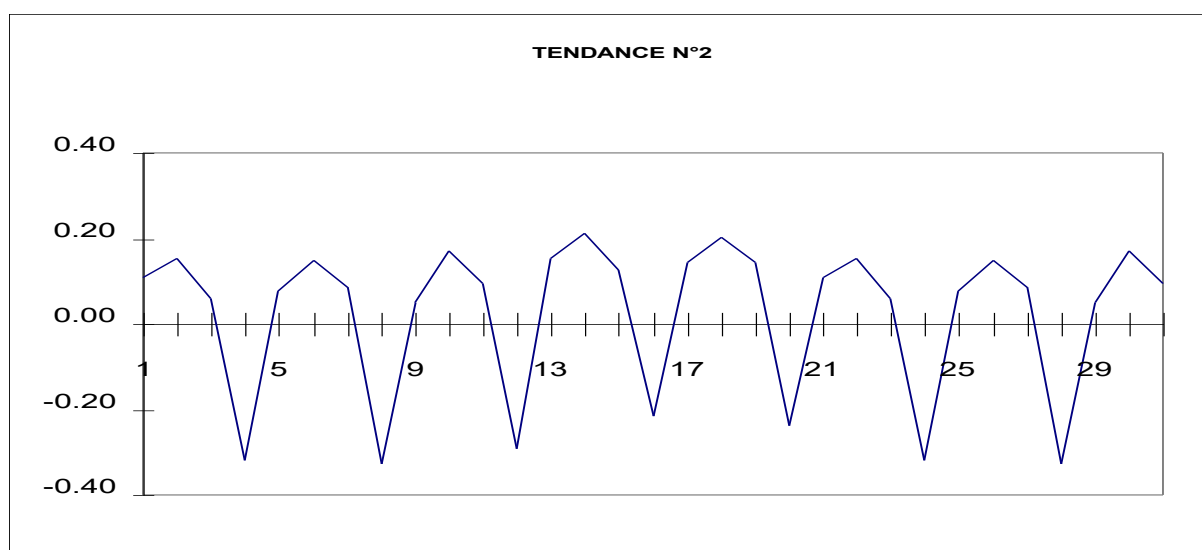
Présentation de la composante temporelle n°1

COMPOSANTE TEMPORELLE N°2

Valeur propre 45135

Les 31 valeurs du vecteur propre correspondant avec les répétitions à partir du rang 21

1	2	3	4	5	6	7	8	9	10	11
0.11	0.15	0.06	-0.32	0.08	0.15	0.09	-0.33	0.05	0.17	0.09
12	13	14	15	16	17	18	19	20		
-0.29	0.15	0.21	0.13	-0.22	0.14	0.20	0.14	-0.24		
21	22	23	24	25	26	27	28	29	30	31
0.11	0.15	0.06	-0.32	0.08	0.15	0.09	-0.33	0.05	0.17	0.09



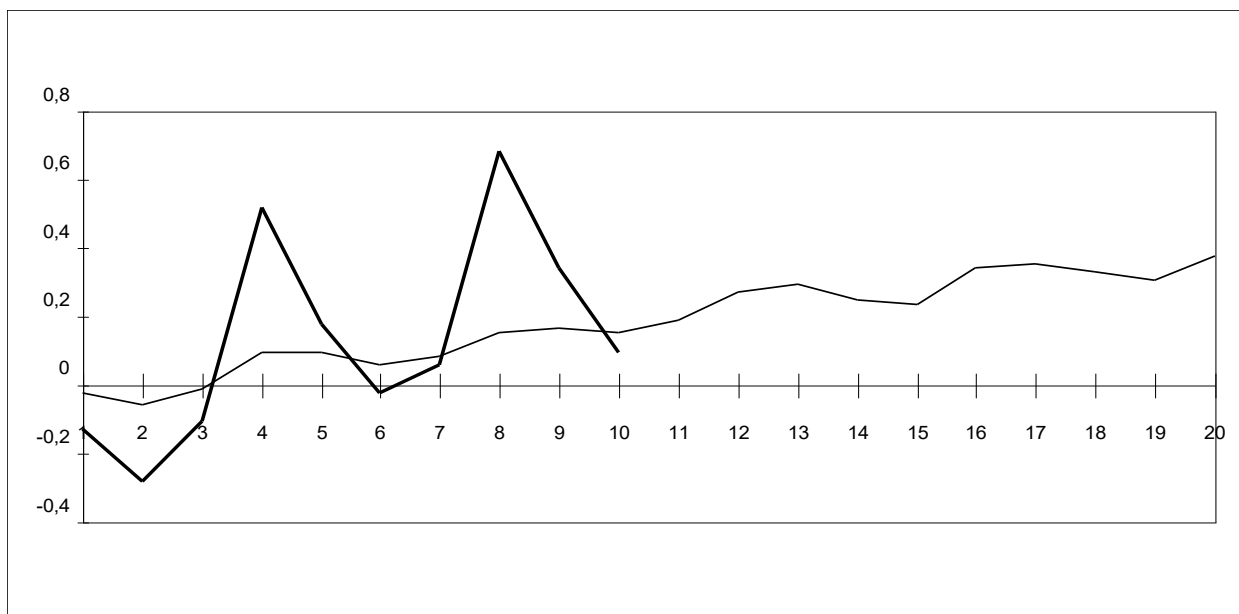
Présentation de la composante temporelle N°2

Remarque importante:

Une idée fausse consisterait à croire que les composantes temporelles d'un processus avec répétition ne sont que la juxtaposition des composantes partielles. Cette idée suppose que les composantes temporelles sont calculées période après période, alors que c'est l'ensemble de instants qui rentrent en compte dans les calculs. Ce n'est que s'il y a périodicité parfaite (IV-2-d), et non répétition parfaite (IV-2-e), que les composantes temporelles sont les répliques des composantes du processus de base.

Dans le graphique ci-après nous avons représenté la première composante temporelle de chaque processus restreint (20 et 10 trimestres) issus du processus à répétition parfaite. Les écarts et les variations sont très significatifs.

Ci dessous les premières composantes temporelles calculées sur 20 trimestres (courbe fine) puis avec les mêmes données sur les 10 premiers trimestres (courbe épaisse).



Aucune répétition n'apparaît comme c'était le cas sur les composantes temporelles calculées sur le processus global.

VI-3-e) Les différentes prévisions:

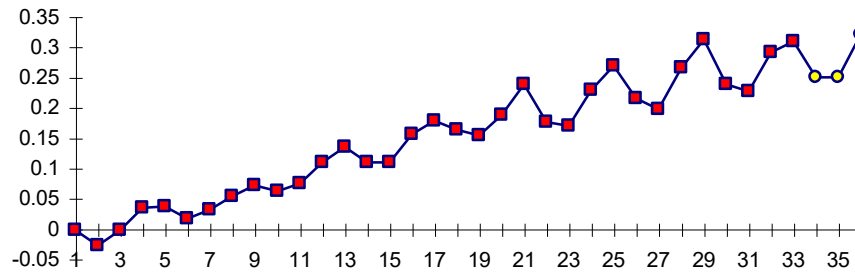
Pour réaliser les prévisions pratiques il suffit de prédire les valeurs futures des composantes temporelles et de la moyenne, le logiciel calculant alors la reconstitution des données et leurs prévisions. Nous sommes ramenés à un problème de prolongement de 7 séries chronologiques. Notre but est de comparer la prévision courbe de données par courbe de données à la prévision faite à l'aide de l'analyse. Nous avons utilisé la décomposition saisonnière qui s'impose pour ces données.

A titre d'exemple voici la prévision sur trois trimestres de la première composante temporelle illustrée par ce tableau de valeurs et ces cinq graphiques:

La composante temporelle et ses quatre saisons.

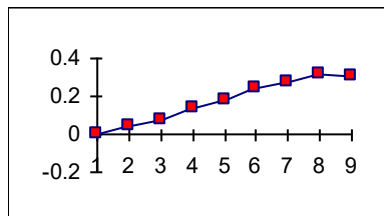
		Valeurs et prévisions								
		1980	1981	1982	1983	1984	1985	1986	1987	1988
Trimestre N°1		-0	0.037	0.071	0.135	0.178	0.238	0.269	0.312	0.309
Trimestre N°2		-0.03	0.017	0.062	0.109	0.163	0.176	0.215	0.238	0.25
Trimestre N°3		-0	0.031	0.075	0.109	0.154	0.17	0.197	0.226	0.25
Trimestre N°4		0.034	0.053	0.109	0.156	0.187	0.229	0.266	0.291	0.32

Ensemble de la composante temporelle

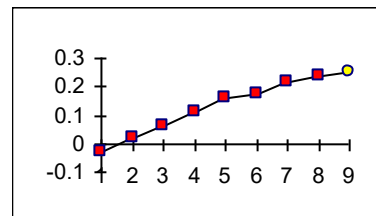


Les trois dernières valeurs sur la courbe (points ronds) représentent les trois prévisions faites grâce aux courbes saisonnières ci dessous.

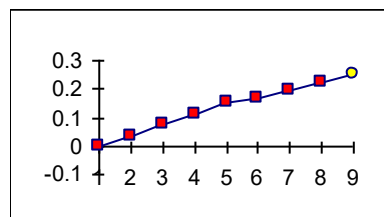
Premiers trimestres



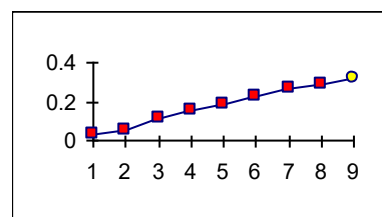
Seconds trimestres



Troisièmes trimestres



Quatrièmes trimestres



En annexe I on trouvera trois tableaux récapitulatifs (un par saison) des prévisions faites correspondant respectivement à la prévision du 34^{ème}, 35^{ème} et 36^{ème} trimestre. Ces tableaux comportent 6 colonnes significatives, les autres colonnes représentant soit les écarts soit un calcul intermédiaire:

1. Les données brutes c'est à dire les valeurs réelles du processus une fois le temps écoulé.
2. Les données reconstituées sur 6 composantes temporelles; c'est à dire les valeurs projetées des données brutes sur l'espace engendré par six composantes temporelles. On obtient ainsi un indicateur de perte minimale lié au bruit. Ici pour le premier trimestre prévu il s'agit de 1,3 %.
3. La prévision brute, c'est à dire la prévision sans correction
4. Une première correction liée à une interprétation du théorème de Principal. Nous avons observé que les composantes temporelles se déforment chaque fois que l'intervalle augmente. Nous avons fait une projection de cette déformation pour corriger les prévisions des composantes temporelles.

5. Une seconde erreur systématique vient du fait de l'utilisation dans la recombinaison des données du coefficient de projection sur la composante temporelle présente et non future. Plus précisément lors du calcul nous utilisons la formule:
$$X(\omega, t) = m + \sum_1^6 \langle X(\omega), f_i \rangle_{T'} f_i(t)$$
 avec T' intervalle de 33 trimestres alors que le calcul réel devrait être fait avec le coefficient calculé sur T' intervalle de 34, 35 ou 36 trimestres selon le cas. La correction de seconde espèce correspond à une multiplication de l'écart à la moyenne par 34/33, 35/33 ou 36/33 suivant le trimestre prévu.
6. L'avant dernière colonne représente les données prévues si nous avons fait la prévision courbe par courbe, soit au total 28 prévisions.

Les résultats de cette étude montrent que nous avons eu, pour le premier trimestre près de 1,9% de gain dans la prévision en moyenne et 0,4 %, 0,3 % pour les deux autres trimestres.

Pour terminer ce paragraphe nous ferons une remarque. Le prolongement des composantes principales relève d'un prolongement de séries sans interprétation probabiliste. En effet les composantes temporelles ne sont pas des réalisations d'un processus mais sont la structure même du processus étudié. Cette remarque pose tout le problème de la légitimité de la méthode utilisée lors du prolongement des composantes.

VI-4) LA CONSOMMATION D'ÉLECTRICITÉ EN FRANCE, PAR TRIMESTRE, DANS LES DIFFÉRENTS SECTEURS ÉCONOMIQUES DE 1972 À 1980

VI-4-a) Présentation des données :

Les données qui ont été choisies pour illustrer notre travail par un second exemple ont été proposées par EDF. Il s'agit d'essayer d'analyser et de prévoir la consommation d'électricité par secteur économique. Les données portent sur la consommation en électricité de 27 secteurs d'activités au plan national. Les enregistrements sont faits sur 20 trimestres.

VI-4-b) Le traitement des données

Ce second exemple a été traité comme le précédent et le détail des résultats et des analyses se trouve en annexe. Nous laisserons le lecteur s'y reporter. Cependant nous ferons quelques remarques sur les résultats qui du premier abord ne sont pas très satisfaisants.

VI-4-c) L'attente minimale d'erreur:

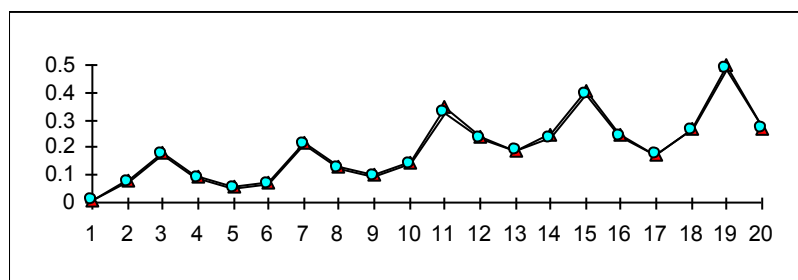
L'analyse à l'aide de l'indice fort nous amène à garder 4 composantes temporelles; le fait de garder uniquement 4 axes induit automatiquement une erreur. En effet quand nous reconstituons les données à l'aide des 4 composantes temporelles la perte se fait sur les composantes non prise en compte. Pour mesurer cette perte plusieurs indices peuvent être mis en place. Par exemple en utilisant la norme de l'indice fort de perte. Pour être plus proche des données, dans la mesure où nous connaissons les valeurs du processus sur le 20^{ème} trimestre que nous voulons prévoir (donc les valeurs réelles des composantes sur le trimestre numéro 20), nous avons reconstitué avec les valeurs vraies le processus à l'aide des 4 premières composantes connues sur la totalité des 20 trimestres. L'erreur moyenne minimale prévisible est alors de moins de 1% (voir tableau récapitulatif en annexe)

VI-4-d) La particularité de l'exemple :

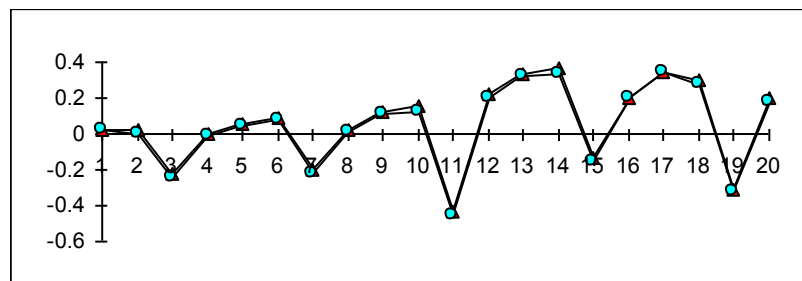
Quand on réalise la prévision en prolongeant les tendances l'erreur moyenne est alors de 9%. Une telle erreur aussi loin de l'erreur prévisible est inattendue. Elle ne peut venir que d'un prolongement maladroit des tendances. C'est pourquoi nous avons eu l'idée de regarder les résultats en utilisant seulement trois axes. La constatation est là : les résultats auraient été meilleurs. Nous avons eu alors la curiosité de comparer les tendances réelles et celles prévues. Voici dans les quatre tableaux ci-dessous les résultats.

Les données réelles comportent sur la courbe un point rond et les prévisions un point triangulaire

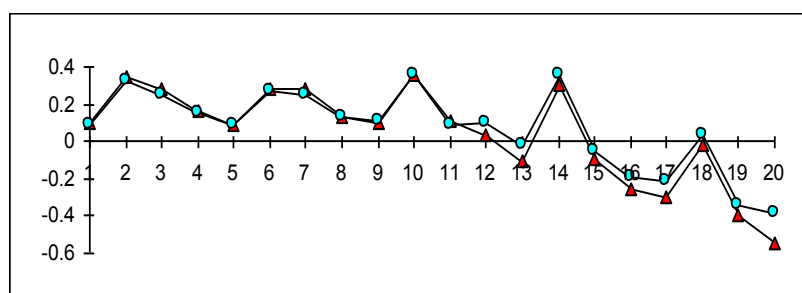
Tendance N°1 : prévision et réalité (il y a deux courbes superposées)



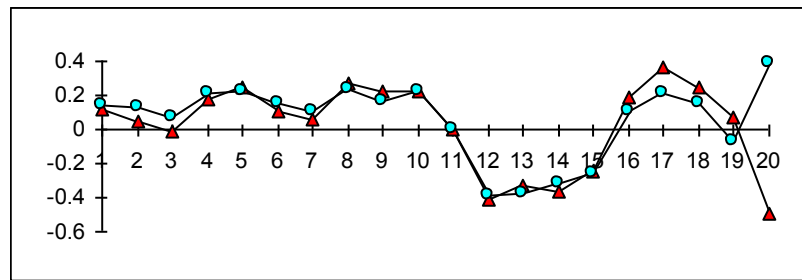
Tendance N°2 : prévision et réalité (il y a deux courbes superposées)



Tendance N°3 : prévision et réalité (il y a deux courbes superposées)



Tendance N°4 : prévision et réalité
(il y a deux courbes superposées)



La question qui se pose alors est de savoir, comment, sur le plan pratique, nous aurions pu nous douter de cette erreur et la corriger. Car notre vérification est basée sur le fait que nous connaissons les valeurs que nous prétendons prévoir !!!.

VI-4-e) Une proposition pour éviter des erreurs de prévision

Comment peut on sortir de cette situation ?

Rappelons que la méthode de prévision que nous avons adoptée est basée sur la prévision la plus « exacte possible » de la moyenne et des tendances sélectionnées.

La première précaution à prendre est de traiter des processus comportant un nombre important de séries chronologiques pour éviter l'impact trop important de phénomènes totalement aléatoires en particulier sur la fonction moyenne.

La difficulté reste cependant entière lorsque l'erreur est faite, comme c'est le cas dans cet exemple, sur la prévision d'une tendance. Sans information supplémentaire la difficulté reste entière.

Imaginons que nous connaissions des informations a priori sur la prévision de certaines réalisations du processus. Par exemple dans les prévisions que nous avons faites avec 4 tendances, la fonction N° 2 , Métaux rares (voir tableau récapitulatif), prend la valeur -527 alors qu'elle prend la valeur 206 avec 3 tendances. La conclusion est alors que la prévision de la tendance N°4 est suspecte. En fonction des données la valeur attendue peut être aux environs de 400 et le coefficient de projection des Métaux rares sur la tendance N°4 est de 761 (voir sorties logiciel); d'où l'égalité :

$$400 \cong 206 + 761 \times f_4(20) \quad \text{soit} \quad f_4(20) \cong 0.26$$

Ce n'est pas la valeur attendue qui est de 0.39 mais c'est loin de l'erreur faite en ayant prévu - 0.5 . D'ailleurs une prévision avec cette valeur donne 4% d'erreur moyenne alors que celle faite avec -0.5 donne 9%.

Il est à remarquer que la connaissance de 5 prolongements de données pour un choix de 4 tendances détermine un système d'équations dont les inconnues sont les prolongements des tendances . Un calcul automatique de prévision pourrait ainsi être mis en place.

L'utilisation de méthodes de prévision de séries chronologiques, appliquées aux tendances, pourrait aussi être une façon automatique de prévision. Sur le plan pratique et sur le plan de la programmation il reste beaucoup de chemins à explorer, et tout un savoir-faire à mettre en place.

CHAPITRE VII

CONCLUSIONS ET PROSPECTIVES

VII-1) PREMIÈRES CONCLUSIONS

- Les savoir-faire usuels

Traditionnellement, dans les problèmes de prévision on ne possède qu'une série chronologique. Cette situation, relativement pauvre en données a amené les chercheurs à considérer la série chronologique comme une réalisation d'un processus. Ce processus joue le rôle du modèle. Dans cette lignée là, la méthode de Box et Jenkins fait actuellement référence pour les séries chronologiques longues. Le lissage exponentiel simple étant habituellement préféré pour les séries courtes.

- L'essentiel de notre méthode de prévision

Notre méthode de prévision est basée sur la connaissance d'une famille de séries chronologiques. En enrichissant ainsi les données nous avons pu utiliser la décomposition de Karhunen et Loeve. Le processus étudié a pu ainsi être décomposés en un nombre restreint de processus élémentaires. La prévision se ramenant au prolongement de ces processus élémentaires.

L'efficacité de la description d'un processus à l'aide de l'analyse harmonique a été confirmée et enrichie pour devenir un outil de prévision global d'un processus.

L'avantage de cette méthode provient de l'utilisation de l'information sur l'espace individu/temps alors que les autres méthodes ne s'appuient que sur l'information donnée par une seule réalisation. Sur le plan technique on peut remarquer que le nombre de composantes temporelles significatives est généralement restreint par rapport aux nombre de trajectoires fournies. C'est pour cette raison que nous recommandons l'analyse harmonique lorsque les données comportent un nombre de trajectoires importantes.

Pour traiter les différents exemples nous avons mis au point un programme qui est à la disposition de tout lecteur qui en ferait la demande.

La limite de la méthode proposée pour analyser les processus à l'aide de l'analyse harmonique se trouve dans la nature des données. Si les données se résument à quelques trajectoires cette méthode est tout à fait à proscrire et le recours à un modèle semble indispensable. (on peut imaginer cependant des méthodes mixtes que nous n'avons pas développées).

Le champ d'investigation du rapport entre l'analyse Harmonique et la prévision des séries chronologiques étant très important nous n'avons pas exploré tous les domaines. Nous proposons dans les paragraphes suivants quelques pistes possibles.

VII-2) AUTRES PERSPECTIVES LIÉES À L'ANALYSE HARMONIQUE:

VII-2-a) Composantes temporelles et données manquantes

Les données manquantes sont des valeurs d'une série qui ne sont pas connues pour des raisons matérielles ou humaines. Cette situation se rencontre assez fréquemment dans le domaine médical. Par exemple le processus évalue une variable sur des malades et ses valeurs sont au temps t les résultats d'une analyse médicale depuis le début de la maladie. Si le malade n'est pas venu en consultation régulièrement, le laboratoire dispose des données incomplètes qui rendent le diagnostic difficile. La reconstitution des données manquantes est alors une priorité.

Voici une démarche que nous proposons pour compléter les données manquante d'une série chronologique:

Nous suggérons de « plonger » les données de la série $X(t)$ dans un processus $Y(\omega, t)$ de données similaires et complètes.

Plus précisément si les données de la série à compléter se présentent sous la forme: $X(t_i)$ avec i compris entre 1 et n dates et si les valeurs de cette séries sont connues sauf pour les dates de t_a à t_b on recherche un processus $Y(\omega, t_i)$ pour i compris entre 1 et n et $\omega \in \Omega$. On plonge les données de la série dans celles du processus, c'est à dire: on pose $Y'(\omega, t_i) = X(t_i)$ et $\Omega' = \Omega \cup \{\omega_o\}$. La probabilité sur chaque élément de Ω' est définie naturellement par $1/(p+1)$ si p désigne le nombre d'éléments de Ω .

Nous pouvons définir un nouveau processus Y' défini sur Ω' par $Y'(\omega, t_i) = Y(\omega, t_i)$ si $\omega \in \Omega$ et $Y'(\omega_o, t_i) = X(t_i)$. Ce processus est alors à données incomplètes.

Pour simplifier les écritures nous noterons T' les nombres de 1 à t_a-1 (resp. T'' de t_a à t_b et T''' de t_b+1 à n) et par abus de langage Y désignera aussi sa restriction sur chacun des intervalles T', T'', T''' .

Alors une question se pose :

Le processus Y' est-il un prolongement acceptable de Y ?

Pour répondre à cette question nous proposons de faire l'analyse harmonique de Y et projeter les données de $X(t_i)$ sur les vecteurs propres associés à cette analyse sur l'intervalle T' puis sur l'intervalle T''' .

On obtient ainsi deux séries de nombres, une sur T' et une sur T'' . Il faudrait pouvoir les comparer (sur quel critère ?) afin de déterminer si le processus Y' est un prolongement satisfaisant de Y . En cas de réponse négative on serait amené à remettre en cause le choix de Y . Par contre, si le prolongement de Y par Y' était acceptable, il est alors raisonnable de reconstituer les données en utilisant la formule:

$$X(t) = C_1 f_1(t) + C_2 f_2(t) + C_3 f_3(t) + C_4 f_4(t) + \dots$$

Avec t élément de T'' et C_i le coefficient de projection de la fonction $X(t)$ soit sur les fonctions propres de Y sur T' soit sur celles de T'' (coefficients qui devraient être proches).

Cette proposition qui n'a pas été testée demande une analyse plus approfondie avant toute mise en oeuvre.

VII-2-b) Utilisation des vecteurs propres lors d'une ACP:

Lors d'une ACP lorsque on a calculé l'ensemble des vecteurs propres et des valeurs propres de la matrice de covariance on étudie essentiellement le comportement des valeurs propres sous forme de pourcentage d'inertie. Nous proposons en annexe une utilisation des vecteurs propres pour séparer les variables entre elles. Cette proposition peut apporter quelques commodités lors d'une analyse de questionnaires par exemple

VII-2-c) Nécessité de généraliser l'analyse harmonique en « analyse des besoins » et en stratégie « qualité » :

La gestion des ressources humaines utilise de plus en plus de questionnaires. L'évolution dans le temps de ces questionnaires représente une information précieuse mais difficile à utiliser sans un traitement approprié. L'enregistrement des réponses sur chaque question peut être interprétée comme un processus aléatoire. L'analyse de l'évolution de chaque question au cours du temps est insuffisante. Un cadre plus général que l'analyse harmonique semble faire défaut. On pourra trouver quelques éléments de réponse dans G.Saporta [26]page 78 au chapitre analyse factorielle d'un processus vectoriel.

La mise en place d'un outil d'analyse des données multi-variables évoluant dans le temps représente un besoin; elle permettrait le rapprochement et la mise en commun des savoir-faire de l'analyse des données et des techniques de management (analyse des besoins, techniques d'audit...). Si de telles procédures étaient utilisées nous avons la conviction qu'il serait possible d'optimiser les investissements et les subventions aussi bien dans les secteurs publics que privés. Les plus grosses difficultés résideront certainement dans la résistance des mentalités.

ORGANISATION DES ANNEXES

Les annexes comportent cinq parties:

Les deux premières annexes sont les détails de l'étude de deux jeux de données; l'un sur les indices de la production industrielle en France de 1980 à 1988 et l'autre sur la consommation d'électricité en France de 1972 à 1980. Ces études proposent des éléments d'analyse qui ne pouvaient pas trouver leurs places dans la partie théorique. Ces deux annexes comportent un grand nombre de tableaux qui permettent au lecteur de retrouver et d'analyser tous les calculs dans les moindres détails. Ces tableaux sont précédés d'un commentaire général explicatif. Lorsque la lecture d'une série de tableaux demande des précisions nous les avons mises juste avant pour éviter des reports trop lointains.

La troisième annexe propose une utilisation des techniques de l'analyse harmonique lors d'une ACP. Le texte est présenté plutôt comme un article qui se termine par la présentation des calculs sous forme de deux séries de tableaux

La quatrième annexe donne des rappels des démonstrations sur les résultats utilisés couramment en analyse harmonique. Cette partie évitera au lecteur non coutumier à cette théorie de rechercher des résultats souvent épars.

La cinquième annexe présente simplement le programme informatique qui a été réalisé et qui a permis de traiter les données.

ANNEXE I

LA PRODUCTION INDUSTRIELLE EN FRANCE DE 1980 À 1988 (DONNÉES INSEE)

Dans cette annexe nous présentons l'ensemble des calculs que nous avons effectués pour analyser et faire une prévision sur les données recueillies dans l'annuaire rétrospectif de la France publiées par l'INSEE. Ces données portent sur la production de 28 secteurs industriels français de 1980 à 1988. L'intérêt de cette annexe dépasse une simple compilation de résultats complétant l'étude du chapitre VI-2. Nous avons essayé de mettre en place des savoir-faire pour réaliser des prévisions. Ces savoir-faire ne peuvent pas rentrer dans le cadre théorique de ce travail, mais peuvent donner des éléments concrets pour de futurs praticiens. Les tableaux de nombres ont été fournis dans le détail, ils sont le plus souvent illustrés de graphiques pour faciliter la lecture.

Vous trouverez les documents suivants:

An-I_1) Liste des produits concernés

Il s'agit des 28 secteurs d'activité dont la production a été relevée par l'INSEE sur 36 trimestres de 1980 à 1988. Nous utiliserons pour faire les prévisions les données que sur 33 trimestres. Les trois derniers trimestre nous permettrons de vérifier la pertinence de nos prévisions.

An-I_2) Sorties logiciel

Le logiciel permet de visualiser l'indice fort de perte d'information et les composantes temporelles afin de réaliser les prévisions; mais il permet aussi de sortir les résultats et les prévisions sous forme de deux fichiers textes facilement récupérables.

Un des fichier contient les composantes temporelles.

Nous avons édité le second fichier. Ce fichier contient les données brutes, les prévisions et les données reconstituées sur le nombre de composantes temporelles choisies (ici 6) grâce à l'indice fort de perte d'information.

Ce fichier contient aussi les coefficients de projection de chaque réalisation du processus sur les différentes composantes temporelles. On peut ainsi vérifier la pertinence du choix fait à l'aide de l'indice fort. En effet à partir du rang 7 les projections des différentes données sont relativement faibles. On peut aussi grâce à ces résultats regarder sur chaque réalisation l'impact de chaque composante temporelle. Par l'observation de ces résultats il est possible de rendre les prévisions plus fines. Pour les spécialistes de la prévision ces informations peuvent être précieuses pour ajuster des prévisions de composantes temporelles.

Enfin , pour faciliter la lecture des erreurs faites celles-ci sont affichées en pourcentage.

An-I_3) Vérification expérimentale du théorème principal [IV-I-e]

Un des résultats les plus significatifs de ce travail a été la démonstration du théorème principal [page 54]. L'idée majeure de ce théorème est la conservation de la « forme » des composantes temporelles indépendamment du temps écoulé. Il était nécessaire, en fonction de l'indice fort, de faire le calcul au moins pour les six premières composantes. On pourra observer au passage que la proximité des composantes temporelles est d'autant plus grande que les valeurs propres sont distantes les unes des autres.

An-I_4) Prolongements de tendances

Les composantes temporelles et la moyenne ont été prolongées d'une façon que nous avons qualifié de « naïve », c'est à dire par un ajustement linéaire appliqué aux courbes saisonnières. Une automatisation, par exemple par ajustement polynomial, pourrait être envisagée. Nous verrons lors du prochain exemple qu'il faut être prudent dans ce domaine.

An-I_5) Correction de la prévision des composantes temporelles

Nous proposons une méthode de correction des erreurs faites sur les composantes. Ces corrections sont basées sur l'information obtenue par le calcul des composantes temporelles sur les périodes antérieures à la prévision.

An-I_8) Prévisions « naïves » trajectoire par trajectoire

Proposer une nouvelle méthode de prévision suppose implicitement que l'on ait l'espoir d'améliorer les prévisions faites antérieurement. Les méthodes de prévision utilisées généralement, s'appliquent sur une seule réalisation du processus. Pour pouvoir comparer, les résultats obtenus par l'analyse harmonique et les résultats obtenus par une analyse classique nous avons dû réaliser une prévision trajectoire par trajectoire

An-I_7) Tableaux récapitulatifs des différents résultats

Nous avons rassemblé dans un même tableau pour chaque trimestre prévu, les données brutes, les données reconstituées sur six composantes temporelles, les prévisions brutes, les prévisions corrigées, les prévisions naïves. Ces tableaux récapitulatifs permettent de se faire une idée de la pertinence de la méthode proposée.

An-I-1) Liste des produits concernés

1	Industrie des biens intermédiaires	16	Produits de la parapharmacie et de la pharmacie
2	Biens d'équipements professionnels	17	Produits de la fonderie et du travail des métaux
3	Biens d'équipements ménager	18	Produits de la mécanique
4	Biens de consommation courante	19	Matériels électriques et électroniques professionnel
5	Bâtiment, génie civil	20	Matériels électriques et électroniques ménagers
6	Viandes et produits laitiers	21	Matériels du transport terrestre
7	Autres produits agricoles	22	Construction navale et aéronautique
8	Combustibles minéraux solides	23	Produits textiles et habillement
9	Pétrole brut, gaz naturel	24	Cuirs et chaussures
10	Électricité, gaz et eau	25	Bois et meubles
11	Minerais, métaux ferreux	26	Papiers et cartons
12	Minerais, métaux non ferreux	27	Imprimerie, presse
13	Matériaux de construction	28	Caoutchouc et matières plastiques
14	Produits de l'industrie du verre		
15	Produits chimiques de base		

NOMBRE DE REALISATIONS 28
 NOMBRE DE TRIMESTRES PRIS EN COMPTE 33
 NOMBRE DE DONNEES RECONSTITUEES OU PREVUES 36

FONCTION N° 1

DONNEES BRUTES

1094	1058	860	988	1005	1012	838	997	998	1021
806	960	982	1002	807	968	1007	978	826	972
958	1000	850	967	962	1007	847	969	963	1024
877	1013	1056	1077	929	1065				

DONNEES RECONSTITUEES ET PREVISION

1090	1057	867	986	1001	1012	838	994	999	1017
809	961	989	1004	807	960	1006	974	821	973
959	995	846	968	962	1014	854	980	966	1024
874	1009	1048	1061	906	1054				

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

0	0	-1	0	0	0	0	0	0	0
0	0	-1	0	0	1	0	0	1	0
0	1	1	0	0	-1	-1	-1	0	0
0	0	1	2	3	1				

COEFFICIENT DE LA FONCTION SUR CHAQUE TENDANCE

-182.773	74	-63.587	16.7	-18.264	9.3	17	-1	-10	
0	12	2	-5	4	75	75	75	75	75
75	75	75	75	75	75	75	75	75	75
75	75	75							

FONCTION N° 2

DONNEES BRUTES

1028	1024	884	1063	1023	1051	907	1081	1039	1060
895	1064	1018	1022	878	1045	1001	983	868	1034
978	992	887	1040	970	994	874	1017	951	987
893	1043	1013	1016	926	1075				

DONNEES RECONSTITUEES ET PREVISION

1030	1025	895	1050	1025	1040	908	1093	1040	1054
894	1062	1020	1022	872	1046	1018	992	871	1032
967	990	883	1036	962	997	873	1028	952	986
896	1041	1012	1016	934	1060				

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

0	0	-1	1	0	1	0	-1	0	1
0	0	0	0	1	0	-2	-1	0	0
1	0	0	0	1	0	0	-1	0	0
0	0	0	0	-1	1				

COEFFICIENT DE LA FONCTION SUR CHAQUE TENDANCE

-73	-80	78	135	-19	19	27	-4	-5	-19
-8	4	-24	-1	7	-80	-80	-80	-80	-80
-80	-80	-80	-80	-80	-80	-80	-80	-80	-80
-80	-80	-80							

FONCTION N° 3

DONNEES BRUTES

1130	1065	785	1021	991	983	754	1016	1002	992
745	1046	1036	1039	753	1008	1011	961	755	1030
976	960	773	1006	984	1042	826	1094	1082	1115
850	1153	1173	1208	904	1208				

DONNEES RECONSTITUEES ET PREVISION

1123	1057	785	1034	999	979	739	1019	997	998
747	1043	1041	1034	750	1008	1020	962	748	1017
969	969	774	1017	992	1048	824	1084	1068	1113
864	1149	1178	1149	885	1209				

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

1	1	0	-1	-1	0	2	0	0	-1
0	0	-1	0	0	0	-1	0	1	1
1	-1	0	-1	-1	-1	0	1	1	0
-2	0	0	5	2	0				

COEFFICIENT DE LA FONCTION SUR CHAQUE TENDANCE

-6	-11	-134	-205	170	35	-57	11	-18	18
-18	-20	-1	11	2	-11	-11	-11	-11	-11
-11	-11	-11	-11	-11	-11	-11	-11	-11	-11
-11	-11	-11							

NOMBRE DE REALISATIONS 28
 NOMBRE DE TRIMESTRES PRIS EN COMPTE 33
 NOMBRE DE DONNEES RECONSTITUEES 36

FONCTION N° 4

DONNEES BRUTES

1095	1039	860	1006	1036	1011	854	1028	1062	1057
857	1028	1105	1059	855	1020	1101	1037	879	1034
1080	1040	903	1039	1076	1071	904	1043	1079	1054
923	1053	1133	1089	945	1089				

DONNEES RECONSTITUEES ET PREVISION

1100	1039	858	1004	1051	1024	855	1029	1058	1049
856	1022	1087	1062	864	1031	1110	1039	880	1036
1081	1042	895	1034	1071	1063	903	1044	1081	1065
920	1058	1129	1096	960	1099				

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

0	0	0	0	-1	-1	0	0	0	1
0	1	2	0	-1	-1	-1	0	0	0
0	0	1	0	1	1	0	0	0	-1
0	0	0	-1	-2	-1				

COEFFICIENT DE LA FONCTION SUR CHAQUE TENDANCE

148	122	-7	48	-3	-16	-9	-22	3	-16
-16	-3	7	-6	-11	122	122	122	122	122
122	122	122	122	122	122	122	122	122	122
122	122	122							

NOMBRE DE REALISATIONS 28
 NOMBRE DE TRIMESTRES PRIS EN COMPTE 33
 NOMBRE DE DONNEES RECONSTITUEES OU PREVUES 36

FONCTION N° 7

DONNEES BRUTES

920	898	841	1341	903	899	845	1402	958	925
855	1392	908	924	844	1334	938	936	847	1368
922	930	877	1382	900	935	872	1359	931	953
914	1416	969	1007	941	1476				

DONNEES RECONSTITUEES ET PREVISION

917	895	844	1343	917	903	827	1387	966	928
859	1398	920	925	843	1320	933	921	847	1362
920	935	877	1390	909	938	883	1361	917	945
920	1426	961	955	919	1400				

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

0	0	0	0	-2	0	2	1	-1	0
0	0	-1	0	0	1	0	2	0	0
0	-1	0	-1	-1	0	-1	0	1	1
-1	-1	1	5	2	5				

COEFFICIENT DE LA FONCTION SUR CHAQUE TENDANCE

220	-999	-153	99	-23	-71	-42	25	-28	16
-6	-6	3	-17	0	-999	-999	-999	-999	-999
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999
-999	-999	-999							

FONCTION N° 8

DONNEES BRUTES

1121	1037	814	1028	1149	1045	804	1107	1088	982
745	989	1051	951	712	970	1017	900	755	937
911	831	682	840	845	820	653	836	806	780
609	761	778	635	559	673				

DONNEES RECONSTITUEES ET PREVISION

1119	1036	815	1030	1077	1022	812	1080	1086	1018
782	1004	1039	966	733	964	1016	878	703	915
933	843	678	872	868	822	641	821	811	754
607	762	786	716	623	754				

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

0	0	0	0	6	2	-1	2	0	-4
-5	-2	1	-2	-3	1	0	2	7	2
-2	-1	1	-4	-3	0	2	2	-1	3
0	0	-1	-13	-12	-12				

COEFFICIENT DE LA FONCTION SUR CHAQUE TENDANCE

-773	68	-270	227	7	-30	27	78	81	37
17	-50	-6	-1	-3	68	68	68	68	68
68	68	68	68	68	68	68	68	68	68
68	68	68							

FONCTION N° 9

DONNEES BRUTES

1074	966	960	1000	978	823	845	921	859	753
741	789	740	709	758	757	847	692	698	772
774	715	742	803	744	670	748	702	707	636
698	745	716	707	712	759				

DONNEES RECONSTITUEES ET PREVISION

1076	967	953	1005	977	845	851	937	853	744
724	790	740	698	748	771	837	707	713	784
781	720	756	787	739	654	732	680	706	643
701	745	736	720	731	800				

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

0	0	1	0	0	-3	-1	-2	1	1
2	0	0	2	1	-2	1	-2	-2	-2
-1	-1	-2	2	1	2	2	3	0	-1
0	0	-3	-2	-3	-5				

COEFFICIENT DE LA FONCTION SUR CHAQUE TENDANCE

-1241	-44	27	-193	-437	10	-76	-54	2	22
9	10	9	19	12	-44	-44	-44	-44	-44
-44	-44	-44	-44	-44	-44	-44	-44	-44	-44
-44	-44	-44							

FONCTION N° 10

DONNEES BRUTES

1185	874	782	1159	1246	957	831	1209	1281	977
817	1172	1322	1022	872	1286	1513	1114	945	1308
1185	874	782	1159	1246	957	831	1209	1281	977
817	0	0							

DONNEES RECONSTITUEES ET PREVISION

1195	877	781	1147	1252	945	828	1222	1280	981
824	1172	1316	1024	880	1280	1502	1114	944	1314
1195	877	781	1147	1252	945	828	1222	1280	981
824	0	0							

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

-0.805	-0.368	0.155	1.013	-0.450	1.251	0.380	-1.063	0.060	-0.367
-0.810	0.040	0.491	-0.218	-0.957	0.485	0.738	-0.028	0.059	-0.461
-0.805	-0.368	0.155	1.013	-0.450	1.251	0.380	-1.063	0.060	-0.367
-0.810	DIVZERO	DIVZERO							

COEFFICIENT DE LA FONCTION SUR CHAQUE TANDENCE

701.155	-214.081	-519.687	-193.187	243.052	35.138	-12.103	5.448	-7.460	-27.662
-21.238	-10.140	1.849	13.233	-5.187	-0.851	1.342	-0.308	0.006	-193.194
-193.194	-193.194	-193.194	-193.194	-193.194	-193.194	-193.194	-193.194	-193.194	-193.194
-193.194									

FONCTION N° 11

DONNEES BRUTES

1147	1080	871	902	948	998	838	954	953	933
662	755	776	808	649	814	870	849	698	839
1147	1080	871	902	948	998	838	954	953	933
662	0	0							

DONNEES RECONSTITUEES ET PREVISION

1132	1081	876	911	965	1008	831	945	946	931
654	748	777	814	648	813	869	852	695	850
1132	1081	876	911	965	1008	831	945	946	931
654	0	0							

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

1.306	-0.048	-0.569	-1.019	-1.749	-1.029	0.776	0.897	0.758	0.240
1.172	0.958	-0.190	-0.762	0.192	0.151	0.077	-0.334	0.419	-1.304
1.306	-0.048	-0.569	-1.019	-1.749	-1.029	0.776	0.897	0.758	0.240
1.172	DIVZERO	DIVZERO							

COEFFICIENT DE LA FONCTION SUR CHAQUE TANDENCE

-694.217	73.684	-41.033	-75.707	-8.116	201.243	-22.884	-30.320	28.021	14.021
1.105	7.440	-0.757	2.022	-11.778	-1.505	-1.865	0.717	-4.313	-75.683
-75.683	-75.683	-75.683	-75.683	-75.683	-75.683	-75.683	-75.683	-75.683	-75.683
-75.683									

FONCTION N° 12

DONNEES BRUTES

1095	1075	843	987	997	1036	840	1032	1057	1074
814	1013	1024	1088	810	1045	1068	1017	761	1016
1095	1075	843	987	997	1036	840	1032	1057	1074
814	0	0							

DONNEES RECONSTITUEES ET PREVISION

1094	1069	842	994	1004	1045	831	1030	1048	1082
822	1007	1043	1061	805	1018	1058	1027	762	1038
1094	1069	842	994	1004	1045	831	1030	1048	1082
822	0	0							

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

0.087	0.543	0.084	-0.719	-0.674	-0.836	1.081	0.231	0.867	-0.704
-0.980	0.581	-1.872	2.480	0.568	2.626	0.895	-0.938	-0.170	-2.202
0.087	0.543	0.084	-0.719	-0.674	-0.836	1.081	0.231	0.867	-0.704
-0.980	DIVZERO	DIVZERO							

COEFFICIENT DE LA FONCTION SUR CHAQUE TANDENCE

-8.447	78.580	-60.237	96.176	-63.382	40.418	-88.976	-7.673	28.551	20.938
-17.093	-42.039	-1.603	7.399	8.789	-7.630	0.934	-2.494	3.932	96.152
96.152	96.152	96.152	96.152	96.152	96.152	96.152	96.152	96.152	96.152
96.152									

NOMBRE DE REALISATIONS 28
 NOMBRE DE TRIMESTRES PRIS EN COMPTE 33
 NOMBRE DE DONNEES RECONSTITUEES OU PREVUES 36

FONCTION N° 13

DONNEES BRUTES

1027	1062	909	1003	959	1005	842	970	900	963
776	915	850	895	734	883	828	840	723	856
703	887	772	880	740	904	771	889	737	917
808	921	870	977	844	959				

DONNEES RECONSTITUEES ET PREVISION

1033	1070	901	997	931	1001	851	988	909	969
779	914	831	905	746	873	812	850	731	871
721	877	764	867	745	892	766	875	743	917
798	931	881	965	818	971				

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

-1	-1	1	1	3	0	-1	-2	-1	-1
0	0	2	-1	-2	1	2	-1	-1	-2
-3	1	1	1	-1	1	1	2	-1	0
1	-1	-1	1	3	-1				

COEFFICIENT DE LA FONCTION SUR CHAQUE TENDANCE

-774	-108	189	-150	28	38	31	-19	39	13
28	11	-8	-2	-10	-108	-108	-108	-108	-108
-108	-108	-108	-108	-108	-108	-108	-108	-108	-108
-108	-108	-108							

FONCTION N° 14

DONNEES BRUTES

1076	1030	892	1002	1038	1001	801	1030	999	999
854	1037	1042	1043	867	1046	1051	1026	870	1060
1054	1045	865	983	974	989	854	933	919	998
890	1024	1035	1046	911	1014				

DONNEES RECONSTITUEES ET PREVISION

1091	1035	858	1016	1010	985	826	1021	999	1009
849	1039	1047	1052	871	1025	1060	1018	874	1041
1026	1019	880	996	971	999	882	984	960	996
881	987	1002	999	898	1010				

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

-1	-1	4	-1	3	2	-3	1	0	-1
1	0	-1	-1	0	2	-1	1	0	2
3	2	-2	-1	0	-1	-3	-5	-4	0
1	4	3	4	1	0				

COEFFICIENT DE LA FONCTION SUR CHAQUE TENDANCE

-91	40	40	88	-16	-98	-81	52	-5	-78
61	-9	15	30	3	40	40	40	40	40
40	40	40	40	40	40	40	40	40	40
40	40	40							

FONCTION N° 15

DONNEES BRUTES

1127	1027	845	1001	1033	1025	875	1012	1009	1050
863	1018	1061	1096	961	1083	1148	1132	1021	1098
1095	1145	1044	1102	1106	1124	1033	1136	1125	1178
1088	1161	1223	1184	1149	1241				

DONNEES RECONSTITUEES ET PREVISION

1078	1036	892	994	1007	1004	873	998	1006	1044
899	1034	1077	1101	946	1064	1130	1118	987	1111
1131	1148	1019	1104	1118	1154	1040	1122	1136	1178
1070	1162	1213	1228	1108	1214				

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

4	-1	-6	1	3	2	0	1	0	1
-4	-2	-1	0	2	2	2	1	3	-1
-3	0	2	0	-1	-3	-1	1	-1	0
2	0	1	-4	4	2				

COEFFICIENT DE LA FONCTION SUR CHAQUE TENDANCE

491	97	192	-3	-41	-82	-59	70	48	36
-31	61	-1	-21	13	97	97	97	97	97
97	97	97	97	97	97	97	97	97	97
97	97	97							

NOMBRE DE REALISATIONS 28
 NOMBRE DE TRIMESTRES PRIS EN COMPTE 33
 NOMBRE DE DONNEES RECONSTITUEES OU PREVUES 36

FONCTION N° 16

DONNEES BRUTES

1047	1013	937	1001	1071	1036	970	1068	1104	1118
976	1082	1166	1164	1010	1142	1228	1215	1092	1195
1247	1258	1149	1242	1291	1292	1143	1251	1307	1303
1204	1307	1404	1420	1295	1420				

DONNEES RECONSTITUEES ET PREVISION

1049	1020	925	1005	1058	1066	963	1053	1094	1125
984	1078	1156	1155	1024	1148	1233	1213	1093	1191
1261	1247	1133	1235	1290	1282	1153	1268	1318	1298
1207	1312	1390	1382	1278	1372				

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

0	-1	1	0	1	-3	1	1	1	-1
-1	0	1	1	-1	-1	0	0	0	0
-1	1	1	1	0	1	-1	-1	-1	0
0	0	1	3	1	3				

COEFFICIENT DE LA FONCTION SUR CHAQUE TENDANCE

1074	126	234	71	-71	-18	28	2	12	-20
14	-8	29	-30	4	126	126	126	126	126
126	126	126	126	126	126	126	126	126	126
126	126	126							

FONCTION N° 17

DONNEES BRUTES

1076	1064	843	1017	1020	1017	830	1020	1028	1049
818	997	1003	998	771	948	969	925	768	928
924	935	788	936	918	942	771	906	895	928
793	963	977	1007	846	1009				

DONNEES RECONSTITUEES ET PREVISION

1084	1056	852	1008	1024	1028	839	1030	1026	1028
809	987	1001	995	781	959	982	931	769	938
911	926	777	927	906	947	775	931	904	945
788	942	971	964	817	971				

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

-1	1	-1	1	0	-1	-1	-1	0	2
1	1	0	0	-1	-1	-1	-1	0	-1
1	1	1	1	1	-1	0	-3	-1	-2
1	2	1	4	3	4				

COEFFICIENT DE LA FONCTION SUR CHAQUE TENDANCE

-405	42	-28	26	48	27	20	-16	-1	-50
-13	17	-3	6	-1	42	42	42	42	42
42	42	42	42	42	42	42	42	42	42
-44	-44	-44							

FONCTION N° 18

DONNEES BRUTES

1034	1034	852	1080	978	1035	840	1058	981	1009
802	1025	944	961	786	985	906	915	762	976
871	918	800	995	874	920	766	919	829	914
783	987	913	966	824	1050				

DONNEES RECONSTITUEES ET PREVISION

1046	1038	866	1050	984	1010	845	1075	992	1006
809	1022	941	959	777	987	929	909	770	981
860	917	785	971	853	922	776	961	835	913
791	977	907	933	819	998				

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

-1.169	-0.435	-1.591	2.803	-0.593	2.457	-0.562	-1.643	-1.102	0.341
-0.902	0.257	0.368	0.189	1.09	-0.223	-2.497	0.701	-1.09	-0.54
1.265	0.114	1.927	2.383	2.413	-0.235	-1.272	-4.589	-0.734	0.104
-0.965	0.999	0.708	3.414	0.621	4.913				

COEFFICIENT DE LA FONCTION SUR CHAQUE TENDANCE

-467.731	-144.483	12.096	27.074	28.86	-9.563	28.922	0.105	-10.302	-50.381
3.181	19.782	-50.371	-22.155	-10.329	-144.476	-144.476	-144.476	-144.476	-144.476
-144.476	-144.476	-144.476	-144.476	-144.476	-144.476	-144.476	-144.476	-144.476	-144.476
-144.476	-144.476	-144.476							

NOMBRE DE REALISATIONS 28
 NOMBRE DE TRIMESTRES PRIS EN COMPTE 33
 NOMBRE DE DONNEES RECONSTITUEES OU PREVUES 36

FONCTION N° 19

DONNEES BRUTES

1008	1010	935	1048	1032	1056	979	1084	1060	1088
979	1085	1061	1071	985	1112	1096	1085	1030	1155
1125	1141	1052	1150	1110	1128	1068	1177	1112	1125
1093	1181	1156	1139	1100	1172				

DONNEES RECONSTITUEES ET PREVISION

1005	1006	945	1043	1027	1050	975	1095	1053	1084
975	1083	1058	1074	987	1117	1113	1112	1028	1141
1110	1135	1058	1167	1116	1133	1052	1159	1104	1123
1088	1187	1165	1143	1138	1220				

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

0.279	0.394	-1.057	0.443	0.499	0.605	0.417	-1.031	0.669	0.371
0.437	0.206	0.257	-0.276	-0.239	-0.433	-1.542	-2.477	0.211	1.213
1.333	0.531	-0.563	-1.474	-0.556	-0.462	1.53	1.521	0.683	0.179
0.416	-0.501	-0.816	-0.33	-3.428	-4.082				

COEFFICIENT DE LA FONCTION SUR CHAQUE TANDENCE

530.433	-55.544	244.202	152.888	-113.252	-23.914	34.845	-9.721	9.419	7.586
-19.203	-20.662	-9.824	16.858	15.205	-55.474	-55.474	-55.474	-55.474	-55.474
-55.474	-55.474	-55.474	-55.474	-55.474	-55.474	-55.474	-55.474	-55.474	-55.474
-55.474	-55.474	-55.474							

FONCTION N° 20

DONNEES BRUTES

1059	1018	820	1102	963	981	754	1137	1058	1071
810	1155	1046	1048	784	1125	1088	1029	877	1238
1094	1148	932	1321	1178	1241	1005	1389	1170	1223
1027	1421	1261	1218	1051	1471				

DONNEES RECONSTITUEES ET PREVISION

1065	1021	810	1104	958	1004	793	1149	1039	1054
785	1139	1023	1043	789	1172	1086	1073	880	1258
1103	1146	933	1306	1159	1218	969	1361	1189	1240
1025	1419	1280	1300	1059	1455				

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

-0.54	-0.307	1.239	-0.175	0.511	-2.311	-5.12	-1.029	1.777	1.55
3.138	1.366	2.213	0.459	-0.678	-4.138	0.163	-4.255	-0.292	-1.611
-0.778	0.174	-0.127	1.109	1.622	1.814	3.558	2.03	-1.631	-1.395
0.199	0.107	-1.476	-6.706	-0.789	1.064				

COEFFICIENT DE LA FONCTION SUR CHAQUE TANDENCE

702.877	-284.445	-76.023	-188.718	141.519	-100.83	89.309	-92.529	50.754	-11.179
-9.656	-2.935	21.017	27.522	16.995	-284.473	-284.473	-284.473	-284.473	-284.473
-284.473	-284.473	-284.473	-284.473	-284.473	-284.473	-284.473	-284.473	-284.473	-284.473
-284.473	-284.473	-284.473							

FONCTION N° 21

DONNEES BRUTES

1141	1072	779	1008	996	984	754	996	995	979
737	1029	1034	1038	748	989	999	950	735	997
957	931	748	956	953	1010	798	1047	1068	1098
821	1110	1159	1206	881	1166				

DONNEES RECONSTITUEES ET PREVISION

1132	1063	782	1023	1006	976	731	998	991	989
741	1028	1045	1033	744	982	1009	944	727	978
947	941	748	971	966	1021	801	1040	1049	1093
838	1106	1161	1127	858	1168				

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

1	1	0	-2	-1	1	3	0	0	-1
-1	0	-1	0	1	1	-1	1	1	2
1	-1	0	-2	-1	-1	0	1	2	0
-2	0	0	7	3	0				

COEFFICIENT DE LA FONCTION SUR CHAQUE TENDANCE

-119	33	-143	-206	174	57	-80	26	-29	22
-19	-23	-4	9	0	33	33	33	33	33
33	33	33	33	33	33	33	33	33	33
33	33	33							

NOMBRE DE REALISATIONS 28
 NOMBRE DE TRIMESTRES PRIS EN COMPTE 33
 NOMBRE DE DONNEES RECONSTITUEES OU PREVUES 36

FONCTION N° 22

DONNEES BRUTES

1059	1034	851	1056	1124	1083	918	1140	1148	1134
950	1119	1121	1071	875	1049	1036	926	773	907
923	844	732	901	899	879	707	908	899	857
713	869	947	860	793	910				

DONNEES RECONSTITUEES ET PREVISION

1043	1034	854	1069	1133	1103	921	1140	1145	1115
933	1120	1144	1070	856	1040	1035	935	772	912
919	845	735	900	897	879	714	898	906	859
727	870	934	852	768	865				

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

2	0	0	-1	-1	-2	0	0	0	2
2	0	-2	0	2	1	0	-1	0	-1
0	0	0	0	0	0	-1	1	-1	0
-2	0	1	1	3	5				

COEFFICIENT DE LA FONCTION SUR CHAQUE TENDANCE

-427	34	-139	389	78	197	9	-3	-23	2
-11	20	16	14	39	34	34	34	34	34
34	34	34	34	34	34	34	34	34	34
34	34	34							

FONCTION N° 23

DONNEES BRUTES

1138	1051	820	992	1021	983	795	996	1063	1022
786	962	1089	993	768	934	1060	933	769	948
1039	931	781	915	993	945	768	893	968	881
733	845	963	848	702	834				

DONNEES RECONSTITUEES ET PREVISION

1151	1044	823	983	1067	998	797	1001	1059	1003
777	957	1060	999	773	953	1081	949	773	945
1033	940	770	912	982	933	759	886	958	899
737	862	958	894	764	888				

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

-1	1	0	1	-4	-2	0	-1	0	2
1	1	3	-1	-1	-2	-2	-2	-1	0
1	-1	1	0	1	1	1	1	1	-2
-1	-2	1	-5	-9	-7				

COEFFICIENT DE LA FONCTION SUR CHAQUE TENDANCE

-387	207	-181	70	-23	-60	-13	-50	-13	-21
-46	-16	-3	-8	-9	207	207	207	207	207
207	207	207	207	207	207	207	207	207	207
207	207	207							

FONCTION N° 24

DONNEES BRUTES

1201	1059	762	978	1087	1017	743	985	1103	1069
791	1011	1189	1113	782	989	1209	1021	762	981
1157	1011	762	960	1061	1021	773	950	1032	943
715	830	949	836	645	758				

DONNEES RECONSTITUEES ET PREVISION

1206	1066	761	967	1097	1019	741	1004	1117	1064
773	1008	1181	1104	779	1010	1190	1035	793	1010
1154	1018	767	936	1062	1002	754	917	1022	946
714	842	981	889	726	851				

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

0	-1	0	1	-1	0	0	-2	-1	0
2	0	1	1	0	-2	2	-1	-4	-3
0	-1	-1	3	0	2	2	3	1	0
0	-1	-3	-6	-13	-12				

COEFFICIENT DE LA FONCTION SUR CHAQUE TENDANCE

-171	336	-302	187	89	-164	-39	-73	2	35
6	15	-21	6	-8	336	336	336	336	336
336	336	336	336	336	336	336	336	336	336
336	336	336							

SORTIE LOGICIEL PREVISION SANS CORRECTION

Prévision sur 6 composantes
temporelles

-45.760	-15.881	-2.768	-7.656	-9.022	206.749	206.749	206.749	206.749	206.749
206.749	206.749	206.749	206.749	206.749	206.749	206.749	206.749	206.749	206.749
206.749	206.749	206.749							

FONCTION N° 24

DONNEES BRUTES

1201	1059	762	978	1087	1017	743	985	1103	1069
791	1011	1189	1113	782	989	1209	1021	762	981
1157	1011	762	960	1061	1021	773	950	1032	943
715	830	949	836	645	758				

DONNEES RECONSTITUEES ET PREVISION

1206	1066	761	967	1097	1019	741	1004	1117	1064
773	1008	1181	1104	779	1010	1190	1035	793	1010
1154	1018	767	936	1062	1002	754	917	1022	946
714	842	981	886	703	839				

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

-0.449	-0.646	0.099	1.123	-0.875	-0.231	0.202	-1.895	-1.312	0.473
2.289	0.320	0.641	0.797	0.393	-2.158	1.603	-1.399	-4.127	-2.983
0.270	-0.668	-0.608	2.518	-0.080	1.821	2.410	3.471	0.936	-0.273
0.160	-1.413	-3.343	-6.027	-9.020	-10.727				

COEFFICIENT DE LA FONCTION SUR CHAQUE COMPOSANTE TEMPORELLE

-170.870	335.829	-302.043	186.607	88.970	-163.676	-39.235	-72.842	2.032	35.343
6.414	15.154	-21.198	5.965	-7.796	335.738	335.738	335.738	335.738	335.738
335.738	335.738	335.738	335.738	335.738	335.738	335.738	335.738	335.738	335.738
335.738	335.738	335.738							

FONCTION N° 25

DONNEES BRUTES

1050	1041	866	1043	987	997	833	1053	995	993
821	1050	993	976	769	930	916	880	745	898
807	872	759	902	841	884	762	894	839	902
806	937	947	949	823	961				

DONNEES RECONSTITUEES ET PREVISION

1052	1044	862	1042	1008	1005	842	1042	983	993
825	1023	972	977	794	955	917	884	743	909
823	859	750	895	819	886	758	907	846	915
782	945	941	942	782	959				

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

-0.202	-0.296	0.441	0.124	-2.141	-0.807	-1.034	1.026	1.248	-0.033
-0.459	2.616	2.067	-0.135	-3.255	-2.703	-0.115	-0.486	0.225	-1.261
-2.014	1.535	1.174	0.735	2.573	-0.186	0.501	-1.425	-0.868	-1.438
2.958	-0.894	0.590	0.708	5.027	0.230				

COEFFICIENT DE LA FONCTION SUR CHAQUE COMPOSANTE TEMPORELLE

-563.510	-71.848	10.867	17.862	47.346	96.791	-43.850	-27.709	21.401	-32.388
-18.418	2.082	32.319	-24.456	-30.969	-71.840	-71.840	-71.840	-71.840	-71.840
-71.840	-71.840	-71.840	-71.840	-71.840	-71.840	-71.840	-71.840	-71.840	-71.840
-71.840	-71.840	-71.840							

FONCTION N° 26

DONNEES BRUTES

1092	1056	886	965	1028	1048	897	1020	1055	1077
910	1034	1093	1106	932	1055	1155	1129	994	1097
1133	1139	1010	1072	1139	1172	1050	1120	1199	1199
1098	1177	1288	1263	1168	1228				

DONNEES RECONSTITUEES ET PREVISION

1079	1046	895	981	1037	1039	899	997	1043	1078
914	1022	1107	1115	946	1055	1145	1123	981	1080
1138	1143	1013	1094	1152	1178	1039	1131	1189	1210
1081	1177	1277	1268	1112	1235				

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

1.231	0.993	-1.017	-1.631	-0.906	0.887	-0.169	2.291	1.107	-0.072
-0.423	1.133	-1.256	-0.838	-1.469	-0.004	0.830	0.504	1.293	1.578
-0.408	-0.371	-0.288	-2.032	-1.175	-0.529	1.087	-0.982	0.848	-0.938
1.552	-0.036	0.825	-0.359	4.834	-0.559				

SORTIE LOGICIEL PREVISION SANS CORRECTION

Prévision sur 6 composantes
temporelles

COEFFICIENT DE LA FONCTION SUR CHAQUE COMPOSANTE TEMPORELLE

563.542	167.080	180.549	-5.536	-14.471	-1.701	-21.813	31.526	2.232	3.501
-26.086	-10.936	33.676	4.822	-15.352	167.130	167.130	167.130	167.130	167.130
167.130	167.130	167.130	167.130	167.130	167.130	167.130	167.130	167.130	167.130
167.130	167.130	167.130							

FONCTION N° 27

DONNEES BRUTES

1058	1042	869	1031	1041	1060	886	1041	1030	1107
911	1097	1121	1117	950	1153	1132	1183	1001	1158
1153	1145	1022	1207	1186	1254	1072	1301	1270	1315
1179	1400	1428	1450	1275	1517				

DONNEES RECONSTITUEES ET PREVISION

1048	1037	877	1038	1018	1044	890	1056	1046	1096
920	1110	1119	1139	945	1132	1140	1144	986	1160
1137	1170	1032	1209	1200	1258	1085	1292	1287	1332
1164	1382	1421	1410	1200	1454				

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

0.935	0.483	-0.889	-0.666	2.177	1.545	-0.476	-1.466	-1.544	1.004
-0.969	-1.205	0.194	-2.005	0.487	1.855	-0.664	3.279	1.548	-0.183
1.412	-2.200	-0.963	-0.134	-1.150	-0.300	-1.259	0.672	-1.329	-1.313
1.252	1.287	0.524	2.760	5.905	4.159				

COEFFICIENT DE LA FONCTION SUR CHAQUE COMPOSANTE TEMPORELLE

911.913	-8.570	166.439	-100.024	81.322	65.123	-12.223	47.756	9.811	-1.296
-2.054	23.011	-26.892	47.610	-15.604	-8.526	-8.526	-8.526	-8.526	-8.526
-8.526	-8.526	-8.526	-8.526	-8.526	-8.526	-8.526	-8.526	-8.526	-8.526
-8.526	-8.526	-8.526							

FONCTION N° 28

DONNEES BRUTES

1111	1061	837	992	1018	985	789	960	998	1029
810	970	1075	1089	842	1006	1108	1060	871	1041
1106	1118	917	1059	1155	1196	974	1123	1222	1287
1003	1186	1341	1340	1083	1254				

DONNEES RECONSTITUEES ET PREVISION

1137	1069	831	964	1013	1000	793	945	1009	1032
800	979	1083	1088	838	1002	1115	1072	876	1042
1102	1105	919	1059	1140	1182	976	1135	1219	1253
1026	1212	1338	1316	1047	1293				

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

-2.306	-0.712	0.735	2.811	0.498	-1.572	-0.561	1.531	-1.141	-0.274
1.261	-0.891	-0.717	0.114	0.471	0.399	-0.650	-1.161	-0.585	-0.122
0.334	1.164	-0.269	-0.027	1.305	1.176	-0.225	-1.074	0.244	2.661
-2.265	-2.170	0.229	1.782	3.337	-3.104				

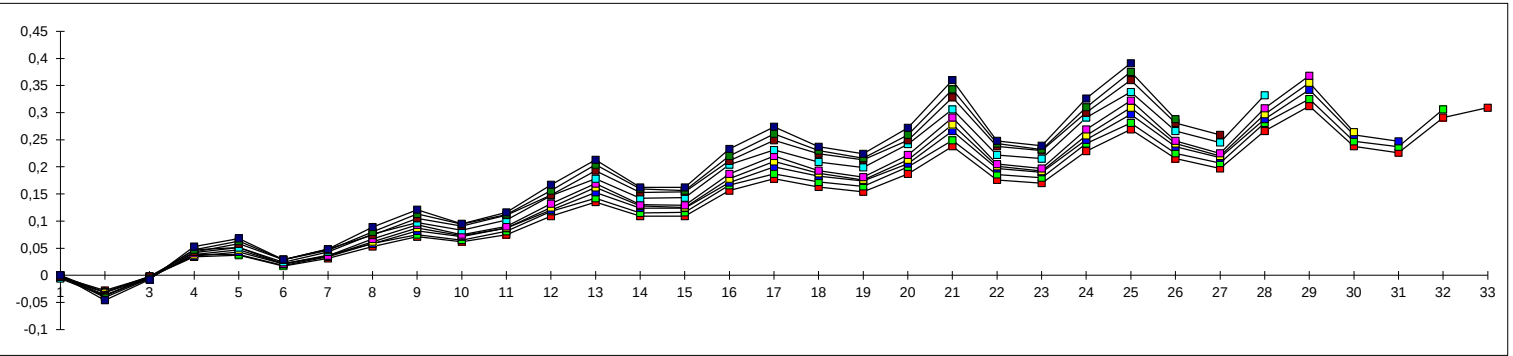
COEFFICIENT DE LA FONCTION SUR CHAQUE COMPOSANTE TEMPORELLE

456.028	202.041	41.569	-245.019	91.367	7.185	-43.994	-15.973	-9.418	-10.864
27.742	-25.035	-17.222	-41.561	38.330	202.051	202.051	202.051	202.051	202.051
202.051	202.051	202.051	202.051	202.051	202.051	202.051	202.051	202.051	202.051
202.051	202.051	202.051							

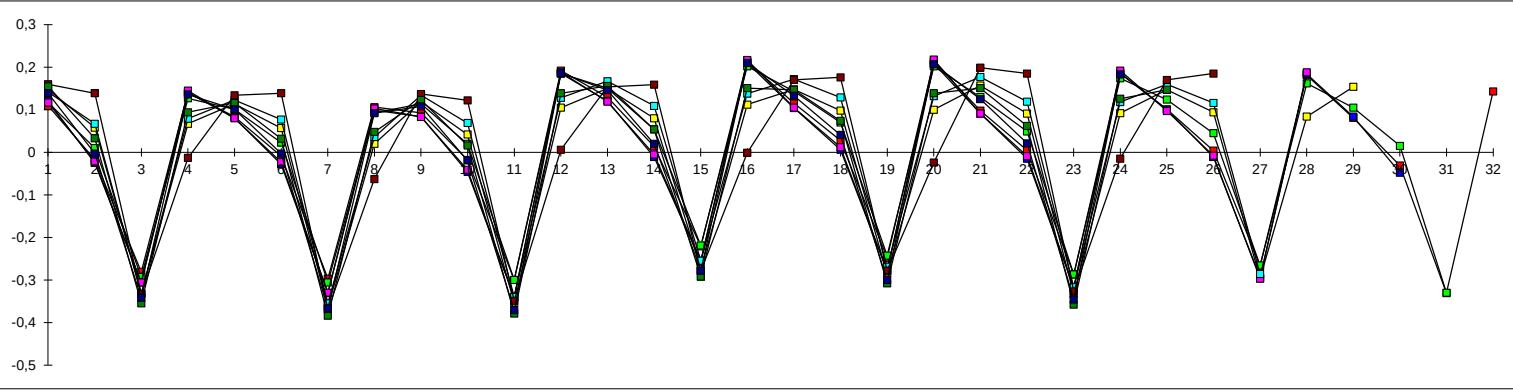
An I-3-a)

VERIFICATION DU THEOREME PRINCIPAL PAR LE CALCUL DE 25 A 33 TRIMESTRES

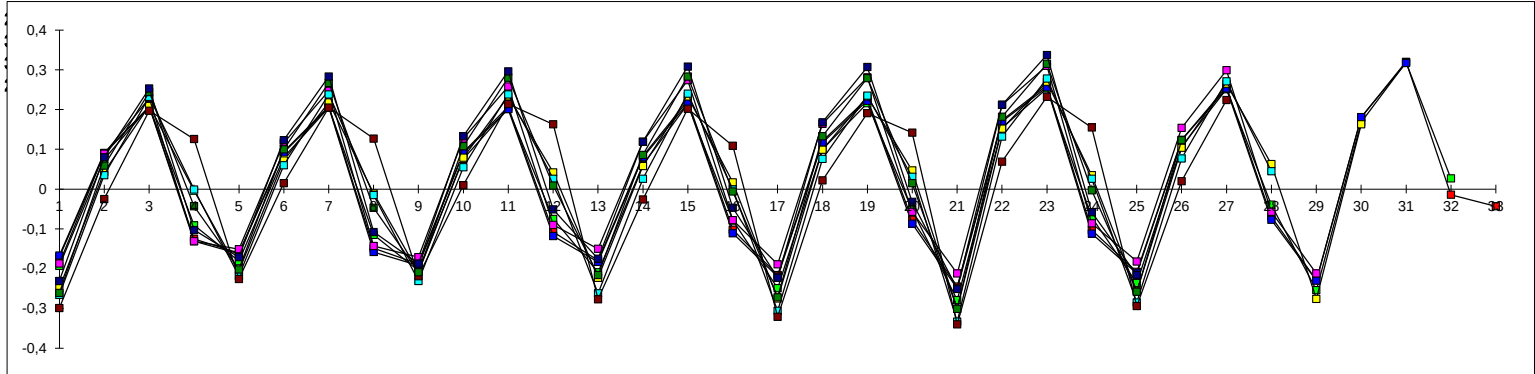
COMPOSANTE TEMPORELLE N°1																																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	
33	-0	-0,03	-0	0,034	0,037	0,017	0,031	0,053	0,071	0,062	0,075	0,109	0,135	0,109	0,109	0,156	0,178	0,163	0,154	0,187	0,238	0,176	0,17	0,229	0,269	0,215	0,197	0,266	0,312	0,238	0,226	0,291	0,309	
32	-0,01	-0,03	-0	0,038	0,038	0,018	0,034	0,059	0,075	0,065	0,081	0,118	0,142	0,115	0,116	0,166	0,187	0,172	0,164	0,2	0,249	0,186	0,18	0,243	0,281	0,225	0,207	0,281	0,325	0,247	0,237	0,306		
31	-0	-0,03	-0	0,035	0,043	0,021	0,037	0,058	0,082	0,071	0,087	0,12	0,153	0,124	0,124	0,171	0,2	0,183	0,174	0,206	0,266	0,197	0,19	0,25	0,297	0,237	0,217	0,288	0,342	0,259	0,247			
30	0	-0,03	-0,01	0,038	0,047	0,021	0,035	0,062	0,087	0,073	0,087	0,125	0,162	0,128	0,125	0,178	0,21	0,188	0,176	0,213	0,278	0,201	0,192	0,258	0,309	0,243	0,22	0,297	0,355	0,264				
29	0	-0,04	-0,01	0,042	0,051	0,021	0,036	0,067	0,093	0,075	0,09	0,132	0,169	0,13	0,129	0,187	0,22	0,193	0,181	0,222	0,291	0,205	0,197	0,269	0,322	0,248	0,225	0,308	0,368					
28	-0,01	-0,04	-0	0,047	0,051	0,024	0,043	0,076	0,097	0,083	0,102	0,147	0,178	0,142	0,143	0,204	0,231	0,209	0,199	0,242	0,306	0,222	0,215	0,291	0,338	0,266	0,245	0,332						
27	-0	-0,04	-0	0,043	0,058	0,029	0,048	0,074	0,105	0,091	0,111	0,148	0,193	0,153	0,154	0,211	0,249	0,224	0,213	0,249	0,328	0,238	0,23	0,3	0,36	0,281	0,259							
26	0	-0,04	-0,01	0,047	0,063	0,029	0,046	0,08	0,113	0,094	0,112	0,156	0,204	0,159	0,156	0,22	0,261	0,23	0,216	0,259	0,343	0,243	0,232	0,31	0,375	0,288								
25	0	-0,05	-0,01	0,053	0,068	0,029	0,048	0,089	0,121	0,095	0,116	0,167	0,213	0,162	0,162	0,233	0,274	0,237	0,224	0,272	0,36	0,248	0,239	0,326	0,391									



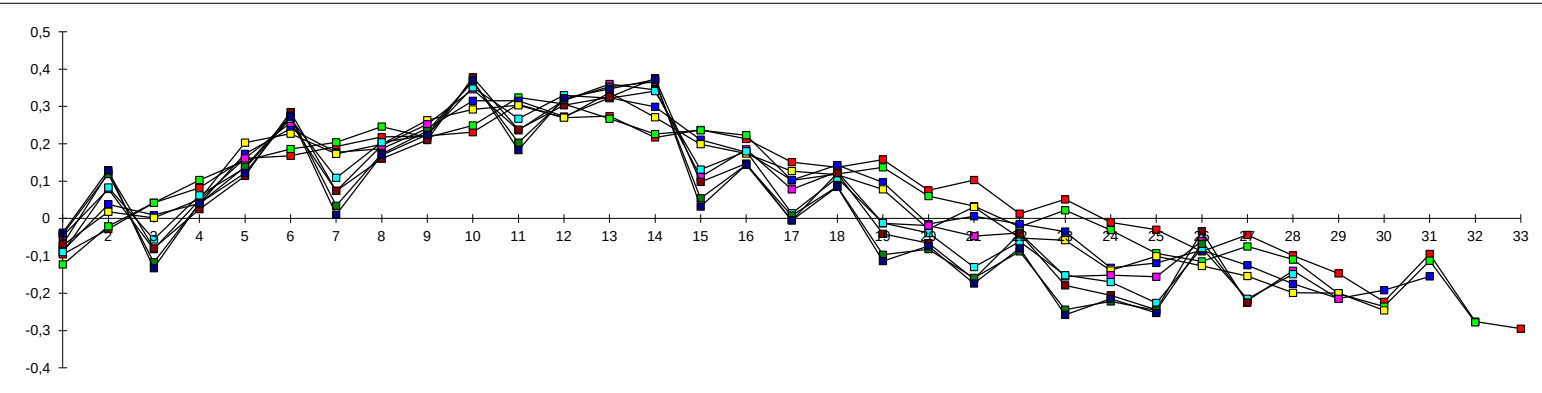
COMPOSANTE TEMPORELLE N°2 OU N° 3																																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
33	0,19	0,109	-0,02	-0,28	0,14	0,084	-0,01	-0,3	0,106	0,093	-0,02	-0,3	0,192	0,13	0,014	-0,22	0,213	0,115	0,028	-0,25	0,214	0,098	0,005	-0,29	0,186	0,101	0,004	-0,27	0,181	0,081	-0,03	-0,33	0,143
32	0,166	0,117	0,01	-0,29	0,127	0,101	0,023	-0,31	0,093	0,114	0,02	-0,3	0,185	0,151	0,054	-0,22	0,202	0,143	0,07	-0,24	0,202	0,13	0,049	-0,29	0,174	0,124	0,045	-0,27	0,162	0,105	0,015	-0,33	
31	0,211	0,121	-0,03	-0,31	0,14	0,081	-0,03	-0,33	0,099	0,084	-0,05	-0,34	0,185	0,121	-0,01	-0,26	0,21	0,104	0,006	-0,28	0,209	0,091	-0,02	-0,32	0,185	0,1	-0,01	-0,29	0,184	0,084	-0,05		
30	0,128	0,148	0,058	-0,33	0,067	0,114	0,057	-0,36	0,02	0,119	0,042	-0,35	0,105	0,15	0,08	-0,27	0,112	0,148	0,098	-0,29	0,1	0,157	0,091	-0,33	0,092	0,147	0,094	-0,3	0,084	0,154			
29	0,211	0,117	-0,02	-0,31	0,145	0,08	-0,02	-0,33	0,103	0,083	-0,04	-0,34	0,188	0,119	-0,01	-0,26	0,217	0,105	0,012	-0,28	0,218	0,091	-0,01	-0,32	0,192	0,097	-0,01	-0,3	0,188				
28	0,122	0,142	0,067	-0,33	0,079	0,123	0,077	-0,36	0,036	0,134	0,069	-0,34	0,128	0,167	0,109	-0,26	0,138	0,172	0,129	-0,27	0,132	0,177	0,119	-0,32	0,118	0,16	0,116	-0,29					
27	0,034	0,16	0,139	-0,33	-0,01	0,134	0,139	-0,37	-0,06	0,137	0,122	-0,35	0,006	0,154	0,159	-0,28	-0	0,17	0,176	-0,28	-0,02	0,199	0,185	-0,33	-0,02	0,17	0,185						
26	0,166	0,155	0,033	-0,36	0,094	0,116	0,032	-0,38	0,048	0,123	0,016	-0,38	0,139	0,156	0,054	-0,29	0,151	0,147	0,074	-0,31	0,139	0,151	0,062	-0,36	0,126	0,147							
25	0,207	0,139	-0	-0,34	0,136	0,101	-0	-0,37	0,092	0,108	-0,02	-0,37	0,186	0,146	0,02	-0,28	0,21	0,133	0,041	-0,3	0,207	0,125	0,021	-0,35	0,183								



COMPOSANTE TEMPORELLE N° 2 OU 3																																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
33	-0,17	0,09	0,21	-0,13	-0,16	0,09	0,21	-0,15	-0,19	0,09	0,21	-0,11	-0,18	0,08	0,22	-0,1	-0,22	0,12	0,23	-0,08	-0,25	0,17	0,26	-0,1	-0,21	0,12	0,26	-0,07	-0,23	0,18	0,32	-0,01	-0,04
32	-0,19	0,08	0,21	-0,09	-0,18	0,08	0,21	-0,12	-0,21	0,08	0,2	-0,08	-0,21	0,06	0,21	-0,08	-0,25	0,1	0,22	-0,05	-0,28	0,15	0,25	-0,07	-0,24	0,11	0,25	-0,04	-0,25	0,17	0,32	0,03	
31	-0,17	0,09	0,21	-0,13	-0,16	0,09	0,21	-0,16	-0,19	0,09	0,2	-0,12	-0,18	0,08	0,22	-0,11	-0,22	0,12	0,22	-0,09	-0,25	0,17	0,26	-0,11	-0,21	0,12	0,25	-0,08	-0,23	0,18	0,32		
30	-0,25	0,04	0,21	-0	-0,21	0,07	0,22	-0,01	-0,21	0,08	0,23	0,04	-0,22	0,06	0,23	0,02	-0,28	0,1	0,23	0,05	-0,3	0,15	0,27	0,04	-0,26	0,1	0,27	0,06	-0,28	0,16			
29	-0,19	0,09	0,23	-0,13	-0,15	0,12	0,25	-0,14	-0,17	0,13	0,26	-0,09	-0,15	0,12	0,27	-0,08	-0,19	0,16	0,28	-0,06	-0,21	0,21	0,31	-0,09	-0,18	0,15	0,3	-0,06	-0,21				
28	-0,27	0,04	0,23	-0	-0,22	0,06	0,24	-0,02	-0,23	0,06	0,24	0,03	-0,26	0,03	0,24	-0	-0,31	0,08	0,24	0,03	-0,33	0,13	0,28	0,03	-0,29	0,08	0,27	0,05					
27	-0,3	-0,03	0,2	0,13	-0,23	0,02	0,21	0,13	-0,22	0,01	0,21	0,16	-0,28	-0,03	0,2	0,11	-0,32	0,02	0,19	0,14	-0,34	0,07	0,23	0,16	-0,29	0,02	0,22						
26	-0,26	0,06	0,24	-0,04	-0,2	0,1	0,27	-0,05	-0,21	0,11	0,28	0,01	-0,22	0,09	0,28	-0,01	-0,27	0,13	0,28	0,02	-0,3	0,18	0,32	-0	-0,26	0,12							
25	-0,23	0,08	0,25	-0,1	-0,17	0,12	0,28	-0,11	-0,19	0,13	0,3	-0,05	-0,18	0,12	0,31	-0,05	-0,22	0,17	0,31	-0,03	-0,25	0,21	0,34	-0,06	-0,22								



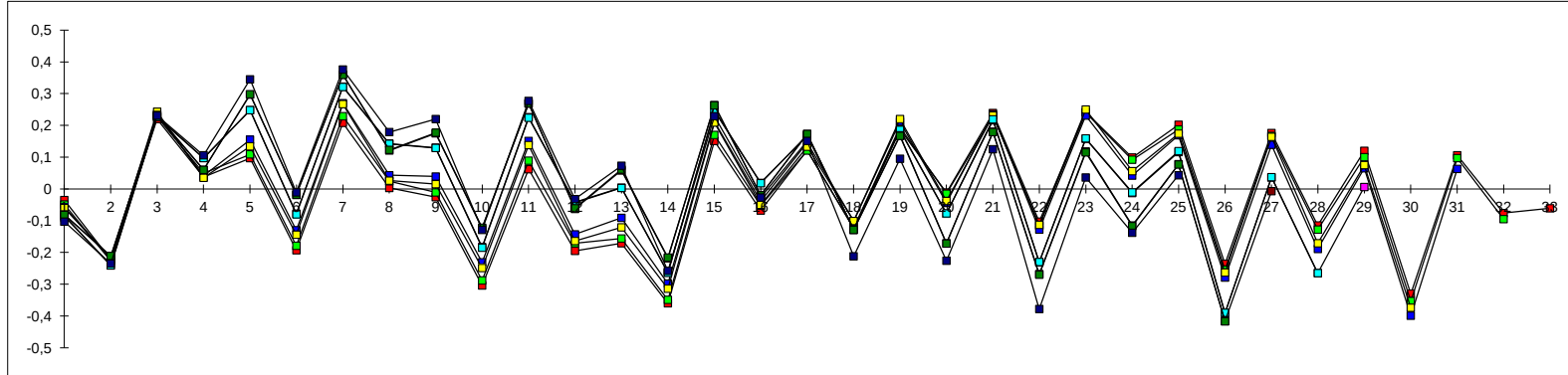
COMPOSANTE TEMPORELLE N°4																																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
33	-0,1	-0,03	0,04	0,08	0,16	0,17	0,19	0,22	0,22	0,23	0,31	0,27	0,27	0,22	0,24	0,21	0,15	0,14	0,16	0,08	0,1	0,01	0,05	-0,01	-0,03	-0,09	-0,04	-0,1	-0,15	-0,22	-0,1	-0,28	-0,3
32	-0,12	-0,02	0,04	0,1	0,16	0,19	0,2	0,25	0,22	0,25	0,32	0,31	0,27	0,23	0,24	0,22	0,1	0,12	0,14	0,06	0,03	-0,02	0,02	-0,03	-0,09	-0,12	-0,08	-0,11	-0,2	-0,24	-0,11	-0,28	
31	-0,09	0,04	0,01	0,04	0,17	0,24	0,18	0,19	0,24	0,32	0,32	0,27	0,32	0,3	0,21	0,18	0,1	0,14	0,1	-0,02	0,01	-0,02	-0,04	-0,13	-0,12	-0,09	-0,13	-0,18	-0,21	-0,19	-0,16		
30	-0,07	0,02	0	0,05	0,2	0,23	0,17	0,2	0,26	0,29	0,3	0,27	0,34	0,27	0,2	0,17	0,13	0,12	0,08	-0,03	0,03	-0,05	-0,06	-0,14	-0,1	-0,13	-0,15	-0,2	-0,2	-0,25			
29	-0,05	0,08	-0,08	0,05	0,16	0,26	0,07	0,2	0,25	0,35	0,24	0,32	0,36	0,34	0,11	0,19	0,08	0,13	-0,01	-0,02	-0,05	-0,04	-0,16	-0,15	-0,16	-0,06	-0,22	-0,14	-0,22				
28	-0,09	0,08	-0,06	0,06	0,14	0,27	0,11	0,2	0,22	0,35	0,27	0,33	0,32	0,34	0,13	0,18	0,01	0,11	-0,01	-0,04	-0,13	-0,06	-0,15	-0,17	-0,23	-0,08	-0,22	-0,15					
27	-0,07	0,12	-0,08	0,03	0,11	0,29	0,07	0,16	0,21	0,38	0,24	0,3	0,33	0,38	0,1	0,15	-0,01	0,12	-0,04	-0,07	-0,16	-0,04	-0,18	-0,21	-0,24	-0,03	-0,23						
26	-0,04	0,12	-0,12	0,04	0,14	0,28	0,03	0,17	0,23	0,37	0,2	0,32	0,35	0,37	0,05	0,14	0,01	0,09	-0,1	-0,08	-0,16	-0,09	-0,24	-0,22	-0,25	-0,07							
25	-0,04	0,13	-0,13	0,04	0,12	0,27	0,01	0,17	0,22	0,37	0,18	0,32	0,35	0,37	0,03	0,14	-0,01	0,09	-0,11	-0,07	-0,17	-0,08	-0,26	-0,22	-0,25								



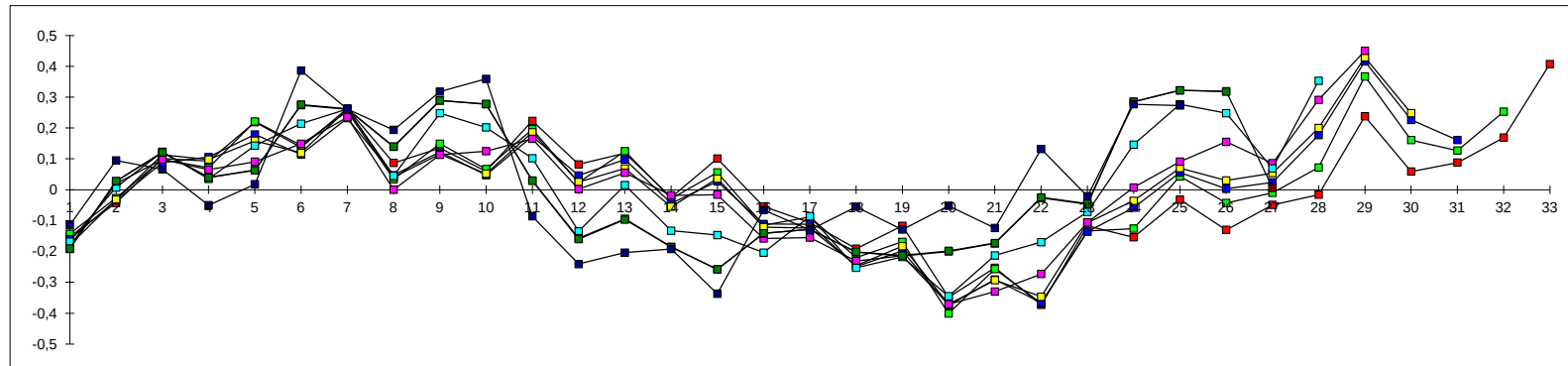
Ann I-3-c)

VERIFICATION DU THEOREME PRINCIPAL PAR LE CALCUL DE 25 A 33 TRIMESTRESS

COPMPOSANTE TEMPORELLE N°5																																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
33	-0,04	-0,22	0,22	0,04	0,1	-0,19	0,21	0	-0,03	-0,3	0,06	-0,2	-0,17	-0,36	0,15	-0,07	0,12	-0,13	0,19	-0,01	0,24	-0,1	0,24	0,1	0,2	-0,24	0,18	-0,12	0,12	-0,33	0,11	-0,08	-0,06
32	-0,05	-0,23	0,23	0,05	0,11	-0,18	0,23	0,02	-0,01	-0,29	0,09	-0,17	-0,16	-0,35	0,17	-0,05	0,12	-0,12	0,2	-0,02	0,23	-0,12	0,24	0,09	0,19	-0,25	0,17	-0,13	0,1	-0,35	0,1	-0,1	
31	-0,06	-0,22	0,24	0,04	0,16	-0,13	0,27	0,04	0,04	-0,23	0,15	-0,14	-0,09	-0,3	0,21	-0,04	0,15	-0,1	0,21	-0,04	0,24	-0,13	0,23	0,04	0,17	-0,28	0,14	-0,19	0,07	-0,4	0,06		
30	-0,06	-0,22	0,24	0,04	0,13	-0,14	0,27	0,03	0,02	-0,25	0,14	-0,16	-0,12	-0,31	0,21	-0,05	0,13	-0,1	0,22	-0,04	0,23	-0,11	0,25	0,06	0,17	-0,26	0,16	-0,17	0,08	-0,37			
29	-0,09	-0,24	0,23	0,1	0,25	-0,08	0,32	0,14	0,13	-0,18	0,23	-0,04	0	-0,26	0,25	0,02	0,17	-0,13	0,19	-0,08	0,22	-0,23	0,16	-0,01	0,12	-0,39	0,04	-0,27	0,01				
28	-0,08	-0,24	0,23	0,1	0,25	-0,08	0,32	0,14	0,13	-0,19	0,22	-0,04	0	-0,26	0,25	0,02	0,17	-0,13	0,19	-0,08	0,22	-0,23	0,16	-0,01	0,12	-0,39	0,04	-0,27					
27	-0,08	-0,21	0,23	0,06	0,3	-0,02	0,36	0,12	0,18	-0,12	0,27	-0,06	0,06	-0,22	0,26	-0,02	0,17	-0,13	0,17	-0,17	0,18	-0,27	0,12	-0,12	0,08	-0,42	-0,01						
26	-0,08	-0,21	0,23	0,06	0,3	-0,02	0,36	0,12	0,18	-0,12	0,27	-0,06	0,06	-0,22	0,26	-0,02	0,17	-0,13	0,17	-0,17	0,18	-0,27	0,12	-0,12	0,08	-0,42							
25	-0,1	-0,24	0,23	0,11	0,35	-0,01	0,38	0,18	0,22	-0,13	0,28	-0,03	0,07	-0,26	0,23	-0,03	0,15	-0,21	0,1	-0,23	0,13	-0,38	0,04	-0,14	0,04								

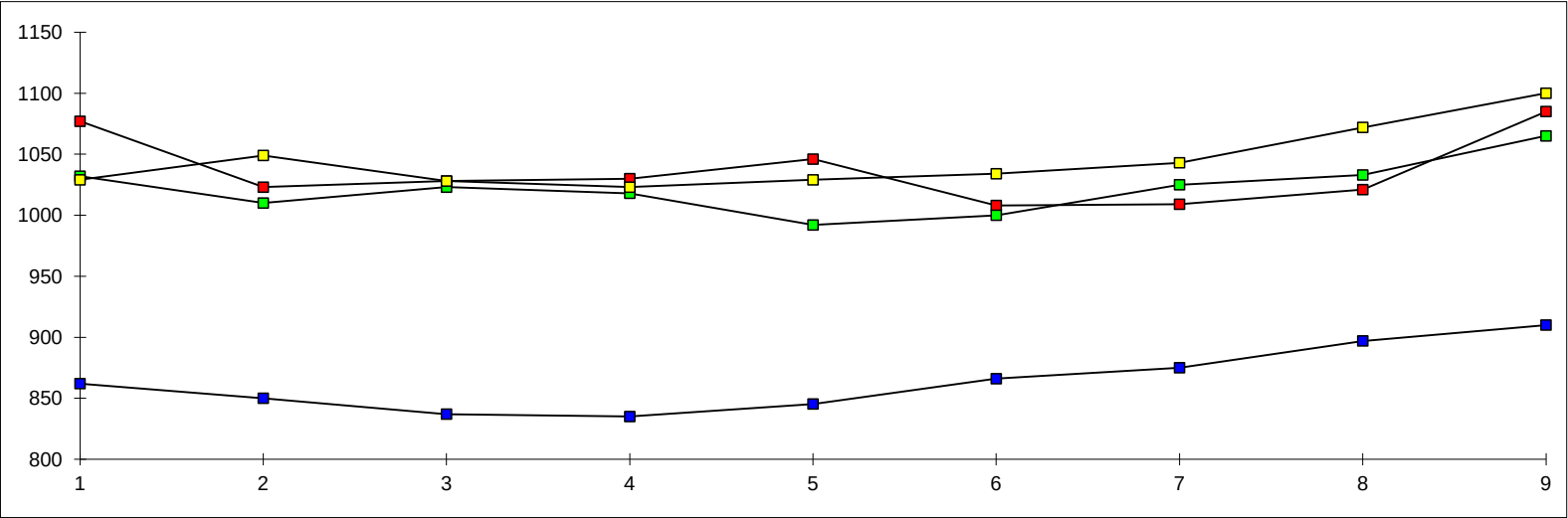


COMPOSANTE TEMPORELLE N° 6																																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
33	-0,15	-0,04	0,1	0,09	0,22	0,14	0,26	0,09	0,14	0,06	0,22	0,08	0,12	-0,03	0,1	-0,05	-0,11	-0,19	-0,12	-0,35	-0,25	-0,37	-0,12	-0,15	-0,03	-0,13	-0,05	-0,02	0,24	0,06	0,09	0,17	0,41
32	-0,14	-0,03	0,1	0,07	0,22	0,14	0,25	0,03	0,15	0,07	0,2	0,02	0,13	-0,03	0,06	-0,12	-0,09	-0,22	-0,17	-0,4	-0,26	-0,37	-0,13	-0,13	0,04	-0,04	-0,01	0,07	0,37	0,16	0,13	0,25	
31	-0,16	-0,03	0,09	0,11	0,18	0,11	0,23	0,05	0,12	0,05	0,17	0,05	0,1	-0,04	0,03	-0,11	-0,11	-0,25	-0,2	-0,37	-0,29	-0,37	-0,14	-0,06	0,06	0	0,02	0,18	0,42	0,23	0,16		
30	-0,18	-0,03	0,11	0,1	0,16	0,12	0,26	0,04	0,12	0,05	0,19	0,03	0,07	-0,06	0,04	-0,12	-0,12	-0,25	-0,18	-0,37	-0,29	-0,35	-0,11	-0,04	0,07	0,03	0,05	0,2	0,43	0,25			
29	-0,18	0,02	0,1	0,07	0,09	0,15	0,24	0	0,11	0,13	0,17	0	0,06	-0,02	-0,02	-0,16	-0,16	-0,23	-0,21	-0,37	-0,33	-0,27	-0,11	0,01	0,09	0,16	0,09	0,29	0,45				
28	-0,17	0,01	0,12	0,04	0,14	0,21	0,26	0,05	0,25	0,2	0,1	-0,14	0,02	-0,13	-0,15	-0,2	-0,09	-0,25	-0,22	-0,35	-0,21	-0,17	-0,07	0,15	0,28	0,25	0,07	0,35					
27	-0,19	0,03	0,12	0,04	0,06	0,27	0,26	0,14	0,29	0,28	0,03	-0,16	-0,09	-0,19	-0,26	-0,14	-0,13	-0,2	-0,22	-0,2	-0,17	-0,03	-0,05	0,29	0,32	0,32	0,01						
26	-0,19	0,03	0,12	0,04	0,06	0,28	0,26	0,14	0,29	0,28	0,03	-0,16	-0,1	-0,19	-0,26	-0,14	-0,13	-0,2	-0,21	-0,2	-0,17	-0,03	-0,05	0,29	0,32	0,32							
25	-0,12	0,1	0,07	-0,05	0,02	0,39	0,26	0,19	0,32	0,36	-0,09	-0,24	-0,2	-0,19	-0,34	-0,07	-0,13	-0,06	-0,13	-0,05	-0,12	0,13	-0,02	0,28	0,27								



MOYENNE ET PREVISION PAR TRIMESTRE

ANNEES	1980	1981	1982	1983	1984	1985	1986	1987	1988
TRIMESTRE N° 1	1077	1023	1028	1030	1046	1008	1009	1021	1085
TRIMESTRE N° 2	1032	1010	1023	1018	992	1000	1025	1033	1065
TRIMESTRE N° 3	862	850	837	835	845,3	866	875	897	910
TRIMESTRE N° 4	1029	1049	1028	1023	1029	1034	1043	1072	1100

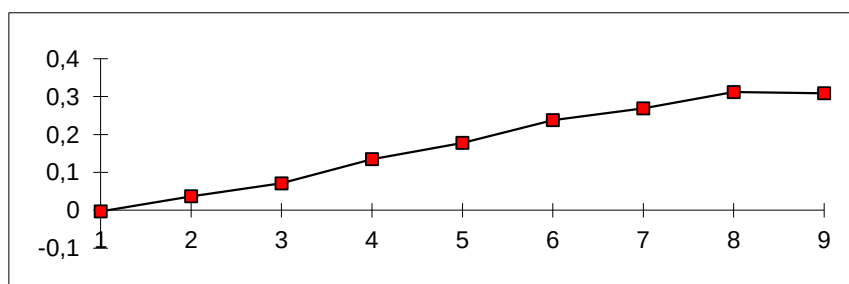


REMARQUE : la prévision de la moyenne est délicate car elle intervient sur l'ensemble des résultats. Plus le nombre d'individus est important plus la prévision sera facile. Les méthodes de prévision plus précises seraient souhaitables

COMPOSANTE TEMPORELLE N°1 PREVISION PAR TRIMESTRE

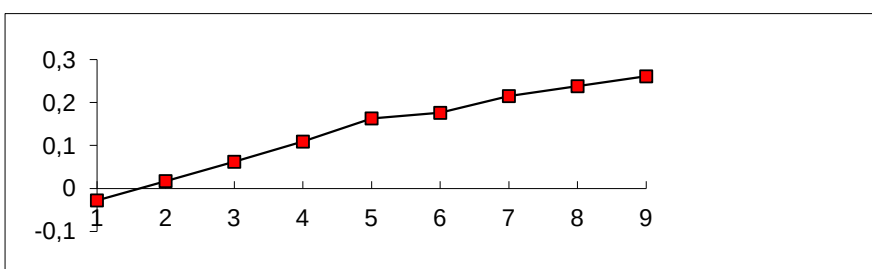
TRIMESTRE N°1

1980	1981	1982	1983	1984	1985	1986	1987	1988
-0,003	0,037	0,071	0,135	0,178	0,238	0,269	0,312	0,309



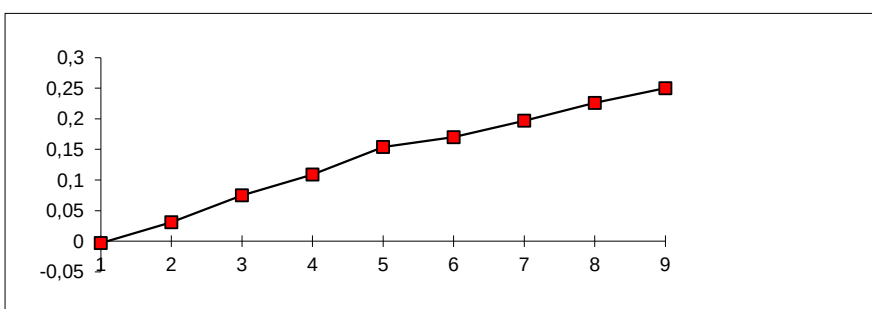
TRIMESTRE N°2

1980	1981	1982	1983	1984	1985	1986	1987	1988
-0,028	0,017	0,062	0,109	0,163	0,176	0,215	0,238	0,261



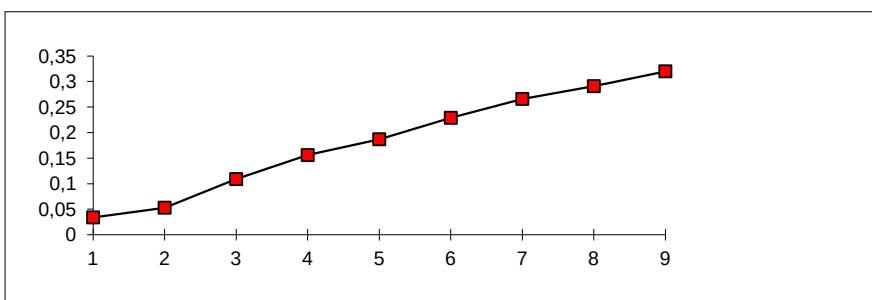
TRIMESTRE N°3

1980	1981	1982	1983	1984	1985	1986	1987	1988
-0,003	0,031	0,075	0,109	0,154	0,17	0,197	0,226	0,25



TRIMESTRE N°4

1980	1981	1982	1983	1984	1985	1986	1987	1988
0,034	0,053	0,109	0,156	0,187	0,229	0,266	0,291	0,32

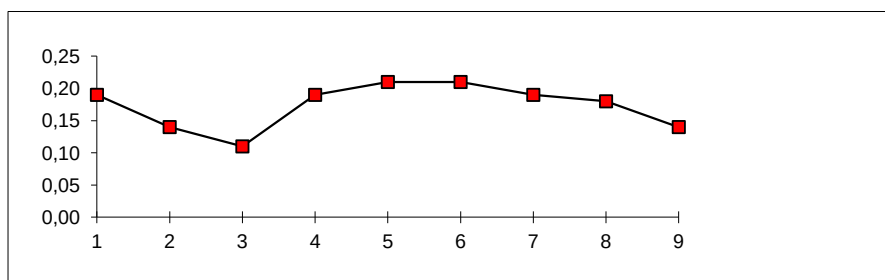


COMPOSANTE TEMPORELLE N°2

PREVISION PAR TRIMESTRE

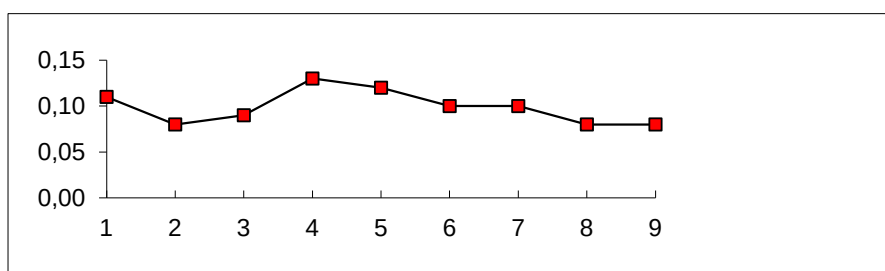
TRIMESTRE N°1

1980	1981	1982	1983	1984	1985	1986	1987	1988
0,19	0,14	0,11	0,19	0,21	0,21	0,19	0,18	0,14



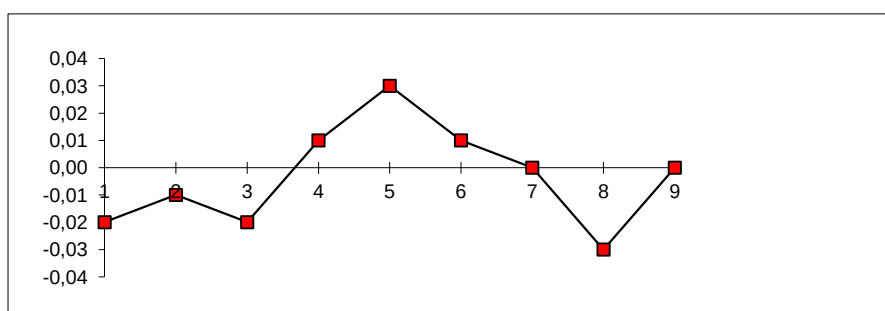
TRIMESTRE N°2

1980	1981	1982	1983	1984	1985	1986	1987	1988
0,11	0,08	0,09	0,13	0,12	0,10	0,10	0,08	0,08



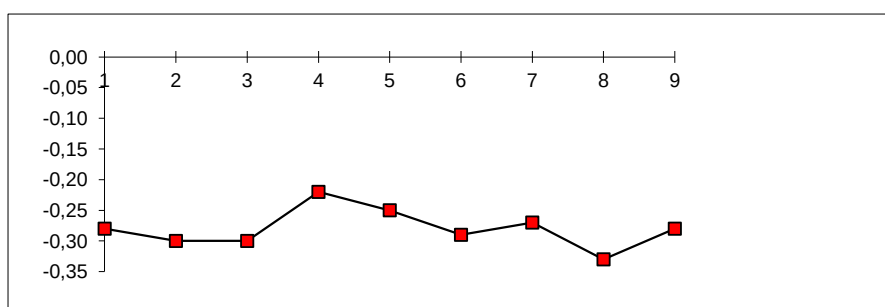
TRIMESTRE N°3

1980	1981	1982	1983	1984	1985	1986	1987	1988
-0,02	-0,01	-0,02	0,01	0,03	0,01	0,00	-0,03	0



TRIMESTRE N°4

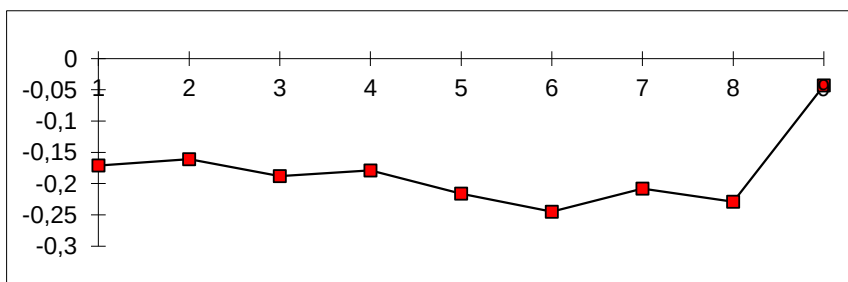
1980	1981	1982	1983	1984	1985	1986	1987	1988
-0,28	-0,30	-0,30	-0,22	-0,25	-0,29	-0,27	-0,33	-0,28



COMPOSANTE TEMPORELLE N°3 PREVISION PAR TRIMESTRE

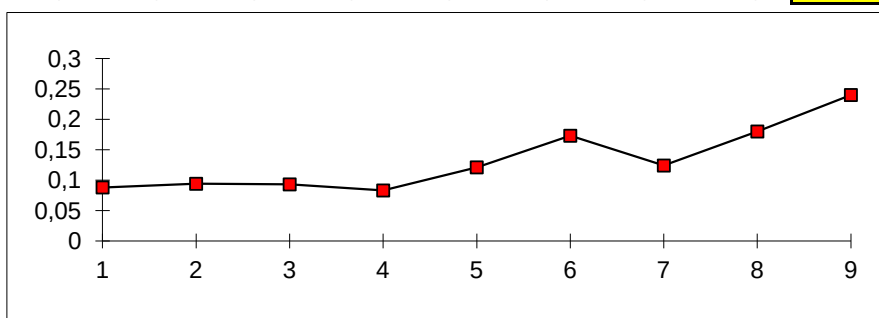
TRIMESTRE N°1

1980	1981	1982	1983	1984	1985	1986	1987	1988
-0,171	-0,161	-0,188	-0,179	-0,216	-0,245	-0,208	-0,229	-0,043



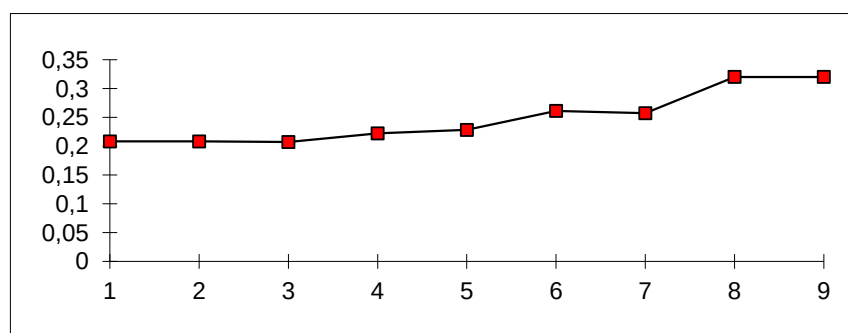
TRIMESTRE N°2

1980	1981	1982	1983	1984	1985	1986	1987	1988
0,088	0,094	0,093	0,083	0,121	0,173	0,124	0,18	0,24



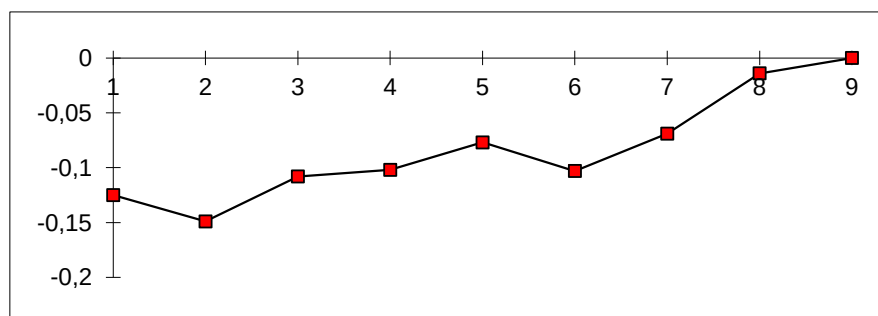
TRIMESTRE N°3

1980	1981	1982	1983	1984	1985	1986	1987	1988
0,208	0,208	0,207	0,222	0,228	0,261	0,257	0,32	0,32



TRIMESTRE N°4

1980	1981	1982	1983	1984	1985	1986	1987	1988
-0,125	-0,149	-0,108	-0,102	-0,077	-0,103	-0,069	-0,014	0

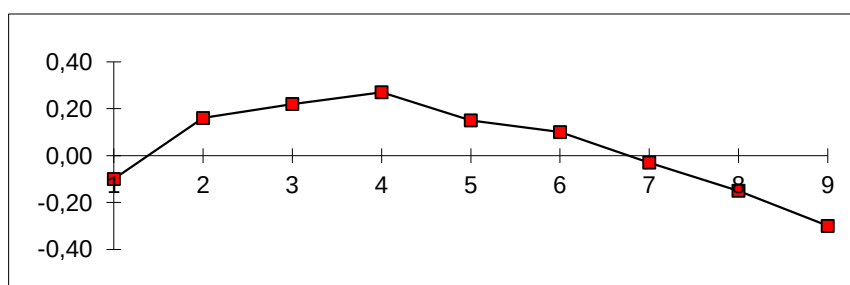


COMPOSANTES TEMPORELLE N°4

PREVISION PAR TRIMESTRE

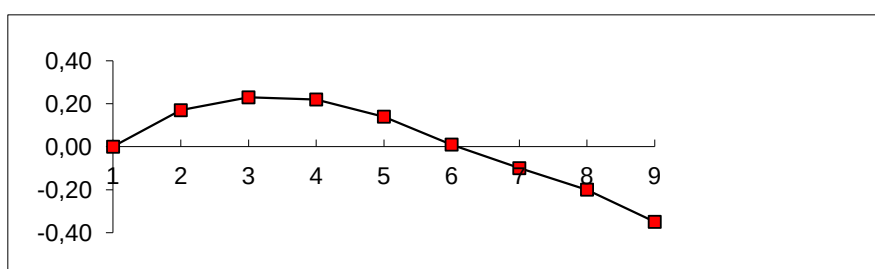
TRIMESTRE N°1

1980	1981	1982	1983	1984	1985	1986	1987	1988
-0,10	0,16	0,22	0,27	0,15	0,10	-0,03	-0,15	-0,30



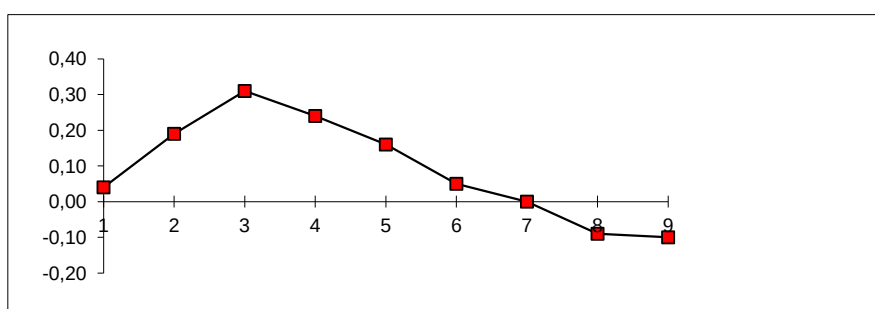
TRIMESTRE N°2

1980	1981	1982	1983	1984	1985	1986	1987	1988
0,00	0,17	0,23	0,22	0,14	0,01	-0,10	-0,20	-0,35



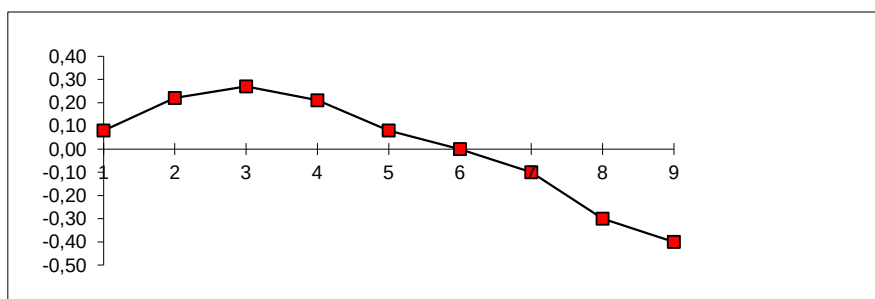
TRIMESTRE N°3

1980	1981	1982	1983	1984	1985	1986	1987	1988
0,04	0,19	0,31	0,24	0,16	0,05	0,00	-0,09	-0,10



TRIMESTRE N°4

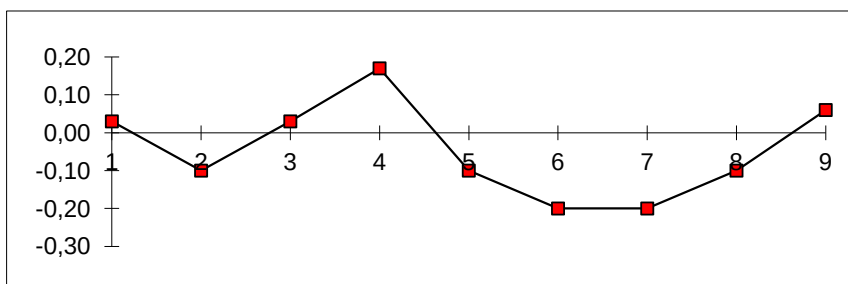
1980	1981	1982	1983	1984	1985	1986	1987	1988
0,08	0,22	0,27	0,21	0,08	0,00	-0,10	-0,30	-0,40



COMPOSANTE TEMPORELLE N°5 PREVISION PAR TRIMESTRE

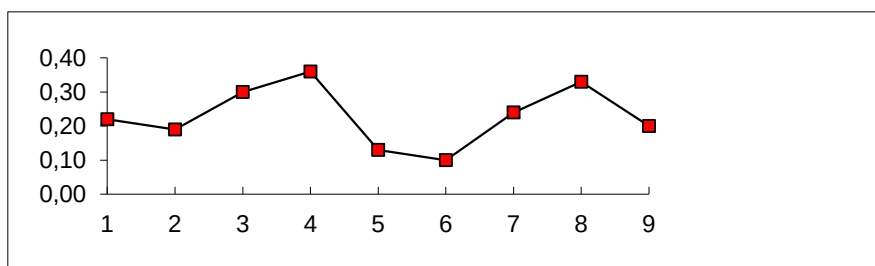
TRIMESTRE N°1

1980	1981	1982	1983	1984	1985	1986	1987	1988
0,03	-0,10	0,03	0,17	-0,10	-0,20	-0,20	-0,10	0,06



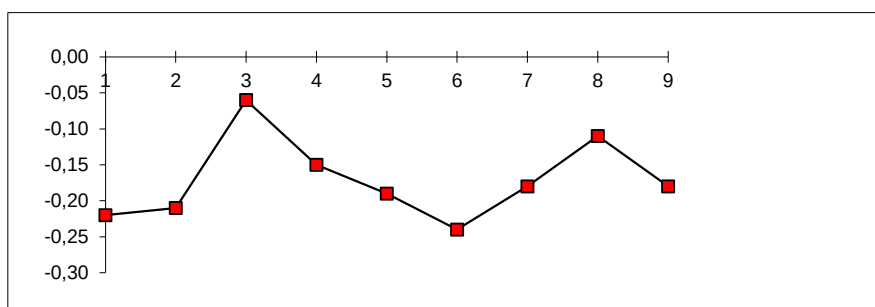
TRIMESTRE N°2

1980	1981	1982	1983	1984	1985	1986	1987	1988
0,22	0,19	0,30	0,36	0,13	0,10	0,24	0,33	0,2



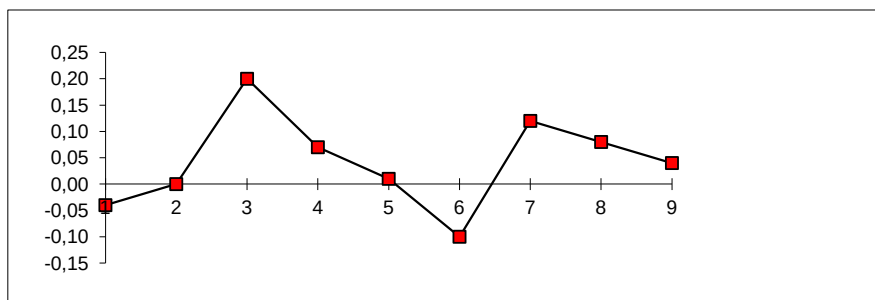
TRIMESTRE N°3

1980	1981	1982	1983	1984	1985	1986	1987	1988
-0,22	-0,21	-0,06	-0,15	-0,19	-0,24	-0,18	-0,11	-0,18



TRIMESTRE N°4

1980	1981	1982	1983	1984	1985	1986	1987	1988
-0,04	0,00	0,20	0,07	0,01	-0,10	0,12	0,08	0,04

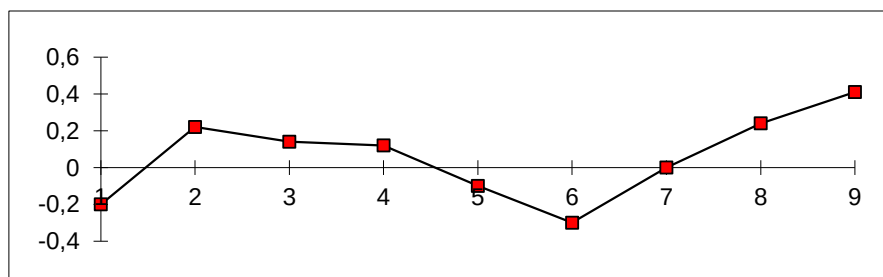


1 1 1 1 1 1 1 1 1 2 2 2

COMPOSANTE TEMPORELLE N°6 PREVISION PAR TRIMESTRE

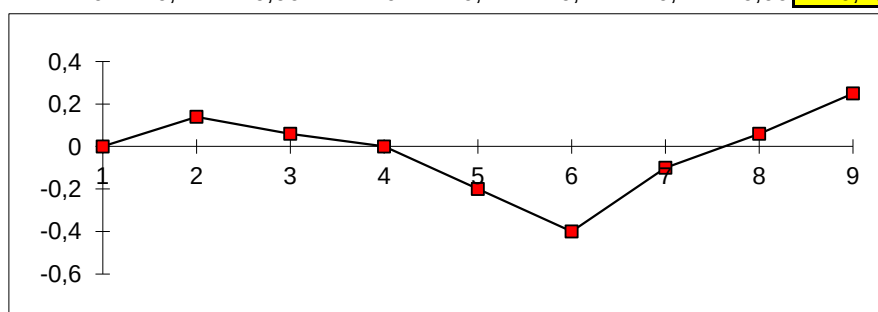
TRIMESTRE N°1

1980	1981	1982	1983	1984	1985	1986	1987	1988
-0,2	0,22	0,14	0,12	-0,1	-0,3	0	0,24	0,41



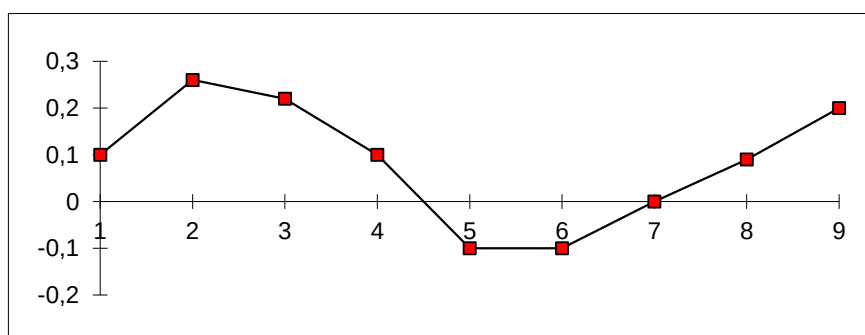
TRIMESTRE N°2

1980	1981	1982	1983	1984	1985	1986	1987	1988
0	0,14	0,06	0	-0,2	-0,4	-0,1	0,06	0,25



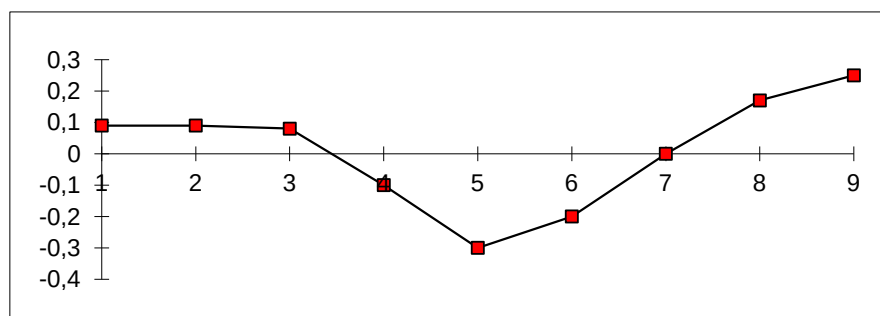
TRIMESTRE N°3

1980	1981	1982	1983	1984	1985	1986	1987	1988
0,1	0,26	0,22	0,1	-0,1	-0,1	0	0,09	0,2

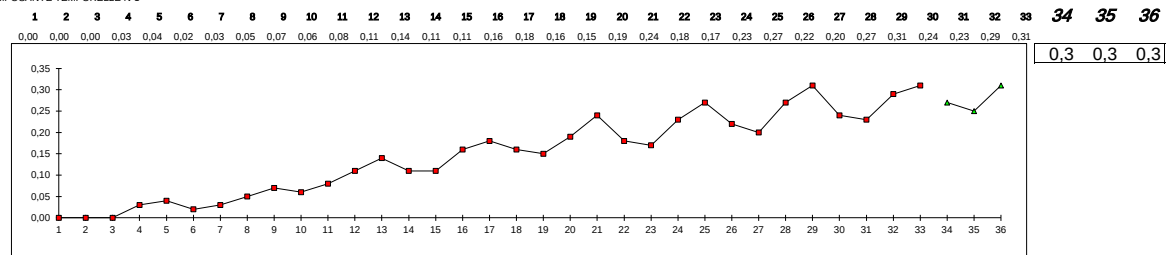


TRIMESTRE N°4

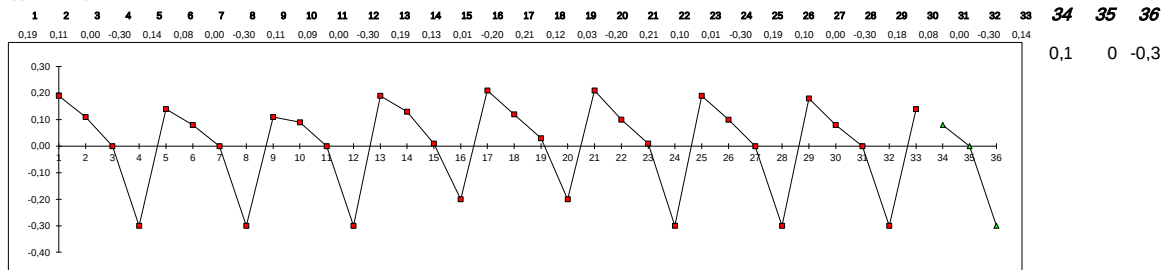
1980	1981	1982	1983	1984	1985	1986	1987	1988
0,09	0,09	0,08	-0,1	-0,3	-0,2	0	0,17	0,25



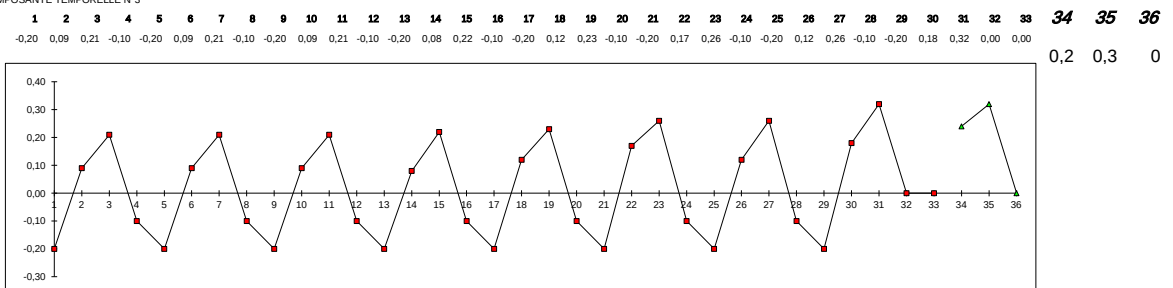
COMPOSANTE TEMPORELLE N°1



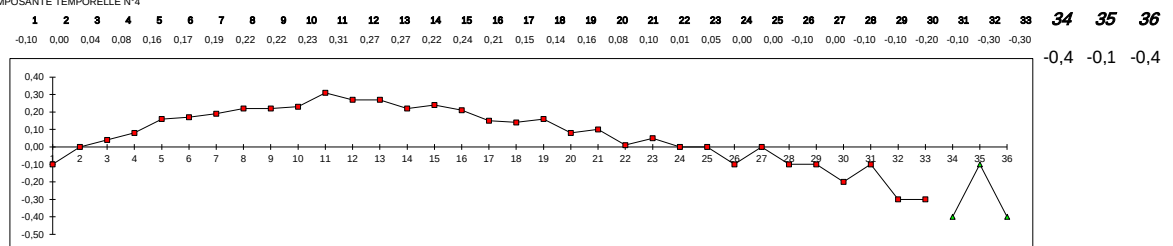
COMPOSANTE TEMPORELLE N°2



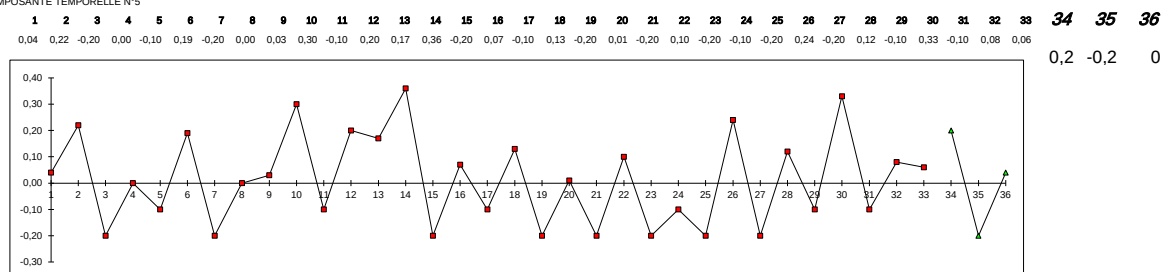
COMPOSANTE TEMPORELLE N°3



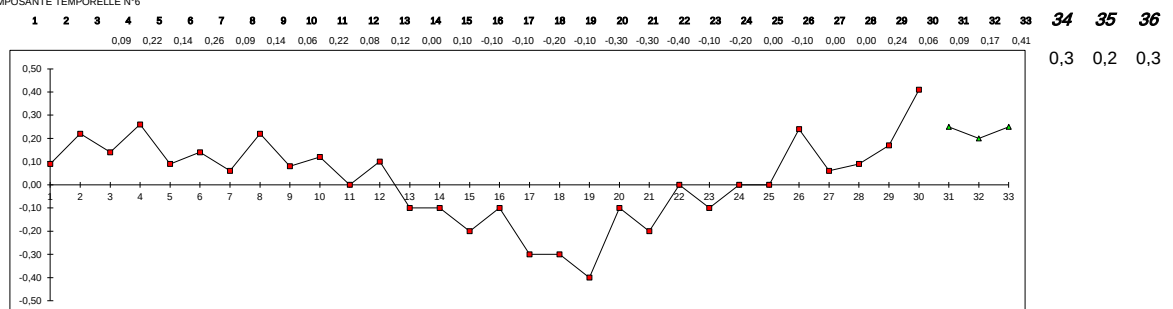
COMPOSANTE TEMPORELLE N°4



COMPOSANTE TEMPORELLE N°5



COMPOSANTE TEMPORELLE N°6



Annexe I-5) Correction de la prévision des composantes temporelles

La plus grosse erreur possible de prévision des composantes temporelles est certainement liée à un changement de direction inattendu. Contre ce genre de situation il est difficile de réagir, sauf comme nous l'avons vu lors du second exemple [VI-4-d], si on dispose d'informations sur le comportement avenir de certaines trajectoires.

Dans le texte qui suit, il ne s'agira pas de ce type d'erreur, mais des erreurs qui sont prévisibles. Nous proposons une correction liée à la déformation en fonction du temps des composantes temporelles .

Les composantes temporelles sont un couple vecteur propre / valeur propre. Pour prévoir « l'avenir » on tient compte de la « forme » de la fonction propre. A chaque période de temps on calcule les composantes temporelles. Si les formes de deux composantes temporelles consécutives sont proches au fur et à mesure de l'écoulement du temps les déformations s'accroissent [Annexes I-3-a et suivantes]. Nous avons essayé, tout à fait empiriquement, de prévoir les déformations pour introduire des corrections. Pour cela nous avons regroupé par saison les résultats dans les tableaux des pages suivantes [Annexes I-5-a à I-5-f]. Chaque colonne de ces tableaux représente la valeur de la même composante temporelle, au même instant mais calculée sur des intervalles différents. Le premier intervalle pris pour le calcul comporte 25 trimestres et le dernier 33. Les graphiques représentent ces tableaux; les longueurs des intervalles étudiés étant représentées en abscisse et les valeurs prises en ordonnée.

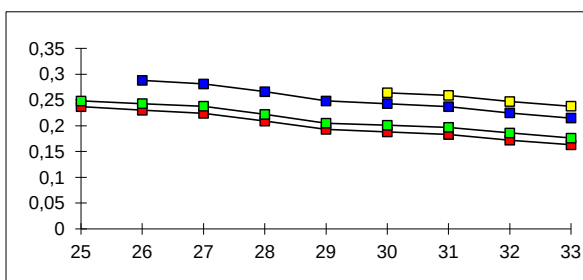
Une synthèse des corrections est proposée dans le tableau de l'annexe I-5_g et les résultats sont exploités en annexes I-7-a à I-7-c.

Remarque:

Nous avons senti la nécessité de faire ces corrections pour tester les gains que cette méthode engendre. Mais compte tenu des erreurs non maîtrisées que l'on fait, le gain, s'il est réel, n'est cependant pas décisif. Ces observations doivent rester comme un possible intéressant sans plus.

ETUDE DES EVOLUTIONS DES COMPOSANTES TEMPORELLES N°1

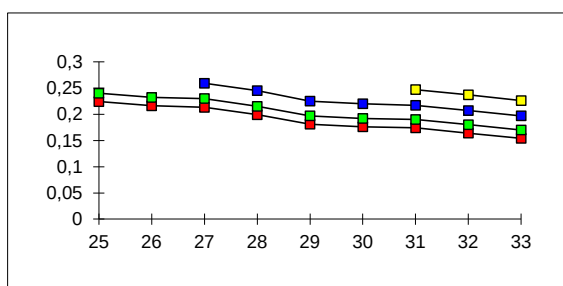
		CORRECTION POUR LA 34 eme VALEUR			
NOMBRE DE TRIMESTRES		18	22	26	30
25		0,237	0,248		
26		0,23	0,243	0,288	
27		0,224	0,238	0,281	
28		0,209	0,222	0,266	
29		0,193	0,205	0,248	
30		0,188	0,201	0,243	0,264
31		0,183	0,197	0,237	0,259
32		0,172	0,186	0,225	0,247
33		0,163	0,176	0,215	0,238



VARIATION MOYENNE	-0,009	-0,009	-0,010	-0,009
DERNIER ECART	-0,009	-0,010	-0,010	-0,009

CORRECTION PROPOSEE	-0,010
------------------------	--------

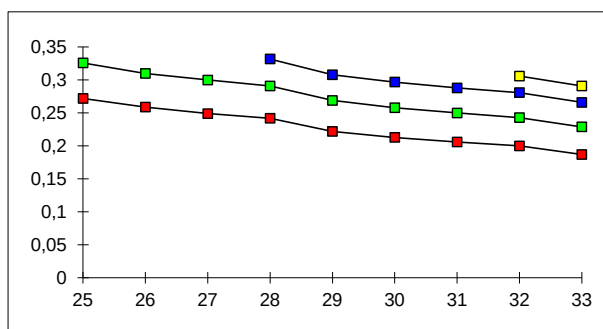
		CORRECTION POUR LA 35 eme VALEUR			
NOMBRE DE TRIMESTRES		19	23	27	31
25		0,224	0,24		
26		0,216	0,232		
27		0,213	0,23	0,259	
28		0,199	0,215	0,245	
29		0,181	0,197	0,225	
30		0,176	0,192	0,22	
31		0,174	0,19	0,217	0,247
32		0,164	0,18	0,207	0,237
33		0,154	0,17	0,197	0,226



VARIATION MOYENNE	-0,009	-0,009	-0,008	-0,119
DERNIER ECART	-0,010	-0,010	-0,010	-0,011

CORRECTION PROPOSEE	-0,010
------------------------	--------

		CORRECTION POUR LA 36 eme VALEUR			
NOMBRE DE TRIMESTRES		20	24	28	32
25		0,272	0,326		
26		0,259	0,31		
27		0,249	0,3		
28		0,242	0,291	0,332	
29		0,222	0,269	0,308	
30		0,213	0,258	0,297	
31		0,206	0,25	0,288	
32		0,2	0,243	0,281	0,306
33		0,187	0,229	0,266	0,291

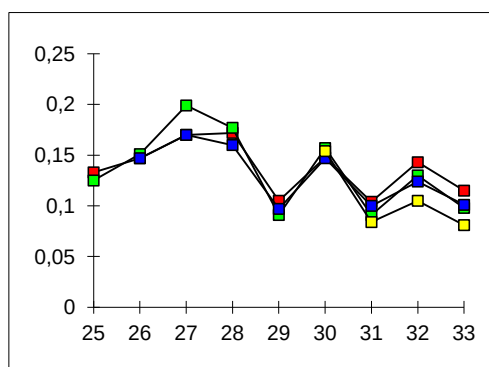


VARIATION MOYENNE	-0,011	-0,012	-0,013	-0,015
DERNIER ECART	-0,013	-0,014	-0,015	-0,015

CORRECTION PROPOSEE	-0,016
------------------------	--------

ETUDE DES EVOLUTIONS DES COMPOSANTES TEMPORELLES N°2

NOMBRE DE TRIMESTRES	CORRECTION POUR LA 34 ^{eme} VALEUR			
	18	22	26	30
25	0,133	0,125		
26	0,147	0,151	0,147	
27	0,17	0,199	0,17	
28	0,172	0,177	0,16	
29	0,105	0,091	0,097	
30	0,148	0,157	0,147	0,154
31	0,104	0,091	0,1	0,084
32	0,143	0,13	0,124	0,105
33	0,115	0,098	0,101	0,081



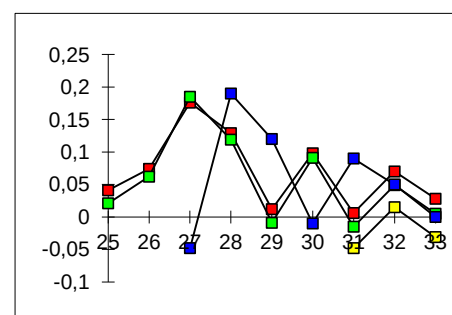
VARIATION
MOYENNE
DERNIER
ECART

-0,002	-0,003	-0,00575	-0,01825
-0,028	-0,032	-0,023	-0,024

CORRECTION
PROPOSEE

0,01

NOMBRE DE TRIMESTRES	CORRECTION POUR LA 35 ^{eme} VALEUR			
	19	23	27	31
25	0,041	0,021		
26	0,074	0,062		
27	0,176	0,185	-0,048	
28	0,129	0,119	0,19	
29	0,012	-0,009	0,12	
30	0,098	0,091	-0,01	
31	0,006	-0,015	0,09	-0,048
32	0,07	0,049	0,05	0,015
33	0,028	0,005	0,00	-0,031



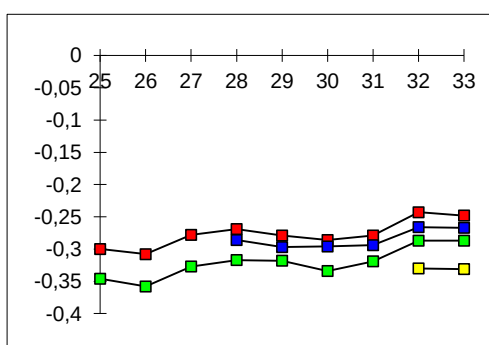
VARATION
MOYENNE
DERNIER
ECART

-0,00144	-0,00178	0,00686	0,00567
-0,042	-0,044	-0,050	-0,046

CORRECTION
PROPOSEE

0,02

NOMBRE DE TRIMESTRES	CORRECTION POUR LA 36 ^{eme} VALEUR			
	20	24	28	32
25	-0,3	-0,346		
26	-0,308	-0,358		
27	-0,278	-0,327		
28	-0,269	-0,317	-0,286	
29	-0,279	-0,318	-0,297	
30	-0,286	-0,334	-0,296	
31	-0,279	-0,319	-0,294	
32	-0,243	-0,287	-0,266	-0,33
33	-0,248	-0,287	-0,267	-0,331



VARATION
MOYENNE
DERNIER
ECART

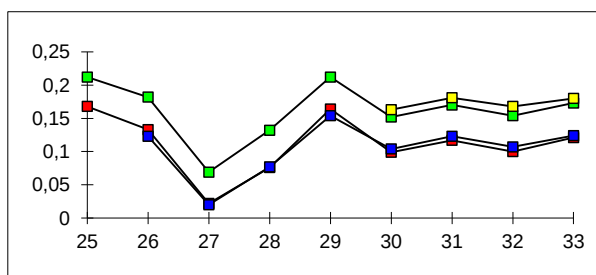
0,00578	0,00656	0,00317	-0,001
-0,005	0,000	-0,001	-0,001

CORRECTION
PROPOSEE

-0,001

ETUDE DES EVOLUTIONS DES COMPOSANTES TEMPORELLES N°3

		CORRECTION POUR LA 34 ^{eme} VALEUR			
NOMBRE DE TRIMESTRES		18	22	26	30
25		0,168	0,212		
26		0,133	0,182	0,123	
27		0,022	0,069	0,02	
28		0,076	0,132	0,077	
29		0,164	0,212	0,154	
30		0,099	0,152	0,104	0,163
31		0,117	0,17	0,123	0,181
32		0,1	0,154	0,107	0,168
33		0,121	0,173	0,124	0,18



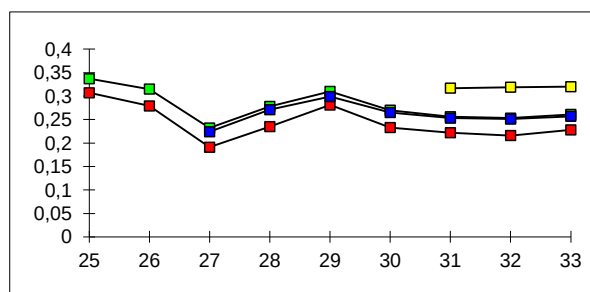
VARIATION
MOYENNE
DERNIER
ECART

-0,006	-0,005	0,000	0,006
0,021	0,019	0,017	0,012

CORRECTION
PROPOSEE

-0,01

		CORRECTION POUR LA 35 ^{eme} VALEUR			
NOMBRE DE TRIMESTRES		19	23	27	31
25		0,307	0,337		
26		0,279	0,315		
27		0,191	0,232	0,224	
28		0,235	0,278	0,271	
29		0,281	0,31	0,299	
30		0,233	0,27	0,265	
31		0,222	0,256	0,253	0,317
32		0,216	0,253	0,251	0,319
33		0,228	0,261	0,257	0,32



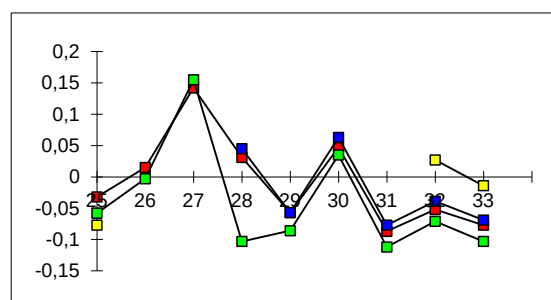
VARIATION
MOYENNE
DERNIER
ECART

-0,010	-0,010	0,043	0,160
0,012	0,008	0,006	0,001

CORRECTION
PROPOSEE

0,02

		CORRECTION POUR LA 36 ^{eme} VALEUR			
NOMBRE DE TRIMESTRES		20	24	28	32
25		-0,032	-0,058		-0,077
26		0,015	-0,003		
27		0,142	0,155		
28		0,031	-0,103	0,045	
29		-0,057	-0,086	-0,057	
30		0,047	0,035	0,063	
31		-0,087	-0,112	-0,077	
32		-0,052	-0,071	-0,039	0,027
33		-0,077	-0,103	-0,069	-0,014



VARIATION
MOYENNE
DERNIER
ECART

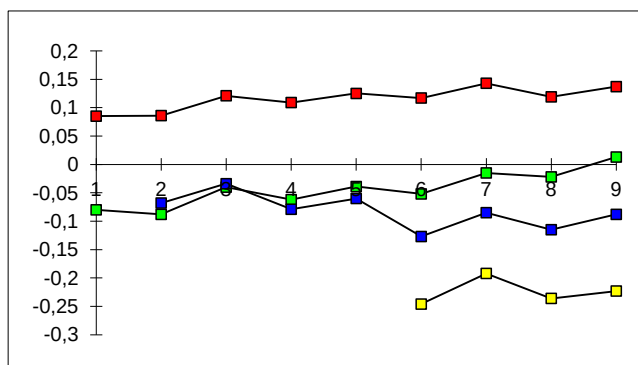
-0,006	-0,006	-0,023	-0,041
-0,025	-0,032	-0,030	-0,041

CORRECTION
PROPOSEE

0,04

ETUDE DES EVOLUTION DES COMPOSANTES TEMPORELLES N°4

		CORRECTION POUR LA 34 ^{eme} VALEUR			
NOMBRE DE TRIMESTRES		18	22	26	30
25		0,085	-0,08		
26		0,086	-0,088	-0,068	
27		0,121	-0,04	-0,034	
28		0,109	-0,062	-0,079	
29		0,125	-0,039	-0,06	
30		0,117	-0,052	-0,127	-0,246
31		0,143	-0,015	-0,085	-0,192
32		0,119	-0,022	-0,115	-0,236
33		0,137	0,013	-0,088	-0,223



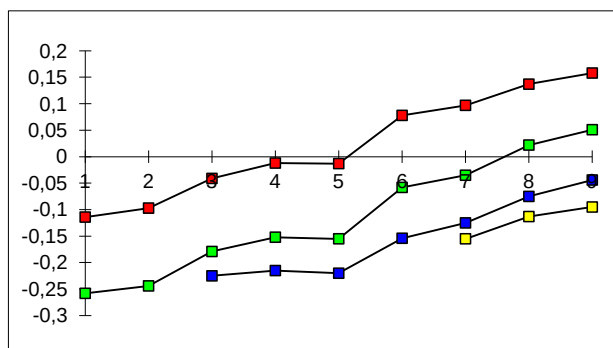
VARIATION
MOYENNE
DERNIER
ECART

0,007	0,012	-0,003	0,008
0,018	0,035	0,027	0,013

CORRECTION
PROPOSEE

-0,008

		CORRECTION POUR LA 35 ^{eme} VALEUR			
NOMBRE DE TRIMESTRES		19	23	27	31
25		-0,114	-0,258		
26		-0,097	-0,244		
27		-0,041	-0,179	-0,225	
28		-0,012	-0,152	-0,215	
29		-0,013	-0,155	-0,22	
30		0,078	-0,058	-0,154	
31		0,097	-0,035	-0,125	-0,155
32		0,137	0,022	-0,075	-0,113
33		0,158	0,051	-0,044	-0,095



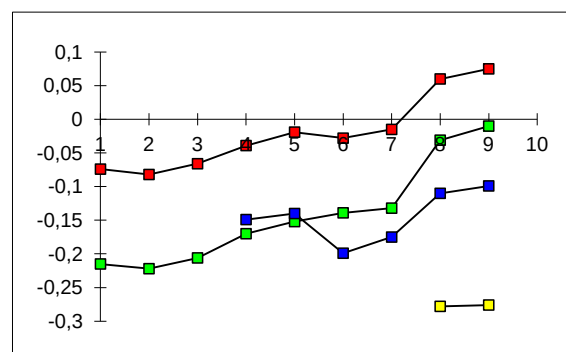
VARIATION
MOYENNE
DERNIER
ECART

0,034	0,039	-0,007	-0,048
0,021	0,029	0,031	0,018

CORRECTION
PROPOSEE

0,018

		CORRECTION POUR LA 36 ^{eme} VALEUR			
NOMBRE DE TRIMESTRES		20	24	28	32
25		-0,074	-0,215		
26		-0,082	-0,222		
27		-0,066	-0,206		
28		-0,039	-0,17	-0,149	
29		-0,019	-0,152	-0,14	
30		-0,028	-0,139	-0,199	
31		-0,015	-0,132	-0,175	
32		0,06	-0,031	-0,11	-0,278
33		0,075	-0,01	-0,099	-0,276



VARIATION
MOYENNE
DERNIER
ECART

0,008	-0,004	-0,022	0,002
0,015	0,021	0,011	0,002

CORRECTION
PROPOSEE

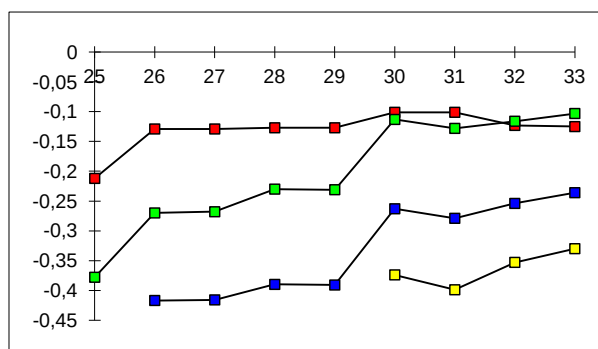
0,004

ETUDE DES EVOLUTIONS DES COMPOSANTES TEMPORELLES N°5

NOMBRE DE TRIMESTRES		CORRECTION POUR LA 34 ^{eme} VALEUR			
		18	22	26	30
25		-0,212	-0,378		
26		-0,129	-0,27	-0,417	
27		-0,129	-0,268	-0,416	
28		-0,127	-0,23	-0,39	
29		-0,127	-0,231	-0,391	
30		-0,101	-0,113	-0,263	-0,374
31		-0,101	-0,128	-0,279	-0,399
32		-0,123	-0,116	-0,254	-0,353
33		-0,125	-0,103	-0,236	-0,33

VARIATION
MOYENNE
DERNIER
ECART

0,011	0,034	0,026	0,015
-0,002	0,013	0,018	0,023

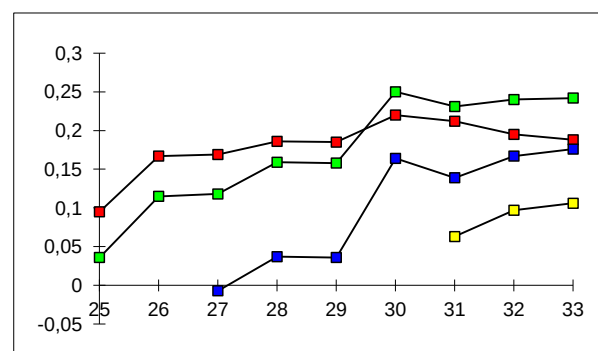


CORRECTION PROPOSEE 0,02

NOMBRE DE TRIMESTRES		CORRECTION POUR LA 35 ^{eme} VALEUR			
		19	23	27	31
25		0,095	0,036		
26		0,167	0,115		
27		0,169	0,118	-0,007	
28		0,186	0,159	0,037	
29		0,185	0,158	0,036	
30		0,22	0,25	0,164	
31		0,212	0,231	0,139	0,063
32		0,195	0,24	0,167	0,097
33		0,188	0,242	0,176	0,106

VARIATION
MOYENNE
DERNIER
ECART

0,012	0,026	0,031	0,022
-0,007	0,002	0,009	0,009

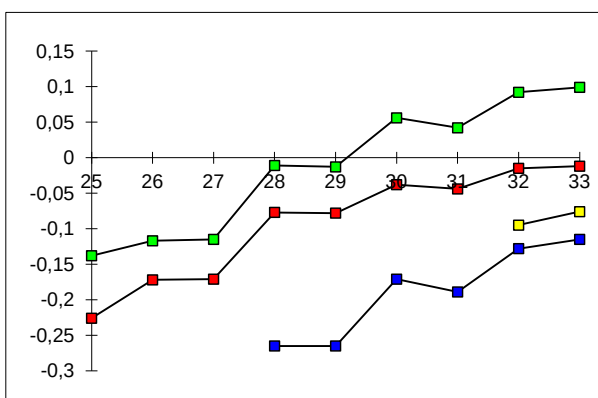


CORRECTION PROPOSEE 0,01

NOMBRE DE TRIMESTRES		CORRECTION POUR LA 36 ^{eme} VALEUR			
		20	24	28	32
25		-0,226	-0,138		
26		-0,172	-0,117		
27		-0,171	-0,115		
28		-0,077	-0,011	-0,265	
29		-0,078	-0,013	-0,265	
30		-0,038	0,056	-0,171	
31		-0,044	0,042	-0,189	
32		-0,015	0,092	-0,128	-0,095
33		-0,012	0,099	-0,115	-0,076

VARIATION
MOYENNE
DERNIER
ECART

0,027	0,030	0,030	0,019
0,003	0,007	0,013	0,019

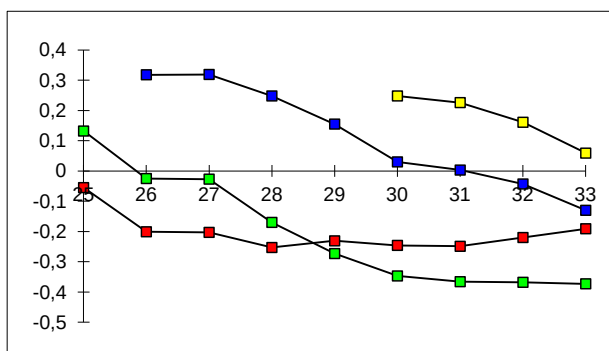


CORRECTION PROPOSEE 0,02

ETUDE DES EVOLUTIONS DES COMPOSANTES TEMPORELLES N°6

		CORRECTION POUR LA 34 eme VALEUR			
NOMBRE DE TRIMESTRES		18	22	26	30
25		-0,055	0,132		
26		-0,201	-0,025	0,318	
27		-0,203	-0,027	0,319	
28		-0,253	-0,17	0,248	
29		-0,231	-0,273	0,155	
30		-0,246	-0,347	0,03	0,248
31		-0,249	-0,366	0,003	0,226
32		-0,22	-0,368	-0,043	0,161
33		-0,191	-0,373	-0,13	0,059

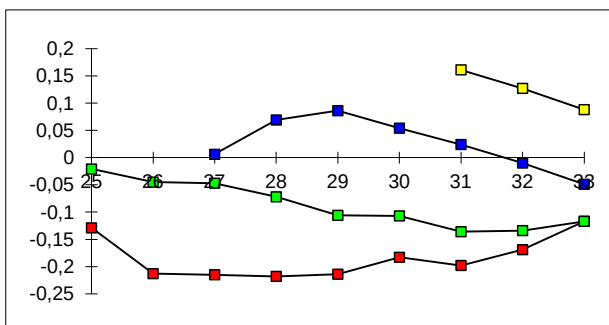
VARIATION MOYENNE	-0,017	-0,063	-0,064	-0,063
DERNIER ECART	0,029	-0,005	-0,087	-0,102



CORRECTION PROPOSEE	-0,08
------------------------	-------

		CORRECTION POUR LA 35 eme VALEUR			
NOMBRE DE TRIMESTRES		19	23	27	31
25		-0,129	-0,021		
26		-0,213	-0,045		
27		-0,215	-0,047	0,006	
28		-0,218	-0,072	0,069	
29		-0,214	-0,106	0,086	
30		-0,183	-0,107	0,054	
31		-0,198	-0,136	0,024	0,161
32		-0,169	-0,134	-0,01	0,127
33		-0,117	-0,117	-0,049	0,088

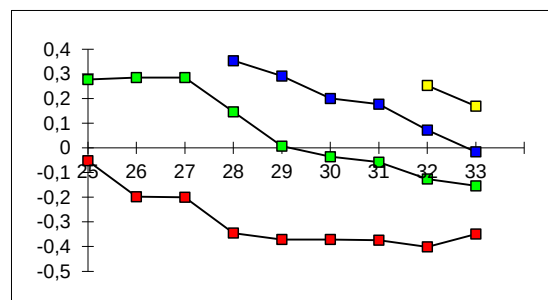
VARIATION MOYENNE	0,002	-0,012	-0,009	-0,037
DERNIER ECART	0,052	0,017	-0,039	-0,039



CORRECTION PROPOSEE	-0,03
------------------------	-------

		CORRECTION POUR LA 36 eme VALEUR			
NOMBRE DE TRIMESTRES		20	24	28	32
25		-0,052	0,277		
26		-0,198	0,285		
27		-0,2	0,285		
28		-0,345	0,146	0,353	
29		-0,371	0,007	0,291	
30		-0,371	-0,036	0,2	
31		-0,374	-0,058	0,177	
32		-0,401	-0,126	0,072	0,253
33		-0,349	-0,154	-0,016	0,169

VARIATION MOYENNE	-0,037	-0,054	-0,074	-0,084
DERNIER ECART	0,052	-0,028	-0,088	-0,084



CORRECTION PROPOSEE	-0,08
------------------------	-------

RECAPITULATIF
DES PROLONGEMENTS DES 6 PREMIERES COMPOSANTES TEMPORELLES ET PROPOSITION DE CORRECTION

TENDANCE N° 1 [page 114]	PREVISION BRUTE	CORRECTION	PREVISION CORRIGEE
Trimestre N°34	0,261	-0,01	0,251
Trimestre N°35	0,25	-0,01	0,24
Trimestre N°36	0,32	-0,016	0,304

TENDANCE N° 2 [page 115]	PREVISION BRUTE	CORRECTION	PREVISION CORRIGEE
Trimestre N°34	0,08	0,01	0,09
Trimestre N°35	0	0,02	0,02
Trimestre N°36	-0,28	-0,001	-0,281

TENDANCE N° 3 [page 116]	PREVISION BRUTE	CORRECTION	PREVISION CORRIGEE
Trimestre N°34	0,24	-0,01	0,23
Trimestre N°35	0,32	0,02	0,34
Trimestre N°36	0	0,04	0,04

TENDANCE N° 4 [page 117]	PREVISION BRUTE	CORRECTION	PREVISION CORRIGEE
Trimestre N°34	-0,35	-0,008	-0,358
Trimestre N°35	-0,1	0,018	-0,082
Trimestre N°36	-0,4	0,004	-0,396

TENDANCE N° 5 [page 118]	PREVISION BRUTE	CORRECTION	PREVISION CORRIGEE
Trimestre N°34	0,2	0,02	0,22
Trimestre N°35	-0,18	0,01	-0,17
Trimestre N°36	0,04	0,02	0,06

TENDANCE N° 6 [page 119]	PREVISION BRUTE	CORRECTION	PREVISION CORRIGEE
Trimestre N°34	0,25	-0,08	0,17
Trimestre N°35	0,2	-0,03	0,17
Trimestre N°36	0,25	-0,08	0,17

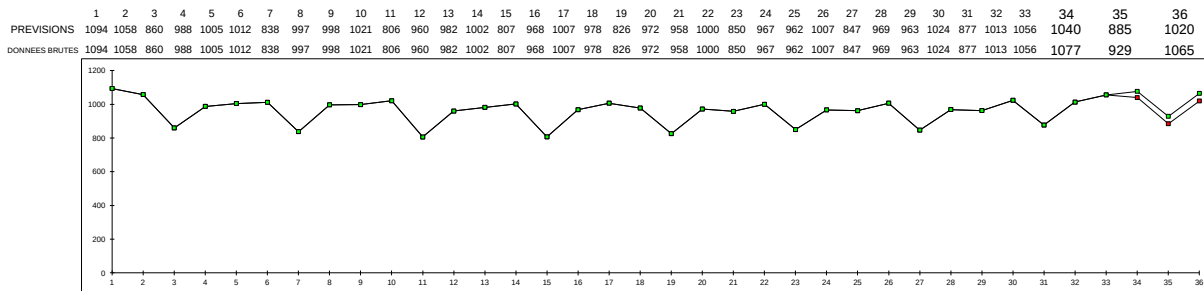
Ann-1-6) Pr vision trajectoire par trajectoire des donn es :

Dans les pages suivantes nous proposons de prolonger, avec une m thode identique   celles que nous avons utilis e pour les composantes temporelles, chaque s rie chronologique. Le but de ce travail est de comparer la performance de notre m thode   celle qui consiste   faire une pr vision graphique trajectoire par trajectoire. Nous avons obtenu de meilleurs r sultats en prolongeant le processus   l'aide de notre m thode, que ceux obtenu par la m thode de pr vision trajectoire par trajectoire. Ces r sultats  taient attendus et confortent l'id e que la pr vision   l'aide de l'analyse harmonique int gre plus d'information que celle faite sur chaque s rie chronologique

Les tableaux ci apr s sont une  num ration des r sultats.

Vous trouverez apr s ces tableaux trois tableaux comparant les diff rents r sultats (Ann I-7-a , Ann I-7-b et Ann I-7-c)

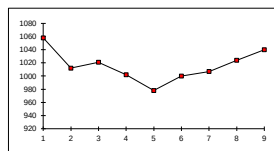
DONNEES BRUTES ET PREVISION "NAIVE" DE LA FONCTION N° 1



PREVISION PAR SAISONS

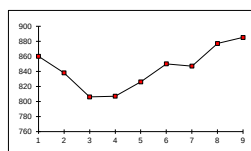
2ème TRIMESTRE

2 6 10 14 18 22 26 30 34
1058 1012 1002 978 1000 1024 1040



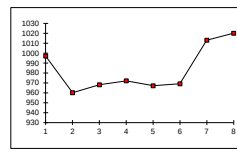
3ème TRIMESTRE

3 7 11 15 19 23 27 31
860 838 806 807 826 850 847 877 885

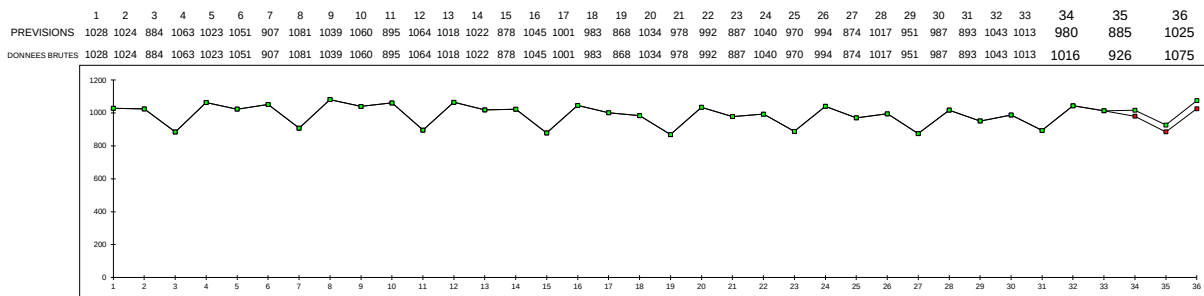


4ème TRIMESTRE

8 12 16 20 24 28 32 36
997 960 968 972 967 969 1013 1020



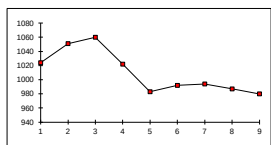
DONNEES BRUTES ET PREVISION "NAIVE" DE LA FONCTION N° 2



PREVISION PAR SAISONS

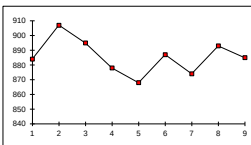
2ème TRIMESTRE

2 6 10 14 18 22 26 30 34
1024 1051 1060 1022 983 992 994 987 980



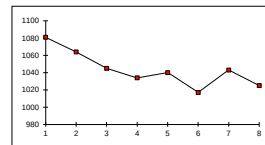
3ème TRIMESTRE

3 7 11 15 19 23 27 31 35
884 907 895 878 868 887 874 893 885

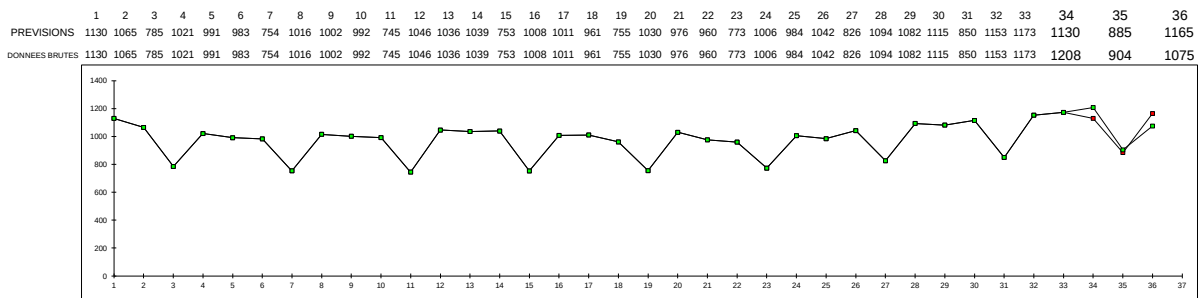


4ème TRIMESTRE

8 12 16 20 24 28 32 36
1081 1064 1045 1034 1040 1017 1043 1025



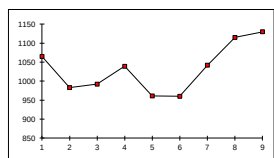
DONNEES BRUTES ET PREVISION "NAIVE" DE LA FONCTION N° 3



PREVISION PAR SAISONS

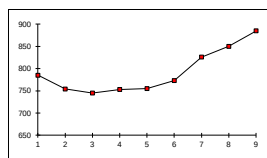
2ème TRIMESTRE

2 6 10 14 18 22 26 30 34
1065 983 992 1039 961 960 1042 1115 1130



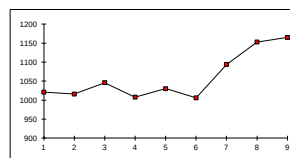
3ème TRIMESTRE

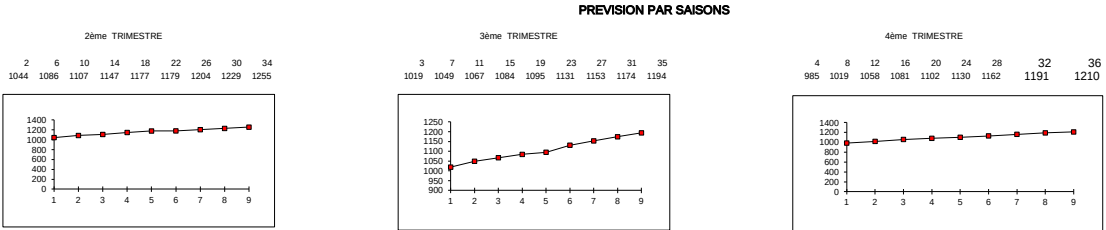
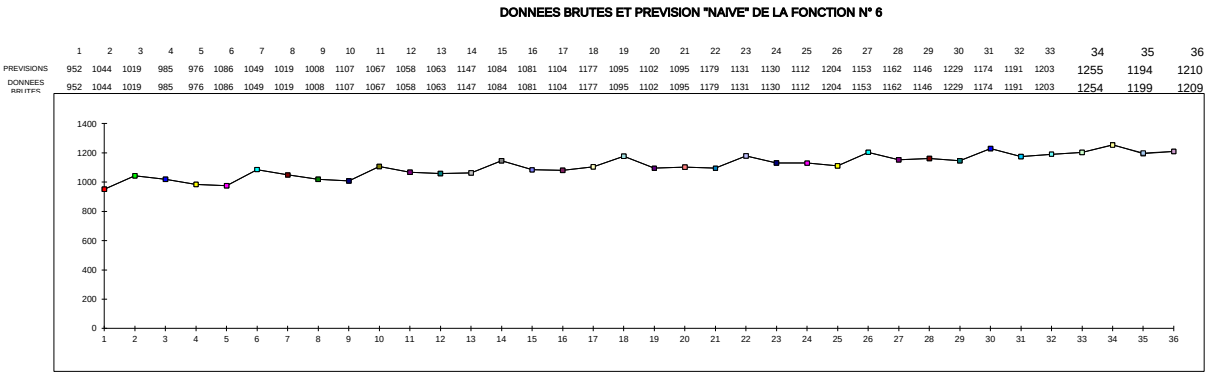
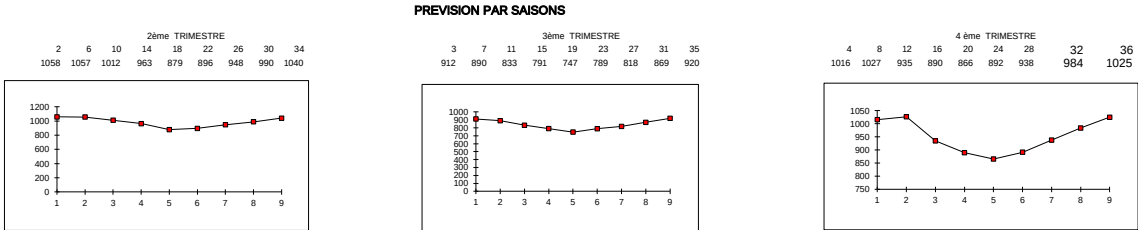
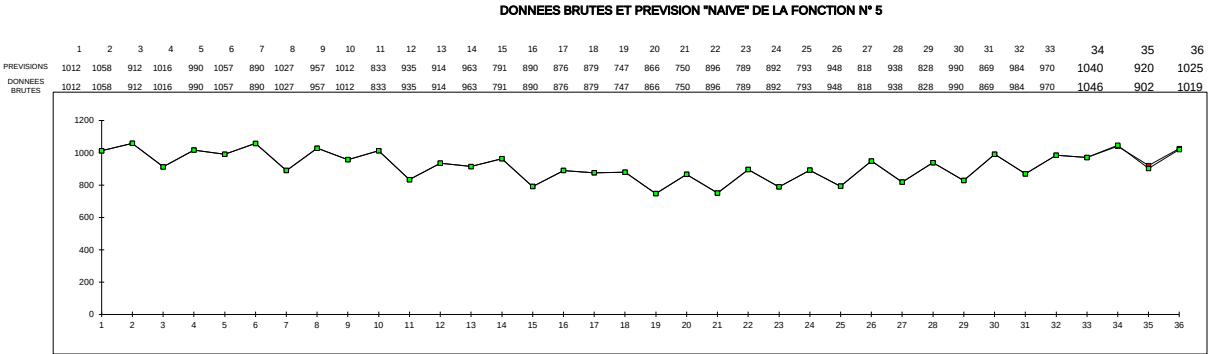
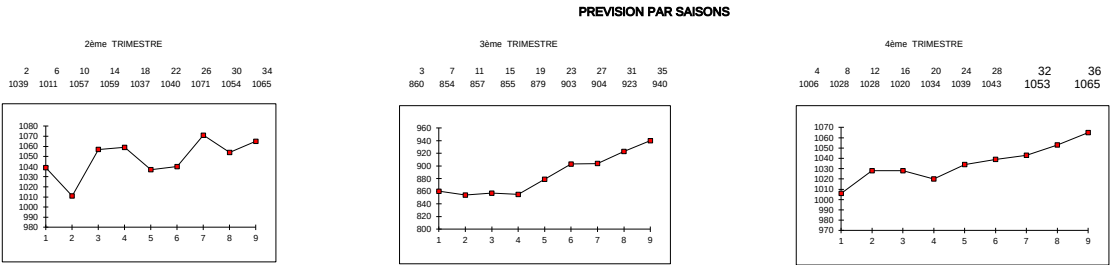
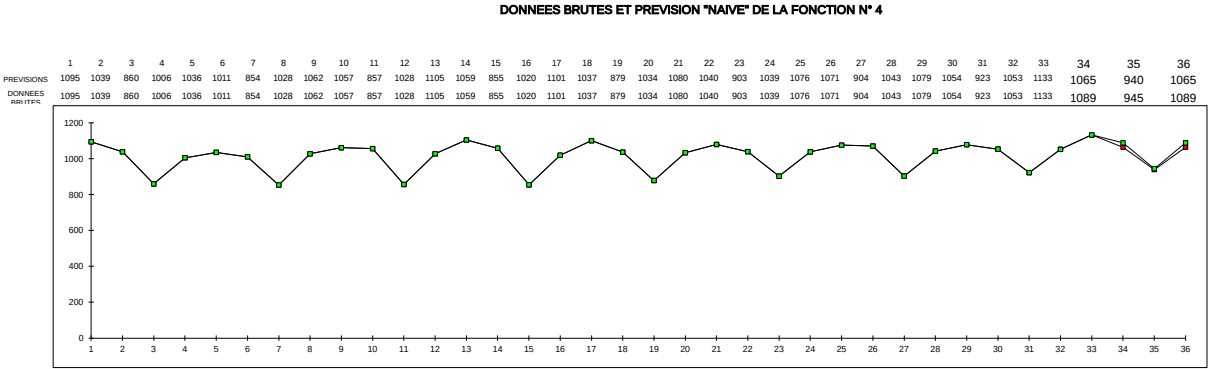
3 7 11 15 19 23 27 31 35
785 754 753 753 755 773 826 850 885



4ème TRIMESTRE

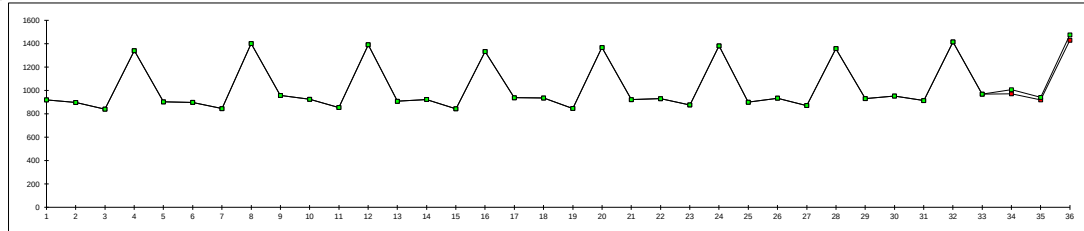
4 8 12 16 20 24 28 32 36
1021 1016 1046 1008 1030 1006 1094 1153 1165



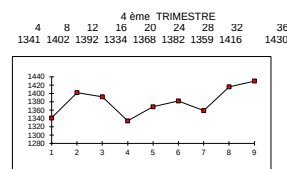
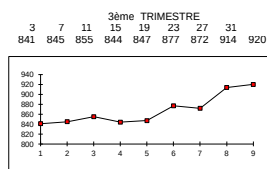
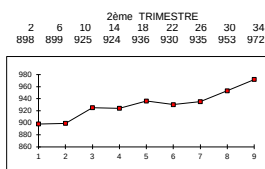


DONNEES BRUTES ET PREVISION "NAÏVE" DE LA FONCTION N° 7

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
PREVISIONS	920	898	841	1341	903	899	845	1402	958	925	855	1392	908	924	844	1334	938	936	847	1368	922	930	877	1382	900	935	872	1359	931	953	914	1416	969	972	920	1430
DONNEES BRUTES	920	898	841	1341	903	899	845	1402	958	925	855	1392	908	924	844	1334	938	936	847	1368	922	930	877	1382	900	935	872	1359	931	953	914	1416	969	1007	941	1476

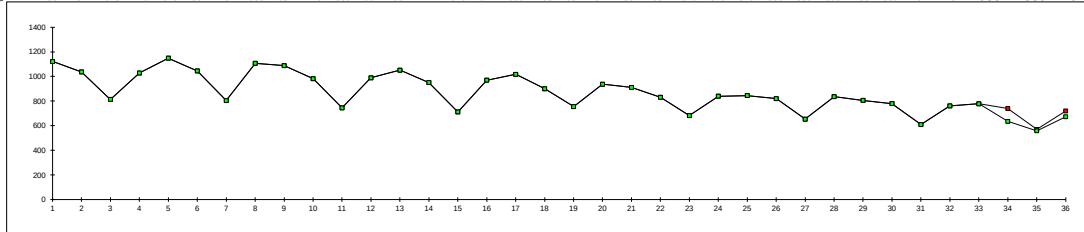


PREVISION PAR SAISONS

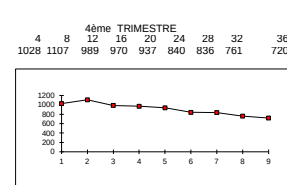
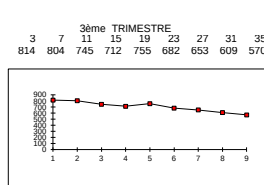
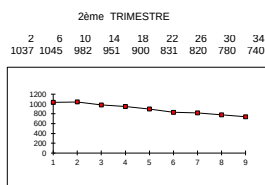


DONNEES BRUTES ET PREVISION "NAÏVE" DE LA FONCTION N° 8

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
PREVISIONS	1121	1037	814	1028	1149	1045	804	1107	1088	982	745	989	1051	951	712	970	1017	900	755	937	911	831	682	840	845	820	653	836	806	780	609	761	778	740	570	720
DONNEES BRUTES	1121	1037	814	1028	1149	1045	804	1107	1088	982	745	989	1051	951	712	970	1017	900	755	937	911	831	682	840	845	820	653	836	806	780	609	761	778	635	559	673

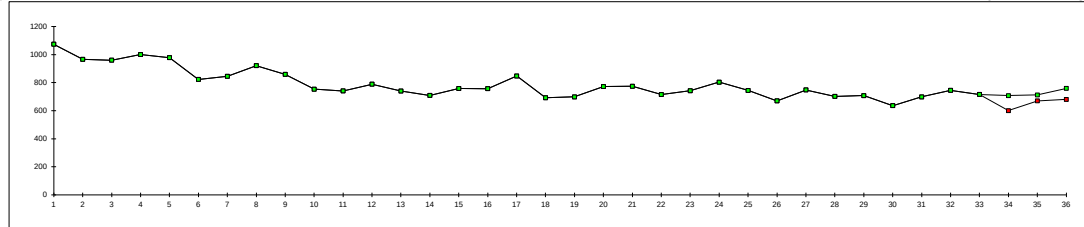


PREVISION PAR SAISONS

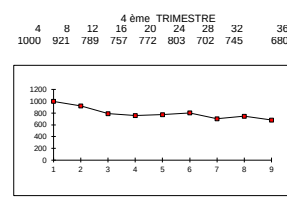
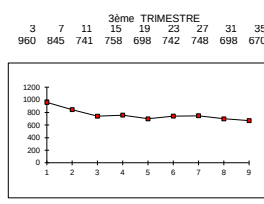
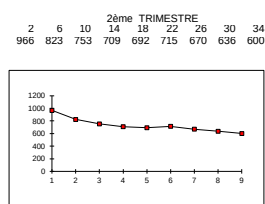


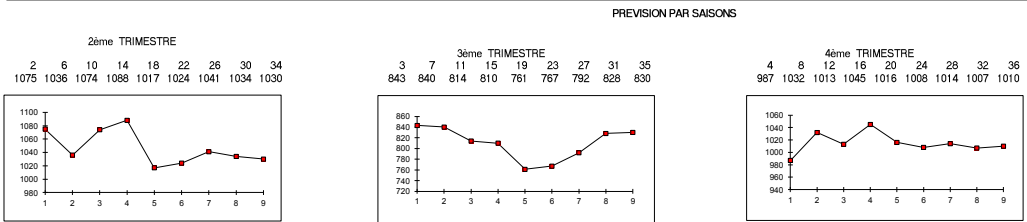
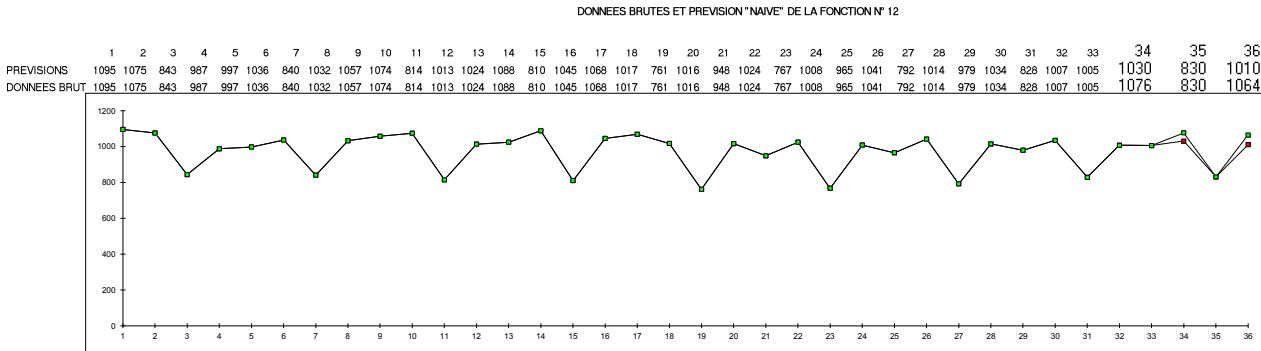
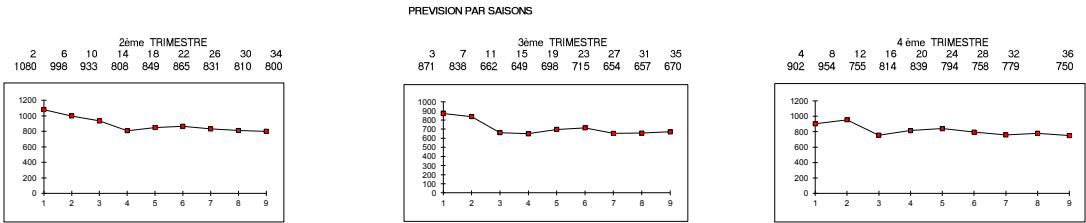
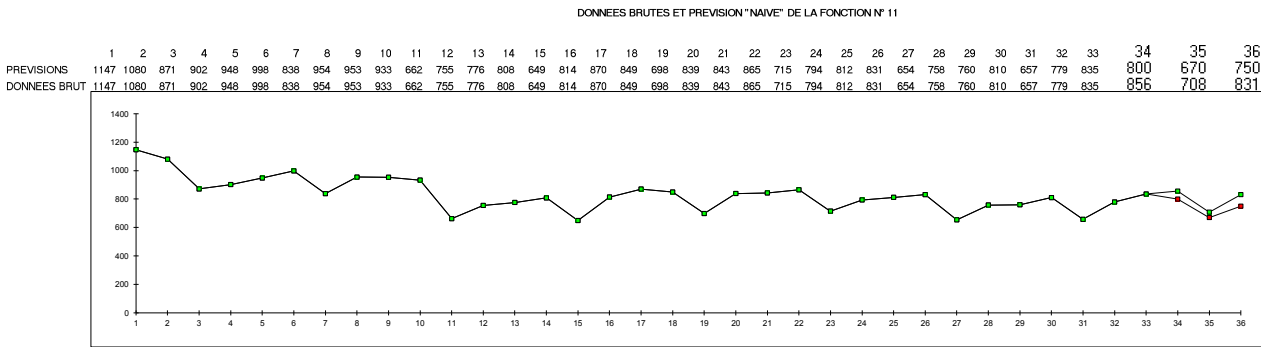
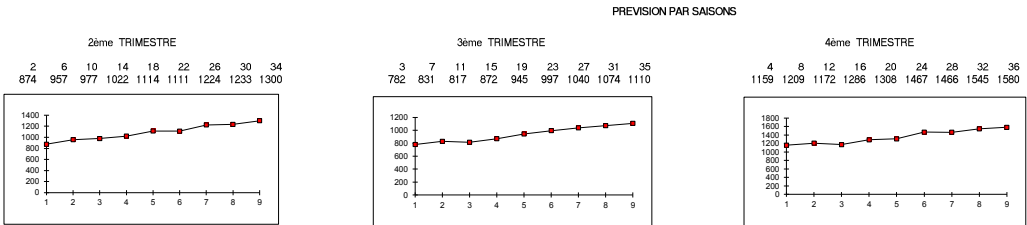
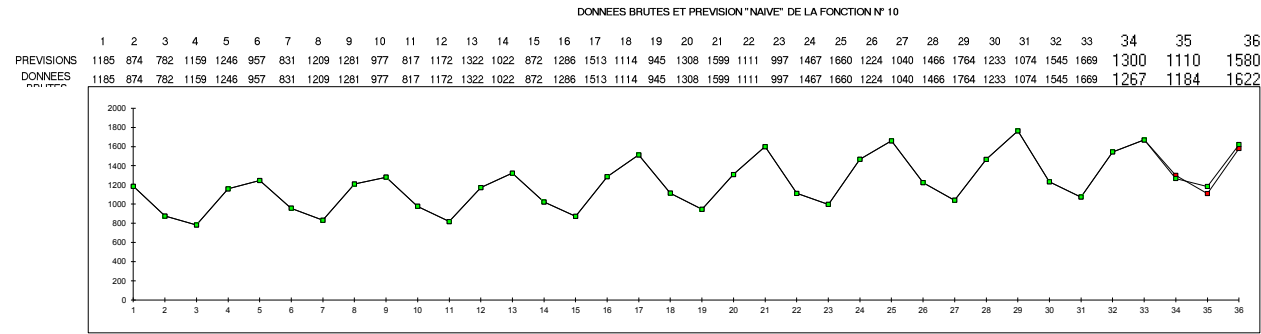
DONNEES BRUTES ET PREVISION "NAÏVE" DE LA FONCTION N° 9

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
PREVISIONS	1074	966	960	1000	978	823	845	921	859	753	741	789	740	709	758	757	847	692	698	772	774	715	742	803	744	670	748	702	707	636	698	745	716	600	670	680
DONNEES BRUTES	1074	966	960	1000	978	823	845	921	859	753	741	789	740	709	758	757	847	692	698	772	774	715	742	803	744	670	748	702	707	636	698	745	716	707	712	759

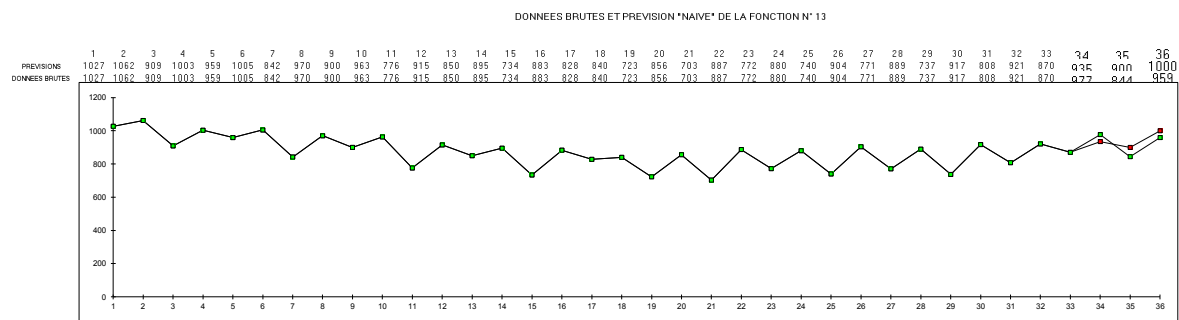


PREVISION PAR SAISONS

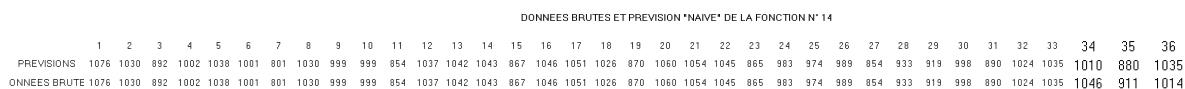
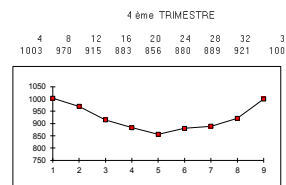
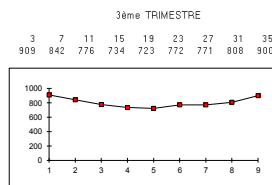
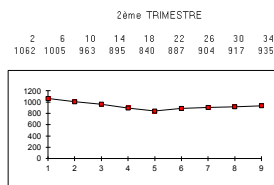




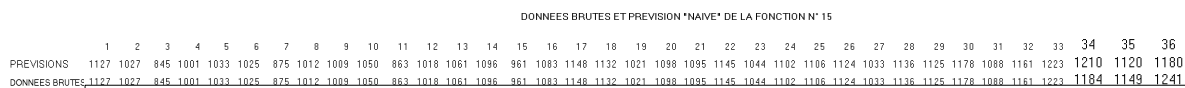
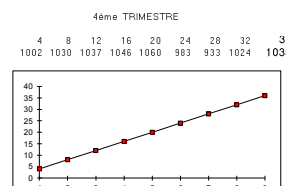
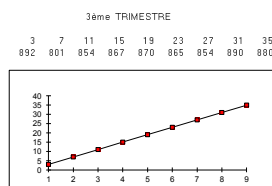
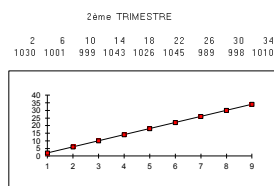
PREVISION "NAÏVE" FONCTIONS 13 A 15



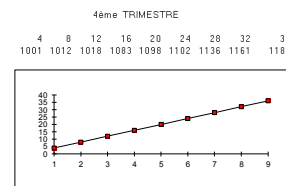
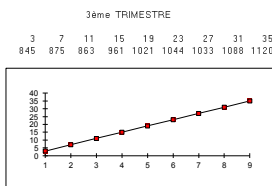
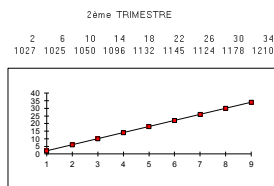
PREVISION PAR SAISONS

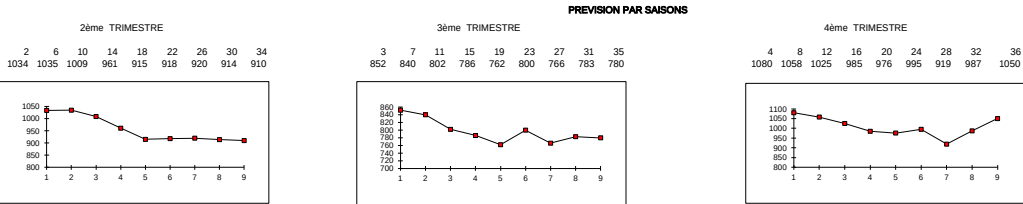
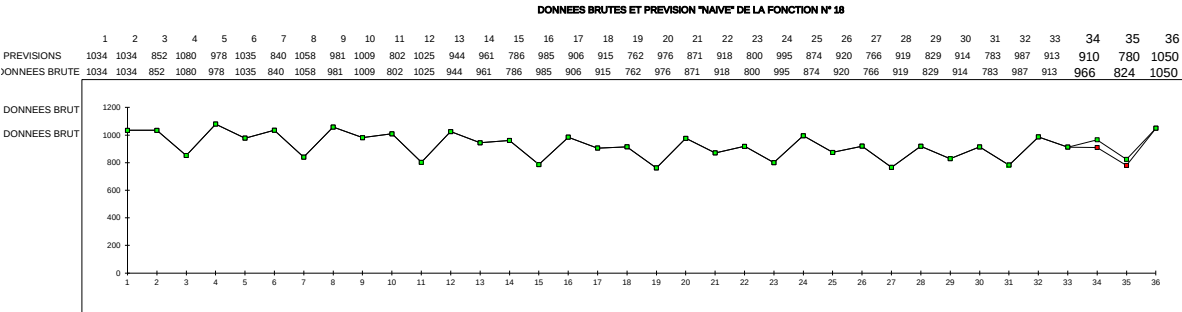
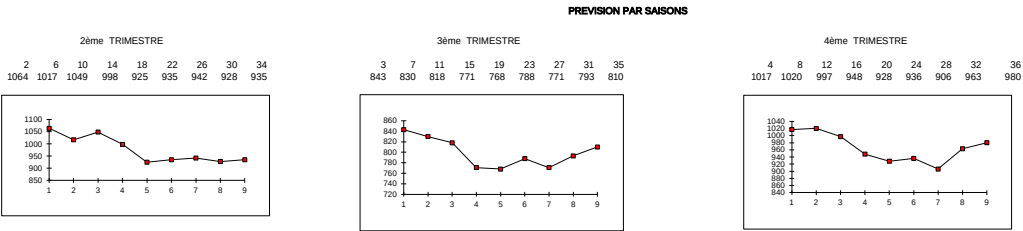
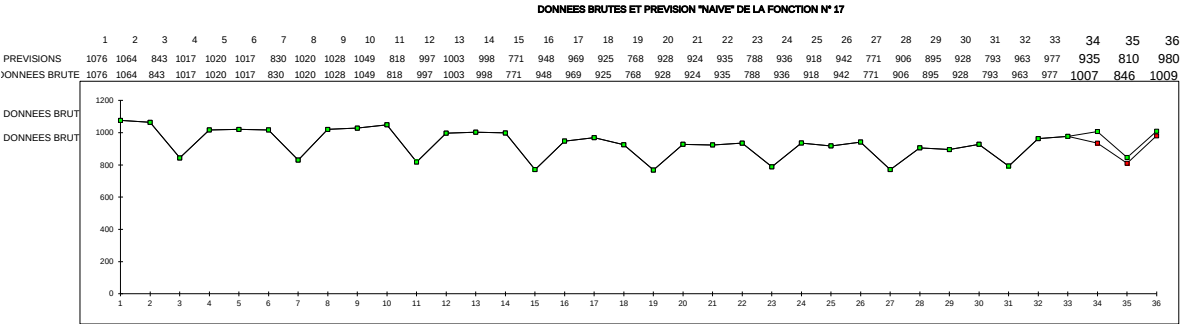
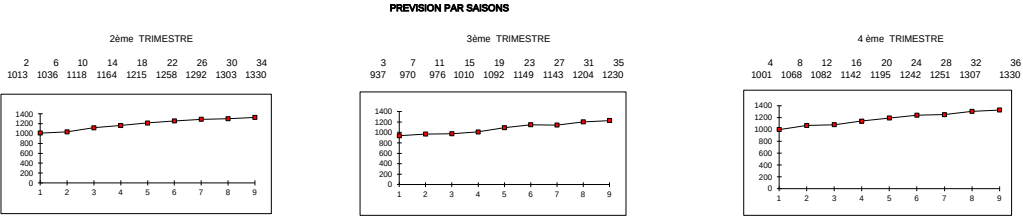
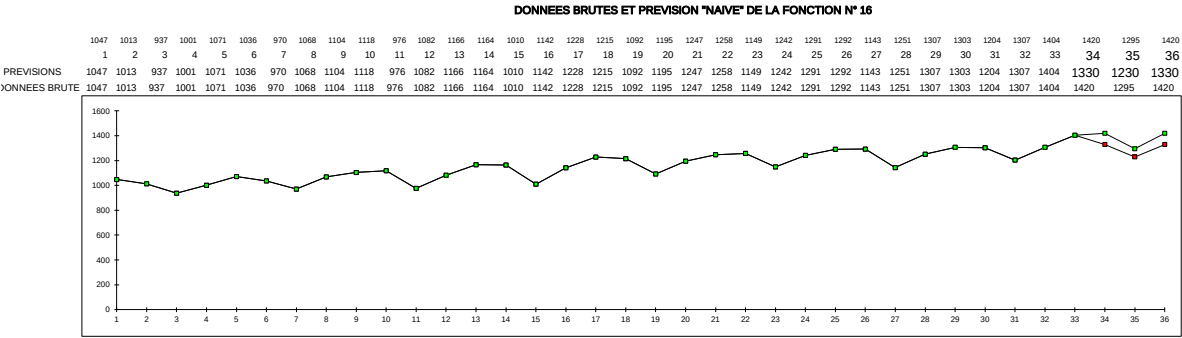


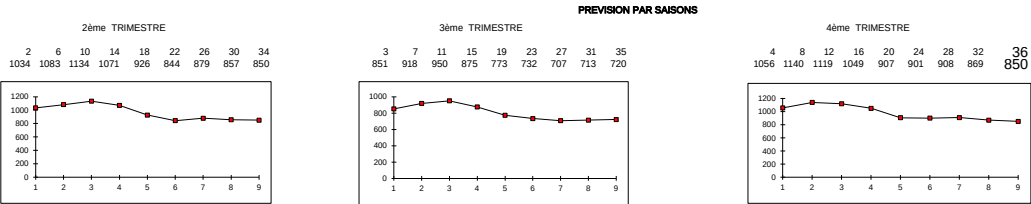
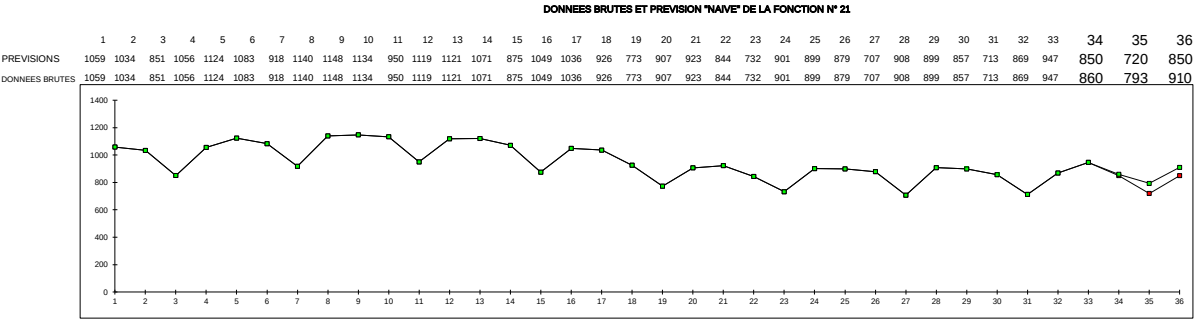
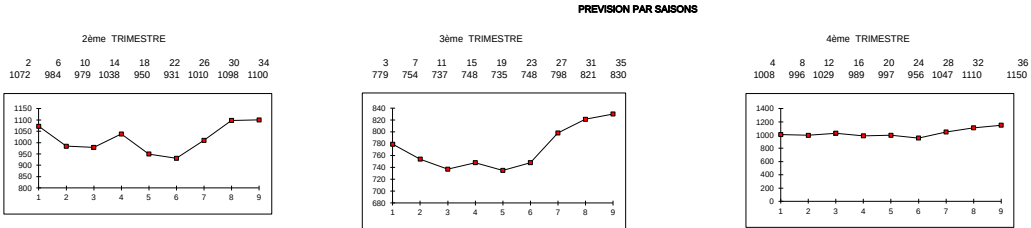
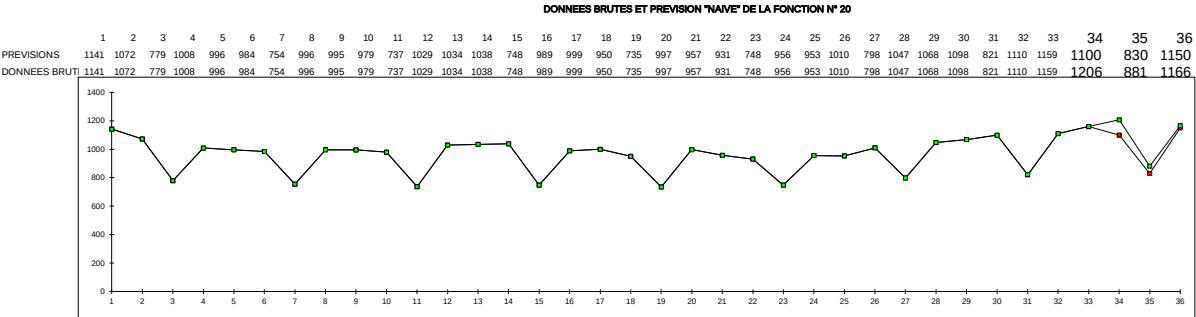
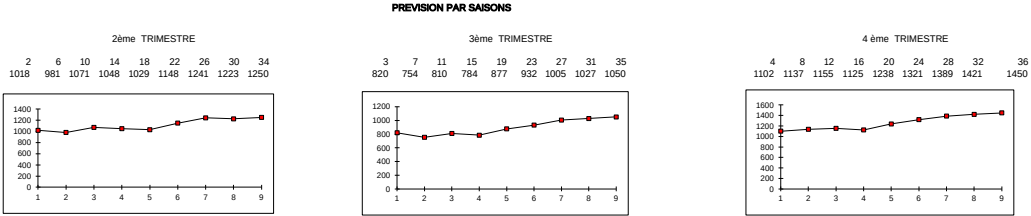
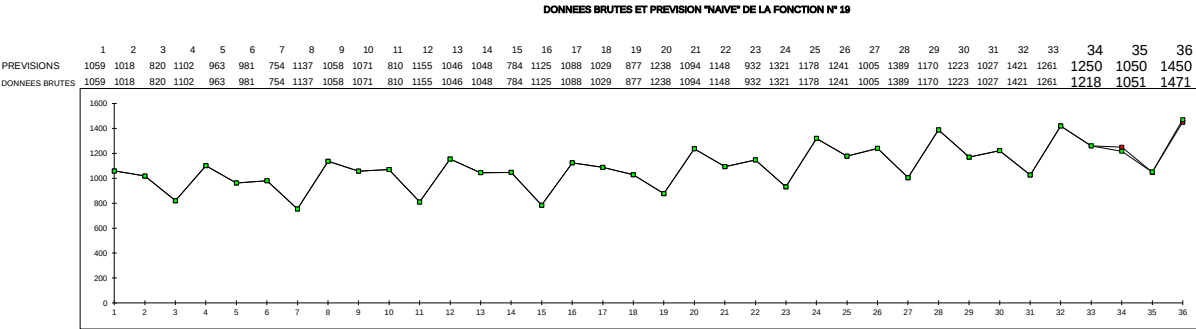
PREVISION PAR SAISONS

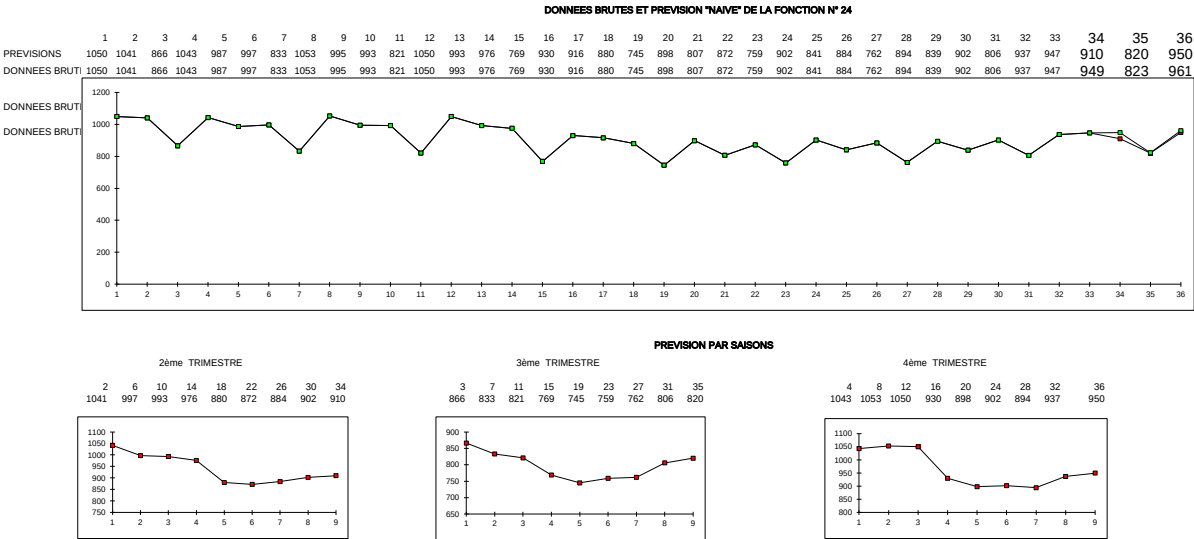
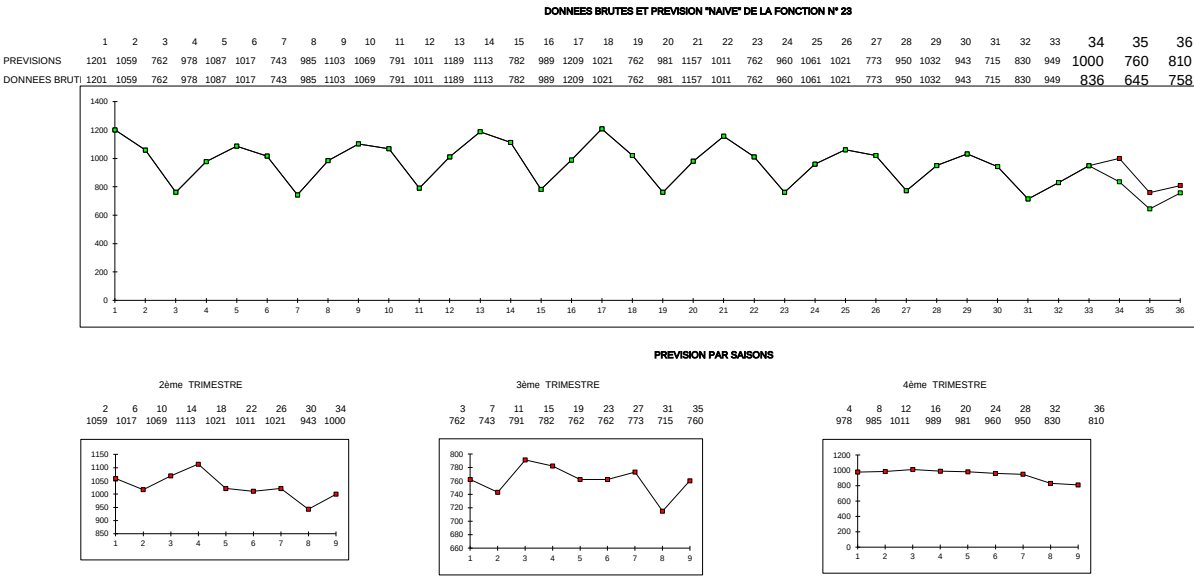
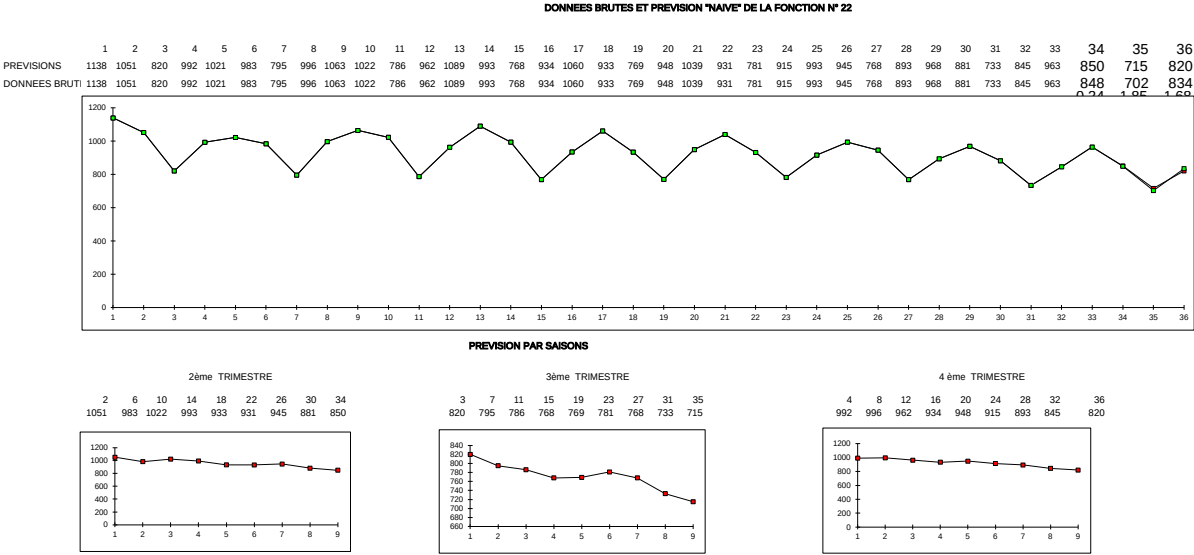


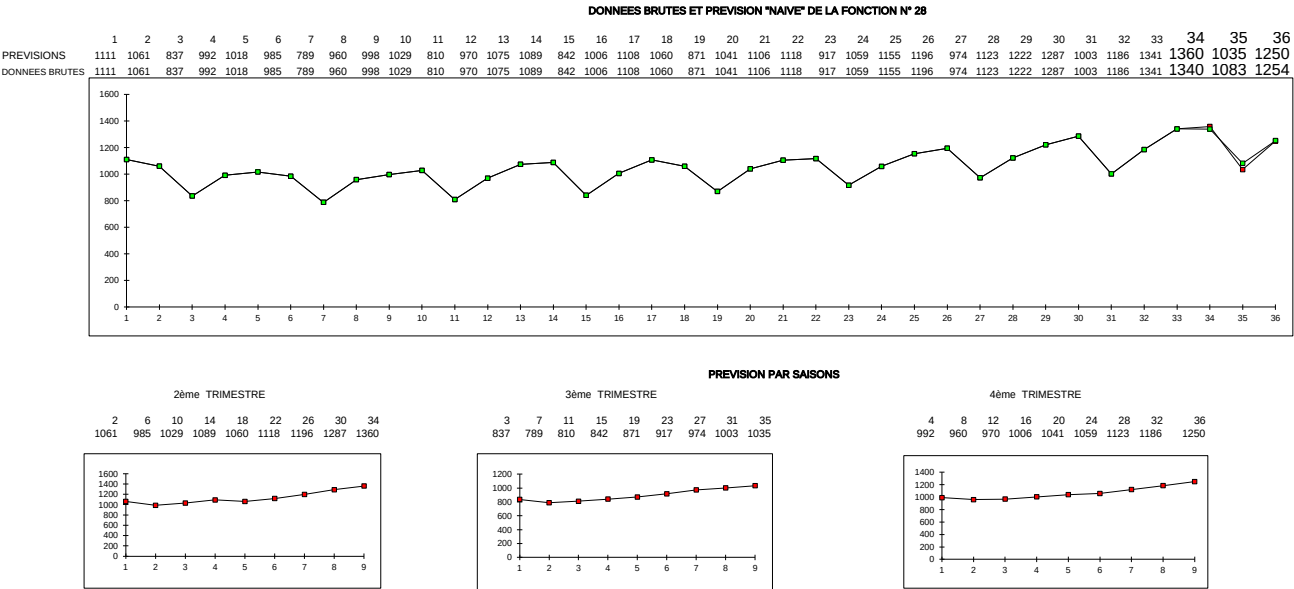
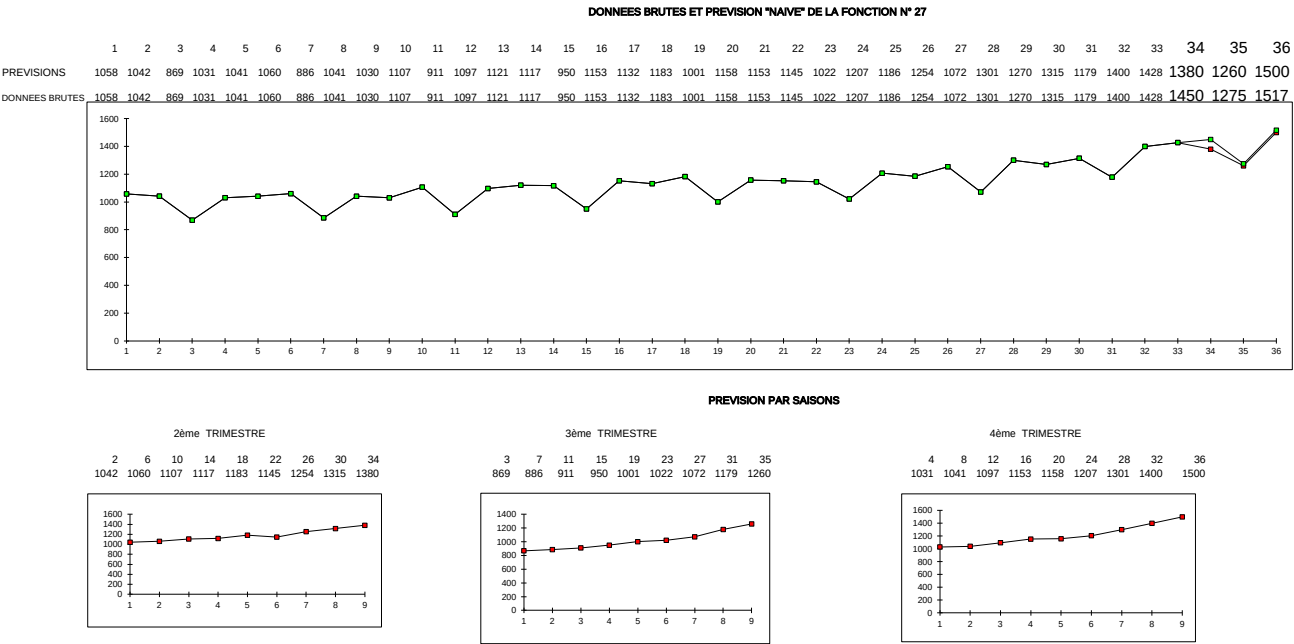
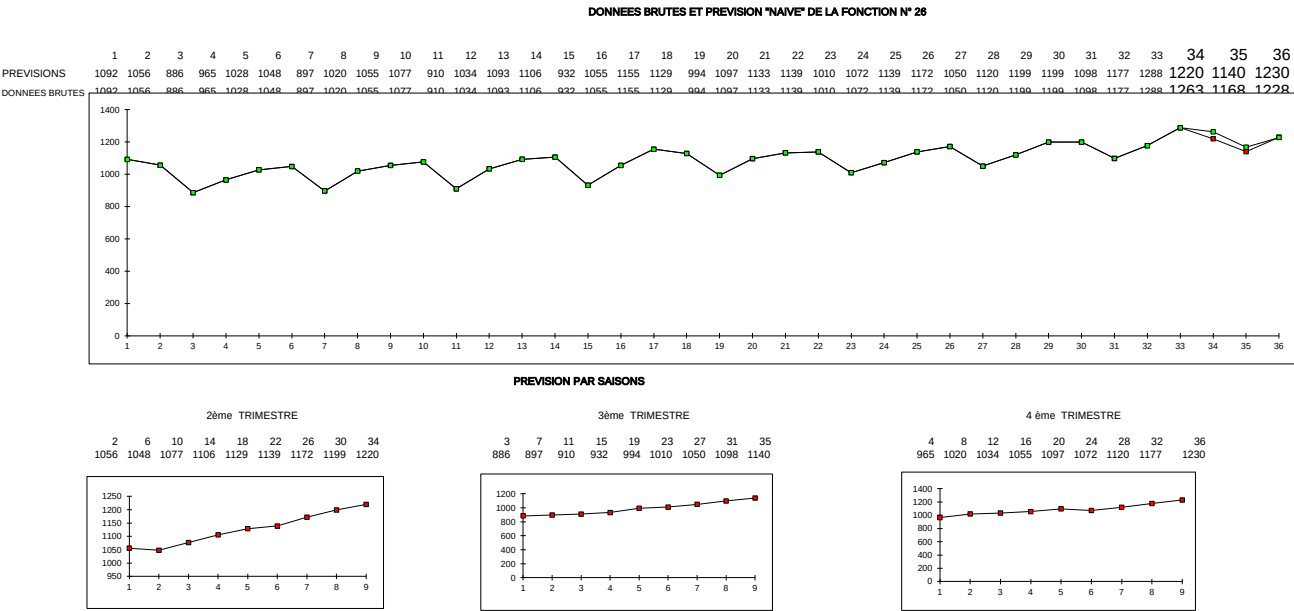
PREVISION PAR SAISONS











Ann-1-7) Comparaison des différentes prévisions et des données brutes :

Les tableaux qui suivent donnent un récapitulatif de tous les résultats obtenus lors des différentes prévisions. En fin du tableau sont donnés la moyenne, les écartypes et les erreurs moyennes des écarts, par colonne . L'indicateur « erreur moyenne » est défini comme le pourcentage de la moyenne des erreurs en fonction de la moyenne des données brutes.

- La première colonne (DB) reprend les données brutes du 34^{ème} trimestre (resp. 35^{ème}, 36^{ème}) (Ann I-2)

- La seconde colonne (D6T) donne la projection des données brutes sur l'espace engendré par les six premières composantes temporelles.

- La troisième colonne donne l'écart entre la première et la seconde colonne. Il est intéressant d'observer l'erreur moyenne du 34^{ème} trimestre (resp. 35^{ème}, 36^{ème}) qui est de 1,3 % (resp. 1,6 % , 1,2 %) . Cette erreur moyenne est un indicateur qui peut être interprété comme la meilleure précision attendue lors de la prévision. En effet puisque nous reconstituons les données dans l'espace engendré par les 6 premières composantes temporelles il est naturel de s'attendre à perdre l'information contenue dans l'espace engendré par les composantes temporelles n°7 à n° 33. Nous pouvons remarquer, au passage, que la notion d'indice fort permet de définir une attente de précision des résultats sans faire aucune hypothèse sur la loi du processus. Dans le cas de notre étude nous pouvons attendre une erreur moyenne supérieure à 1,3 %.

- La quatrième colonne (PB) donne la prévision obtenue après prolongement des 6 premières composantes temporelles et reconstitution des données. (Ann I-2)

- La cinquième colonne est la différence entre les données brutes et les prévisions brutes. L'erreur moyenne entre les données brutes et les prévisions brutes est de 3,34% pour le 34^{ème} trimestre; 3,9% pour le 35^{ème} trimestre; 3,2% pour le 36^{ème} trimestre, c'est à dire plus du double de la prévision minimale attendue.

- La sixième colonne donne les prévision après corrections (Ann I-5)

- La septième colonne est la différence entre les données brutes et les prévisions corrigées. Il est intéressant de comparer les erreurs moyennes des prévisions corrigées (3,13% pour le 34^{ème} trimestre; 3,92 % pour le 35^{ème} trimestre; 3.1% pour le 36^{ème} trimestre) avec celles des prévisions brutes. L'intérêt de cette correction semble diminuer avec la longueur de la prévision. Il ne s'agit cependant que d'empirisme sur peu de résultats.

- La huitième colonne donne les prévisions par prolongement graphique, trajectoire par trajectoire en fonction des résultats de l'annexe I-6.

- La neuvième colonne donne l'écart entre les données brutes et la prévision naïve. L'écart entre la prévision corrigé et la prévision naïve peut être considéré comme le gain de notre méthode par rapport à une méthode basée sur une prévision de chaque série chronologique . Pour le 34^{ème} trimestre (resp. 35^{ème}, 36^{ème}) le gain est de 1,87 % (resp. 0,26% , 0,2%).

COMPARAISON DES DIFFERENTES PREVISIONS ET DES DONNEES BRUTES

TRIMESTRE N°34		DB	D6T	DB-D6T	PB	DB-PB	PC1	DB-PC1	PN	DB-PN
		DONNEES BRUTES	DONNEES RECONSTITUEES SUR 6 COMPOSANTES	ERREUR	PREVISION BRUTE	ERREUR	PREV. CORRIGEEES	ERREUR	PREVISION NAIVE	ERREUR
1	Industrie des biens intermédiaires	1077	1071	6	1059	18	1061	16	1040	37
2	Biens d'équipement industriel	1016	1005	11	1012	4	1016	0	980	36
3	Biens d'équipement ménager	1208	1198	10	1145	63	1149	59	1030	178
4	Biens de consommation courantes	1089	1080	9	1091	2	1096	7	1065	24
5	Bâtiment, génie civil	1046	1056	10	1048	2	1048	2	1040	6
6	Vielandes et produits laitiers	1254	1253	1	1295	41	1305	51	1255	1
7	Autres produits agricoles	1007	990	17	949	58	955	52	972	35
8	Combustibles minéraux solides	635	651	16	718	83	716	81	740	105
9	Pétrole brut,gaz naturel	707	705	2	727	20	720	13	600	107
10	Electricité gaz et eau	1267	1262	5	1326	59	1284	17	1300	33
11	Minerais,métaux ferreux	856	826	30	842	14	851	5	800	56
12	Minerais, métaux non ferreux	1076	1044	32	1026	50	1032	44	1030	46
13	Matériaux de construction	977	982	5	967	10	979	2	935	42
14	Produits de l'industrie du verre	1046	1023	23	996	50	1003	43	1010	36
15	Produit chimiques de base	1184	1220	36	1219	35	1219	35	1210	26
16	Produit de la parapharmacie et de la pharmacie	1420	1376	44	1368	52	1355	65	1333	87
17	Produit de la fonderie et du travail des métaux	1007	988	19	963	44	964	43	935	72
18	Produit de la mécanique	966	951	15	928	38	940	26	910	56
19	Matériel électrique et électrotechnique professionnel	1139	1126	13	1176	37	1175	36	1120	19
20	Matériel électrique et électronique ménager	1218	1246	28	1277	59	1291	73	1250	32
21	Matériel de transport terrestre	1206	1190	16	1123	83	1120	86	1100	106
22	Construction navale et aéronautique	860	861	1	852	8	840	20	850	10
23	Produits textiles et habillement	848	850	2	893	45	891	43	850	2
24	Cuir et chaussures	836	831	5	886	50	886	50	1000	164
25	Bois et meubles	949	962	13	942	7	944	5	910	39
26	Papier et carton	1263	1269	6	1268	5	1257	6	1220	43
27	Imprimerie, presse	1450	1453	3	1410	40	1399	51	1380	70
28	Caoutchouc et matières plastiques	1340	1338	2	1316	24	1305	35	1360	20
MOYENNE		1070	1065	14	1065	35,75	1064	34,50	1044	53
ECARTYPE		200,8	200,0		192,0		188,6		194	
ERREUR MOYENNE				1,3%		3,34%		3,23%		5,0%

COMPARAISON DES DIFFERENTES PREVISIONS
ET DES DONNEES BRUTES

		DB	D6T	DB-D6T	PB	DB-PB	PC1	DB-PC1	PN	DB-PN
	TRIMESTE N°35	DONNEES BRUTES	DONNEES RECONSTITUEES SUR 6 COMPOSANTES	ERREUR	PREVISION BRUTE	ERREUR	PREVISION CORRIGEEES	ERREUR	PREVISION NAIVE	ERREUR
1	Industrie des biens intermédiaires	929	916	13	892	37	897	32	885	44
2	Biens d'équipement industriel	926	928	2	928	2	927	1	885	41
3	Biens d'équipement ménager	904	920	16	867	37	869	35	885	19
4	Biens de consommation courantes	945	949	4	941	4	948	3	940	5
5	âtiment, génie civil	902	913	11	885	17	884	18	920	18
6	Viandes et produits laitiers	1199	1240	41	1261	62	1264	65	1194	5
7	Autres produits agricoles	941	939	2	950	9	911	30	920	21
8	Combustibles minéraux solides	559	561	2	615	56	620	61	570	11
9	Pétrole brut,gaz naturel	712	723	11	725	13	725	13	670	42
10	Electricité gaz et eau	1184	1186	2	1129	55	1134	50	1110	74
11	Minerais,métaux ferreux	708	695	13	686	22	693	15	670	38
12	Minerais, métaux non ferreux	830	843	13	833	3	837	7	830	0
13	Matériaux de construction	844	836	8	817	27	815	29	900	56
14	Produits de l'industrie du verre	911	894	17	889	22	892	19	880	31
15	Produit chimiques de base	1149	1110	39	1092	57	1098	51	1120	29
16	Produit de la parapharmacie et de la pharmacie	1295	1292	3	1255	40	1262	33	1230	65
17	Produit de la fonderie et du travail des métaux	846	834	12	806	40	809	37	810	36
18	Produit de la mécanique	824	823	1	810	14	806	18	780	44
19	Matériel électrique et électrotechnique professionnel	1100	1111	11	1136	36	1135	35	1050	50
20	Matériel électrique et électronique ménager	1051	1050	1	1050	1	1040	11	1035	16
21	Matériel de transport terrestre	881	899	18	838	43	842	39	830	51
22	Construction navale et aéronautique	793	779	14	757	36	761	32	720	73
23	Produits textiles et habillement	702	717	15	744	42	755	53	715	13
24	Cuir et chaussures	645	649	4	703	58	719	74	760	115
25	Bois et meubles	823	819	4	799	24	798	25	820	3
26	Papier et carton	1168	1143	25	1112	56	1121	47	1140	28
27	Imprimerie, presse	1254	1270	16	1201	53	1203	51	1260	6
28	Caoutchouc et matières plastiques	1196	1089	107	1040	156	1051	145	1035	161
	MOYENNE	936	933	15	920	37	922	37	913	39
	ECARTYPE	195,4	191,7		178,2		177,8		180	35
	ERREUR MOYENNE			1,62%		3,90%		3,92%		4,18%

		DB	D6T	DB-D6T	PB	DB-PB	PC1	DB-PC1	PN	DB-PN
	TRIMESTE N°36	DONNEES BRUTES	DONNEES RECONSTITUEES SUR 6 COMPOSANTES	ERR E U R	PREVISION BRUTE	ERR E U R	PREV. CORRIGEEES 1ERE ESPECE	ERR E U R	PREVISION NAIVE	ERR E U R
1	Industrie des biens Intermédiaires	1065	1057	8	1049	16	1047	18	1020	45
2	Biens d'équipement industriel	1075	1073	2	1052	23	1049	26	1025	50
3	Biens d'équipement ménager	1208	1215	7	1193	15	1199	9	1165	43
4	Biens de consommation courantes	1089	1085	4	1089	0	1089	0	1025	64
5	Bâtiment, génie civil	1019	1046	27	1027	8	1020	1	1025	6
6	Viandes et produits laitiers	1209	1225	16	1244	35	1220	11	1210	1
7	Autres produits agricoles	1476	1476	0	1388	88	1393	83	1430	46
8	Combustibles minéraux solides	673	695	22	741	68	756	83	720	47
9	Pétrole brut, gaz naturel	759	755	4	788	29	792	33	680	79
10	Electricité gaz et eau	1622	1627	5	1570	52	1595	27	1580	42
11	Minerais, métaux ferreux	831	813	18	818	13	822	9	750	81
12	Minerais, métaux non ferreux	1064	1043	21	1024	40	1026	38	1010	54
13	Matériaux de construction	959	976	17	970	11	965	6	1000	41
14	Produits de l'industrie du verre	1014	1004	10	1010	4	1009	5	1035	21
15	Produit chimiques de base	1241	1211	30	1215	26	1205	36	1180	61
16	Produit de la parapharmacie et de la pharmacie	1420	1403	17	1365	55	1352	68	1330	90
17	Produit de la fonderie et du travail des métaux	1009	991	18	961	48	964	45	980	29
18	Produit de la mécanique	1050	1024	26	988	62	989	61	1050	0
19	Matériel électrique et électrotechnique professionnel	1172	1203	31	1215	43	1203	31	1190	18
20	Matériel électrique et électronique ménager	1471	1476	5	1456	15	1457	14	1450	21
21	Matériel de transport terrestre	1166	1174	8	1151	15	1158	8	1150	16
22	Construction navale et aéronautique	910	895	15	839	71	847	63	850	60
23	Produits textiles et habillement	834	833	1	878	44	888	54	820	14
24	Cuir et chaussures	758	767	9	846	88	860	102	810	52
25	Bois et meubles	961	974	13	962	1	963	2	950	11
26	Papier et carton	1228	1234	6	1233	5	1223	5	1230	2
27	Imprimerie, presse	1517	1501	16	1442	75	1431	86	1500	17
28	Caoutchouc et matières plastiques	1254	1278	24	1288	34	1284	30	1250	4
	MOYENNE	1109	1109	14	1100	35	1100	34	1086	36
	ECARTYPE	241,9	241,5		220,2		217,7		235	26
	ERREUR MOYENNE			1,2%		3,2%		3,1%		3,3%

ANNEXE II

LA CONSOMMATION D'ÉLECTRICITÉ EN FRANCE, PAR TRIMESTRE, DANS LES DIFFÉRENTS SECTEURS ÉCONOMIQUES DE 1972 À 1980

La seconde annexe présente un complément de résultats sur l'étude de la consommation d'électricité dans les différents secteurs économiques traitée dans les applications pratiques [V-4].

Nous nous contenterons dans cette annexe de tableaux de présentation des données et des tableaux de résultats comparatifs.

Vous trouverez:

Ann-II-1) Liste des secteurs concernés

Ann-II-2) Édition des sorties du logiciel

Nous donnerons les deux fichiers fournis par le logiciel: celui des données et des prévisions et celui des calculs des composantes temporelles de la moyenne et des valeurs propres

Ann-2-1) Consommation d'électricité, par trimestre : liste des variables traitées.

N° 1	MÉTALLURGIE GÉNÉRALE	N° 19	TEXTILES ARTIFICIELS
N° 2	MÉTAUX RARES	N° 20	CUIR
N° 3	ALUMINIUM	N° 21	BOIS ET AMEUBLEMENT
N° 4	FERO- ALLIAGES	N° 22	INDUSTRIE DU PAPIER
N° 5	ÉLECTRODES	N° 23	INDUSTRIES DIVERSES
N° 6	FONDERIES	N° 24	TRANSPORT ET COMMUNICATION
N° 7	AUTOMOBILE	N° 25	TRANSPORT FERROVIAIRE
N° 8	CONSTRUCTION ÉLECTRIQUE	N° 26	COMMERCE ET SERVICES
N° 9	PRÉCISION HORLOGERIE	N° 27	ADMINISTRATION
N° 10	CHAUX ET CEMENTS		
N° 11	BÂTIMENTS ET TP		
N° 12	MATÉRIAUX DE CONSTRUCTION		
N° 13	INDUSTRIE CHIMIQUE		
N° 14	ELECTRO-CHIMIE		
N° 15	INDUSTRIE DU VERRE		
N° 16	AGRICULTURE PISCICULTURE		
N° 17	INDUSTRIE AGRO-ALIMENTAIRES		
N° 18	INDUSTRIE TEXTILES		

NOMBRE DE FONCTIONS 27

NOMBRE DE DONNEES UTILISEES DANS LES CALCULS 19

NOMBRE DE DONNEES RECONSTITUEES 20

FONCTION N° 1

DONNEES BRUTES

2971	2188	2386	3029	3134	2070	2574	3263	3419	3352
2770	3560	3468	3068	2299	3000	3552	3298	2653	3411

DONNEES RECONSTITUEES ET PREVISION

2924	2297	2309	2996	3162	2075	2565	3329	3512	3363
2712	3361	3534	3093	2490	3279	3411	3126	2582	3711

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

2	-5	3	1	-1	0	0	-2	-3	0
2	6	-2	-1	-8	-9	4	5	3	-9

COEFFICIENT DE LA FONCTION SUR CHAQUE COMPOSANTE TEMPORELLE

-2438	-128	78	-208	365	-31	157	155	-396	47
-119	47	-133	10	65	-56	-35	13	15	

FONCTION N° 2

DONNEES BRUTES

2932	2581	2748	2962	2954	1908	2907	3084	3016	2954
2916	405	465	352	268	404	389	351	282	412

DONNEES RECONSTITUEES ET PREVISION

3025	2470	2795	3001	3068	1998	2877	3073	2923	2907
2993	618	391	288	216	470	426	313	217	-527

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

-3	4	-2	-1	-4	-5	1	0	3	2
-3	-53	16	18	20	-16	-9	11	23	228

COEFFICIENT DE LA FONCTION SUR CHAQUE COMPOSANTE TEMPORELLE

-8348	-4688	2339	761	621	-226	-256	30	92	-39
30	29	18	21	51	-24	0	-2	3	

FONCTION N° 3

DONNEES BRUTES

2968	2969	2912	2851	3010	2015	3054	3100	3041	3042
3107	3233	3176	3038	2984	3033	3046	2993	2994	3036

DONNEES RECONSTITUEES ET PREVISION

3033	2746	2969	3015	2907	2037	2984	3137	3004	3095
3113	3313	3059	3077	2966	3186	2960	2978	2960	3601

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

-2	8	-2	-6	3	-1	2	-1	1	-2
0	-2	4	-1	1	-5	3	1	1	-19

COEFFICIENT DE LA FONCTION SUR CHAQUE COMPOSANTE TEMPORELLE

-2046	-1228	399	-570	-443	-246	-47	168	73	207
49	-87	3	-99	-8	30	-3	12	-1	

FONCTION N° 4

DONNEES BRUTES

3265	4569	5399	4926	4283	3985	6269	5733	5719	6830
7411	8121	7635	9073	8731	7204	6416	8301	9447	7223

DONNEES RECONSTITUEES ET PREVISION

3341	4434	5400	4832	4342	4060	6261	5700	5612	6879
7563	8262	7530	8987	8854	7273	6538	8269	9235	7785

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

-2	3	0	2	-1	-2	0	1	2	-1
-2	-2	1	1	-1	-1	-2	0	2	-8

COEFFICIENT DE LA FONCTION SUR CHAQUE COMPOSANTE TEMPORELLE									
14355	2576	2269	-613	303	-44	-328	272	29	-112
17	20	15	15	33	-19	-2	-9	-14	
FONCTION N° 5									

DONNEES BRUTES

3031	2756	3047	3497	3749	2835	3598	4000	4197	4487
4276	4733	4559	4441	4221	3992	3929	4039	3779	4031

DONNEES RECONSTITUEES ET PREVISION

2965	2776	3090	3519	3563	2699	3617	4044	4270	4493
4197	4608	4588	4621	3952	3910	3825	4153	4012	4399

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

2	-1	-1	-1	5	5	-1	-1	-2	0
2	3	-1	-4	6	2	3	-3	-6	-9

COEFFICIENT DE LA FONCTION SUR CHAQUE COMPOSANTE TEMPORELLE

1611	339	1181	-423	886	132	281	-234	178	297
1	-61	-39	-1	-8	52	30	-2	10	

FONCTION N° 6

DONNEES BRUTES

2929	2066	1949	2923	3124	2007	2190	3221	3373	3152
2112	3459	3582	3054	2300	3422	3503	3158	2412	3642

DONNEES RECONSTITUEES ET PREVISION

2893	2137	1994	2849	3074	1937	2209	3164	3407	3146
2214	3369	3662	3059	2290	3388	3622	3154	2340	3992

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

1	-3	-2	3	2	3	-1	2	-1	0
-5	3	-2	0	0	1	-3	0	3	-10

COEFFICIENT DE LA FONCTION SUR CHAQUE COMPOSANTE TEMPORELLE

-2961	400	-406	-308	261	210	-84	-23	-53	22
-33	38	-12	36	-93	23	-50	-27	4	

FONCTION N° 7

DONNEES BRUTES

2886	1843	1945	3126	3210	2036	2139	3373	3567	3283
2338	3927	4144	3518	2705	4138	4194	3657	2765	4193

DONNEES RECONSTITUEES ET PREVISION

2870	2053	1885	2901	3185	1983	2196	3322	3651	3334
2283	3821	4236	3501	2697	3936	4245	3716	2879	4693

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

1	-11	3	7	1	3	-3	2	-2	-2
2	3	-2	0	0	5	-1	-2	-4	-12

COEFFICIENT DE LA FONCTION SUR CHAQUE COMPOSANTE TEMPORELLE

-1727	1197	-824	-264	384	346	125	-170	8	-143
-95	-29	13	13	49	-47	30	21	2	

FONCTION N° 8

DONNEES BRUTES

2855	2063	1937	2979	3182	2006	2188	3249	3302	3089
2125	3385	3391	3015	2370	3526	3552	3198	2469	3677

DONNEES RECONSTITUEES ET PREVISION

2920	2164	2032	2860	3073	1917	2225	3168	3366	3107
2217	3290	3563	2967	2307	3454	3686	3197	2429	3963

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

-2	-5	-5	4	3	4	-2	2	-2	-1
-4	3	-5	2	3	2	-4	0	2	-8

COEFFICIENT DE LA FONCTION SUR CHAQUE COMPOSANTE TEMPORELLE

-2927	309	-494	-172	141	289	-135	-128	-51	44
-89	4	-7	-63	-100	-18	-11	-48	-12	

FONCTION N° 9

DONNEES BRUTES

3208	2334	2083	3000	3146	2126	2417	3542	3708	3730
2688	3980	4063	3854	3000	4208	4334	3979	3062	4333

DONNEES RECONSTITUEES ET PREVISION

2981	2371	2223	3083	3279	2184	2521	3498	3725	3607
2547	3943	4213	3783	2964	4104	4353	4015	3116	4569

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

7	-2	-7	-3	-4	-3	-4	1	0	3
5	1	-4	2	1	2	0	-1	-2	-5

COEFFICIENT DE LA FONCTION SUR CHAQUE COMPOSANTE TEMPORELLE

-912	1203	-373	-71	176	23	80	-250	-114	-59
253	165	120	4	-19	14	-33	18	23	

FONCTION N° 10

DONNEES BRUTES

3666	3608	4123	3934	3785	3100	4256	4079	3900	4393
4430	4455	3587	4458	4154	4314	4207	4570	4542	4434

DONNEES RECONSTITUEES ET PREVISION

3387	3700	4024	3854	3585	2930	4255	4195	3876	4582
4410	4173	3641	4498	4299	4356	4086	4708	4491	3675

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

8	-3	2	2	5	5	0	-3	1	-4
0	6	-1	-1	-4	-1	3	-3	1	17

COEFFICIENT DE LA FONCTION SUR CHAQUE COMPOSANTE TEMPORELLE

2478	-381	1763	563	-582	300	375	264	-91	144
-59	140	75	-2	-73	-22	27	-6	8	

FONCTION N° 11

DONNEES BRUTES

3040	2307	2173	2533	2573	1587	2066	2653	2933	2520
2067	2974	3213	2360	2027	2919	3093	2520	2067	3106

DONNEES RECONSTITUEES ET PREVISION

2884	2154	2095	2618	2720	1638	2076	2728	2798	2507
2016	2952	3055	2498	2098	2975	3018	2531	2066	3766

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

5	7	4	-3	-6	-3	0	-3	5	1
2	1	5	-6	-4	-2	2	0	0	-21

COEFFICIENT DE LA FONCTION SUR CHAQUE COMPOSANTE TEMPORELLE

-4250	-611	-627	-715	-286	-182	144	187	30	-2
205	29	-19	173	75	27	5	-31	-2	

FONCTION N° 12

DONNEES BRUTES

3175	2948	2929	3464	3309	2433	3145	3712	3649	4083
3392	4227	3866	3980	3433	4403	4319	4144	3485	4525

DONNEES RECONSTITUEES ET PREVISION

3158	2937	2955	3381	3377	2477	3199	3745	3734	3965
3227	4020	3949	4071	3477	4204	4239	4291	3612	4196

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

1	0	-1	2	-2	-2	-2	-1	-2	3
5	5	-2	-2	-1	5	2	-4	-4	7

COEFFICIENT DE LA FONCTION SUR CHAQUE COMPOSANTE TEMPORELLE

360	588	491	165	-219	153	259	-164	16	-109
-----	-----	-----	-----	------	-----	-----	------	----	------

205	3	-170	-78	25	-36	20	-4	-43	
FONCTION N° 13									
DONNEES BRUTES									
3108	2936	3153	3260	3344	2565	3475	4068	3851	4250
3872	4434	4353	4623	4695	4721	5026	5322	4666	5033
DONNEES RECONSTITUEES ET PREVISION									
3184	3039	3226	3529	3471	2545	3524	3975	3905	4165
3780	4585	4464	4599	4405	4979	4923	5051	4829	5077
ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE									
-2	-4	-2	-8	-4	1	-1	2	-1	2
2	-3	-3	1	6	-5	2	5	-3	-1
COEFFICIENT DE LA FONCTION SUR CHAQUE COMPOSANTE TEMPORELLE									
2638	903	61	234	-370	-384	-75	-385	-214	229
-117	-51	76	99	55	-21	21	23	-21	

FONCTION N° 14

DONNEES BRUTES									
3222	3235	3895	3434	3218	2656	4042	3717	3542	4166
4182	3932	3527	4190	3960	4023	3500	3893	3528	3807
DONNEES RECONSTITUEES ET PREVISION									
3244	3447	3755	3559	3299	2675	3898	3784	3543	4113
4048	4013	3505	4199	3885	3754	3419	3968	3809	3610
ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE									
-1	-7	4	-4	-3	-1	4	-2	0	1
3	-2	1	0	2	7	2	-2	-8	5
COEFFICIENT DE LA FONCTION SUR CHAQUE COMPOSANTE TEMPORELLE									
883	-759	1643	-186	-475	-119	311	-281	213	-224
-110	-61	73	26	-13	-13	-42	-27	1	

FONCTION N° 15

DONNEES BRUTES									
2940	2795	2650	2796	3156	2132	2867	3699	3722	3916
3061	3747	3855	3843	2867	3759	4133	4193	3144	3710
DONNEES RECONSTITUEES ET PREVISION									
3075	2675	2617	3242	3334	2353	2886	3616	3716	3797
2899	3809	3886	3774	3034	3924	4061	3948	3132	4051
ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE									
-5	4	1	-16	-6	-10	-1	2	0	3
5	-2	-1	2	-6	-4	2	6	0	-9
COEFFICIENT DE LA FONCTION SUR CHAQUE COMPOSANTE TEMPORELLE									
-717	629	306	94	42	-545	-42	-26	-331	-112
157	-169	19	-64	-100	-2	24	-17	17	

FONCTION N° 16

DONNEES BRUTES									
2867	2800	6067	4266	3600	2067	6733	5067	4134	4000
10734	5801	4600	4066	10133	6800	4934	5867	13600	8133
DONNEES RECONSTITUEES ET PREVISION									
2819	2830	6102	4184	3576	2043	6801	5008	4179	4021
10710	5697	4664	4078	10150	6731	4999	5870	13606	7966
ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE									
2	-1	-1	2	1	1	-1	1	-1	-1
0	2	-1	0	0	1	-1	0	0	2
COEFFICIENT DE LA FONCTION SUR CHAQUE COMPOSANTE TEMPORELLE									
13983	-6214	-2134	145	103	192	16	-51	-65	-9
54	-20	7	1	-6	9	-6	-6	-13	

FONCTION N° 17

DONNEES BRUTES

2959	3028	3324	3331	3076	2200	3517	3489	3247	3765
3669	3745	3525	3779	4165	4055	3807	4021	4338	4241

DONNEES RECONSTITUEES ET PREVISION

3145	3034	3405	3310	3124	2260	3495	3541	3299	3535
3790	3905	3553	3748	3991	4121	3823	4000	4304	4336

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

-6	0	-2	1	-2	-3	1	-1	-2	6
-3	-4	-1	1	4	-2	0	1	1	-2

COEFFICIENT DE LA FONCTION SUR CHAQUE COMPOSANTE TEMPORELLE

743	-786	242	-185	-677	-155	-251	-179	85	16
-91	102	-191	-39	9	21	-7	6	31	

FONCTION N° 18

DONNEES BRUTES

3095	2555	2142	3162	3209	2162	2250	3359	3331	3135
1926	3337	3392	3068	2210	3494	3594	3379	2445	3703

DONNEES RECONSTITUEES ET PREVISION

3052	2465	2219	2969	3115	2066	2342	3229	3325	3250
2091	3231	3393	3097	2252	3529	3771	3386	2257	3675

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

1	4	-4	6	3	4	-4	4	0	-4
-9	3	0	-1	-2	-1	-5	0	8	1

COEFFICIENT DE LA FONCTION SUR CHAQUE COMPOSANTE TEMPORELLE

-2879	483	-106	73	-207	323	-232	194	-49	107
51	-14	-2	59	-42	16	-7	9	-29	

FONCTION N° 19

DONNEES BRUTES

2982	2436	2392	3312	3248	2249	2551	3491	3445	3422
2526	3473	3503	3495	2719	3760	3633	3712	2936	3931

DONNEES RECONSTITUEES ET PREVISION

3047	2546	2462	3096	3204	2174	2657	3411	3493	3475
2617	3508	3606	3373	2723	3710	3858	3616	2828	3897

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

-2	-5	-3	7	1	3	-4	2	-1	-2
-4	-1	-3	3	0	1	-6	3	4	1

COEFFICIENT DE LA FONCTION SUR CHAQUE COMPOSANTE TEMPORELLE

-1709	339	39	44	-73	289	-285	-54	-99	-75
-126	15	63	-21	30	43	21	-27	-13	

FONCTION N° 20

DONNEES BRUTES

3065	2419	2161	3258	3194	2032	2290	3291	3194	3000
2226	3419	3452	3162	2516	3871	3775	3581	2614	3968

DONNEES RECONSTITUEES ET PREVISION

3045	2448	2267	2951	3074	1999	2371	3210	3269	3161
2195	3347	3489	3145	2520	3751	3945	3539	2637	4008

ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE

1	-1	-5	9	4	2	-4	2	-2	-5
1	2	-1	1	0	3	-4	1	-1	-1

COEFFICIENT DE LA FONCTION SUR CHAQUE COMPOSANTE TEMPORELLE

-2377	464	-425	23	-343	407	-94	-27	-19	-2
-3	-100	70	-64	156	11	-25	-8	22	

FONCTION N° 21

DONNEES BRUTES									
3111	2639	2417	3333	3500	2223	2723	3694	3722	3416
2611	3833	4028	3444	2944	4305	4389	3973	3361	4750
DONNEES RECONSTITUEES ET PREVISION									
3067	2499	2406	3175	3335	2208	2674	3587	3693	3624
2689	3707	3902	3553	3003	4211	4446	4112	3339	4317
ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE									
1	5	0	5	5	1	2	3	1	-6
-3	3	3	-3	-2	2	-1	-3	1	9
COEFFICIENT DE LA FONCTION SUR CHAQUE COMPOSANTE TEMPORELLE									
-739	846	-362	373	-54	277	24	246	177	74
86	-164	15	5	-19	-57	-46	36	13	

FONCTION Nø 22

DONNEES BRUTES									
2978	2458	2383	3004	3026	2033	2543	3273	3248	3293
2566	3175	3310	3168	2544	3359	3263	3013	2573	3359
DONNEES RECONSTITUEES ET PREVISION									
2985	2461	2450	2972	3054	2044	2590	3212	3284	3218
2605	3287	3334	3062	2509	3272	3331	3099	2521	3644
ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE									
0	0	-3	1	-1	-1	-2	2	-1	2
-2	-4	-1	3	1	3	-2	-3	2	-8
COEFFICIENT DE LA FONCTION SUR CHAQUE COMPOSANTE TEMPORELLE									
-2608	-291	120	-278	-32	44	-152	-94	72	-87
43	84	38	-45	-40	23	14	86	-9	

FONCTION Nø 23

DONNEES BRUTES									
3241	2885	2701	3760	3943	2817	3212	4452	4635	4913
3231	4788	5192	4924	4115	5827	6211	6067	4778	6847
DONNEES RECONSTITUEES ET PREVISION									
3260	2874	2659	3746	3971	2823	3224	4482	4661	4834
3245	4793	5122	4998	4124	5781	6248	6062	4781	5315
ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE									
-1	0	2	0	-1	0	0	-1	-1	2
0	0	1	-1	0	1	-1	0	0	22
COEFFICIENT DE LA FONCTION SUR CHAQUE COMPOSANTE TEMPORELLE									
3409	2978	-270	1369	100	-9	4	-1	46	-59
18	-24	-93	55	11	84	-10	-4	5	

FONCTION Nø 24

DONNEES BRUTES									
2778	2148	2334	2852	3074	1815	2593	3407	3592	3222
2667	3592	4110	3259	3444	4333	4481	3888	3925	4704
DONNEES RECONSTITUEES ET PREVISION									
2915	2130	2247	2959	3142	1857	2497	3368	3492	3147
2834	3763	4026	3267	3310	4337	4472	3949	3925	5005
ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE									
-5	1	4	-4	-2	-2	4	1	3	2
-6	-5	2	0	4	0	0	-2	0	-6
COEFFICIENT DE LA FONCTION SUR CHAQUE COMPOSANTE TEMPORELLE									
-580	408	-1353	-2	-27	-220	-145	-24	225	-31
-83	-14	-41	109	-105	-63	14	8	9	

FONCTION Nø 25

DONNEES BRUTES									
2898	2051	2617	2790	3013	1824	2550	2961	3182	2808
2772	3027	3288	2526	2153	3181	3334	2828	2820	3225
DONNEES RECONSTITUEES ET PREVISION									
2893	2159	2312	2847	2982	1797	2464	3109	3182	2893
2718	3011	3134	2521	2465	3193	3233	2813	2776	3716
ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE									
0	-5	12	-2	1	2	3	-5	0	-3
2	1	5	0	-14	0	3	1	2	-15
COEFFICIENT DE LA FONCTION SUR CHAQUE COMPOSANTE TEMPORELLE									
-2988	-921	-486	-212	90	-93	271	316	-61	-222
-196	-48	3	7	-17	68	37	28	1	
FONCTION N° 26									

DONNEES BRUTES									
2872	2231	2364	2826	3352	2012	3029	3521	4128	3707
3401	4211	5037	4075	4087	4933	5632	4867	4909	5822
DONNEES RECONSTITUEES ET PREVISION									
2910	2133	2314	3180	3419	2071	2764	3784	4010	3673
3330	4485	4861	4029	4095	5139	5336	4862	4951	5870
ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE									
-1	4	2	-13	-2	-3	9	-7	3	1
2	-7	3	1	0	-4	5	0	-1	-1
COEFFICIENT DE LA FONCTION SUR CHAQUE TENDANCE									
1835	1344	-1586	202	285	-659	158	151	146	66
-99	97	58	-101	42	1	-54	-9	-26	
FONCTION N° 27									

DONNEES BRUTES									
2892	2090	1767	2848	3153	1745	2013	3153	3592	2983
2023	3363	3955	3063	2614	4126	4534	3668	3090	4695
DONNEES RECONSTITUEES ET PREVISION									
2910	1959	1785	2827	3130	1789	2019	3229	3460	3021
2056	3465	3915	3013	2580	4141	4504	3755	3049	4707
ECART ENTRE DONNEES ET PREVISION EN POURCENTAGE									
-1	6	-1	1	1	-2	0	-2	4	-1
-2	-3	1	2	1	0	1	-2	1	0
COEFFICIENT DE LA FONCTION SUR CHAQUE COMPOSANTE TEMPORELLE									
-2090	999	-1485	160	31	-73	-77	109	154	32
52	70	41	-55	43	-42	86	-34	20	

FONCTION MOYENNE

3035	2627	2815	3239	3280	2246	3081	3617	3644	3663
3374	3864	3862	3663	3469	4041	4029	3944	3803	

FONCTION VARIANCE

31839	313766	1000399	250860	115082	235288	1274086	402155	318030	709683
3281848	1489997	1230175	1880335	3721564	1533281	1234092	1876482	6004998	

COMPOSANTE TEMPORELLE N° 1

VALEUR PROPRE CORRESPONDANTE

21892928

VALEURS

0,01	0,078	0,182	0,092	0,056	0,069	0,218	0,127	0,1	0,144
0,346	0,241	0,191	0,243	0,407	0,243	0,177	0,266	0,5	

VARIANCE RESIDUELLE

29430	179089	276545	67238	46409	131786	234596	49031	97949	255160
668477	219818	429836	583708	100745	244069	549080	324432	523637	

COMPOSANTE TEMPORELLE N° 2

VALEUR PROPRE CORRESPONDANTE

3308127

VALEURS

0,026	0,018	-0,222	0,001	0,056	0,092	-0,203	0,026	0,119	0,152
-0,437	0,225	0,336	0,367	-0,135	0,202	0,35	0,3	-0,313	

VARIANCE RESIDUELLE

27271	177959	113869	67237	35957	103578	98149	46823	50952	178766
37208	51727	56698	137114	40039	108938	142948	27649	200026	

COMPOSANTE TEMPORELLE N° 3

VALEUR PROPRE CORRESPONDANTE

1176294

VALEURS

0,095	0,348	0,281	0,168	0,091	0,285	0,278	0,133	0,096	0,361
0,113	0,039	-0,108	0,301	-0,095	-0,257	-0,296	-0,024	-0,399	

VARIANCE RESIDUELLE

16709	35637	20952	34179	26263	8074	7416	26142	40061	25880
22162	49971	43076	30193	29473	31188	39780	26962	12495	

COMPOSANTE TEMPORELLE N° 4

VALEUR PROPRE CORRESPONDANTE

182223

VALEURS

0,115	0,044	-0,006	0,174	0,247	0,104	0,062	0,271	0,22	0,219
-0,005	-0,409	-0,332	-0,365	-0,248	0,193	0,368	0,243	0,069	

VARIANCE RESIDUELLE

14301	35277	20946	28686	15179	6087	6711	12727	31274	17125
22157	19478	23018	5909	18260	24384	15068	16174	11627	

COMPOSANTE TEMPORELLE N° 5

VALEUR PROPRE CORRESPONDANTE

127002

VALEURS

-0,179	-0,424	-0,313	0,008	0,192	0,021	-0,054	0,194	0,455	0,24
0,314	0,069	0,328	-0,078	-0,134	-0,231	-0,115	-0,214	0,01	

VARIANCE RESIDUELLE

10245	12440	8521	28677	10494	6032	6337	7929	4981	9822
9620	18882	9337	5134	15982	17580	13389	10373	11614	

COMPOSANTE TEMPORELLE N° 6

VALEUR PROPRE CORRESPONDANTE

73151

VALEURS

	0,162	-0,155	-0,081	0,581	0,259	0,205	-0,182	0,147	-0,129	-0,203
	-0,096	0,372	-0,165	-0,045	-0,036	0,306	-0,284	-0,169	0,056	
VARIANCE RESIDUELLE										
	8319	10693	8043	3988	5583	2951	3925	6349	3759	6817
	8949	8772	7350	4983	15887	10731	7475	8283	11382	
COMPOSANTE TEMPORELLE N°				7						
VALEUR PROPRE CORRESPONDANTE										
37537										
VALEURS										
	0,309	-0,194	0,271	-0,171	0,073	0,094	0,119	-0,266	-0,054	-0,066
	0,369	0,347	0,027	-0,21	-0,146	0,139	0,406	-0,185	-0,343	
VARIANCE RESIDUELLE										
	4744	9281	5280	2894	5380	2621	3390	3696	3650	6652
	3840	4248	7323	3335	15083	10009	1297	6998	6978	
COMPOSANTE TEMPORELLE N°				8						
VALEUR PROPRE CORRESPONDANTE										
34475										
VALEURS										
	0,103	0,301	0,213	0,1	0,154	0,004	0,127	-0,147	0,207	-0,322
	-0,181	0,076	0,339	-0,043	-0,503	-0,249	0,043	-0,064	0,398	
VARIANCE RESIDUELLE										
	4382	6157	3710	2551	4558	2621	2836	2948	2175	3077
	2708	4048	3350	3271	6360	7873	1233	6858	1509	
COMPOSANTE TEMPORELLE N°				9						
VALEUR PROPRE CORRESPONDANTE										
21994										
VALEURS										
	-0,099	0,186	0,029	0,106	0,131	0,026	0,241	-0,14	0,203	-0,017
	-0,196	-0,32	0,326	-0,116	0,339	0,403	0,009	-0,489	-0,173	
VARIANCE RESIDUELLE										
	4165	5398	3692	2303	4182	2606	1560	2518	1269	3071
	1860	1802	1020	2976	3833	4299	1231	1592	854	
COMPOSANTE TEMPORELLE N°				10						
VALEUR PROPRE CORRESPONDANTE										
15228										
VALEURS										
	0,132	0,243	-0,307	-0,109	0,378	0,262	-0,031	0,008	-0,012	-0,168
	-0,092	0,148	-0,058	-0,233	0,477	-0,472	0,187	0,039	-0,043	
VARIANCE RESIDUELLE										
	3901	4501	2253	2122	2010	1560	1545	2517	1266	2640
	1732	1468	969	2150	367	910	698	1570	826	
COMPOSANTE TEMPORELLE N°				11						
VALEUR PROPRE CORRESPONDANTE										
12407										
VALEURS										
	0,308	0,571	-0,313	-0,123	-0,236	-0,281	-0,255	0,185	0,122	0,145
	0,221	0,201	0,004	-0,142	-0,082	0,193	-0,072	-0,192	0,022	
VARIANCE RESIDUELLE										
	2726	451	1041	1933	1321	578	735	2094	1082	2378
	1124	968	969	1898	283	448	635	1114	820	0
COMPOSANTE TEMPORELLE N°				12						
VALEUR PROPRE CORRESPONDANTE										
6377										
VALEURS										
	0,438	-0,177	-0,17	0,145	-0,194	0,104	-0,022	-0,419	0,136	0,325
	-0,15	-0,111	-0,228	0,268	0,064	-0,152	0,148	-0,291	0,294	
VARIANCE RESIDUELLE										
	1503	252	856	1798	1081	509	732	976	964	1704
	979	890	639	1438	256	300	496	575	270	
COMPOSANTE TEMPORELLE N°				13						
VALEUR PROPRE CORRESPONDANTE										
5311										
VALEURS										

	0,412	-0,109	-0,085	-0,266	0,081	0,122	-0,016	0,176	0,257	-0,524
	0,181	-0,326	-0,04	0,392	0,014	0,134	-0,151	0,074	-0,082	
VARIANCE RESIDUELLE										
	602	189	818	1424	1047	430	731	813	612	246
	806	324	630	622	256	206	375	546	234	
COMPOSANTE TEMPORELLE N°				14						
VALEUR PROPRE CORRESPONDANTE										
3941										
VALEURS										
	0,332	-0,189	0,374	0,013	-0,426	0,139	-0,146	0,248	0,238	0,036
	-0,272	0,043	0,227	-0,37	0,231	-0,108	-0,147	0,138	-0,042	
VARIANCE RESIDUELLE										
	166	48	266	1423	332	354	647	570	388	241
	513	317	427	82	44	160	290	471	227	
COMPOSANTE TEMPORELLE N°				15						
VALEUR PROPRE CORRESPONDANTE										
3629										
VALEURS										
	0,054	0,079	-0,148	0,61	-0,216	-0,118	-0,097	-0,262	-0,021	-0,179
	0,362	-0,239	0,241	-0,042	0,071	-0,161	0,092	0,286	-0,237	
VARIANCE RESIDUELLE										
	156	26	186	72	163	304	613	321	386	125
	38	110	216	76	26	66	259	174	23	
COMPOSANTE TEMPORELLE N°				16						
VALEUR PROPRE CORRESPONDANTE										
1402										
VALEURS										
	0,165	0,041	-0,025	-0,218	0,25	0,33	-0,257	-0,335	-0,292	0,283
	0,116	-0,21	0,306	-0,167	-0,085	0,17	-0,357	0,223	0,12	
VARIANCE RESIDUELLE										
	117	24	185	5	76	151	520	163	267	13
	19	48	85	37	16	25	80	104	3	
COMPOSANTE TEMPORELLE N°				17						
VALEUR PROPRE CORRESPONDANTE										
969										
VALEURS										
	-0,268	0,132	0,414	0,013	0,119	0,087	-0,725	-0,089	0,241	-0,031
	0,126	-0,073	-0,146	0,145	0,098	-0,04	0,152	-0,163	0,012	
VARIANCE RESIDUELLE										
	48	7	19	5	62	143	11	155	211	12
	4	42	64	17	7	24	58	78	3	
COMPOSANTE TEMPORELLE N°				18						
VALEUR PROPRE CORRESPONDANTE										
682										
VALEURS										
	0,179	-0,046	0,089	0,072	0,04	-0,014	-0,121	0,474	-0,553	0,075
	0,044	-0,248	0,301	0,157	-0,033	-0,177	0,279	-0,328	0,062	
VARIANCE RESIDUELLE										
	26	5	14	1	61	143	1	2	2	8
	3	0	2	0	6	2	4	5	0	
COMPOSANTE TEMPORELLE N°				19						
VALEUR PROPRE CORRESPONDANTE										
280										
VALEURS										
	0,292	-0,112	0,209	0,036	0,475	-0,72	-0,048	-0,069	0,039	0,147
	-0,113	0,001	0,009	0,039	0,148	-0,077	-0,136	0,122	-0,083	
VARIANCE RESIDUELLE										
	2	2	1	1	-2	-2	1	0	2	2
	-1	0	2	-1	0	1	-1	1	-2	

REMARQUE: à partir de ce rang les valeurs propres sont nulle ainsi que la variance résiduelle

ANNEXE III

APPLICATION DE L'ANALYSE HARMONIQUE AU CHOIX DU NOMBRE D'AXE EN ANALYSE EN COMPOSANTES PRINCIPALES

Ann III-1) Rappel sur la méthode du coude en ACP :

L'analyse d'un questionnaire à l'aide d'une Analyse en Composantes Principales sur p variables (les questions) sélectionnées fait intervenir un espace de dimension p . La complexité des données oblige l'analyste à réduire l'information pour mieux la maîtriser.

L'idée à l'origine des techniques de l'ACP est de faire l'hypothèse que le nuage de points constitué par les réponses est dans sa grande majorité situé dans un espace de dimension relativement réduite.

Les calculs d'inerties montrent que le nuage de points est concentré sur les axes engendrés par les premiers vecteurs propres de la matrice de covariance; or l'inertie de la projection du nuage sur un vecteur propre est égale à la valeur propre associée à ce vecteur propre.

D'où la stratégie pratique suivante :

- a) classer les valeurs propres par ordre décroissant
- b) isoler les m premières valeurs propres contenant l'information (inertie) significative
- c) faire l'analyse dans l'espace engendré par les m premiers vecteurs propres.

Si on représente les valeurs propres ordonnées sur un diagramme en bâton, le passage des valeurs propres significatives aux valeurs propres à rejeter forme un coude; d'où le nom de la méthode.

Lorsque les valeurs propres ne sont pas naturellement séparables en deux groupes l'analyse devient difficile.

C'est le cas de l'exemple traité ci-après.

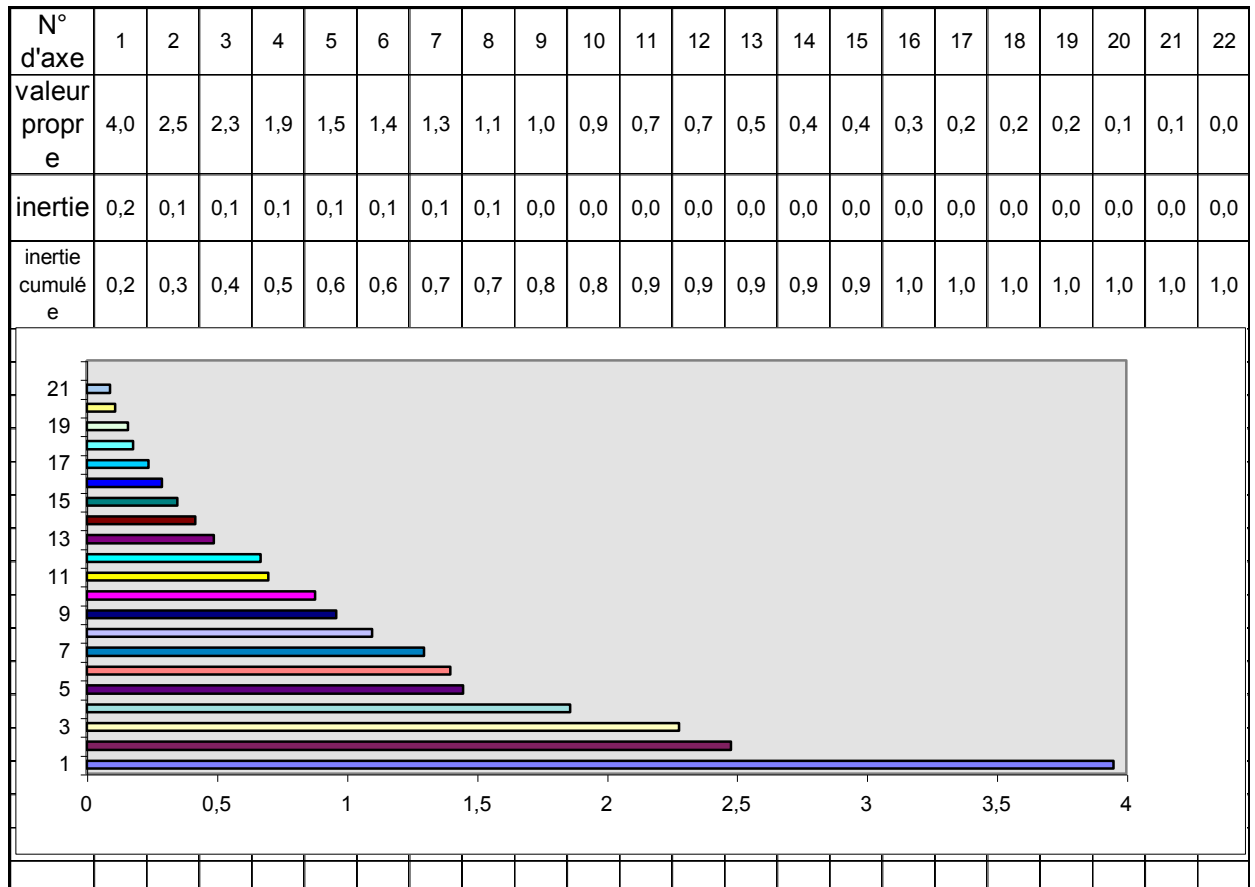
Ann III-2) Un exemple illustrant les difficultés pratiques liées à la méthode du coude

Ann III-2-a) Présentation des données

Les données que nous avons utilisées pour illustrer la difficulté liée au choix du nombre d'axes significatifs en Analyse en Composantes Principales (A.C.P.) sont issues d'une enquête commandée en 1994 à l'institut BVA par la presse. Cette enquête intitulée « Les Français sont-ils tolérants » comporte 22 questions [page 153] pouvant être traitées par une A.C.P..

Ann III-2-b) Traitement des données

Les données ont été centrées et réduites. La matrice de covariance ainsi obtenue a été diagonalisée, donnant naissance à 22 valeurs propres que nous présentons dans le diagramme ci-dessous



Ann III-2-c) Les difficultés de l'analyse

Pour faire une analyse correcte de ces données il nous faut déterminer les valeurs propres qui participent fortement à la description de l'information et celle qui véhiculent une information négligeable. Il faut donc faire une séparation entre deux groupes de valeurs. Malheureusement cet exercice est périlleux la séparation n'étant pas évidente.

Faut-il prendre 1,4,7 axes ou plus ?

Ann III-2-d) Utilisation des vecteurs propres de la matrice de covariance

Nous allons essayer d'éclaircir cette situation en faisant intervenir en plus des valeurs propres, les vecteurs propres.

Dans le cas des processus aléatoires nous avons utilisé la formule :

$$V(X) = \sum_{i \in N} \lambda_i f_i^2$$

Cette formule reste valable dans le cas d'une analyse en composante principale et donne à ce moment des renseignements sur la variance de chaque variable.

Si $\{X(\omega, i)\}_{i \leq p}$ représente le tableau X des données à p colonnes (pour plus de précision se reporter à l'annexe III-3-b) alors la variance de chaque variable s'écrit :

$$V(X(i)) = \sum_{j \leq p} \lambda_j f_j^2(i)$$

où $f_j(i)$ représente la valeur i du vecteur propre de rang j . Plus précisément on a la variance de la projection du tableau des données sur le vecteur f_j s'écrit :

$$V\left(\left\langle X, f_j \right\rangle\right) = \lambda_j f_j^2$$

Il ne s'agit que d'une transcription des résultats de l'analyse harmonique qu'il ne faut pas interpréter trop rapidement sans faire attention au sens des notations ci-dessus.

Les processus élémentaires ne sont pas ici les $X(i)$, comme une lecture hâtive pourrait le laisser croire, mais les $Y_j(i) = \left\langle X, f_j \right\rangle f_j(i)$; si nous désirions refaire la démonstration il faudrait utiliser la base canonique $\{e_i\}$ de l'espace R^p engendré par les p variables du tableau pour la différencier sans risquer d'ambiguïté avec la base des vecteurs propres $\{f_j\}$.

Proposition pratique :

Nous proposons de représenter par étapes successives les variances des projections du tableau X sur les vecteurs f_j et simultanément les valeurs cumulées; c'est à dire par valeur décroissantes de valeurs propres la quantité $\lambda_j f_j^2$ et la quantité cumulative $\sum_{h \leq j} \lambda_h f_h^2$ ce que nous avons fait dans les pages suivantes pour l'exemple traité.

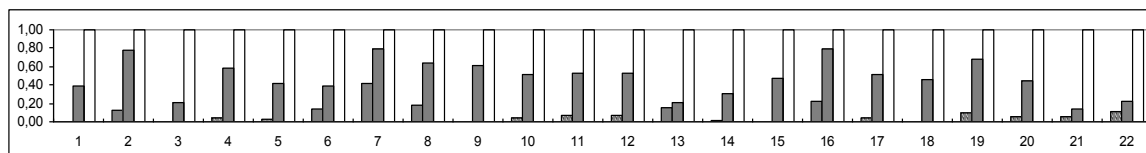
Ann-III-3) Proposition pour améliorer la méthode du coude

Ann-III-3-a) Intervention de la variance des vecteurs propres dans l'analyse de l'enquête sur la tolérances

Nous avons représenté dans le diagramme ci-dessous la variance associée au quatrième vecteur propre et le cumul des quatre premiers.

Il apparaît qu'à ce stade de l'analyse les questions N° 3, N°13, N°21 et N°22 ont une variance cumulée faible, comme le montre le graphique suivant.

COMPOSANTE PRINCIPALE N°4																						
VALEUR PROPRE CORRESPONDANTE				1,86																		
FONCTION PROPRE	0,04	-0,26	0,01	0,15	0,13	-0,27	0,47	-0,31	0,04	0,16	0,20	-0,19	-0,29	0,08	0,00	-0,35	0,14	-0,06	0,22	0,17	0,17	
PART DE VARIANCE	0,00	0,12	0,00	0,04	0,03	0,13	0,41	0,18	0,00	0,05	0,07	0,07	0,16	0,01	0,00	0,22	0,04	0,01	0,09	0,06	0,06	
VARIANCE CUMULEE	0,39	0,77	0,21	0,58	0,42	0,39	0,79	0,64	0,61	0,51	0,53	0,53	0,21	0,30	0,47	0,79	0,51	0,45	0,68	0,44	0,14	
VARIANCE totale	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	



Les nombres en abscisse représentent les questions et en ordonnée respectivement, la part de variance de la composante (à gauche), la variance cumulée (au centre) et la variance totale (à droite)

On est amené à penser que ces quatre questions N° 3 ,N°13 N°21 et N°22 sont des questions « étrangères » aux autres. Cela nous conduira à séparer ces quatre question du questionnaire et à réaliser une nouvelle ACP sur les 18 questions restantes. Nous allons apporter quelques éléments théoriques pour justifier cette pratique.

Ann-III-3-b) Interprétation des résultats ci-dessus

Rappelons que les réponses aux 22 questions, c'est à dire les données de l'ACP se présentent sous la forme d'un tableau $\{X(\omega, i)\}$ où ω parcourt l'ensemble des individus Ω et i

l'ensemble Q des p premiers entiers correspondant aux p questions (ici $p=22$). Si on identifie l'ensemble des réponses d'un individu à un élément de R^{22} et si $\{e_i\}_{i \in \{1, \dots, p\}}$ représente la base canonique de R^{22} alors le tableau des données peut aussi s'écrire : $\{ \langle X(\omega), e_i \rangle \mid \omega \in \Omega \text{ et } i \in Q \}$ et où $X(\omega)$ représente la ligne des p réponses de l'individu ω . De se fait nous pourrions dire que $X(\omega)(i) = X(\omega, i) = \langle X(\omega), e_i \rangle = e_i$.

Les vecteurs propres de la matrice de covariance, éventuellement complétés par une base du noyau de covariance, forment une base de R^{22} . Aux quatre premières valeurs propres $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$ nous pouvons associer quatre vecteurs propres normés $\{f_1, f_2, f_3, f_4\}$

Considérons le projecteur orthogonal P de R^{22} sur le sous espace vectoriel de dimension quatre $[f_1, f_2, f_3, f_4]$. Tout $X(\omega)$ peut alors s'écrire :

$$X(\omega) = P(X(\omega)) + R(X(\omega))$$

et nous savons alors que :

$$V(X) = V(P(X)) + V(R(X)) = \sum_1^4 \lambda_i f_i^2 + \sum_5^{22} \lambda_i f_i^2$$

Dire que la variance cumulée de la question N°21 représente 14% de variance cumulée s'écrit :

$$V(P(X))(21) = 14\% .$$

On peut comparer cette valeur à la participation moyenne sur l'ensemble des questions. Pour cela montrons en premier que la somme des valeurs propres est égale au nombre de questions dans la mesure où chaque question est normalisée (d'écartype égal à 1).

En effet pour tout i on a :

$$\sum_{j=1}^{22} \lambda_j f_j^2(i) = V(X)(i) = 1(i) = 1 \text{ donc } \sum_{i=1}^{22} \sum_{j=1}^{22} \lambda_j f_j^2(i) = 22 = \sum_{j=1}^{22} \lambda_j \sum_{i=1}^{22} f_j^2(i) = \sum_{j=1}^{22} \lambda_j$$

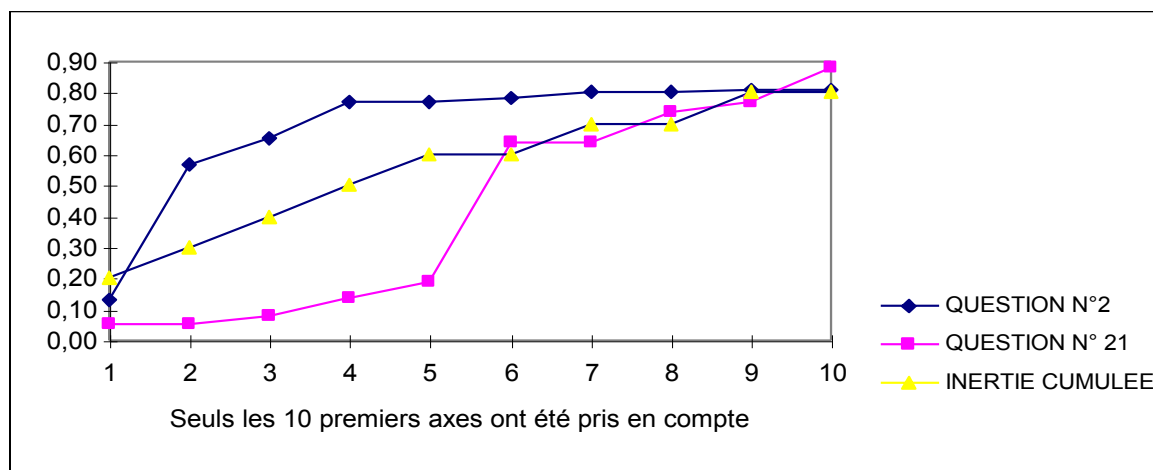
Calculons maintenant la moyenne des variances cumulées sur les quatre premiers vecteurs propres

$$\frac{1}{22} \sum_{i=1}^{22} \sum_{j=1}^4 \lambda_j f_j^2(i) = \sum_{j=1}^4 \frac{\lambda_j}{22} \sum_{i=1}^{22} f_j^2(i) = \sum_{j=1}^4 \frac{\lambda_j}{\sum_{k=1}^{22} \lambda_k} \sum_{i=1}^{22} f_j^2(i) = \frac{\sum_{j=1}^4 \lambda_j}{\sum_{k=1}^{22} \lambda_k} = 48\%$$

Cette moyenne est égale à l'inertie cumulée que nous avons déjà rencontrée et calculée (arrondie à 0,5) dans le tableau de l'annexe III-2-b .

Ann-III-3-c) Représentation de la variance cumulée des question N°2 ET N°21 ainsi que de la moyenne des variances cumulées de chaque question en fonction du nombre d'axes de l'ACP

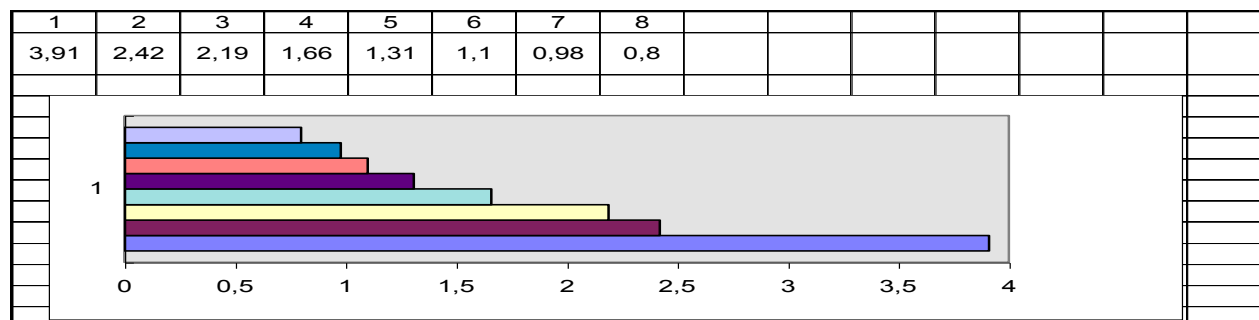
Pour permettre de mieux réaliser ce qui se passe dans le détail pour chaque question au fur et à mesure que l'on augmente le nombre d'axes dans une ACP nous proposons de représenter sur un graphique deux questions, dont le comportement est différent, et la moyenne qui n'est autre que l'inertie cumulée.



Comme nous l'avons annoncé cela nous amène à proposer la suppression des questions 3,13,21 et 22 et de faire une ACP sur les 18 questions restantes.

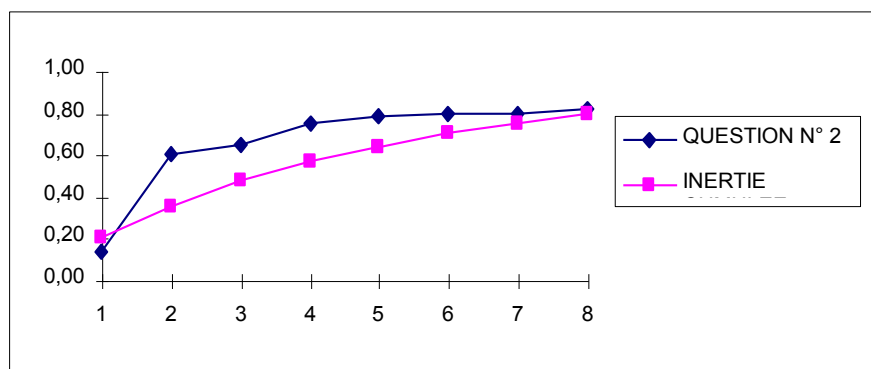
Ann-III-4-c) Reprise de l'analyse en composantes principales après avoir enlevé les question n° 3,13,21,22.

L'Analyse en Composantes Principales faite sur les 18 dernières questions donne le diagramme des valeurs propres suivantes (seules les 8 premières valeurs propres sont représentées)



Le choix de trois axes semble plus aisé que lors de la première analyse mais le doute peut encore rester ; c'est pour quoi, dans les annexes, nous avons représenté les graphiques des parts de variances; dès la composante numéro 3 chaque question participe au moins à 30% de sa propre variance alors que lors de la première analyse cette situation ne se rencontrait qu'à partir de l'axe numéro 6.

Nous représentons ci-dessous la variance cumulée de la question N°2 et la moyenne des variances cumulées en fonction du nombre d'axes principaux (nous nous sommes limité à 8 sur un total de 18, pour des raisons de commodité graphiques). La question N°2 est toujours une question dominante mais plus proche de la fonction moyenne.



Cet outil s'avère un « plus », relativement commode permettant de préciser une stratégie d'analyse; il ne répond pas aux deux questions fondamentales: combien d'axes faut-il prendre et faut-il éliminer des questions? Mais de telles interrogations ont-elle une réponse claire? nous pouvons en douter !!!

Ann-III-4) Enquête sur la tolérance des français : détails de l'étude

Dans ce paragraphe nous donnons les compléments de résultats sur l'enquête qui ne figuraient pas dans les paragraphes précédents et la liste des questions. Plus précisément:

- Au paragraphe Ann III-4-a) la liste des questions
- Au paragraphe Ann III-4-b) la variances et les variances cumulées sur chaque composante pour l'ensemble des questions
- Au paragraphe Ann III-4-c) la variances et les variances cumulées sur chaque composante une fois les questions numéro 3, 13, 21 et 22 enlevées de l'analyse.

Ann-III-4-a) Enquête sur la tolérance des Français

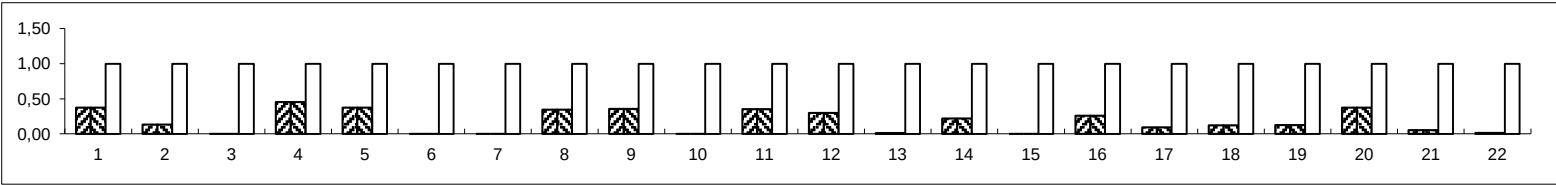
Les 22 variables qui ont été prises en compte sur un total de 35 ont été celles qui pouvaient être traitées à l'aide d'une A.C.P. .

- | | |
|------|---|
| N°1 | Droit de vote pour les travailleurs européens résidant en France depuis plus de 5 ans |
| N°2 | Distribution des seringues gratuites |
| N°3 | Impérialisme culturel américain sur le cinéma et la télévision |
| N°4 | Ne plus importer de produits des pays à bas salaires |
| N°5 | Contre les mariages des personnes de races différentes |
| N°6 | Les magazines porno et les films X sont des distractions inoffensives |
| N°7 | Il faut interdire le voile islamique |
| N°8 | J. TOUBON a raison de vouloir éliminer les mots étrangers |
| N°9 | Les immigrés renforcent notre pays grâce à leur travail |
| N°10 | La libre expression ne devrait pas être reconnue aux extrémistes |
| N°11 | Les accords de libre échange sont une source de prospérité réciproque |
| N°12 | L'homosexualité est un mode de vie que la société devra accepter |
| N°13 | Le droit de la police d'exercer les contrôles d'identité est grave |
| N°14 | La culture française s'enrichit des échanges avec les autres cultures |
| N°15 | On doit totalement interdire de fumer dans les lieux publics |
| N°16 | Les femmes devraient revenir à leur rôle traditionnel dans la société |
| N°17 | Les marques d'alcool et de tabac ne devraient pas faire de la publicité |
| N°18 | A la radio on entend trop de rock et de R.A.T.P |
| N°19 | On devrait faire plus d'efforts pour améliorer la situation des minorités en France |
| N°20 | On devrait interdire les fast food en France |
| N°21 | Je ne me sens pas grand chose en commun avec les autres races |
| N°22 | Il est normal que les médias parlent de la vie privée des hommes politiques |

FONCTIONS VARIANCES SUR A.C.P.

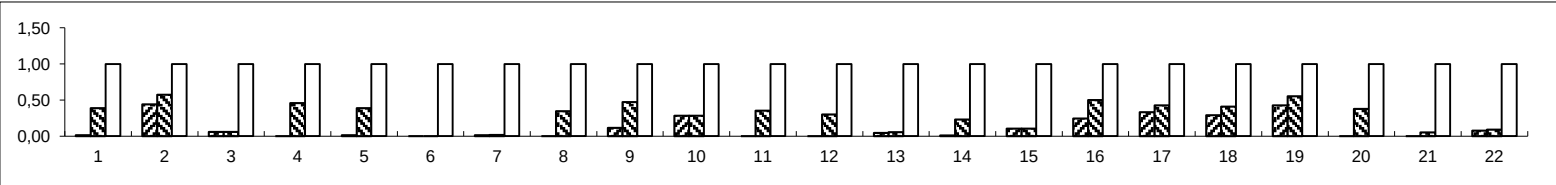
COMPOSANTE TEMPORELLE N° 1
VALEUR PROPRE CORRESPONDANTE 3,95

VALEURS	0,31	0,18	0,008	0,34	0,31	0	0,02	0,3	0,3	-0,02	0,3	0,28	0,05	0,24	-0,02	0,26	-0,15	0,18	0,18	0,31	0,11	-0,06
PART DE																						
VARIANCE	0,37	0,13	0,00	0,45	0,37	0,00	0,00	0,34	0,36	0,00	0,35	0,30	0,01	0,22	0,00	0,26	0,09	0,12	0,13	0,37	0,05	0,01
PART DE																						
CUMULEE	0,37	0,13	0,00	0,45	0,37	0,00	0,00	0,34	0,36	0,00	0,35	0,30	0,01	0,22	0,00	0,26	0,09	0,12	0,13	0,37	0,05	0,01
VARIANCE																						
totale	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1



COMPOSANTE TEMPORELLE N° 2
VALEUR PROPRE CORRESPONDANTE 2,48

VALEURS	-0,07	-0,42	0,153	0,04	0,08	-0,04	-0,08	-0,02	0,22	0,34	0,02	-0,03	0,13	-0,06	0,2	0,31	0,37	0,34	-0,42	0,02	-0,02	-0,18
PART DE																						
VARIANCE	0,01	0,44	0,06	0,00	0,01	0,00	0,01	0,00	0,12	0,28	0,00	0,00	0,04	0,01	0,10	0,24	0,33	0,29	0,43	0,00	0,00	0,08
PART DE																						
CUMULEE	0,39	0,57	0,06	0,46	0,39	0,00	0,01	0,34	0,47	0,28	0,35	0,30	0,05	0,23	0,10	0,50	0,43	0,41	0,55	0,38	0,05	0,09
VARIANCE																						
totale	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1



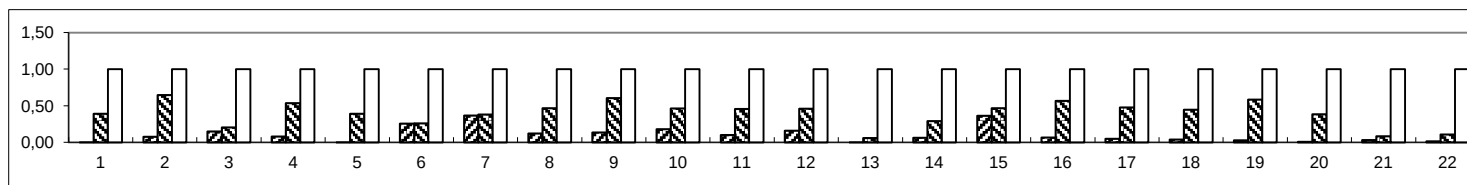


Ann III-4-b)

FONCTIONS VARIANCES SUR A.C.P.

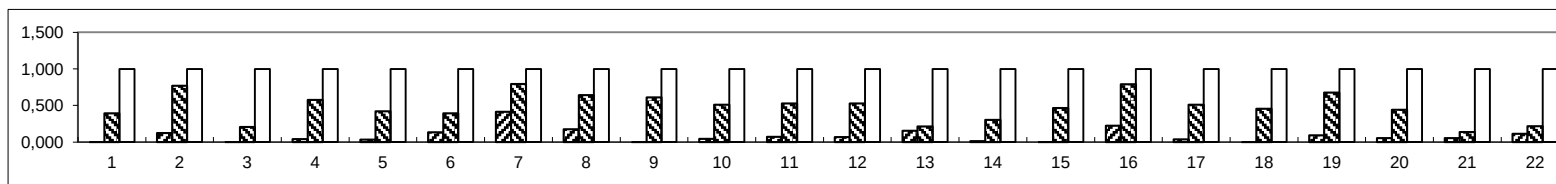
COMPOSANTE TEMPORELLE N° 3
VALEUR PROPRE CORRESPONDANTE 2,28

VALEURS	-0,03	0,18	0,254	-0,19	-0,03	0,33	0,4	0,23	-0,24	0,28	-0,21	0,26	0,04	0,16	0,4	-0,17	0,15	0,13	0,11	0,06	0,12	0,08
PART DE																						
VARIANCE	0,00	0,07	0,15	0,08	0,00	0,25	0,36	0,12	0,13	0,18	0,10	0,16	0,00	0,06	0,36	0,06	0,05	0,04	0,03	0,01	0,03	0,01
VARIANCE																						
CUMULEE	0,39	0,65	0,21	0,54	0,39	0,26	0,38	0,46	0,61	0,46	0,46	0,46	0,06	0,29	0,47	0,57	0,48	0,45	0,58	0,38	0,08	0,11
VARIANCE																						
totale	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1



COMPOSANTE TEMPORELLE N° 4
VALEUR PROPRE CORRESPONDANTE 1,86

VALEURS	0,04	-0,26	0,009	0,15	0,13	-0,27	0,47	-0,31	0,04	0,16	0,2	-0,19	-0,29	0,08	-0	-0,35	0,14	-0,06	0,22	0,17	0,17	-0,25
PART DE																						
VARIANCE	0,003	0,124	0,000	0,041	0,033	0,135	0,413	0,176	0,003	0,045	0,071	0,069	0,155	0,012	0,000	0,224	0,035	0,007	0,092	0,056	0,055	0,112
VARIANCE																						
CUMULEE	0,39	0,77	0,21	0,58	0,42	0,39	0,79	0,64	0,61	0,51	0,53	0,53	0,21	0,30	0,47	0,79	0,51	0,45	0,68	0,44	0,14	0,22
VARIANCE																						
totale	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1



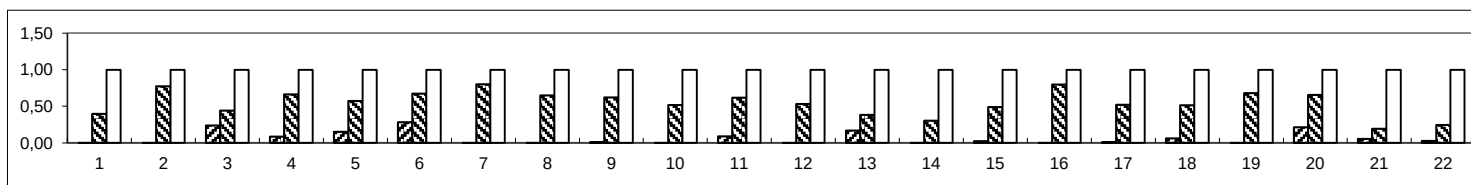


Ann III-4-b)

COMPOSANTE TEMPORELLE N° 5

VALEUR PROPRE CORRESPONDANTE 1,45

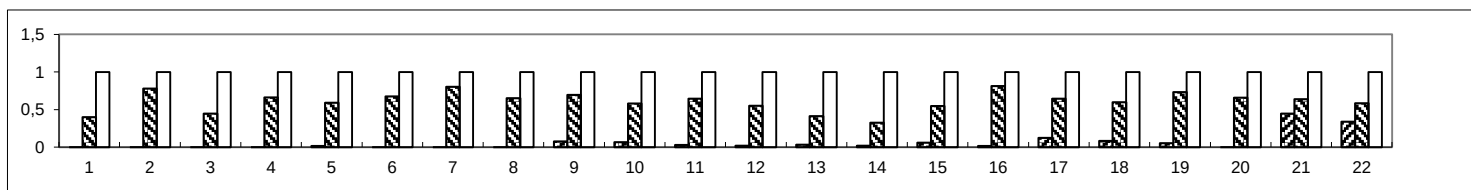
VALEURS	0,05	-0,01	0,40	0,24	-0,32	-0,44	0,07	0,07	-0,09	0,07	0,25	0,06	0,34	0,02	0,13	0,07	-0,08	-0,20	0,05	-0,39	0,19	0,13
PART DE																						
VARIANCE	0,00	0,00	0,24	0,09	0,15	0,28	0,01	0,01	0,01	0,01	0,09	0,00	0,17	0,00	0,02	0,01	0,01	0,06	0,00	0,21	0,05	0,03
VARIANCE																						
CUMULEE	0,40	0,77	0,44	0,66	0,57	0,67	0,80	0,65	0,62	0,52	0,62	0,53	0,38	0,30	0,49	0,80	0,52	0,51	0,68	0,65	0,19	0,24
VARIANCE																						
totale	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00



COMPOSANTE TEMPORELLE N° 6

VALEUR PROPRE CORRESPONDANTE 1,4

VALEURS	0,03	-0,07	-0,05	-0,02	-0,11	0,01	0,04	0,05	0,23	0,22	-0,14	-0,12	-0,15	-0,12	-0,2	-0,11	-0,3	0,24	-0,19	0,05	0,57	0,49
PART DE																						
VARIANCE	0	0,01	0,0035	0	0,02	0	0	0	0,08	0,07	0,03	0,02	0,03	0,02	0,06	0,02	0,12	0,08	0,05	0	0,45	0,34
VARIANCE																						
CUMULEE	0,40	0,78	0,45	0,66	0,59	0,67	0,80	0,65	0,70	0,58	0,64	0,55	0,41	0,32	0,55	0,81	0,64	0,60	0,73	0,66	0,64	0,58
VARIANCE																						
totale	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1



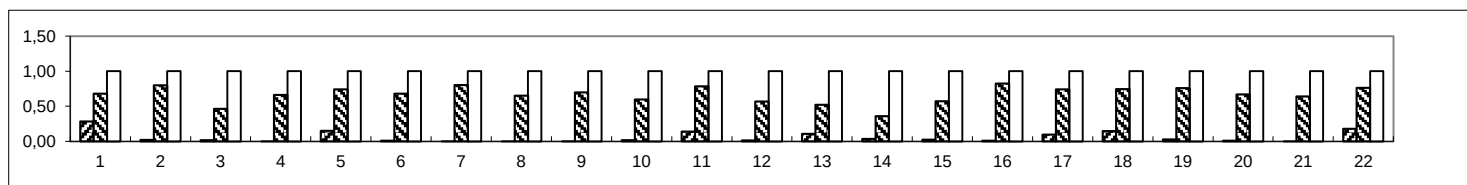


Ann III-4-b)

FONCTIONS VARIANCES SUR A.C.P.

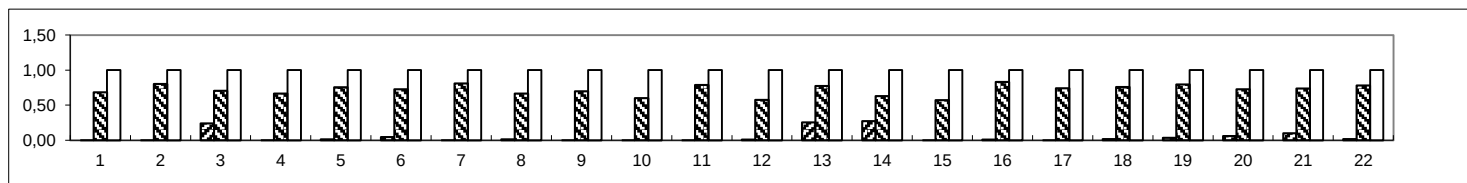
COMPOSANTE TEMPORELLE N° 7
VALEUR PROPRE CORRESPONDANTE 1,3

VALEURS	-0,47	0,13	0,12	0,03	0,34	0,08	-0,05	0,01	0,02	0,11	0,33	0,11	-0,29	0,16	-0,14	0,09	0,27	-0,34	-0,15	-0,09	0,02	0,37
PART DE VARIANCE	0,28	0,02	0,02	0,00	0,15	0,01	0,00	0,00	0,00	0,02	0,14	0,01	0,11	0,03	0,02	0,01	0,10	0,15	0,03	0,01	0,00	0,18
VARIANCE CUMULEE	0,68	0,80	0,46	0,66	0,74	0,68	0,80	0,65	0,70	0,60	0,78	0,57	0,52	0,36	0,57	0,82	0,74	0,74	0,76	0,67	0,64	0,76
VARIANCE totale	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00



COMPOSANTE TEMPORELLE N° 8
VALEUR PROPRE CORRESPONDANTE 1,1

VALEURS	-0,03	-0,05	-0,468	-0,02	0,11	0,2	0,07	-0,11	-0,02	0,02	0,06	-0,09	0,48	0,5	-0,02	-0,09	0	-0,12	-0,18	-0,23	0,3	-0,13
PART DE VARIANCE	0,00	0,00	0,24	0,00	0,01	0,04	0,01	0,01	0,00	0,00	0,00	0,01	0,25	0,27	0,00	0,01	0,00	0,02	0,04	0,06	0,10	0,02
VARIANCE CUMULEE	0,68	0,80	0,71	0,66	0,76	0,73	0,81	0,66	0,70	0,60	0,79	0,58	0,78	0,63	0,57	0,83	0,74	0,76	0,80	0,73	0,74	0,78
VARIANCE totale	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

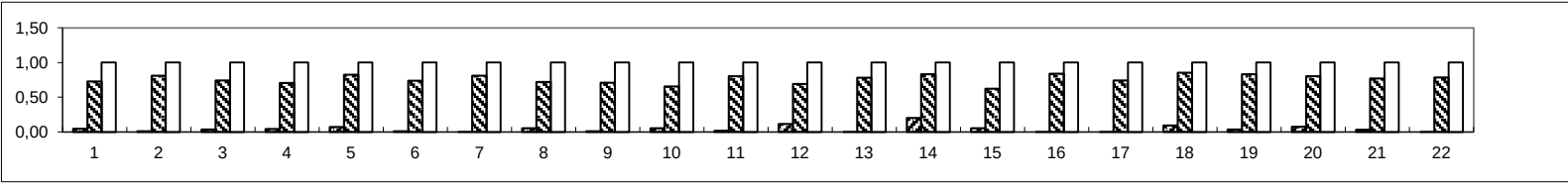




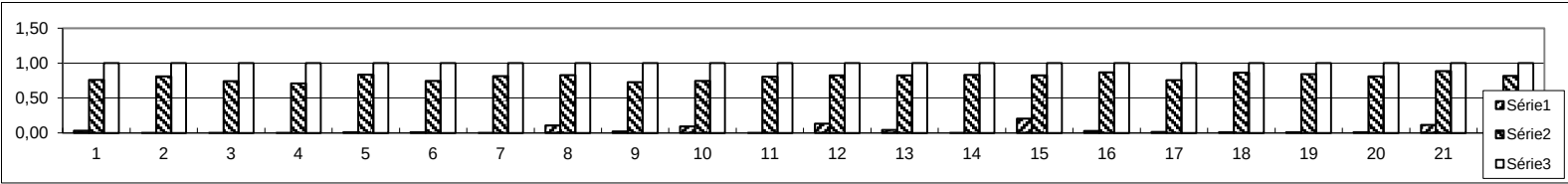
Ann III-4-b)

FONCTIONS VARIANCES SUR A.C.P.

COMPOSANTE TEMPORELLE N°	9																					
VALEUR PROPRE CORRESPONDANTE	0,96																					
VALEURS	-0,22	-0,09	-0,194	0,21	0,27	0,1	0,01	-0,24	-0,11	-0,24	0,13	0,34	0,08	-0,46	0,23	-0,07	-0,02	0,31	0,19	-0,28	0,18	0,07
PART DE VARIANCE	0,05	0,01	0,04	0,04	0,07	0,01	0,00	0,05	0,01	0,05	0,02	0,11	0,01	0,20	0,05	0,00	0,00	0,09	0,04	0,07	0,03	0,00
VARIANCE CUMULEE	0,73	0,81	0,74	0,71	0,83	0,74	0,81	0,72	0,71	0,65	0,80	0,69	0,78	0,83	0,62	0,84	0,74	0,85	0,83	0,80	0,77	0,78
VARIANCE totale	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00



COMPOSANTE TEMPORELLE N°	10																					
VALEUR PROPRE CORRESPONDANTE	0,88																					
VALEURS	0,19	-0,02	0,032	0,08	-0,11	0,11	0,05	-0,35	0,15	0,32	-0,04	0,39	0,22	0,04	-0,48	-0,18	0,13	0,11	0,11	-0,1	-0,36	0,19
PART DE VARIANCE	0,03	0,00	0,00	0,01	0,01	0,01	0,00	0,11	0,02	0,09	0,00	0,13	0,04	0,00	0,20	0,03	0,01	0,01	0,01	0,01	0,11	0,03
VARIANCE CUMULEE	0,76	0,81	0,74	0,71	0,84	0,75	0,81	0,83	0,73	0,75	0,81	0,82	0,82	0,83	0,83	0,87	0,75	0,86	0,84	0,81	0,88	0,82
VARIANCE totale	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00

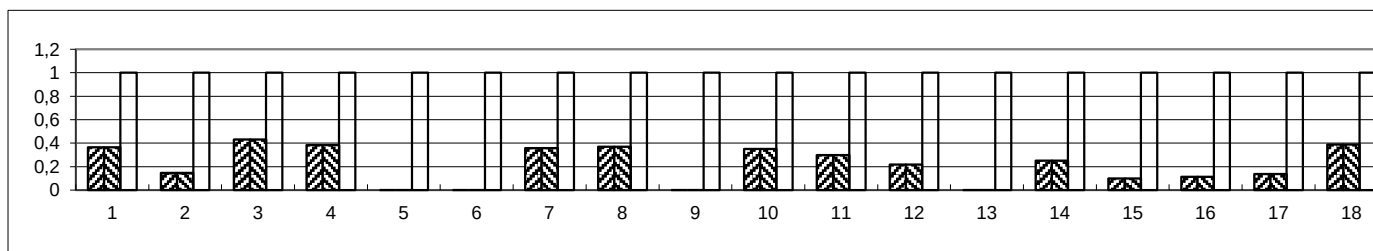


FONCTIONS VARIANCES **APRES SUPPRESSION DES QUESTIONS N° 3,13,21,22**

COMPOSANTE N° 1

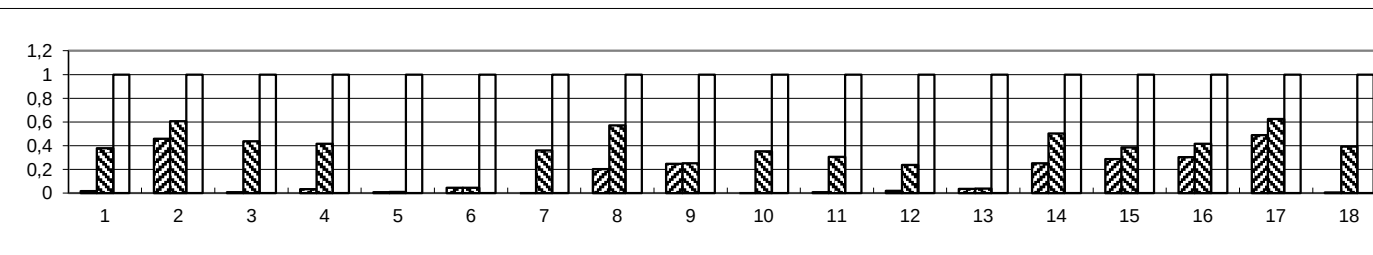
VALEUR PROPRE CORRESPONDANTE 3,91

VALEURS	0,31	0,19	0,33	0,31	0,02	0	0,3	0,31	-0,03	0,3	0,28	0,24	-0,03	0,25	-0,16	0,17	0,19	0,32
PART DE	0,36	0,15	0,43	0,38	0	0	0,36	0,37	0	0,35	0,3	0,22	0	0,25	0,1	0,11	0,14	0,39
VARIANCE	0,36	0,15	0,43	0,38	0	0	0,36	0,37	0	0,35	0,3	0,22	0	0,25	0,1	0,11	0,14	0,39
CUMULEE																		
VARIANCE																		
totale	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

**COMPOSANTE N° 2**

VALEUR PROPRE CORRESPONDANTE 2,42

VALEURS	-0,08	-0,44	0,05	0,12	-0,06	-0,14	-0,04	0,29	0,32	0,04	-0,05	-0,09	0,12	0,32	0,35	0,35	-0,45	0,04
PART DE	0,02	0,46	0,01	0,03	0,01	0,04	0	0,2	0,25	0	0,01	0,02	0,03	0,25	0,29	0,3	0,49	0
VARIANCE	0,38	0,61	0,44	0,42	0,01	0,04	0,36	0,57	0,25	0,35	0,3	0,24	0,04	0,5	0,39	0,41	0,62	0,39
CUMULEE																		
VARIANCE																		
totale	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1





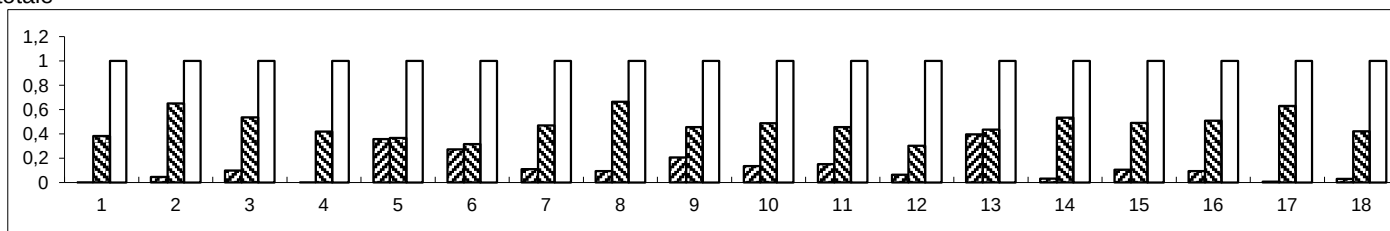
Ann III-4-c)

FONCTIONS VARIANCES **APRES SUPPRESSION DES QUESTIONS N° 3,13,21,22**

COMPOSANTE N° 3

VALEUR PROPRE CORRESPONDANTE 2,19

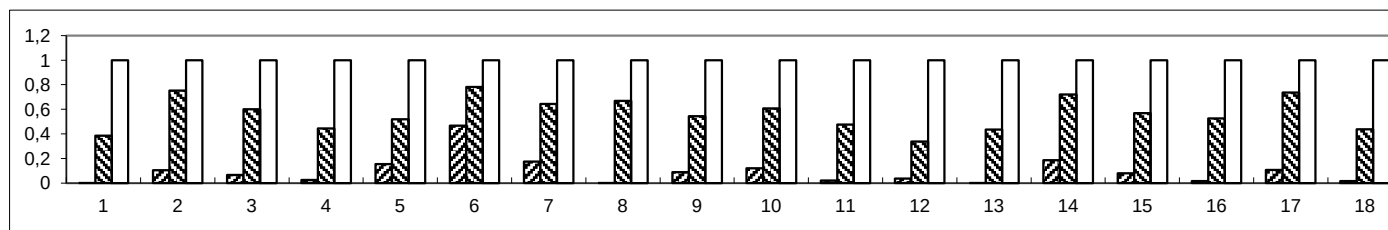
VALEURS	-0,04	0,14	-0,21	0,03	0,4	0,35	0,22	-0,21	0,31	-0,25	0,26	0,17	0,43	-0,12	0,22	0,21	0,05	0,12
PART DE	0	0,04	0,1	0	0,36	0,27	0,11	0,09	0,21	0,13	0,15	0,06	0,4	0,03	0,11	0,09	0	0,03
VARIANCE																		
CUMULEE	0,38	0,65	0,54	0,42	0,36	0,32	0,47	0,66	0,46	0,49	0,45	0,3	0,43	0,53	0,49	0,51	0,63	0,42
VARIANCE																		
totale	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1



COMPOSANTE N° 4

VALEUR PROPRE CORRESPONDANTE 1,66

VALEURS	0,04	-0,25	0,2	0,12	-0,31	0,53	-0,32	0,05	0,23	0,27	-0,11	0,15	0,01	-0,34	0,22	-0,1	0,25	0,1
PART DE	0	0,1	0,06	0,03	0,15	0,47	0,17	0	0,09	0,12	0,02	0,04	0	0,19	0,08	0,02	0,11	0,01
VARIANCE																		
CUMULEE	0,38	0,75	0,6	0,44	0,52	0,78	0,64	0,67	0,54	0,61	0,48	0,34	0,43	0,72	0,57	0,52	0,74	0,44
VARIANCE																		
totale	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1





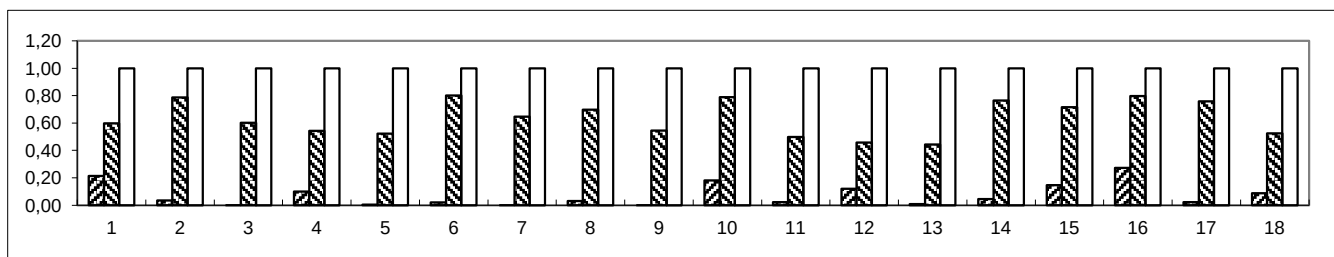
Ann III-4-c)

FONCTIONS VARIANCES **APRES SUPPRESSION DES QUESTIONS N° 3,13,21,22**

COMPOSANTE N° 5

VALEUR PROPRE CORRESPONDANTE 1,31

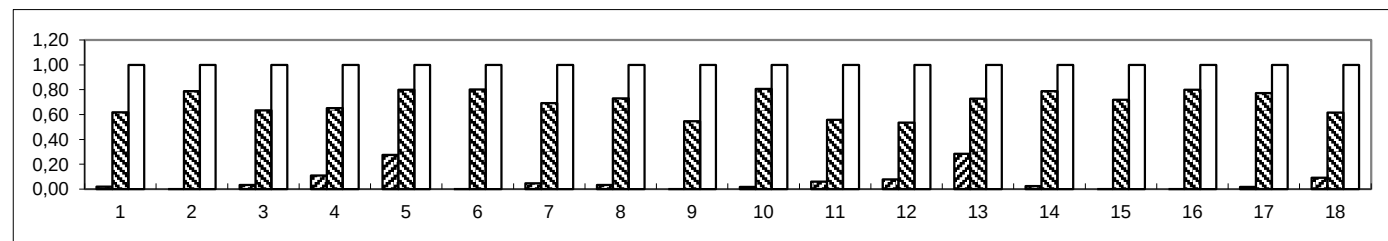
VALEURS	-0,4	0,16	0,03	0,28	0,05	-0,12	0,05	-0,15	-0,04	0,37	0,13	0,3	0,08	0,18	0,33	-0,46	-0,13	-0,26
PART DE VARIANCE	0,21	0,04	0,00	0,10	0,00	0,02	0,00	0,03	0,00	0,18	0,02	0,12	0,01	0,04	0,15	0,27	0,02	0,09
VARIANCE CUMULEE	0,6	0,79	0,6	0,54	0,52	0,8	0,65	0,7	0,54	0,79	0,5	0,46	0,44	0,76	0,71	0,8	0,76	0,52
VARIANCE totale	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

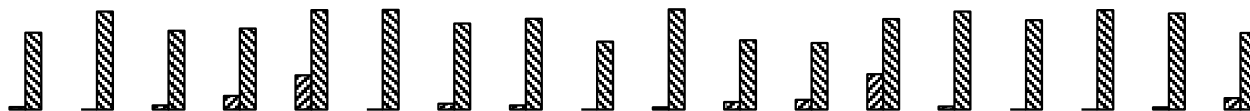


COMPOSANTE N° 6

VALEUR PROPRE CORRESPONDANTE 1,1

VALEURS	0,14	0,03	0,17	-0,32	-0,5	0,01	0,21	-0,18	-0,04	0,13	0,24	-0,27	0,51	0,15	0,06	0,05	0,12	-0,29
PART DE VARIANCE	0,02	0,00	0,03	0,11	0,28	0,00	0,05	0,03	0,00	0,02	0,06	0,08	0,28	0,02	0,00	0,00	0,02	0,09
VARIANCE CUMULEE	0,62	0,79	0,63	0,65	0,80	0,80	0,69	0,73	0,55	0,81	0,56	0,54	0,73	0,79	0,72	0,80	0,77	0,62
VARIANCE totale	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00





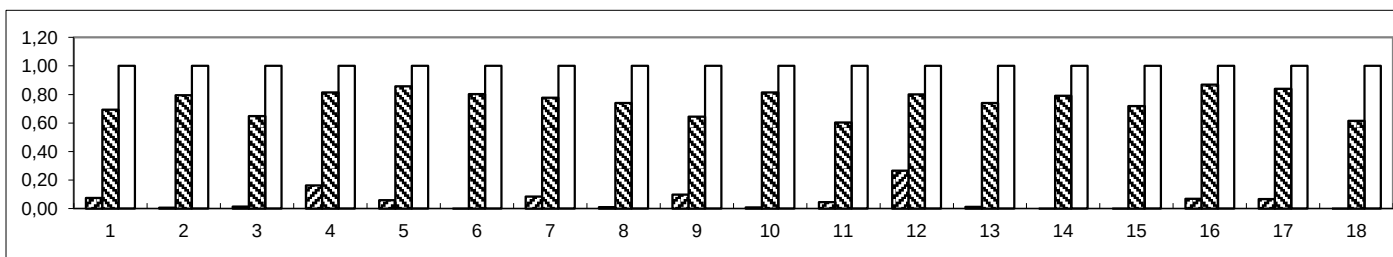
Ann III-4-c)

FONCTIONS VARIANCES APRES SUPPRESSION DES QUESTIONS N° 3,13,21,22

COMPOSANTE N° 7

VALEUR PROPRE CORRESPONDANTE 0,98

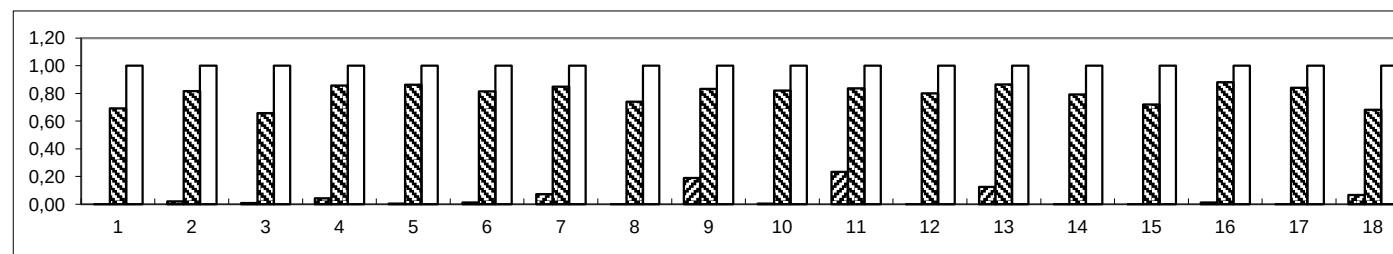
VALEURS	-0,28	-0,09	0,12	0,41	0,25	-0,02	-0,29	-0,1	-0,32	0,1	0,22	-0,52	0,11	-0,05	0,01	0,27	0,26	-0,01
PART DE VARIANCE	0,07	0,01	0,01	0,16	0,06	0,00	0,08	0,01	0,10	0,01	0,05	0,26	0,01	0,00	0,00	0,07	0,07	0,00
VARIANCE CUMULEE	0,69	0,80	0,65	0,81	0,86	0,80	0,78	0,74	0,64	0,82	0,60	0,80	0,74	0,79	0,72	0,87	0,84	0,62
VARIANCE totale	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00



COMPOSANTE N° 8

VALEUR PROPRE CORRESPONDANTE 0,8

VALEURS	0,01	0,16	0,11	-0,23	0,08	-0,12	-0,3	-0,01	0,49	0,08	0,54	-0,03	-0,4	-0,05	-0,05	0,12	0,02	-0,29
PART DE VARIANCE	0,00	0,02	0,01	0,04	0,00	0,01	0,07	0,00	0,19	0,01	0,23	0,00	0,13	0,00	0,00	0,01	0,00	0,07
VARIANCE CUMULEE	0,69	0,82	0,66	0,86	0,86	0,81	0,85	0,74	0,83	0,82	0,84	0,80	0,87	0,79	0,72	0,88	0,84	0,68
VARIANCE totale	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00



ANNEXE IV

DEMONSTRATION DE QUELQUES CALCULS CONNUS ET UTILES

Nous ne donnerons pas de démonstration sur les résultats analytiques généraux énoncés au chapitre II-2). Nous nous contenterons de démontrer les résultats spécifiques à l'analyse harmonique.

Ann-4-1) Les théorèmes de base de l'analyse harmonique

Ann-4-1-a) Théorème fondamental

La norme de la projection des trajectoires de X sur N vecteurs orthogonaux est de variance maximale si les N vecteurs choisis sont les N premiers vecteurs propres de l'opérateur de covariance.

Démonstration :

Désignons par g_i une suite de N vecteurs orthonormés et P le projecteur sur le sous espace engendré par ces N vecteurs.

$$\begin{aligned} V(\|P(X(\omega))\|) &= \int_{\Omega} \left\langle \sum_{i \leq N} \langle X(\omega), g_i \rangle g_i, \sum_{i \leq N} \langle X(\omega), g_i \rangle g_i \right\rangle dP\Omega \\ &= \int_{\Omega} \sum_{i \leq N} \langle X(\omega), g_i \rangle^2 dP\Omega = \sum_{i \leq N} \int_T \int_T \int_{\Omega} X(\omega, t) X(\omega, t') g_i(t) g_i(t') dt dt' dP\Omega \\ &= \sum_{i \leq N} \int_T \int_T N(t, t') g_i(t) g_i(t') dt dt' = \sum_{i \leq N} \langle U(g_i), g_i \rangle \end{aligned}$$

Si $\{f_j\}_{j \in N}$ est la base Hilbertienne des vecteurs propres de U alors : $g_i = \sum_{j \in N} \langle g_i, f_j \rangle f_j$

$$\begin{aligned} \text{on a: } \sum_{i \leq N} \langle U(g_i), g_i \rangle &= \sum_{i \leq N} \int_T U(g_i)(t) g_i(t) dt \\ &= \sum_{i \leq N} \int_T \left(\sum_{j \in N} \langle g_i, f_j \rangle U(f_j)(t) \right) \left(\sum_{j \in N} \langle g_i, f_j \rangle f_j(t) \right) dt \end{aligned}$$

$$= \sum_{i \leq N} \int_T \left(\sum_{j \in N} \langle g_i, f_j \rangle \lambda_j f_j(t) \right) \left(\sum_{j \in N} \langle g_i, f_j \rangle f_j(t) \right) dt = \sum_{j \in N} \sum_{i \leq N} \lambda_j \langle g_i, f_j \rangle^2$$

Pour que la dernière série à terme positif soit maximale il suffit que $f_i = g_i$;

en effet nous allons montrer que cette dernière s'écrit $\sum_{i \in N} \lambda_i k_i$ avec $\sum_{i \leq N} k_i = N$ et $k_j \leq 1$; cette série étant maximale lorsque $k_1 = 1, k_2 = 1, \dots, k_N = 1, k_{N+1} = 0, k_{N+2} = 0, \dots$

En effet si $k_j = \sum_{i \leq N} \langle g_i, f_j \rangle^2$ alors

$$\sum_{j \in N} k_j = \sum_{j \in N} \sum_{i \leq N} \langle g_i, f_j \rangle^2 = \sum_{i \leq N} \sum_{j \in N} \langle g_i, f_j \rangle^2 = \sum_{i \leq N} \|g_i\|^2 = N$$

et d'autre part

$$k_j = \sum_{i \leq N} \langle g_i, f_j \rangle^2 \leq \|f_j\|^2 = 1$$

Ann-4-1-b) Théorème

Notons $\|X\|$ la fonction définie par $\|X\|(\omega) = \|X(\omega)\|$ avec les notations de (II-2-c) on a les égalités

$$V(\|X\|) = \sum_{i \in N} \lambda_i \quad \text{et} \quad V(\|Y\|) = \sum_{i < p} \lambda_i \quad \text{et} \quad V(\|Z\|) = \sum_{i \geq p} \lambda_i$$

Démonstration :

Presque pour tout $\omega \in \Omega$ $\|X(\omega)\|^2 = \sum_{i \in N} \langle X(\omega), f_i \rangle^2$.

Comme $X \in L^2(\Omega)$, $\|X\|$ composée de fonctions de carrés intégrables est de carré intégrable :

$$\int_{\Omega} \|X(\omega)\|^2 dP\omega = \int_{\Omega} \sum_{i \in N} \langle X(\omega), f_i \rangle^2 dP\omega = \sum_{i \in N} \int_{\Omega} \langle X(\omega), f_i \rangle^2 dP\omega$$

La dernière série est convergente car positive et majorée par la série précédente, ce qui justifie les égalités.

Pour conclure il suffit d'appliquer le théorème suivant :

Ann-4-1-c) Théorème

Avec les notations habituelles :

$$\int_{\Omega} X(\omega) \left\langle X(\omega), f_i \right\rangle dP\omega = \lambda_i f_i$$

et

$$\int_{\Omega} \left\langle X(\omega), f_i \right\rangle \left\langle X(\omega), f_j \right\rangle dP\omega = \delta_{ij} \lambda_i$$

Démonstration :

D'après le théorème de Fubini :

$$\int_{\Omega} X(\omega) \left\langle X(\omega), f_i \right\rangle dP\omega = \int_{\Omega} \int_T X(\omega, t) X(\omega, t') f_i(t') d\mu(t') dP\omega =$$

$$\int_T \int_{\Omega} X(\omega, t) X(\omega, t') f_i(t') d\mu(t') dP\omega = \int_T N(t, t') f_i(t') d\mu(t') = \lambda_i f_i(t)$$

D'autre part :

$$\int_{\Omega} \left\langle X(\omega), f_i \right\rangle \left\langle X(\omega), f_j \right\rangle dP\omega = \left\langle f_i, \int_{\Omega} X(\omega) \left\langle X(\omega), f_j \right\rangle dP\omega \right\rangle = \lambda_i \left\langle f_i, f_j \right\rangle = \delta_{ij} \lambda_i$$

Le premier théorème est totalement démontré en appliquant ce lemme ; il suffit de recommencer la démonstration pour Z ; le processus Y ne posant aucun problème de convergence.

Ann-4-2) Résultats sur les processus élémentaires

Ann-4-2-a) Lemme :

Si le processus donné est centré alors chaque processus élémentaire est centré.

Démonstration :

$$E(X_i) = \int_{\Omega} \left\langle X(\omega), f_i \right\rangle f_i = \int_{\Omega} \left[\int_T X(\omega, t) f_i(t) dt \right] dP\omega f_i = \int_T \left[\int_{\Omega} X(\omega, t) dP\omega \right] f_i(t) dt f_i = 0$$

Ann-4-2-b) Lemme

La variance totale d'un processus élémentaire est la quantité : $\lambda_i f_i^2$

Démonstration :

Calculons

$$V(X_i) = \int_{\Omega} \langle X(\omega), f_i \rangle^2 f_i^2 dP\omega = \int_{\Omega} \langle X(\omega), f_i \rangle^2 dP\omega f_i^2 = \lambda_i f_i^2$$

Ann-4-2-c) Lemme

La covariance de deux processus élémentaires différents est nulle

Démonstration :

$$\begin{aligned} \text{COV } X_i, X_j &= \int_{\Omega} X_i(\omega) X_j(\omega) dP\omega = \int_{\Omega} \langle X_i(\omega), f_i \rangle f_i \langle X_j(\omega), f_j \rangle f_j dP\omega = \\ &= \int_{\Omega} \langle X_i(\omega), f_i \rangle \langle X_j(\omega), f_j \rangle dP\omega f_i f_j = 0 \cdot f_i f_j \end{aligned}$$

Ann-4-2-d) Calcul formel

Ces deux lemmes nous amènent à calculer formellement la variance fonctionnelle des processus X , Y et Z

(1)

$$V(X) = V\left(\sum_{i \in N} X_i\right) = \sum_{i \in N} V(X_i) + \sum_{i \in N, j \in N} \text{COV}(X_i, X_j) = \sum_{i \in N} \lambda_i f_i^2$$

(2)

$$V(Y) = V\left(\sum_{i \leq p} X_i\right) = \sum_{i \leq p} V(X_i) + \sum_{i \leq p, j \leq p} \text{COV}(X_i, X_j) = \sum_{i \leq p} \lambda_i f_i^2$$

(3)

$$V(Z) = V\left(\sum_{i > p} X_i\right) = \sum_{i > p} V(X_i) + \sum_{i > p, j > p} \text{COV}(X_i, X_j) = \sum_{i > p} \lambda_i f_i^2$$

Ann-4-2-e) Comportement du noyau de l'opérateur de covariance

Il ne faut pas confondre le noyau de l'opérateur avec le noyau de covariance $N(t,t')$ du processus. Rappelons sa définition.

Définition :

Le noyau de covariance désigne l'ensemble des vecteurs $\{f\}$ de $L^2(T)$ tels que $U(f) = 0$; il est noté $\text{Ker}(U)$.

Remarque:

L'espace des fonctions de carré intégrable est la somme directe de l'espace engendré par les vecteurs propres associés aux valeurs propres non nulles et du noyau. Lorsque on projette le processus sur les vecteurs propres, on peut imaginer qu'une partie de l'information n'est pas prise en compte; c'est à dire que la projection du processus sur le noyau est significative. Le lemme suivant permet d'écarter cette inquiétude.

Ann-4-3) Théorème:

Presque pour tout $\omega \in \Omega$ la projection sur le sous espace vectoriel $\text{Ker}(U)$ de $X(\omega)$ est nulle

Démonstration :

Pour tout $\varphi \in \text{Ker}(U)$ on a :

$$\int_{\Omega} \langle X(\omega), \varphi \rangle^2 dP_{\Omega} = \int_{\Omega} \int_T \int_T X(\omega, t) \varphi(t) X(\omega, t') \varphi(t') dt dt' dP_{\Omega}$$

Appliquons le théorème de Fubini et développons :

$$\begin{aligned} &= \int_T \int_T \int_{\Omega} X(\omega, t) X(\omega, t') dP_{\Omega} \varphi(t) \varphi(t') dt dt' = \int_T \int_T N(t, t') \varphi(t) \varphi(t') dt dt' \\ &= \int_T U(\varphi)(t') \varphi(t') dt' = \langle U(\varphi), \varphi \rangle = 0 \end{aligned}$$

ANNEXE V

LE TRAITEMENT INFORMATIQUE

Le traitement des données a nécessité la création d'un programme en langage C++. Les textes et les formules ont été édités sous Word. Le tableur Excel a été utilisé pour l'édition des tableaux et des graphiques; le logiciel assiste l'analyste dans son travail en présentant toute les opérations sous forme de graphiques.

Ann V-1) la structure du programme

Le programme informatique qui a permis de faire les études pratiques propose

- La lecture des données
- le calcul des composantes temporelles sur la totalité ou sur une partie des données
- La représentation de l'indice fort
- La représentation et le prolongement des composantes temporelles
- La reconstitution des données et leurs prolongements
- La sauvegarde des différents résultats dans des fichiers

Ann V-1-a) La lecture des données

Les données doivent se présenter sous forme d'un fichier texte. Sa création peut se faire sous Excel. Les deux premiers nombres représentent respectivement le nombre d'instants et le nombre d'individus. Les données proprement dites du processus sont rangées fonction après fonction. Il est commode de présenter ce fichier texte de telle façon qu'une ligne (ou un regroupement de lignes) représente une fonction du processus. Alors, excepté pour les deux premiers nombres, les données sont présentées suivant l'habitude sous forme d'un tableau double entrées, où les lignes représentent les valeurs prises par chaque individu et les colonnes les différents instants enregistrés.

Ann V-1-b) Le traitement des données

Le tableau est d'abord centré et la matrice de covariance calculée à partir de ce tableau. Cette matrice est alors diagonalisée pour rechercher ses vecteurs propres et ses valeurs propres. Tous ces tableaux sont créés en mémoire vive. Attention les résultats numériques ne sont valides que si les valeurs propres sont différentes de zéro.

Ann V-1-c) La représentation de l'indice fort

La fonction variance est calculée directement à partir des valeurs centrées. Les variances partielles cumulées sont calculées à partir des valeurs propres et des vecteurs propres. La représentation graphique simultanée de la fonction variance et de la fonction variance d'indice n (indice fort de rang

n) permet de prendre la décision sur le rang au delà duquel la reconstitution du processus n'est qu'un bruit.

Ann V-1-d) La prévision des composantes temporelles

Cette partie est la plus délicate sur le plan pratique. Chaque composante temporelle est représentée sur un écran et l'opérateur peut prolonger les données en contrôlant ce qu'il fait sous forme graphique et numérique.

Ann V-1-e) La reconstitution et la prévision des données:

La reconstitution et la prévision des données se fait à l'aide de la formule :

$$X(\omega, t) \approx \sum_{i \leq n} \left\langle X(\omega), f_i \right\rangle f_i(t)$$

Pour un individu, les valeurs brutes, les valeurs reconstituées et les prévisions sont représentées simultanément sur un écran. Cela permet à l'opérateur de contrôler la validité de son travail d'un simple coup d'oeil.

Ann V-1-f) La sauvegarde des résultats

Les résultats sont sauvegardés dans deux fichiers textes.

Contenu du fichier de format « fichier.RES »

- Les données brutes
- les données reconstituées et les prévisions
- les pourcentages d'erreurs entre données brutes et reconstituées
- Le coefficient de projection des données brutes sur chaque composantes temporelles

Contenu du fichier de format « fichier.VEC »

- Les valeurs propres
- Les vecteurs propres
- Le prolongement des composantes temporelles

Tous ces résultats peuvent être repris par un tableau de type Excel comme nous l'avons fait , nous même, lors du traitement des exemples pratiques.

Ann V-2) Les différentes versions du programme

La première version du programme a été réalisé sous dos avec le logiciel TURBO C++ de chez BORLAND.

Ce programme comportait 2160 lignes pour le programme principal et 136 lignes pour le fichier, de format « fichier.H », contenant la définition de certaines variables.

La principale difficulté de ce programme était l'utilisation du format 16 bits qui limitait la dimension des matrices actives à environ 50x50.

Pour les données que nous avons traitées ce format était suffisant mais pour un travail professionnel il ne l'était pas.

Une nouvelle version sous Windows 95 a été écrite à l'aide du logiciel WINDEV. L'utilisation des 32 bits du micro processeur permet de traiter des données importantes sans aucune limitation pratique (matrices 1000x1000).

Une nouvelle version orienté objet est actuellement à l'écriture. Elle permettra de présenter un produit évolutif aux professionnels de la prévision. Ce dernier programme comporte plus de 15000 lignes.

Les différentes versions du programme sont déposées à l'agence de la protection des programmes: 119, avenue de Flandre à Paris, France

BIBLIOGRAPHIE

- [1] Anderson O.D. *Time Series Analysis and Forecasting*, Butterworths, Londres, 1976
- [2] Aguilera A.M., Ocaña F.A., Valderrama M. J. 1995
Regression dinamica de un proceso estocatico mediante analisis en componentes principales
(publication) *University of Granada España*
- [3] Aguilera A.M., Ocaña F.A., Valderrama M. J. 1995 Principal component prédiction
of a stochastic process (publication) *University of Granada España*
- [4] Aguilera A.M., Ocaña F.A., Valderrama M. J. 1995
Discussion of two principal component approaches to forecast continuous time stochastic
processes (publication) *University of Granada España*
- [5] Besse P. 1979 Etude Descriptive d'un Processus (approximation et interpolation)
(thèse) *Université P. Sabatier Toulouse*
- [6] Boumaza, Dauxois J., Pousse A. 1979 Analyse descriptive d'un processus qualitatif
Publication du laboratoire de statistiques et de probabilité Toulouse n° 04-79
- [7] Box G.E.P Jenkins G.M. *Time Series Analysis: Forecasting and control*, Holden-Day, San Francisco, 1976.
- [8] Cailliez F., Pages J.P. 1976 *Introduction à l'analyse des données* Smash, Paris
- [9] Chretienne PH. & Faure R. 1974 *Processus stochastiques, leurs graphes, leurs usages* Gauthier-Villars
- [10] Combs C. H. 1964 *A theory of data* New York : Wiley
- [11] Dauxois J., Pousse A. 1976 Les analyses factorielles en calcul des probabilités et en statistiques : essai d'étude synthétique Statistiques et Analyse des données (thèse) *Université P. Sabatier Toulouse*
- [12] Dagun E.B. Seasonal Factor Forecast from ARIMA Models, *communication présentée à la 40e session de l'Institut international de statistiques, Varsovie, Pologne, 1975*
- [13] Dieudonné J. 1972 *Eléments d'Analyse, fondements de l'analyse moderne* Gauthier-Villars
- [14] Deville J. C. 1977 Analyse harmonique et calendrier de constitution des familles
Annales de l'INSEE

- [15] Deville J. C. 1974 Méthodes statistiques et numériques de l'analyse harmonique *Annales de l'INSEE n° 15*
- [16] Deville J. C. & Saporta G. 1979 Analyse harmonique qualitative *Data Analysis and Informatics E.Diday eds North-Holland, 375-389*
- [17] Escoufier Y. 1970 Echantillon dans une population de variables aléatoires réelles *Publication de l'I.S.U.P. Volume XIX Fascicule 4*
- [18] Goodrich R.L. *Applied Statistical Forecasting*, Business Forecast Systems, Belmont, 1989.
- [19] INSEE 1990 Annuaire rétrospectif de la France 1948-1988 *Publication INSEE*
- [20] Janssen J. 1989 Finances Stochastique (*publication*) *INSUP / CNAM Programme Erasmus*
- [21] Kendal M.G., *time-series*, Charles Griffin, Londre, 1973
- [22] Ladiray D. Désaisonnalisation de séries chronologiques, le programme Census-X, *Note ENSAE, 1987*
- [23] Ladiray D. et Roth N. « Lissage robustes de séries chronologiques. une étude expérimentale », *Annales d'économie et de statistique, n°5, 1987, 147-181*
- [24] Loeve M. 1945 Fonctions aléatoires du second ordre C.R *Académie des Sciences*
- [25] Melard G. *Méthodes de prévision à court terme*, Edition de l'université de Belgique, Ellipses Paris, 1990
- [26] Saporta G. 1981 Méthodes Exploratoires d'analyse des données temporelles *Cahiers du bureau universitaire de recherche opérationnelle n° 37-38 Université P.M. Curie Paris*
- [27] Saporta G. 1985 « Data analysis for numerical and catégorical individual times series » *Applied Stochastic Models and Data Analysis vol 1 n°2 109-119*
- [28] Saporta G. 1990 *Probabilités, Analyse des données et statistique* Editions Technip
- [29] Ramsay J.O. 1982 When the data are functions *Psychometrika Vol 47 n°4*
- [30] Tenenhaus M. 1989 Décomposition d'une série chronologique (*publication*) *Groupe HEC D1340k*
- [31] Tenenhaus M 1989 La Méthodologies de Box-Jenskins pour l'analyse de la prévision d'une série chronologique (*publication*) *Groupe HEC D1362k*

- [32] Tenenhaus M. 1994 *Méthodes statistiques en gestion* Dunod
- [33] Winsberg S. et Ramsay J.O. Monotone spline transformations for ordered categorical data, *Paper present at the Psychometric Society Annual Meeting, Montréal Canada*