

MAI 1993

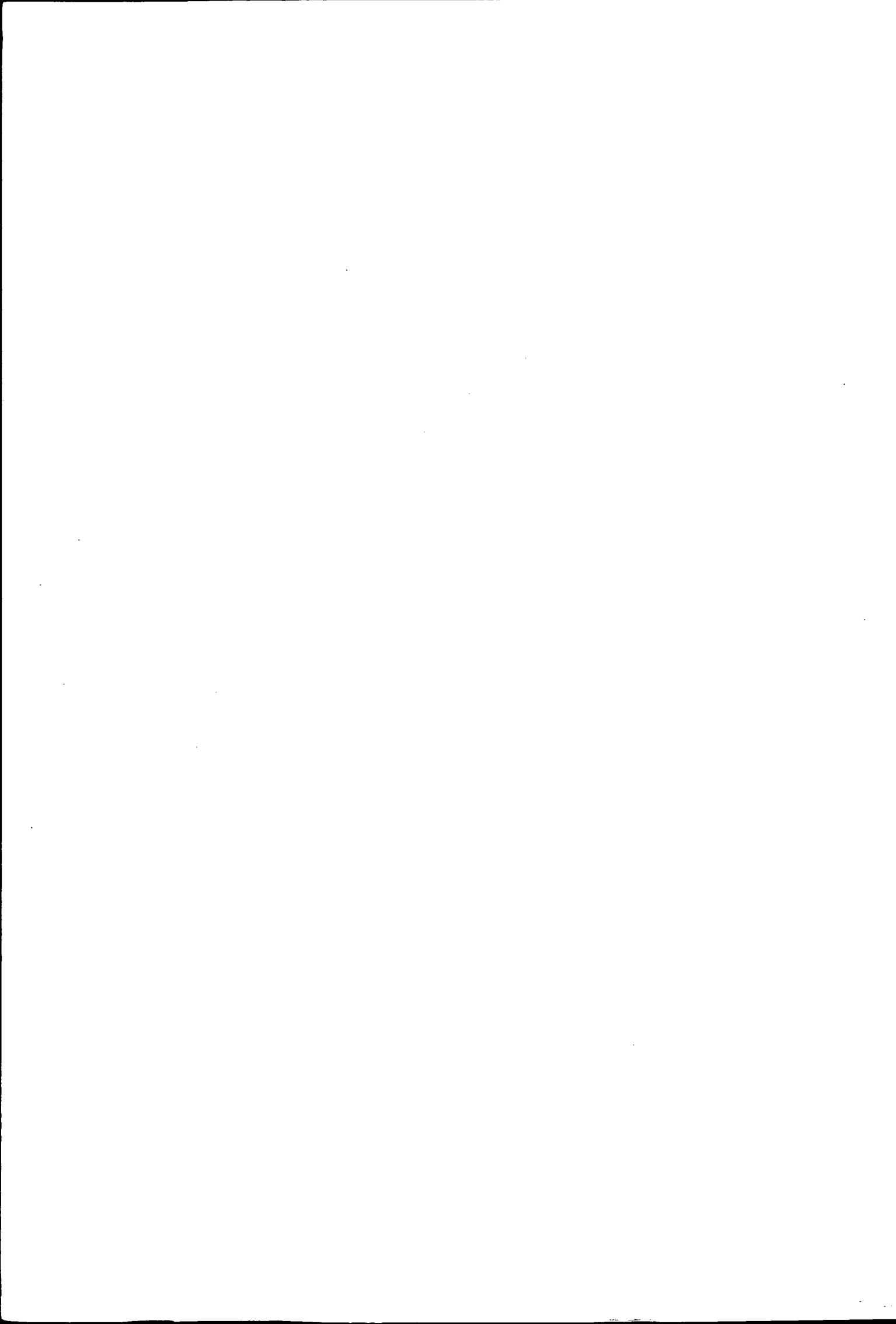
Exploration d'un modèle  
poissonnien  
d'indicateurs journaliers  
d'insécurité routière

Cette étude a été réalisée par : Robert AZENCOTT  
Anne RICORDEAU

Responsable de l'étude à l'OEST : Ruth BERGEL,  
dans le cadre d'un groupe de travail associant :

Elvyne FEVRIER  
Sylvain LASSARRE  
Patrick LE BRETON

**D I A M**  
RECHERCHES



## AVANT-PROPOS

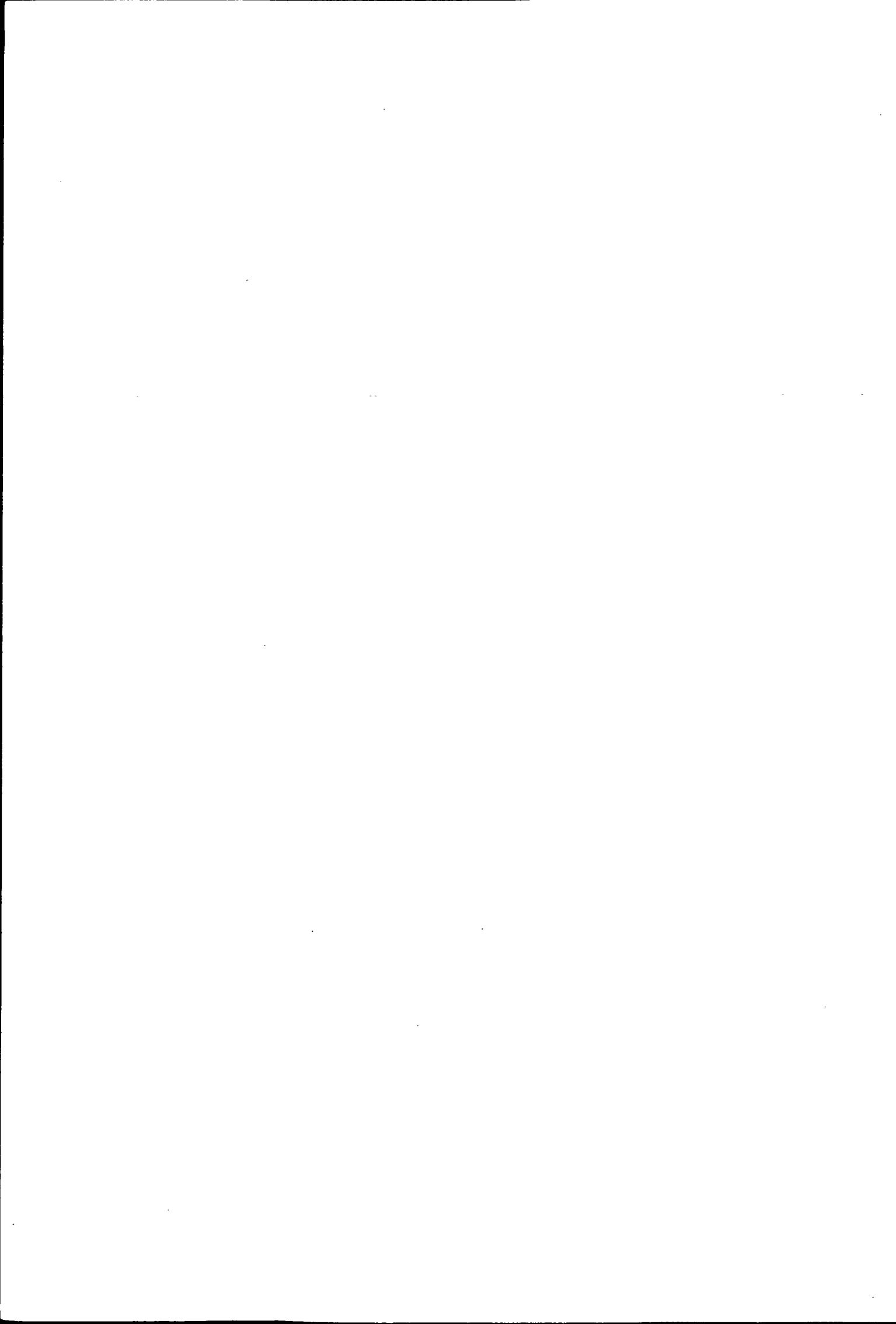
Cette étude a permis la mise en oeuvre exploratoire d'un modèle poissonnien d'indicateurs journaliers de l'insécurité routière.

Elle constitue un complément aux travaux de conception d'un modèle gaussien de suivi des indicateurs journaliers, réalisés à la demande de la D.S.C.R. en vue de la mise en oeuvre opérationnelle d'un tableau de bord de l'insécurité routière. Dans le cas où les indicateurs prennent des valeurs faibles, on tente en effet de mettre en oeuvre des modèles théoriquement plus adaptés qu'un modèle gaussien.

---

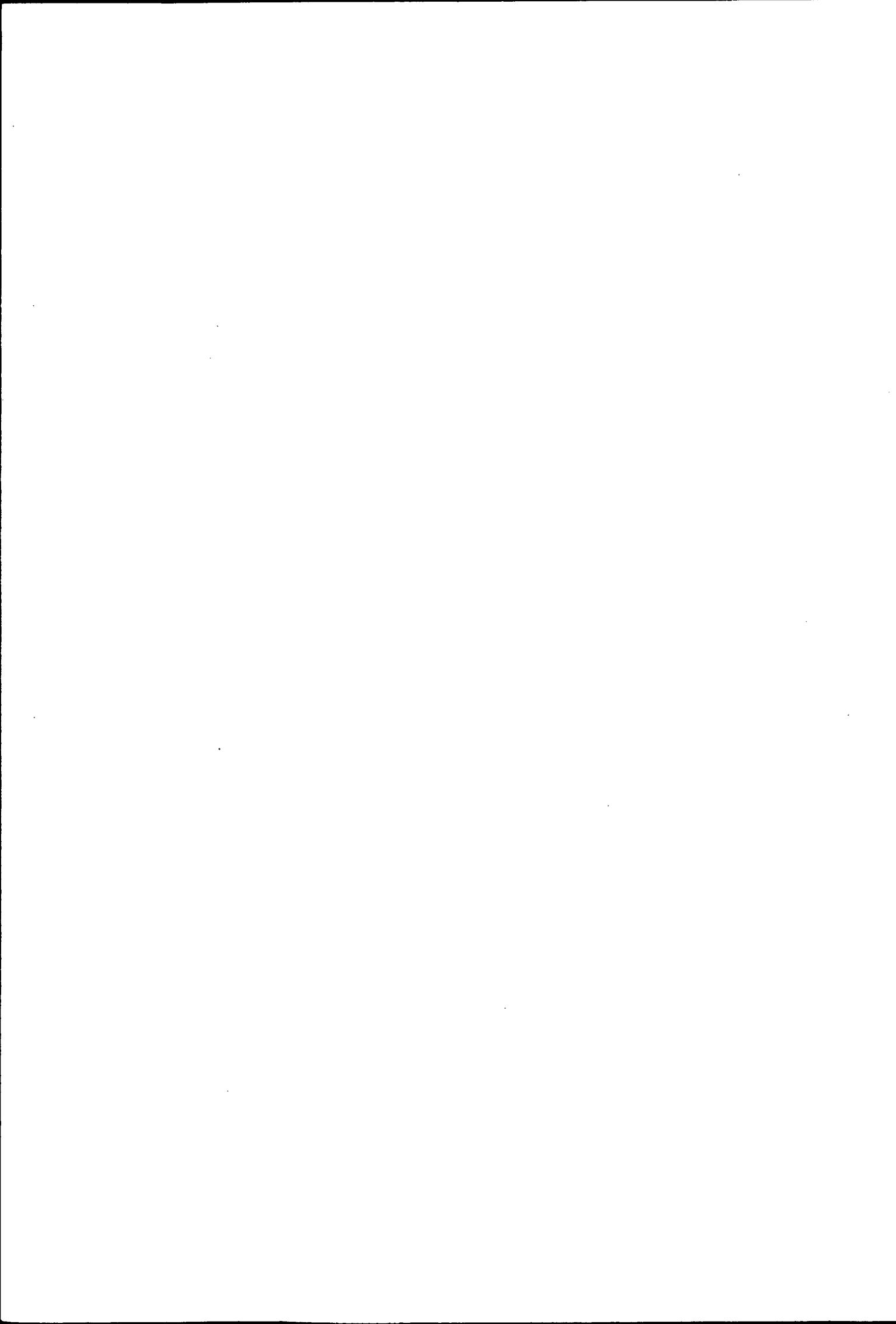
Les travaux présentés dans cette étude ont été réalisés par DIAM-Recherches, dans le cadre d'un groupe de travail animé par l'O.E.S.T., regroupant la D.S.C.R., l'I.N.R.E.T.S. et le S.E.T.R.A..

---



## TABLE DES MATIERES

	INTRODUCTION	1
1	AJUSTEMENT DU MODELE POISSONNIEN	2
	1.1 Le modèle paramétrique	2
	1.2 le choix des liaisons	4
	1.3 Intégration de variables explicatives	5
2	ESTIMATION ET COMPARAISON DE MODELES	7
	2.1 La moyenne conditionnelle comme critère de comparaison	8
	2.2 Résultats pour la série simple des Accidents	9
	2.3 Résultats pour la série simple des Tués	14
	2.4 Résultats pour la série double des Accidents et Tués	15
3	LA PREDICTION	20
	3.1 Présentation de la méthode	20
	3.2 Résultats de la prédiction	21
4	CONCLUSIONS	26
	4.1 Sur les performances du modèle poissonnien	26
	4.2 Sur les améliorations envisagées	28
	ANNEXES	29
	A.1 Passage des données brutes aux données corrigées	30
	A.2 Détail des résultats de la prédiction poissonnienne sur données corrigées	36
	A.3 Ajustement du modèle poissonnien sur données initiales	39
	A.4 Le prédicteur gaussien	44



## INTRODUCTION

Les travaux présentés dans cette étude portent sur la conception et la mise en oeuvre exploratoire d'un **modèle poissonnier d'indicateurs journaliers de l'insécurité routière**. Les indicateurs qui se prêtent à ce type de modélisation sont les nombres journaliers d'accidents et de victimes de la circulation routière, désagrégés par réseau ou par typologie, qui prennent des valeurs entières et faibles. Seuls deux indicateurs ont été modélisés: les nombres journaliers d'accidents et de tués comptabilisés sur les routes nationales, sur une période de 6 ans, allant du 1/1/85 au 31/8/90; on dispose ainsi de deux séries chronologiques de 2100 observations environ chacune.

Rappelons que ces indicateurs, en particulier le nombre d'accidents, ont fait l'objet de nombreuses modélisations parmi lesquelles on peut distinguer les modèles explicatifs linéaire, log-linéaire ou poissonnier des modèles prédictifs essentiellement de type Arma; mais à notre connaissance une modélisation multivariée n'a pas encore été exploitée tout particulièrement dans le cadre des séries temporelles. Pour mesurer l'influence de phénomènes étalés dans le temps ces modélisations ont été faites sur des données annuelles ou mensuelles, et plus rarement sur des données journalières. Le cadre journalier permet pour sa part de s'intéresser à des phénomènes plus ponctuels et d'étudier plus en détail la variabilité des indicateurs.

La présente étude est consacrée à l'ajustement d'un **modèle poissonnier** qui appartient à la classe des modèles markoviens. Pour justifier le choix poissonnier rappelons que les variables étudiées ici sont entières et ne prennent pas de valeurs trop importantes ( de l'ordre de 7 en moyenne pour le nombre de tués sur routes nationales, et de l'ordre de 50 en moyenne pour le nombre d'accidents sur routes nationales ). Quant aux modèles markoviens précisons qu'ils utilisent une approche locale et mettent l'accent sur les comportements relationnels : une variable  $y$  est décrite en prenant en compte tout le contexte l'environnant.

Ce type de modèle poissonnier markovien fait actuellement l'objet d'une réflexion approfondie dans le cadre des séries temporelles : en plus de l'étude sur son ajustement, différentes méthodes de prédiction ont été proposées et simulées. Nous présentons ici l'application d'une de ces méthodes qui utilise l'information apportée par un modèle linéaire approché de type auto-projectif. Précision qu'une prédiction n'est bien sur envisageable que si l'on prend en compte au moins pour une des séries des relations temporelles significatives, en plus des relations instantanées entre les séries.

En ce qui concerne l'intégration au modèle de **variables explicatives** qui permettront d'améliorer son ajustement et de lui apporter plus d'intérêt descriptif, précisons que dans les fichiers de données actuellement disponibles, seul le type de jour de la semaine peut être directement intégré. Toutefois, ces mêmes séries d'indicateurs journaliers ont fait récemment l'objet de deux études descriptives : la première (e1) s'est intéressée au choix de variables météorologiques pertinentes et à l'ajustement d'un modèle linéaire permettant la correction du seul effet météorologique; dans le cadre de la seconde étude (e2) un modèle de type Armax a ensuite été ajusté sur ces données corrigées de la météo en intégrant des variables explicatives décrivant les phénomènes calendaires exceptionnels qui correspondent aux jours fériés et de

départs et retours de vacances par opposition à des phénomènes calendaires hebdomadaires ( l'annexe (1) présente ce passage des données brutes aux données corrigées ).

Nous disposons donc des séries de **données corrigées** des effets météo et de ces phénomènes calendaires exceptionnels, ainsi que d'un modèle Arma ajusté sur chacune de ces séries. Et puisque les variables explicatives utilisées dans ces études antérieures sont encore trop nombreuses pour être intégrées au modèle poissonnien nous avons choisi de travailler dans un premier temps sur ces séries corrigées artificiellement rendues entières. Ce choix permettra également de tester le modèle étudié en comparant ses performances prédictives avec celles du modèle Arma déjà disponible pour chaque série d'indicateurs. Enfin, en plus des résultats d'ajustement et de prédiction sur les données corrigées, dont les principaux figurent dans le rapport et les résultats détaillés en annexe (2), nous présenterons en annexe (3) l'ajustement sur les données initiales brutes.

---

(e1) " Constitution d'indicateurs météorologiques locaux adaptés à l'insécurité routière "  
O.E.S.T / C.E.M.S juin 1991

(e2) " Modélisation Arima d'une dizaine d'indicateurs de l'insécurité routière "  
O.E.S.T / C.E.M.S janvier 1992

---

## 1 AJUSTEMENT DU MODELE POISSONNIEN

### 1.1 Le modèle paramétrique

Le modèle que nous allons mettre en place appartient à la large classe des **modèles markoviens** qui ont fait l'objet de nombreuses études théoriques et pratiques, tout particulièrement en traitement d'image. Ces modèles fournissent une description locale et cohérente des phénomènes intervenant dans un système aléatoire d'éléments interagissant. Cette approche est particulièrement naturelle dans le cadre spatial, pour la modélisation d'une image par exemple, et peut être étendue au cadre temporel où l'on cherche à modéliser une ou plusieurs séries temporelles interagissant. Notons que la structure de ces modèles impose de considérer la ou les séries étudiées dans leur ensemble comme une unique observation, nous pourrions donc en parler comme d'un réseau de variables interagissant.

Nous disposons ici de l'observation d'un réseau régulier puisque les variables étudiées sont observées à intervalles de temps réguliers. On note  $X$  la variable associée au réseau  $R$  pris dans sa totalité et  $X_i$  la variable associée au site  $n^o i$  du réseau, plus généralement on pourra noter  $X_{i,t}$  la variable correspondant à l'observation à l'instant  $t$  de la série  $n^o i$ .

Dans la classe des modèles markoviens on peut définir une structure probabiliste cohérente pour  $X$  uniquement en choisissant convenablement les probabilités conditionnelles décrivant les comportements de dépendance entre les variables  $X_i$ .

Puisque nous étudions des variables discrètes associées à l'apparition d'événements rares, il est naturel de supposer ces lois conditionnelles poissonniennes. Nous avons donc choisi d'étudier le modèle "auto-poissonnien" qui appartient à une sous classe des modèles markoviens, et pour lequel on définit :

$$\text{Loi} ( X_i / Y_i=y_i ) \equiv \text{Poisson} ( \lambda_i (y_i) )$$

avec

$$\lambda_i (y_i) = \exp [ \alpha_i + \sum_{j \in V_i} \beta_{ij} x_j ]$$

On note  $Y_i = R^{-i}$  l'ensemble de toutes les variables du réseau privé de  $X_i$  et où  $V_i$  à définir est appelé le voisinage de la variable  $X_i$  et comprend toutes les variables  $X_j$  dont  $X_i$  dépend conditionnellement, on a donc :  $\text{Loi} ( X_i / Y_i ) = \text{Loi} ( X_i / V_i )$ . Sans définir précisément la loi jointe nous pouvons dire qu'elle est de la forme :

$$P( X=x ) = \frac{1}{Z} \exp [ Q(x) ]$$

avec

$$Q(x) = \sum_i \{ \alpha_i x_i - \log(x_i!) \} + \sum_i \sum_{j \in V_i} \beta_{ij} x_i x_j$$

Plus précisément, on définit en fait cette loi sur l'ensemble des points intérieurs au réseau, c'est à dire associés à des variables dont on connaît le voisinage. Notons que le paramètre de liaison  $\beta_{ij}$  est non nul si et seulement si les lois de  $X_i$  et  $X_j$  sont conditionnellement dépendantes, et remarquons également l'écriture symétrique des liaisons dans  $Q(x)$ .

Nous allons donc étudier l'ajustement du modèle "auto-poissonnien", tout d'abord sur la série simple associée à la fréquence des accidents notée  $(A_t)$ , puis en y associant une variable indicatrice de la gravité, sur la série double des accidents et tués  $(A_t, T_t)$ .

Pour la série simple  $(A_t)$  par exemple, rappelons simplement que le modèle se définit avec :

$$\text{Loi} ( A_t / V_{A_t} ) \equiv \text{Poisson} ( \lambda_{A_t} )$$

où  $\lambda_{A_t}$  s'exprime en fonction des variables liées à  $A_t$ . Il reste donc à définir le système des liaisons entre variables qui caractérise la structure probabiliste, ainsi que l'intégration de variables explicatives au modèle.

## 1.2 Le choix des liaisons

Les liaisons définissent les relations de voisinage entre les variables et donc les lois conditionnelles. Le voisinage de  $A_t$  par exemple est noté  $V_{A_t}$  et défini par les variables liées à  $A_t$  dans la loi conditionnelle, variables qui peuvent appartenir à la même série ou à une série "parallèle",  $(T_t)$  par exemple lorsque l'on modélise le couple  $(A_t, T_t)$   $1 \leq t \leq T$ . Ces relations internes ou relations de voisinage définissent ce que l'on appellera le **modèle de base**, où l'on ne prend pas encore en compte l'influence éventuelle de variables explicatives.

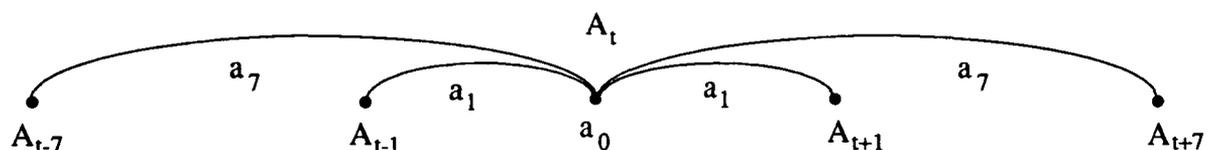
Remarquons que les liaisons temporelles sur une série peuvent permettre d'intégrer des facteurs externes à variation plus ou moins lente, tels qu'un effet météorologique saisonnier ou un effet calendaire hebdomadaire, ou plus généralement une tendance. Elles permettent également d'envisager un calcul de prédiction.

Pour choisir un modèle simple, une étude fruste conduit à ne prendre en compte que des liaisons temporelles sur la série  $(A_t)$ , en s'inspirant par exemple du modèle Arma, des périodicités détectées, et l'unique liaison instantanée entre  $A_t$  et  $T_t$  pour la série double.

Rappelons le modèle Arma obtenu pour la série journalière des accidents corrigés de la météo et du calendrier exceptionnel sur la période de 6 ans, avec  $B$  l'opérateur retard et  $W_t$  le bruit blanc gaussien d'innovation :

$$(1 - B^7)(1 - 0,91B) A_t = (1 - 0,80B)(1 - 0,98B^7) W_t$$

• Pour la **série simple**  $(A_t)$ , on étudiera essentiellement le modèle de base décrit par le graphe suivant :

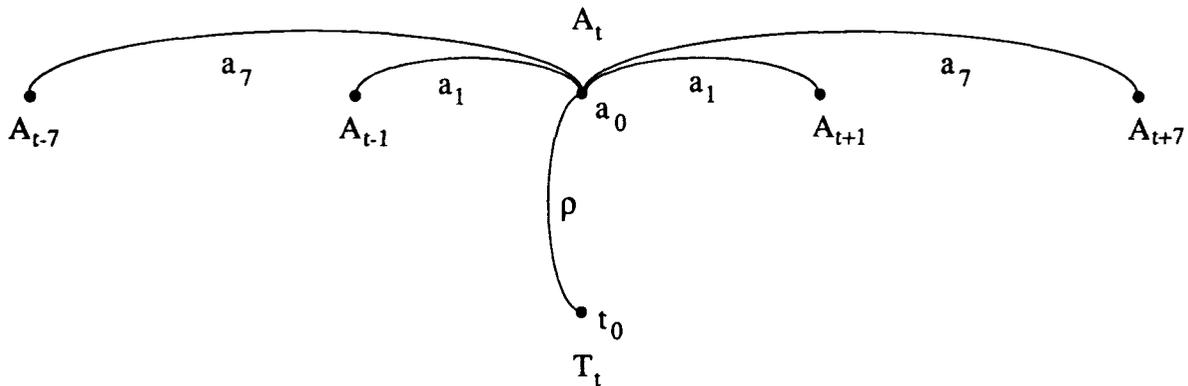


où Loi  $(A_t / V_{A_t}) \equiv$  Poisson  $(\lambda_{A_t})$

$$\lambda_{A_t} = \exp [ a_0 + a_1(A_{t-1} + A_{t+1}) + a_7(A_{t-7} + A_{t+7}) ]$$

Nous examinerons également à titre indicatif des modèles faisant intervenir d'autres liaisons temporelles sur la série  $(A_t)$ , une liaison d'ordre 2 par exemple reliant  $A_t$  aux variables  $A_{t-2}$  et  $A_{t+2}$  et décrite par le paramètre  $a_2$ . Plus généralement on notera  $a_i$  le paramètre reliant  $A_t$  aux variables  $A_{t-i}$  et  $A_{t+i}$  et on parlera du paramètre d'ordre  $i$ .

• Et pour la série double  $(A_t, T_t)$ , on étudiera essentiellement le modèle de base décrit par le graphe suivant :



où  $\text{Loi} ( A_t / V_{A_t} ) \equiv \text{Poisson} ( \lambda_{A_t} )$   
 $\lambda_{A_t} = \exp [ a_0 + a_1(A_{t-1}+A_{t+1}) + a_7(A_{t-7}+A_{t+7}) + \rho T_t ]$

et  $\text{Loi} ( T_t / V_{T_t} ) \equiv \text{Poisson} ( \lambda_{T_t} )$   
 $\lambda_{T_t} = \exp [ t_0 + \rho A_t ]$

### 1.3 Intégration de variables explicatives

Pour intégrer des variables de "contexte" au modèle, rappelons que nous supposons que le contexte n'a pas d'influence sur la structure du modèle, c'est à dire sur la fonctionnelle définissant la loi jointe, mais qu'il peut modifier quantitativement les paramètres du modèle de base; on peut dire que seul les "poids" des liaisons peuvent être perturbés.

Pour une variable Z qualitative prenant deux modalités et dont l'influence s'exerce sur  $a_0$  par exemple, on posera :  $a_0 (Z) = a_{0(1)} 1_{(Z="1")} + a_{0(2)} 1_{(Z="2")}$  .

Les variables explicatives sont supposées connues. Lorsque les modalités que peut prendre une variable qualitative Z sont trop nombreuses, on cherchera une agrégation appropriée afin de réduire la dimension du paramètre de base. Il faut alors choisir des modes d'agrégation adaptés à chaque paramètre. Le modèle finalisé doit prendre en compte des paramètres significatifs en nombre raisonnable. Les liaisons internes, ainsi que les facteurs externes et leur intervention sur les paramètres du modèle de base, doivent donc être choisis avec soin.

A ce niveau, la seule variable explicative pouvant raisonnablement être intégrée au modèle est la **variable type de jour** qui prend au maximum 7 modalités (L,M,M,J,V,S,D). Il faut donc définir la nature de son influence sur les différents paramètres du modèle soit  $a_0, t_0, \rho, a_1$  et  $a_7$  .

Notons que l'on distingue trois types de paramètres :  $a_0$  et  $t_0$  apportent une information instantanée propre à une série,  $\rho$  est un paramètre de liaison instantané entre les deux séries,  $a_1$  et  $a_7$  caractérisent des liaisons temporelles. Bien que leur interprétation en terme d'espérance, corrélation et autocorrélation soit abusive, ces caractéristiques fournissent des critères simples de comparaison. Les tableaux suivants donnent les estimations de ces caractéristiques calculées pour chaque type de jour ( toujours sur les données corrigées ).

• Variables accidents (A) et tués (T) : moyennes et variances

	global	L	M	M	J	V	S	D
E(A)	48,26	45,87	42,66	42,56	43,85	53,54	54,64	54,42
Var(A)	127,04	93,44	74,13	82,56	86,44	105,58	118,75	134,27
$\sigma(A)$	11,27	9,67	8,61	9,09	9,30	10,28	10,90	11,59
E(T)	7,60	6,98	6,54	6,19	6,68	8,35	9,07	9,25
Var(T)	13,18	9,78	8,72	9,07	10,22	13,97	14,73	16,02
$\sigma(T)$	3,63	3,13	2,95	3,01	3,20	3,74	3,84	4,00

La demi longueur  $2\sigma / \sqrt{N}$  des intervalles de confiance pour E(A) est de l'ordre de 1,14 et pour E(T) de l'ordre de 0,3 ( pour des calculs empiriques avec N = 311 ou 312 ).

• Variables accidents (A) et tués (T) : corrélations

	L	M	Me	J	V	S	D
Corr (A,T)	0,50	0,44	0,42	0,51	0,42	0,51	0,49

• Variable accidents (A) : autocorrélations d'ordre 1 à 7

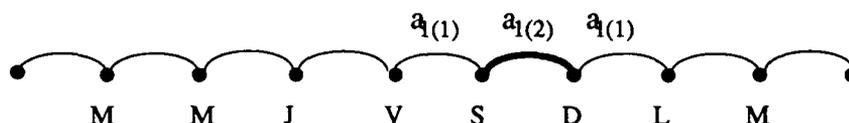
Jour et Jour +	1	2	3	4	5	6	7
L	<b>0,42</b>	0,28	0,34	0,25	0,42	0,30	<b>0,38</b>
M	0,26	0,30	0,06	0,25	0,25	0,28	0,11
Me	0,34	0,18	0,25	0,21	0,29	0,15	0,22
J	0,29	0,30	0,32	0,36	0,22	0,33	0,22
V	0,35	0,26	0,28	0,19	0,19	0,14	0,11
S	<b>0,44</b>	<b>0,45</b>	0,28	0,27	0,27	0,32	<b>0,30</b>
D	<b>0,39</b>	0,23	0,19	0,25	0,15	0,29	<b>0,39</b>

Sur la première ligne par exemple, 0,42 correspond à la corrélation (ordre1) entre le lundi et le mardi suivant, 0,38 à la corrélation (ordre7) entre le lundi et le lundi suivant.

( Les autocorrélations pour la variable T peuvent être négligées )

A chaque paramètre on peut faire correspondre une **décomposition particulière** fonction du type de jour. Sur le premier tableau, on observe un écart significatif entre les moyennes du L,M,M,J et celles du V,S,D pour les variables A et T. Pour les paramètres  $a_0$  et  $t_0$  on choisira donc une décomposition déduite de l'agrégation en deux classes (semaine, week-end) de types de jour. Plus généralement il semble que ce type de décomposition semaine, week-end peut être utilisé pour tous les paramètres en y apportant quelques nuances. On pourra par exemple choisir parmi les décompositions suivantes: LMMJVD / S , LMMJV / SD , LMMJ / VSD .

Notons que quelques précautions doivent être prises lorsque l'on veut décomposer un paramètre de liaison non instantanée, les variables liées pouvant ne pas appartenir à la même classe de type de jour. Ce problème ne se pose pas pour le paramètre  $a_7$  qui relie deux jours de même type. Pour les autres comme par exemple  $a_1$ , on choisit de rattacher le paramètre à la classe d'ordre inférieur (on introduit un ordre sur les classes définissant le type de jour par exemple "semaine"=1, "week-end"=2). Ceci permet de particulariser une liaison précise, la décomposition LMMJV / SD pour le paramètre  $a_1$  donnera par exemple la décomposition suivante :



## 2 ESTIMATION ET COMPARAISON DE MODELES

Chaque modèle est donc décrit par son graphe de liaison et par les décompositions choisies pour chacun des paramètres. Ainsi dans les modules programmés :

Le graphe est caractérisé par une matrice "Lien" où l'on codifie les liaisons retenues, par exemple pour le modèle simple sur  $(A_t)$  défini avec les trois paramètres  $a_0, a_1, a_7$  numérotés de 1 à 3 et caractérisant les liaisons d'ordre 0,1 et 7 :

t	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7
$A_t$	3	0	0	0	0	0	2	1	2	0	0	0	0	0	3

Et les décompositions des paramètres sont définies à l'aide des codes numériques suivants avec la convention précédente :

- 0 → aucune décomposition
- 1 → LMMJVD / S
- 2 → LMMJV / SD
- 3 → LMMJ / VSD

...

Rappelons que dans le module d'estimation on utilise la méthode classique de maximisation de la pseudo-vraisemblance fonction uniquement des probabilités conditionnelles de la vraie loi soit:

$$L^*(\theta) = \prod_{i \in D} P_{\theta}(x_i / V_i)$$

Cette méthode est généralement choisie lorsque la loi jointe est insuffisamment précisée ou que certaines fonctionnelles nécessaires au calcul de l'estimateur du maximum de vraisemblance ne sont pas elles même calculables. On peut remarquer que lorsque  $D$  est restreint à un sous ensemble de points non voisins, on retrouve la vraie vraisemblance et donc le calcul de l'estimateur du maximum de vraisemblance.

Notons que  $D$  désigne ici l'ensemble des sites intérieurs au réseau, c'est à dire les sites pour lesquels le voisinage est observable et rappelons que l'on observe une unique réalisation du réseau constitué d'une ou deux séries pour cette application.

Le module d'estimation permet donc d'approcher les valeurs des coordonnées de  $\theta$  qui maximisent cette pseudo-vraisemblance, et de calculer, en plus de ces estimateurs, leurs intervalles de confiance (de l'ordre de deux écart-types), enfin quelques indicateurs destinés à comparer les modèles étudiés.

## 2.1 La moyenne conditionnelle comme critère de comparaison

A voisinage connu, on peut approcher la valeur de  $A_t$  par exemple par sa moyenne conditionnelle :

$$\lambda_{A_t} = E(A_t / V_{A_t})$$

qui représente la meilleure approximation au sens des moindres carrés, c'est à dire l'approximation qui minimise l'écart quadratique connaissant tout le voisinage de  $A_t$ . Notons que pour ce modèle qui suppose les lois conditionnelles poissonniennes,  $\lambda_{A_t}$  représente aussi la valeur la plus probable connaissant tout le voisinage de  $A_t$ .

Les modèles peuvent donc être comparés en utilisant un critère mesurant la qualité de l'interpolation de  $A_t$  par  $\lambda_{A_t}$  et l'on note :  $A_t = \lambda_{A_t} + e_{A_t}$  où  $e_{A_t} = (A_t - \lambda_{A_t})$  représente l'erreur d'approximation. Cette approximation permet d'interpoler une donnée manquante et peut également servir pour mesurer l'effet d'un phénomène occasionnel, un phénomène climatique exceptionnel par exemple. La valeur ajustée donne alors une idée de la valeur qu'aurait du prendre la variable si ce phénomène n'avait pas eu lieu.

Pour chacun des modèles étudiés nous donnons une évaluation de la moyenne de cet écart, de la moyenne de l'écart absolu et de sa variance. Ces trois indicateurs, qui mesurent la qualité de l'approximation à voisinage connu, sont présentés dans un tableau. Ils sont calculés

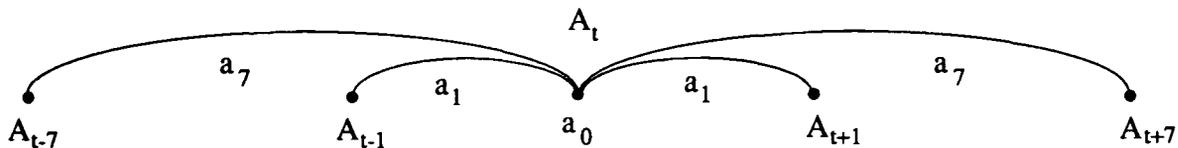
globalement et, pour plus de précision, conditionnellement aux variables explicatives, soit ici le type de jour. Tous ces résultats numériques sont ensuite complétés par des courbes de suivi permettant de visualiser en chaque site l'approximation par la moyenne conditionnelle à voisinage connu.

Remarquons également que l'utilisation de la propriété du modèle poissonnien sur la variance soit :  $\text{Var}(A_t / V_{A_t}) = \lambda_{A_t}$ , permet de définir un intervalle de confiance autour de la moyenne conditionnelle. On pourra utiliser ces intervalles de confiance pour détecter les observations aberrantes, ou pour valider le modèle en comptabilisant le pourcentage d'observations appartenant à cet intervalle. Notons que pour  $\lambda_{A_t} \geq 30$  l'approximation par une loi normale est acceptée et l'on peut examiner à chaque instant si

$$A_t \in \text{IC}(A_t) = ] \lambda_{A_t} - 2 \sqrt{\lambda_{A_t}} ; \lambda_{A_t} + 2 \sqrt{\lambda_{A_t}} [ .$$

## 2.2 Résultats pour la série simple ( $A_t$ )

Le modèle de base étudié est représenté par le graphe :



Ce choix en partie intuitif semble validé lorsque l'on compare les ordres de grandeur des paramètres de liaison, dans un premier temps non décomposés suivant le type de jour de semaine. En particulier pour chacun des 7 modèles ne prenant en compte qu'un seul paramètre de liaison, l'estimation a donné les résultats suivants :

$a_1 = 0,0066$	[ 0,0061 ; 0,0071 ]
$a_2 = 0,0032$	[ 0,0027 ; 0,0038 ]
$a_3 = - 0,0004$	[ - 0,0008 ; 0,0001 ]
$a_4 = - 0,0003$	[ - 0,0008 ; 0,0002 ]
$a_5 = 0,0030$	[ 0,0024 ; 0,0036 ]
$a_6 = 0,0056$	[ 0,0051 ; 0,0061 ]
$a_7 = 0,0059$	[ 0,0054 ; 0,0063 ]

En choisissant les liaisons associées aux paramètres les plus importants, soit ici  $a_1$  et  $a_7$  par exemple, on définit un modèle plus sensible aux variations de voisinage.

Notons que différents modèles prenant en compte deux paramètres de liaisons parmi les sept ont également été examinés. Le modèle ( $a_0, a_1, a_7$ ) a encore été validé pour sa sensibilité et par comparaison des indicateurs mesurant la qualité de l'ajustement à voisinage connu.

Parmi les **décompositions** des paramètres  $a_i$  testées pour ce modèle de base, il apparaît que les 3 modèles où l'on choisit de décomposer tour à tour un des paramètres en fonction du type de jour ( décomposition semaine, week-end ) sont très semblables quant à la qualité de l'ajustement. Précisons toutefois que les estimations des paramètres décomposés sont **significativement distinctes**.

Notons également que les décompositions multiples n'apportent aucun gain et que les estimations de  $a_1$  et  $a_7$  décomposés en fonction du type de jour ne sont même plus **significativement distinctes**.

Nous présentons donc uniquement ici les résultats pour le modèle (  $a_0, a_1, a_7$  ) sans décomposition de ces paramètres suivant le type de jour. Puis, pour ce même modèle de base où l'on choisit la décomposition n°3 pour  $a_0$  soit LMMJ / VSD pour les deux classes de type de jour, et sans décomposer  $a_1$  et  $a_7$  :

• **modèle** :  $a_0(0), a_1(0), a_7(0)$

**estimation** :

$a_0$	3.0574	[ 3.0059 ; 3.1090 ]
$a_1$	0.0045	[ 0.0039 ; 0.0050 ]
$a_7$	0.0039	[ 0.0034 ; 0.0044 ]

**écarts sur**  $e = (A_t - \lambda_{A_t})$  :

	global	L	M	Me	J	V	S	D
E ( e )	<b>0,00</b>	-1,46	-1,80	-1,40	-2,46	2,86	1,34	2,92
E   e	<b>7,34</b>	6,70	6,52	6,58	7,43	7,88	7,89	8,36
Var ( e )	<b>86,19</b>	71,16	63,05	67,83	79,08	91,41	97,95	101,00

% de variance globale expliquée : **32.15**

• **modèle** :  $a_0(3), a_1(0), a_7(0)$

**estimation** :

$a_0(1)$	3.1815	[ 3.1282 ; 3.2348 ]
$a_0(2)$	3.3037	[ 3.2411 ; 3.3663 ]
$a_1$	0.0043	[ 0.0038 ; 0.0049 ]
$a_7$	0.0022	[ 0.0017 ; 0.0028 ]

**écarts sur**  $e = (A_t - \lambda_{A_t})$  :

	global	L	M	Me	J	V	S	D
E ( e )	<b>0,00</b>	0,78	-0,26	0,09	-0,62	0,35	-0,95	0,61
E   e	<b>7,07</b>	6,70	6,16	6,42	6,96	7,46	7,72	8,09
Var ( e )	<b>80,06</b>	71,41	60,61	68,16	75,70	87,88	92,54	101,64

% de variance globale expliquée : **36.98**

Quelques résultats déduits des modèles estimés permettent de donner une idée des variations prises en compte par ces modèles :

En notant  $y$  la moyenne du nombre d'accidents et en posant :  $y = e^{[\alpha + \beta x]}$ , une variation du voisinage noté  $x$  ou du paramètre  $\alpha$  entraîne une variation  $\Delta y$  sur  $y$  :

$$\Delta y = y_2 - y_1 = y_1 ( e^{\Delta\alpha + \beta \Delta x} - 1 ).$$

On quantifie alors les résultats suivants :

1- La décomposition du paramètre  $a_0$  entraîne à elle seule, c'est à dire en supposant le voisinage inchangé, une augmentation de 13% du nombre moyen d'accidents le week-end par rapport au nombre moyen d'accidents en semaine.

2- Pour être plus précis, si l'on suppose qu'un jour de week-end, tel que le samedi par exemple, tout le voisinage a augmenté de 10 accidents par rapport à celui d'un jour de semaine tel que le mercredi par exemple, l'augmentation du nombre moyen d'accidents entre ces deux jours est de l'ordre de :

18 % pour le premier modèle, sans décomposition, puisque :  $e^{(2(a_1+a_7)).10} - 1 = 0.18$  ;

28 % pour le second modèle puisque :  $e^{[a_0(2)-a_0(1)] + 2(a_1+a_7).10} - 1 = 0.28$  .

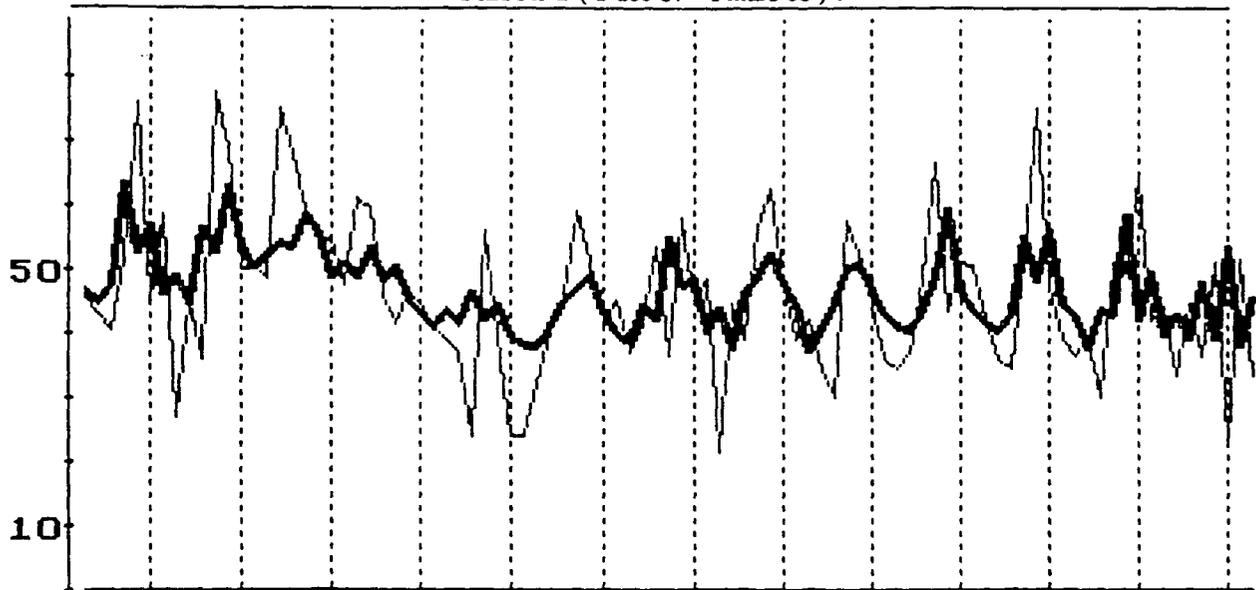
**Les courbes qui suivent** permettent d'apprécier sur un tronçon de trois mois l'apport de cette décomposition. Précisons que les courbes en gras représentent la moyenne conditionnelle calculée à voisinage connu comparativement aux courbes des données à modéliser ( corrigées des effets météorologiques et calendaires exceptionnels ) et que l'on représente en trait fin. Les traits pointillés verticaux, placés toutes les sept observations, correspondent au dimanche. Les intervalles de confiance, non représentés, ont une demi longueur de l'ordre de  $2\sqrt{50} = 14$  pour une moyenne de 50 accidents.

Modèle univarié sur la série simple  $(A_t)$  des données corrigées :

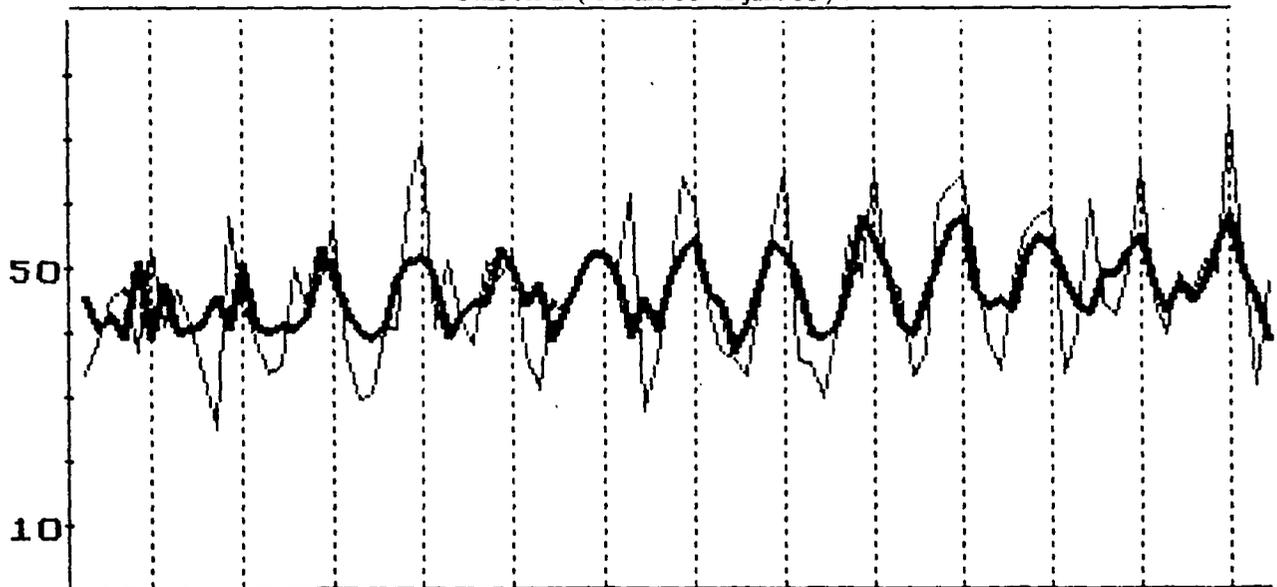
$$a_0(0), a_1(0), a_7(0)$$

Observation  $(A_t)$  et moyenne conditionnelle  $(\lambda_{A_t})$

Saison 1 ( 1 dec 87 - 1 mars 88 ) :



Saison 2 ( 1 mars 88 - 1 juin 88 ) :

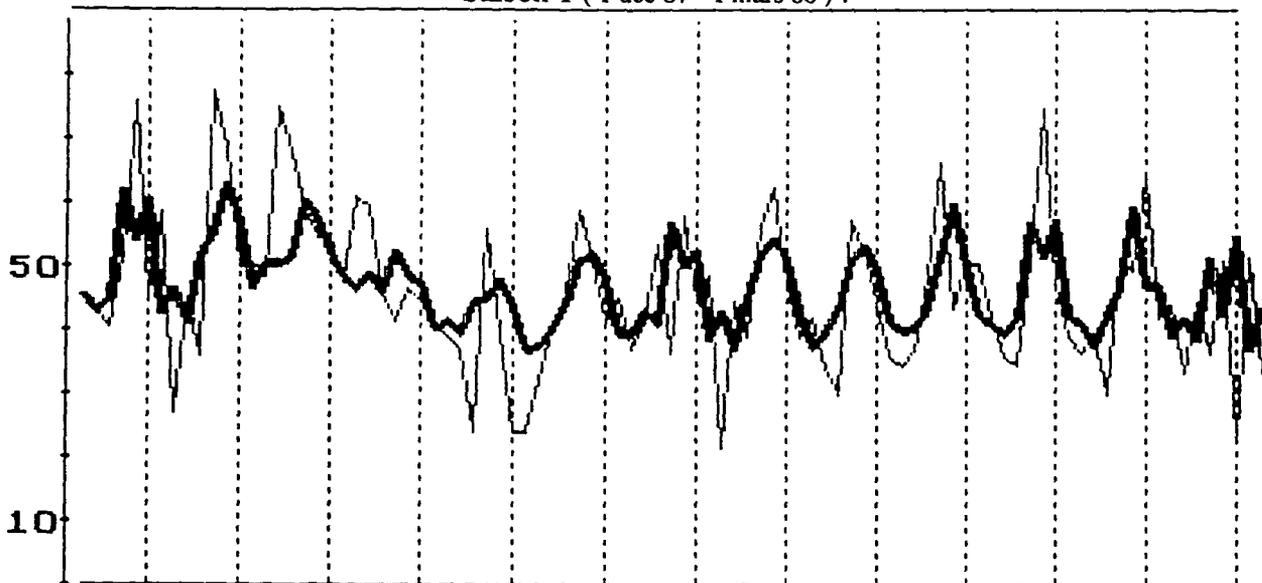


Modèle univarié sur la série simple ( $A_t$ ) des données corrigées :

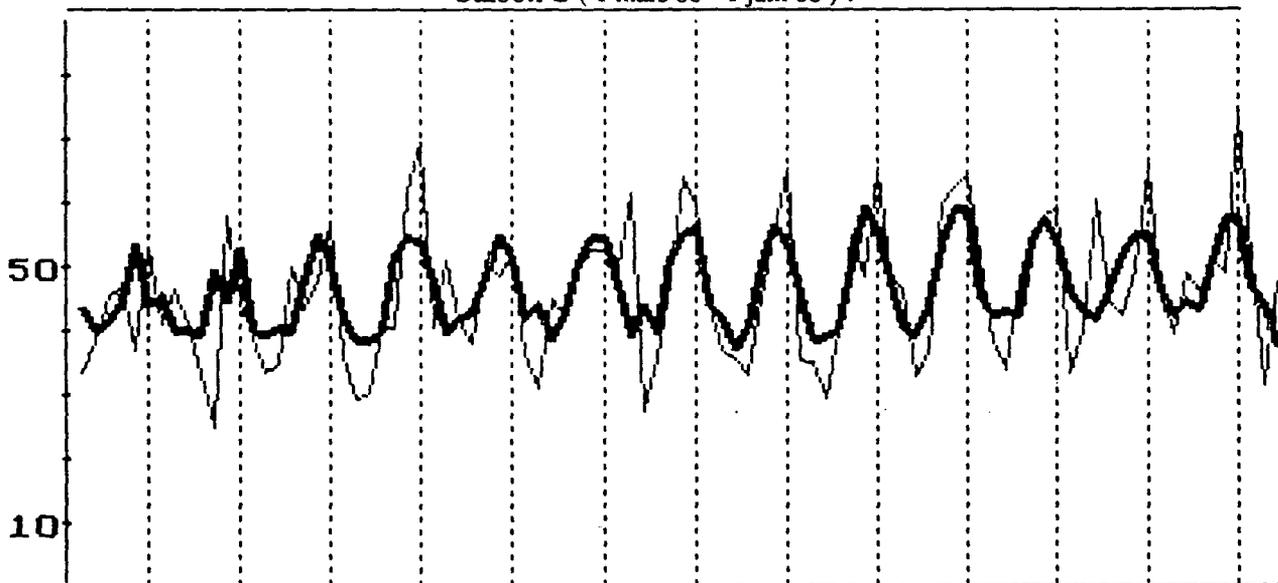
$$a_0(3), a_1(0), a_7(0)$$

Observation ( $A_t$ ) et moyenne conditionnelle ( $\lambda_{A_t}$ )

Saison 1 ( 1 dec 87 - 1 mars 88 ) :

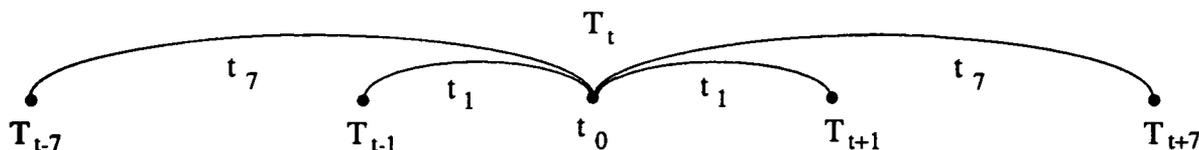


Saison 2 ( 1 mars 88 - 1 juin 88 ) :



## 2.2 Résultats pour la série simple ( $T_t$ )

Le modèle de base étudié est représenté par le graphe :



Nous présentons les résultats pour le modèle ( $t_0, t_1, t_7$ ) sans décomposition de ces paramètres suivant le type de jour. Puis, pour ce même modèle de base où l'on choisit la décomposition n°3 pour  $t_0$  soit LMMJ / VSD pour les deux classes de type de jour, et sans décomposer  $t_1$  et  $t_7$  :

- modèle :**  $t_0(0), t_1(0), t_7(0)$

<b>estimation :</b>	$t_0$	1.5714	[ 1.5009 ; 1.6418 ]
	$t_1$	0.0129	[ 0.0092 ; 0.0166 ]
	$t_7$	0.0166	[ 0.0132 ; 0.0200 ]

**écarts sur**  $e = (T_t - \lambda T_t)$  :

	global	L	M	Me	J	V	S	D
E(e)	0,00	-0,47	-0,59	-0,86	-0,59	0,50	0,85	1,15
E e	2,73	2,50	2,43	2,52	2,56	2,81	2,98	3,31
Var(e)	12,20	9,59	8,69	8,88	9,82	14,33	13,62	16,53

% de variance globale expliquée : 7,45

- modèle :**  $t_0(3), t_1(0), t_7(0)$

<b>estimation :</b>	$t_0(1)$	1.6260	[ 1.5565 ; 1.6955 ]
	$t_0(2)$	1.8591	[ 1.7720 ; 1.9462 ]
	$t_1$	0.0100	[ 0.0065 ; 0.0136 ]
	$t_7$	0.0089	[ 0.0053 ; 0.0125 ]

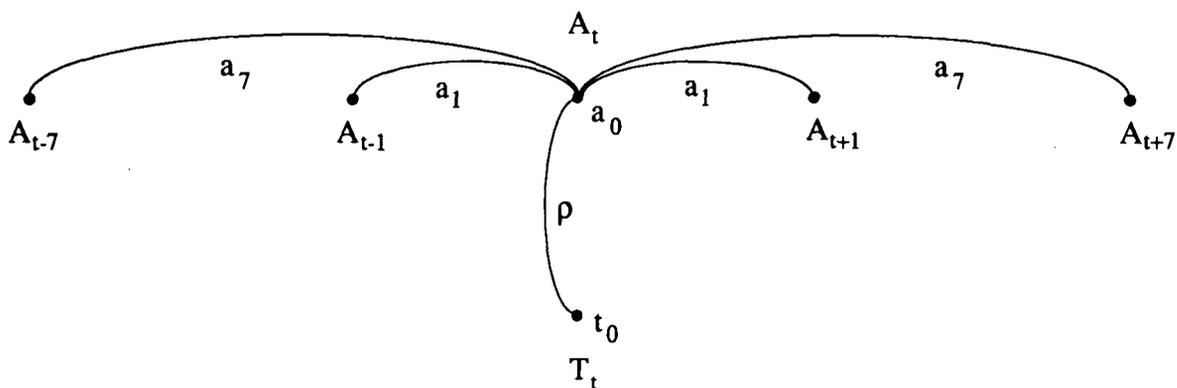
**écarts sur**  $e = (T_t - \lambda T_t)$  :

	global	L	M	Me	J	V	S	D
E(e)	0,00	0,25	-0,01	-0,31	0,06	-0,40	0,06	0,35
E e	2,66	2,42	2,32	2,42	2,47	2,84	2,96	3,19
Var(e)	11,56	9,46	8,57	8,88	9,85	13,88	13,71	16,10

% de variance globale expliquée : 12,36

## 2.3 Résultats pour la série double $(A_t, T_t)$

Pour la série double, on étudie le modèle de base déduit du modèle sur  $(A_t)$  augmenté naturellement de la liaison instantanée entre accidents et tués :



Précisons que l'examen de modèles comportant des liaisons temporelles sur la série  $(T_t)$  a montré que les paramètres associés à ces liaisons n'étaient pas significativement différents de zéro, et que leur présence n'augmentait pas la qualité de l'ajustement du modèle, pas plus que sa capacité prédictive. Sans décomposition des paramètres, le modèle de base a fourni les résultats suivants :

- **modèle :**  $a_0(0), t_0(0), \rho(0), a_1(0), a_7(0)$
- estimation :**

$a_0$	3.0802	[ 3.0320 ; 3.1284 ]
$t_0$	0.9089	[ 0.8329 ; 0.9849 ]
$\rho$	0.0225	[ 0.0211 ; 0.0239 ]
$a_1$	0.0034	[ 0.0029 ; 0.0039 ]
$a_7$	0.0029	[ 0.0025 ; 0.0034 ]

**écarts sur  $e = (A_t - \lambda_{A_t})$  :**

	global	L	M	Me	J	V	S	D
E(e)	0,00	-0,95	-1,52	-0,88	-1,90	2,65	0,82	1,79
E e	6,68	6,10	5,95	6,02	6,41	7,58	7,30	7,43
Var(e)	70,60	57,76	52,63	57,54	62,14	77,90	84,69	83,03

% de variance globale expliquée : 44.43

**écarts sur  $e = (T_t - \lambda_{T_t})$  :**

	global	L	M	Me	J	V	S	D
E(e)	0,00	-0,13	-0,06	-0,41	-0,11	-0,14	0,33	0,52
E e	2,39	2,14	2,06	2,25	2,22	2,65	2,61	2,83
Var(e)	9,32	7,29	6,96	7,61	7,66	11,46	11,28	12,36

% de variance globale expliquée : 29.32

En ce qui concerne ( $T_t$ ) rappelons que pour ce modèle de base la moyenne du nombre de tués connaissant le nombre d'accidents à un jour  $t$  donné s'écrit :

$$E(T_t / A_t) = \lambda_{T_t} = e^{t_0 + \rho A_t}$$

Le paramètre  $\rho$  dans ce sens est donc relié à une notion de gravité. Pour le modèle sans décomposition par exemple une augmentation de de 10 accidents un jour donné entraîne une augmentation de 25% du nombre moyen de tués.

Nous nous sommes donc tout particulièrement intéressés à la décomposition de ce paramètre en fonction des variables explicatives ici disponibles. En se référant au modèle sur la série simple nous avons étudié le modèle où l'on décompose  $a_0$  et  $\rho$  en fonction des deux types de jour semaine(1) et week-end(2) :

• <u>modèle</u> :	$a_0$ (3) , $t_0$ (0) , $\rho$ (3) , $a_1$ (0) , $a_7$ (0)
<u>estimation</u> :	
	$a_{0(1)}$ 3.1760            [ 3.1259 ; 3.2261 ]
	$a_{0(2)}$ 3.2616            [ 3.2033 ; 3.3199 ]
	$t_0$ 0.9732            [ 0.8884 ; 1.0579 ]
	$\rho(1)$ 0.0208            [ 0.0189 ; 0.0226 ]
	$\rho(2)$ 0.0216            [ 0.0201 ; 0.0231 ]
	$a_1$ 0.0033            [ 0.0029 ; 0.0038 ]
	$a_7$ 0.0017            [ 0.0013 ; 0.0022 ]

écarts sur  $e = (A_t - \lambda A_t)$  :

	global	L	M	Me	J	V	S	D
E (e)	0,00	0,64	-0,35	0,27	-0,56	0,68	-0,84	0,15
E   e	6,53	6,14	5,78	5,92	6,02	7,10	7,09	7,45
Var (e)	67,35	58,07	51,58	58,29	59,99	75,24	78,24	83,84

% de variance globale expliquée : 46,98

écarts sur  $e = (T_t - \lambda T_t)$  :

	global	L	M	Me	J	V	S	D
E (e)	0,00	0,03	0,13	-0,22	0,07	-0,41	0,11	0,30
E   e	2,39	2,13	2,05	2,24	2,21	2,65	2,59	2,82
Var (e)	9,27	7,32	6,96	7,66	7,67	11,30	11,08	12,21

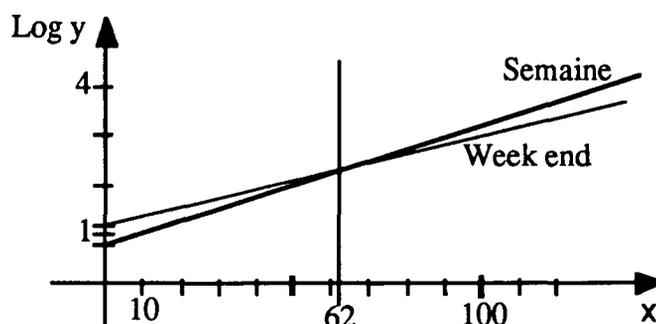
% de variance globale expliquée : 29,68

**Remarque :** nous avons également examiné le modèle où l'on décompose de surcroît le paramètre  $t_0$ . On obtient les résultats suivants :

• <u>modèle :</u>	$a_0(3)$ , $t_0(3)$ , $\rho(3)$ , $a_1(0)$ , $a_7(0)$
<u>estimation :</u>	
	$a_{0(1)}$ 3.1468            [ 3.0961 ; 3.1976 ]
	$a_{0(2)}$ 3.2866            [ 3.2278 ; 3.3453 ]
	$t_{0(1)}$ 0.7935            [ 0.6848 ; 0.9022 ]
	$t_{0(2)}$ 1.1676            [ 1.0411 ; 1.2942 ]
	$\rho(1)$ 0.0244            [ 0.0221 ; 0.0267 ]
	$\rho(2)$ 0.0184            [ 0.0162 ; 0.0206 ]
	$a_1$ 0.0034            [ 0.0029 ; 0.0038 ]
	$a_7$ 0.0018            [ 0.0013 ; 0.0023 ]

et un pourcentage de variance globale expliquée pour ( $A_t$ ) de 47,44 et pour ( $T_t$ ) de 30,06 .

Pour ce modèle qui n'est pas significativement meilleur que le précédent on peut remarquer un phénomène de compensation entre les paramètres  $t_0$  et  $\rho$ . Une augmentation de 10 accidents un jour de semaine entraîne une augmentation de 28% du nombre moyen de tués alors que pour un jour de week-end elle n'est que de 20%. Mais n'oublions pas que le paramètre  $t_0$  est nettement plus important le week-end ce qui compense généralement ce phénomène par ailleurs peut significatif ici. Pour être plus précis, on peut représenter linéairement le log du nombre moyen de tués ( $y$ ) en fonction du nombre d'accidents ( $x$ ), on obtient ici les deux droites semaine et week-end suivantes :



Précisons que ces courbes visualisées pour les trois modèles présentés précédemment ne présentent pas de distinction notable sur ce tronçon, des nuances pourront apparaître dans le cadre de la prédiction.

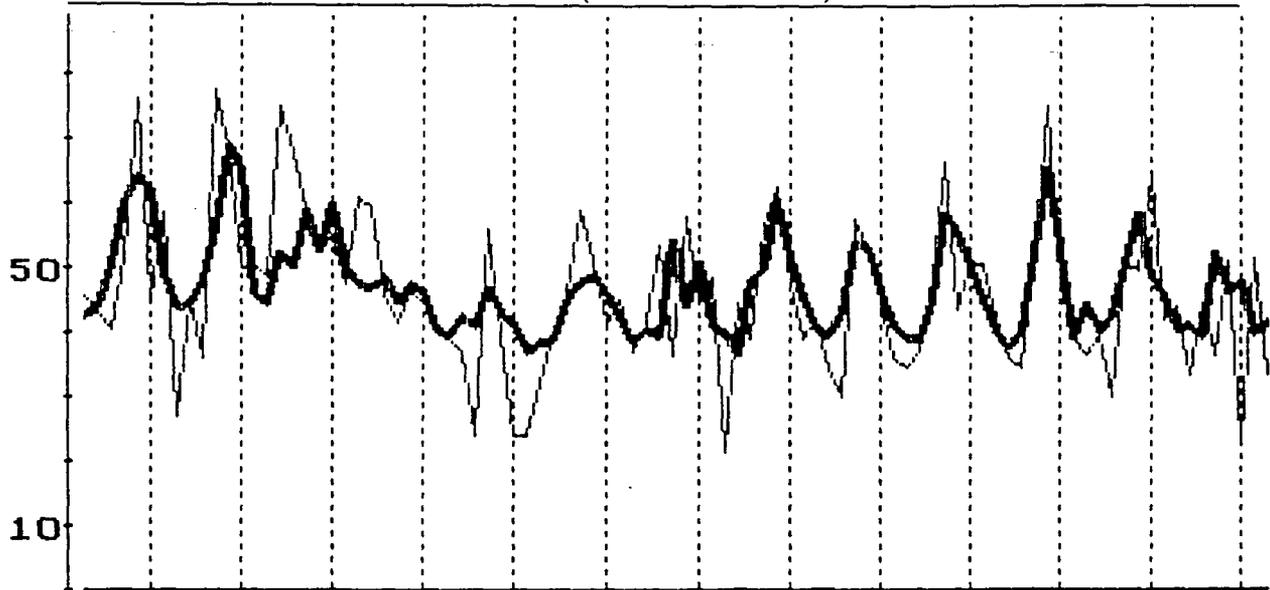
Nous donnons maintenant un aperçu de la qualité de l'ajustement sur le modèle avec décomposition sur  $a_0$  et  $\rho$  uniquement.

Modèle bivarié sur la série double  $(A_t, T_t)$  des données corrigées :

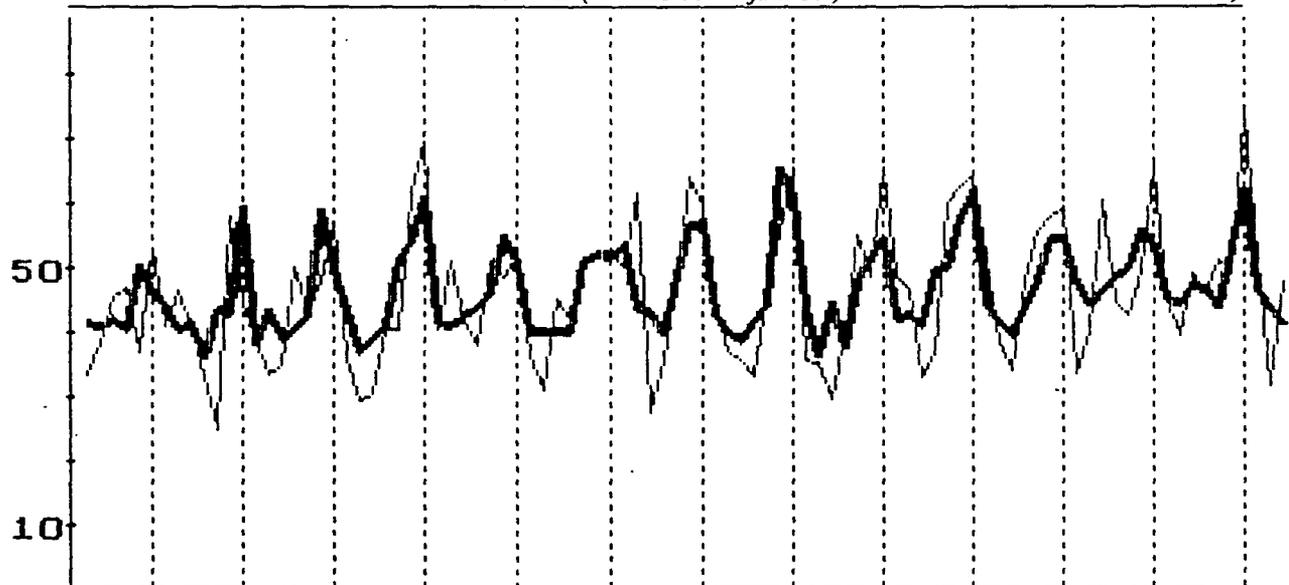
$$a_0(3), t_0(0), \rho(3), a_1(0), a_7(0)$$

Observation  $(A_t)$  et moyenne conditionnelle  $(\lambda_{A_t})$

Saison 1 ( 1 dec 87 - 1 mars 88 ) :



Saison 2 ( 1 mars 88 - 1 juin 88 ) :

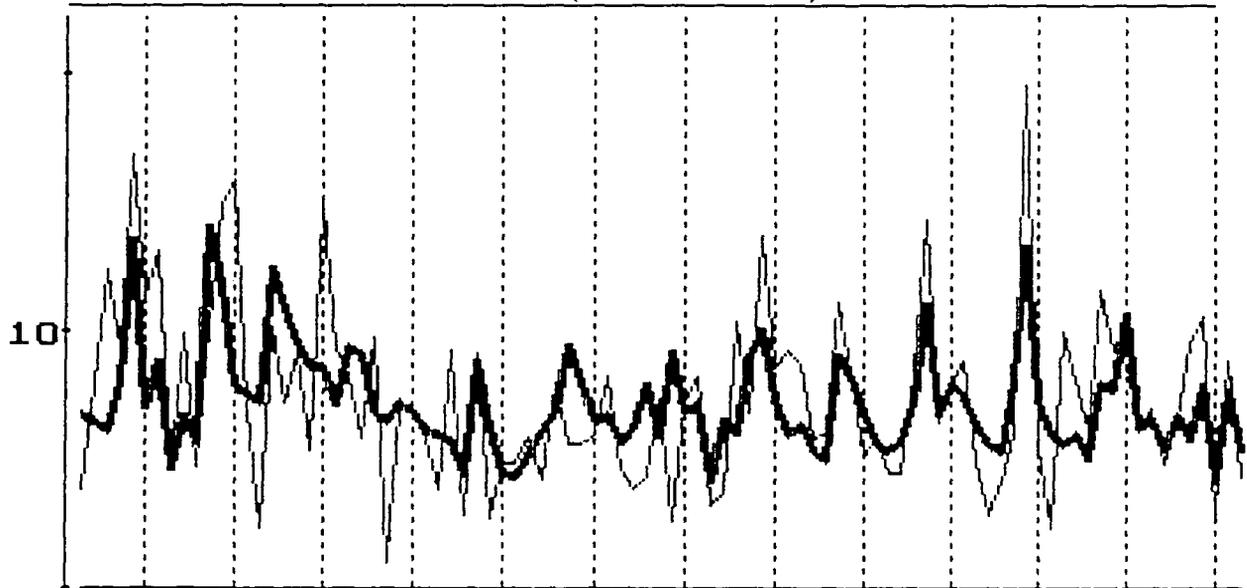


Modèle bivarié sur la série double  $(A_t, T_t)$  des données corrigées :

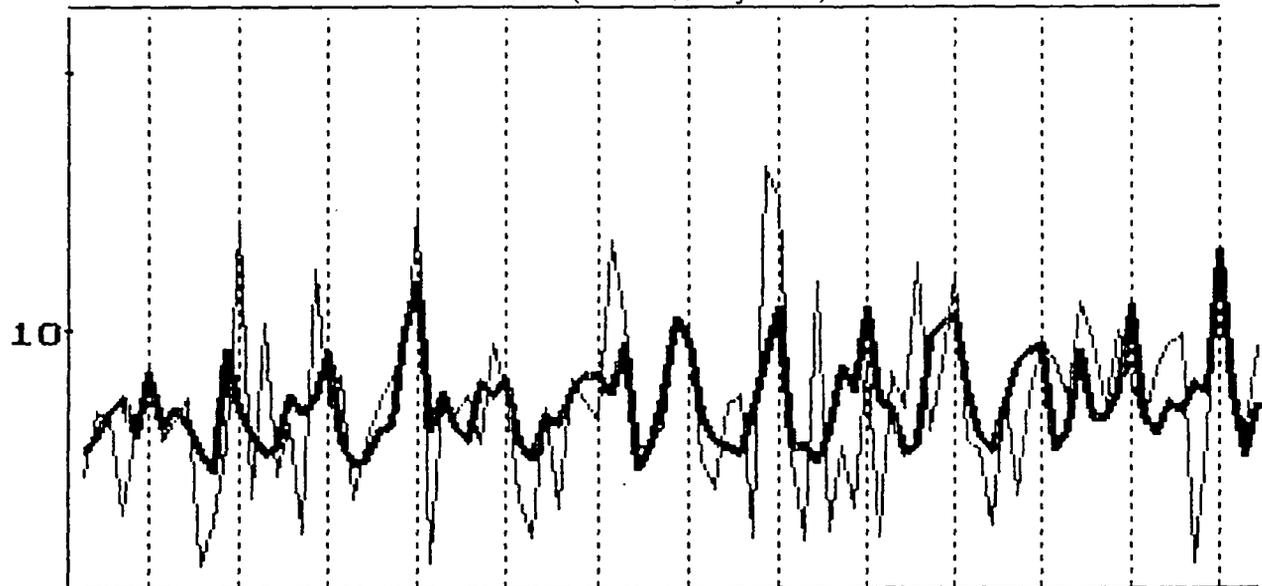
$a_0(3)$ ,  $t_0(0)$ ,  $\rho(3)$ ,  $a_1(0)$ ,  $a_7(0)$

Observation  $(T_t)$  et moyenne conditionnelle  $(\lambda_{T_t})$

Saison 1 ( 1 dec 87 - 1 mars 88 ) :



Saison 2 ( 1 mars 88 - 1 juin 88 ) :



### 3 LA PREDICTION

#### 3.1 Présentation de la méthode

Lorsque l'on ajuste un modèle auto-poissonien sur une ou plusieurs séries, la présence de liaisons temporelles significatives permet d'envisager un calcul de prédiction. Une **prédiction à un pas** peut être obtenue en estimant l'espérance suivante :

$$\hat{X}_t = E ( X_t / F_{t-1} )$$

$\hat{X}_t$  est le prédicteur optimal au sens des moindres carrés, et on note  $F_{t-1}$  le passé des observations antérieures à l'instant  $t$ . Afin d'alléger les notations, nous oublierons le conditionnement fonction des variables explicatives supposées connues.

De par sa **structure symétrique**, le modèle poissonien n'est pas particulièrement adapté à un calcul de prédiction. Ajusté sur une série simple ( $X_t$ ), il fournit l'expression de  $E ( X_t / V_t )$  soit la meilleure estimation de  $X_t$  connaissant son passé et son futur, mais le prédicteur optimal ne peut être calculé directement puisqu'il s'écrit :

$$E ( X_t / F_{t-1} ) = E [ E ( X_t / V_t ) / F_{t-1} ] = E ( \lambda_{X_t} / F_{t-1} )$$

Pour le modèle étudié sur la série univariée ( $A_t$ ) on a par exemple :

$$E ( A_t / F_{t-1} ) = e^{a_0} e^{[ a_1 A_{t-1} + a_7 A_{t-7} ]} E ( e^{[ a_1 A_{t+1} + a_7 A_{t+7} ]} / F_{t-1} )$$

Un estimateur de cette espérance qui fait intervenir des variables du futur fournit donc directement un prédicteur associé au modèle poissonien.

Dans cette méthode nous avons choisi d'utiliser un **modèle approché simple** permettant d'estimer les variables du futur intervenant dans l'espérance conditionnelle.

Rappelons que les données corrigées des effets de la météo et du calendrier exceptionnel ont été obtenues en ajustant sur chacune des séries préalablement corrigées de la météo un modèle Armax où les variables explicatives correspondent aux effets du calendrier "spécial". On dispose donc, pour chaque série de données corrigées, d'un modèle Arma tout particulièrement adapté au calcul de prédiction.

Pour la série ( $A_t$ ) par exemple ce modèle s'écrit :

$$( 1 - B^7 ) ( 1 - 0,91 B ) A_t = ( 1 - 0,8 B ) ( 1 - 0,98 B^7 ) W_t$$

et sa forme développée

$$A_t = 0,91 A_{t-1} + A_{t-7} - 0,91 A_{t-8} + W_t - 0,8 W_{t-1} - 0,98 W_{t-7} + 0,784 W_{t-8}$$

donne le prédicteur gaussien de  $A_t$  à un pas:

$$\text{Pred}_{t-1}^G (A_t) = 0,91 A_{t-1} + A_{t-7} - 0,91 A_{t-8} - 0,8 W_{t-1} - 0,98 W_{t-7} + 0,784 W_{t-8}$$

rappelons que les  $W_t$ , indépendants les uns des autres, représentent l'innovation à l'instant  $t$ .

Une approximation des variables du futur qui interviennent dans l'exponentielle à estimer conditionnellement au passé peut alors être obtenue en utilisant ce modèle approché sous sa forme développée. Pour la série simple  $(A_t)$  par exemple, on approxime  $A_{t+1}$  et  $A_{t+7}$  en utilisant la formule récursive ce qui donne :

$$A_{t+h} = \sum_{i=1}^8 \alpha_{i,h} A_{t-i} + \sum_{i=1}^8 u_{i,h} W_{t-i} + \sum_{j=0}^h v_{j,h} W_{t+j}$$

Les  $W_j$  étant supposés gaussiens, indépendants et de variance  $\sigma_w$  on obtient alors simplement le prédicteur associé au modèle poissonnien ajusté à la série simple :

$$A_t^P = K_A e^{A_t^{(-)}} e^{\left[ \sum_{i=1}^8 \alpha_i A_{t-i} \right]} e^{\left[ \sum_{i=1}^8 u_i W_{t-i} \right]} e^{\frac{1}{2} \left[ \sum_{i=0}^7 v_i^2 \right] \sigma_w^2}$$

Notons que l'utilisation de la formule récursive linéaire permet de conserver la forme exponentielle des espérances conditionnelles.

Pour la **série double**  $(A_t, T_t)$  un travail préliminaire est nécessaire pour prédire  $A_t$  qui consiste à "éliminer" la variable  $T_t$  inconnue en utilisant les propriétés du modèle. On retrouve alors dans l'espérance de  $A_t$  une formule exponentielle ne dépendant plus que des variables appartenant à  $(A_t)$ , ce qui permet d'utiliser la technique précédente pour déterminer le prédicteur de  $A_t$ ,  $T_t$  est ensuite prédit directement à partir de ce prédicteur.

## 3.2 Résultats de la prédiction

Pour quelques uns des modèles étudiés, nous avons estimé la moyenne, la moyenne absolue et la variance de l'écart à la prédiction poissonnienne soit  $(X_t - X_t^P)$ . Parmi ces résultats, qui sont par ailleurs détaillés dans annexe (2), nous ne présentons ici que l'estimation de la variance de l'écart à la prédiction en rappelant à titre de comparaison la variance de l'écart de l'approximation par la moyenne déduite du modèle, qui peut être vue comme une limite des capacités prédictives. On donne également les pourcentages de variance expliquée qui

correspondent aux gains réalisés par rapport à la variance globale qui est aussi la variance de l'écart de l'approximation par la moyenne globale.

Rappelons que pour les **données corrigées** on a :  $\text{Var} (A) = 127,04$   
 $\text{Var} (T) = 13,18$ .

• $(A_t)$ modèle univarié $a_0(0), a_1(0), a_7(0)$ :	$\text{Var} (A_t - \lambda_{A_t}) = 86,19$	32,15 %
	$\text{Var} (A_t - A_t^P) = 86,37$	32,01 %
• $(A_t)$ modèle univarié $a_0(3), a_1(0), a_7(0)$ :	$\text{Var} (A_t - \lambda_{A_t}) = 80,06$	36,98 %
	$\text{Var} (A_t - A_t^P) = 81,75$	35,65 %
• $(T_t)$ modèle univarié $t_0(0), t_1(0), t_7(0)$ :	$\text{Var} (T_t - \lambda_{T_t}) = 12,20$	7,45 %
	$\text{Var} (T_t - T_t^P) = 12,07$	8,46 %
• $(T_t)$ modèle univarié $t_0(3), t_1(0), t_7(0)$ :	$\text{Var} (T_t - \lambda_{T_t}) = 11,56$	12,36 %
	$\text{Var} (T_t - T_t^P) = 11,54$	12,49 %
• $(A_t, T_t)$ modèle bivarié $a_0(0), t_0(0), \rho(0), a_1(0), a_7(0)$ :		
	$\text{Var} (A_t - \lambda_{A_t}) = 70,60$	44,43 %
	$\text{Var} (A_t - A_t^P) = 83,95$	33,92 %
	$\text{Var} (T_t - \lambda_{T_t}) = 9,32$	29,32 %
	$\text{Var} (T_t - T_t^P) = 11,36$	13,83 %
• $(A_t, T_t)$ modèle bivarié $a_0(3), t_0(0), \rho(3), a_1(0), a_7(0)$ :		
	$\text{Var} (A_t - \lambda_{A_t}) = 67,35$	46,98 %
	$\text{Var} (A_t - A_t^P) = 80,91$	36,31 %
	$\text{Var} (T_t - \lambda_{T_t}) = 9,27$	29,68 %
	$\text{Var} (T_t - T_t^P) = 11,14$	15,49 %

Rappelons que le gain réalisé avec l'approximation par la moyenne conditionnelle peut être vu comme le gain maximum réalisable par la prédiction. En ce qui concerne la prédiction de  $A_t$  on observe donc un gain plus proche des limites du modèle pour les modèles univariés que bivariés, ce qui provient bien évidemment de la non connaissance de  $T_t$ . Les calculs de prédiction dans le cas bivarié sont toutefois aussi bons voire meilleurs que dans le cas univarié, ce qui valide la méthode de calcul du prédicteur et surtout l'utilisation du modèle bivarié, par ailleurs meilleur quant à son ajustement aux données observées.

A titre de comparaison précisons que l'on obtient pour le **modèle gaussien** les résultats suivants :

$$\text{Var} ( A_t - A_t^G ) = 79,95 \quad 37,07 \%$$

$$\text{Var} ( T_t - T_t^G ) = 11,30 \quad 14,26 \%$$

Quoiqu'il n'y ait pas d'amélioration sensible comparativement au modèle gaussien, on peut noter que le prédicteur poissonnien calculé sur la variable  $T_t$  pour le modèle bivarié est comparable au prédicteur gaussien alors qu'il ne dépend que du prédicteur poissonnien calculé sur la variable  $A_t$ . Il est aussi important de rappeler que la précision de ce modèle gaussien est en partie liée à la méthode de correction des effets météorologiques et calendaires exceptionnels.

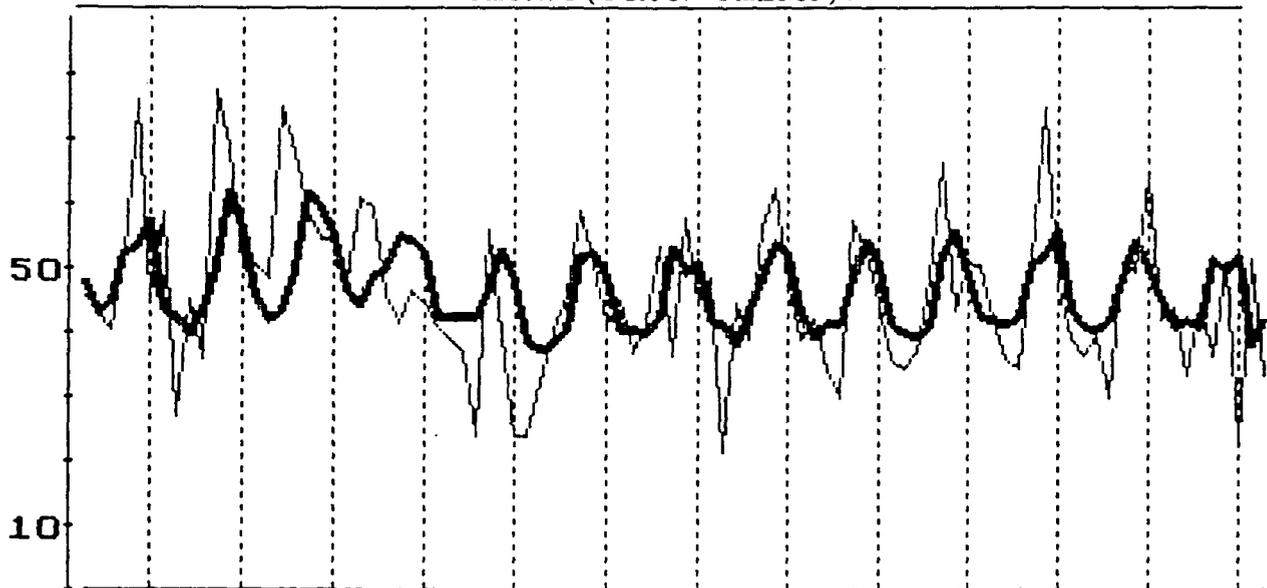
Les courbes qui suivent permettent de visualiser pour le meilleur modèle poissonnien bivarié et sur le même tronçon de trois mois déjà utilisé pour l'ajustement par la moyenne conditionnelle, les prédictions à un pas calculées et représentées par la courbe en gras. Des courbes similaires sont présentées en annexe (4) pour le modèle gaussien, toujours sur les données corrigées.

Modèle bivarié sur la série double  $(A_t, T_t)$  des données corrigées :

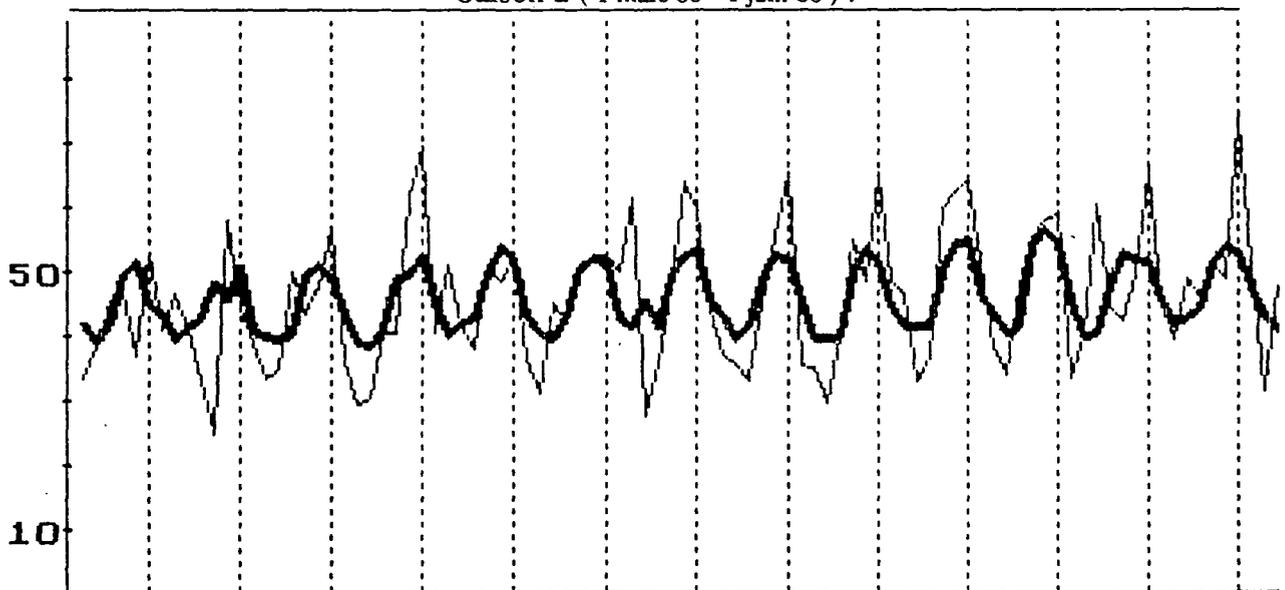
$$a_0 (3) , t_0 (0) , \rho (3) , a_1 (0) , a_7 (0)$$

Observation  $(A_t)$  et prédiction à un pas  $(A_t^P)$

Saison 1 ( 1 dec 87 - 1 mars 88 ) :



Saison 2 ( 1 mars 88 - 1 juin 88 ) :

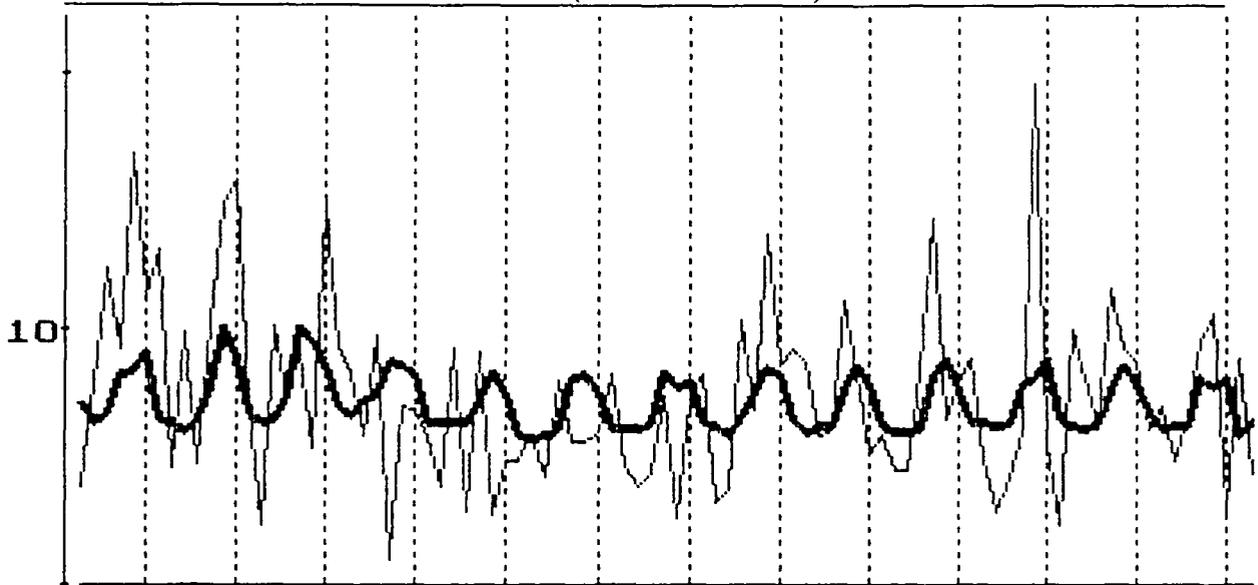


Modèle bivarié sur la série double  $(A_t, T_t)$  des données corrigées :

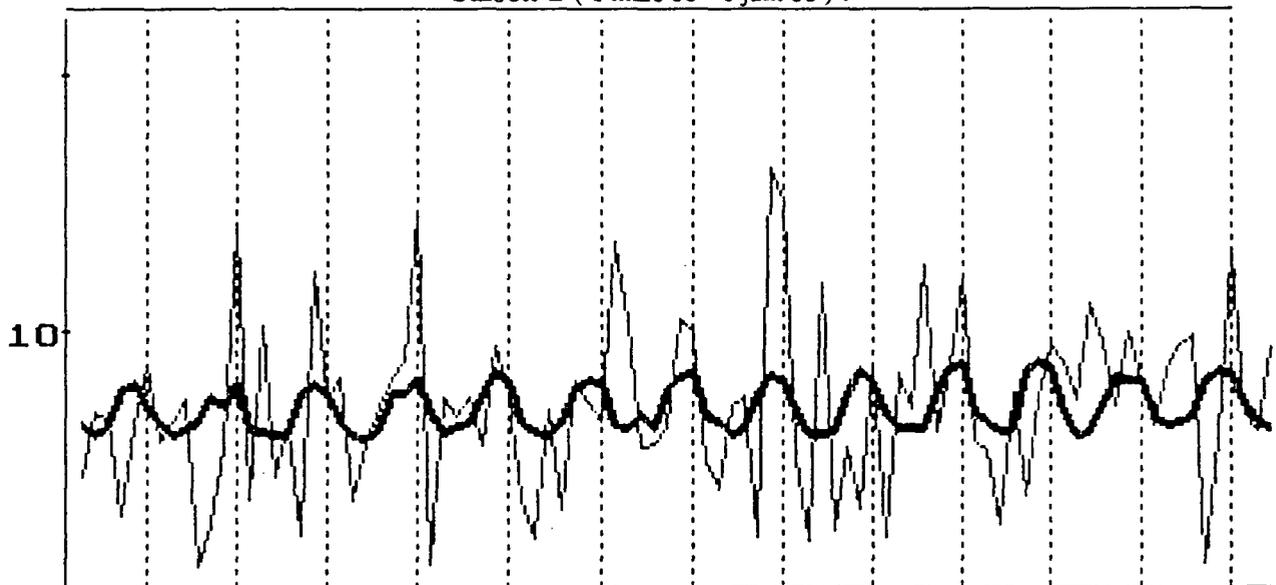
$a_0$  (3) ,  $t_0$  (0) ,  $\rho$  (3) ,  $a_1$  (0) ,  $a_7$  (0)

Observation  $(T_t)$  et prédiction à un pas  $(T_t^P)$

Saison 1 ( 1 dec 87 - 1 mars 88 ) :



Saison 2 ( 1 mars 88 - 1 juin 88 ) :



## 4 CONCLUSIONS

Le modèle poissonnien étudié, qui prend en compte pour décrire une variable tout le contexte environnant (c'est à dire dans le cas général du modèle bivarié, la liaison instantanée avec l'autre variable ainsi que les liaisons avec les variables du passé et du futur si elles existent), est essentiellement un **modèle descriptif**.

La mise en oeuvre d'une méthode de prédiction à un pas (c'est à dire ici à un jour connaissant les valeurs prises par les variables du passé), permet de l'utiliser comme un **modèle de suivi**, pour apprécier un changement dans l'évolution de la série, en comparant les dernières observations aux valeurs prédites.

Rappelons que le modèle est testé sur des séries journalières d'accidents et de tués corrigées des effets des conditions météorologiques et des effets calendaires exceptionnels. La méthode de correction est rappelée dans l'annexe (1).

### 4.1 Sur les performances du modèle poissonnien

Nous allons maintenant faire quelques remarques sur les qualités descriptives et prédictives du modèle poissonnien ajusté sur les séries des indicateurs.

Soit  $\hat{X}$  une valeur approchée de  $X$  déduite du modèle et pour laquelle on a :

$$X = \hat{X} + \varepsilon$$

avec  $\text{Var}(X) = E(X - EX)^2$  et  $\text{Var}(X - \hat{X}) = E(X - \hat{X})^2$ .

On rappelle que l'on mesure la part de variance expliquée par le modèle, que l'on note  $\alpha$ , en posant  $\text{Var}(X - \hat{X}) = \text{Var}(X) - \alpha\% \text{Var}(X)$  soit :

$$\alpha\% = \frac{\text{Var}(X) - \text{Var}(X - \hat{X})}{\text{Var}(X)}$$

Pour décrire la qualité descriptive du modèle on choisit comme valeur approchée  $\hat{X}$  la moyenne conditionnelle  $\lambda_{X_t} = E(X_t / V_t)$  que l'on calcule exactement, c'est à dire par exemple connaissant le passé et le futur de la série pour une modélisation univariée.

Pour décrire la qualité prédictive du modèle, la valeur approchée  $\hat{X}$  est la valeur prédite à un pas soit, avec la méthode de prédiction utilisée ici, une estimation de l'espérance de la moyenne conditionnelle :  $\hat{E}(\lambda_{X_t} / F_{t-1})$ .

Pour commenter les performances du modèle, nous rappelons les résultats des parts de variance expliquées dans le tableau suivant :

	Modèle [ décomposition ]	Modèle Poissonnien		Arma gaussien
		Ajustement	Prédiction	Prédiction
$A_t$	Univarié [ - ]	32,15 %	32,01 %	37,07 %
	Univarié [ $a_0(3)$ ]	36,98 %	35,65 %	
	Bivarié [ - ]	44,43 %	33,92 %	
	Bivarié [ $a_0(3), \rho(3)$ ]	46,98 %	36,31 %	
$T_t$	Univarié [ - ]	7,45 %	8,46 %	14,26 %
	Univarié [ $t_0(3)$ ]	12,36 %	12,49 %	
	Bivarié [ - ]	29,32 %	13,83 %	
	Bivarié [ $a_0(3), \rho(3)$ ]	29,68 %	15,49 %	

### 1. Sur la qualité descriptive :

C'est en prenant en compte la liaison instantanée entre  $A_t$  et  $T_t$  que l'on réalise la plus forte amélioration, tant pour la série ( $A_t$ ) que pour la série ( $T_t$ ). On passe par exemple de 32,15% à 44,43% de variance expliquée pour ( $A_t$ ) en introduisant cette liaison instantanée au modèle univarié. Mais c'est pour la série ( $T_t$ ) que le gain est le plus important puisque l'on passe de 7,45% à 29,32% de variance expliquée.

On soulignera que la prise en compte des liaisons temporelles de la série ( $T_t$ ) présente peu d'intérêt descriptif pour le modèle univarié. Ces liaisons ne sont donc pas conservées dans le modèle bivarié, qui ne prend en compte que les liaisons temporelles sur la série ( $A_t$ ).

On soulignera aussi que les décompositions des paramètres, pour les modèles univariés, ne permettent pas de réaliser un gain aussi important que l'introduction de la liaison instantanée entre  $A_t$  et  $T_t$ . La décomposition des paramètres permet néanmoins de réaliser un gain supplémentaire pour le modèle bivarié.

## 2. Sur la qualité prédictive :

Globalement on observe une perte entre l'ajustement descriptif et prédictif. Pour la série ( $A_t$ ) en particulier, cette perte est plus importante avec le meilleur modèle bivarié ( on passe de 46,98% à 36,31% de variance expliquée ) qu'avec le meilleur modèle univarié ( on passe de 36,98% à 35,65% de variance expliquée ). Ceci s'explique par le fait que la prédiction de  $A_t$ , faite en utilisant le modèle bivarié, fait intervenir la variable supplémentaire  $T_t$  inconnue à l'instant  $t-1$ . Cependant la prédiction de  $A_t$  en utilisant le modèle bivarié est de qualité équivalente à celle que l'on obtient en utilisant le modèle univarié.

On remarquera surtout que l'on obtient avec le modèle bivarié poissonnien des résultats de qualité équivalente à ceux que l'on obtient avec des modèles ARMA gaussiens, pour ( $A_t$ ) comme pour ( $T_t$ ) : il ne faut pas oublier que les outils statistiques et numériques dont on dispose actuellement sont plus performants dans le cadre ARMA. Il doit donc être possible d'améliorer les performances du modèle poissonnien.

## 4.2 Sur les améliorations envisagées

L'utilisation de lois poissonniennes apparaît naturelle et même nécessaire lorsque l'on s'intéresse à des variables de comptage journalières. Pour le modèle bivarié étudié, nous avons vérifié que la loi du nombre de tués sachant le nombre d'accidents a bien une allure poissonnienne, mais nous sommes dans une situation limite en ce qui concerne la loi du nombre d'accidents conditionné par le nombre de tués et par le passé et le futur du nombre d'accidents. Il serait donc plus intéressant d'ajuster ce modèle sur une série d'accidents plus rares. On envisage de travailler par la suite à un **niveau désagrégé**, en étudiant des séries correspondant à des typologies particulières d'accidents : accidents mortels, accidents de jour et de nuit par exemple, et de même typologie pour les tués pour lesquels on peut penser que le relation Accidents-Tués sera encore plus significative.

Par ailleurs, dans les modèles poissonniens que nous avons étudiés, seule la variable explicative type de jour a été prise en compte. On peut envisager d'intégrer d'autres **variables explicatives** dans le modèle poissonnien telles que la saison (découpage été-hiver) par exemple. Pour cela on procédera comme cela a été fait pour la variable type de jour, c'est à dire par décomposition des paramètres du modèle.

Enfin, la méthode de prédiction que nous avons présentée et testée ici n'utilise pas toutes les propriétés du modèle poissonnien. Il est possible de simuler les observations futures du champ connaissant les observations passées jusqu'à la date ( $t-1$ ). Pour les modèles markoviens, on dispose en effet d'un outil efficace de simulation : l'échantillonneur de Gibbs; ce **calcul empirique d'un prédicteur exact** est en cours d'étude et sera présenté dans un cadre universitaire. Cet outil pourrait être utilisé dans ce cas particulier de la modélisation multivariée poissonnienne d'indicateurs de l'insécurité routière.

# ANNEXES

## ANNEXE 1

### Passage des données brutes aux données corrigées

Deux études réalisées par le C.E.M.S - Centre d'Etude sur la Modélisation et les Statistiques - portent sur la **conception d'un modèle gaussien explicatif de l'insécurité routière**, prenant en compte les effets des conditions météorologiques et de la configuration du calendrier sur les indicateurs d'insécurité routière. La première étude (1) est consacrée au traitement des données météorologiques afin d'extraire des variables significatives adaptées aux indicateurs d'insécurité. Elle fournit un mode de correction du seul effet météorologique. La seconde étude (2) finalise le modèle en intégrant en plus les effets calendaires exceptionnels et hebdomadaires. Au total on dénombre 10 à 30 variables, par indicateur et par saison, pour décrire les effets météorologiques, et une quarantaine de variables pour décrire les effets calendaires exceptionnels.

Nous allons maintenant rappeler les modèles retenus dans ces études et la méthode de correction des données.

Pour chaque indicateur d'insécurité  $Y$ , des **modèles de régression** ont été mis en place saison par saison ( $s$ ) :

$$Y_t = \overline{Y}_{(s)} + \sum_i b_i (M_{i,t} - \overline{M}_{i(s)}) + YC_t$$

$\frac{Y_t}{\overline{Y}_{(s)}}$  est l'indicateur à l'instant  $t$   
sa moyenne pour la saison  $s$

$\frac{M_{i,t}}{\overline{M}_{i(s)}}$  la valeur de la variable météo  $M_i$  à l'instant  $t$   
sa moyenne pour la saison  $s$

La partie située dans le signe somme représente l'effet local des variations des variables météo pour la saison considérée et donc l'écart à la moyenne saisonnière dû aux conditions météorologiques.

Un **modèle Arima avec variables exogènes** correspondant aux variables du calendrier exceptionnel permet ensuite d'isoler les différents effets calendaires. Le modèle obtenu pour la série journalière des accidents sur routes nationales par exemple s'écrit :

$$(1 - B^7)(1 - 0.91 B)(YC_t - \sum_i \alpha_i Z_{i,t}) = (1 - 0.80 B)(1 - 0.98 B^7) W_t$$

$YC_t$  est la série des accidents corrigés des effets météo en t  
 $Z_{i,t}$  la valeur de la variable calendaire exceptionnel  $Z_i$  à l'instant t  
 $W_t$  le bruit blanc gaussien d'innovation et B l'opérateur retard.

Plus généralement, la structure Arima retenue pour chacune des séries étudiées est la même soit :

$$(1 - B^7)(1 - a B)(YC_t - \sum_i \alpha_i Z_{i,t}) = (1 - b B)(1 - c B^7) W_t$$

Ces modèles permettent d'obtenir les séries des **indicateurs corrigés des effets météo et calendaires exceptionnels** (notées  $YC_{MCS}$ ).

En posant

$$Y_{M,t} = \sum_i b_i (M_{i,t} - \overline{M_i(s)})$$

et

$$Y_{CS,t} = \sum_i \alpha_i Z_{i,t}$$

l'indicateur corrigé s'écrit

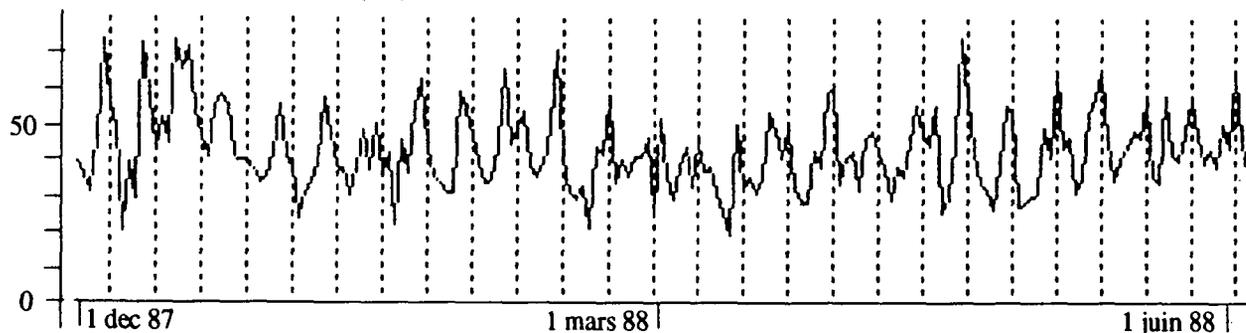
$$YC_{MCS,t} = Y_t - Y_{M,t} - Y_{CS,t}$$

Les **graphiques** qui suivent permettent de visualiser les effets météo et calendaires exceptionnels ainsi déterminés sur quatre saisons consécutives pour les variables  $A_t$  et  $T_t$ , et finalement les séries corrigées utilisées dans la présente étude.

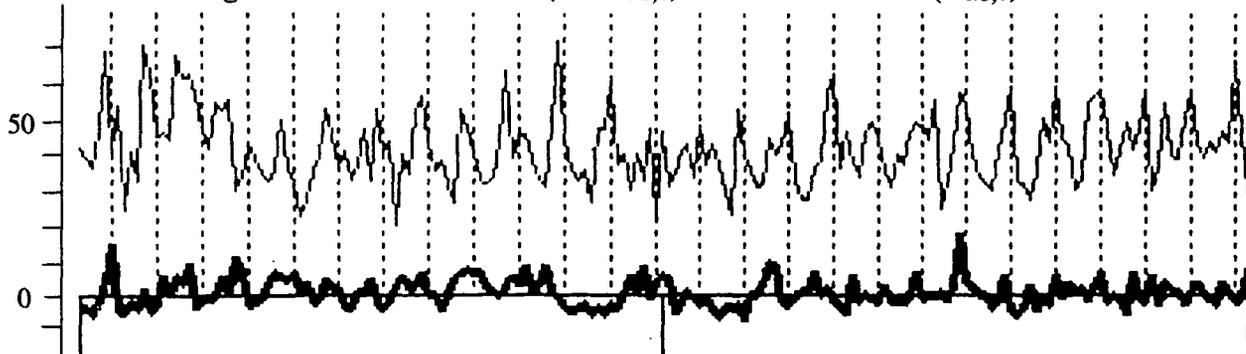
- 
- (1) " Constitution d'indicateurs météorologiques locaux adaptés à l'insécurité routière "  
O.E.S.T / C.E.M.S juin 1991
  - (2) " Modélisation Arima d'une dizaine d'indicateurs de l'insécurité routière "  
O.E.S.T / C.E.M.S janvier 1992

**Passage des données brutes ( $Y_t$ ) aux données corrigées ( $Y_{MCS,t}$ )  
Série des ACCIDENTS - saisons 1 et 2 -**

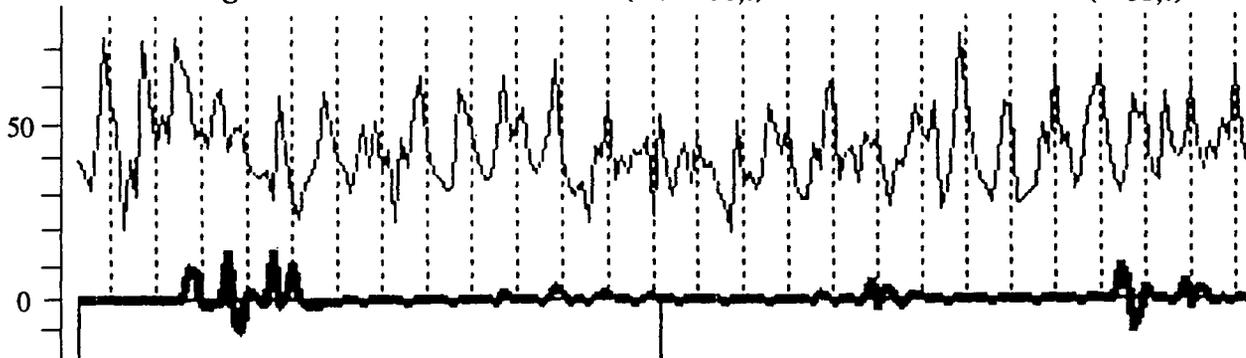
**La série brute initiale ( $Y_t$ )**



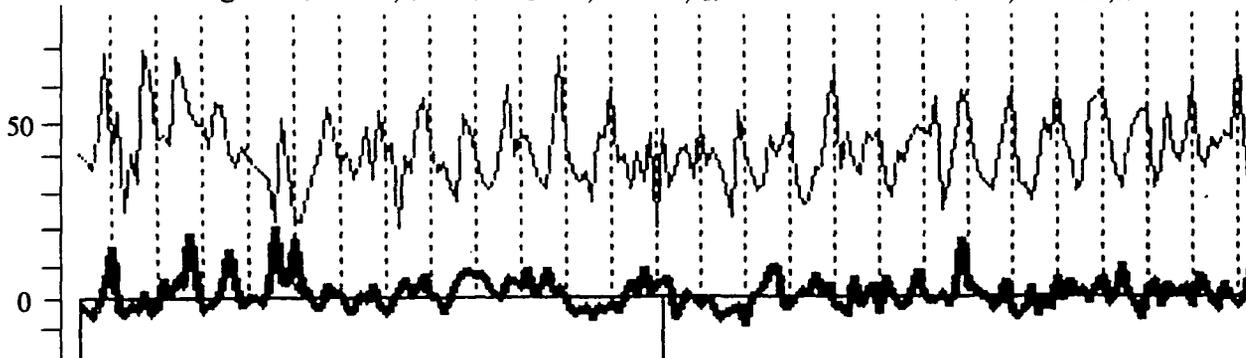
**La série corrigée des effets météo ( $Y_t - Y_{M,t}$ ) et l'effet météo ( $Y_{M,t}$ )**



**La série corrigée des effets calendaires ( $Y_t - Y_{CS,t}$ ) et l'effet calendaire ( $Y_{CS,t}$ )**

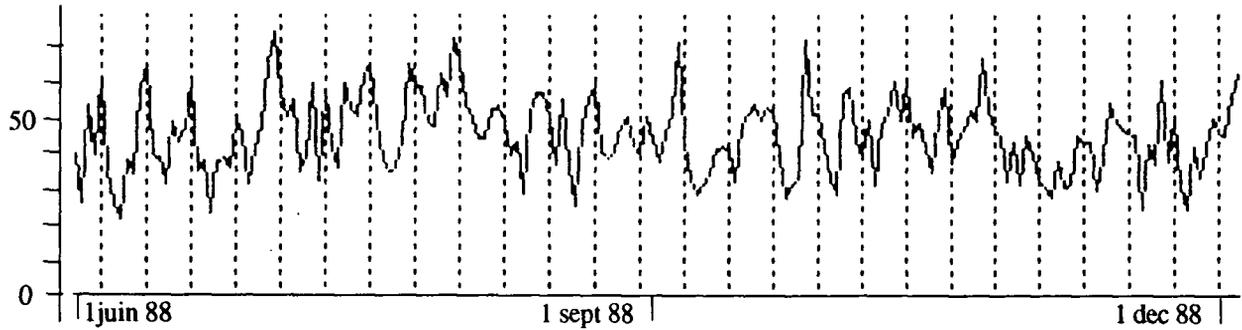


**La série corrigée ( $Y_{MCS,t} = (Y_t - [Y_{M,t} + Y_{CS,t}])$ ) et correction ( $Y_{M,t} + Y_{CS,t}$ )**

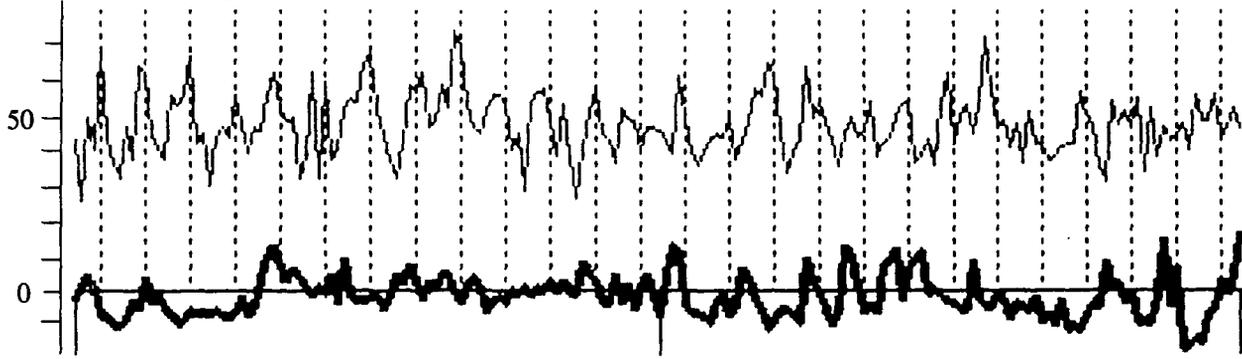


**Passage des données brutes ( $Y_t$ ) aux données corrigées ( $Y_{MCS,t}$ )**  
**Série des ACCIDENTS - saisons 3 et 4 -**

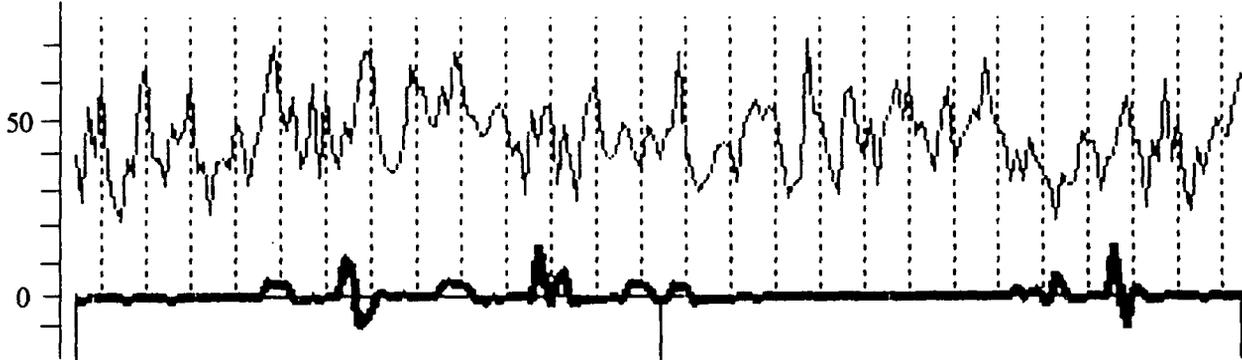
**La série brute initiale ( $Y_t$ )**



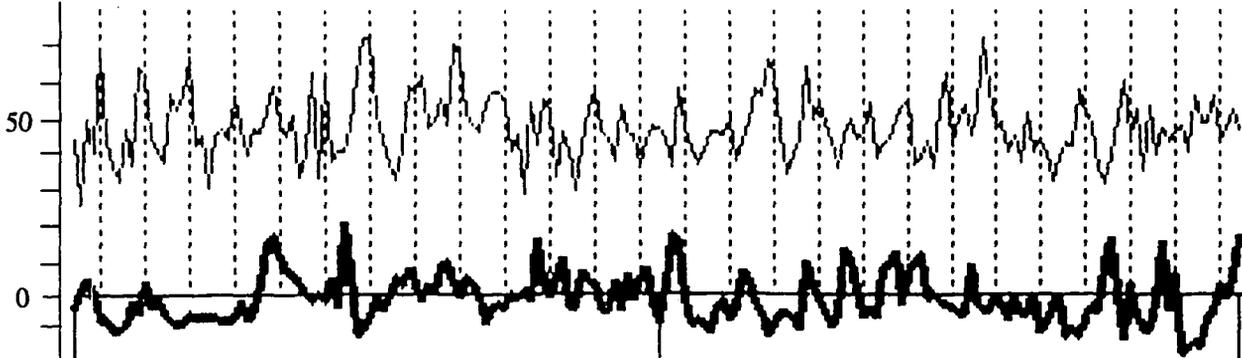
**La série corrigée des effets météo ( $Y_t - Y_{M,t}$ ) et l'effet météo ( $Y_{M,t}$ )**



**La série corrigée des effets calendaires ( $Y_t - Y_{CS,t}$ ) et l'effet calendaire ( $Y_{CS,t}$ )**

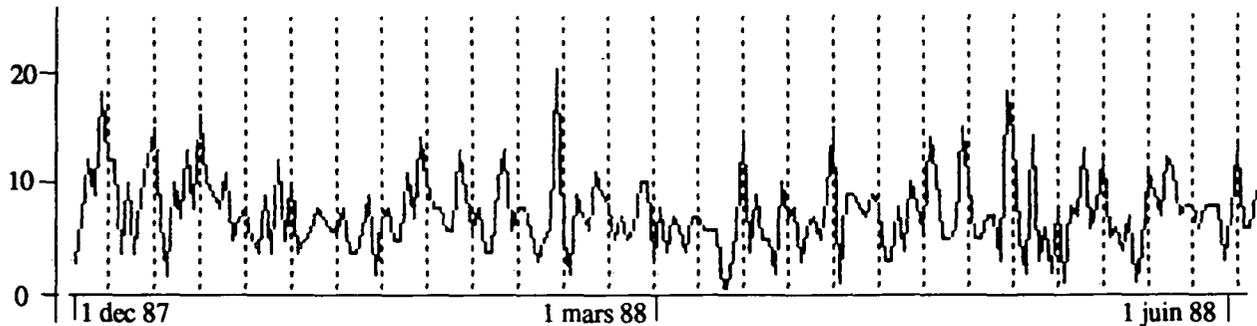


**La série corrigée ( $Y_{MCS,t} = (Y_t - [Y_{M,t} + Y_{CS,t}])$ ) et correction ( $Y_{M,t} + Y_{CS,t}$ )**

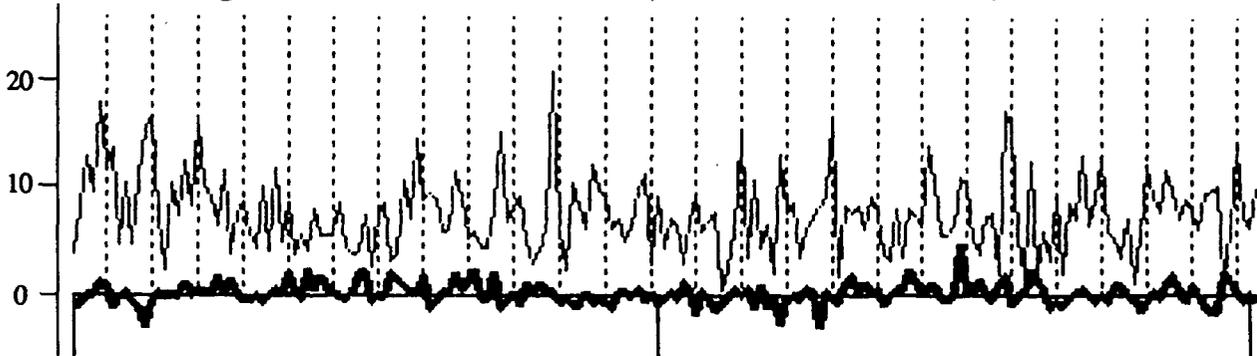


**Passage des données brutes ( $Y_t$ ) aux données corrigées ( $Y_{MCS,t}$ )  
Série des TUES - saisons 1 et 2 -**

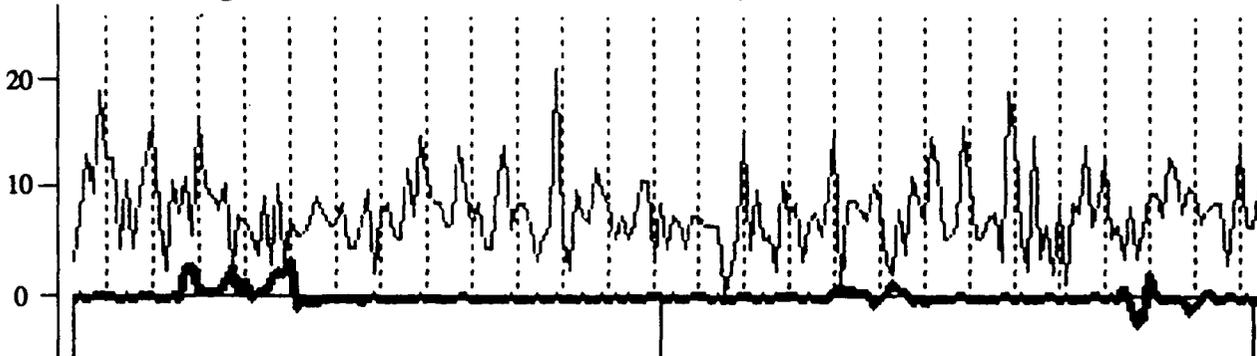
**La série brute initiale ( $Y_t$ )**



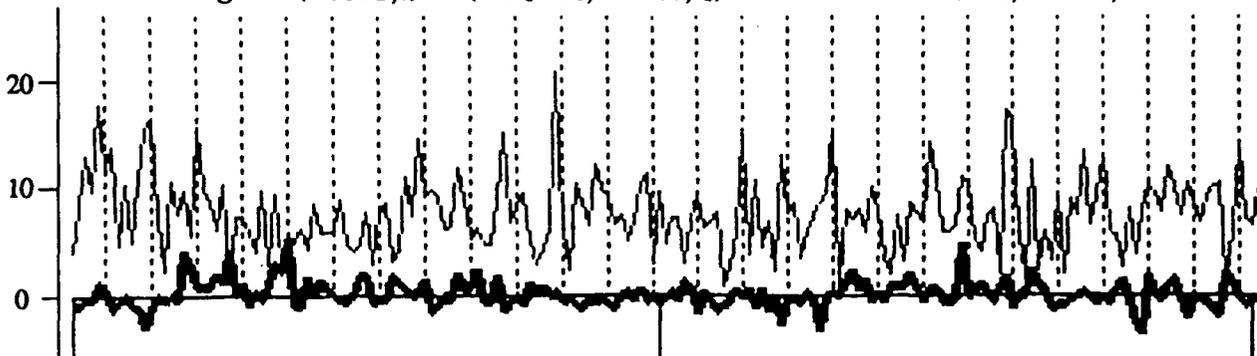
**La série corrigée des effets météo ( $Y_t - Y_{M,t}$ ) et l'effet météo ( $Y_{M,t}$ )**



**La série corrigée des effets calendaires ( $Y_t - Y_{CS,t}$ ) et l'effet calendaire ( $Y_{CS,t}$ )**

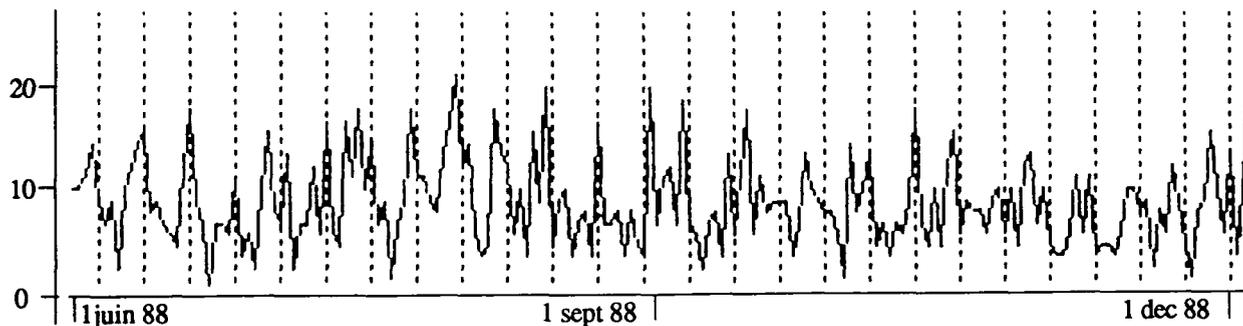


**La série corrigée ( $Y_{MCS,t} = (Y_t - [Y_{M,t} + Y_{CS,t}])$ ) et correction ( $Y_{M,t} + Y_{CS,t}$ )**

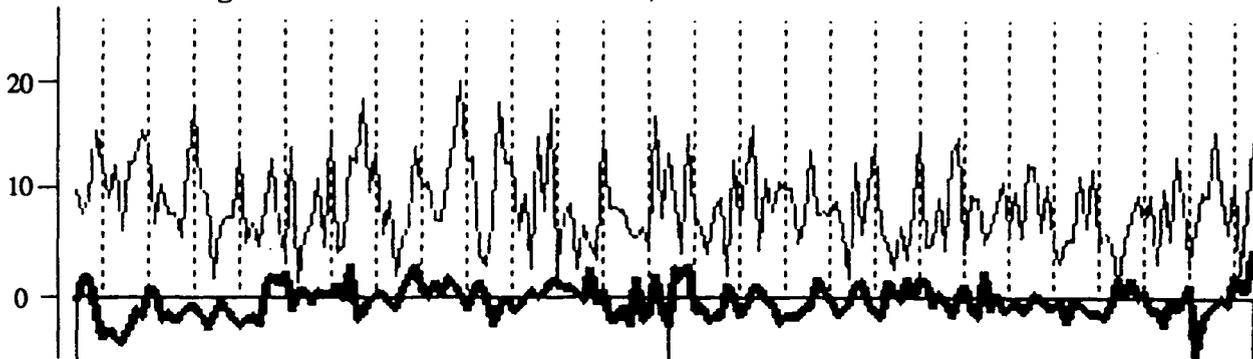


**Passage des données brutes ( $Y_t$ ) aux données corrigées ( $Y_{MCS,t}$ )  
Série des TUES - saisons 3 et 4 -**

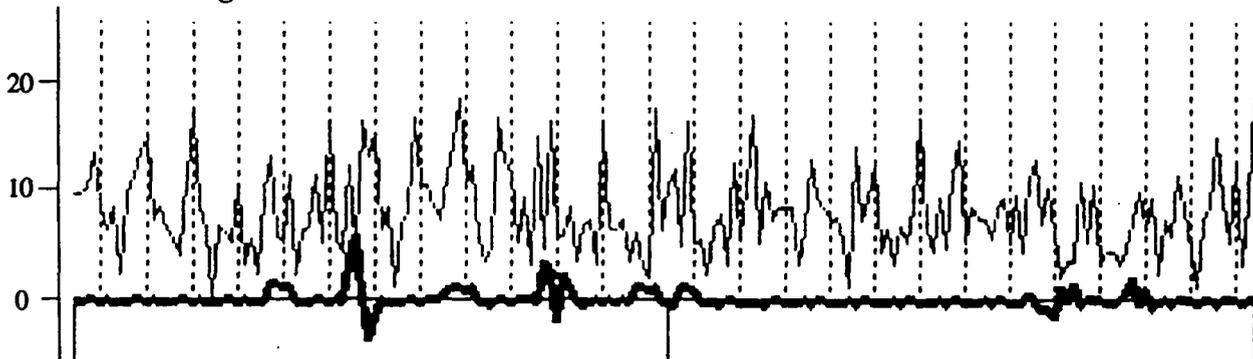
**La série brute initiale ( $Y_t$ )**



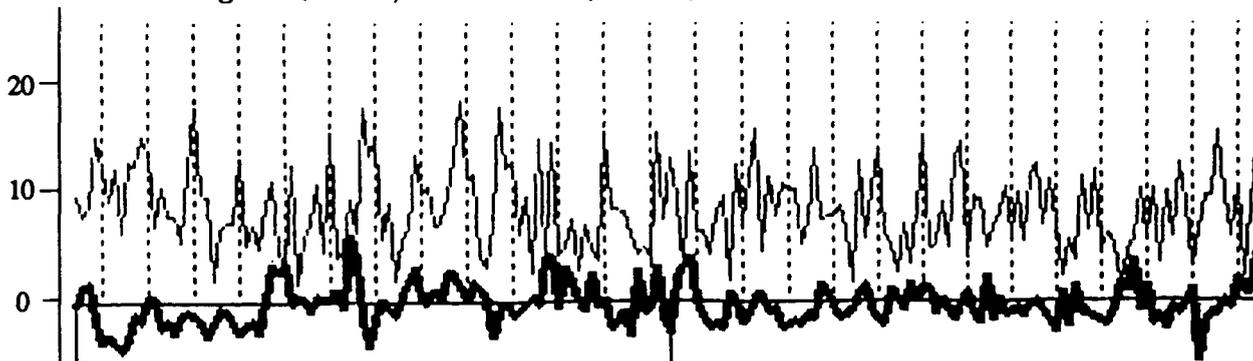
**La série corrigée des effets météo ( $Y_t - Y_{M,t}$ ) et l'effet météo ( $Y_{M,t}$ )**



**La série corrigée des effets calendaires ( $Y_t - Y_{CS,t}$ ) et l'effet calendaire ( $Y_{CS,t}$ )**



**La série corrigée ( $Y_{MCS,t} = (Y_t - [Y_{M,t} + Y_{CS,t}])$ ) et correction ( $Y_{M,t} + Y_{CS,t}$ )**



## ANNEXE 2

### Détail des résultats de la prédiction poissonnienne sur données corrigées

• Modèle univarié :  $a_0(0)$ ,  $a_1(0)$ ,  $a_7(0)$

écart à la moyenne conditionnelle ( capacités prédictives ) :  $e = ( A_t - \lambda A_t )$

	global	L	M	Me	J	V	S	D
E ( e )	0,00	-1,46	-1,80	-1,40	-2,46	2,86	1,34	2,92
E   e	7,34	6,70	6,52	6,58	7,43	7,88	7,89	8,36
Var ( e )	86,19	71,16	63,05	67,83	79,08	91,41	97,95	101,00

% de variance globale expliquée : 32.15

écarts à la prédiction :  $e = ( A_t - A_t^P )$

	global	L	M	Me	J	V	S	D
E ( e )	-0,01	-1,36	-1,77	-1,52	-2,59	2,93	1,36	2,91
E   e	7,35	6,82	6,28	6,72	7,52	8,05	7,71	8,33
Var ( e )	86,37	73,52	58,62	70,18	78,82	93,48	94,75	102,31

% de variance globale expliquée : 32.01

• Modèle univarié :  $a_0(3)$ ,  $a_1(0)$ ,  $a_7(0)$

écart à la moyenne conditionnelle ( capacités prédictives ) :  $e = ( A_t - \lambda A_t )$

	global	L	M	Me	J	V	S	D
E ( e )	0,00	0,78	-0,26	0,09	-0,62	0,35	-0,95	0,61
E   e	7,07	6,70	6,16	6,42	6,96	7,46	7,72	8,09
Var ( e )	80,06	71,41	60,61	68,16	75,70	87,88	92,54	101,64

% de variance globale expliquée : 36.98

écarts à la prédiction :  $e = ( A_t - A_t^P )$

	global	L	M	Me	J	V	S	D
E ( e )	0,00	0,86	-0,26	0,00	-0,69	0,42	-0,96	0,61
E   e	7,14	6,83	6,03	6,55	7,04	7,69	7,74	8,13
Var ( e )	81,75	74,90	58,54	70,41	76,74	91,72	92,76	104,41

% de variance globale expliquée : 35.65

• Modèle bivarié :  $a_0(0)$ ,  $t_0(0)$ ,  $\rho(0)$ ,  $a_1(0)$ ,  $a_7(0)$

écart à la moyenne conditionnelle ( capacités prédictives ) :  $e = (A_t - \lambda A_t)$

	global	L	M	Me	J	V	S	D
E (e)	0,00	-0,95	-1,52	-0,88	-1,90	2,65	0,82	1,79
E   e	6,68	6,10	5,95	6,02	6,41	7,58	7,30	7,43
Var (e)	70,60	57,76	52,63	57,54	62,14	77,90	84,69	83,03

% de variance globale expliquée : 44.43

écarts à la prédiction :  $e = (A_t - A_t^P)$

	global	L	M	Me	J	V	S	D
E (e)	0,03	-0,97	-1,53	-1,42	-2,15	2,54	1,37	2,38
E   e	7,23	6,71	6,13	6,68	7,34	7,88	7,60	8,25
Var (e)	83,95	72,06	57,72	69,93	77,40	92,24	92,64	101,76

% de variance globale expliquée : 33.92

écart à la moyenne conditionnelle ( capacités prédictives ) :  $e = (T_t - \lambda T_t)$

	global	L	M	Me	J	V	S	D
E (e)	0,00	-0,13	-0,06	-0,41	-0,11	-0,14	0,33	0,52
E   e	2,39	2,14	2,06	2,25	2,22	2,65	2,61	2,83
Var (e)	9,32	7,29	6,96	7,61	7,66	11,46	11,28	12,36

% de variance globale expliquée : 29.32

écarts à la prédiction :  $e = (T_t - T_t^P)$

	global	L	M	Me	J	V	S	D
E (e)	0,19	-0,14	-0,19	-0,51	-0,31	0,50	0,79	1,19
E   e	2,63	2,38	2,26	2,47	2,45	2,78	2,89	3,21
Var (e)	11,36	9,07	8,14	8,91	9,51	13,28	13,05	15,12

% de variance globale expliquée : 13.83

- Modèle bivarié :  $a_0(3)$ ,  $t_0(3)$ ,  $\rho(3)$ ,  $a_1(0)$ ,  $a_7(0)$

écart à la moyenne conditionnelle ( capacités prédictives ) :  $e = (A_t - \lambda A_t)$

	global	L	M	Me	J	V	S	D
E(e)	<b>0,00</b>	0,64	-0,35	0,27	-0,56	0,68	-0,84	0,15
E e	<b>6,50</b>	6,14	5,78	5,92	6,02	7,10	7,09	7,45
Var(e)	<b>66,77</b>	58,07	51,58	58,29	59,99	75,24	78,24	83,84

% de variance globale expliquée : **47,44**

écarts à la prédiction :  $e = (A_t - A_t^P)$

	global	L	M	Me	J	V	S	D
E(e)	<b>0,02</b>	0,78	-0,18	-0,07	-0,60	0,36	-0,62	0,47
E e	<b>7,09</b>	6,73	5,95	6,53	7,00	7,64	7,66	8,12
Var(e)	<b>80,79</b>	73,06	57,65	70,19	76,20	91,19	91,75	103,78

% de variance globale expliquée : **36,40**

écart à la moyenne conditionnelle ( capacités prédictives ) :  $e = (T_t - \lambda T_t)$

	global	L	M	Me	J	V	S	D
E(e)	<b>0,00</b>	0,03	0,13	-0,22	0,07	-0,41	0,11	0,30
E e	<b>2,38</b>	2,13	2,05	2,24	2,21	2,65	2,59	2,82
Var(e)	<b>9,22</b>	7,32	6,96	7,66	7,67	11,30	11,08	12,21

% de variance globale expliquée : **30,06**

écarts à la prédiction :  $e = (T_t - T_t^P)$

	global	L	M	Me	J	V	S	D
E(e)	<b>0,15</b>	0,32	0,22	-0,10	0,13	-0,22	0,16	0,55
E e	<b>2,60</b>	2,35	2,23	2,41	2,40	2,78	2,90	3,13
Var(e)	<b>11,12</b>	9,10	8,14	8,90	9,56	13,33	13,28	15,11

% de variance globale expliquée : **15,67**

## ANNEXE 3

### Ajustement du modèle poissonnien sur les données brutes initiales

modèle :  $a_0(3)$ ,  $t_0(0)$ ,  $\rho(3)$ ,  $a_1(0)$ ,  $a_7(0)$

estimation :

$a_0(1)$	3.1258	[ 3.0788 ; 3.1729 ]
$a_0(2)$	3.2185	[ 3.1623 ; 3.2747 ]
$t_0$	0.9891	[ 0.9105 ; 1.0677 ]
$\rho(1)$	0.0200	[ 0.0183 ; 0.0218 ]
$\rho(2)$	0.0212	[ 0.0199 ; 0.0225 ]
$a_1$	0.0044	[ 0.0039 ; 0.0048 ]
$a_7$	0.0012	[ 0.0007 ; 0.0017 ]

écart à la moyenne conditionnelle ( capacités prédictives ) :  $e = ( A_t - \lambda A_t )$

	global	L	M	Me	J	V	S	D
E (e)	<b>0,00</b>	0,90	-0,67	0,51	-0,74	1,13	-1,09	-0,04
E   e	<b>7,11</b>	6,81	6,17	6,54	7,07	7,66	7,86	7,66
Var (e)	<b>81,34</b>	72,38	60,61	70,61	79,40	89,30	97,14	95,44

% de variance globale expliquée : **52,02**

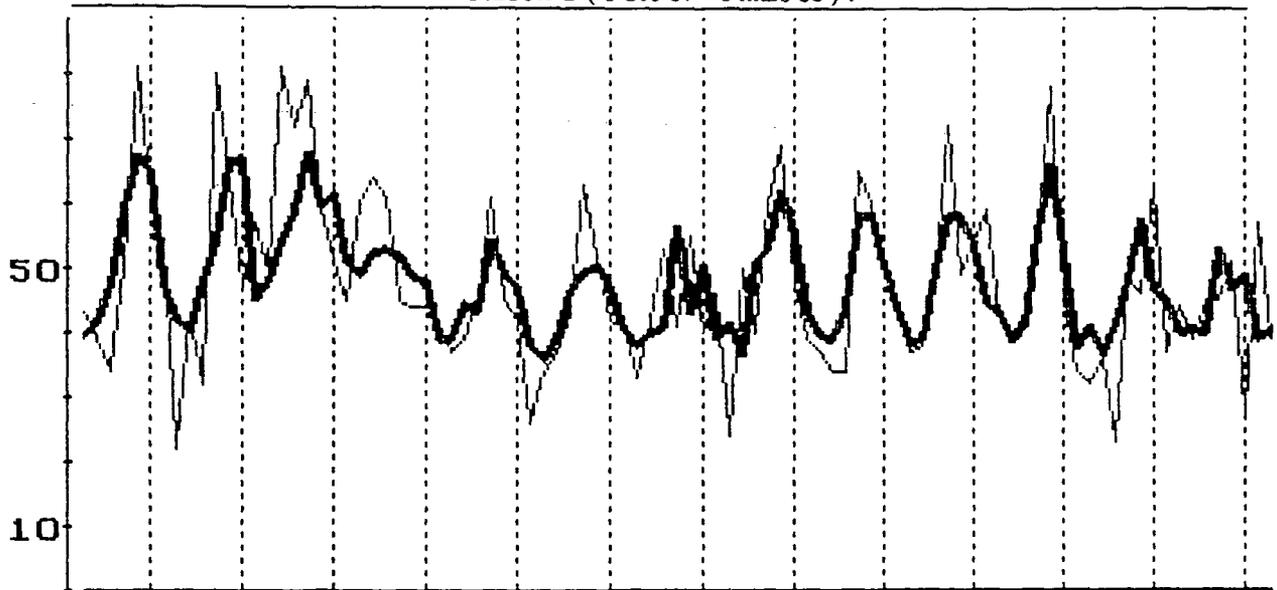
écart à la moyenne conditionnelle ( capacités prédictives ) :  $e = ( T_t - \lambda T_t )$

	global	L	M	Me	J	V	S	D
E (e)	<b>0,00</b>	-0,07	-0,11	-0,29	-0,03	-0,25	0,24	0,50
E   e	<b>2,51</b>	2,32	2,12	2,30	2,27	2,69	2,87	3,01
Var (e)	<b>10,41</b>	8,97	7,38	8,22	8,78	11,78	13,62	13,63

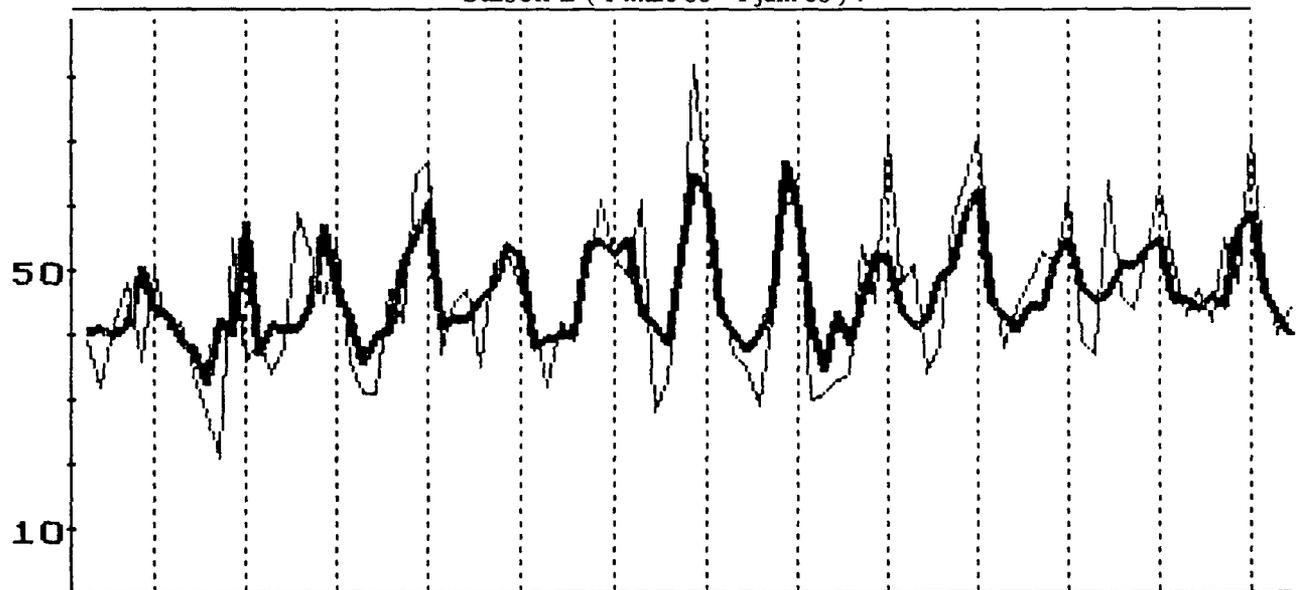
% de variance globale expliquée : **33,76**

DONNEES BRUTES  
Observation ( $A_t$ ) et ajustement poissonnien ( $\lambda_{A_t}$ )

Saison 1 ( 1 dec 87 - 1 mars 88 ) :

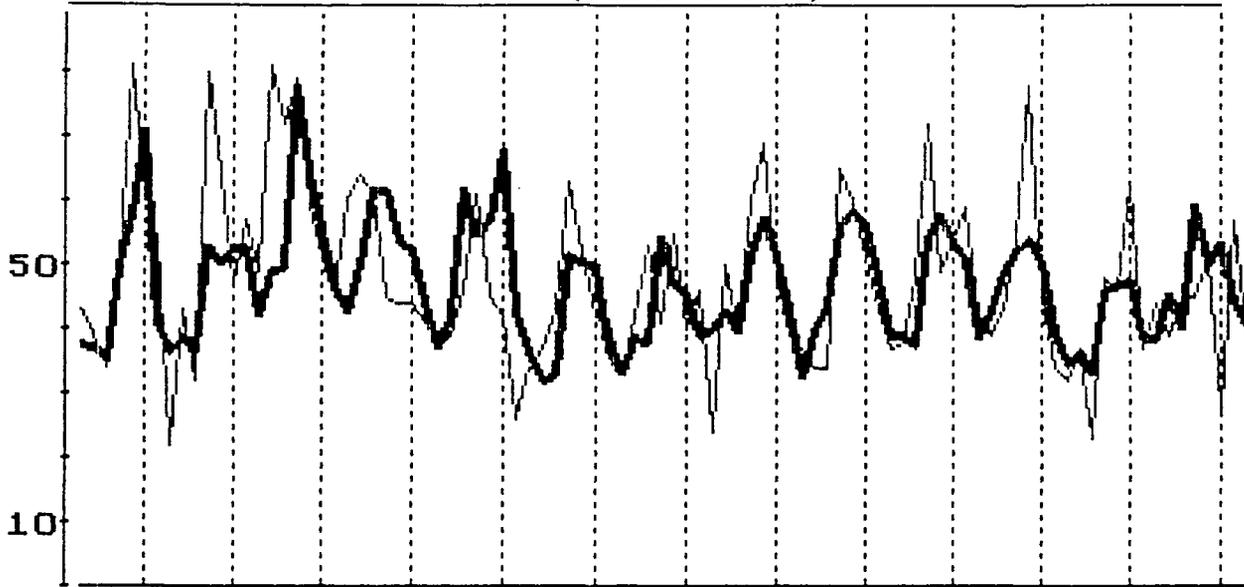


Saison 2 ( 1 mars 88 - 1 juin 88 ) :

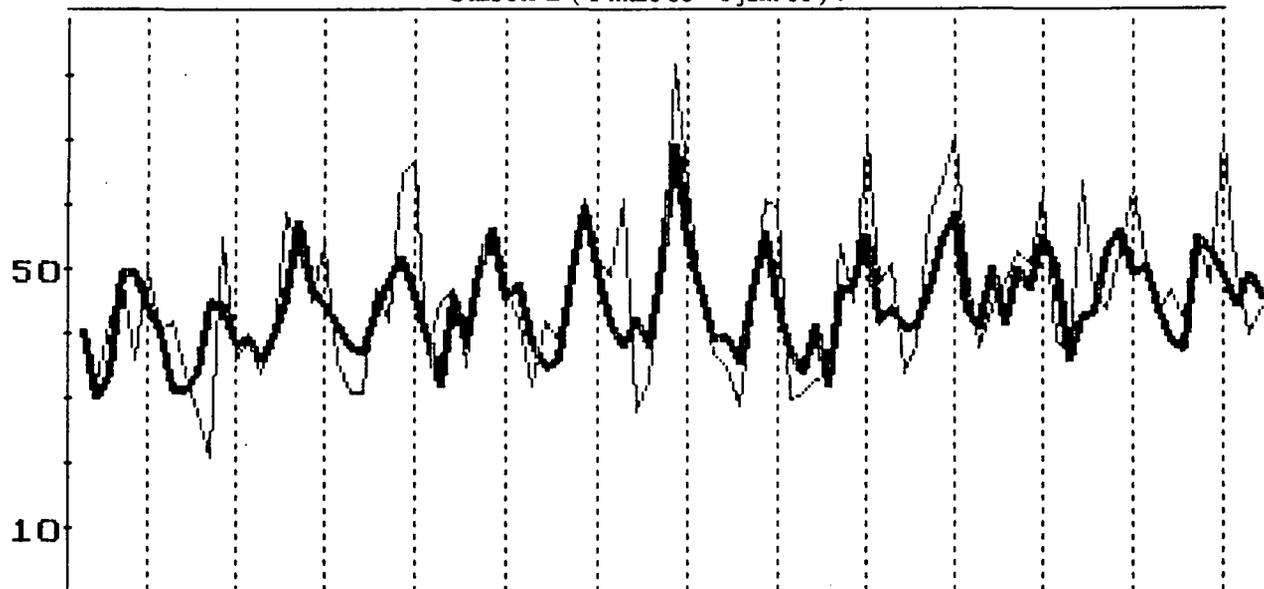


DONNEES BRUTES  
Observation ( $A_t$ ) et ajustement gaussien ( $A_t^G$ )

Saison 1 ( 1 dec 87 - 1 mars 88 ) :

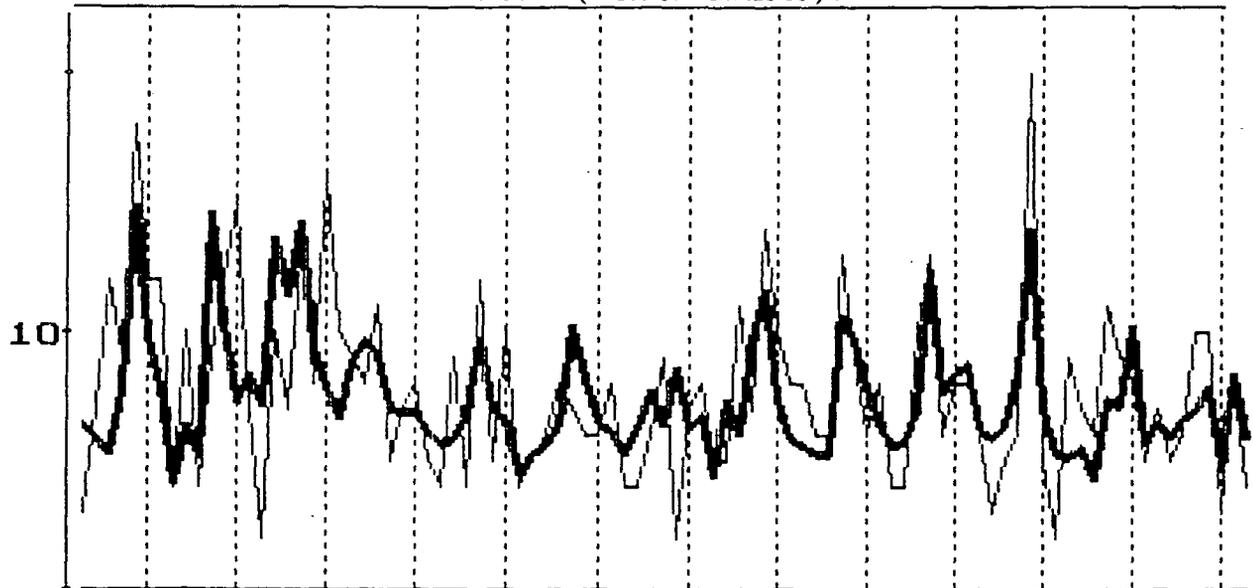


Saison 2 ( 1 mars 88 - 1 juin 88 ) :

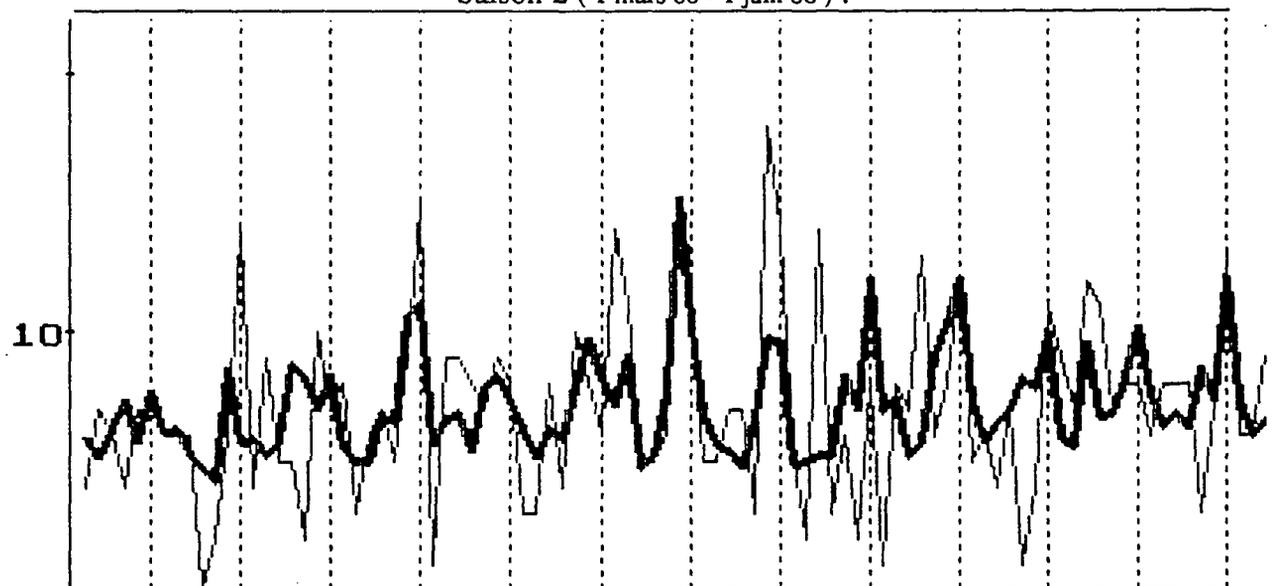


DONNEES BRUTES  
Observation ( $T_D$ ) et ajustement poissonnien ( $\lambda_{T_D}$ )

Saison 1 ( 1 dec 87 - 1 mars 88 ) :

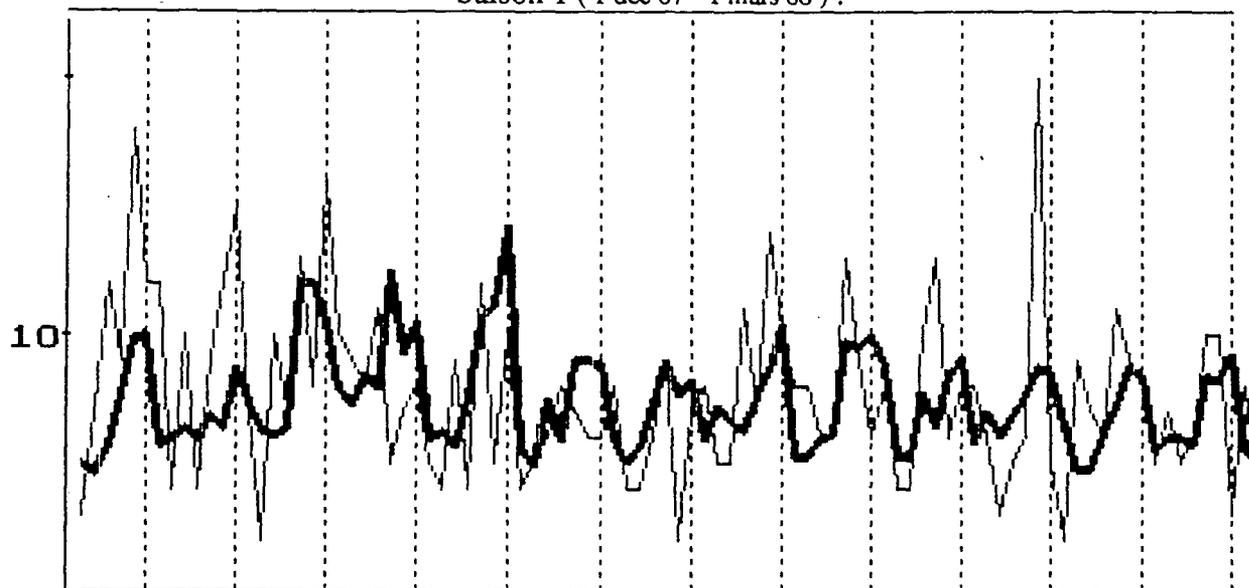


Saison 2 ( 1 mars 88 - 1 juin 88 ) :

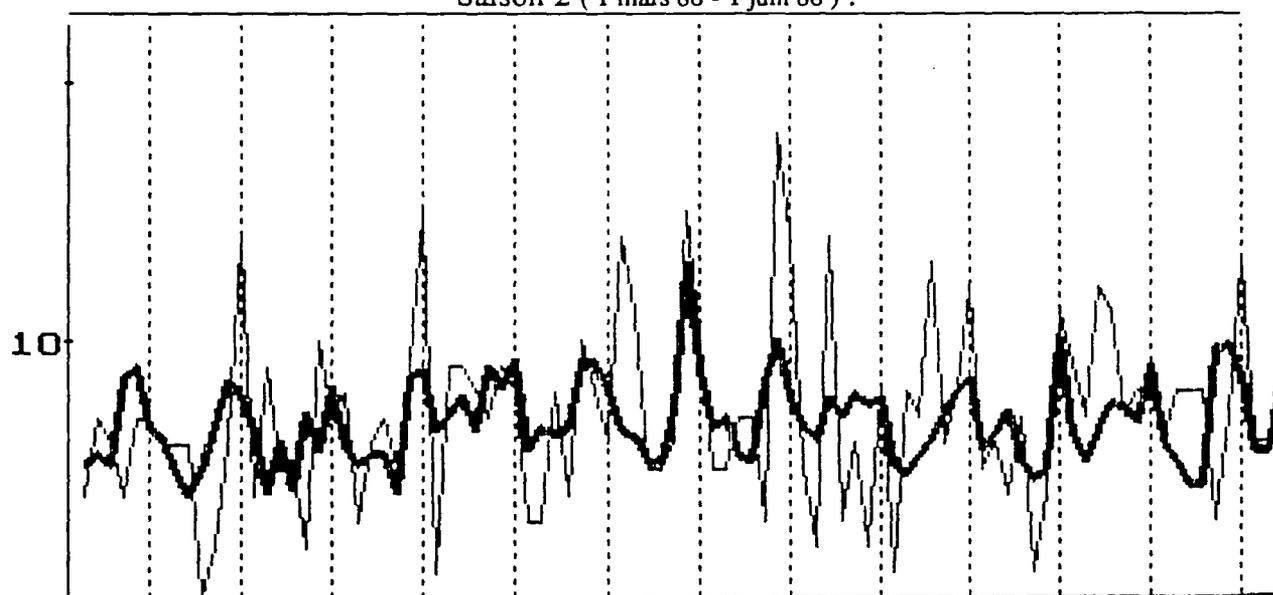


DONNEES BRUTES  
Observation ( $T_t$ ) et ajustement gaussien ( $T_t^G$ )

Saison 1 ( 1 dec 87 - 1 mars 88 ) :



Saison 2 ( 1 mars 88 - 1 juin 88 ) :



## ANNEXE 4

### Le prédicteur gaussien

Pour les données corrigées de la météo et du calendrier exceptionnel ou "spécial", on dispose pour chaque série d'un modèle Arma de la forme :

$$(1 - B^7)(1 - aB)(YC_t - \sum_i \alpha_i Z_{i,t}) = (1 - bB)(1 - cB^7)W_t$$

où  $W_t$  représente le bruit blanc gaussien d'innovation,  $B$  l'opérateur retard. On en déduit un **prédicteur gaussien** de la forme :

$$Y_t^G = Y_t - W_t$$

Pour les séries ( $A_t$ ) et ( $T_t$ ) on observe les écarts à la prédiction gaussienne suivants :

$$e = (A_t - A_t^G):$$

	global	L	M	Me	J	V	S	D
E(e)	0,52	0,74	1,00	0,34	0,10	0,51	0,75	0,21
E e	6,98	6,51	5,80	6,49	6,94	7,56	7,38	8,18
Var(e)	79,95	68,86	56,70	70,85	77,83	91,74	89,52	103,48

% de variance globale expliquée : 37.07

$$e = (T_t - T_t^G):$$

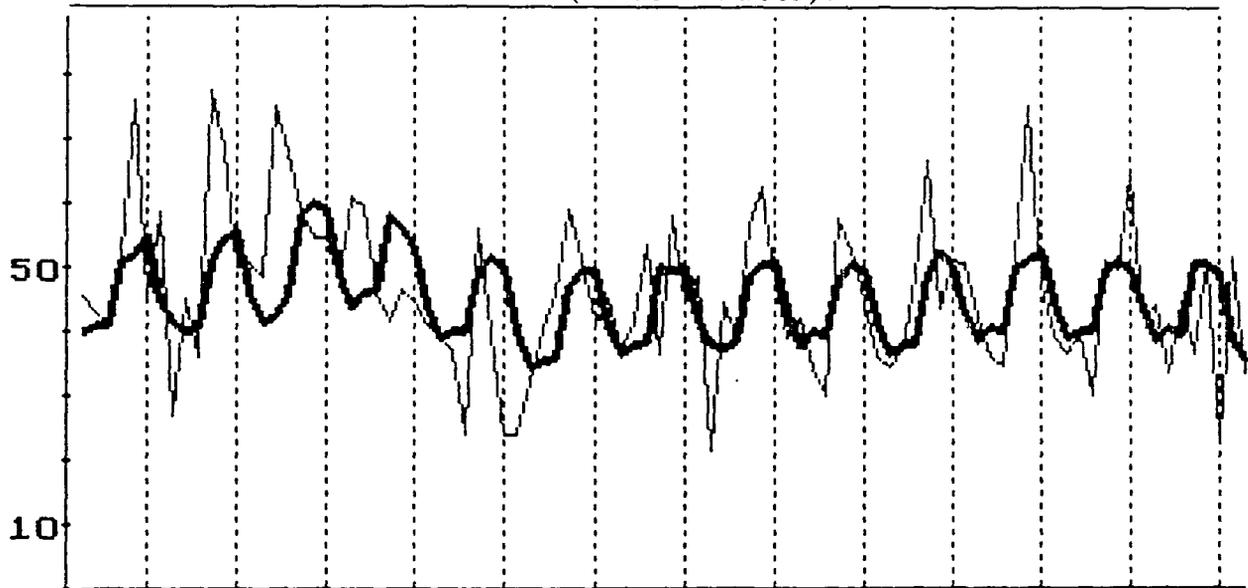
	global	L	M	Me	J	V	S	D
E(e)	0,00	0,06	0,10	-0,25	0,07	0,10	-0,03	-0,03
E e	2,63	2,39	2,22	2,43	2,44	2,77	2,94	3,19
Var(e)	11,30	9,31	8,14	9,06	9,56	13,57	13,49	15,92

% de variance globale expliquée : 14.26

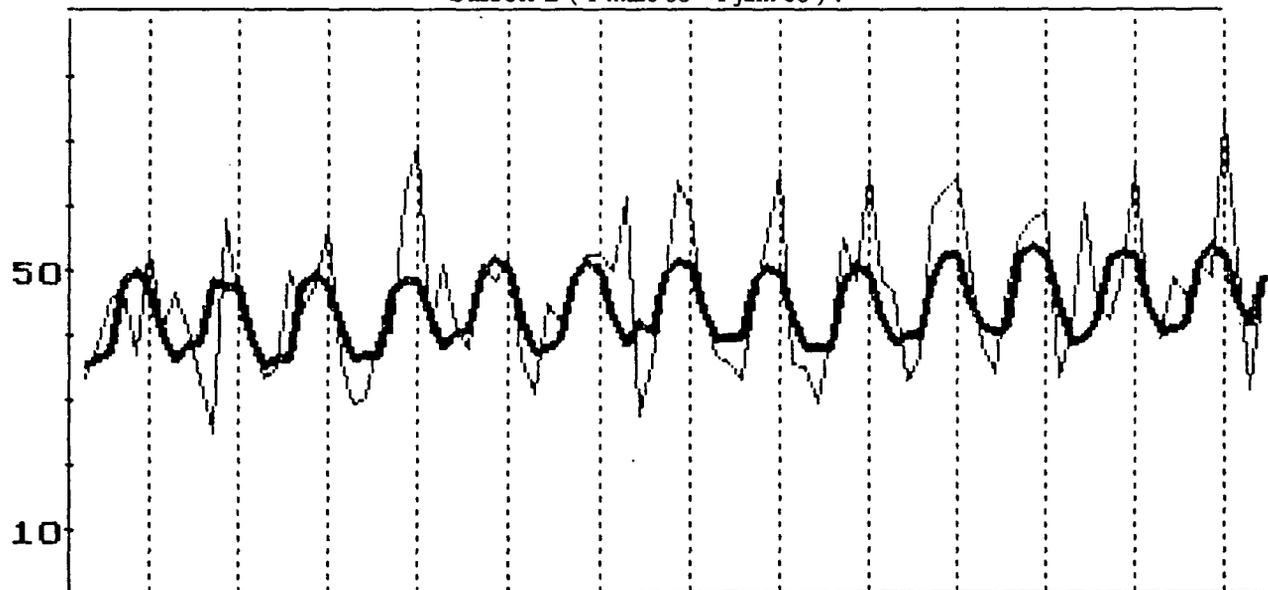
Les courbes qui suivent permettront de les apprécier plus finement .

Observation ( $A_t$ ) et prédiction gaussienne à un pas ( $A_t^G$ )

Saison 1 ( 1 dec 87 - 1 mars 88 ) :

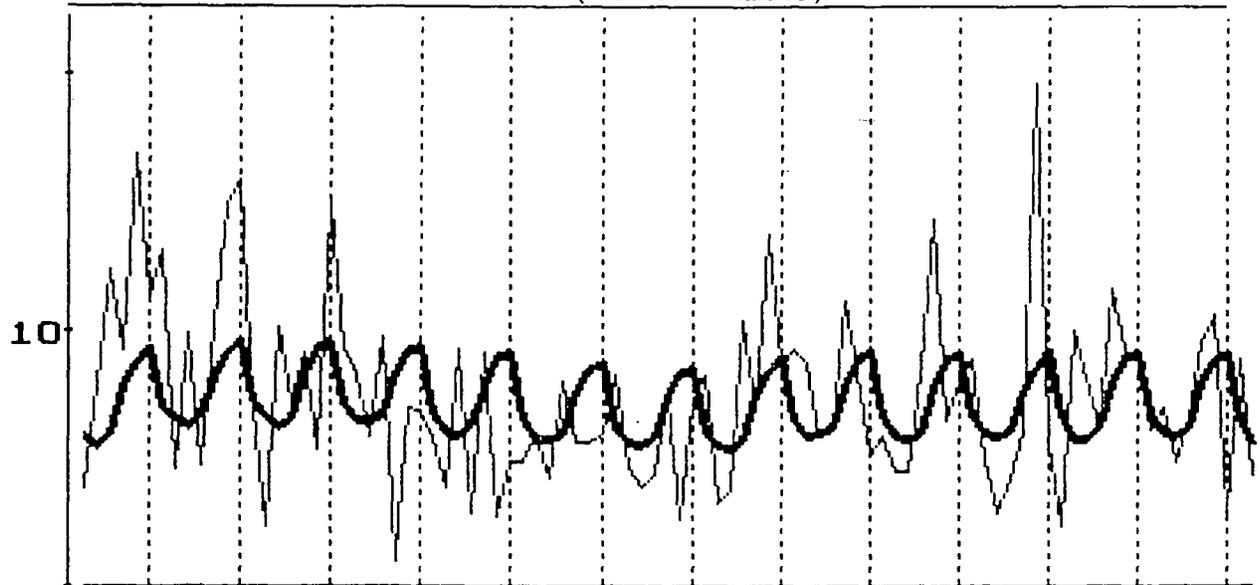


Saison 2 ( 1 mars 88 - 1 juin 88 ) :



Observation ( $T_t$ ) et prédiction gaussienne à un pas ( $T_t^G$ )

Saison 1 ( 1 dec 87 - 1 mars 88 ) :



Saison 2 ( 1 mars 88 - 1 juin 88 ) :

