



HAL
open science

Séries chronologiques vectorielles à composantes binaires : application en climatologie

Belkheir Essebbar

► **To cite this version:**

Belkheir Essebbar. Séries chronologiques vectorielles à composantes binaires : application en climatologie. Modélisation et simulation. Université Joseph-Fourier - Grenoble I, 1984. Français. NNT : . tel-00311718

HAL Id: tel-00311718

<https://theses.hal.science/tel-00311718>

Submitted on 20 Aug 2008

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 à

l' Université Scientifique et Médicale de Grenoble

pour obtenir le grade de
DOCTEUR DE 3ème CYCLE
«Mathématiques appliquées»

par

ESSEBBAR Belkheir



**SERIES CHRONOLOGIQUES VECTORIELLES A
COMPOSANTES BINAIRES.**

APPLICATION EN CLIMATOLOGIE.



Thèse soutenue le 17 mai 1984 devant la commission d'examen.

**A. LE BRETON
J.W. DELLEUR
S. MARTIN
G. GREGOIRE**

Président

Examineurs

UNIVERSITE SCIENTIFIQUE ET MEDICALE DE GRENOBLE

Année universitaire 1982-1983

Président de l'Université : M. TANCHE

MEMBRES DU CORPS ENSEIGNANT DE L'U.S.M.G.

(RANG A)

SAUF ENSEIGNANTS EN MEDECINE ET PHARMACIE

PROFESSEURS DE 1ère CLASSE

ARNAUD Paul	Chimie organique
ARVIEU Robert	Physique nucléaire I.S.N.
AUBERT Guy	Physique C.N.R.S.
AYANT Yves	Physique approfondie
BARBIER Marie-Jeanne	Electrochimie
BARBIER Jean-Claude	Physique expérimentale C.N.R.S. (labo de magnétisme)
BARJON Robert	Physique nucléaire I.S.N.
BARNOUD Fernand	Biosynthèse de la cellulose-Biologie
BARRA Jean-René	Statistiques - Mathématiques appliquées
BELORISKY Elie	Physique
BENZAKEN Claude (M.)	Mathématiques pures
BERNARD Alain	Mathématiques pures
BERTRANDIAS Françoise	Mathématiques pures
BERTRANDIAS Jean-Paul	Mathématiques pures
BILLET Jean	Géographie
BONNIER Jean-Marie	Chimie générale
BOUCHEZ Robert	Physique nucléaire I.S.N.
BRAVARD Yves	Géographie
CARLIER Georges	Biologie végétale
CAUQUIS Georges	Chimie organique
CHIBON Pierre	Biologie animale
COLIN DE VERDIERE Yves	Mathématiques pures
CRABBE Pierre (détaché)	C.E.R.M.O.
CYROT Michel	Physique du solide
DAUMAS Max	Géographie
DEBELMAS Jacques	Géologie générale
DEGRANGE Charles	Zoologie
DELOBEL Claude (M.)	M.I.A.G. Mathématiques appliquées
DEPORTES Charles	Chimie minérale
DESRE Pierre	Electrochimie
DOLIQUE Jean-Michel	Physique des plasmas
DUCROS Pierre	Cristallographie
FONTAINE Jean-Marc	Mathématiques pures
GAGNAIRE Didier	Chimie physique

.../...

GASTINEL Noël	Analyse numérique - Mathématiques appliquées
GERBER Robert	Mathématiques pures
GERMAIN Jean-Pierre	Mécanique
GIRAUD Pierre	Géologie
IDELMAN Simon	Physiologie animale
JANIN Bernard	Géographie
JOLY Jean-René	Mathématiques pures
JULLIEN Pierre	Mathématiques appliquées
KAHANE André (détaché DAFCO)	Physique
KAHANE Josette	Physique
KOSZUL Jean-Louis	Mathématiques pures
KRAKOWIAK Sacha	Mathématiques appliquées
KUPTA Yvon	Mathématiques pures
LACAZE Albert	Thermodynamique
LAJZEROWICZ Jeannine	Physique
LAJZEROWICZ Joseph	Physique
LAURENT Pierre	Mathématiques appliquées
DE LEIRIS Joël	Biologie
LLIBOUTRY Louis	Géophysique
LOISEAUX Jean-Marie	Sciences nucléaires I.S.N.
LOUP Jean	Géographie
MACHE Régis	Physiologie végétale
MAYNARD Roger	Physique du solide
MICHEL Robert	Minéralogie et pétrographie (géologie)
MOZIERES Philippe	Spectrométrie - Physique
OMONT Alain	Astrophysique
OZENDA Paul	Botanique (biologie végétale)
PAYAN Jean-Jacques (détaché)	Mathématiques pures
PEBAY PEYROULA Jean-Claude	Physique
PERRIAUX Jacques	Géologie
PERRIER Guy	Géophysique
PIERRARD Jean-Marie	Mécanique
RASSAT André	Chimie systématique
RENARD Michel	Thermodynamique
RICHARD Lucien	Biologie végétale
RINAUDO Marguerite	Chimie CERMAV
SENGEL Philippe	Biologie animale
SERGERAERT Francis	Mathématiques pures
SOUTIF Michel	Physique
VAILLANT François	Zoologie
VALENTIN Jacques	Physique nucléaire I.S.N.
VAN CUTSEN Bernard	Mathématiques appliquées
VAUQUOIS Bernard	Mathématiques appliquées
VIALON Pierre	Géologie
PROFESSEURS DE 2ème CLASSE	
ADIBA Michel	Mathématiques pures
ARMAND Gilbert	Géographie

AURIAULT Jean-Louis	Mécanique
BEGUIN Claude (M.)	Chimie organique
BOEHLER Jean-Paul	Mécanique
BOITET Christian	Mathématiques appliquées
BORNAREL Jean	Physique
BRUN Gilbert	Biologie
CASTAING Bernard	Physique
CHARDON Michel	Géographie
COHENADDAD Jean-Pierre	Physique
DENEUVILLE Alain	Physique
DEPASSEL Roger	Mécanique des fluides
DOUCE Roland	Physiologie végétale
DUFRESNOY Alain	Mathématiques pures
GASPARD François	Physique
GAUTRON René	Chimie
GIDON Maurice	Géologie
GIGNOUX Claude (M.)	Sciences nucléaires I.S.N.
GUITTON Jacques	Chimie
HACQUES Gérard	Mathématiques appliquées
HERBIN Jacky	Géographie
HICTER Pierre	Chimie
JOSELEAU Jean-Paul	Biochimie
KERCKOVE Claude (M.)	Géologie
LE BRETON Alain	Mathématiques appliquées
LONGEQUEUE Nicole	Sciences nucléaires I.S.N.
LUCAS Robert	Physiques
LUNA Domingo	Mathématiques pures
MASCLE Georges	Géologie
NEMOZ Alain	Thermodynamique (CNRS - CRTBT)
OUDET Bruno	Mathématiques appliquées
PELMONT Jean	Biochimie
PERRIN Claude (M.)	Sciences nucléaires I.S.N.
PFISTER Jean-Claude (détaché)	Physique du solide
PIBOULE Michel	Géologie
PIERRE Jean-Louis	Chimie organique
RAYNAUD Hervé	Mathématiques appliquées
ROBERT Gilles	Mathématiques pures
ROBERT Jean-Bernard	Chimie physique
ROSSI André	Physiologie végétale
SAKAROVITCH Michel	Mathématiques appliquées
SARROT REYNAUD Jean	Géologie
SAXOD Raymond	Biologie animale
SOUTIF Jeanne	Physique
SCHOOL Pierre-Claude	Mathématiques appliquées
STUTZ Pierre	Mécanique
SUBRA Robert	Chimie
VIDAL Michel	Chimie organique
VIVIAN Robert	Géographie

à Anissa ...

à mes frères et soeurs ...

à mes parents ...

Je remercie vivement Messieurs DELLEUR et GREGOIRE d'avoir accepté de participer au Jury et pour l'intérêt qu'ils ont porté à ce travail. Je remercie aussi vivement :

Monsieur MARTIN, membre du Jury, qui a suggéré le sujet de cette thèse et m'a apporté des conseils avisés, et par son intermédiaire l'équipe de CNRS ER30, pour les moyens informatiques offerts ;

Monsieur LE BRETON, président du Jury et directeur de cette thèse, de m'avoir accueilli au sein de l'équipe statistique du laboratoire Tim3. Je le remercie tout particulièrement pour toute l'aide qu'il m'a apportée tout au long de ce travail ;

Madame HOTTELLIER qui m'a aidé à dactylographier cette thèse ;

Les membres du service de reprographie qui ont effectué le tirage de cette thèse.

Enfin, un souvenir chaleureux à MARGUERIE DE ROTROU et à mes collègues de travail pour leur soutien moral.

" Que la loi invite à l'observation rationnelle des êtres existants et à la recherche d'une connaissance de ces êtres par la raison "

Ibn Rochd (dit Averroès).

TABLE DES MATIERES

INTRODUCTION	1
<u>Chapitre 1</u> : VECTEURS ALEATOIRES A COMPOSANTES BINAIRES	5
I. Différentes représentations de la loi de probabilité P_Y	8
I.1. Première représentation	8
I.2. Deuxième représentation	10
II. Etude de deux types particuliers de vecteurs aléatoires	12
II.1. Premier modèle	12
II.2. Deuxième modèle	18
<u>Chapitre 2</u> : SERIES CHRONOLOGIQUES UNIDIMENTIONNELLES BINAIRES	25
I. Caractéristiques d'une série chronologique unidimensionnelle binaire	25
I.1. Lois des temps de séjour et persis - tances des états	28
I.2. Lois des temps de séjour résiduel, moments et fonctions génératrices ..	33
I.3. Lois marginales de dimensions finies et statistiques de comptage	36
II. Modèles de séries chronologiques binaires	38
II.1. Processus de renouvellement alterné à temps discret	38
II.2. Chaines de Markov binaires d'ordre 1 et 2	40
II.3. Processus de renouvellement alterné basés sur des lois binômiales négatives translatées	43
II.4. Autres modèles de séries chronologiques binaires	44

<u>Chapitre 3</u> :	CALCUL DE CERTAINES FONCTIONS GENERATRICES ASSOCIEES A UN PROCESSUS DE RENOUVELLEMENT ALTERNE	45
I.	Algorithme de calcul de la loi de la variable aleatoire S_n	47
II.	Fonctions génératrices des suites $\{R_n^i(k); n \geq 1\}; i \in \{0,1\}, k \in \mathbb{N}$	52
III.	Fonctions génératrices des suites $\{\tilde{R}_n^i(k); n \geq 1\}; i \in \{0,1\}, k \in \mathbb{N}$	55
IV.	Fonctions génératrices des suites $\{R_t^i(u); u \in \mathbb{N}\}$ et $\{\tilde{R}_t^i(u); u \in \mathbb{N}\}, i \in \{0,1\}, t \leq 1$	59
<u>Chapitre 4</u> :	ETUDE D'UN MODELE DE SERIES CHRONOLOGIQUES VECTORIELLES A MARGINALES CONSTRUITES PAR AMINCISSEMENTS D'UN PROCESSUS DE RENOUVELLEMENT ALTERNE	63
I.	Etude des caractéristiques du modèle .	67
I.1.	Propriétés du processus vectoriel $(\vec{X}_t; t \geq 1)$	68
I.2.	Propriétés des séries chronologiques marginales $(X_t^j; t \geq 1), j=1, \dots, q$..	73
II.	Analyse statistique et étude expérimentale	87
II.1.	Cas d'une station test observable	87
II.2.	Cas d'une station test non observable	100
<u>Chpitre 5</u> :	ETUDE D'UN MODELE DE SERIES CHRONOLOGIQUES VECTORIELLES A MARGINALES CONSTRUITES PAR SUPERPOSITION D'AMINCISSEMENTS D'UN PROCESSUS DE RENOUVELLEMENT ALTERNE ET DE SON "COMPLEMENTAIRE"	113

I. Etudes des caracteristiques du modèle	116
I.1. Propriétés du processus vectoriel $(\vec{X}_t; t \geq 1)$	116
I.2. Propriétés des séries chronologiques marginales $(X_t^j; t \geq 1), j=1, \dots, q$	119
II. Analyse statistique et étude expérimentale	129
II.1. Cas d'une station test observable	129
II.2. Cas d'une station test non observable	140
 <u>Chapitre 6</u> : APPLICATIONS A LA MODELISATION DU PHENOMENE CLIMATOLOGIQUE.	 155
I. Essai de modélisation 1	158
II. Essai de modélisation 2	165
III. Essai de modélisation 3	171
IV. Essai de modélisation 4	177
 Conclusion	 183
 Bibliographie	 185

INTRODUCTION

La modélisation du phénomène climatologique de la succession des jours selon leur caractère sec ou humide dans les différentes stations de mesure d'un réseau au cours d'une certaine période de l'année a été l'objet de diverses études dans lesquelles chaque station est étudiée séparément. Les classes de modèles de séries chronologiques binaires qui ont été le plus utilisées à cette fin et dont on trouve une étude comparée dans Bousseboua (1983) sont :

- d'une part la classe des modèles de processus de renouvellement alternés qui contient les chaînes de Markov d'ordre 1 et 2 (cf. Gabriel-Neuman (1962), Green (1964), Katz (1974) et Galloy-Le Breton-Martin (1984))
- d'autre part la classe des modèles de processus binaires autorégressifs, moyennes mobiles et mixtes qui sont des cas particuliers des modèles DARMA introduits par Jacob-Lewis (1978a, 1978b) (cf. Buishand (1978), Chang-Delleur-Kavvas (1982, 1983a, 1983b)).

Nous envisageons ici une modélisation plus globale du phénomène, pour un réseau supposé assez homogène, dans laquelle l'interdépendance entre les évolutions spécifiques du phénomène dans les différentes stations est prise en compte. Un modèle adapté est alors un modèle de série chronologique vectorielle à composantes binaires dont les séries chronologiques marginales représentent bien le phénomène dans les différentes stations et tel que la structure de liaison entre ces séries rende bien compte des concomitances dans le réseau. Pour être utilisable un modèle doit se prêter à une analyse assez complète en particulier

de caractéristiques liées aux longueurs des séquences sèches ou humides, de probabilités d'apparition de certains événements...

En vue de répondre à ces objectifs nous avons proposé et étudié deux modèles dont les possibilités d'utilisation sont illustrées par des simulations et des essais de modélisation de certains sous-réseaux du réseau français de stations météorologiques.

Dans le premier chapitre, après avoir décrit les caractéristiques essentielles des vecteurs aléatoires à composantes binaires, nous examinons deux modèles particuliers en vue de la représentation du phénomène climatologique.

Dans le deuxième chapitre nous précisons les propriétés importantes des modèles de séries chronologiques unidimensionnelles binaires qui permettent d'apprécier leur aptitude à représenter les données climatologiques concernant une station et nous présentons quelques modèles classiques.

Dans le troisième chapitre, en préparation à l'analyse des modèles vectoriels proposés dans les chapitres suivants, nous calculons certaines fonctions génératrices associées à des processus de renouvellement alternés à temps discret.

Dans les quatrième et cinquième chapitres, d'une part nous faisons une étude complète des caractéristiques de deux modèles différents de séries chronologiques multidimensionnelles à composantes binaire dont les séries chronologiques marginales sont définies par certaines transformations d'un même processus de renouvellement alterné et d'autre part nous menons une étude expérimentale par simulation de ces modèles.

Le sixième chapitre contient l'application des modèles proposés dans des essais de modélisation concernant des réseaux de stations extraits du réseau météorologique français pour certaines périodes de l'année.

CHAPITRE 1 :

VECTEURS ALEATOIRES A COMPOSANTES BINAIRES



Lorsqu'on s'intéresse à décrire le phénomène climatologique de la succession des jours selon leur caractère sec ou humide on est amené naturellement à considérer des observations binaires i.e. à valeurs dans l'ensemble $\{0,1\}$. Un seuil α de quantité de précipitations étant choisi (souvent $\alpha = 1\text{mm}$ dans les applications), un jour étant déclaré sec (resp. humide) s'il a été recueilli ce jour moins (resp. plus) de α mm de pluie, on définit pour une station donnée une variable prenant la valeur 1 (resp. 0) si le jour est sec (resp. humide) dans la station. Dans le cadre d'une modélisation probabiliste du phénomène on est alors conduit en particulier à définir des vecteurs aléatoires à valeurs dans $E_q = \{0,1\}^q$ pour représenter les observations correspondant à un jour donné dans les q stations d'un réseau météorologique ou correspondant à q jours consécutifs dans une station donnée.

Dans ce chapitre nous décrivons des caractéristiques essentielles de tels vecteurs aléatoires et nous présentons deux modèles particuliers qui interviennent dans l'étude des modèles de séries chronologiques vectorielles des chapitres IV et V.

Etant donné un vecteur aléatoire $Y = (Y_1, \dots, Y_q)'$, défini sur un certain espace probabilisé $(\Omega; \mathcal{A}, P)$, à valeurs dans $\{0,1\}^q$ nous noterons P_Y sa loi de probabilité, $\text{cov}(Y_1, Y_j)$ la covariance entre les composantes Y_1 et Y_j ($\text{var}(Y_1) = \text{cov}(Y_1, Y_1)$) et S la variable aléatoire somme de ses composantes $S = \sum_{i=1}^q Y_i$. Il sera parfois commode d'identifier un élément $Y = (Y_1, Y_2, \dots, Y_q)'$ de E_q à la partie $\{j \in \{1, \dots, q\} : Y_j = 1\}$ de $\{1, \dots, q\}$; nous noterons encore y cette partie et nous utiliserons les notations usuelles de la théorie des ensembles pour décrire des propriétés d'éléments de E_q .

I. DIFFERENTES REPRESENTATIONS DE LA LOI DE PROBABILITE P_Y

Suivant Cox (1972) et Kedem (1980), plusieurs représentations de la loi P_Y , commodes pour certains calculs, sont possibles.

Nous notons V_q l'espace vectoriel de dimension $\text{card}(E_q) = 2^q$ des applications définies sur E_q à valeurs dans \mathbb{R} .

I.1. Première représentation

Munissons V_q du produit scalaire $\langle \cdot, \cdot \rangle_1$, défini par :

$$\langle \mu_1, \mu_2 \rangle_1 = \sum_{y \in E_q} \mu_1(y) \mu_2(y) \quad ; \quad \mu_1, \mu_2 \in V_q$$

Alors la famille $\{\xi_r\}_{r \in E_q}$ des éléments de V_q définis par :

$$\begin{aligned} \xi_r(y) &= \frac{1}{2^{q/2}} \cdot (-1)^{r'y} \quad ; \quad r=(r_1, \dots, r_q) \in E_q \\ &= \frac{1}{2^{q/2}} \cdot (-1)^{r_1 y_1 + \dots + r_q y_q} \quad ; \quad y=(y_1, \dots, y_q) \in E_q \end{aligned}$$

constitue une base orthonormale de V_q .

En effet on a, pour $r \in E_q$,

$$\langle \xi_r, \xi_r \rangle_1 = \frac{1}{2^q} \sum_{y \in E_q} (-1)^{2r'y} = \frac{\text{card } E_q}{2^q} = 1$$

et, pour $r \neq s$, $r, s \in E_q$

$$\langle \xi_r, \xi_s \rangle_1 = \frac{1}{2^q} \sum_{y \in E_q} (-1)^{y'(r+s)} = \frac{1}{2^q} \sum_{y \in E_q} (-1)^{y'k} = 0$$

car la somme $r_i + s_i, i=1, \dots, q$ est bien sûr entendue modulo 2 de sorte que $k \in E_q$.

On peut alors déduire :

Lemme 1.

La loi de probabilité P_Y de $Y = (Y_1, \dots, Y_q)$ peut s'écrire sous la forme :

$$(1) \quad P_Y(\{y\}) = (1-2y_1)(1-2y_2)\dots(1-2y_q) \times \\ \times E\{(1-y_1-Y_1)(1-y_2-Y_2)\dots(1-y_q-Y_q)\} \\ Y = (Y_1, \dots, Y_q) \in E_q.$$

Preuve - Comme $\{\xi_r\}_{r \in E_q}$ est une base orthonormale de V_q il suffit de vérifier que si h est l'application de E_q dans \mathbb{R} définie par :

$$h(y) = (1-2y_1)\dots(1-2y_q) E\{(1-y_1-Y_1)\dots(1-y_q-Y_q)\}$$

alors, pour tout $r \in E_q$, on a :

$$\langle h, \xi_r \rangle_1 = \langle P_Y, \xi_r \rangle_1 = E(\xi_r(Y)).$$

Or, comme pour $x \in \{0, 1\}$ on a $(-1)^x = 1-2x$, on obtient

$$\langle h, \xi_r \rangle_1 = E\left\{ \sum_{y \in E_q} \frac{1}{2^{q/2}} (1-2y_1)(1-2y_2)\dots(1-2y_q)^{r_1} \dots \right. \\ \left. \dots (1-2y_q)^{r_q} (1-y_1-Y_1)\dots(1-y_q-Y_q) \right\}$$

Alors, comme $\prod_{i=1}^q (1-2y_i)(1-2y_i)^{r_i} (1-y_i-Y_i)$ vaut 0 ou $\prod_{i=1}^q (1-2y_i)^{r_i}$

selon que $Y \neq y$ ou $Y=y$, on a

$$\langle h, \xi_r \rangle_1 = E \left\{ \frac{1}{2^{q/2}} \prod_{i=1}^q (1-2Y_i)^{r_i} \right\} = E(\xi_r(Y)) .$$

En fait, plus directement, constatant que l'évènement $[Y=y]$ est réalisé si et seulement si la variable aléatoire binaire

$$\prod_{i=1}^q (1-2Y_i)^{(1-Y_i-Y_i)} \text{ vaut } 1, \text{ on obtient le résultat. } \square$$

Remarque 1.

On a en particulier : $P[Y_i=1] = E(Y_i)$, $P[Y_i=0] = 1-E(Y_i)$
 et $P[Y_i=1, Y_j=1] = E(Y_i Y_j)$; $P[Y_i=1, Y_j=0] = E(Y_i) - E(Y_i Y_j)$;
 $P[Y_i=0, Y_j=0] = 1 - E(Y_i) - E(Y_j) + E(Y_i Y_j)$.

On peut envisager une autre représentation de la loi P_Y .

I.2. Deuxième représentation

Posons $p_i = E(Y_i) = P[Y_i=1]$, munissons V_q du produit scalaire \langle , \rangle_2 défini par :

$$\langle \mu_1, \mu_2 \rangle_2 = \sum_{y \in E_q} \mu_1(y) \mu_2(y) \left\{ \prod_{i=1}^q p_i^{y_i} (1-p_i)^{1-y_i} \right\} ; \mu_1, \mu_2 \in V_q .$$

et introduisons les versions normalisées de la variable Y_i :

$$x_i = \frac{Y_i - p_i}{\sqrt{p_i(1-p_i)}}$$

et de la composante y_i de $y = (y_1, \dots, y_q)'$:

$$x_i = \frac{y_i - p_i}{\sqrt{p_i(1-p_i)}}$$

Il est clair que les 2^q fonctions qui à (y_1, \dots, y_q) associent respectivement 1 ; x_1, \dots, x_q ; $x_1 x_2, \dots, x_{q-1} x_q$;

.....; $x_1 x_2 x_3 \dots x_q$ forment une base orthonormale de V_q . On a alors

Lemme 2.

La loi de probabilité P_Y de $Y = (Y_1, \dots, Y_q)$ peut s'écrire sous la forme :

$$(2) \quad P_Y(\{y\}) = \prod_{i=1}^q p_i^{y_i} (1-p_i)^{1-y_i} \left\{ 1 + \sum_{i < j} \rho_{ij} x_i x_j + \sum_{i < j < k} \rho_{ijk} x_i x_j x_k + \dots + \rho_{12\dots q} x_1 x_2 \dots x_q \right\}$$

où $\rho_{ij\dots k} = E(X_i X_j \dots X_k)$

Preuve - l'application $y \rightarrow P_Y(y) / \prod_{i=1}^q p_i^{y_i} (1-p_i)^{1-y_i}$ est un élément de V_q qui admet la représentation :

$$1 + E(X_1)x_1 + \dots + E(X_q)x_q + E(X_1 X_2)x_1 x_2 + \dots + E(X_{q-1} X_q)x_{q-1} x_q + \dots + E(X_1 X_2 \dots X_q)x_1 x_2 \dots x_q$$

selon la base orthonormale définie précédemment. Cela fournit l'assertion du lemme. \square

Remarque 2.

la représentation (2) a l'avantage de mettre en évidence le facteur entre accolades qui est un indice "de l'écart de la loi de P_Y par rapport à celle correspondant à l'indépendance des composantes Y_1, \dots, Y_q ".

On a en particulier

$$P[Y_i=y_i, Y_j=y_j] = P[Y_i=y_i] P[Y_j=y_j] \{ 1 + \rho_{ij} x_i x_j \}$$

ce qui montre que Y_i et Y_j sont indépendantes si et seulement si elles sont non corrélées.

Notons enfin que $\text{Cov}(Y_i, Y_j) = P[Y_i=1, Y_j=1] - P[Y_i=1] P[Y_j=1]$.

Pour d'autres représentations de la loi P_Y nous renvoyons à Cox(1972), Kedem(1980).

II. ETUDE DE DEUX TYPES PARTICULIERS DE VECTEURS ALEATOIRES.

Nous étudions maintenant deux classes de vecteurs aléatoires que nous rencontrerons dans les modèles que nous proposons aux chapitres IV et V pour le phénomène climatologique. Chacun des deux types de vecteurs aléatoires que nous considérons est construit à partir de la donnée d'un vecteur aléatoire $Y^* = (Y_1^*, \dots, Y_q^*)'$ de la loi P_{Y^*} et de q variables aléatoires $\epsilon_1, \epsilon_2, \dots, \epsilon_q$ de Bernoulli indépendantes de paramètres respectifs $\alpha_j = E(\epsilon_j) = P[\epsilon_j=1]$, $j=1, \dots, q$ tels que les vecteurs aléatoires Y^* et $(\epsilon_1, \dots, \epsilon_q)$ sont indépendants.

Nous noterons $p_j^* = P[Y_j^*=1]$. De même S^* (resp. Z , resp. Z_i) désignera la variable aléatoire somme des composantes de $Y^* = (Y_1^*, \dots, Y_q^*)'$ (resp. $\epsilon = (\epsilon_1, \dots, \epsilon_q)'$, resp. $(\epsilon_j, j \in \{1, \dots, q\} \setminus \{i\})$).

II.1. Premier modèle.

On s'intéresse ici au vecteur aléatoire Y dont les composantes binaires Y_j , $j=1, \dots, q$ sont définies par :

$$Y_j = \epsilon_j Y_j^* \quad ; \quad j=1, \dots, q.$$

On a évidemment

$$\begin{aligned} P[Y_j=1] &= E(\epsilon_j Y_j^*) = E(\epsilon_j) E(Y_j^*) \\ &= \alpha_j p_j^* \quad ; \quad j=1, \dots, q. \end{aligned}$$

En ce qui concerne la loi conjointe de $(Y_1, \dots, Y_q)'$ on a :

Lemme 3.

La loi P_Y du vecteur $Y = (Y_1, \dots, Y_q)'$ est donnée par :

$$\begin{aligned} (3) \quad P_Y(\{y\}) &= (-1)^{|y|} \sum_{z \in E_q} \prod_{i=1}^q (1 - y_i - \alpha_i z_i) P_{Y^*}(\{z\}) \\ &= \left(\prod_{i \in y} \alpha_i \right) \sum_{z \supset y} \prod_{j \in z \setminus y} (1 - \alpha_j) P_{Y^*}(\{z\}) \quad ; \quad y \in E_q \end{aligned}$$

La matrice de covariance de Y a pour terme général

$$(4) \quad \text{Cov}(Y_i, Y_j) = \begin{cases} \alpha_i p_i^* (1 - \alpha_i p_i^*) & \text{si } i=j \\ \alpha_i \alpha_j \text{Cov}(Y_i^*, Y_j^*) & \text{si } i \neq j . \end{cases}$$

Preuve - D'après le lemme 1, on a :

$$P_Y(\{y\}) = \left[\prod_{i=1}^q (1 - 2y_i) \right] E \left\{ \prod_{i=1}^q (1 - y_i - \epsilon_i Y_i^*) \right\}$$

d'où, compte tenu des propriétés d'indépendance :

$$P_Y(\{y\}) = (-1)^{|y|} E \left\{ \prod_{i=1}^q (1 - y_i - \alpha_i Y_i^*) \right\}$$

ce qui nous donne la première égalité dans (3).

Remarquant que

$$\begin{aligned} P[Y=y] &= P[Y=y, Y^* \supseteq y] = \sum_{z \supseteq y} P[Y=y, Y^*=z] \\ &= \sum_{z \supseteq y} P[y^*=z; \epsilon_j=1, j \in y; \epsilon_j=0, j \in z \setminus y] \end{aligned}$$

et utilisant les hypothèses d'indépendance on obtient immédiatement la deuxième égalité dans (3).

La dernière assertion du lemme est une conséquence facile de la définition de Y. \square

Les propriétés suivantes de la variable S sont immédiates :

- $P[S=q] = \left[\prod_{j=1}^q \alpha_j \right] P[S^*=q]$
- $E(S) = \sum_{j=1}^q \alpha_j p_j^*$
- $\text{Var}(S) = \sum_{j=1}^q \alpha_j p_j^* (1 - \alpha_j p_j^*) + 2 \sum_{i < j} \alpha_i \alpha_j \text{Cov}(Y_i^*, Y_j^*)$

Nous considérons maintenant deux cas particuliers de ce modèle.

Cas particulier 1.

On suppose que $Y_1^* = Y_2^* = \dots = Y_q^*$ avec $P[Y_j^*=1] = p_j^* = p$

Dans le cadre du modèle envisagé au chapitre IV le vecteur aléatoire Y représentera l'observation, pour un jour donné, des états des q différentes stations d'un réseau et S représentera le nombre de stations "sèches" dans le réseau ce jour-là.

Comme Y^* ne prend alors que les deux valeurs $(0, \dots, 0)$ et $(1, \dots, 1)$ avec les probabilités respectives $1-p^*$ et p^* , la formule (3) s'écrit sous la forme :

$$P_Y(\{y\}) = (-1)^{|y|} \left\{ p^* \left[\prod_{i=1}^q (1-\alpha_i - y_i) \right] + (1-p^*) \left[\prod_{i=1}^q (1-y_i) \right] \right\}; y \in E_C$$

$$= \begin{cases} (1-p^*) + p^* \prod_{i=1}^q (1-\alpha_i) & \text{si } y = \phi \\ p^* \prod_{i=1}^q \alpha_i^{y_i} (1-\alpha_i)^{1-y_i} & \text{si } y \neq \phi \end{cases}$$

De même l'égalité (4) s'écrit :

$$(4') \quad \text{Cov}(Y_i, Y_j) = \begin{cases} \alpha_i p^* (1-\alpha_i p^*) & \text{si } i=j \\ \alpha_i \alpha_j p^* (1-p^*) & \text{si } i \neq j \end{cases}$$

Quant à la variable aléatoire S on déduit aisément des résultats précédents :

Lemme 4.

On a :

$$P[S = k] = \begin{cases} (1-p^*) + p^* \left[\prod_{j=1}^q (1-\alpha_j) \right] & \text{si } k=0 \\ p^* P[Z = k] & \text{si } 1 \leq k \leq q \end{cases}$$

$$\text{où } P[Z = k] = \sum_{y: |y|=k} \prod_{j=1}^q \alpha_j^{y_j} (1-\alpha_j)^{1-y_j}$$

$|y|$ désignant le cardinal de y .

On a aussi

$$E(S) = p^* \sum_{j=1}^q \alpha_j \quad ; \quad \text{Var}(S) = p^* \sum_{j=1}^q \alpha_j (1-\alpha_j p^*) + 2p^* (1-p^*) \sum_{i < j} \alpha_i \alpha_j$$

On s'intéresse maintenant aux lois conditionnelles des variables aléatoires Y_i sachant la variable aléatoire S qui fournissent les probabilités que les différents sous-réseaux soient "secs" sachant le nombre de stations "sèches" dans le réseau un jour donné

Lemme 5.

On a :

$$P[Y_1 = 1/S = k] = \begin{cases} \frac{\alpha_1 P[Z_1 = k-1]}{P[Z = k]} & \text{si } 1 \leq k \leq q \\ 0 & \text{si } k=0 \end{cases}$$

et

$$P[Y = y/S = k] = \begin{cases} \frac{\prod_{i=1}^q \alpha_i^{y_i} (1-\alpha_i)^{1-y_i}}{P[Z = k]} & \text{si } 1 \leq k \leq q \text{ et } |y|=k \\ 1 & \text{si } k=0 \text{ et } |y|=0 \end{cases}$$

Preuve - On a bien sûr d'une part

$$P[Y_1 = 1, S = k] = \begin{cases} P[Y_1 = 1] P[Z_1 = k-1] & \text{si } 1 \leq k \leq q \\ 0 & \text{si } k=0 \end{cases}$$

et d'autre part

$$P[Y = y, S = k] = \begin{cases} P[Y = y] & \text{si } |y| = k \\ 0 & \text{sinon} \end{cases}$$

En vertu du lemme 4 on a alors les résultats annoncés. \square

Remarquons que $P[Z_i = \ell] = \sum_{\substack{r \in \{0,1\}^{q-1} \\ |r| = \ell}} \left[\prod_{j \neq i} \alpha_j^{r_j} (1-\alpha_j)^{1-r_j} \right];$
 $\ell \in \{0,1,\dots,q-1\}.$

Cas particulier 2.

On suppose ici que les variables aléatoires ϵ_j $j=1,\dots,q$ sont de même loi avec $\alpha_j = \alpha$, $j=1,\dots,q$. Dans le cadre du modèle envisagé au chapitre IV le vecteur aléatoire Y représentera alors l'observation, pour une station donnée, des états successifs au cours de q jours consécutifs et S représentera le nombre de jours secs parmi ces q jours.

On traduit immédiatement les formules (3) et (4) par :

$$(3'') \quad P_Y(\{y\}) = (-1)^{|y|} \sum_{z \in E_q} \left[\prod_{i=1}^q (1-y_i - \alpha z_i) \right] P_{Y^*}(\{z\})$$

$$= \alpha^{|y|} \sum_{z > y} (1-\alpha)^{|z-y|} P_{Y^*}(\{z\}) \quad ; \quad y \in E_q,$$

$$(4'') \quad \text{Cov}(Y_i, Y_j) = \begin{cases} \alpha p_i^* (1-\alpha p_i^*) & \text{si } i=j \\ \alpha^2 \text{Cov}(Y_i^*, Y_j^*) & \text{si } i \neq j \end{cases}$$

En ce qui concerne la variable S on peut énoncer :

Lemme 6.

On a :

$$P[S = k] = \left(\frac{\alpha}{1-\alpha} \right)^k \sum_{s=k}^q C_s^k (1-\alpha)^s P[S^* = s] \quad ; \quad 0 \leq k \leq q$$

et

$$E(S) = \alpha E(S^*) \quad ; \quad m_{[2]}(S) = \alpha^2 m_{[2]}(S^*)$$

où $m_{[2]}(S) = E(S(S-1))$ est le moment factoriel d'ordre 2 de S .

Preuve - On peut écrire

$$P[S = k] = \sum_{s=k}^q P[S = k, S^* = s]$$

où pour $s \geq k$ on a :

$$\begin{aligned} P[S = k, S^* = s] &= \sum_{y: |y|=s} P[S = k, Y^* = y] \\ &= \sum_{y: |y|=s} \sum_{z: y: |z|=k} P[Y^* = y; \epsilon_j = 1, j \in z; \epsilon_j = 0, j \in y \setminus z] \\ &= \sum_{y: |y|=s} \sum_{z: y: |z|=k} \alpha^{|z|} (1-\alpha)^{|y-z|} P[Y^* = y] \\ &= C_s^k \alpha^k (1-\alpha)^{s-k} \sum_{y: |y|=s} P[Y^* = y] \\ &= C_s^k \alpha^k (1-\alpha)^{s-k} P[S^* = s] \end{aligned}$$

La première assertion du lemme en découle.

Comme $S = \sum_{j=1}^q \epsilon_j Y_j^*$ on a aussi

$$E(S) = \alpha \sum_{j=1}^q E(Y_j^*) = \alpha E(S^*)$$

et

$$\begin{aligned} E(S^2) &= E\left(\sum_j \epsilon_j Y_j^*\right) + E\left(\sum_{i \neq j} \epsilon_i \epsilon_j Y_i^* Y_j^*\right) \\ &= \alpha E(S^*) + \alpha^2 \sum_{i \neq j} E(Y_i^* Y_j^*) \\ &= \alpha E(S^*) + \alpha^2 \{ E((S^*)^2) - E(S^*) \} \end{aligned}$$

$$\text{d'où } m_{[2]}(S) = \alpha^2 m_{[2]}(S^*) \quad \square$$

Remarquons que $\text{Var}(S) = \alpha^2 \text{Var}(S^*) + \alpha(1-\alpha)E(S^*)$.

II.2. Deuxième modèle.

On considère maintenant un vecteur aléatoire Y dont les composantes binaires Y_j , $j=1, \dots, q$ sont définies par :

$$Y_j = \varepsilon_j Y_j^* + (1-\varepsilon_j)(1-Y_j^*) \quad ; \quad j=1, \dots, q.$$

On a évidemment

$$\begin{aligned} P[Y_j = 1] &= E(\varepsilon_j Y_j^*) + E((1-\varepsilon_j)(1-Y_j^*)) \\ &= \alpha_j p_j^* + (1-\alpha_j)(1-p_j^*) \quad ; \quad j=1, \dots, q. \end{aligned}$$

En ce qui concerne la loi conjointe de Y_1, \dots, Y_q on a :

Lemme 7.

La loi P_Y du vecteur aléatoire $Y = (Y_1, \dots, Y_q)$ est donnée par :

$$\begin{aligned} (5) \quad P_Y(\{y\}) &= (-1)^{|y|} \sum_{z \in E_q} \left[\prod_{i=1}^q (\alpha_i + z_i - 2\alpha_i z_i - y_i) \right] P_{Y^*}(\{z\}) \\ &= \sum_{z \in E_q} \left(\prod_{i \in Y \Delta Z} \alpha_i \right) \left(\prod_{i \in Y \Delta z} (1-\alpha_i) \right) P_{Y^*}(\{z\}) \end{aligned}$$

La matrice de covariance de Y a pour terme général

$$(6) \quad \text{Cov}(Y_i, Y_j) = \begin{cases} [\alpha_i p_i^* + (1-\alpha_i)(1-p_i^*)][1-\alpha_i p_i^* - (1-\alpha_i)(1-p_i^*)] & \text{si } i=j \\ (2\alpha_i - 1)(2\alpha_j - 1) \text{Cov}(Y_i^*, Y_j^*) & \text{si } i \neq j. \end{cases}$$

Preuve - La première égalité dans (5) résulte immédiatement du lemme 1.

La seconde égalité dans (5) s'obtient en remarquant que :

$$\begin{aligned}
 P[Y=y] &= \sum_{z \in E_q} P[Y=y, Y^* = z] \\
 &= \sum_{z \in E_q} P[Y^* = z; \varepsilon_j = 0, j \in \gamma \Delta z, \varepsilon_j = 1, j \in \overline{\gamma \Delta z}]
 \end{aligned}$$

et en utilisant les hypothèses d'indépendance.

La dernière assertion est une conséquence directe de la définition de Y . \square

Les propriétés suivantes de la variable S sont immédiates :

$$\begin{aligned}
 \bullet \quad E(S) &= \sum_{j=1}^q \alpha_j p_j^* + (1-\alpha_j)(1-p_j^*) \\
 \bullet \quad \text{Var}(S) &= \sum_{j=1}^q \alpha_j p_j^* + (1-\alpha_j)(1-p_j^*) [1-\alpha_j p_j^* - (1-\alpha_j)(1-p_j^*)] + \\
 &\quad + 2 \sum_{i < j} (2\alpha_i - 1)(2\alpha_j - 1) \text{Cov}(Y_i^*, Y_j^*)
 \end{aligned}$$

Nous considérons maintenant deux cas particuliers de ce modèle.

Cas particulier 3.

On fait la même hypothèse sur Y^* que celle faite dans le cas particulier 1 ci-dessus. Dans le cadre du modèle envisagé au chapitre V le vecteur Y représentera l'observation, pour un jour donné, des états des q différentes stations d'un réseau et S représentera le nombre de stations "sèches" dans le réseau ce jour-là.

On traduit immédiatement les formules (5) et (6) par

$$\begin{aligned}
 (5') \quad P_Y(\{y\}) &= (-1)^{|y|} \{ p^* [\prod_{i=1}^q (1-y_i - \alpha_i)] + (1-p^*) [\prod_{i=1}^q (\alpha_i - y_i)] \} \\
 &= p^* \prod_{i=1}^q \alpha_i^{y_i} (1-\alpha_i)^{1-y_i} + (1-p^*) \prod_{i=1}^q \alpha_i^{1-y_i} (1-\alpha_i)^{y_i}
 \end{aligned}$$

$$(6') \quad \text{Cov}(Y_i, Y_j) = \begin{cases} [\alpha_i p^* + (1-\alpha_i)(1-p^*)][1-\alpha_i p^* - (1-\alpha_i)(1-p^*)] & \text{si } i=j \\ (2\alpha_i-1)(1\alpha_j-1)p^*(1-p^*) & \text{si } i \neq j \end{cases}$$

Quant à la variable aléatoire S on déduit aisément des résultats précédents :

Lemme 8.

On a :

$$P[S = k] = p^* P[Z = k] + (1-p^*) P[Z = q-k].$$

On a aussi

$$\begin{aligned} E(S) &= p^* \sum_{j=1}^q \alpha_j + (1-p^*) (q - \sum_{j=1}^q \alpha_j) \\ \text{Var}(S) &= \sum_{j=1}^q [\alpha_j p^* + (1-\alpha_j)(1-p^*)][1-\alpha_j p^* - (1-\alpha_j)(1-p^*)] + \\ &\quad + 2p^*(1-p^*) \sum_{i < j} (2\alpha_i-1)(2\alpha_j-1). \end{aligned}$$

En ce qui concerne les lois conditionnelles des variables aléatoires Y_i sachant la variable aléatoire S on peut énoncer :

Lemme 9.

On a :

$$\begin{aligned} P[Y_i=1/S=k] &= \begin{cases} \frac{p^* \alpha_i P[Z_i=k-1] + (1-p^*)(1-\alpha_i) P[Z_i=q-k-1]}{p^* P[Z=k] + (1-p^*) P[Z=q-k]} & \text{si } 1 \leq k \leq q \\ 0 & \text{si } k=0 \end{cases} \\ P[Y=y/S=k] &= \begin{cases} \frac{p^* \prod_{i=1}^q \alpha_i^{y_i} (1-\alpha_i)^{1-y_i} + (1-p^*) \prod_{i=1}^q \alpha_i^{1-y_i} (1-\alpha_i)^{y_i}}{p^* P[Z=k] + (1-p^*) P[Z=q-k]} & \text{si } 1 \leq k \leq q \\ 1 & \text{si } k=0 \text{ et } |y|=0 \end{cases} \end{aligned}$$

Preuve

On a bien sûr d'une part :

$$P[Y_1 = 1, S = k] = \begin{cases} p^* \alpha_1 P[Z_1 = k-1] + (1-p^*) (1-\alpha_1) P[Z_1 = q-k-1] & \text{si } 1 \leq k \leq q \\ 0 & \text{si } k = 0 \end{cases}$$

et d'autre part :

$$P[Y = y, S = k] = \begin{cases} P[Y=y] & \text{si } |y| = k \\ 0 & \text{sinon} \end{cases}$$

En vertu du lemme 8 on a alors les résultats annoncés. \square

Cas particulier 4.

On fait la même hypothèse sur $\varepsilon_1, \dots, \varepsilon_q$ que celle faite dans le cas particulier 2 ci-dessus. Dans le cadre du modèle envisagé au chapitre V le vecteur aléatoire Y représentera alors l'observation, pour une station donnée, des états successifs au cours de q jours consécutifs et S représentera le nombre de jours secs parmi ces q jours.

Les formules (5) et (6) deviennent :

$$(5'') \quad P_Y(\{y\}) = (-1)^{|y|} \sum_{z \in E_q} \left[\prod_{i=1}^q (\alpha + z_i - 2\alpha z_i - y_i) \right] P_{Y^*}(\{z\})$$

$$= \sum_{z \in E_q} \alpha^{|\overline{y \Delta z}|} (1-\alpha)^{|y \Delta z|} P_{Y^*}(\{z\})$$

$$(6'') \quad \text{cov}(Y_i, Y_j) = \begin{cases} [\alpha p_i^* + (1-\alpha)(1-p_i^*)][1 - \alpha p_i^* - (1-\alpha)(1-p_i^*)] & \text{si } i = j \\ (2\alpha-1)^2 \text{cov}(Y_i^*, Y_j^*) & \text{si } i \neq j. \end{cases}$$

En ce qui concerne la variable S on peut énoncer :

Lemme 10.

On a :

$$P[S=k] = \sum_{s=0}^q \sum_{i=0}^{s \wedge k} C_s^i C_{q-s}^{k-i} \alpha^{2i-k-s+q} (1-\alpha)^{k+s-2i} P[S^*=s]$$

pour $k \in \{0, \dots, q\}$ et

$$\begin{aligned} E(S) &= (2\alpha-1) E(S^*) + q(1-\alpha) \\ m_{[2]}(S) &= (2\alpha-1)^2 m_{[2]}(S^*) + 2(2\alpha-1)(1-\alpha)(q-1)E(S^*) + q(q-1)(1-\alpha)^2. \end{aligned}$$

Preuve

On peut écrire :

$$P[S = k] = \sum_{s=0}^q P[S = k, S^* = s]$$

où pour $s \in \{0, \dots, q\}$ on a :

$$\begin{aligned} P[S=k, S^*=s] &= \sum_{y: |y|=s} P[S=k, Y^*=y] \\ &= \sum_{y: |y|=s} \sum_{z: |z|=k} P[Y^*=y; \varepsilon_j=0, j \in z\Delta y; \varepsilon_j=1, j \in \overline{z\Delta y}] \\ &= \sum_{y: |y|=s} \sum_{z: |z|=k} \alpha^{|\overline{z\Delta y}|} (1-\alpha)^{|z\Delta y|} P[Y^*=y]. \end{aligned}$$

Or on a $|z\Delta y| = |z| + |y| - 2|y \cap z|$ et $|\overline{z\Delta y}| = q - |z| - |y| + 2|y \cap z|$;
alors il vient :

$$P[S=k, S^*=s] = \alpha^{q-k-s} (1-\alpha)^{k+s} \sum_{y: |y|=s} \sum_{z: |z|=k} \left(\frac{\alpha}{1-\alpha}\right)^{2|y \cap z|} P[Y^*=y].$$

Evaluant maintenant, pour $y \in E_q$ tel que $|y| = s$, la quantité :

$$\begin{aligned} \sum_{z: |z|=k} \left(\frac{\alpha}{1-\alpha}\right)^{|y \cap z|} &= \sum_{i=0 \vee (k+s-q)}^{s \wedge k} \text{card} \{z \in E_q : |z|=k, |y \cap z|=i\} \left(\frac{\alpha}{1-\alpha}\right)^{2i} \\ &= \sum_{i=0 \vee (k+s-q)}^{s \wedge k} C_s^i C_{q-s}^{k-i} \left(\frac{\alpha}{1-\alpha}\right)^{2i} \end{aligned}$$

on obtient :

$$P[S=k, S^*=s] = \alpha^{q-k-s} (1-\alpha)^{k+s} \sum_{i=0 \vee (k+s-q)}^{s \wedge k} C_s^i C_{q-s}^{k-i} \left(\frac{\alpha}{1-\alpha}\right)^{2i} \times P[S^*=s].$$

La première assertion du lemme en découle.

Comme $S = \sum_{j=1}^q [\epsilon_j Y_j^* + (1-\epsilon_j)(1-Y_j^*)]$ on a aussi

$$\begin{aligned} E(S) &= \alpha \sum_{j=1}^q E(Y_j^*) + (1-\alpha) \left(q - \sum_{j=1}^q E(Y_j^*)\right) \\ &= \alpha E(S^*) + (1-\alpha) (q - E(S^*)) \\ &= (2\alpha-1) E(S^*) + q(1-\alpha) \end{aligned}$$

et

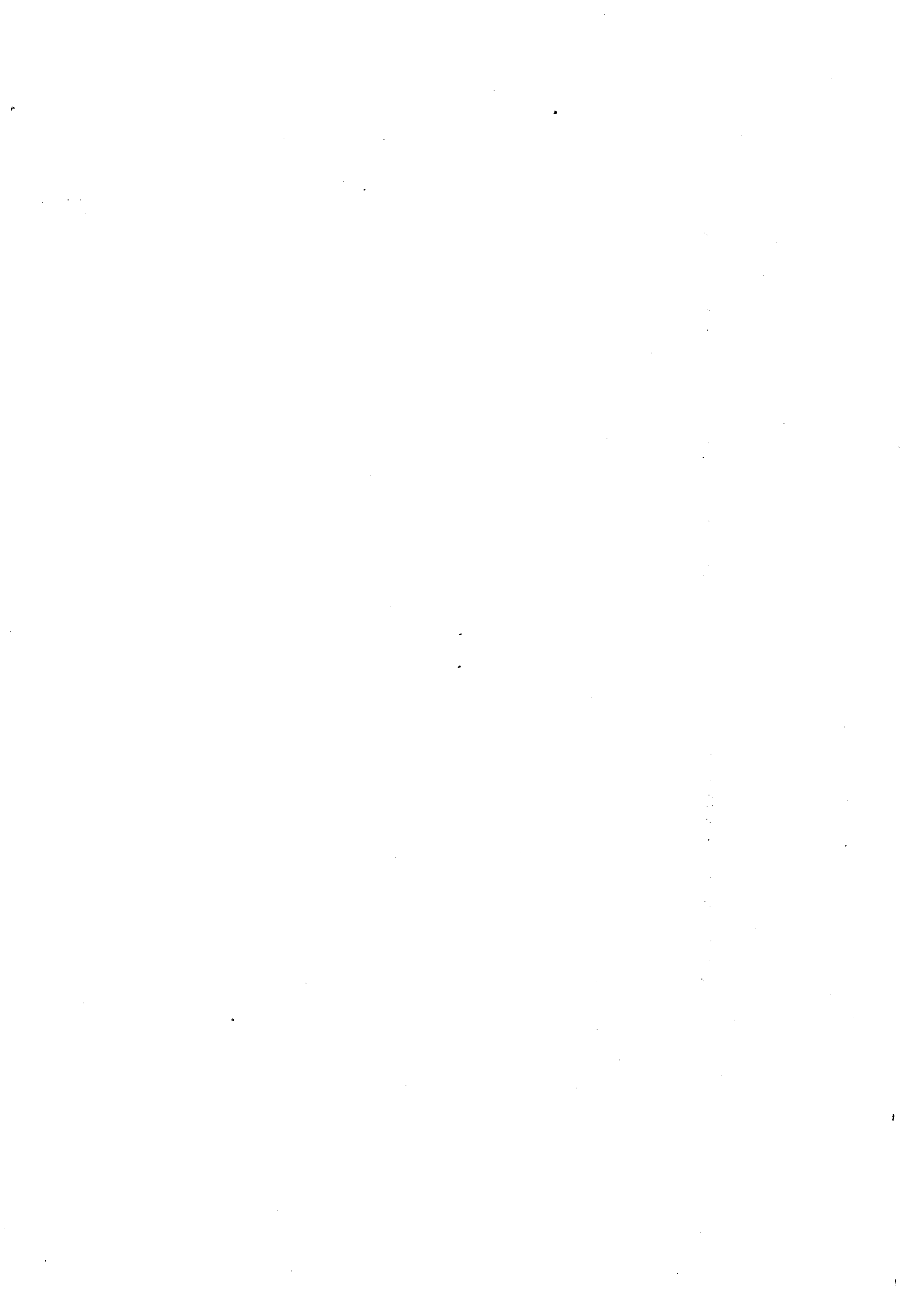
$$\begin{aligned} E(S^2) &= E\left[\sum_{j=1}^q \epsilon_j Y_j^* + (1-\epsilon_j)(1-Y_j^*)\right]^2 \\ &\quad + E\left[\sum_{i \neq j} \{\epsilon_i Y_i^* + (1-\epsilon_i)(1-Y_i^*)\} \{\epsilon_j Y_j^* + (1-\epsilon_j)(1-Y_j^*)\}\right] \\ &= (2\alpha-1) E(S^*) + q(1-\alpha) \\ &\quad + \sum_{i \neq j} \text{cov}(Y_i^*, Y_j^*) + \sum_{i \neq j} E(Y_i^*) E(Y_j^*) \\ &= (2\alpha-1) E(S^*) + q(1-\alpha) + (2\alpha-1)^2 \sum_{i \neq j} \text{cov}(Y_i^*, Y_j^*) \\ &\quad + \sum_{i \neq j} [(2\alpha-1) E(Y_i^*) + (1-\alpha)] [(2\alpha-1) E(Y_j^*) + (1-\alpha)] \end{aligned}$$

$$\begin{aligned} &= (2\alpha-1) E(S^*) + q(1-\alpha) + (2\alpha-1)^2 \sum_{i \neq j} E(Y_i^* Y_j^*) + q(q-1)(1-\alpha)^2 \\ &\quad + 2(2\alpha-1)(1-\alpha)(q-1) \sum_i E(Y_i^*) \\ &= (2\alpha-1) E(S^*) + q(1-\alpha) + (2\alpha-1)^2 [E((S^*)^2) - E(S^*)] \\ &\quad + q(q-1)(1-\alpha)^2 + 2(2\alpha-1)(1-\alpha)(q-1) E(S^*) \end{aligned}$$

d'où le résultat annoncé concernant $m_{[2]}(S)$. \square

Remarquons que $\text{Var}(S) = (2\alpha-1)^2 \text{Var}(S^*) + \alpha(1-\alpha)q$.

CHAPITRE 2 :
SERIES CHRONOLOGIQUES UNIDIRECTIONNELLES BINAIRES



Dans la suite, nous appelons, série chronologique binaire tout processus stochastique $(X_t; t \in \mathbb{N}^*)$, basé sur un certain espace probabilisé (Ω, \mathcal{A}, P) , tel que pour tout $t \in \mathbb{N}^*$ X_t est une variable aléatoire (binaire) à valeurs dans l'ensemble $\{0,1\}$. Une telle série chronologique représentera, pour une station climatologique, la suite des états successifs de jours à partir d'un premier jour $t=1$. Nous ne ferons pas toujours explicitement la distinction entre le processus stochastique $(\Omega, \mathcal{A}, P(X_t; t \in \mathbb{N}^*))$ et le processus canonique qui lui correspond défini sur $(\{0,1\}^{\mathbb{N}^*}, B_{\{0,1\}^{\mathbb{N}^*}})$ où $B_{\{0,1\}^{\mathbb{N}^*}}$ est la tribu sur $\{0,1\}^{\mathbb{N}^*}$ engendrée par les cylindres à base finie; en particulier X_t désignera indifféremment la variable aléatoire "état du processus à l'instant t " et la $t^{\text{ème}}$ application coordonnée du processus canonique.

Dans ce chapitre nous examinons d'abord les caractéristiques essentielles de la structure d'un modèle de série chronologique unidimensionnelle binaire puis nous présentons quelques modèles particuliers de processus de renouvellement alternés à partir desquels nous construisons dans les chapitres IV et V des modèles de séries chronologiques vectorielles.

I. CARACTERISTIQUES D'UNE SERIE CHRONOLOGIQUE UNIDIMENSIONNELLE BINAIRE.

Etant donné une suite infinie $(x_1, x_2, \dots, x_t, \dots) \in \{0,1\}^{\mathbb{N}^*}$, on appelle séquence de 1 (resp. 0) de longueur $k \geq 1$ commençant à l'instant $t \geq 2$ une suite $(t, \dots, t+k-1)$ d'instants telle que $x_{t-1} = x_{t+k} = 0$ (resp. 1) et $x_{t+i} = 1$ (resp. 0) pour $i=0, \dots, k-1$. Ainsi si (x_1, \dots, x_t, \dots) représente les états successifs des jours dans une station climatologique à partir d'un jour initial 1, une séquence de 1 (resp. 0) commençant au jour t est une séquence sèche (resp. humide) au sens où tous les jours secs (resp. humides) et où elle est immédiatement précédée et suivie par un jour humide (resp. sec).

La description des trajectoires d'une série chronologique binaire $(X_t, t \in \mathbb{N}^*)$ peut s'effectuer à l'aide des dates successives T_n^0, T_n^1 d'entrée dans les états 0 et 1 et des durées D_n^0, D_n^1 des séjours successifs dans ces états. De façon précise, les suites $(T_n^i; n \geq 1)$ et $(D_n^i; n \geq 1)$, $i = 0, 1$ sont définies par :

$$T_1^i = \inf \{t \geq 1 : X_t = i\} ;$$

$$D_1^i = \inf \{t > T_1^i : X_t = 1-i\} - T_1^i ;$$

$$T_n^i = \inf \{t > T_{n-1}^i + D_{n-1}^i : X_t = i\} , n > 1 ;$$

$$D_n^i = \inf \{t > T_n^i : X_t = 1-i\} - T_n^i , n > 1 .$$

Remarquons que dans le contexte climatologique, pour $n \geq 2$ la suite d'instants $T_n^i, \dots, T_n^i + D_n^i$ est une séquence sèche (resp. humide) de longueur D_n^i si $i = 1$ (resp. 0). Dans la suite nous utiliserons la terminologie correspondante.

Nous supposons désormais que la série chronologique $(X_t, t \in \mathbb{N}^*)$ est stationnaire.

I.1. Lois des temps de séjour et persistances des états

On définit la persistance au $n^{\text{ème}}$ jour de l'état i dans une séquence de i (i.e. la persistance de la sécheresse (resp. pluviosité) dans une séquence sèche (resp. humide) si $i = 1$ (resp. $i = 0$)) comme étant la probabilité que le séjour dans l'état i dure strictement plus de n jours, sachant qu'il a déjà duré n jours. On la note :

$$q_n^i = P[X_{t+n+1} = i \mid X_t = 1-i, X_{t+1} = i, \dots, X_{t+n} = i] ;$$

$$n \geq 1 ; t \geq 1 .$$

On a évidemment :

$$q_n^i = \frac{P[X_{t+n+1}=i, X_{t+n}=i, \dots, X_{t+1}=i/X_t=1-i, X_{t+1}=i]}{P[X_{t+n}=i, \dots, X_{t+1}=i/X_t=1-i, X_{t+1}=i]}$$

$$= 1 - \frac{P[X_{t+n+1}=1-i, X_{t+n}=i, \dots, X_{t+1}=i/X_t=1-i, X_{t+1}=i]}{\sum_{k \geq n} P[X_{t+k+1}=1-i, X_{t+k}=i, \dots, X_{t+1}=i/X_t=1-i, X_{t+1}=i]}$$

soit

$$(1) \quad q_n^i = 1 - \frac{p_n^i}{\sum_{k \geq n} p_k^i} ; n \geq 1$$

si on pose :

$$p_n^i = P[X_{t+n+1}=1-i, X_{t+n}=i, \dots, X_{t+1}=i/X_t=1-i, X_{t+1}=i]$$

$$= P[D_1^i = n \mid T_1^i = 2] , n \geq 1.$$

$p_n^i = (p_n^i, n \geq 1)$ est la "loi de la durée d'une séquence de i " (i.e. d'une séquence sèche (resp. humide) si $i = 1$ (resp. $i = 0$)) et est appelée loi des temps de séjour dans l'état i .

On a bien sûr la relation inverse de la relation (1) :

$$(2) \quad p_n^i = \begin{cases} 1 - q_1^i & \text{si } n = 1 \\ q_1^i q_2^i \dots q_{n-1}^i (1 - q_n^i) & \text{si } n \geq 2. \end{cases}$$

Les suites $(q_n^i ; n \geq 1)$ de persistances sont des caractéristiques importantes d'un modèle de série chronologique binaire et leur comportement (monotonie ou non, convergence ...) conditionne beaucoup son intérêt pour les applications.

Désignons par Q_n^i la probabilité qu'au cours de n instants consécutifs donnés le processus soit dans l'état i :

$$Q_n^i = P[X_t=i, \dots, X_{t+n-1}=i] ; t \geq 1 ; n \geq 1 ,$$

$$\text{soit } Q_n^1 = E\left(\prod_{t=1}^n X_t\right) \text{ et } Q_n^0 = E\left(\prod_{t=1}^n (1-X_t)\right).$$

$$\text{On a } Q_1^1 = P[X_t=1] = E(X_t) ; Q_1^0 = P[X_t=0] = 1-E(X_t)$$

$$\begin{aligned} Q_1^1 - Q_2^1 &= P[X_t=1] - P[X_t = X_{t+1} = 1] \\ &= P[X_t = 1, X_{t+1} = 0] = E(X_t) - E(X_t X_{t+1}) \end{aligned}$$

de même que

$$\begin{aligned} Q_1^0 - Q_2^0 &= P[X_t = 0] - P[X_t = X_{t+1} = 0] \\ &= P[X_t=0, X_{t+1} = 1] = E(X_{t+1}) - E(X_t X_{t+1}) \end{aligned}$$

soit

$$Q_1^1 - Q_2^1 = Q_1^0 - Q_2^0 = \frac{1}{2} P[X_t \neq X_{t+1}] .$$

La suite $(Q_n^i, n \geq 1)$ peut être reliée aux suites $(q_n^i, n \geq 1)$ et $(p_n^i, n \geq 1)$ de la manière suivante :

Lemme 1

On a, pour tout $n \geq 1$:

$$(4) \quad p_n^i = \frac{Q_n^i - 2Q_{n+1}^i + Q_{n+2}^i}{Q_1^i - Q_2^i}$$

$$(5) \quad q_n^i = \frac{Q_{n+1}^i - Q_{n+2}^i}{Q_n^i - Q_{n+1}^i}$$

Preuve

En effet, on a :

$$\begin{aligned}
 & P[X_{n+2} = 1-i, X_{n+1} = \dots = X_2 = i, X_1 = 1-i] \\
 &= P[X_{n+1} = \dots = X_2 = i] - \{P[X_{n+1} = \dots = X_2 = i, X_{n+2} = i, X_1 = 1-i] + \\
 &\quad + P[X_{n+1} = \dots = X_2 = i, X_{n+2} = 1-i, X_1 = i] + P[X_{n+1} = \dots = X_2 = i, X_{n+2} = X_1 = i]\} \\
 &= Q_n^i - \{(Q_{n+1}^i - Q_{n+2}^i) + (Q_{n+1}^i - Q_{n+2}^i) + Q_{n+2}^i\} \\
 &= Q_n^i - 2Q_{n+1}^i + Q_{n+2}^i
 \end{aligned}$$

d'où la relation (4), dont se déduit immédiatement la relation (5) d'après (1). \square

Soit $p^{1,1-i} = (p_{n,m}^{1,1-i} ; n \geq 1, m \geq 1)$ la "loi de probabilité du couple des durées d'une séquence de i et de la séquence de $1-i$ qui la suit" c'est-à-dire :

$$\begin{aligned}
 p_{n,m}^{1,1-i} &= P[X_{t+n+m+1} = i, X_{t+n+m} = \dots = X_{t+n+1} = 1-i, \\
 &\quad X_{t+n} = \dots = X_{t+1} = i | X_{t+1} = i, X_t = 1-i] \\
 &= P[D_1^i = n, D_2^{1-i} = m | T_1^i = 2] \quad n \geq 1 ; m \geq 1
 \end{aligned}$$

Désignons par $Q_{n,m}^{1,1-i}$ la probabilité qu'au cours de $n+m$ instants consécutifs donnés le processus soit d'abord n fois dans l'état i puis m fois dans l'état $1-i$:

$$Q_{n,m}^{1,1-i} = P[X_t = \dots = X_{t+n-1} = i, X_{t+n} = \dots = X_{t+n+m-1} = 1-i]$$

$$\text{soit } Q_{n,m}^{1,0} = E\left(\prod_{t=1}^n X_t \cdot \prod_{s=1}^m (1-X_{n+s})\right)$$

et

$$Q_{n,m}^{0,1} = E\left(\prod_{t=1}^n (1-X_t) \cdot \prod_{s=1}^m X_{n+s}\right)$$

On a :

$$Q_{1,1}^{1,0} = P[X_t = 1, X_{t+1} = 0] = Q_1^1 - Q_2^1 = Q_1^0 - Q_2^0$$

$$Q_{1,1}^{0,1} = P[X_t = 0, X_{t+1} = 1] = Q_1^1 - Q_2^1 = Q_1^0 - Q_2^0$$

On peut alors relier les lois $(p_{n,m}^{i,1-i} ; n \geq 1, m \geq 1) i = 0,1$ aux suites $(Q_{n,m}^{i,1-i} ; n \geq 1, m \geq 1) i = 0,1$ comme suit :

Lemme 2

On a pour tout $n \geq 1$ et $m \geq 1$

$$(7) \quad p_{n,m}^{i,1-i} = \frac{Q_{n,m}^{i,1-i} - Q_{n+1,m}^{i,1-i} - Q_{n,m+1}^{i,1-i} + Q_{n+1,m+1}^{i,1-i}}{Q_1^i - Q_2^i}$$

Preuve

En effet, on a :

$$p_{n,m}^{i,1-i} = \frac{P[X_t=1-i, X_{t+1}=...=X_{t+n}=i, X_{t+n+1}=...=X_{t+n+m}=1-i, X_{t+n+m+1}=i]}{P[X_t=1-i, X_{t+1}=i]}$$

dont le dénominateur vaut $Q_1^i - Q_2^i$ et le numérateur peut s'écrire :

$$\begin{aligned} & P[X_{t+1} = \dots = X_{t+n} = i, X_{t+n+1} = \dots = X_{t+n+m} = 1-i] \\ & - P[X_t = \dots = X_{t+n} = i, X_{t+n+1} = \dots = X_{t+n+m} = 1-i, X_{t+n+m+1} = 1-i] \\ & - P[X_t = \dots = X_{t+n} = i, X_{t+n+1} = \dots = X_{t+n+m} = 1-i, X_{t+n+m+1} = i] \\ & - P[X_t = 1-i, X_{t+1} = \dots = X_{t+n} = i, X_{t+n+1} = \dots = X_{t+n+m} = 1-i, X_{t+n+m+1} = 1-i] \\ & = Q_{n,m}^{i,1-i} - Q_{n+1,m+1}^{i,1-i} - [Q_{n+1,m}^{i,1-i} - Q_{n+1,m+1}^{i,1-i}] - [Q_{n,m+1}^{i,1-i} - Q_{n+1,m+1}^{i,1-i}] \\ & = Q_{n,m}^{i,1-i} - Q_{n+1,m}^{i,1-i} - Q_{n,m+1}^{i,1-i} + Q_{n+1,m+1}^{i,1-i} \end{aligned}$$

Cela fournit la relation (7). \square

I.2. Lois des temps de séjour résiduel, moments et fonctions génératrices

Posons, pour $i = 0, 1$

$$\begin{aligned} \hat{p}_n^i &= P[X_{t+n} = 1-i, X_{t+n-1} = \dots = X_t = i \mid X_t = i] \\ &= P[D_1^i = n \mid X_1 = i] \quad ; \quad n \geq 1. \end{aligned}$$

$\hat{p}^i = (\hat{p}_n^i ; n \geq 1)$ est la "loi du temps de séjour résiduel dans l'état i ".

Supposons que les lois des temps de séjour $(p_n^i ; n \geq 1)$, $i = 0, 1$ admettant des moments d'ordre 1 notés $m^i = \sum_{n \geq 1} n p_n^i$, $i = 0, 1$.

On peut alors relier les suites $(\hat{p}_n^i ; n \geq 1)$ aux suites $(p_n^i ; n \geq 1)$ et $(Q_n^i ; n \geq 1)$:

Lemme 3

On a les relations :

$$(8) \quad \hat{p}_n^i = \frac{1}{m^i} \sum_{k \geq n} p_k^i$$

$$(9) \quad Q_n^i = \frac{m^i}{m^0 + m^1} \sum_{k \geq n} \hat{p}_k^i = \frac{1}{m^0 + m^1} \sum_{k \geq 1} k p_{n+k-1}^i$$

Preuve

On peut bien sûr écrire :

$$\hat{p}_n^i = \frac{P[X_{t+n} = 1-i, X_{t+n-1} = \dots = X_t = i]}{P[X_t = i]} \quad ; \quad n \geq 1$$

d'où il vient :

$$Q_n^i - 2Q_{n+1}^i + Q_{n+2}^i = (\hat{p}_n^i - \hat{p}_{n+1}^i) Q_1^i \quad ; \quad n \geq 1 .$$

Alors d'après l'égalité (4), on a :

$$\hat{p}_n^i - \hat{p}_{n+1}^i = \frac{Q_1^i - Q_2^i}{Q_1^i} p_n^i ; n \geq 1 ;$$

puis

$$\hat{p}_n^i = \frac{Q_1^i - Q_2^i}{Q_1^i} \sum_{k \geq n} p_k^i ; n \geq 1 .$$

Comme $\sum_{k \geq 1} \hat{p}_n^i = 1$, on obtient :

$$\sum_{n \geq 1} \sum_{k \geq n} p_k^i = \sum_{k \geq 1} k p_k^i = \frac{Q_1^i}{Q_1^i - Q_2^i}$$

$$\text{soit } \frac{Q_1^i}{Q_1^i - Q_2^i} = m^i$$

Puisque $Q_1^1 - Q_2^1 = Q_1^0 - Q_2^0$ et $Q_1^1 + Q_1^0 = 1$, on en déduit que

$$Q_1^0 = \frac{m^0}{m^0 + m^1} ; Q_1^1 = \frac{m^1}{m^0 + m^1} ; Q_1^1 - Q_2^1 = \frac{1}{m^0 + m^1} .$$

Cela démontre l'égalité (8). L'égalité (9) s'en déduit immédiatement car d'après (4), on a :

$$\sum_{k \geq n} p_k^i = \frac{Q_n^i - Q_{n+1}^i}{Q_1^i - Q_2^i} . \quad \square$$

Remarquons qu'on a en particulier :

$$P[X_t = i] = \frac{m^i}{m^0 + m^1} \text{ et } P[X_t = i, X_{t+1} = 1-i] = \frac{1}{m^0 + m^1} ; i = 0, 1 .$$

Désignons par P^i et \tilde{P}^i les fonctions génératrices respectives des lois $p^i = (p_n^i ; n \geq 1)$ et $\hat{p}^i = (\hat{p}_n^i ; n \geq 1)$:

$$P^i(t) = \sum_{n \geq 1} t^n p_n^i \text{ et } \tilde{P}^i(t) = \sum_{n \geq 1} t^n \hat{p}_n^i , t \in [-1, +1] .$$

Remarquons que, d'après la relation (9) la fonction génératrice Q^1 de la suite $(Q_n^1; n \geq 1)$, $Q^1(t) = \sum_{n \geq 1} t^n Q_n^1$, est bien définie pour $t \in [-1, +1]$ lorsque la loi $\sum_{n \geq 1} (p_n^1; n \geq 1)$ admet un moment d'ordre 2 (la loi $\hat{p}^1 = (\hat{p}_n^1, n \geq 1)$ admettant alors, d'après (8), un moment d'ordre 1).

Les relations du lemme 3 se traduisent alors par :

Lemme 4

On a :

$$(10) \quad \hat{P}^1(t) = \frac{t}{m^1(1-t)} (1 - P^1(t))$$

$$(11) \quad Q^1(t) = \frac{m^1}{m^0 + m^1} \frac{t}{1-t} (1 - \hat{P}^1(t))$$

$$(12) \quad P^1(t) = (m^0 + m^1) \cdot \frac{(1-t)^2}{t^2} \cdot Q^1(t) - m^1 \cdot \frac{1-t}{t} + 1.$$

Preuve

D'après (8), on a :

$$\begin{aligned} \hat{P}^1(t) &= \frac{1}{m_1} \sum_{n \geq 1} \left(\sum_{k \geq n} p_k^1 \right) t^n = \frac{1}{m_1} \sum_{k \geq 1} \left(\sum_{n=1}^k t^n \right) p_k^1 \\ &= \frac{1}{m_1} \frac{t}{(1-t)} \sum_{k \geq 1} (1-t^k) p_k^1 \\ &= \frac{1}{m_1} \frac{t}{1-t} \left(1 - \sum_{k \geq 1} t^k p_k^1 \right) \end{aligned}$$

d'où la relation (10).

Un calcul analogue conduit à la relation (11) à partir de (9). La relation (12) s'obtient immédiatement par "inversion" de (10) et (11). \square

Notons par $m_{[k]}^i$ le moment factoriel d'ordre k de la loi $p^i = (p_n^i; n \geq 1)$ lorsqu'il existe i.e.

$$m_{[k]}^i = \frac{d^k}{dt^k} P^i(t) /_{t=1} = \sum_{n \geq k} n(n-1)\dots(n-k+1) p_n^i$$

(on a bien sûr $m_{[1]}^i = m^i$). De même, notons par $\hat{m}_{[k]}^i$ le moment factoriel d'ordre k de la loi $\hat{p}^i = (\hat{p}_n^i; n \geq 1)$ lorsqu'il existe.

On déduit immédiatement du lemme 4

Lemme 5

On a :

$$\hat{m}_{[k]}^i = \frac{1}{m^i} (m_{[k]}^i + \frac{1}{k+1} m_{[k+1]}^i) ; k \geq 1$$

et

$$m_{[k]}^i = 2(m^0 + m^1) Q^i(1) - 2 m^i$$

lorsque ces moments existent.

I.3. Lois marginales de dimensions finies et statistiques de comptage

L'analyse statistique d'un modèle de série chronologique binaire conduit à étudier les propriétés de certaines caractéristiques liées à l'observation du processus au cours d'une suite finie d'instantanés $1, 2, \dots, T$, le calcul de la fonction de vraisemblance correspondante (i.e. de la loi marginale du vecteur aléatoire (X_1, \dots, X_T)) ayant éventuellement permis de mettre en évidence l'exhaustivité de ces statistiques.

Les statistiques les plus importantes tant du point de vue de l'estimation des paramètres d'un modèle que de la comparaison entre un modèle estimé et les observations sont les statistiques de comptage :

$$S_T = \sum_{t=1}^T X_t \quad \text{et} \quad T - S_T$$

qui comptent respectivement le nombre de jours secs et humides sur la période $\{1, \dots, T\}$, et

$$N_T^{ij} = \sum_{t=2}^T (-1)^{i+j} (X_{t-1} + i - 1) (X_t + j - 1) ; \quad i, j \in \{0, 1\}$$

qui comptent le nombre de passages de l'état i à l'état j entre le jour 1 et le jour T .

Notons que les probabilités $P[S_T = T]$ et $P[S_T = 0]$ ne sont autres que les quantités Q_T^1 et Q_T^0 définies par (3).

On est alors amené à étudier la loi de la variable somme des composantes de certains vecteurs aléatoires (cf. chapitre I).

Notons aussi que les caractéristiques du second ordre des statistiques S_T et $T - S_T$ sont données par :

$$(13) \quad \begin{aligned} E(S_T) &= Tp ; \quad E(T - S_T) = T(1-p) \\ \text{Var}(S_T) &= \text{Var}(T - S_T) = T C(0) + 2 \sum_{k=1}^{T-1} (T-k) C(k) \end{aligned}$$

lorsque la série chronologique $(X_t ; t \geq 1)$ est stationnaire avec $p = P[X_t = 1]$ et pour fonction de covariance :

$$C(k) = \text{cov}(X_t, X_{t+k}) = p \{P[X_{k+1} = 1 \mid X_1 = 1] - p\} ; \quad k \in \mathbb{N} .$$

II - MODELES DE SERIES CHRONOLOGIQUES BINAIRES

Nous présentons ici brièvement quelques modèles particuliers qui ont été utilisés avec succès pour la modélisation de données climatologiques notamment ceux dont nous ferons usage dans les chapitres IV et V pour construire des modèles de séries chronologiques vectorielles. Nous renvoyons à BOUSSEBOUA (1983) pour une étude détaillée de ces modèles.

II.1. Processus de renouvellement alterné à temps discret

Les processus de renouvellement alternés ont déjà été abondamment utilisés pour la modélisation en climatologie (cf. par exemple GREEN (1964), BUIHAND (1977), GALLOY - LE BRETON - MARTIN (1983)).

Ils traduisent l'hypothèse a priori qu'à partir de chaque instant de changement d'état le phénomène à représenter ne dépend du passé qu'à travers l'état à cet instant.

De façon précise, étant donné $0 \leq p \leq 1$ et $\tilde{p}^i = (\tilde{p}_n^i ; n \geq 1)$, $p^i = (p_n^i ; n \geq 1)$, $i = 0, 1$ quatre lois de probabilité sur \mathbb{N}^* on pose :

Définition 1

La série chronologique binaire $(X_t ; t \geq 1)$ est dite processus de renouvellement alterné de paramètres $(p, p^i, \tilde{p}^i, i = 0, 1)$ si $P[X_1 = 1] = p$ et pour $i \in \{0, 1\}$, conditionnellement à $X_1 = i$, les durées de séjour $(D_n^i ; n \geq 1)$ (resp. $(D_n^{1-i} ; n \geq 1)$) sont indépendantes de loi \tilde{p}^i pour $n = 1$ et de même loi p^i pour $n \geq 2$ (resp. de même loi p^{1-i} pour $n \geq 1$) les suites $(D_n^0 ; n \geq 1)$ et $(D_n^1 ; n \geq 1)$ étant de plus indépendantes entre elles.

Selon la terminologie déjà utilisée, les lois p^0 et p^1 sont les lois des temps de séjour dans l'état 0 et l'état 1 respectivement.

Etant donné deux lois p^0 et p^1 sur \mathbb{N}^* , apériodiques et admettant des moments d'ordre 1 m^0 et m^1 , on peut montrer (cf. par exemple GREGOIRE (1982)) :

Lemme 6

Le processus de renouvellement alterné $(X_t, t \geq 1)$ de paramètres $(p, \tilde{p}^0, \tilde{p}^1, p^0, p^1)$ est stationnaire si et seulement si les deux conditions suivantes sont satisfaites :

$$i) p = \frac{m^1}{m^0 + m^1}$$

$$ii) \tilde{p}_k^i = \tilde{p}_k^i = \frac{1}{m_i} \sum_{\ell \geq 0} p_{k+\ell}^i ; k \geq 1 ; i \in \{0, 1\}.$$

Nous dirons alors que la série chronologique binaire $(X_t, t \in \mathbb{N}^*)$ est un processus de renouvellement alterné stationnaire de paramètres p^0 et p^1 et noterons $(X_t, t \in \mathbb{N}^*) \sim R.A. (p^0, p^1)$.

Les différentes caractéristiques du modèle (Cf. I de ce chapitre) sont alors complètement connues. Notons qu'en particulier les lois $p^{i, 1-i}$ ne sont autres que les lois produits $p^i \otimes p^{1-i}$.

Signalons que l'on dispose d'un algorithme de calcul de loi de la variable aléatoire S_T (cf. chapitre III) et du résultat asymptotique suivant concernant sa variance (cf. par exemple BOUSSEBOUA (1983)) :

$$(14) \text{Var}(S_T) = \frac{(m^1)^2 m_2^0 + (m^0)^2 m_2^1}{(m^0 + m^1)^3} \cdot T + \frac{(m_1^0 m_2^0 - m^0 m_2^1)^2}{2(m^0 + m^1)^4} - \\ - \frac{(m^1)^2 m_3^0 + (m^0)^2 m_3^1}{3(m^0 + m^1)^3} + \frac{2m^0 m^1 + (m^0 m^1)^2}{6(m^0 + m^1)^2} + o(1)$$

où m_k^i est le moment centré d'ordre k (i.e. la variance si $k = 2$) de la loi p^i ; $i = 0, 1$.

II.2. Chaînes de Markov binaires d'ordre 1 et 2

GABRIEL - NEUMAN (1962), CASKEY (1963), WEISS (1964), HOPKINS - ROBILLARD (1964), KATZ (1974), TODOROVIC - WOOLHISER (1975), BERGER - GOOSSENS (1983) ... ont utilisé les chaînes de Markov binaires pour des applications en climatologie. Jusqu'à l'ordre 2 de tels processus sont (cf. par exemple BOUSSEBOUA (1983)) des processus de renouvellement alternés.

Si $(X_t ; t \geq 1)$ est une chaîne de Markov binaire d'ordre 1, stationnaire de paramètres $p = P[X_t = 1]$ et $\lambda = P[X_t = 1 / X_{t-1} = 1]$ où $0 < p < 1$, $\max(0, \frac{2p-1}{p}) \leq \lambda < 1$ (on notera $(X_t ; t \geq 1) \sim M(1 ; p, \lambda)$) alors $(X_t, t \geq 1) \sim R.A.(G(p(\frac{1-\lambda}{1-p})), G(1-\lambda))$ où $G(\mu)$ pour $0 < \mu < 1$ désigne la loi géométrique de paramètre μ i.e.

$$P^0(t) = \frac{1 - \lambda}{1 - p - t(1 - 2p + \lambda p)} \cdot p^t, \quad P^1(t) = \frac{1 - \lambda}{1 - \lambda t} t$$

Notons que les lois de temps de séjour résiduel \hat{p}^i coïncident avec les lois des temps de séjour p^i correspondantes et que les suites des persistances sont constantes égales à :

$$q_1^1 = \lambda \quad \text{et} \quad q_1^0 = \frac{1 - 2p + \lambda p}{1 - p} .$$

On a aussi :

$$m^0 = \frac{1 - p}{p(1 - \lambda)} \quad ; \quad m^1 = \frac{1}{1 - \lambda}$$

$$m_2^0 = \frac{(1-p)(1-2p+\lambda p)}{p^2(1-\lambda)^2} \quad ; \quad m_2^1 = \frac{\lambda}{(1-\lambda)^2} .$$

L'estimation des paramètres peut être menée à l'aide des statistiques S_T et N_T^{ij} .

Si $(X_t ; t \geq 1)$ est une chaîne de Markov binaire d'ordre 2 stationnaire de paramètres $P[X_t = 1] = p$, $P[X_{t+1} = 1 | X_t = 1] = \lambda_1$, $P[X_{t+2} = 1 | X_t = 1] = \lambda_2$ et $P[X_{t+2} = 1 | X_{t+1} = X_t = 1] = \mu$ où $0 < p < 1$, $\max\{0, \frac{2p-1}{p}\} \leq \lambda_1 \leq 1$, $\max\{0, \frac{2\lambda_1-1}{1}\} \leq \mu < 1$ et $\max \frac{3p-2p\lambda_1-1+p\lambda_1\mu}{p+\lambda_1\mu}, \lambda_1\mu \leq \lambda_2 < 1-\lambda_1(1-\mu)$ (on notera $(X_t ; t \geq 1) \sim M(2, p, \lambda_1, \lambda_2, \mu)$) alors $(X_t ; t \geq 1) \sim R.A.(p^0, p^1)$ où

$$p_n^0 = \begin{cases} (\lambda_2 - \lambda_1\mu)/(1 - \lambda_1) & \text{si } n = 1 \\ \frac{p(1-\lambda_2-\lambda_1+\lambda_1\mu)^2(1-p(3-\lambda_2-2\lambda_1+\lambda_1\mu))^{n-2}}{(1-\lambda_1) \cdot (1-2p+p\lambda_1)^{(n-1)}} & \text{si } n \geq 2 \end{cases}$$

$$p_n^1 = \begin{cases} (1-2\lambda_1+\lambda_1\mu)/(1-\lambda_1) & \text{si } n = 1 \\ \lambda_1(1-\mu)^2 \mu^{n-2}/(1-\lambda_1) & \text{si } n \geq 2 \end{cases}$$

$$\text{i.e. } P^0(t) = \frac{\lambda_0 - \lambda_1\mu}{1-\lambda_1} t + \frac{(1-\lambda_2-\lambda_1+\lambda_1\mu)^2}{1-2p+p\lambda_1-t(1-p(3-\lambda_2-2\lambda_1+\lambda_1\mu))} \cdot \frac{1}{1-\lambda_1} \cdot pt^2$$

$$P^1(t) = \frac{1-2\lambda_1+\lambda_1\mu}{1-\lambda_1} t + \frac{(1-\mu)^2}{(1-\lambda_1)(1-t\mu)} \cdot \lambda_1 t^2$$

Notons que les suites de persistances sont constantes au-delà du deuxième jour :

$$q_n^0 = \begin{cases} \frac{1-\lambda_2-\lambda_1+\lambda_1\mu}{1-\lambda_1} & \text{si } n = 1 \\ \frac{1-p(3-\lambda_2-2\lambda_1+\lambda_1\mu)}{1-2p+p\lambda_1} & \text{si } n \geq 2 \end{cases}$$

$$q_n^1 = \begin{cases} \frac{\lambda_1(1-\mu)}{1-\lambda_1} & \text{si } n = 1 \\ \mu & \text{si } n \geq 2 \end{cases}$$

et les lois de temps de séjour résiduels \hat{p}^0 et \hat{p}^1 s'écrivent :

$$\hat{p}_n^0 = \begin{cases} p(1-\lambda_1)/(1-p) & \text{si } n = 1 \\ \frac{p(1-\lambda_2-\lambda_1+\lambda_1\mu)}{(1-p)} \cdot \frac{(1-p(3-\lambda_2-2\lambda_1+\lambda_1\mu))^{n-2}}{(1-2p+p\lambda_1)} & \text{si } n \geq 2 \end{cases}$$

$$\hat{p}_n^1 = \begin{cases} 1-\lambda_1 & \text{si } n = 1 \\ \lambda_1(1-\mu) \cdot \mu^{n-2} & \text{si } n \geq 2 \end{cases}$$

On a aussi :

$$m^0 = \frac{1-p}{p(1-\lambda_1)} ; m^1 = \frac{1}{1-\lambda_1}$$

$$m_2^0 = \frac{1-2p+p\lambda_1}{p^2(1-\lambda_1)^2(1-\lambda_2-\lambda_1+\lambda_1\mu)} \{ 1-\lambda_1+\lambda_2-\lambda_1\mu - p(3-5\lambda_1+\lambda_2+2(\lambda_1)^2-\lambda_1\mu) \} ;$$

$$m_2^1 = \frac{\lambda_1(1-2\lambda_1+\mu)}{(1-\mu)(1-\lambda_1)^2}$$

L'estimation des paramètres peut être fondée sur les statistiques S_T et $N_T^{ijk} = \text{card} \{ t \in \{1, \dots, T-2\}, X_t = i, X_{t+1} = j, X_{t+2} = k \}$ $i, j, k \in \{0, 1\}$.

II.3. Processus de renouvellement alternés basés sur des lois binômiales négatives translatées

GRISOLET - GUILMET - ARLERY (1962), BUIHAND (1977), LE BRETON - MARTIN (1979), GALLOY (1982), GALLOY - LE BRETON - MARTIN (1983), ont noté la bonne adéquation des distributions de longueurs de séquences climatologiques à des lois binômiales négatives translatées (B.N.T.).

GALLOY - LE BRETON - MARTIN (1983) ont modélisé par des processus R.A. basés sur de telles lois la succession des séquences sèches et humides dans plusieurs stations climatologiques françaises.

Etant donné h et d deux paramètres réels positifs la loi $(p_k; k \geq 1)$ sur \mathbb{N}^* est dite B.N.T. (h, d) si :

$$p_k = \frac{1}{(1+d)^{h/d}} \frac{(h/d)_{k-1}}{(h-1)!} \left(\frac{d}{1+d}\right)^{k-1}; k \geq 1$$

où, pour a réel et n entier on a posé :

$$(a)_n = \begin{cases} (a+n-1)(a+n-2)\dots(a+1)a & \text{si } n \geq 1, \\ 1 & \text{si } n = 0. \end{cases}$$

Elle a pour fonction génératrice :

$$P(t) = \frac{t}{(1+d)^{h/d}} \left(1 - \frac{dt}{1+d}\right)^{-h/d}$$

pour moyenne $m = h + 1$ et pour variance $m_2 = h(d+1)$.

Si $(X_t ; t \geq 1) \sim \text{R.A. (B.N.T. } (h_0, d_0), \text{ B.N.T. } (h_1, d_1))$ alors la suite des persistances de l'état $i, i = 0, 1$ est monotone croissante (resp. décroissante) de limite $\frac{d_i}{1+d_i}$ si $\frac{d_i}{h_i} > 1$ (resp. $\frac{d_i}{h_i} < 1$) et constante égale à $\frac{d_i}{d_i+1}$ si $\frac{d_i}{h_i} = 1$.

On peut construire des estimations des paramètres à l'aide de la statistique S_T et des statistiques concernant les longueurs de séquences observées.

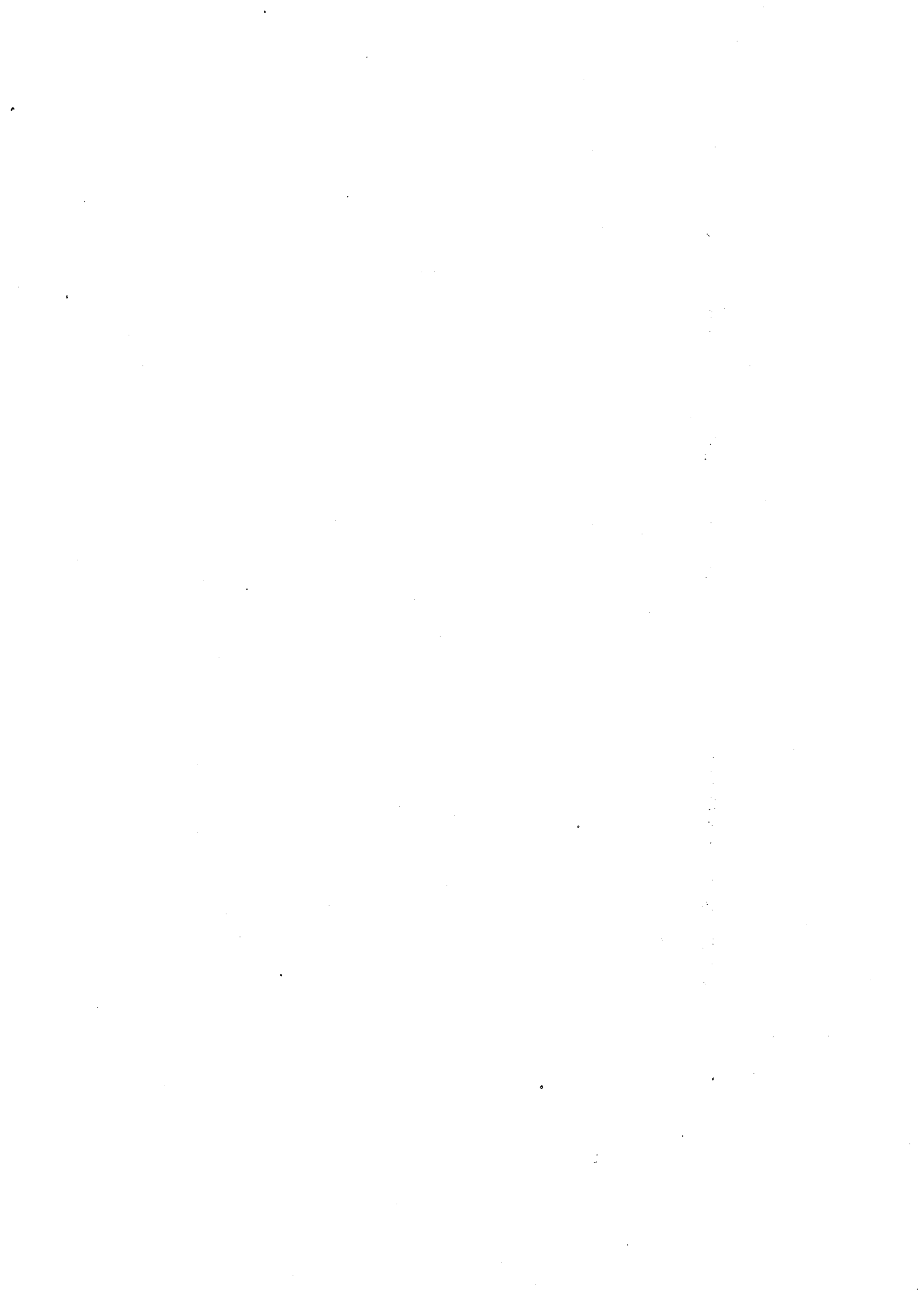
II.4. Autres modèles de séries chronologiques binaires

Certains auteurs ont eu recours à des chaînes de Markov d'ordre 3, 4 ... pour modéliser le phénomène climatologique (cf. par exemple BERGER - GOOSSENS (1982)) mais ces modèles (qui ne conduisent plus à des processus de renouvellement alternés) deviennent lourds à manipuler compte tenu en particulier du nombre élevé des paramètres qu'ils contiennent.

Des analogues des modèles linéaires classiques autorégressifs, moyennes mobiles et mixtes pour les processus à valeurs réelles ont été proposés par JOCOB-LEWIS (1978 a, 1978 b) pour les séries chronologiques à valeurs discrètes. BUSHAND (1978), CHANG - DELLEUR - KAVVAS (1982, 1983 a, 1983 b) ont utilisé avec succès des versions de ces modèles dans des études en climatologie. Nous renvoyons à ces auteurs et à BOUSSEBOUA (1983) pour une étude approfondie de ces modèles et des applications.

CHAPITRE 3 :

CALCUL DE CERTAINES FONCTIONS GENERATRICES ASSOCIEES
A UN PROCESSUS DE RENOUVELLEMENT ALTERNE



Les séries chronologiques marginales des séries chronologiques vectorielles à composantes binaires que nous définissons dans les chapitres IV et V sont construites à partir d'un même processus de renouvellement alterné. L'étude de leurs caractéristiques, telles qu'elles ont été définies dans le chapitre II, passe en particulier par le calcul de certaines fonctions génératrices associées au dit processus de renouvellement alterné.

Dans ce chapitre, nous décrivons d'abord un algorithme de calcul de la loi de la statistique de comptage $S_n = \sum_{t=1}^n X_t$ dans un processus $(X_t; t \geq 1) \sim \text{R.A.}(p^0, p^1)$ puis nous déterminons des fonctions génératrices liées à des suites intervenant dans cet algorithme.

I. ALGORITHME DE CALCUL DE LA LOI DE LA VARIABLE ALEATOIRE S_n .

On considère donc une série chronologique $(X_t; t \geq 1) \sim \text{R.A.}(p^0, p^1)$ dont les lois p^0 et p^1 de temps de séjour ont pour moyennes respectives m^0 et m^1 . On s'intéresse au calcul de la loi de la statistique de comptage

$$S_n = X_1 + X_2 + \dots + X_n$$

on a évidemment , pour $m \in \{0, \dots, n\}$

$$P[S_n = m] = P[S_n = m/X_1 = 1] P[X_1 = 1] + P[S_n = m/X_1 = 0] P[X_1 = 0]$$

soit

$$P[S_n = m] = \frac{1}{m^0 + m^1} \{ m^0 \tilde{R}_n^0(m) + m^1 \tilde{R}_n^1(m) \}$$

si on pose, pour $i \in \{0, 1\}$

$$(1) \quad \tilde{R}_n^i(m) = P[S_n = m / X_1 = i]$$

Le calcul de la loi de la statistique S_n repose alors sur celui des suites $\{\tilde{R}_n^0(m); 0 \leq m \leq n, n \geq 1\}$ et $\{\tilde{R}_n^1(m); 0 \leq m \leq n, n \geq 1\}$.

Buishand (1977) a proposé un algorithme de calcul de ces suites (cf. aussi Bousseboua (1983)) fondé sur le calcul intermédiaire des suites $\{R_n^0(m); 0 \leq m \leq n, n \geq 1\}$ et $\{R_n^1(m); 0 \leq m \leq n, n \geq 1\}$ où, pour $i \in \{0, 1\}$:

$$(2) \quad R_n^i(m) = P[S_n = m / X_1 = i, X_2 = 1-i]$$

Lemme 1.

Les suites $\{R_n^i(m); 0 \leq m \leq n, n \geq 1\}$, $i \in \{0, 1\}$ définies par (2) vérifient :

. pour $n=1$

$$R_1^i(1) = \delta_{i1} \quad , \quad R_1^i(0) = 1 - \delta_{i1} \quad ,$$

. pour $n \geq 2$

$$R_n^0(0) = R_n^0(n) = R_n^1(0) = R_n^1(n) = 0 \quad ,$$

$$R_n^0(n-1) = m^1 \hat{p}_{n-1}^1 \quad , \quad R_n^1(1) = m^0 \hat{p}_{n-1}^0$$

$$R_n^0(1) = \begin{cases} 1 & \text{si } n=2 \\ m^0 p_1^0 \hat{p}_{n-2}^0 & \text{si } n \geq 3 \end{cases} \quad R_n^1(n-1) = \begin{cases} 1 & \text{si } n=2 \\ m^1 p_1^0 \hat{p}_{n-2}^1 & \text{si } n \geq 3 \end{cases}$$

. pour $n \geq 3$

$$R_n^0(m) = \sum_{j=1}^m p_j^1 R_{n-j}^1(m-j+1) \quad \text{si } 1 \leq m \leq n-2$$

$$R_n^1(m) = \sum_{j=1}^{n-m} p_j^0 R_{n-j}^0(m-1) \quad \text{si } 2 \leq m \leq n-1.$$

Preuve - Les résultats des deux premiers cas s'obtiennent facilement. Pour $n \geq 3$, $1 \leq m \leq n-2$, on peut écrire

$$\begin{aligned}
 R_n^0(m) &= \sum_{j=1}^m P[X_2=1, \dots, X_{j+1}=1, X_{j+2}=0, \sum_{k=j+1}^n X_k=m-j+1 / X_1=0, X_2=1] \\
 &= \sum_{j=1}^m P[X_2=1, \dots, X_{j+1}=1, X_{j+2}=0 / X_1=0, X_2=1] \times \\
 &\quad \times P[\sum_{k=j+1}^n X_k=m-j+1 / X_{j+1}=1, X_{j+2}=0] \\
 &= \sum_{j=1}^m p_j^1 R_{n-j}^1(m-j+1) .
 \end{aligned}$$

De même pour $n \geq 3$, $2 \leq m \leq n-1$, on peut écrire

$$\begin{aligned}
 R_n^1(m) &= \sum_{j=1}^{n-m} P[X_2=0, \dots, X_{j+1}=0, X_{j+2}=1, \sum_{k=j+1}^m X_k=m-1 / X_1=1, X_2=0] \\
 &= \sum_{j=1}^{n-m} P[X_2=0, \dots, X_{j+1}=0, X_{j+2}=1 / X_1=1, X_2=0] \times \\
 &\quad \times P[\sum_{k=j+1}^n X_k=m-1 / X_{j+1}=0, X_{j+2}=1] \\
 &= \sum_{j=1}^{n-m} p_j^0 R_{n-j}^0(m-1) . \quad \square
 \end{aligned}$$

On peut maintenant relier les suites $\{\tilde{R}_n^i(m); 0 \leq m \leq n, n \geq 1\}$, $i \in \{0,1\}$ définies par (1) aux suites $\{R_n^i(m); 0 \leq m \leq n, n \geq 1\}$ définies par (2) et fournies par le lemme 1 :

Lemme 2.

On a :

. pour $n = 1$

$$\tilde{R}_1^1(0) = 1 - \delta_{11} \quad ; \quad \tilde{R}_1^1(1) = \delta_{11} ,$$

. pour $n \geq 2$

$$\tilde{R}_n^0(n) = \tilde{R}_n^1(0) = 0 ,$$

$$\begin{aligned} \tilde{R}_n^0(0) &= \sum_{j \geq n} \hat{p}_j^0 & ; & \quad \tilde{R}_n^1(n) = \sum_{j \geq n} \hat{p}_j^1 , \\ \tilde{R}_n^0(n-1) &= \frac{1}{m^0} \cdot \sum_{j \geq n-1} p_j^1 = \frac{m^1}{m^0} \hat{p}_{n-1}^1 , \\ \tilde{R}_n^1(1) &= \frac{1}{m^1} \cdot \sum_{j \geq n-1} p_j^0 = \frac{m^0}{m^1} \hat{p}_{n-1}^0 , \end{aligned}$$

. pour $n \geq 3$ et $1 \leq m \leq n-1$

$$\tilde{R}_n^0(m) = \sum_{j=1}^{n-m} \hat{p}_j^0 R_{n-j+1}^0(m) \quad ; \quad \tilde{R}_n^1(m) = \sum_{j=1}^m \hat{p}_j^1 R_{n-j+1}^1(m-j+1) .$$

Preuve - Les résultats des deux premiers cas s'obtiennent immédiatement. Pour $n \geq 3$, $1 \leq m \leq n-1$, on peut écrire

$$\begin{aligned} \tilde{R}_n^0(m) &= \sum_{j=1}^{n-m} P[X_1=0, \dots, X_j=0, X_{j+1}=1, \sum_{k=j}^n X_k=m/X_1=0] \\ &= \sum_{j=1}^{n-m} P[X_1=0, \dots, X_j=0, X_{j+1}=1/X_1=0] \times \\ &\quad \times P[\sum_{k=j}^n X_k=m/X_j=0, X_{j+1}=1] \\ &= \sum_{j=1}^{n-m} \hat{p}_j^0 R_{n-j+1}^0(m) . \end{aligned}$$

De même, on peut écrire

$$\begin{aligned} \tilde{R}_n^1(m) &= \sum_{j=1}^m P[X_1=1, \dots, X_j=1, X_{j+1}=0, \sum_{k=j}^n X_k=m-j+1/X_1=1] \\ &= \sum_{j=1}^m P[X_1=1, \dots, X_j=1, X_{j+1}=0/X_1=1] \times \\ &\quad \times P[\sum_{k=j}^n X_k=m-j+1/X_j=1, X_{j+1}=0] \\ &= \sum_{j=1}^m \hat{p}_j^1 R_{n-j+1}^1(m-j+1) . \quad \square \end{aligned}$$

Remarque 1 - Dans le cas particulier d'un processus $(X_t; t \geq 1) \sim MA(1; p, \lambda)$ ou $(X_t; t \geq 1) \sim MA(1; p, \lambda_1, \lambda_2, \mu)$ (cf. chapitre II §II.) on

dispose d'algorithmes (spécifiques) équivalents à celui fourni par les lemmes 1 et 2 (cf. par exemple Kedem (1980), Bousseboua (1983)).

Remarquons aussi que dans Gabriel (1959), Helgert (1970), Klotz (1972), Kedem (1977) on trouve, pour ces modèles markoviens, la loi de S_n sous forme explicite.

II. FONCTIONS GENERATRICES DES SUITES $\{R_n^i(k); n \geq 1\}; i \in \{0,1\}, k \in \mathbb{N}$.

On considère, pour $i \in \{0,1\}$ et $k \in \mathbb{N}$ la fonction génératrice $t \rightarrow R_t^i(k)$ de la suite $\{R_n^i(k); n \geq 1\}$ définie par (2) i.e.

$$(3) \quad R_t^i(k) = \sum_{n \geq 1} R_n^i(k) t^n \quad ; t \in [-1,+1].$$

A l'aide du lemme 1 on peut montrer :

Lemme 3.

Les fonctions génératrices $t \rightarrow R_t^1(k)$ et $t \rightarrow R_t^0(k)$, $k \in \mathbb{N}$ définies par (3) vérifient les relations :

$$R_t^1(k) = \begin{cases} 0 & \text{si } k=0 \\ t(1 + m^0 \hat{P}^0(t)) & \text{si } k=1 \\ P^0(t) R_t^0(k-1) & \text{si } k \geq 2 \end{cases}$$

$$R_t^0(k) = \begin{cases} t & \text{si } k=0 \\ m^1 t^{k+1} \hat{P}_{k+1}^1 + \sum_{j=1}^k t^j p_j^1 R_t^1(k-j+1) & \text{si } k \geq 1 \end{cases}$$

où P^0 et \hat{P}^0 sont les fonctions génératrices respectives des lois $p^0 = (p^0(n); n \geq 1)$ et $\hat{p}^0 = (\hat{p}^0(n); n \geq 1)$ (cf. chapitre II §I.2)

Preuve - Remarquons que le fait que $R_t^i(k) < +\infty$ pour $|t|=1$ découlera de la démonstration.

Pour $k=0$, comme $R_n^1(0) = 0$ pour tout $n \geq 1$, on a

$$R_t^1(0) = \sum_{n \geq 1} t^n R_n^1(0) = 0$$

et comme $R_n^0(0) = \begin{cases} 1 & \text{si } n=1 \\ 0 & \text{si } n \geq 2 \end{cases}$

on a

$$R_t^0(0) = \sum_{n \geq 1} t^n R_n^0(0) = t.$$

• Pour $k=1$, puisque $R_n^1(1) = \begin{cases} 1 & \text{si } n=1 \\ m^0 \hat{p}_{n-1}^0 & \text{si } n \geq 2 \end{cases}$

$$\begin{aligned} R_t^1(1) &= \sum_{n \geq 1} t^n R_n^1(1) \\ &= t + \sum_{n \geq 2} t^n m^0 \hat{p}_{n-1}^0 \\ &= t(1 + m^0 \sum_{n \geq 2} t^n \hat{p}_{n-1}^0) \\ &= t(1 + m^0 \hat{P}^0(t)) \end{aligned}$$

• De même, pour $k=1$, comme $R_n^0(1) = \begin{cases} 0 & \text{si } n=1 \\ 1 & \text{si } n=2 \\ m^0 p_1^1 \hat{p}_{n-2}^0 & \text{si } n \geq 3 \end{cases}$

on a

$$\begin{aligned} R_t^0(1) &= \sum_{n \geq 1} t^n R_n^0(1) \\ &= t^2 + \sum_{n \geq 3} t^n m^0 p_1^1 \hat{p}_{n-2}^0 \\ &= t^2(1 + m^0 p_1^1 \sum_{n \geq 3} t^{n-2} \hat{p}_{n-2}^0) \\ &= t^2(1 + m^0 p_1^1 \hat{P}^0(t)) \end{aligned}$$

• Pour $k \geq 2$, puisque $R_n^1(k) = \begin{cases} 0 & \text{si } n \leq k \\ \sum_{j=1}^{n-k} p_j^0 R_{n-j}^0(k-1) & \text{si } n \geq k+1 \end{cases}$

on a

$$\begin{aligned} R_t^1(k) &= \sum_{n \geq 1} t^n R_n^1(k) \\ &= \sum_{n \geq k+1} t^n \left[\sum_{j=1}^{n-k} p_j^0 R_{n-j}^0(k-1) \right] \\ &= \sum_{j \geq 1} t^j p_j^0 \sum_{n \geq j+k} t^{n-j} R_{n-j}^0(k-1) \\ &= P^0(t) R_t^0(k-1) \end{aligned}$$

la dernière égalité découlant du fait que $R_n^0(k-1) = 0$ si $n \leq k-1$

. De même, pour $k \geq 2$, puisque

$$R_n^0(k) = \begin{cases} 0 & \text{si } n \leq k \\ m^1 \hat{p}_k^1 & \text{si } n = k+1 \\ \sum_{j=1}^k p_j^1 R_{n-j}^1(k-j+1) & \text{si } n \geq k+2 \end{cases}$$

on a

$$\begin{aligned} R_t^0(k) &= \sum_{n \geq 1} t^n R_n^0(k) \\ &= t^{k+1} m^1 \hat{p}_k^1 + \sum_{n \geq k+2} t^n \left[\sum_{j=1}^k p_j^1 R_{n-j}^1(k-j+1) \right] \\ &= t^{k+1} m^1 \hat{p}_k^1 + \sum_{j=1}^k t^j p_j^1 \sum_{n \geq k+2} t^{n-j} R_{n-j}^1(k-j+1) \end{aligned}$$

où

$$\begin{aligned} \sum_{n \geq k+2} t^{n-j} R_{n-j}^1(k-j+1) &= \sum_{u \geq k-j+2} t^u R_u^1(k-j+1) \\ &= \begin{cases} R_t^1(1) - t & \text{si } j=k \\ R_t^1(k-j+1) & \text{si } 1 \leq j \leq k-1 \end{cases} \end{aligned}$$

Il vient donc

$$\begin{aligned} R_t^0(k) &= t^{k+1} m^1 \hat{p}_k^1 + \sum_{j=1}^{k-1} t^j p_j^1 R_t^1(k-j+1) + t^k p_k^1 (R_t^1(1) - t) \\ &= t^{k+1} (m^1 \hat{p}_k^1 - p_k^1) + \sum_{j=1}^k t^j p_j^1 R_t^1(k-j+1) \\ &= t^{k+1} \left(\sum_{j \geq k} p_j^1 - p_k^1 \right) + \sum_{j=1}^k t^j p_j^1 R_t^1(k-j+1) \\ &= m^1 t^{k+1} \hat{p}_{k+1}^1 + \sum_{j=1}^k t^j p_j^1 R_t^1(k-j+1) . \end{aligned}$$

On vérifie enfin que l'expression précédente de $R_t^0(k)$ est valable pour $k=1$: en effet on a

$$\begin{aligned} m^1 t^2 \hat{p}_2^1 + t p_1^1 R_t^1(k) &= t^2 \left[\sum_{k \geq 2} p_k^1 + p_1^1 + m^0 p_1^1 \hat{p}^0(t) \right] \\ &= t^2 [1 + m^0 p_1^1 \hat{p}^0(t)] \end{aligned}$$

ce qui est bien la valeur trouvée ci-dessus de $R_t^0(1)$. Cela achève la démonstration du lemme. \square

III. FONCTIONS GENERATRICES DES SUITES $\{\tilde{R}_n^1(k); n \geq 1\}; i \in \{0,1\}, k \in \mathbb{N}$

On considère, pour $i \in \{0,1\}$ et $k \in \mathbb{N}$, la fonction génératrice $t \rightarrow \tilde{R}_t^i(k)$ de la suite $\{\tilde{R}_n^i(k); n \geq 1\}$ définie par (1) i.e.

$$(4) \quad \tilde{R}_t^i(k) = \sum_{n \geq 1} \tilde{R}_n^i(k) t^n \quad ; \quad t \in [-1,+1]$$

A l'aide du lemme 2 on peut montrer :

Lemme 4.

Les fonctions génératrices $t \rightarrow \tilde{R}_t^1(k)$ et $t \rightarrow \tilde{R}_t^0(k)$ définies par (4) vérifient les relations :

$$\tilde{R}_t^1(k) = \begin{cases} 0 & \text{si } k=0 \\ t^k \sum_{j \geq k+1} \tilde{p}_j^1 + \sum_{j=1}^k t^{j-1} \tilde{p}_j^1 R_t^1(k-j+1) & \text{si } k \geq 1 \end{cases}$$

$$\tilde{R}_t^0(k) = \begin{cases} \frac{t}{1-t} (1 - \tilde{p}^0(t)) & \text{si } k=0 \\ \frac{1}{t} \tilde{p}^0(t) R_t^0(k) & \text{si } k \geq 1 \end{cases}$$

où $t \rightarrow R_t^1(k)$ et $t \rightarrow R_t^0(k)$ sont fournies par le lemme 3.

Preuve.

. Pour $k=0$, comme $\tilde{R}_n^1(0) = 0$ pour tout $n \geq 1$, on a

$$\tilde{R}_t^1(0) = \sum_{n \geq 1} t^n \tilde{R}_n^1(0) = 0$$

et comme $\tilde{R}_n^0(0) = \sum_{k \geq n} \tilde{p}_k^0$; $n \geq 1$ on a :

$$\begin{aligned} \tilde{R}_t^0(0) &= \sum_{n \geq 1} t^n \tilde{R}_n^0(0) \\ &= \sum_{n \geq 1} t^n \sum_{k \geq n} \hat{p}_k^0 \\ &= \sum_{k \geq 1} \hat{p}_k^0 \sum_{n \geq 1}^k t^n \end{aligned}$$

soit si $|t| \neq 1$

$$\begin{aligned} \tilde{R}_t^0(0) &= \sum_{k \geq 1} \hat{p}_k^0 \frac{t}{1-t} (1 - t^k) \\ &= \frac{t}{1-t} (1 - \sum_{k \geq 1} t^k \hat{p}_k^0) \\ &= \frac{t}{1-t} (1 - \hat{P}^0(t)) \end{aligned}$$

Il est facile de voir que l'expression est encore valable pour $|t| = 1$.

. Pour $k=1$, compte tenu de ce que

$$\tilde{R}_n^1(1) = \begin{cases} 1 & \text{si } n=1 \\ \frac{m^0}{m^1} \hat{p}_{n-1}^0 & \text{si } n \geq 2 \end{cases}$$

on a

$$\begin{aligned} \tilde{R}_t^1(1) &= \sum_{n \geq 1} t^n \tilde{R}_n^1(1) \\ &= t + \sum_{n \geq 2} t^n \frac{m^0}{m^1} \hat{p}_{n-1}^0 \\ &= t(1 + \frac{m^0}{m^1} \hat{P}^0(t)) \end{aligned}$$

. De même, pour $k=1$, puisque

$$\tilde{R}_n^0(1) = \begin{cases} 0 & \text{si } n=1 \\ \frac{1}{m^0} & \text{si } n=2 \\ \sum_{j=1}^{n-1} \hat{p}_j^0 \tilde{R}_{n-j+1}^0(1) & \text{si } n \geq 3 \end{cases}$$

on a

$$\begin{aligned}
 \tilde{R}_t^0(1) &= \sum_{n \geq 1} t^n \tilde{R}_n^0(1) \\
 &= \frac{t^2}{m^0} + \sum_{n \geq 3} t^n \left[\sum_{j=1}^{n-1} \hat{p}_j^0 R_{n-j+1}^0(1) \right] \\
 &= t^2 \hat{p}_1^0 + \sum_{n \geq 3} t^n \hat{p}_1^0 R_n^0(1) + \sum_{n \geq 3} t^n \sum_{j=2}^{n-1} \hat{p}_j^0 R_{n-j+1}^0(1) \\
 &= t^2 \hat{p}_1^0 + \hat{p}_1^0 \sum_{n \geq 3} t^n R_n^0(1) + \frac{1}{t} \sum_{j \geq 2} t^j \hat{p}_j^0 \sum_{n \geq j+1} t^{n-j+1} R_{n-j+1}^0(1) \\
 &= t^2 \hat{p}_1^0 + \hat{p}_1^0 [R_t^0(1) - t^2] + \frac{1}{t} [\hat{P}^0(t) - \hat{p}_1^0 t] R_t^0(1) \\
 &= \frac{1}{t} \hat{P}^0(t) R_t^0(1)
 \end{aligned}$$

. Pour $k \geq 2$, puisque

$$\tilde{R}_n^0(k) = \begin{cases} 0 & \text{si } n \leq k \\ \sum_{j=1}^{n-k} \hat{p}_j^0 R_{n-j+1}^0(k) & \text{si } n \geq k+1 \end{cases}$$

on a

$$\begin{aligned}
 \tilde{R}_t^0(k) &= \sum_{n \geq 1} t^n \tilde{R}_n^0(k) \\
 &= \sum_{n \geq k+1} t^n \sum_{j=1}^{n-k} \hat{p}_j^0 R_{n-j+1}^0(k) \\
 &= \frac{1}{t} \sum_{j \geq 1} t^j \hat{p}_j^0 \sum_{n \geq k+j} R_{n-j+1}^0(k) t^{n-j+1} \\
 &= \frac{1}{t} \hat{P}^0(t) R_t^0(k)
 \end{aligned}$$

La dernière égalité découlant du fait que $R_n^0(k) = 0$ si $n \leq k$.

. De même, pour $k \geq 2$, puisque

$$\tilde{R}_n^1(k) = \begin{cases} 0 & \text{si } n < k \\ \sum_{j \geq k} \hat{p}_j^1 & \text{si } n = k \\ \sum_{j=1}^k \hat{p}_j^1 R_{n-j+1}^1(k-j+1) & \text{si } n \geq k+1 \end{cases}$$

on a

$$\begin{aligned}
 \tilde{R}^1(k) &= \sum_{n \geq 1} t^n \tilde{R}_n^1(k) \\
 &= t^k \tilde{R}_k^1(k) + \sum_{n \geq k+1} t^n \sum_{j=1}^k \hat{p}_j^1 R_{n-j+1}^1(k-j+1) \\
 &= t^k \sum_{j \geq k} \hat{p}_j^1 + \sum_{j=1}^k t^{j-1} \hat{p}_j^1 \sum_{n \geq k+1} t^{n-j+1} R_{n-j+1}^1(k-j+1)
 \end{aligned}$$

où

$$\begin{aligned}
 \sum_{n \geq k+1} t^{n-j+1} R_{n-j+1}^1(k-j+1) &= \sum_{v \geq k-j+2} t^v R_v^1(k-j+1) \\
 &= \begin{cases} R_t^1(1) - t & \text{si } j=k \\ R_t^1(k-j+1) & \text{si } 1 \leq j \leq k-1 \end{cases}
 \end{aligned}$$

Il vient donc

$$\begin{aligned}
 \tilde{R}_t^1(k) &= t^k \sum_{j \geq k} \hat{p}_j^1 + \sum_{j=1}^k t^{j-1} \hat{p}_j^1 R_t^1(k-j+1) + \\
 &\quad + t^{k-1} \hat{p}_k^1 [R_t^1(1) - t] \\
 &= t^k \sum_{j \geq k+1} \hat{p}_j^1 + \sum_{j=1}^k t^{j-1} \hat{p}_j^1 R_t^1(k-j+1)
 \end{aligned}$$

Il est facile de voir que, pour $k=1$, cette expression coïncide avec la valeur de $\tilde{R}_t^1(k)$ trouvée plus haut. Cela achève la démonstration du lemme. \square

IV. FONCTIONS GENERATRICES DES SUITES $\{R_t^i(u); u \in \mathbb{N}\}$ ET $\{\tilde{R}_t^i(u); u \in \mathbb{N}\}$
 $i \in \{0,1\}, |t| \leq 1.$

On considère, pour $i \in \{0,1\}$ et $|t| \leq 1$, les fonctions génératrices $x \rightarrow \delta_t^i(x)$ et $x \rightarrow \tilde{\delta}_t^i(x)$ respectives des suites $\{R_t^i(u); u \in \mathbb{N}\}$ et $\{\tilde{R}_t^i(u); u \in \mathbb{N}\}$ définies par (3) et (4) i.e.

$$(5) \quad \delta_t^i(x) = \sum_{u \geq 0} x^u R_t^i(u) ;$$

$$(6) \quad \tilde{\delta}_t^i(x) = \sum_{u \geq 0} x^u \tilde{R}_t^i(u) .$$

Notons que $(t,x) \rightarrow \delta_t^i(x)$ (resp. $\tilde{\delta}_t^i(x)$) n'est autre que la fonction génératrice de la suite à double indice $\{R_n^i(m); n \geq 1, 0 \leq m \leq n\}$ (resp. $\{\tilde{R}_n^i(m); n \geq 1, 0 \leq m \leq n\}$). Ces fonctions sont définies pour $|tx| \leq 1$: en effet, on a par exemple

$$|t^n \sum_{m=0}^n R_n^i(m) x^m| \leq |t|^n \vee |t x|^n .$$

A l'aide du lemme 3 on peut montrer :

Lemme 5.

Les fonctions génératrices $x \rightarrow \delta_t^1(x)$ et $x \rightarrow \delta_t^0(x)$ définies par (5) sont données par :

$$\delta_t^1(x) = m^0 x \tilde{P}^0(t) + x P^0(t) \delta_t^0(x)$$

$$\delta_t^0(x) = \frac{m^1}{x} \tilde{P}^1(tx) + \frac{1}{x} P^1(tx) \delta_t^1(x)$$

Preuve - Comme d'après le lemme 3 on a

$$R_t^1(u) = \begin{cases} 0 & \text{si } u=0 \\ t(1 + m^0 \tilde{P}^0(t)) & \text{si } u=1 \\ P^0(t) R_t^0(u-1) & \text{si } u \geq 2 \end{cases}$$

il vient

$$\begin{aligned}
 \delta_t^1(x) &= \sum_{u \geq 1} x^u R_t^1(u) \\
 &= xt(1 + m^0 \hat{P}^0(t)) + \sum_{u \geq 2} x^u P^0(t) R_t^0(u-1) \\
 &= xt(1 + m^0 \hat{P}^0(t)) + x P^0(t) [\delta_t^0(x) - t] \\
 &= xt(1 - P^0(t) + m^0 \hat{P}^0(t)) + x P^0(t) \delta_t^0(x)
 \end{aligned}$$

et compte tenu de ce que

$$\hat{P}^0(t) = \frac{1}{m^0} \cdot \frac{t}{1-t} (1 - P^0(t))$$

on a immédiatement l'expression de $\delta_t^1(x)$.

De même, puisque

$$R_t^0(u) = \begin{cases} t & \text{si } u=0 \\ t^2(1 + m^0 p_1^1 \hat{P}^0(t)) & \text{si } u=1 \\ m^1 t^{u+1} \hat{p}_{u+1}^1 + \sum_{j=1}^n t^j p_j^1 R_t^1(u-j+1) & \text{si } u \geq 2 \end{cases}$$

on a

$$\begin{aligned}
 \delta_t^0(x) &= \sum_{u \geq 0} x^u R_t^0(u) \\
 &= t + xt^2[1 + m^0 p_1^1 \hat{P}^0(t)] + \\
 &\quad + \sum_{u \geq 2} x^u \{ m^1 t^{u+1} \hat{p}_{u+1}^1 + \sum_{j=1}^n t^j p_j^1 R_t^1(u-j+1) \} \\
 &= t + xt^2[1 + m^0 p_1^1 \hat{P}^0(t)] + \\
 &\quad + m^1 \sum_{u \geq 2} x^u t^{u+1} \hat{p}_{u+1}^1 + \sum_{u \geq 2} x^u \sum_{j=1}^n t^j p_j^1 R_t^1(u-j+1)
 \end{aligned}$$

où

$$\sum_{u \geq 2} x^u t^{u+1} \hat{p}_{u+1}^1 = \frac{1}{x} (\hat{P}^1(xt) - xt \hat{p}_1^1 - x^2 t^2 \hat{p}_2^1)$$

et

$$\begin{aligned}
 \sum_{u \geq 2} x^u \sum_{j=1}^u t^j p_j^1 R_t^1(u-j+1) &= \sum_{u \geq 2} x^u t p_1^1 R_t^1(u) + \\
 &\quad + \sum_{j \geq 2} t^j p_j^1 \sum_{u \geq j} R_t^1(u-j+1) x^u \\
 &= t p_1^1 (\delta_t^1(x) - x R_t^1(1)) + \frac{1}{x} \sum_{j \geq 2} t^j x^j p_j^1 \sum_{u \geq j} x^{u-j+1} R_t^1(u-j+1) \\
 &= t p_1^1 (\delta_t^1(x) - x R_t^1(1)) + \frac{1}{x} (P^1(xt) - x t p_1^1) \delta_t^1(x) \\
 &= \frac{1}{x} \delta_t^1(x) P^1(xt) - x t p_1^1 R_t^1(1) .
 \end{aligned}$$

Comme on a

$$R_t^1(1) = t(1 + m^0 \hat{P}^0(t))$$

et de plus

$$\hat{P}_2^1 = \frac{1}{m} (1 - p_1^1)$$

on achève alors facilement la démonstration du lemme . \square

A partir du lemme 4 on obtient :

Lemme 6.

Les fonctions génératrices $x \rightarrow \tilde{\delta}_t^1(x)$ et $x \rightarrow \tilde{\delta}_t^0(x)$ définies par (6) sont données par :

$$\tilde{\delta}_t^1(x) = \frac{1}{xt} \hat{P}^1(xt) \delta_t^1(x) + \frac{1}{1-xt} (xt - \hat{P}^1(xt))$$

$$\tilde{\delta}_t^0(x) = \frac{1}{t} \hat{P}^0(t) \delta_t^0(x) + \frac{1}{1-t} (t - \hat{P}^0(t))$$

où $x \rightarrow \delta_t^1(x)$ et $x \rightarrow \delta_t^0(x)$ sont fournies par le lemme 5.

Preuve - Comme d'après le lemme 4 on a

$$\tilde{R}_t^0(u) = \begin{cases} \frac{t}{1-t} (1 - \hat{P}^0(t)) & \text{si } u=0 \\ \frac{1}{t} \hat{P}^0(t) R_t^0(u) & \text{si } u \geq 1 \end{cases}$$

il vient

$$\tilde{\delta}_t^0(x) = \frac{t}{1-t} (1 - \hat{P}^0(t)) + \sum_{u \geq 1} x^u \tilde{R}_t^0(u)$$

où

$$\begin{aligned} \sum_{u \geq 1} x^u \tilde{R}_t^0(u) &= \sum_{u \geq 1} x^u \frac{1}{t} \hat{P}^0(t) R_t^0(u) \\ &= \frac{1}{t} \hat{P}^0(t) \sum_{u \geq 1} x^u R_t^0(u) \\ &= \frac{1}{t} \hat{P}^0(t) \delta_t^0(x) \end{aligned}$$

On en déduit alors le résultat annoncé pour $\tilde{\delta}_t^0(x)$.

De même puisque

$$\tilde{R}_t^1(u) = t^u \sum_{j \geq u+1} \hat{p}_j^1 + \sum_{j=1}^n t^{j-1} \hat{p}_j^1 R_t^1(u-j+1) ; \quad u \geq 1$$

on a

$$\begin{aligned} \tilde{\delta}_t^1(x) &= \sum_{u \geq 1} x^u \tilde{R}_t^1(u) \\ &= \sum_{u \geq 1} x^u t^u \sum_{j \geq u+1} \hat{p}_j^1 + \sum_{u \geq 1} x^u \sum_{j=1}^n t^{j-1} \hat{p}_j^1 R_t^1(u-j+1) \end{aligned}$$

où

$$\begin{aligned} \sum_{u \geq 1} x^u t^u \sum_{j \geq u+1} \hat{p}_j^1 &= \sum_{j \geq 2} \hat{p}_j^1 \sum_{u=1}^{j-1} x^u t^u \\ &= \sum_{j \geq 2} \frac{xt}{1-xt} (1 - (xt)^{j-1}) \\ &= \frac{xt}{1-xt} \left(\sum_{j \geq 2} \hat{p}_j^1 - \sum_{j \geq 2} \hat{p}_j^1 (xt)^{j-1} \right) \\ &= \frac{xt}{1-xt} \left(1 - \hat{p}_1^1 - \frac{1}{xt} (\hat{P}^1(xt) - xt \hat{p}_1^1) \right) \\ &= \frac{xt}{1-xt} \left(1 - \frac{1}{xt} \hat{P}^1(xt) \right) \end{aligned}$$

et

$$\begin{aligned} \sum_{u \geq 1} x^u \sum_{j=1}^n t^{j-1} \hat{p}_j^1 R_t^1(u-j+1) &= \frac{1}{xt} \sum_{j \geq 1} \{ (xt)^j \hat{p}_j^1 [\sum_{u \geq j} x^{u-j+1} R_t^1(u-j+1)] \} \\ &= \frac{1}{xt} \hat{P}^1(xt) \delta_t^1(x) . \end{aligned}$$

On en déduit l'expression annoncée de $\tilde{\delta}_t^1(x)$. □

CHAPITRE 4 :

ETUDE D'UN MODELE DE SERIES CHRONOLOGIQUES VECTORIELLES
A MARGINALES CONSTRUITES PAR AMINCISSEMENTS D'UN
PROCESSUS DE RENOUVELLEMENT ALTERNE

La construction des modèles de séries chronologiques vectorielles que nous envisageons pour décrire le phénomène climatologique dans un réseau de stations repose sur les considérations suivantes :

- d'une part un modèle adapté doit être tel que les séries chronologiques marginales correspondantes représentent bien le phénomène dans les différentes stations et aient une structure voisine de celle de processus de renouvellement alterné (au moins pour les données auxquelles nous les destinons, compte tenu d'études déjà effectuées (cf. Galloy- Le Breton-Martin (1983))),

- d'autre part le modèle doit rendre bien compte de concomitances lors des évolutions du phénomène dans les différentes stations,

- enfin il est souhaitable que le nombre des paramètres du modèle soit relativement faible et si possible moindre que celui nécessaire à une modélisation dans laquelle les stations sont traitées indépendamment les unes des autres.

En vue de répondre à ces objectifs nous avons retenu comme éléments d'une modélisation :

- un processus de renouvellement alterné $(X_t^*; t \geq 1)$ destiné à représenter le phénomène dans une station test qui est soit une station "privilégiée" dans le réseau soit une station fictive (et donc non observable) et devant rendre compte des simultanités dans le réseau,

- des suites $(\varepsilon_t^j; t \geq 1)$, $j=1; \dots, q$ de variables aléatoires de Bernoulli indépendantes entre elles et indépendantes de $(X_t^*; t \geq 1)$, associées aux q différentes stations du réseau, devant permettre de traduire la spécificité des évolutions du phénomène dans ces stations "autour" de l'évolution dans la station test modélisée par $(X_t^*; t \geq 1)$.

Dans ce chapitre nous nous intéressons à un modèle de la forme

$$(1) \quad X_t^j = \varepsilon_t^j X_t^* \quad , \quad j=1; \dots, q \quad ; \quad t \geq 1 \quad ,$$

où - pour $j=1, \dots, q$, $(X_t^j; t \geq 1)$ est la série chronologique représentant la succession des jours dans la station j ,

- $(X_t^*; t \geq 1)$ est un processus RA($p^0(*)$, $p^1(*)$) ,

- pour $j=1, \dots, q$, $(\varepsilon_t^j; t \geq 1)$ est une suite de variables aléatoires indépendantes de Bernoulli de même paramètre α_j ,

- Les suites de variables aléatoires $(X_t^*; t \geq 1)$, $(\varepsilon_t^1; t \geq 1)$, \dots , $(\varepsilon_t^q; t \geq 1)$ sont indépendantes entre elles.

Remarquant que dans ce modèle, l'évolution décrite par $(X_t^*; t \geq 1)$ est un "majorant" de toutes les évolutions du phénomène dans les différentes stations au sens que la station test (éventuellement fictive) qui lui est associée est "plus sèche que toutes les autres". Lorsque $(X_t^*; t \geq 1)$ est attaché à une station "priviliégiée" du réseau, auquel cas nous supposerons qu'il s'agit de celle correspondant à l'indice $j=1$, alors bien sûr $X_t^* = X_t^1$ et $\alpha_1 = 1$.

Notons que considérant une série chronologique binaire $(X_t; t \geq 1)$ comme le processus ponctuel $\sum_{t \geq 1} X_t \delta_t$ sur \mathbb{N}^* , chaque série chronologique $(X_t^j; t \geq 1)$ marginale de la série chronologique vectorielle $(\vec{X}_t; t \geq 1)$, où $\vec{X}_t = (X_t^1, \dots, X_t^q)'$, apparait comme le processus ponctuel obtenu par amincissement du processus ponctuel $(X_t; t \geq 1)$ selon la règle consistant à supprimer ou conserver indépendamment les uns des autres chaque point de ce processus avec les probabilités respectives $1-\alpha_j$ et α_j .

Signalons aussi que l'on pourrait envisager un modèle dans lequel le processus $(X_t^*; t \geq 1)$ représenterait une évolution "minorant" toutes les évolutions dans le réseau i.e. l'évolution du phénomène dans une station test "plus humide que toutes les autres". On serait

alors amené à définir

$$X_t^j = 1 + \varepsilon_t^j (X_t^* - 1)$$

ou, de façon équivalente,

$$(1 - X_t^j) = \varepsilon_t^j (1 - X_t^*).$$

Il est bien clair que l'étude de ce modèle se ramène à celle du modèle (1).

Dans une première partie de ce chapitre nous étudions la structure du modèle défini par (1), d'abord les propriétés du processus vectoriel $(\vec{X}_t^*; t \geq 1)$ des états dans le réseau pour les jours successifs, puis celles des séries chronologiques marginales. Dans une deuxième partie nous examinons le problème statistique de l'estimation des paramètres du modèle en distinguant selon que la station test est une station dans laquelle le phénomène est observable ou non, puis nous faisons une étude expérimentale par simulation des possibilités d'utilisation du modèle.

I. ETUDE DES CARACTERISTIQUES DU MODELE .

Il est clair que, par construction même, la série chronologique vectorielle $(\vec{X}_t^*; t \geq 1)$ est stationnaire. Nous faisons porter le symbole * (resp. l'indice j ou (j,k)) à toute caractéristique du processus $(X_t^*; t \geq 1)$ (resp. $(X_t^j; t \geq 1)$ ou $((X_t^j, X_t^k); t \geq 1)$), réservant l'indice i $\in \{0,1\}$ à l'indexation d'une quantité liée "à la sécheresse (i=1) ou à l'humidité (i=0)".

Ainsi $(C^*(h); h \in \mathbb{N})$ désignera la fonction de covariance de $(X_t^*; t \geq 1)$ i.e.

$$C^*(h) = \text{Cov}(X_t^*, X_{t+h}^*) = p^* \{P[X_{h+1}^* / X_1^* = 1] - p^*\}; h \geq 0$$

où

$$p^* = P[X_t^* = 1] = \frac{m^1(*)}{m^0(*) + m^1(*)}$$

$m^i(*)$ étant l'espérance de la loi $p_n^i(*) = (p_n^i(*); n \geq 1)$; $i=0,1$ des

temps de séjour dans l'état i du processus $(X_t^*; t \geq 1)$.

De même $(C(h); h \in \mathbb{N})$ désignera la fonction de covariance de $(\vec{X}_t; t \geq 1)$ i.e. $C(h) = (C_{jk}(h); (j,k) \in \{1, \dots, q\}^2)$.

$$\begin{aligned} C_{jk}(h) &= \text{Cov}(X_t^j, X_{t+h}^k) \\ &= P[X_t^j = X_{t+h}^k = 1] - P[X_t^j = 1] P[X_t^k = 1]; h \geq 0 \end{aligned}$$

où $P[X_t^j = 1] = p(j) = \frac{m^1(j)}{m^0(j) + m^1(j)}$, $m^i(j)$ étant l'espérance de la loi $p^i(j) = (p_n^i(j); n \geq 1)$ des temps de séjour dans l'état i du processus $(X_t^j; t \geq 1)$.

Remarquons que, sauf mention du contraire, les considérations qui suivent sont valables pour un processus de base $(X_t^*; t \geq 1)$ qui n'est, pas nécessairement de renouvellement alterné, pourvu qu'il soit stationnaire.

I.1. Propriétés du processus vectoriel $(\vec{X}_t; t \geq 1)$.

Il est clair que le vecteur aléatoire \vec{X}_t est du type qui a été étudié dans le cas particulier 1 du §II.1. du chapitre I. Compte tenu des résultats obtenus dans ce chapitre on a alors :

$$(2) \quad p(j) = \alpha_j p^* ; j=1, \dots, q$$

$$(3) \quad P[\vec{X}_t = x] = \begin{cases} (1-p^*) + p^* \prod_{j=1}^q (1-\alpha_j) & \text{si } x = \phi \\ p^* \prod_{j=1}^q \alpha_j^{x_j} (1-\alpha_j)^{1-x_j} & \text{si } x \neq \phi \end{cases}$$

où rappelons que $x = (x_1, \dots, x_q)' \in \{0, 1\}^q$ est identifié à $\{j \in \{1, \dots, q\} : x_j = 1\}$.

De même si N_t désigne le nombre de stations "sèches" dans le réseau

le jour t i.e. $N_t = \sum_{j=1}^q X_t^j$ on a

$$(4) \quad P [N_t = k] = \begin{cases} (1-p^*) + p^* \prod_{j=1}^q (1-\alpha_j) & \text{si } k = 0 \\ p^* \sum_{x: |x|=k} \prod_{j=1}^q \alpha_j^{x_j} (1-\alpha_j)^{1-x_j} & \text{si } 1 \leq k \leq q, \end{cases}$$

$$E(N_t) = p^* \sum_{j=1}^q \alpha_j ;$$

$$\text{Var}(N_t) = p^* \sum_{j=1}^q \alpha_j (1-\alpha_j p^*) + 2p^*(1-p^*) \sum_{i < j} \alpha_i \alpha_j ,$$

et

$$(5) \quad P[\vec{X}_t = x / N_t = k] = \begin{cases} \frac{\prod_{j=1}^q \alpha_j^{x_j} (1-\alpha_j)^{1-x_j}}{\sum_{z: |z|=k} \prod_{\ell=1}^q \alpha_{\ell}^{z_{\ell}} (1-\alpha_{\ell})^{1-z_{\ell}}} & \text{si } 1 \leq k \leq q, |x|=k \\ 1 & \text{si } k=0 \text{ et } |x|=0 \end{cases}$$

En ce qui concerne la structure d'ordre 2 du processus $(\vec{X}_t; t \geq 1)$ on obtient

Proposition 1.

La fonction de covariance $(C(h); h \geq 0)$ s'écrit :

$$(6) \quad C_{jk}(h) = \begin{cases} \alpha_j \alpha_k C^*(h) & \text{si } j \neq k, h=0, 1, \dots \text{ ou } j=k \text{ et } h=1, 2, \dots \\ \alpha_j p^* (1-\alpha_j p^*) & \text{si } j=k, h=0. \end{cases}$$

La loi conjointe des vecteurs aléatoires \vec{X}_{t+h} et \vec{X}_t , $h \geq 1$, est donnée par :

$$(7) \quad P[\vec{X}_{t+h} = y, \vec{X}_t = x] = \begin{cases} \left[\prod_{j=1}^q \alpha_j^{y_j + x_j} (1-\alpha_j)^{2-y_j-x_j} \right] P[X_{h+1}^* = 1, X_1^* = 1] & \text{si } |y| \neq 0 \text{ et } |x| \neq 0 \\ \left[\prod_{j=1}^q \alpha_j^{y_j} (1-\alpha_j)^{1-y_j} \right] \{ P[X_{h+1}^* = 1, X_1^* = 0] + \left[\prod_{j=1}^q (1-\alpha_j) \right] P[X_{h+1}^* = 1, X_1^* = 1] \} & \text{si } |y| \neq 0 \text{ et } |x| = 0 \\ \sum_{u, v=0}^1 \left(\prod_{j=1}^q (1-\alpha_j) \right)^{u+v} P[X_{h+1}^* = u, X_1^* = v] & \text{si } |y| = |x| = 0 \end{cases}$$

Preuve - L'égalité (6) est une conséquence immédiate de la définition (1).

On a bien sûr, pour x et y dans $\{0,1\}^q$:

$$P[\vec{X}_{t+h} = y, \vec{X}_t = x] = \begin{cases} P[\vec{X}_{t+h} = y, X_{t+h}^* = 1, \vec{X}_t = x] & \text{si } |y| \neq 0 \\ P[X_{t+h}^* = 0, \vec{X}_t = x] + \prod_{j=1}^q (1-\alpha_j) P[X_{t+h}^* = 1, \vec{X}_t = x] & \text{si } |y| = 0 \end{cases}$$

Or si $|y| \neq 0$ on a

$$P[\vec{X}_{t+h} = y, X_{t+h}^* = 1, \vec{X}_t = x] = \begin{cases} \prod_{j=1}^q \alpha_j^{x_j} (1-\alpha_j)^{1-x_j} \prod_{j=1}^q \alpha_j^{y_j} (1-\alpha_j)^{1-y_j} P[X_{t+h}^* = 1, X_t^* = 1] & \text{si } |x| \neq 0 \\ \prod_{j=1}^q \alpha_j^{y_j} (1-\alpha_j)^{1-y_j} \{P[X_{t+h}^* = 1, X_t^* = 0] + \prod_{j=1}^q (1-\alpha_j) P[X_{t+h}^* = 1, X_t^* = 1]\} & \text{si } |x| = 0 \end{cases}$$

De même on a pour $i \in \{0,1\}$:

$$P[X_{t+h}^* = i, \vec{X}_t = x] = \begin{cases} \prod_{j=1}^q \alpha_j^{x_j} (1-\alpha_j)^{1-x_j} P[X_{t+h}^* = i, X_t^* = 1] & \text{si } |x| \neq 0 \\ P[X_{t+h}^* = i, X_t^* = 0] + \prod_{j=1}^q (1-\alpha_j) P[X_{t+h}^* = i, X_t^* = 1] & \text{si } |x| = 0 \end{cases}$$

On en déduit alors l'égalité (7). \square

Remarque 1

Notons que la probabilité $P[\vec{X}_{t+h} = 0, \vec{X}_t = x]$ pour $|x| \neq 0$ est fournie par changement de $h+1, 1$ et y en $1, h+1$ et x respectivement dans la deuxième ligne de (7).

On peut bien sûr déduire de la proposition 1 la loi conditionnelle de \vec{X}_{t+h} sachant \vec{X}_t de même que des résultats concernant le processus $(N_t; t \geq 1)$.

Pour simplifier les écritures nous utiliserons dans la suite la notation $v(\vec{\alpha}, k)$ pour la quantité

$$(8) \quad P\left[\sum_{j=1}^q \epsilon_j^t = k\right] = \sum_{y: |y|=k} \prod_{j=1}^q \alpha_j^{y_j} (1-\alpha_j)^{1-y_j} \quad 0 \leq k \leq q$$

Corollaire 1.

La fonction de covariance du processus $(N_t; t \geq 1)$ s'écrit

$$(9) \quad \text{Cov}(N_{t+h}, N_t) = \left[\sum_{j=1}^q \alpha_j \right]^2 C^*(h) \quad , \quad h \geq 1$$

La loi conditionnelle de la variable aléatoire N_{t+h} sachant N_t est donnée pour $h \geq 1$ par :

$$(10) \quad P[N_{t+h} = \ell / N_t = k] = \begin{cases} v(\vec{\alpha}, \ell) P[X_{t+h}^* = 1 / X_t^* = 1] & \text{si } k \neq 0 \text{ et } \ell \neq 0 \\ P[X_{h+1}^* = 0 / X_1^* = 1] + v(\vec{\alpha}, 0) P[X_{h+1}^* = 1 / X_1^* = 1] & \text{si } k \neq 0 \text{ et } \ell = 0 \\ \frac{v(\vec{\alpha}, \ell) \{P[X_{h+1}^* = 1, X_1^* = 0] + v(\vec{\alpha}, 0) P[X_{h+1}^* = X_1^* = 1]\}}{(1-p^*) + p^* v(\vec{\alpha}, 0)} & \\ 1 & \text{si } k=0 \text{ et } \ell \neq 0 \\ \frac{\sum_{u,v=0}^1 v(\vec{\alpha}, 0)^{u+v} P[X_{h+1}^* = u, X_1^* = v]}{(1-p^*) + p^* v(\vec{\alpha}, 0)} & \text{si } k=\ell=0 \end{cases}$$

où $v(\vec{\alpha}, k)$ est fournie par (8).

Preuve - Pour $h \geq 1$, on a

$$\begin{aligned} \text{Cov}(N_{t+h}, N_t) &= \text{Cov}\left(\sum_{j=1}^q X_{t+h}^j, \sum_{k=1}^q X_t^k\right) \\ &= \sum_{j,k} C_{jk}(h) \end{aligned}$$

d'où la formule (9) d'après l'égalité (6)

On a bien sûr, pour $h \geq 1$,

$$P[N_{t+h} = \ell, N_t = k] = \sum_{y: |y|=\ell} \sum_{x: |x|=k} P[\vec{X}_{t+h} = y, \vec{X}_t = x]$$

Alors d'après l'égalité (7), si $k \neq 0$ et $\ell \neq 0$, on a :

$$P[N_{t+h}=\ell, N_t=k] = v(\vec{\alpha}, \ell) v(\vec{\alpha}, k) P[X_{h+1}^* = X_1^* = 1]$$

puis, utilisant (3),

$$P[N_{t+h}=\ell / N_t=k] = v(\vec{\alpha}, \ell) \cdot \frac{P[X_{h+1}^* = X_1^* = 1]}{p^*}$$

d'où la première ligne de (10) compte tenu de ce que $p^* = P[X_t^* = 1]$.
De même si $k \neq 0$ et $\ell = 0$ on a, d'après la remarque 1 et (7)

$$P[N_{t+h}=0, N_t=k] = v(\vec{\alpha}, k) \{P[X_{h+1}^* = 0, X_1^* = 1] + v(\vec{\alpha}, 0) P[X_{h+1}^* = 1, X_1^* = 1]\}$$

puis

$$P[N_{t+h}=0 / N_t=k] = \frac{P[X_{h+1}^* = 0, X_1^* = 1] + v(\vec{\alpha}, 0) P[X_{h+1}^* = 1, X_1^* = 1]}{p^*}$$

d'où la deuxième ligne de (10).

Si $k=0$ et $\ell \neq 0$, on a aussi, d'après (7)

$$P[N_{t+h}=\ell, N_t=0] = v(\vec{\alpha}, \ell) P[X_{h+1}^* = 1, X_1^* = 0] + v(\vec{\alpha}, 0) P[X_{h+1}^* = 1, X_1^* = 1]$$

d'où la troisième ligne de (10) d'après (3).

Enfin si $k=0$ et $\ell=0$ on a bien sûr

$$P[N_{t+h}=0, N_t=0] = P[\vec{X}_{t+h} = \phi, \vec{X}_t = \phi]$$

d'où la quatrième ligne de (10) d'après (7) et (3). \square

Remarque 2 - Dans le cas particulier où $h=1$ les probabilités de transition $P[N_{t+1}=\ell / N_t=k]$ se calculent facilement compte tenu de ce que (cf. chapitre II. §I.2) on a pour $i \in \{0, 1\}$:

$$P[X_2^* = i, X_1^* = 1-i] = \frac{1}{m^0(*) + m^1(*)}$$

$$P[X_2^* = i, X_1^* = i] = \frac{m^1(*) - 1}{m^0(*) + m^1(*)}$$

et aussi

$$P[X_2^* = i / X_1^* = i] = 1 - \frac{1}{m^i(*)}$$

$$P[X_2^* = 1 / X_1^* = 1-i] = \frac{1}{m^{1-i}(*)}$$

Lorsque $(X_t^*) \sim RA(p^0(*), p^1(*))$, on dispose (cf. Bousseboua (1983)) d'un algorithme de calcul des probabilités conditionnelles

$P[X_{t+h}^* = u / X_t^* = v]$; $h \geq 1$, s'inspirant des mêmes principes que celui décrit dans le §I. du chapitre III. En fait la famille $(P[X_h^* = 1 / X_1^* = 1])$; $h \geq 1$ admet dans ce cas (cf. Grégoire (1982)) comme fonction génératrice

$$H(s) = \frac{1}{1-s} - \frac{s}{m^1(*) (1-s)^2} \cdot \frac{(1 - P^0_*(s))(1 - P^1_*(s))}{1 - P^0_*(s) P^1_*(s)}$$

où pour $i \in \{0, 1\}$, P^i_* est la fonction génératrice de la loi $p^i(*)$

I.2. Propriétés des séries chronologiques marginales

$$(X_t^j; t \geq 1), \quad j=1, \dots, q.$$

Dans ce paragraphe, pour simplifier les écritures nous omettrons l'indice j dans toutes les caractéristiques du processus

$$(X_t^j; t \geq 1); \text{ ainsi à } p(j) = P[X_t^j = 1] =$$

$$\frac{m^1(j)}{m^0(j) + m^1(j)} = \alpha_j p^* \quad \text{nous substituerons } p = \alpha p^* = \frac{m^1}{m^0 + m^1} \quad \text{où}$$

m^1 est substitué à $m^1(j)$ comme $p^1 = (p_n^1; n \geq 1)$ sera substitué à $p^1(j) = (p_n^1(j); n \geq 1)$.

Les résultats concernant donc un processus $(X_t; t \geq 1)$ défini par

$$X_t = \varepsilon_t X_t^* \quad ; \quad t \geq 1$$

où

- $(X_t^*; t \geq 1)$ est une série chronologique binaire stationnaire

- $(\varepsilon_t; t \geq 1)$ est une suite de variables aléatoires indépendantes de Bernoulli de même paramètre α

- les suites $(X_t^*; t \geq 1)$ et $(\varepsilon_t; t \geq 1)$ sont indépendantes entre elles.

Le processus $(X_t; t \geq 1)$ est bien sûr une série chronologique binaire stationnaire. Il est clair que pour tout n le vecteur aléatoire $X_{(n)} = (X_1, \dots, X_n)'$ est du type qui a été étudié dans le cas particulier 2 du §II.1 du chapitre I. Compte tenu des résultats obtenus dans ce chapitre on a alors :

$$(11) \quad P[X_{(n)}=x] = \alpha^{|x|} \sum_{z \supset x} (1-\alpha)^{|z-x|} P[X_{(n)}=z] \quad ; \quad x \in \{0,1\}^n$$

$$(12) \quad C(h) = \text{Cov}(X_{t+h}, X_t) = \begin{cases} \alpha p^* (1-\alpha p^*) & \text{si } h=0 \\ \alpha^2 C^*(h) & \text{si } h \geq 1 \end{cases}$$

De même si S_n désigne la variable aléatoire comptant le nombre de jours secs au cours de n jours consécutifs $1, \dots, n$ i.e.

$$S_n = \sum_{t=1}^n X_t, \text{ on a}$$

$$(13) \quad P[S_n=k] = \alpha^k \sum_{s=k}^n C_s^k (1-\alpha)^{s-k} P[S_n^*=s] \quad ; \quad 0 \leq k \leq n,$$

où S_n^* est la variable aléatoire $S_n^* = \sum_{t=1}^n X_t^*$ dont la loi est

fournie par l'algorithme décrit dans le §I du chapitre III lorsque $(X_t^*; t \geq 1) \sim \text{RA}(p^0(*), p^1(*))$.

Enfin on a :

$$(14) \quad E(S_n) = \alpha E(S_n^*) \quad ; \quad \text{Var}(S_n) = \alpha^2 \text{Var}(S_n^*) + \alpha(1-\alpha)E(S_n^*)$$

où compte tenu de la stationnarité de $(X_t^*, t \geq 1)$,

$$E(S_n^*) = np^* = n \frac{m^1(*)}{m^0(*) + m^1(*)} \quad ;$$

$$\text{Var}(S_n^*) = nC^*(0) + 2 \sum_{h=1}^{n-1} (n-h) C^*(h) .$$

En fait on a aussi

$$(14') \quad \begin{aligned} E(S_n) &= np = n \frac{m^1}{m^0 + m^1} ; \\ \text{Var}(S_n) &= n C(0) + 2 \sum_{h=1}^{n-1} (n-h) C(h) \end{aligned}$$

où la suite $(C(h); h \geq 1)$ est donnée par (12).

Remarquons que si $(X_t^*; t \geq 1) \sim \text{RA}(p^0(*), p^1(*))$ le processus $(X_t; t \geq 1)$ n'est pas lui même en général un processus de renouvellement alterné; de même lorsque $(X_t^*; t \geq 1) \sim \text{MA}(1; p^*, \lambda^*)$, $(X_t; t \geq 1)$ n'est pas Markovien d'ordre 1 ni Markovien d'ordre 2. En effet, soit $(X_t^*; t \geq 1) \sim \text{MA}(1; p^*, \lambda^*)$; on a pour $t \geq 4$

$$P[X_t=1/X_{t-1}=0, X_{t-2}=1, X_{t-3}=0] = \frac{P[X_t=1, X_{t-1}=0, X_{t-2}=1, X_{t-3}=0]}{P[X_{t-1}=0, X_{t-2}=1, X_{t-3}=0]}$$

où compte tenu de (11), on a

$$P[X_t=1, X_{t-1}=0, X_{t-2}=1, X_{t-3}=0] = \alpha^2 \sum_{z \in \{1, 3\}} (1-\alpha)^{|z|-2} P[X_{(4)}^* = z] ,$$

$$P[X_{t-1}=0, X_{t-2}=1, X_{t-3}=0] = \alpha \sum_{z \in \{2\}} (1-\alpha)^{|z|-1} P[X_{(3)}^* = z]$$

d'où

$$P[X_t=1/X_{t-1}=0, X_{t-2}=1, X_{t-3}=0] = \frac{(1-\alpha\lambda^*)}{1-p^*} \cdot \frac{p^*-2\lambda^*p^*+(\lambda^*)^2(1-\alpha+\alpha p^*)}{1-\alpha+\alpha^2(1-\lambda^*)^2}$$

De même, on a pour tout $t \geq 3$

$$P[X_t=1/X_{t-1}=0, X_{t-2}=1] = \frac{\alpha}{1-p^*} \cdot \frac{p^*(1-2\lambda^*+\alpha(\lambda^*)^2)+(1-\alpha)(\lambda^*)^2}{1-\alpha\lambda^*}$$

Or ces deux dernières expressions ne sont pas égales .

Nous nous intéressons maintenant aux caractéristiques des lois $p^i = (p_n^i; n \geq 1)$ des temps de séjour dans les deux états 0 et 1. Rappelons (cf. (12) chapitre II §I.) que la fonction génératrice P^i de la loi p^i s'exprime par

$$(15) \quad P^i(t) = (m^0 + m^1) \frac{(1-t)^2}{t^2} Q^i(t) - m^i \frac{1-t}{t} + 1$$

en fonction de la fonction génératrice Q^i de la suite $(Q_n^i; n \geq 1)$ où

$$Q_n^i = P[X_1=i, \dots, X_n=i]$$

$$\text{i.e.} \quad Q_n^1 = P[S_n = n] \quad ; \quad Q_n^0 = P[S_n = 0] \quad .$$

On détermine donc d'abord les fonctions génératrices Q^0 et Q^1 en fonction des caractéristiques du processus $(X_t^*; t \geq 1)$:

Lemme 1.

On a

$$(16) \quad Q^1(t) = Q_*^1(\alpha t)$$

où Q_*^1 est la fonction génératrice de la suite $(Q_n^1(*); n \geq 1)$ où $Q_n^1(*) = P[S_n^* = n]$.

Preuve - Constatant qu'en vertu de (13) on a

$$Q_n^1 = \alpha^n Q_n^1(*)$$

on obtient

$$\begin{aligned} Q^1(t) &= \sum_{n \geq 1} Q_n^1 t^n \\ &= \sum_{n \geq 1} Q_n^1(*) (\alpha t)^n = Q_*^1(\alpha t) \quad . \quad \square \end{aligned}$$

Lemme 2.

On a

$$(17) \quad Q^0(t) = p^* \tilde{\delta}_*^1(t, 1-\alpha) + (1-p^*) \tilde{\delta}_*^0(t, 1-\alpha) \quad ; \quad t \geq 1$$

où pour $i \in \{0, 1\}$, $(t, x) \rightarrow \tilde{\delta}_*^i(t, x)$ est la fonction génératrice de la suite à double indice $\{ P[S_n^* = s / X_1^* = i] ; n \geq 1, 0 \leq s \leq n \}$.

Preuve - D'après l'égalité (13) on a

$$Q_n^0 = \sum_{s=0}^n (1-\alpha)^s P[S_n^* = s]$$

d'où

$$Q^0(t) = \sum_{n \geq 1} \left\{ \sum_{s=0}^n (1-\alpha)^s P[S_n^* = s] \right\} t^n$$

$$\text{où } P[S_n^* = s] = p^* P[S_n^* = s / X_1^* = 1] + (1-p^*) P[S_n^* = s / X_1^* = 0] .$$

Alors il vient

$$Q^0(t) = p^* \sum_{n \geq 1} \left\{ \sum_{s=0}^n (1-\alpha)^s P[S_n^* = s / X_1^* = 1] \right\} t^n \\ + (1-p^*) \sum_{n \geq 1} \left\{ \sum_{s=0}^n (1-\alpha)^s P[S_n^* = s / X_1^* = 0] \right\} t^n$$

d'où la formule (17) . \square

On calcule maintenant les moments m^1 des lois p^1 en fonction des moments $m^i(*)$ des lois $p^i(*)$:

Lemme 3

On a

$$(18) \quad m^1 = \frac{m^1(*)}{m^1(*) (1-\alpha) + \alpha} \quad ; \quad m^0 = \frac{m^1(*) (1-\alpha) + m^0(*)}{m^1(*) \alpha (1-\alpha) + \alpha^2}$$

Preuve - D'après la démonstration du lemme 3, chapitre II §I.2, on a

$$(19) \quad m^i = \frac{Q_1^i}{Q_1^i - Q_2^i} \quad ; \quad i \in \{0, 1\} .$$

Alors il vient (cf. la preuve du lemme 1 ci-dessus) :

$$m^1 = \frac{\alpha Q_1^1(*)}{\alpha Q_1^1(*) - \alpha^2 Q_2^1(*)}$$

où (cf. Remarque 2 ci-dessus)

$$Q_1^1(*) = \frac{m^1(*)}{m^0(*) + m^1(*)} \quad \text{et} \quad Q_2^1(*) = \frac{m^1(*) - 1}{m^0(*) + m^1(*)}$$

On a donc

$$m^1 = \frac{m^1(*)}{m^1(*) - \alpha(m^1(*) - 1)} = \frac{m^1(*)}{m^1(*) (1 - \alpha) + \alpha}$$

D'après (19) on a

$$\begin{aligned} m^0 + m^1 &= \frac{Q_1^0 + Q_1^1}{Q_1^1 - Q_2^1} = \frac{1}{Q_1^1 - Q_2^1} \\ &= \frac{1}{\alpha Q_1^1(*) - \alpha^2 Q_2^1(*)} \end{aligned}$$

D'où, d'après le calcul de m^1 , on obtient

$$m^0 = \frac{m^0(*) + m^1(*) - \alpha m^1(*)}{\alpha(1 - \alpha)m^1(*) + \alpha^2} = \frac{m^1(*) (1 - \alpha) + m^0(*)}{m^1(*) \alpha(1 - \alpha) + \alpha^2} \quad \square$$

On est alors en mesure de déterminer les fonctions génératrices

\hat{P}^0 et \hat{P}^1 .

On note $\hat{P}^i(*)$ la fonction génératrice de la loi $\hat{p}^i(*) = (\hat{p}_n^i(*); n \geq 1)$

Proposition 2.

(i) La fonction génératrice \hat{P}^1 est donnée par l'égalité (15) en y remplaçant $Q^1(t)$ par

$$Q^1(t) = p^* \frac{\alpha t}{1 - \alpha t} (1 - \hat{P}_*(\alpha t))^1$$

et m^0, m^1 par les valeurs des égalités (18) .

(ii) Lorsque $(X_t^*; t \geq 1) \sim RA(p^0(*), p^1(*))$ la fonction génératrice P^0 est donnée par l'égalité (15) en y remplaçant m^0, m^1 par les valeurs des égalités (18) et $Q^0(t)$ par

$$Q^0(t) = \frac{\delta_*^0(t, 1-\alpha)}{(m^0(*) + m^1(*)t)} \{ m^0(*) \hat{P}_*(t)^0 + m^1(*) P_*(t)^0 \hat{P}_*((1-\alpha)t)^1 \} \\ + \frac{1}{m^0(*) + m^1(*)} \{ m^0(*) m^1(*) \hat{P}_*(t)^0 \hat{P}_*((1-\alpha)t)^1 + \frac{m^1(*)}{1 - (1-\alpha)t} \times \\ \times ((1-\alpha)t - \hat{P}_*((1-\alpha)t)^1) + \frac{m^0(*)}{1-t} (t - \hat{P}_*(t)^0) \}$$

où

$$(20) \quad \delta_*^0(t, x) = \frac{1}{1 - P_*(t)^0 P_*(xt)^1} \{ m^0(*) \hat{P}_*(t)^0 P_*(xt)^1 + \frac{m^1(*)}{x} \hat{P}_*(xt)^1 \}$$

Preuve - La première assertion (i) est une conséquence immédiate des lemmes 1 et 3 et de l'égalité (11) du chapitre II §I.

Démontrons maintenant la deuxième assertion (ii). D'après les lemmes 5 et 6 du chapitre III, sous l'hypothèse faite concernant $(X_t^*; t \geq 1)$ les fonctions génératrices $\tilde{\delta}_*^1$ intervenant dans l'égalité (17) sont fournies par

$$\tilde{\delta}_*^1(t, x) = \frac{1}{xt} \hat{P}_*(xt)^1 \delta_*^1(t, x) + \frac{1}{1-xt} (xt - \hat{P}_*(xt)^1)$$

$$\tilde{\delta}_*^0(t, x) = \frac{1}{t} \hat{P}_*(t)^0 \delta_*^0(t, x) + \frac{1}{1-t} (t - \hat{P}_*(t)^0)$$

où

$$\delta_*^1(t, x) = m^0(*) x \hat{P}_*(t)^0 + x P_*(t)^0 \delta_*^0(t, x)$$

$$(21) \quad \delta_*^0(t, x) = \frac{m^1(*)}{x} \hat{P}_*(xt)^1 + \frac{1}{x} P_*(xt)^1 \delta_*^1(t, x) .$$

Reportant ces expressions dans (17) on obtient, pour $x = 1-\alpha$

$$\begin{aligned}
 Q^0(t) &= p^* \frac{1}{xt} \hat{P}^1(xt) \{m^0(*) x \hat{P}^0(t) + x P^0(t) \delta_*^0(t,x)\} + \\
 &+ p^* \frac{1}{1-xt} (xt - \hat{P}^1(xt)) + (1-p^*) \frac{1}{t} \hat{P}^0(t) \delta_*^0(t,x) + \\
 &\quad (1-p^*) \frac{1}{1-t} (t - \hat{P}^0(t)) \\
 &= \delta_*^0(t,x) \{p^* \frac{1}{t} P^0(t) \hat{P}^1(xt) + (1-p^*) \frac{1}{t} \hat{P}^0(t)\} + \\
 &+ p^* m^0(*) \frac{1}{t} \hat{P}^0(t) \hat{P}^1(xt) + p^* \frac{1}{1-xt} (xt - \hat{P}^1(xt)) + \\
 &\quad (1-p^*) \frac{1}{1-t} (t - \hat{P}^0(t)) \\
 &= \frac{\delta_*^0(t,x)}{(m^0(*) + m^1(*)t)} \{m^0(*) \hat{P}^0(t) + m^1(*) P^0(t) \hat{P}^1(xt)\} + \\
 &+ \frac{1}{m^0(*) + m^1(*)} \{ m^0(*) m^1(*) \frac{1}{t} \hat{P}^0(t) \hat{P}^1(xt) + \frac{m^1(*)}{1-xt} (xt - \hat{P}^1(xt)) \\
 &\quad + \frac{m^0(*)}{1-t} (t - \hat{P}^0(t)) \}
 \end{aligned}$$

où d'après (21), on a

$$\begin{aligned}
 \frac{1}{x} P^1(xt) \delta_*^1(t,x) &= m^0(*) \hat{P}^0(t) P^1(xt) + P^0(t) P^1(xt) \delta_*^0(t,x) \\
 \delta_*^0(t,x) &= \frac{m^1(*)}{x} \hat{P}^1(xt) + \frac{1}{x} P^1(xt) \delta_*^1(t,x)
 \end{aligned}$$

puis

$$\delta_*^0(t,x) [1 - P^0(t) P^1(xt)] = m^0(*) \hat{P}^0(t) P^1(xt) + \frac{m^1(*)}{x} \hat{P}^1(xt) .$$

On en déduit alors l'expression annoncée de $Q^0(t)$. \square

Il est en fait possible de donner de manière explicite la loi p^1 :

Proposition 3.

La loi des temps de séjour dans l'état 1 est donnée par

$$(22) \quad p_n^1 = \frac{\alpha^{n-1}}{m^1(*) (1-\alpha) + \alpha} \rho_n \quad ; \quad n \geq 1$$

où $\rho_n = \gamma_n - 2\alpha\gamma_{n+1} + \alpha^2 \gamma_{n+2}$

avec $\gamma_n = \sum_{j \geq 1} j p_{n+j-1}^1(*)$.

Preuve - D'après l'égalité (4) du chapitre II §I., on a

$$(23) \quad p_n^1 = \frac{Q_n^1 - 2Q_{n+1}^1 + Q_{n+2}^1}{Q_1^1 - Q_2^1}$$

D'après l'égalité (13) ci dessus on a

$$Q_n^1 = \alpha^n Q_n^1(*)$$

où

$$\begin{aligned} Q_n^1(*) &= P[S_n^* = n] = P[S_n^* = n / X_1^* = 1] P[X_1^* = 1] \\ &= p^* \sum_{k \geq n} \hat{p}_k^1(*) = p^* \frac{1}{m^1(*)} \sum_{k \geq n} \sum_{l \geq k} p_l^1(*) \\ &= p^* \frac{1}{m^1(*)} \sum_{j \geq 1} j p_{n+j-1}^1(*) . \end{aligned}$$

Reportant les valeurs des Q_n^1 dans (23) il est alors facile de conclure . \square

Remarque 3.

Dans le cas particulier où $p^1(*)$ est une loi géométrique $G(\mu)$ alors p^1 est la loi géométrique $G(1-\alpha(1-\mu))$. En effet la quantité ρ_n de la proposition précédente vaut alors :

$$\rho_n = \frac{(1-\mu)^{n-1}}{\mu} [1 - \alpha(1-\mu)]^2 \quad ; \quad n \geq 1$$

et donc

$$p_n^1 = [\alpha(1-\mu)]^{n-1} [1 - \alpha(1-\mu)] \quad ; \quad n \geq 1$$

D'ailleurs d'après la proposition 2 et (15) on a :

$$P^1(t) = p^*(m^0 + m^1) \frac{(1-t)^2}{t(1-\alpha t)} \alpha(1 - \hat{P}^1(\alpha t)) - m^1 \frac{1-t}{t} + 1$$

où m^0 et m^1 sont donnés par (18) i.e.

$$m^0 = \frac{m^1(*) (1-\alpha) + m^0(*)}{m^1(*) \alpha(1-\alpha) + \alpha^2} \quad ; \quad m^1 = \frac{m^1(*)}{m^1(*) (1-\alpha) + \alpha}$$

Alors, tenant compte de ce que $p^1(*)$ est une loi géométrique, il vient

$$P^1(t) = \frac{1 - \alpha(1-\mu)}{1 - t\alpha(1-\mu)} t$$

Ainsi en particulier si $(X_t^*; t \geq 1) \sim MA(1; p^*, \lambda^*)$ alors p^1 est la loi $G(1-\alpha\lambda^*)$.

En ce qui concerne les moments d'ordre 2 des lois p^1 on déduit de la proposition 2

Corollaire 2.

On a :

$$m_{[2]}^1 = \frac{2 m^1(*)}{m^1(*) (1-\alpha)^2 + \alpha(1-\alpha)} (\alpha - \hat{P}^1(\alpha))$$

et, si $(X_t^*; t \geq 1) \sim RA(p^0(*), p^1(*))$:

$$m_{[2]}^0 = \frac{2}{m^1(*) \alpha(1-\alpha) + \alpha^2} \{ \delta_*^0(1, 1-\alpha) [m^0(*) + m^1(*) \hat{P}^1(1-\alpha)] + \\ + m^0(*) m^1(*) \hat{P}^1(1-\alpha) + \frac{m^1(*)}{\alpha} [(1-\alpha)^2 - \hat{P}^1(1-\alpha)] + \\ + \frac{1}{2} m_{[2]}^0(*) - m^0(*) \}$$

$$\text{où } \delta_*^0(1, 1-\alpha) = \frac{1}{1 - P_*(1-\alpha)} \{ m^0(*) P_*^1(1-\alpha) + \frac{m^1(*)}{1-\alpha} \hat{P}_*^1(1-\alpha) \} .$$

Preuve - D'après le lemme 5 du chapitre II §I.2 on a

$$m_{[2]}^1 = 2(m^0 + m^1) Q^1(1) - 2m^1 .$$

Il suffit alors de remplacer les valeurs des différents termes par celles obtenues précédemment : m^1 , m^0 données par (18) et $Q^1(1)$, $Q^0(1)$ données par la proposition 2, la valeur de

$$\frac{m^0(*)}{1-t} (t - \hat{P}_*^0(t)) \text{ en } t=1 \text{ étant } \frac{1}{2} m_{[2]}^0(*) . \quad \square$$

On dispose des approximations suivantes des moments d'ordre 1 et 2 des lois p^i , $i \in \{0, 1\}$:

Corollaire 3.

On a, pour α voisin de 1, les approximations :

$$m^1 \approx m^1(*) + (\alpha-1)m^1(*) (m^1(*) - 1)$$

$$m^0 \approx m^0(*) + (\alpha-1)[m^1(*) (m^0(*) - 1) - 2m^0(*)]$$

$$m_{[2]}^1 \approx m_{[2]}^1(*) + (\alpha-1)[m^1(*) m_{[2]}^1(*) + \frac{1}{3} m_{[3]}^1(*)]$$

et, si de plus $(X_t^*; t \geq 1) \sim RA(p^0(*), p^1(*))$,

$$m_{[2]}^0 \approx m_{[2]}^0(*) + (\alpha-1)[(m^1(*) - 2)m_{[2]}^0(*) - 2m^0(*) (2 + m^0(*) p_1^1(*))]$$

Preuve - Les deux premières approximations se déduisent aisément des expressions (18) et de ce que

$$\frac{1}{m^1(*) (1-\alpha) + \alpha} \approx 1 + (\alpha-1) (m^1(*) - 1) ;$$

$$\frac{1}{m^1(*) \alpha (1-\alpha) + \alpha^2} \approx 1 + (\alpha-1) (m^1(*) - 2)$$

D'après le corollaire 2, on a :

$$m_{[2]}^1 = \frac{2 m^1(*)}{m^1(*) (1-\alpha) + \alpha} \left\{ \frac{1 - \hat{P}^*(\alpha)}{1 - \alpha} - 1 \right\}$$

où, d'après les lemmes 4 et 5 du chapitre II :

$$\begin{aligned} \hat{P}^*(\alpha) \approx & 1 + (\alpha-1) \frac{1}{m^1(*)} [m^1(*) + \frac{1}{2} m_{[2]}^1(*)] + \\ & + (\alpha-1) \frac{2}{m^1(*)} \left[\frac{m_{[2]}^1(*)}{2} + \frac{m_{[3]}^1(*)}{6} \right] . \end{aligned}$$

Alors compte tenu de ce qui précède on obtient la troisième approximation.

On déduit facilement la dernière approximation, du résultat du corollaire 2, de l'approximation ci-dessus de $\frac{1}{m^1(*) \alpha (1-\alpha) + \alpha^2}$ et des approximations suivantes :

$$\delta_*^0(1, 1-\alpha) \approx 1 + (1-\alpha) m^0(*) p_1^1(*) + 1$$

$$m^1(*) \hat{P}^*(1-\alpha) \approx (1-\alpha)$$

$$m^1(*) \frac{(1-\alpha)^2}{\alpha} \approx m^1(*) (1-\alpha)^2 . \quad \square$$

Exemples numériques.

(i) Placons-nous dans le cas où $(X_t^*; t \geq 1) \sim MA(1; p^*, \lambda^*)$; alors on a

$$m^1(*) = \frac{1}{1 - \lambda^*}$$

$$m_{[2]}^1(*) = 2 \frac{\lambda^*}{(1-\lambda^*)^2}$$

$$m^0(*) = \frac{1 - p^*}{p^* (1-\lambda^*)}$$

$$m_{[2]}^0(*) = 2 \frac{(1-p^*) (1-2p^* + p^* \lambda^*)}{(p^*)^2 (1-\lambda^*)^2}$$

On donne dans le tableau suivant quelques valeurs exactes et approchées des moments correspondant à certaines valeurs des paramètres p^* , λ^* et α :

α	1.	0.99	0.975	0.95	0.925	0.90
m^1 ex.	2.	1.98	1.951	1.904	1.860	1.818
m^1 app.	2.	1.98	1.950	1.900	1.850	1.800
m^0 ex.	1.33	1.353	1.384	1.436	1.491	1.548
m^0 app.	1.33	1.353	1.383	1.433	1.483	1.533
$m^1_{[2]}$ ex.	4.	3.881	3.712	3.446	3.201	2.975
$m^1_{[2]}$ app.	4.	3.880	3.700	3.400	3.100	2.800
$m^0_{[2]}$ ex.	0.88	0.961	1.074	1.276	1.498	1.740
$m^0_{[2]}$ app.	0.88	0.960	1.066	1.244	1.422	1.600

Tableau 1 : $p^* = .6$; $\lambda^* = .5$

(ii) Plaçons - nous maintenant dans le cas où $(X_t^*; t \geq 1) \sim \text{RA}(\text{BNT}(h_0, d_0), \text{BNT}(h_1, d_1))$, alors on a

$$m^1(*) = h_1 + 1 \quad m^1_{[2]}(*) = h_1(h_1 + d_1 + 2)$$

$$m^0(*) = h_0 + 1 \quad m^0_{[2]}(*) = h_0(h_0 + d_0 + 2)$$

On obtient les résultats suivants pour différentes valeurs des paramètres :

α	1.	0.99	0.975	0.95	0.925	0.90
m^1 ex.	4.33	4.193	3.999	3.714	3.466	3.249
m^1 app.	4.33	4.188	3.971	3.610	3.249	2.888
m^0 ex.	5.50	5.418	5.309	5.157	5.037	4.944
m^0 app.	5.50	5.415	5.287	5.075	4.862	4.650
$m^1_{[2]}$ ex.	19.99	18.833	17.255	14.987	13.090	11.487
$m^1_{[2]}$ app.	19.99	18.787	16.977	13.960	10.942	7.925
$m^0_{[2]}$ ex.	36.	35.455	34.739	33.782	33.081	32.603
$m^0_{[2]}$ app.	36.	35.427	34.567	33.135	31.703	30.271

Tableau 2 : $\begin{cases} h_1 = 3.33 \\ d_1 = .66 \end{cases} \quad \begin{cases} h_0 = 4.50 \\ d_0 = 1.50 \end{cases}$

α	1.	0.99	0.975	0.95	0.925	0.90
m^1 ex.	4.	3.883	3.720	3.478	3.265	3.076
m^1 app.	4.	3.880	3.70	3.40	3.10	2.80
m^0 ex.	3.	2.981	2.957	2.929	2.912	2.905
m^0 app.	3.	2.98	2.95	2.90	2.85	2.80
$m^1_{[2]}$ ex.	30.	27.657	24.597	20.457	17.214	14.628
$m^1_{[2]}$ app.	30.	27.527	23.80	17.60	11.40	5.20
$m^0_{[2]}$ ex.	16.	15.871	15.713	15.536	15.456	15.464
$m^0_{[2]}$ app.	16.	15.861	15.653	15.307	14.960	14.614

Tableau 3 : $\begin{cases} h_1 = 3. \\ d_1 = 5. \end{cases} \quad \begin{cases} h_0 = 2. \\ d_0 = 4. \end{cases}$

II. ANALYSE STATISTIQUE ET ETUDE EXPERIMENTALE.

Nous étudions ici le problème statistique de l'estimation des paramètres du modèle (1) au vu de l'observation sur un intervalle de temps $\{1, \dots, n\}$ d'une trajectoire du processus $(\vec{X}_t; t \geq 1)$. Nous verrons dans le chapitre VI comment, dans les applications, où on cherche à ajuster un modèle à des données climatologiques concernant un réseau de stations et une période $\{1, \dots, T\}$ (mois, ensemble de mois..) de l'année pour plusieurs années consécutives, il est possible de construire une trajectoire $(\vec{x}_1, \dots, \vec{x}_n)$ dont on puisse admettre qu'elle obéit sur une longue durée n aux "règles" valables sur la période $\{1, \dots, T\}$.

Nous examinons successivement les deux situations qui se présentent selon que la station test est une station privilégiée du réseau ou une station fictive non observable puis nous menons une étude expérimentale par simulations des procédures d'estimation des paramètres du modèle. Nous renvoyons à notre conclusion pour quelques remarques concernant cette étude.

II.1. Cas d'une station test observable.

On observe donc $\vec{X}_t = (X_t^1, \dots, X_t^q)$ pour $t \in \{1, \dots, n\}$ où

- $(X_t^1; t \geq 1) \sim \text{RA}(p^0(*), p^1(*))$

- $X_t^j = \epsilon_t^j X_t^1$; $j=2, \dots, q$; $t \geq 1$

- $(\epsilon_t^j; t \geq 1)$ est une suite de variables aléatoires indépendantes de Bernoulli de même paramètre α_j , $j=2, \dots, q$

- les suites $(X_t^1; t \geq 1)$, $(\epsilon_t^j; t \geq 1)$, $j=2, \dots, q$ sont indépendantes entre elles.

On cherche à estimer au vu de cette observation les paramètres des lois $p^0(*)$ et $p^1(*)$ ainsi que $\alpha_2, \dots, \alpha_q$.

(a) Il est naturel de chercher des estimations des paramètres des lois $p^0(*)$ et $p^1(*)$ à partir de la seule observation (X_1^1, \dots, X_n^1) . On dispose pour ce faire de procédures convergentes (cf. Bousseboua (1983)) :

- Si $S_n^1 = \sum_{t=1}^n X_t^1$ alors $\frac{S_n^1}{n}$ est un estimateur sans biais convergent en probabilité de $p^* = \frac{m^1(*)}{m^0(*) + m^1(*)}$ et de plus la suite $\{\sqrt{n}(\frac{S_n^1}{n} - p^*); n \geq 1\}$ converge en loi vers une variable aléatoire gaussienne centrée de variance

$$\frac{(m^1(*)^2 m_2^0(*) + (m^0(*)^2 m_2^1(*)}{(m^0(*) + m^1(*)^3}$$

où $m_2^i(*)$ est la variance de la loi $p^i(*)$;

- Si $N_n^{0,1} = \sum_{t=2}^n (1 - X_{t-1}^1) X_t^1$ (resp. $N_n^{1,0} = \sum_{t=2}^n X_{t-1}^1 (1 - X_t^1)$)

est le nombre de séquences sèches (resp. humides) débutant sur la période $\{1, \dots, n\}$ dans la station 1, alors

$\frac{N_n^{0,1}}{n}$ et $\frac{N_n^{1,0}}{n}$ sont des estimateurs convergents en probabilité de $\frac{1}{m^0(*) + m^1(*)}$;

- Les statistiques $\frac{S_n^1}{N_n^{0,1}}$ (resp. $\frac{n - S_n^1}{N_n^{1,0}}$) et $\frac{S_n^1}{N_n^{1,0}}$ (resp. $\frac{n - S_n^1}{N_n^{0,1}}$) sont des estimateurs convergents en probabilité de $m^1(*)$ (resp. $m^0(*)$) ;

- Si $m^i(*, n)$ (resp. $m_2^i(*, n)$) est la moyenne (resp. variance) empirique des séquences de type i débutant sur la période $\{1, \dots, n\}$

dans la station 1, alors $m^1(*,n)$ (resp. $m_2^1(*,n)$) est un estimateur convergent en probabilité de $m^1(*)$ (resp. $m_2^1(*)$). Alors

$\frac{m^1(*,n)}{m^0(*,n)+m^1(*,n)}$ est aussi un estimateur convergent de p^* ;

- Dans le cas particulier où $(X_t^1; t \geq 1) \sim \text{RA}(\text{BNT}(h_*^0, d_*^0), \text{BNT}(h_*^1, d_*^1))$, compte tenu de ce que

$$m^1(*) = h_*^1 + 1 \quad ; \quad m_2^1(*) = h_*^1(d_*^1 + 1)$$

on sera amené à estimer h_*^1 et d_*^1 respectivement par

$$h_*^1(n) = m^1(*,n) - 1 \quad , \quad d_*^1(n) = \frac{m_2^1(*,n)}{m^1(*,n) - 1} - 1 \quad ;$$

- Dans le cas particulier où $(X_t^1; t \geq 1) \sim \text{MA}(1; p^*, \lambda^*)$, compte tenu de ce que

$$m^1(*) = \frac{1}{1 - \lambda^*} \quad ; \quad m^0(*) = \frac{(1-p^*)}{p^*} \cdot \frac{1}{1 - \lambda^*}$$

on pourra estimer λ^* par $1 - \frac{1}{m^1(*,n)}$ ou par $\frac{N_n^{11}}{S_n^1}$

où $\frac{N_n^{11}}{S_n^1} = \frac{1}{n} \sum_{t=2}^n X_{t-1}^1 X_t^1$ et p^* par $\frac{m^1(*,n)}{m^0(*,n) + m^1(*,n)}$ ou par $\frac{S_n^1}{n}$;

- Dans le cas particulier où $(X_t^1; t \geq 1) \sim \text{MA}(2; \lambda_1^*, \lambda_2^*, \mu^*)$, compte tenu de ce que

$$m^1(*) = \frac{1}{1 - \lambda_1^*} \quad ; \quad m^0(*) = \frac{1 - p^*}{p^*} \cdot \frac{1}{1 - \lambda_1^*}$$

on estimera λ_1^* par $1 - \frac{1}{m^1(*,n)}$ ou par $\frac{N_n^{11}}{S_n^1}$ et p^* par

$\frac{m^1(*,n)}{m^0(*,n) + m^1(*,n)}$ ou par $\frac{S_n^1}{n}$. De même on estimera λ_2^*

par $\frac{\sum_{t=3}^n X_{t-2}^1 X_t^1}{S_n^1}$ et μ^* par $\frac{\sum_{t=3}^n X_{t-2}^1 X_{t-1}^1 X_t^1}{N_n^{11}}$.

(b) En ce qui concerne l'estimation des paramètres α_j , $j=2, \dots, q$, observant que

$$P[X_t^j = 1 / X_t^1 = 1] = \alpha_j = \frac{P[X_t^j = 1]}{P[X_t^1 = 1]}$$

il est naturel de proposer les estimateurs convergents :

$$\alpha_j(n) = \frac{S_n^j}{S_n^1} \quad ; \quad j=2, \dots, q$$

où pour $j=1, \dots, q$, $S_n^j = \sum_{t=1}^n X_t^j$ compte le nombre de jours secs dans la station j durant la période $\{1, \dots, n\}$.

Une fois les paramètres estimés on dispose, compte tenu de l'étude des caractéristiques du modèle menée dans le §I, de différentes possibilités pour apprécier la qualité de l'ajustement correspondant à des observations. On peut par exemple comparer les lois de probabilité empiriques des temps de séjour aux lois $p^i(j)$ calculées (pour les valeurs estimées des paramètres (cf. proposition 3), comparer les probabilités de transition empiriques aux probabilités $P[N_{t+1} = 1 / N_t = k]$ calculées (cf. corollaire 1), comparer les moments observés des temps passés dans un état donné au cours de k jours consécutifs aux moments calculés $E(S_k^j)$, $\text{var}(S_k^j)$ (cf. (14) ci dessus et aussi chapitre II) ...

On se propose maintenant de cerner les possibilités d'utilisation du modèle par une étude expérimentale des procédures proposées d'estimation et de confrontation entre un modèle et des données. On simule des trajectoires obéissant au modèle (1) avec $q = 4$ pour certaines lois particulières $p^0(*)$ et $p^1(*)$ et certaines valeurs de $\alpha_2, \alpha_3, \alpha_4$.

On choisira n voisin de 1500, dans la mesure où dans les applications cela correspond à l'observation sur 2 mois donnés pour 25 années consécutives, et les α_i voisins de 1 dans la mesure où on s'intéresse en pratique à des réseaux homogènes.

Paramètres de la simulation : $X_t^1 \sim \text{RA}(\text{BNT}(h_1, d_1), \text{BNT}(h_0, d_0))$

avec $\begin{cases} h_1 = 5. \\ d_1 = 7. \end{cases}$ $\begin{cases} h_0 = 1. \\ d_0 = 1.5 \end{cases}$

et $\alpha_2 = .99$ $\alpha_3 = .97$ $\alpha_4 = .95$ $n = 1541$

	h_1	d_1	h_0	d_0	α_2	α_3	α_4
val. d'entrée	5.	7.	1.	1.5	.99	.97	.95
val. estimées	4.635	6.717	1.04	1.402	.987	.976	.944

Tableau 4. Paramètres d'entrée et estimés

Lois de temps de séjour.

station	1	2	3	4
m^1 ex.	6.	5.714	5.217	4.80
m^1 emp.	5.635	5.33	5.074	4.476
m^1 cal.	5.635	5.314	5.070	4.473

Tableau 5. Moments exacts , empiriques et calculés d'ordre 1 des lois des temps de séjour dans l'état 1.

station	1	2	3	4
m^0 ex.	2.	1.981	1.954	1.936
m^0 emp.	2.04	2.014	2.065	2.
m^0 cal.	2.04	2.019	2.005	1.981

Tableau 6. idem tableau 5 : état 0.

station	1	2	3	4
$m_{[2]}^1$ ex.	70.	62.103.	49.716	40.556
$m_{[2]}^1$ emp.	61.89	53.014	48.651	34.751
$m_{[2]}^1$ cal.	61.89	53.47	47.568	34.84

Tableau 7. Moments exacts, empiriques et calculés d'ordre 2 des lois des temps de séjour dans l'état 1.

station	1	2	3	4
$m_{[2]}^0$ ex.	4.50	4.425	4.315	4.250
$m_{[2]}^0$ emp.	4.62	4.497	4.707	4.548
$m_{[2]}^0$ cal.	4.62	4.544	4.496	4.424

Tableau 8. idem tableau 7 : état 0.

$N_{t+1} \backslash N_t$	0	1	2	3
0	0.5000 0.5193 0.5098	0.1667 0.4000 0.1775	0.1667 0.1810 0.1775	0.1667 0.1750 0.1775
1	0.0011 0.0048 0.0011	0.0019 0.0000 0.0019	0.0019 0.0000 0.0019	0.0019 0.0029 0.0019
2	0.0427 0.0531 0.0429	0.0712 0.0000 0.0719	0.0712 0.0857 0.0719	0.0712 0.0728 0.0719
3	0.4562 0.4251 0.4462	0.7602 0.6000 0.7488	0.7602 0.7333 0.7488	0.7602 0.7483 0.7488

Tableau 9. Probabilités de transition du processus stochastique N_t : exactes, empiriques et calculées.

Longueur de la séquence	Nbre. de séquences (N=209)	Probabilités empiriques	Probabilités théorique exactes	Probabilités théoriques calculées
1	56	1.0000	1.0000	1.0000
2	23	0.7321	0.7692	0.7509
3	23	0.6220	0.6235	0.5994
4	15	0.5120	0.5143	0.4880
5	22	0.4402	0.4283	0.4016
6	15	0.3349	0.3588	0.3328
7	8	0.2632	0.3019	0.2770
8	6	0.2249	0.2548	0.2314
9	8	0.1962	0.2156	0.1939
10	3	0.1579	0.1828	0.1628
11	3	0.1435	0.1552	0.1370
12	10	0.1292	0.1320	0.1154
13	4	0.0813	0.1124	0.0974
14	1	0.0622	0.0958	0.0823
15	0	0.0574	0.0818	0.0695
16	3	0.0574	0.0698	0.0589
17	0	0.0431	0.0597	0.0499
18	1	0.0431	0.0510	0.0423
19	3	0.0383	0.0437	0.0358
20	1	0.0239	0.0374	0.0304
21	1	0.0191	0.0320	0.0258
22	1	0.0144	0.0274	0.0219
23	0	0.0096	0.0235	0.0187
24	0	0.0096	0.0202	0.0159
25	0	0.0096	0.0173	0.0135
26	0	0.0096	0.0149	0.0115
27	0	0.0096	0.0128	0.0098
28	0	0.0096	0.0109	0.0083
29	0	0.0096	0.0094	0.0071
30	0	0.0096	0.0081	0.0060
31	0	0.0096	0.0069	0.0051
32	0	0.0096	0.0060	0.0044

Tableau 10. Lois des temps de séjour dans l'état 1 : empirique , théorique exacte, théorique calculée ; $\alpha_2 = .99$, $\alpha_2(n) = .987$ (voir tableau 4 pour les autres caractéristiques).

Longueur de la séquence	Nbre. de séquences (N= 215)	Probabilités empiriques	Probabilités théoriques exactes	Probabilités théoriques calculées
1	59	1.0000	1.0000	1.0000
2	25	0.7256	0.7594	0.7457
3	23	0.6093	0.6054	0.5896
4	16	0.5023	0.4904	0.4752
5	23	0.4279	0.4008	0.3870
6	17	0.3209	0.3295	0.3171
7	12	0.2419	0.2719	0.2611
8	5	0.1860	0.2250	0.2157
9	6	0.1628	0.1867	0.1787
10	2	0.1349	0.1552	0.1484
11	5	0.1256	0.1293	0.1234
12	6	0.1023	0.1078	0.1028
13	4	0.0744	0.0900	0.0858
14	2	0.0558	0.0752	0.0716
15	1	0.0465	0.0629	0.0599
16	1	0.0419	0.0526	0.0501
17	1	0.0372	0.0441	0.0420
18	0	0.0326	0.0369	0.0352
19	1	0.0326	0.0310	0.0295
20	1	0.0279	0.0260	0.0247
21	1	0.0233	0.0218	0.0208
22	1	0.0186	0.0183	0.0174
23	0	0.0140	0.0154	0.0146
24	0	0.0140	0.0129	0.0123
25	0	0.0140	0.0109	0.0104
26	0	0.0140	0.0091	0.0087
27	0	0.0140	0.0077	0.0073
28	0	0.0140	0.0065	0.0062
29	1	0.0140	0.0055	0.0052
30	0	0.0093	0.0046	0.0044
31	0	0.0093	0.0039	0.0037
32	0	0.0093	0.0033	0.0031

Tableau 11. Lois des temps de séjour dans l'état 1 : empirique , théorique exacte , théorique calculée ; $\alpha_3 = .97$, $\alpha_3(n) = .976$ (voir tableau 4 por les autres caractéristiques).

Longueur de la séquence	Nbre. de séquences (N=237)	Probabilités empiriques	Probabilités théoriques exactes	Probabilités théoriques calculées
1	65	1.0000	1.0000	1.0000
2	34	0.7257	0.7485	0.7294
3	33	0.5823	0.5861	0.5607
4	18	0.4430	0.4660	0.4386
5	25	0.3671	0.3735	0.3463
6	19	0.2616	0.3010	0.2750
7	5	0.1814	0.2435	0.2194
8	8	0.1603	0.1976	0.1755
9	4	0.1266	0.1606	0.1408
10	2	0.1097	0.1309	0.1132
11	4	0.1013	0.1068	0.0911
12	6	0.0844	0.0872	0.0735
13	3	0.0591	0.0713	0.0593
14	0	0.0464	0.0584	0.0480
15	2	0.0464	0.0479	0.0388
16	1	0.0380	0.0392	0.0314
17	1	0.0338	0.0322	0.0255
18	1	0.0295	0.0264	0.0206
19	2	0.0253	0.0217	0.0167
20	2	0.0169	0.0179	0.0136
21	1	0.0084	0.0147	0.0110
22	0	0.0042	0.0121	0.0090
23	0	0.0042	0.0099	0.0073
24	0	0.0042	0.0082	0.0059
25	0	0.0042	0.0067	0.0048
26	0	0.0042	0.0055	0.0039
27	0	0.0042	0.0046	0.0032
28	1	0.0042	0.0038	0.0026
29	0	0.0000	0.0031	0.0021
30	0	0.0000	0.0026	0.0017
31	0	0.0000	0.0021	0.0014
32	0	0.0000	0.0017	0.0011

Tableau 12. idem tableau 11 :

$$\alpha_4 = .95 \quad , \quad \alpha_4(n) = .944 .$$

	N_t	S_{30}^2	S_{30}^3	S_{30}^4
exact	2.182	22.275	21.825	21.375
empirique	2.119	21.784	21.333	20.745
calculé	2.135	21.754	21.501	20.796

Tableau 13. Moments d'ordre 1 des statistiques N_t et S_{30}^j ; $j=2,3,4$.

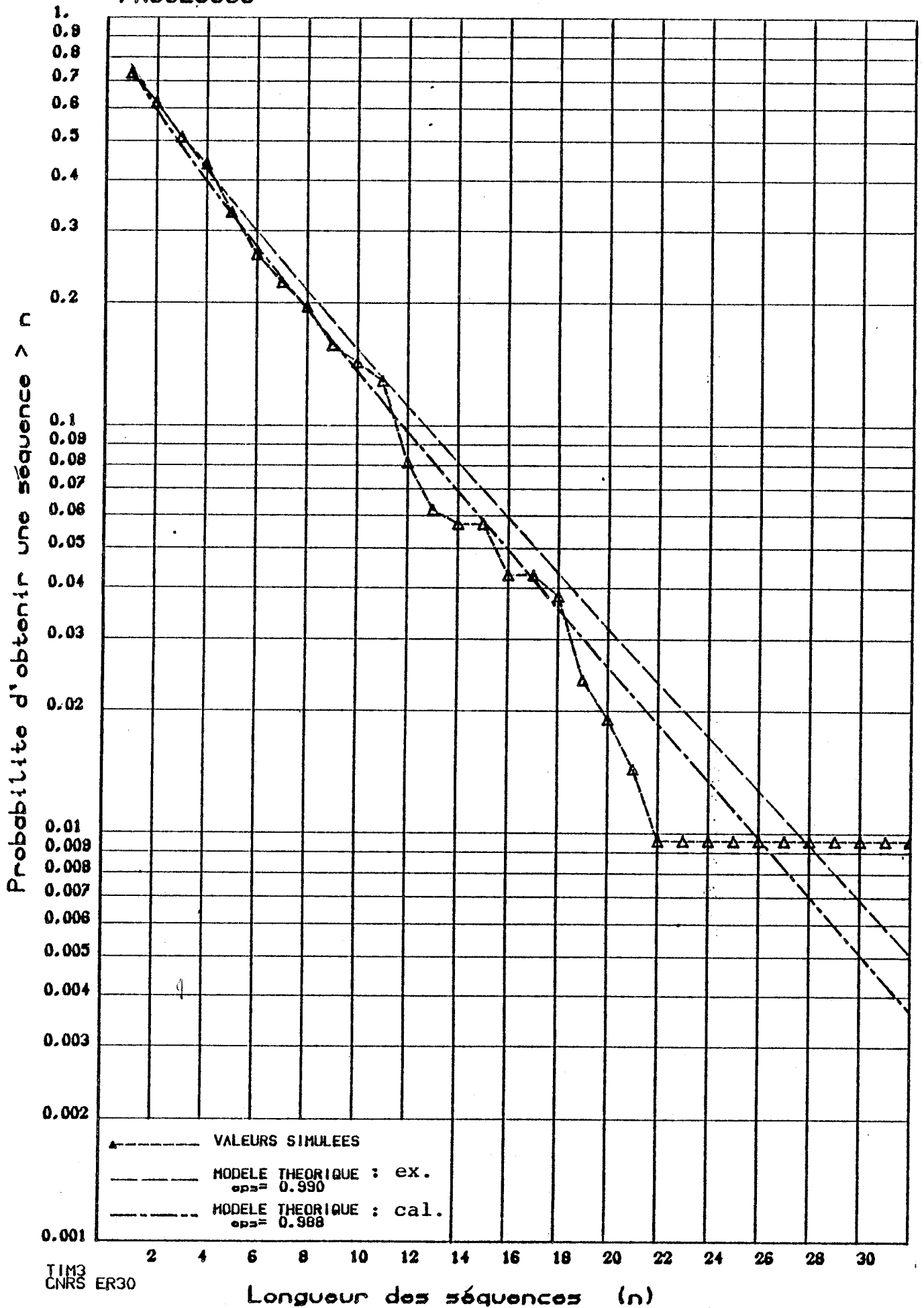
	N_t	S_{30}^2	S_{30}^3	S_{30}^4
exact	1.587	17.440	17.183	16.923
empirique	1.723	17.580	16.339	16.50
calculé	1.650	17.404	17.252	16.821

Tableau 14. idem tableau 13 : ordre 2 .

ss-réseau	fréquences	Prob. th. ex.	Prob. emp.	Prob. th. cal.
111	1017	0.6842	0.6600	0.6683
011	13	0.0069	0.0084	0.0084
101	31	0.0212	0.0201	0.0163
110	61	0.0360	0.0396	0.0395
001	0	0.0002	0.0000	0.0002
010	0	0.0004	0.0000	0.0005
100	5	0.0011	0.0032	0.0010
000	414	0.2500	0.2687	0.2658

Tableau 15. Probabilités pour qu'un sous-réseau soit sec : théoriques exactes, empiriques, théoriques calculées.

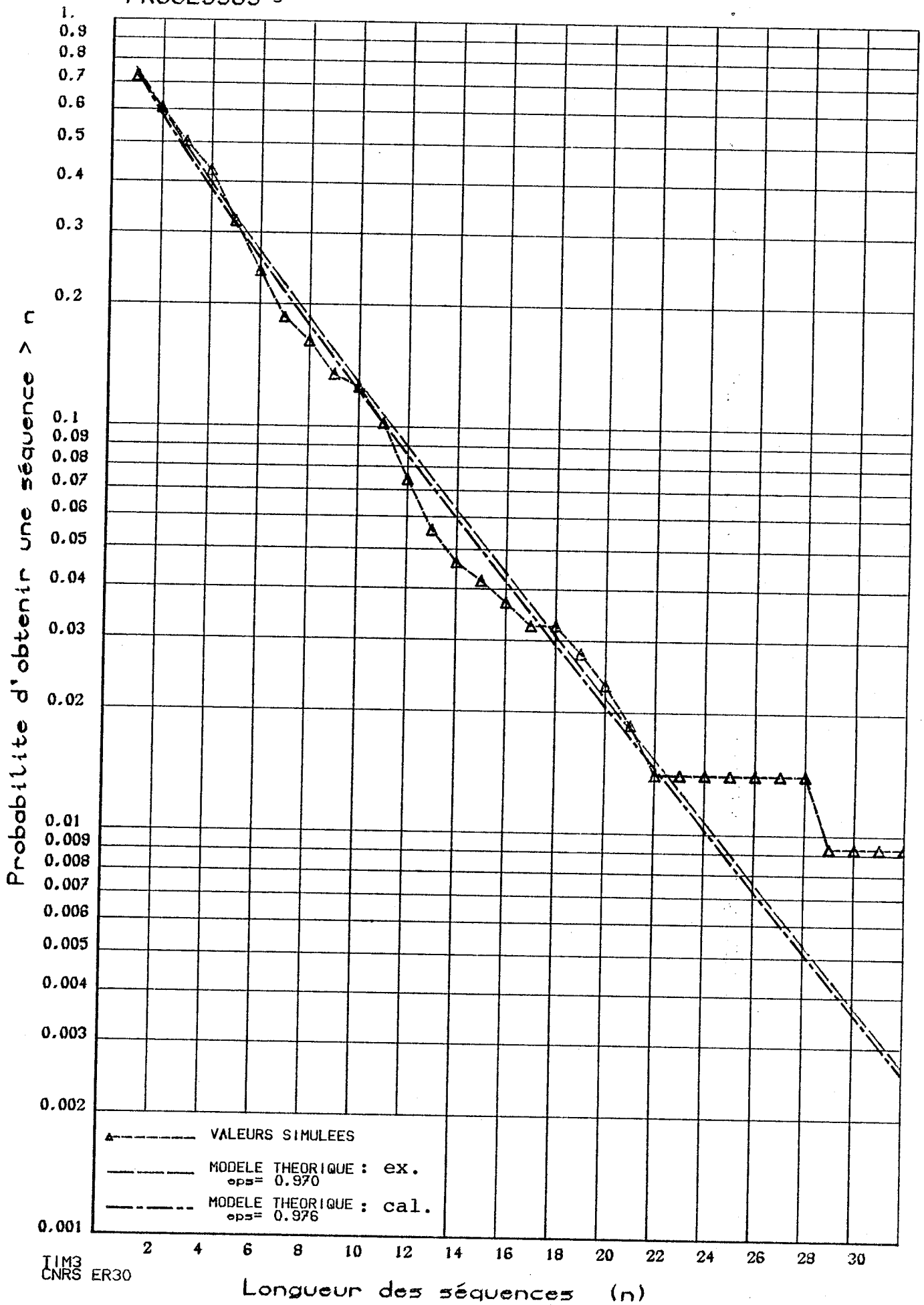
PROCESSUS 2



TIM3
CNRS ER30

Fig. 1. Lois des temps de séquences dans l'état 2.

PROCESSUS 3



IM3
CNRS ER30

Fig 2. idem figure 1.

II.2. Cas d'une station test non observable.

On observe donc $\vec{X}_t = (X_t^1, \dots, X_t^q)'$ pour $t \in \{1, \dots, n\}$
 où $X_t^j = \varepsilon_t^j X_t^*$; $j=1, \dots, q$, $(X_t^* ; t \geq 1)$ et $(\varepsilon_t^j ; t \geq 1)$, $j=1, \dots, q$
 vérifiant les conditions imposées dans le modèle (1).

On cherche à estimer au vu de cette seule observation
 les paramètres du processus $(X_t^* ; t \geq 1)$ et $\alpha_1, \dots, \alpha_q$.

(a) On peut d'abord envisager de faire l'estimation des paramètres $\alpha_1, \dots, \alpha_q$ de la manière suivante : comme pour tout $j \neq k$ on a

$$P[X_t^j = 1 / X_t^k = 1] = \alpha_j = \frac{E(X_t^j X_t^k)}{E(X_t^k)}$$

on a aussi

$$\alpha_j = \frac{\sum_{k \neq j} E(X_t^j X_t^k)}{\sum_{k \neq j} E(X_t^k)} = \frac{E(X_t^j \sum_{k \neq j} X_t^k)}{E(\sum_{k \neq j} X_t^k)}$$

soit, avec la notation $N_t = \sum_{j=1}^q X_t^j$ introduite précédemment pour
 la variable comptant le nombre de stations "sèches" un jour t,

$$\alpha_j = \frac{E[X_t^j (N_t - X_t^j)]}{E(N_t - X_t^j)} ; \quad j=1, \dots, q .$$

Compte tenu de (3), on a aussi

$$\alpha_j = \frac{P[X_t^1 = \dots = X_t^q = 1]}{P[X_t^k = 1, k \neq j]} = \frac{E(\prod_{k=1}^q X_t^k)}{E(\prod_{k \neq j} X_t^k)}$$

Cela suggère les estimateurs $\alpha_j^1(n)$ et $\alpha_j^2(n)$ de α_j définis respectivement par :

$$\alpha_j^1(n) = \frac{\sum_{t=1}^n X_t^j (N_t - X_t^j)}{\sum_{t=1}^n (N_t - X_t^j)}$$

et

$$\alpha_j^2(n) = \frac{\sum_{t=1}^n \left(\prod_{k=1}^q X_t^k \right)}{\sum_{t=1}^n \left(\prod_{k \neq j} X_t^k \right)}$$

(b) On peut conduire ensuite l'estimation des paramètres du processus $(X_t^*; t \geq 1)$ comme suit : comme

$$P[X_t^j = 1] = \alpha_j p^* = \alpha_j \frac{m^1(*)}{m^0(*) + m^1(*)} \quad \text{pour tout } j=1, \dots, q,$$

on a

$$p^* = \frac{\sum_{j=1}^q E(X_t^j)}{\sum_{j=1}^q \alpha_j} = \frac{E(N_t)}{\sum_{j=1}^q \alpha_j}$$

Cela amène à proposer les estimateurs de p^* définis par :

$$p^*(n) = \frac{1}{n} \cdot \frac{\sum_{t=1}^n N_t}{\sum_{j=1}^q \alpha_j(n)}$$

où $\alpha_j(n)$ est l'un des estimateurs $\alpha_j^1(n)$ et $\alpha_j^2(n)$ de α_j définis précédemment.

De même comme d'après (18) ci dessus on a, pour tout $j=1, \dots, q$

$$m^1(*) = \frac{m^1(j) \alpha_j}{1 - (1 - \alpha_j) m^1(j)} ;$$

$$m^0(*) = \frac{\alpha_j [m^0(j) \alpha_j - (1 - \alpha_j) m^1(j)]}{1 - (1 - \alpha_j) m^1(j)}$$

On a aussi

$$m^1(*) = \frac{\sum_{j=1}^q m^1(j) \alpha_j}{q - \sum_{j=1}^q (1-\alpha_j) m^1(j)} ;$$

$$m^0(*) = \frac{\sum_{j=1}^q \alpha_j [m^0(j) \alpha_j - (1-\alpha_j) m^1(j)]}{q - \sum_{j=1}^q (1-\alpha_j) m^1(j)}$$

Cela suggère les estimateurs de $m^1(*)$ et $m^0(*)$ définis par

$$m^1(*,n) = \frac{\sum_{j=1}^q m^1(j,n) \alpha_j(n)}{q - \sum_{j=1}^q (1-\alpha_j(n)) m^1(j,n)} ;$$

$$m^0(*,n) = \frac{\sum_{j=1}^q \alpha_j(n) [m^0(j,n) \alpha_j(n) - (1-\alpha_j(n)) m^1(j)]}{q - \sum_{j=1}^q (1-\alpha_j(n)) m^1(j)}$$

où $m^i(j,n)$ est la moyenne empirique des séquences de type i débutant sur la période $\{1, \dots, n\}$ dans la station j .

Cela fournit d'ailleurs un autre estimateur de p^* :

$$\frac{m^1(*,n)}{m^0(*,n) + m^1(*,n)}$$

Lorsque les lois $p^0(*)$ et $p^1(*)$ dépendent chacune de deux paramètres il reste à fournir une estimation des moments factoriels d'ordre 2 $m_{[2]}^1(*)$ et $m_{[2]}^0(*)$. On utilise pour cela les approximations suivantes que l'on peut facilement démontrer de la même manière qu'ont été démontrées celles du corollaire 3 ci-dessus : lorsque α_j est voisin de 1 on a

$$m_{[2]}^1(*) \approx m_{[2]}^1(j) + (1-\alpha_j) \{ m^1(j) m_{[2]}^1(j) + \frac{1}{3} m_{[3]}^1(j) \}$$

et

$$m_{[2]}^0(*) \approx m_{[2]}^0(j) + (1-\alpha_j) \{ (m^1(j) - 2) m_{[2]}^0(j) - 2 m^0(j) [2 + m^0(j) p_1^1(j)] \}.$$

Il est alors raisonnable d'estimer $m_{[2]}^1(*)$ et $m_{[2]}^0(*)$ respectivement par

$$m_{[2]}^1(*,n) = \frac{1}{q} \sum_{j=1}^q \{ m_{[2]}^1(j,n) + (1-\alpha_j(n)) [m^1(j,n) + \frac{1}{3} m_{[3]}^1(j,n)] \}$$

$$m_{[2]}^0(*,n) = \frac{1}{q} \sum_{j=1}^q \{ m_{[2]}^0(j,n) + (1-\alpha_j(n)) [(m^1(j,n)-2)m_{[2]}^0(j,n) - 2 m^0(j,n) [2 + m^0(j,n)p_1^1(j,n)]] \}$$

où $m_{[k]}^i(j,n)$ est le moment factoriel empirique d'ordre k des séquences de type i débutant sur la période $\{1, \dots, n\}$ dans la station j et

$$p_1^1(j,n) = \frac{\sum_{t=3}^n (1-x_{t-2}^j) x_{t-1}^j (1-x_t^j)}{\sum_{t=2}^n (1-x_{t-1}^j) x_t^j}$$

est l'estimateur naturel de $p_1^1(j)$ défini par le rapport entre le nombre de séquences "sèches" de durée 1 jour observées dans la station j et le nombre total de séquences sèches ayant débuté sur la période $\{1, \dots, n\}$ dans la station j .

Si maintenant on est dans le cadre de l'hypothèse de lois de séjour BNT ou de l'hypothèse markovienne d'ordre 1 ou 2 on peut envisager comme cela a été fait en I.1 de construire des estimateurs des paramètres en jeu dans les lois $p^0(*)$ et $p^1(*)$ correspondantes à partir des estimations précédentes des moments de ces lois.

On envisage maintenant une étude expérimentale dans le même esprit qu'en I.1.

Paramètres de la simulation : $X_t^* \sim RA(BNT(h_1^*, d_1^*), BNT(h_0^*, d_0^*))$

avec $\begin{cases} h_1^* = 5. \\ d_1^* = 7. \end{cases}$ $\begin{cases} h_0^* = 1. \\ d_0^* = 1.5 \end{cases}$

et $\alpha_1 = .99$ $\alpha_2 = .97$ $\alpha_3 = .95$ $n = 1541$.

	h_1	d_1	h_0	d_0	α_1	α_2	α_3
val. d'entrée	5.	7.	1.	1.5	.99	.97	.95
val. estimées	4.555	6.461	1.053	1.376	.987	.970	.943

Tableau 4. Paramètres d'entrée et estimés.

Lois de temps de séjour.

station	1	2	3	4
m^1 ex.	6.	5.714	5.217	4.80
m^1 emp.	5.635	5.330	5.074	4.476
m^1 cal.	5.555	5.244	4.887	4.410

Tableau 5. Moments exacts, empiriques et calculés d'ordre 1 des lois des temps de séjour dans l'état 1.

station	1	2	3	4
m^0 ex.	2.	1.981	1.954	1.936
m^0 emp.	2.04	2.014	2.065	2.0
m^0 cal.	2.053	2.032	2.013	1.994

Tableau 6. idem tableau 5 : état 0.

station	1	2	3	4
$m_{[2]}^1$ ex.	70. . .	62.103	49.716	40.556
$m_{[2]}^1$ emp.	61.89	53.014	48.651	34.751
$m_{[2]}^1$ cal.	59.287	51.394	43.153	33.462

Tableau 7. Moments exacts , empiriques et calculés d'ordre 2 des lois des temps de séjour dans l'état 1.

station	1	2	3	4
$m_{[2]}^0$ ex.	4.50	4.425	4.315	4.250
$m_{[2]}^0$ emp.	4.62	4.497	4.707	4.548
$m_{[2]}^0$ cal.	4.663	4.591	4.526	4.479

Tableau 8. idem tableau 7 : état 0.

$N_{t+1} \backslash N_t$	0	1	2	3
0	0.5000 0.5193 0.5131	0.1667 0.4000 0.1800	0.1667 0.1810 0.1800	0.1667 0.1750 0.1800
1	0.0011 0.0048 0.0013	0.0019 0.0000 0.0022	0.0019 0.0000 0.0022	0.0019 0.0029 0.0022
2	0.0427 0.0531 0.0454	0.0712 0.0000 0.0765	0.0712 0.0857 0.0765	0.0712 0.0728 0.0765
3	0.4562 0.4251 0.4401	0.7602 0.6000 0.7412	0.7602 0.7333 0.7412	0.7602 0.7483 0.7412

Tableau 9. Probabilités de transition du processus stochastique N_t : exactes, empiriques et calculées.

Longueur de la séquence	Nbre. de séquences (N=209)	Probabilités empiriques	Probabilités théoriques exactes	Probabilités théoriques calculées
1	56	1.0000	1.0000	1.0000
2	23	0.7321	0.7692	0.7520
3	23	0.6220	0.6235	0.5990
4	15	0.5120	0.5143	0.4862
5	22	0.4402	0.4283	0.3987
6	15	0.3349	0.3588	0.3290
7	8	0.2632	0.3019	0.2728
8	6	0.2249	0.2548	0.2269
9	8	0.1962	0.2156	0.1893
10	3	0.1579	0.1828	0.1582
11	3	0.1435	0.1552	0.1325
12	10	0.1292	0.1320	0.1111
13	4	0.0813	0.1124	0.0933
14	1	0.0622	0.0958	0.0785
15	0	0.0574	0.0818	0.0661
16	3	0.0574	0.0698	0.0556
17	0	0.0431	0.0597	0.0469
18	1	0.0431	0.0510	0.0396
19	3	0.0383	0.0437	0.0334
20	1	0.0239	0.0374	0.0282
21	1	0.0191	0.0320	0.0238
22	1	0.0144	0.0274	0.0201
23	0	0.0096	0.0235	0.0170
24	0	0.0096	0.0202	0.0144
25	0	0.0096	0.0173	0.0122
26	0	0.0096	0.0149	0.0103
27	0	0.0096	0.0128	0.0088
28	0	0.0096	0.0109	0.0074
29	0	0.0096	0.0094	0.0063
30	0	0.0096	0.0081	0.0053
31	0	0.0096	0.0069	0.0045
32	0	0.0096	0.0060	0.0038

Tableau 10. idem tableau 11 :

$$\alpha_1 = .99, \alpha_1(n) = .987.$$

Longueur de la séquence	Nbre. de séquences (N=215)	Probabilités empiriques	Probabilités théoriques exactes	Probabilités théoriques calculées
1	59	1.0000	1.0000	1.0000
2	25	0.7256	0.7594	0.7439
3	23	0.6093	0.6054	0.5841
4	16	0.5023	0.4904	0.4669
5	23	0.4279	0.4008	0.3768
6	17	0.3209	0.3295	0.3059
7	12	0.2419	0.2719	0.2495
8	5	0.1860	0.2250	0.2042
9	6	0.1628	0.1867	0.1675
10	2	0.1349	0.1552	0.1377
11	5	0.1256	0.1293	0.1134
12	6	0.1023	0.1078	0.0935
13	4	0.0744	0.0900	0.0772
14	2	0.0558	0.0752	0.0638
15	1	0.0465	0.0629	0.0528
16	1	0.0419	0.0526	0.0437
17	1	0.0372	0.0441	0.0362
18	0	0.0326	0.0369	0.0301
19	1	0.0326	0.0310	0.0249
20	1	0.0279	0.0260	0.0207
21	1	0.0233	0.0218	0.0172
22	1	0.0186	0.0183	0.0143
23	0	0.0140	0.0154	0.0119
24	0	0.0140	0.0129	0.0099
25	0	0.0140	0.0109	0.0082
26	0	0.0140	0.0091	0.0068
27	0	0.0140	0.0077	0.0057
28	0	0.0140	0.0065	0.0047
29	1	0.0140	0.0055	0.0040
30	0	0.0093	0.0046	0.0033
31	0	0.0093	0.0039	0.0027
32	0	0.0093	0.0033	0.0023

Tableau 11. Lois des temps de séjour dans l'état 1 : empirique , théorique exacte , théorique calculée ; $\alpha_2 = .97$, $\alpha_2(n) = .970$ (voir tableau 4 pour les autres caractéristiques).

Longueur de la séquence	Nbre. de séquences (N=237)	Probabilités empiriques	Probabilités théoriques exactes	Probabilités théoriques calculées
1	65	1.0000	1.0000	1.0000
2	34	0.7257	0.7485	0.7294
3	33	0.5823	0.5861	0.5589
4	18	0.4430	0.4660	0.4355
5	25	0.3671	0.3735	0.3423
6	19	0.2616	0.3010	0.2706
7	5	0.1814	0.2435	0.2148
8	8	0.1603	0.1976	0.1710
9	4	0.1266	0.1606	0.1365
10	2	0.1097	0.1309	0.1092
11	4	0.1013	0.1068	0.0875
12	6	0.0844	0.0872	0.0702
13	3	0.0591	0.0713	0.0564
14	0	0.0464	0.0584	0.0453
15	2	0.0464	0.0479	0.0365
16	1	0.0380	0.0392	0.0294
17	1	0.0338	0.0322	0.0237
18	1	0.0295	0.0264	0.0191
19	2	0.0253	0.0217	0.0154
20	2	0.0169	0.0179	0.0124
21	1	0.0084	0.0147	0.0100
22	0	0.0042	0.0121	0.0081
23	0	0.0042	0.0099	0.0066
24	0	0.0042	0.0082	0.0053
25	0	0.0042	0.0067	0.0043
26	0	0.0042	0.0055	0.0035
27	0	0.0042	0.0046	0.0028
28	1	0.0042	0.0038	0.0023
29	0	0.0000	0.0031	0.0018
30	0	0.0000	0.0026	0.0015
31	0	0.0000	0.0021	0.0012
32	0	0.0000	0.0017	0.0010

Tableau 12. idem tableau 11 :

$$\alpha_3 = .95 \quad , \quad \alpha_3(n) = .943$$

	N_t	S_{30}^1	S_{30}^2	S_{30}^3
exact	2.182	22.275	21.825	21.375
empirique	2.119	21.784	21.333	20.745
calculé	2.118	21.625	21.254	20.662

Tableau 13. Moments d'ordre 1 des statistiques N_t et S_{30}^j ; $j=1,2,3$.

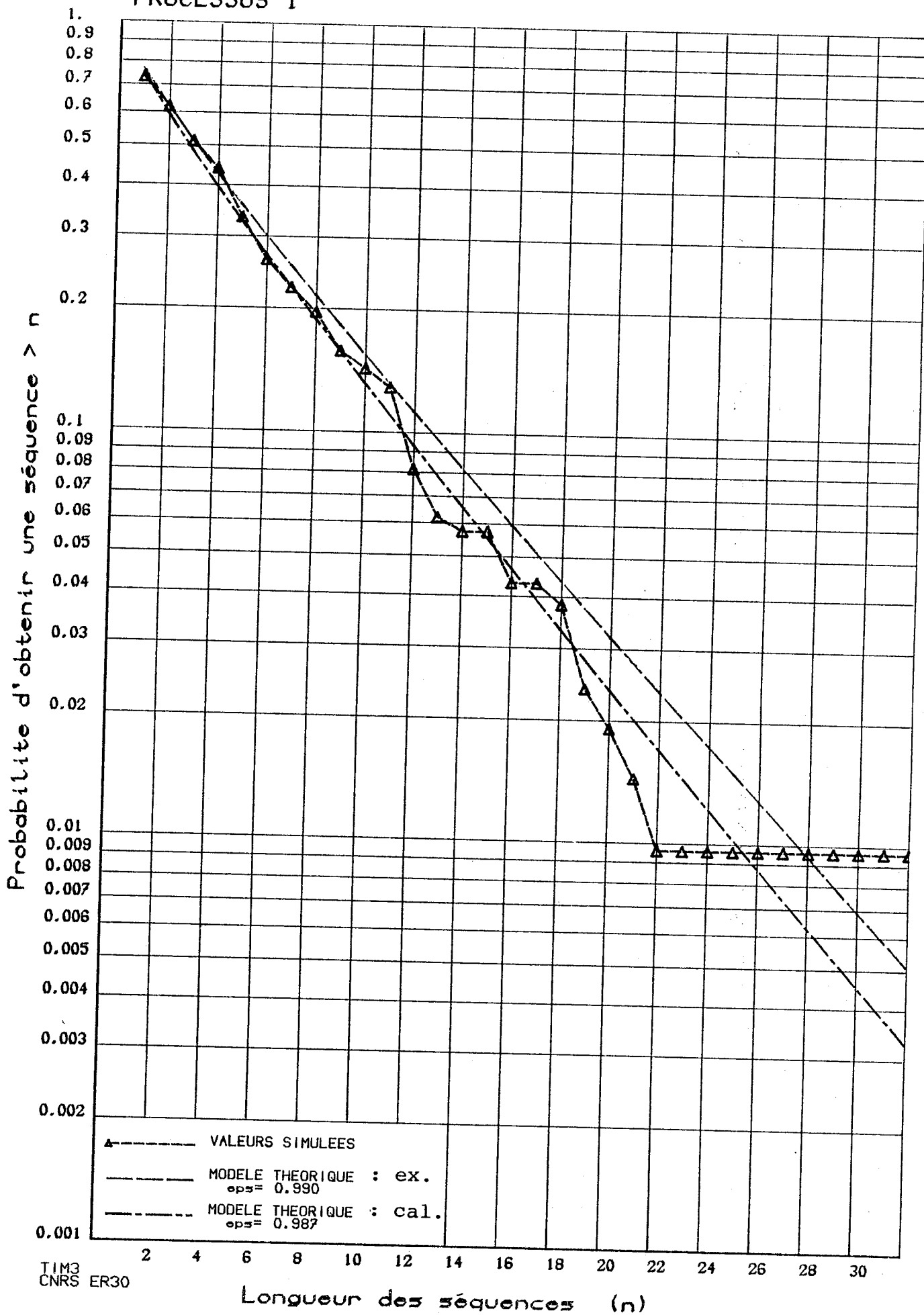
	N_t	S_{30}^1	S_{30}^2	S_{30}^3
exact	1.587	17.458	17.201	16.940
empirique	1.723	17.580	16.339	16.150
calculé	1.658	17.417	17.189	16.821

Tableau 14. idem tableau 13 : ordre 2 .

ss-réseau	fréquences	prob. exactes	prob. emp.	prob. calculées
111	1017	0.6842	0.6600	0.6599
011	13	0.0069	0.0084	0.0084
101	31	0.0212	0.0201	0.0201
110	61	0.0360	0.0396	0.0396
001	0	0.0002	0.0000	0.0003
010	0	0.0004	0.0000	0.0005
100	5	0.0011	0.0032	0.0012
000	414	0.2500	0.2687	0.2699

Tableau 15. Probabilités pour qu'un sous-réseau soit sec : théoriques exactes, empiriques, théoriques calculées.

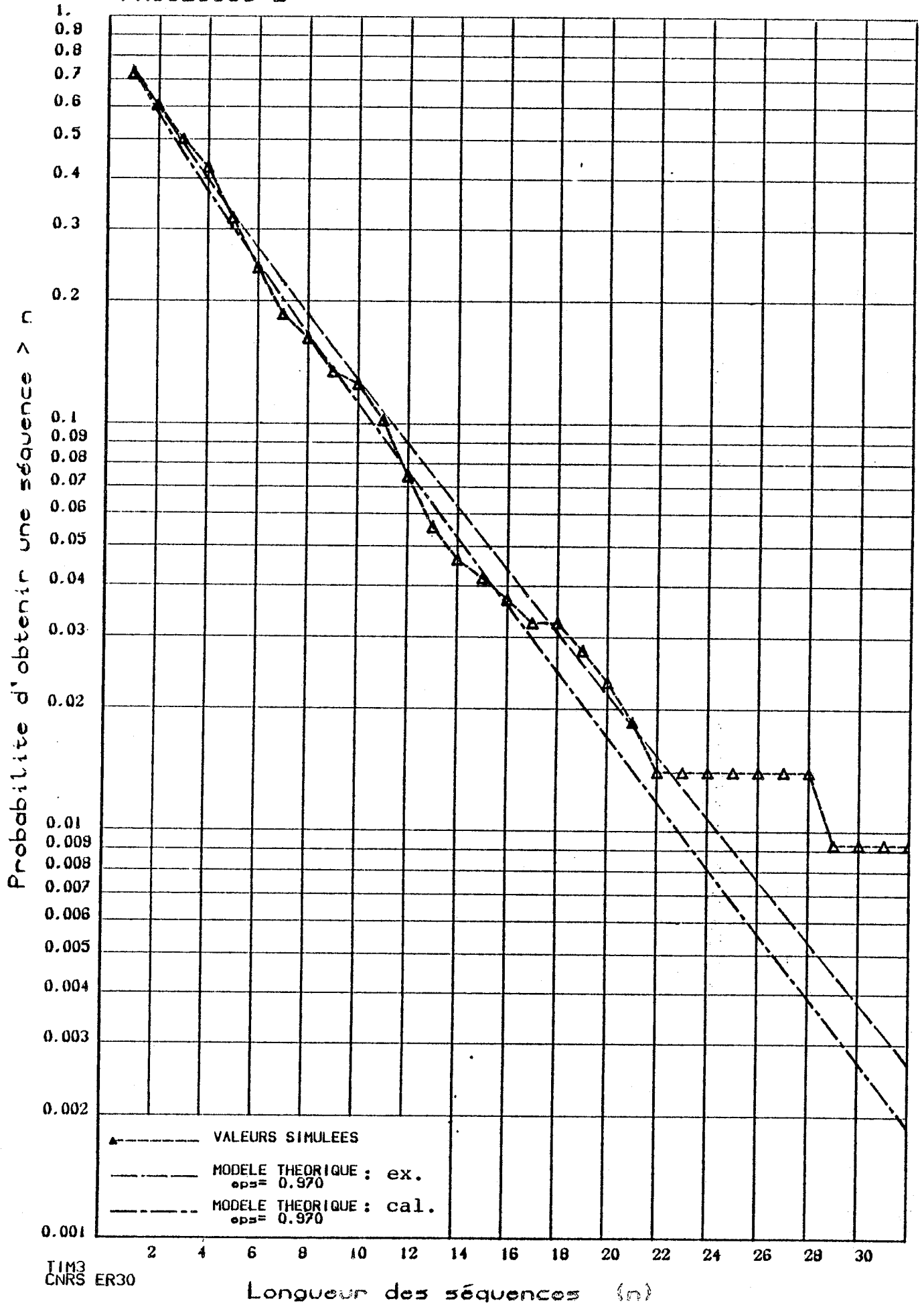
PROCESSUS 1



TIM3
CNRS ER30

Fig 4. Lois des temps de séjour dans l'état 1.

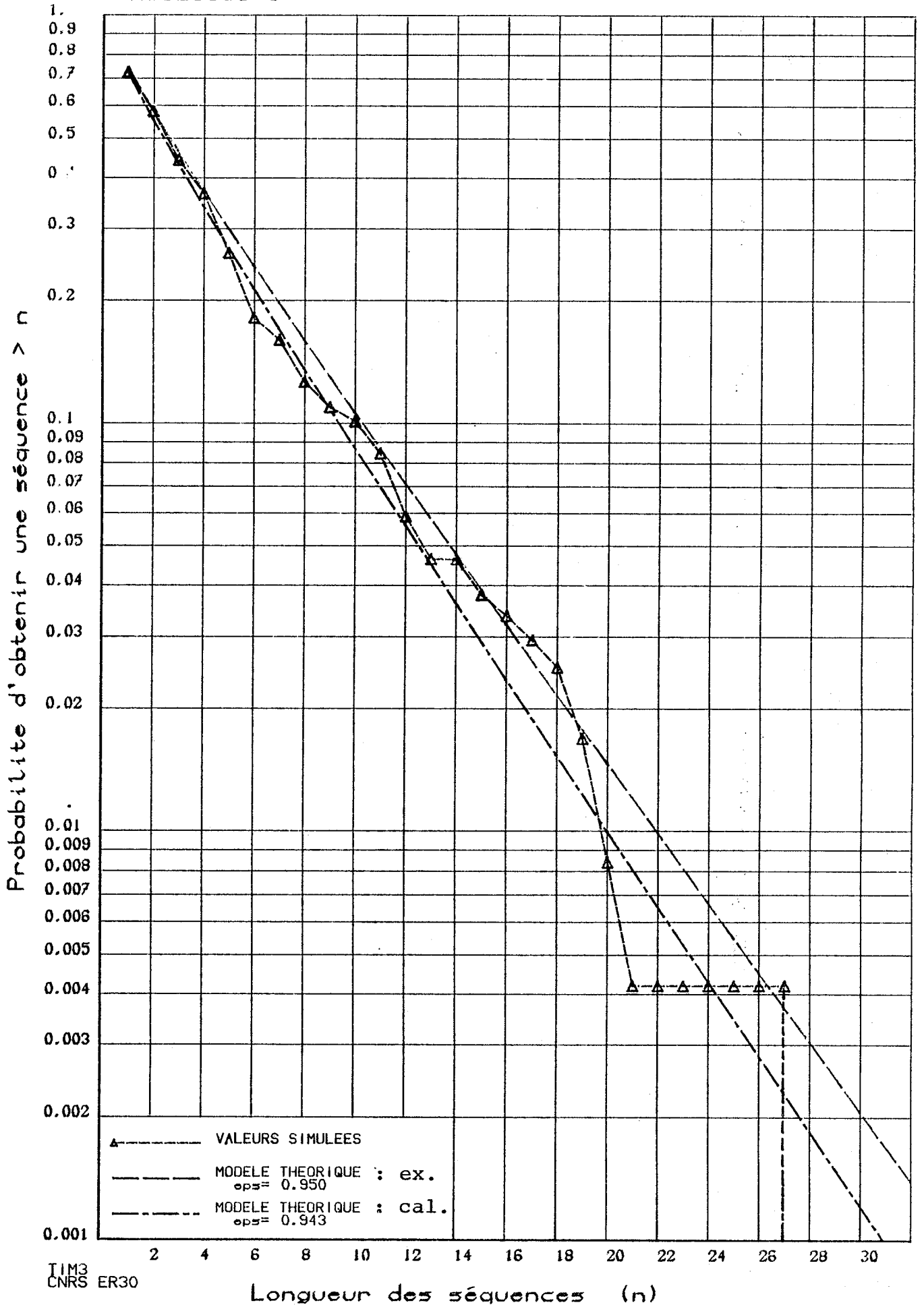
PROCESSUS 2



TIM3
CNRS ER30

Fig 5. idem figure 4

PROCESSUS 3



TIM3
CNRS ER30

Fig 6. idem figure 4.

CHAPITRE 5 :

ETUDE D'UN MODELE DE SERIES CHRONOLOGIQUES VECTORIELLES
A MARGINALES CONSTRUITES PAR SUPERPOSITION D'AMINCISSEMENTS
D'UN PROCESSUS DE RENOUVELLEMENT ALTERNE ET DE SON

"COMPLEMENTAIRE"



Partant des mêmes considérations que celles de l'introduction du chapitre IV, nous envisageons ici un modèle dans lequel l'évolution du phénomène dans la station test ("priviligée" ou fictive) apparaît non plus comme un "extrême" mais comme un certain "juste milieu" des évolutions dans les différentes stations du réseau. De façon précise, nous définissons :

$$(1) \quad X_t^j = \varepsilon_t^j + (1-\varepsilon_t^j)(1-X_t^*) ; j=1, \dots, q ; t \geq 1$$

où les éléments $(X_t^*; t \geq 1)$, $(\varepsilon_t^j; t \geq 1)$, $j=1, \dots, q$ de la modélisation vérifient les mêmes propriétés que dans le chapitre précédent. Lorsque $(X_t^*; t \geq 1)$ est attaché à la station "priviligée" 1 du réseau, on prend bien sûr $X_t^* = X_t^1$ et $\alpha_1 = 1$.

Notons que pour $j=1, \dots, q$ $(X_t^j; t \geq 1)$ est le processus ponctuel obtenu par superposition des processus ponctuels $(\varepsilon_t^j X_t^*; t \geq 1)$ et $(1-\varepsilon_t^j)(1-X_t^*); t \geq 1$ construits respectivement par amincissement du processus ponctuel $(X_t^*; t \geq 1)$ et par amincissement du "complémentaire" $(1-X_t^*; t \geq 1)$ de ce processus.

Dans une première partie de ce chapitre nous étudions la structure du modèle défini par (1), d'abord les propriétés du processus vectoriel $(\vec{X}_t = (X_t^1, \dots, X_t^q); t \geq 1)$ des états dans le réseau pour un jour donné puis celles des séries chronologiques marginales. Dans une deuxième partie nous examinons le problème statistique de l'estimation des paramètres du modèle en distinguant selon que la station test est une station dans laquelle le phénomène est observable ou non, puis nous faisons une étude expérimentale par simulation des possibilités d'utilisation du modèle.

I. ETUDE DES CARACTERISTIQUES DU MODELE.

Nous utilisons les mêmes notations que celles du § I. du chapitre IV. Bien sûr, par construction, la série chronologique vectorielle $(\vec{X}_t; t \geq 1)$ est stationnaire.

I.1. Propriétés du processus vectoriel $(\vec{X}_t; t \geq 1)$.

Il est clair que le vecteur aléatoire \vec{X}_t est du type qui a été étudié dans le cas particulier 3 du § II.2. du chapitre I.

Compte tenu des résultats obtenus dans ce chapitre on a alors :

$$(2) \quad p(j) = \alpha_j p^* + (1-\alpha_j)(1-p^*) \quad ; \quad j=1, \dots, q$$

$$(3) \quad P[\vec{X}_t = x] = p^* \prod_{j=1}^q \alpha_j^{x_j} (1-\alpha_j)^{(1-x_j)} + (1-p^*) \prod_{j=1}^q \alpha_j^{(1-x_j)} (1-\alpha_j)^{x_j}$$

$$P[N_t = k] = p^* v(\vec{\alpha}, k) + (1-p^*) v(\vec{\alpha}, q-k) \quad ; \quad 0 \leq k \leq q,$$

$$E(N_t) = p^* \sum_{j=1}^n \alpha_j + (1-p^*) (q - \sum_{j=1}^n \alpha_j) \quad ;$$

$$(4) \quad \text{Var}(N_t) = \sum_{j=1}^q [\alpha_j p^* + (1-\alpha_j)(1-p^*)] [1-\alpha_j p^* - (1-\alpha_j)(1-p^*)] + 2 p^* (1-p^*) \sum_{i < j} (2\alpha_i - 1)(2\alpha_j - 1)$$

et

$$(5) \quad P[\vec{X}_t = x / N_t = k] = \begin{cases} \frac{p^* \prod_{j=1}^q \alpha_j^{x_j} (1-\alpha_j)^{1-x_j} + (1-p^*) \prod_{j=1}^q \alpha_j^{1-x_j} (1-\alpha_j)^{x_j}}{p^* v(\vec{\alpha}, k) + (1-p^*) v(\vec{\alpha}, q-k)} & \text{si } 1 \leq k \leq q \text{ et } |x| = k \\ 1 & \text{si } k=0 \text{ et } |x| = 0 \end{cases}$$

où v est défini par (8) § I. chapitre IV.

En ce qui concerne la structure d'ordre 2 du processus $(\vec{X}_t; t \geq 1)$ on obtient :

Proposition 1.

La fonction de covariance $(C(h); h \in \mathbb{N})$ s'écrit :

$$(6) \quad C_{jk}(h) = \begin{cases} (2\alpha_j - 1)(2\alpha_k - 1) C^*(h) & \text{si } j \neq k, h=0, 1, \dots \text{ ou } j=k, h=1, \dots \\ [\alpha_j p^* + (1-\alpha_j)(1-p^*)][1-\alpha_j p^* - (1-\alpha_j)(1-p^*)] & \text{si } j=k, h=0 \end{cases}$$

La loi conjointe des vecteurs aléatoires \vec{X}_{t+h} et \vec{X}_t , $h \geq 1$ est donnée par :

$$(7) \quad P[\vec{X}_{t+h}=y, \vec{X}_t=x] = \sum_{u,v=0}^1 \prod_{j=1}^q \alpha_j^{\theta(u,v,y_j,x_j)} (1-\alpha_j)^{2-\theta(u,v,y_j,x_j)} \times P[X_{h+1}^*=v, X_1^*=u] ; \quad x, y \in \{0, 1\}^q$$

où $\theta(v, u, y_j, x_j) = vy_j + (1-v)(1-y_j) + ux_j + (1-u)(1-x_j)$.

Preuve - L'égalité (6) est une conséquence immédiate de la définition 1.

On a bien sûr, pour x et y dans $\{0, 1\}^q$:

$$P[\vec{X}_{t+h}=y, \vec{X}_t=x] = \sum_{u,v=0}^1 P[\vec{X}_{t+h}=y, \vec{X}_t=x / X_{t+h}^*=v, X_t^*=u] \times P[X_{h+1}^*=v, X_1^*=u]$$

où

$$P[\vec{X}_{t+h}=y, \vec{X}_t=x / X_{t+h}^*=v, X_t^*=u] = \prod_{j=1}^q \alpha_j^{\theta(u,v,y_j,x_j)} \times (1-\alpha_j)^{2-\theta(u,v,y_j,x_j)}$$

avec la notation $\theta(u, v, y_j, x_j)$ de l'énoncé.

On en déduit alors l'égalité (7). \square

On peut naturellement déduire de la proposition 1 la loi conditionnelle de \vec{X}_{t+h} sachant \vec{X}_t de même que des résultats concernant le processus $(N_t; t \geq 1)$:

Corollaire 1.

La fonction de covariance de $(N_t; t \geq 1)$ s'écrit

$$(8) \quad \text{Cov}(N_{t+h}, N_t) = \sum_{j=1}^q (2\alpha_j - 1)^2 C^*(h) \quad , \quad h \geq 1 .$$

La loi conditionnelle de la variable aléatoire N_{t+h} sachant N_t est donnée pour $h \geq 1$ par :

$$(9) \quad P[N_{t+h} = \ell / N_t = k] = \frac{\sum_{u,v=0}^1 v(\vec{\alpha}, \omega(v, \ell)) v(\vec{\alpha}, \omega(u, k)) P[X_1^* = u, X_{h+1}^* = v]}{p^* v(\vec{\alpha}, k) + (1-p^*) v(\vec{\alpha}, q-k)}$$

où $\omega(v, \ell) = v\ell + (1-v)(q-\ell)$ et v est donnée par §I. chapitre IV.

Preuve - Pour $h \geq 1$, on a

$$\begin{aligned} \text{Cov}(N_{t+h}, N_t) &= \text{Cov}\left(\sum_{i=1}^q X_{t+h}^i, \sum_{j=1}^q X_t^j\right) \\ &= \sum_{i < j} C_{ij}(h) \end{aligned}$$

d'où la formule (8) d'après l'égalité (6) .

On a bien sûr, pour $h \geq 1$,

$$P[N_{t+h} = \ell / N_t = k] = \frac{P[N_{t+h} = \ell, N_t = k]}{P[N_t = k]}$$

où

$$\begin{aligned} P[N_{t+h} = \ell, N_t = k] &= \sum_{u,v=0}^1 P[N_{t+h} = \ell, N_t = k / X_t^* = u, X_{t+h}^* = v] \times \\ &\quad \times P[X_1^* = u, X_{h+1}^* = v] \end{aligned}$$

et

$$P[N_{t+h}=\ell, N_t=k/X_t^*=u, X_{t+h}^*=v] = P[\sum_{j=1}^q \epsilon_{t+h}^j = \omega(v, \ell), \sum_{j=1}^q \epsilon_t^j = \omega(u, k)]$$

avec la notation $\omega(u, k)$ de l'énoncé.

Alors avec la notation (8), §I. chapitre IV, on a

$$P[N_{t+h}=\ell, N_t=k] = \sum_{u, v=0}^1 v(\vec{\alpha}, \omega(v, \ell)) v(\vec{\alpha}, \omega(u, k)) P[X_1^*=u, X_{h+1}^*=v]$$

et compte tenu de (4) ci-dessus, on obtient (9) . \square

I.2. Propriétés des séries chronologiques marginales

$(X_t^j; t \geq 1), j=1, \dots, q.$

Dans ce paragraphe, pour simplifier les écritures nous omettrons l'indice j dans toutes les caractéristiques du processus

$$(X_t^j; t \geq 1); \text{ ainsi à } p(j) = P[X_t^j = 1] = \frac{m^1(j)}{m^0(j) + m^1(j)} = \alpha_j p^* + (1 - \alpha_j)(1 - p^*) \text{ nous substituerons}$$

$$p = \alpha p^* + (1 - \alpha)(1 - p^*) = \frac{m^1}{m^0 + m^1} \text{ où } m^i \text{ est substitué à } m^1(j) \text{ comme}$$

$p^i = (p_n^i; n \geq 1)$ sera substitué à $p^i(j) = (p_n^i(j); n \geq 1)$. D'une façon générale nous utiliserons les mêmes notations que celles du §I.2. du chapitre IV.

Les résultats concernant donc un processus $(X_t; t \geq 1)$ défini par

$$X_t = \epsilon_t X_t^* + (1 - \epsilon_t)(1 - X_t^*) \quad ; \quad t \geq 1$$

où

- $(X_t^*; t \geq 1)$ est une série chronologique binaire stationnaire

- $(\epsilon_t; t \geq 1)$ est une suite de variables aléatoires indépendantes de Bernoulli de même paramètre α

- Les suites $(X_t^*; t \geq 1)$ et $(\epsilon_t; t \geq 1)$ sont indépendantes entre elles.

Le processus $(X_t; t \geq 1)$ est bien sûr une série chronologique binaire stationnaire. Il est clair que pour tout n , le vecteur aléatoire $X_{(n)} = (X_1, \dots, X_n)$ est du type qui a été étudié dans le cas particulier 4 du §II.2 du chapitre I. Compte tenu des résultats obtenus dans ce chapitre on a alors :

$$(10) \quad P[X_{(n)}=x] = \sum_{z \in E_n} \alpha^{|\mathbf{x} \Delta z|} (1-\alpha)^{|\mathbf{x} \Delta z|} P[X_{(n)}^*=z]$$

$$(11) \quad C(h) = \text{Cov}(X_{t+h}, X_t) = \begin{cases} [\alpha p^* + (1-\alpha)(1-p^*)][1-\alpha p^* - (1-\alpha)(1-p^*)] & \text{si } h=0 \\ (2\alpha-1)^2 C^*(h) & \text{si } h \geq 1 \end{cases}$$

$$(12) \quad P[S_n=k] = \sum_{s=0}^n \sum_{i=0 \vee (k+s-n)}^{s \wedge k} C_s^i C_{n-s}^{k-i} \alpha^{q+2i-k-s} (1-\alpha)^{k+s-2i} \times P[S_n^*=s]$$

$$(13) \quad \begin{aligned} E(S_n) &= (2\alpha-1) E(S_n^*) + n(1-\alpha) \quad ; \\ \text{Var}(S_n) &= (2\alpha-1)^2 \text{Var}(S_n^*) + (1-\alpha)\alpha n \end{aligned}$$

où compte tenu de la stationnarité de $(X_t^*; t \geq 1)$,

$$\begin{aligned} E(S_n^*) &= np^* = n \frac{m^1(*)}{m^0(*) + m^1(*)} \quad ; \\ \text{Var}(S_n^*) &= n C^*(0) + 2 \sum_{h=1}^{n-1} (n-h) C^*(h) \end{aligned}$$

En fait on a aussi

$$(13') \quad E(S_n) = np = n \frac{m^1}{m^0 + m^1} \quad ; \quad \text{Var}(S_n) = n C(0) + 2 \sum_{h=1}^{n-1} (n-h) C(h)$$

où la suite $C(h)$ est donnée par (11).

Remarquons que, de la même façon que pour le modèle (1) du chapitre précédent, si $(X_t^*; t \geq 1) \sim \text{RA}(p^0(*), p^1(*))$ le processus $(X_t; t \geq 1)$ n'est pas lui même en général un processus de renouvellement

alterné; de même lorsque $(X_t^*; t \geq 1) \sim MA(1; p^*, \lambda^*)$, $(X_t; t \geq 1)$ n'est pas de Markov.

Nous nous intéressons maintenant aux caractéristiques des lois $p^i = (p_n^i; n \geq 1)$ des temps de séjour dans les deux états 0 et 1. On détermine d'abord les fonctions génératrices Q^0 et Q^1 en fonction des caractéristiques du processus $(X_t^*; t \geq 1)$.

Lemme 1.

On a, pour $i \in \{0, 1\}$

$$(14) \quad Q^i(t) = p^* \tilde{\delta}_*^1(t(1-\alpha)^i \alpha^{1-i}, (\frac{\alpha}{1-\alpha})^{2i-1}) + \\ + (1-p^*) \tilde{\delta}_*^0(t(1-\alpha)^i \alpha^{1-i}, (\frac{\alpha}{1-\alpha})^{2i-1}).$$

Preuve - D'après l'égalité (12) on a

$$Q_n^1 = \sum_{s=0}^n \alpha^s (1-\alpha)^{n-s} P[S_n^* = s]$$

d'où

$$Q^1(t) = \sum_{n \geq 1} \left\{ \sum_{s=0}^n \alpha^s (1-\alpha)^{n-s} P[S_n^* = s] \right\} t^n$$

$$\text{où } P[S_n = s] = p^* P[S_n^* = s/X_1^* = 1] + (1-p^*) P[S_n^* = s/X_1^* = 0]$$

Alors il vient

$$Q^1(t) = p^* \sum_{n \geq 1} \left\{ \sum_{s=0}^n (\frac{\alpha}{1-\alpha})^s P[S_n^* = s/X_1^* = 1] \right\} (t(1-\alpha))^n \\ + (1-p^*) \sum_{n \geq 1} \left\{ \sum_{s=0}^n (\frac{\alpha}{1-\alpha})^s P[S_n^* = s/X_1^* = 0] \right\} (t(1-\alpha))^n$$

d'où la formule (14) pour $i=1$.

De la même façon, comme, d'après l'égalité (12) on a

$$Q_n^0 = \sum_{s=0}^n \alpha^{n-s} (1-\alpha)^s P[S_n^* = s]$$

on montre la validité de (14) pour $i=0$. \square

Remarquons que, conformément à la remarque précédant l'énoncé du lemme 5, §IV, chapitre III, les quantités $\tilde{\delta}^j(t(1-\alpha)^i \alpha^{1-i}, (\frac{\alpha}{1-\alpha})^{2i-1})$ sont bien définies. Bien sûr pour $\alpha=1$ (resp. $\alpha=0$) on a $Q^i(t) = Q_*^i(t)$ (resp. $Q^i(t) = Q_*^{1-i}(t)$).

On calcule maintenant les moments m^i des lois p^i en fonction des moments $m^i(*)$ des lois $p^i(*)$:

Lemme 2.

On a, pour $i \in \{0,1\}$

$$(15) \quad m^i = \frac{\alpha m^i(*) + (1-\alpha)m^{1-i}(*)}{\alpha(1-\alpha)(m^0(*) + m^1(*) + (2\alpha-1)^2} .$$

Preuve - D'après (19) §I. du chapitre IV, on a :

$$(16) \quad m^i = \frac{Q_1^i}{Q_1^i - Q_2^i} \quad ; \quad i \in \{0,1\}$$

Or d'après la preuve du lemme 1 ci-dessus on a :

$$Q_1^1 = (1-\alpha) P[S_1^*=0] + \alpha P[S_1^*=1] = (1-\alpha) Q_1^0(*) + \alpha Q_1^1(*)$$

$$Q_2^1 = (1-\alpha)^2 P[S_2^*=0] + \alpha(1-\alpha) P[S_2^*=1] + \alpha^2 P[S_2^*=2]$$

$$= (1-\alpha)^2 Q_2^0(*) + \alpha(1-\alpha) \{1 - Q_2^0(*) - Q_2^1(*)\} + \alpha^2 Q_2^1(*)$$

où

$$Q_1^i(*) = \frac{m^i(*)}{m^0(*) + m^1(*)} \quad , \quad Q_2^i(*) = \frac{m^i(*) - 1}{m^0(*) + m^1(*)} \quad ; \quad i \in \{0,1\}$$

On en déduit alors aisément la valeur de m^1 .

$$m^1 = \frac{\alpha m^1(*) + (1-\alpha)m^0(*)}{\alpha(1-\alpha)(m^0(*)+m^1(*) + (2\alpha-1)^2}$$

De la même manière on montre que l'expression (15) est valable pour $i=0$. □

On est alors en mesure de déterminer les fonctions génératrices P^0 et P^1 . Rappelons (cf. (12) chapitre II. §I.) que la fonction génératrice P^1 de la loi p^1 s'exprime par

$$(17) \quad P^1(t) = (m^0+m^1) \frac{(1-t)^2}{t^2} Q^1(t) - m^1 \frac{1-t}{t} + 1.$$

On suppose désormais que $(X_t^*; t \geq 1) \sim RA(p^0(*), p^1(*))$.

Proposition 2.

La fonction génératrice P^0 est donnée par l'égalité (17) en y remplaçant $Q^0(t)$ par :

$$Q^0(t) = \frac{\delta_*^0(\alpha t, \frac{1-\alpha}{\alpha})}{(m^0(*)+m^1(*))\alpha t} \{ m^0(*) \tilde{P}_*^0(\alpha t) + m^1(*) P_*^0(\alpha t) \tilde{P}_*^1((1-\alpha)t) \} \\ + \frac{1}{m^0(*)+m^1(*)} \{ m^0(*)m^1(*) \frac{1}{\alpha t} \tilde{P}_*^0(\alpha t) \tilde{P}_*^1((1-\alpha)t) + \\ \frac{m^1(*)}{1-(1-\alpha)t} ((1-\alpha)t - \tilde{P}_*^1((1-\alpha)t)) + \frac{m^0(*)}{1-\alpha t} (t - \tilde{P}_*^0(\alpha t)) \}$$

où

$$(18) \quad \delta_*^0(t, x) = \frac{1}{1 - P_*(t) P_*(xt)} \{ m^0(*) \tilde{P}_*^0(t) P_*^1(xt) + \frac{m^1(*)}{x} \tilde{P}_*^1(xt) \}$$

L'expression de la fonction génératrice P^1 est donnée par l'égalité (17) en y remplaçant $Q^1(t)$ par l'expression $Q^0(t)$ définie ci-dessus dans laquelle α et $(1-\alpha)$ sont échangés.

Preuve - D'après les lemmes 5 et 6 du chapitre III, les fonctions génératrices $\tilde{\delta}_*^i$ intervenant dans l'égalité (14) sont fournies par

$$\tilde{\delta}_*^1(t, x) = \frac{1}{xt} \hat{P}_*^1(xt) \delta_*^1(t, x) + \frac{1}{1-xt} (xt - \hat{P}_*^1(xt))$$

$$\tilde{\delta}_*^0(t, x) = \frac{1}{t} \hat{P}_*^0(t) \delta_*^0(t, x) + \frac{1}{1-t} (t - \hat{P}_*^0(t))$$

où

$$\delta_*^1(t, x) = m^0(*) x \hat{P}_*^0(t) + x P_*^0(t) \delta_*^0(t, x) \quad (19)$$

$$\delta_*^0(t, x) = \frac{m^1(*)}{x} \hat{P}_*^1(xt) + \frac{1}{x} P_*^1(xt) \delta_*^1(t, x).$$

Reportant ces expressions dans (14) on obtient pour $i=0$:

$$\begin{aligned} Q^0(t) &= p^* \frac{1}{(1-\alpha)t} \hat{P}_*^1((1-\alpha)t) \left\{ m^0(*) \frac{1-\alpha}{\alpha} \hat{P}_*^0(\alpha t) + \frac{1-\alpha}{\alpha} P_*^0(\alpha t) \delta_*^0\left(\alpha t, \frac{1-\alpha}{\alpha}\right) \right\} \\ &\quad + p^* \frac{1}{1-(1-\alpha)t} ((1-\alpha)t - \hat{P}_*^1((1-\alpha)t)) + (1-p^*) \frac{1}{\alpha t} \hat{P}_*^0(\alpha t) \delta_*^0\left(\alpha t, \frac{1-\alpha}{\alpha}\right) \\ &\quad + (1-p^*) \frac{1}{1-\alpha t} (\alpha t - \hat{P}_*^0(\alpha t)) \\ &= \delta_*^0\left(\alpha t, \frac{1-\alpha}{\alpha}\right) \left\{ p^* \frac{1}{\alpha t} P_*^0(\alpha t) \hat{P}_*^1((1-\alpha)t) + (1-p^*) \frac{1}{\alpha t} \hat{P}_*^0(\alpha t) \right\} + \\ &\quad + p^* m^0(*) \frac{1}{\alpha t} \hat{P}_*^0(\alpha t) \hat{P}_*^1((1-\alpha)t) + p^* \frac{1}{1-(1-\alpha)t} ((1-\alpha)t - \hat{P}_*^1((1-\alpha)t)) \\ &\quad + (1-p^*) \frac{1}{1-\alpha t} (\alpha t - \hat{P}_*^0(\alpha t)) \\ &= \frac{\delta_*^0\left(\alpha t, \frac{1-\alpha}{\alpha}\right)}{(m^0(*) + m^1(*) \alpha t)} \left\{ m^0(*) \hat{P}_*^0(\alpha t) + m^1(*) P_*^0(\alpha t) \hat{P}_*^1((1-\alpha)t) \right\} \\ &\quad + \frac{1}{m^0(*) + m^1(*)} \left\{ m^0(*) m^1(*) \frac{1}{\alpha t} \hat{P}_*^0(\alpha t) \hat{P}_*^1((1-\alpha)t) + \right. \\ &\quad \left. \frac{m^1(*)}{1-(1-\alpha)t} ((1-\alpha)t - \hat{P}_*^1((1-\alpha)t)) + \frac{m^0(*)}{1-\alpha t} (\alpha t - \hat{P}_*^0(\alpha t)) \right\} \end{aligned}$$

où, d'après (19), $\delta_*^0(t,x)$ est donnée par la formule (18) de l'énoncé.

Cela assure le résultat annoncé pour $i=0$. Le résultat pour $i=1$ s'obtient en remplaçant dans la précédente preuve α par $(1-\alpha)$. \square

En ce qui concerne les moments d'ordre 2 des lois p^1 on déduit de la proposition 2 :

Corollaire 2.

On a

$$m_{[2]}^0 = \frac{2}{\alpha(1-\alpha)(m^0(*)+m^1(*))+(2\alpha-1)^2} \{ \gamma_*^0(\alpha)[m^0(*)\hat{P}_*(\alpha) + m^1(*)P_*^0(\alpha)\hat{P}_*(1-\alpha)] + m^0(*)m^1(*)\frac{1}{\alpha}\hat{P}_*(\alpha)P_*^1(1-\alpha) + \frac{m^1(*)}{\alpha}[(1-\alpha)^2 - \hat{P}_*(1-\alpha)] + \frac{m^0(*)}{1-\alpha}[\alpha^2 - \hat{P}_*(\alpha)] \}$$

$$\text{où } \gamma_*^0(\alpha) = \frac{1}{1 - \frac{0}{P_*(\alpha)} \frac{1}{P_*(1-\alpha)}} \{ \frac{m^0(*)}{\alpha} \hat{P}_*(\alpha) P_*^1(1-\alpha) + \frac{m^1(*)}{1-\alpha} \hat{P}_*(1-\alpha) \}$$

La valeur de $m_{[2]}^1$ est donnée par la même expression en changeant α en $(1-\alpha)$.

Preuve - D'après le lemme 5 du chapitre II §I.2. on a

$$m_{[2]}^i = 2(m^0+m^1) Q^i(1) - 2m^i$$

Il suffit alors de remplacer les valeurs des différents termes par celles obtenues précédemment: m^0 , m^1 données par (15) et $Q^0(1)$ et $Q^1(1)$ données par la proposition 2. \square

On dispose des approximations suivantes des moments d'ordre 1 et 2 des lois p^i , $i \in \{0,1\}$:

Corollaire 3.

On a, pour $i \in \{0,1\}$ et pour α voisin de 1, les approximations :

$$m^i \approx m^i(*) + (\alpha-1) \{ m^i(*) [m^0(*) + m^1(*) - 3] - m^{1-i}(*) \}$$

$$m_{[2]}^i \approx m_{[2]}^i(*) + (\alpha-1) \left\{ \frac{m_{[3]}^i(*)}{3} + m_{[2]}^i(*) [m^0(*) + m^1(*) - 2] \right. \\ \left. - 2m^i(*) [m^i(*) p_1^{1-i}(*) + 2] \right\} .$$

On a aussi, pour α voisin de 0, les approximations :

$$m^i \approx m^{1-i}(*) - \alpha \{ m^{1-i}(*) [m^0(*) + m^1(*) - 3] - m^i(*) \}$$

$$m_{[2]}^i \approx m_{[2]}^{1-i}(*) - \alpha \left\{ \frac{m_{[3]}^{1-i}(*)}{3} + m_{[2]}^{1-i}(*) [m^0(*) + m^1(*) - 2] \right. \\ \left. - 2m^{1-i}(*) [m^{1-i}(*) p_1^i(*) + 2] \right\} .$$

Preuve - On fait les vérifications pour $i=0$ et α voisin de 1. L'approximation de m^0 se déduit aisément de l'expression (15) et de ce que

$$(20) \quad \frac{1}{\alpha(1-\alpha)(m^0(*) + m^1(*) + (2\alpha-1)^2)} \approx 1 + (\alpha-1)[m^0(*) + m^1(*) - 4]$$

On obtient l'approximation de $m_{[2]}^0$ à partir du résultat du corollaire 2, et de l'approximation (20) et des approximations suivantes :

$$P_{*}^0(\alpha) \approx 1 + (\alpha-1)m^0(*)$$

$$P_{*}^1(1-\alpha) \approx -(\alpha-1)p_1^1(*)$$

$$\hat{P}_{*}^0(\alpha) \approx 1 + (\alpha-1) \frac{1}{m^1(*)} [m^0(*) + \frac{1}{2} m_{[2]}^0(*)] + \\ + \frac{(\alpha-1)^2}{2} \frac{1}{m^0(*)} [m_{[2]}^0(*) + \frac{1}{3} m_{[3]}^0(*)]$$

$$\hat{p}_*^{-1}(1-\alpha) \approx -(\alpha-1)\frac{1}{m^1(*)} + (\alpha-1)^2\frac{1}{m^1(*)}(1-p_1^1(*)).$$

Les autres approximations s'obtiennent de manière analogue. □

Exemples numériques.

(i) Plaçons nous dans le cas où $(X_t^*; t \geq 1) \sim MA(1; p^*, \lambda^*)$. On donne dans le tableau suivant quelques valeurs exactes et approchées des moments correspondant à certaines valeurs des paramètres p^* , λ^* et α :

α	1.	0.99	0.975	0.95	0.925	0.90
m^1 ex.	2.	2.006	2.016	2.030	2.044	2.056
m^1 app.	2.	2.006	2.016	2.033	2.050	2.066
m^0 ex.	1.33	1.348	1.372	1.411	1.450	1.489
m^0 app.	1.33	1.348	1.372	1.411	1.450	1.489
$m_{[2]}^1$ ex.	4.	4.045	4.112	4.215	4.307	4.390
$m_{[2]}^1$ app.	4.	4.046	4.116	4.233	4.350	4.466
$m_{[2]}^0$ ex.	.88	.945	1.032	1.180	1.332	1.488
$m_{[2]}^0$ app.	.88	.945	1.029	1.170	1.311	1.451

Tableau 1. $p^* = .6$, $\lambda^* = .5$

(ii) Plaçons nous maintenant dans le cas où $(X_t^*; t \geq 1) \sim RA(BNT(h_0, d_0, BNT(h_1, d_1)))$:

α	1.	0.99	0.975	0.95	0.925	0.90
m^1 ex.	4.33	4.107	3.819	3.438	3.147	2.917
m^1 app.	4.33	4.091	3.730	3.127	2.524	1.922
m^0 ex.	5.50	5.188	4.789	4.261	3.853	3.530
m^0 app.	5.50	5.167	4.668	3.837	3.006	2.175
$m^1_{[2]}$ ex.	19.99	18.398	16.368	13.709	11.693	10.124
$m^1_{[2]}$ app.	19.99	18.290	15.734	11.473	7.213	2.952
$m^0_{[2]}$ ex.	36.	32.745	28.641	23.339	19.384	16.352
$m^0_{[2]}$ app.	36.	32.493	27.234	18.469	9.703	0.938

Tableau 2. $\begin{cases} h_1 = 3.33 \\ d_1 = .66 \end{cases} \quad \begin{cases} h_0 = 4.50 \\ d_0 = 1.50 \end{cases}$

α	1.	0.99	0.975	0.95	0.925	0.90
m^1 ex.	4.	3.874	3.704	3.457	3.248	3.070
m^1 app.	4.	3.870	3.675	3.350	3.025	2.700
m^0 ex.	3.	2.923	2.818	2.669	2.545	2.440
m^0 app.	3.	2.920	2.80	2.60	2.40	2.20
$m^1_{[2]}$ ex.	30.	27.666	24.640	20.584	17.445	14.969
$m^1_{[2]}$ app.	30	27.523	23.807	17.615	11.423	5.231
$m^0_{[2]}$ ex.	16.	14.921	13.502	11.566	10.039	8.817
$m^0_{[2]}$ app.	16.	14.861	13.153	10.307	7.460	4.614

Tableau 3. $\begin{cases} h_1 = 3. \\ d_1 = 5. \end{cases} \quad \begin{cases} h_0 = 2. \\ d_0 = 4. \end{cases}$

II. ANALYSE STATISTIQUE ET ETUDE EXPERIMENTALE.

De la même façon qu'au chapitre IV. §II, nous examinons successivement les deux situations qui se présentent selon que la station test est une station privilégiée du réseau ou une station fictive non observable et nous menons une étude expérimentale par simulations des procédures d'estimation des paramètres du modèle. Nous renvoyons à notre conclusion pour quelques remarques concernant cette étude.

II.1. Cas d'une station test observable.

On observe donc une trajectoire du processus $(\vec{X}_t; t \geq 1)$ sur un intervalle de temps $\{1, \dots, n\}$: $\vec{X}_t = (X_t^1, \dots, X_t^q)$ pour $t \in \{1, \dots, n\}$ où

$$- (X_t^1; t \geq 1) \sim RA(p^0(*), p^1(*))$$

$$- X_t^j = \varepsilon_t^j X_t^1 + (1 - \varepsilon_t^j)(1 - X_t^1) \quad ; \quad j=2, \dots, q \quad ; \quad t \geq 1$$

- $(\varepsilon_t^j; t \geq 1)$ est une suite de variables aléatoires indépendantes de Bernoulli et de même paramètre α_j ; $j=2, \dots, q$

- les suites $(X_t^j; t \geq 1)$, $(\varepsilon_t^j; t \geq 1)$, $j=2, \dots, q$ sont indépendantes entre elles.

(a) Pour l'estimation des paramètres des lois $p^0(*)$ et $p^1(*)$ on dispose des mêmes procédures que celles décrites en II.1.a. du chapitre IV fondées sur la seule observation (X_1^1, \dots, X_n^1) .

(b) En ce qui concerne l'estimation des paramètres $\alpha_2, \dots, \alpha_q$, observant qu'on a encore $P[X_t^j = 1 / X_t^1 = 1] = \alpha_j$, on peut utiliser les estimateurs

$$\alpha_j(n) = \frac{s_n^j}{s_n^1} \quad ; \quad j=2, \dots, q .$$

On dispose, compte tenu de l'étude des caractéristiques du modèle menée dans le §I ci-dessus, des différentes possibilités évoquées dans le chapitre IV pour apprécier la qualité de l'ajustement correspondant à des observations.

On envisage maintenant une étude expérimentale dans le même esprit qu'en I.1 du chapitre IV.

Paramètres de la simulation : $X_t^1 \sim \text{RA}(\text{BNT}(h_1, d_1), \text{BNT}(h_0, d_0))$

avec $\begin{cases} h_1 = 5. \\ d_1 = 7. \end{cases}$ $\begin{cases} h_0 = 1. \\ d_0 = 1.5 \end{cases}$

et $\alpha_2 = .99$ $\alpha_3 = .97$ $\alpha_4 = .95$ $n = 1541$.

	h_1	d_1	h_0	d_0	α_2	α_3	α_4
val. d'entrée	5.	7.	1.	1.5	.99	.97	.95
val. estimées	4.635	6.717	1.04	1.402	.978	.965	.936

Tableau 4. Paramètres d'entrée et estimés.

Lois des temps de séjour.

station	1	2	3	4
m^1 ex.	6.	5.732	5.266	4.873
m^1 emp.	5.635	5.358	5.211	4.50
m^1 cal.	5.635	5.148	4.900	4.429

Tableau 5. Moments exacts, empiriques et calculés d'ordre 1 des lois des temps de séjour dans l'état 1.

station	1	2	3	4
m^0 ex.	2.	1.962	1.898	1.848
m^0 emp.	2.04	1.985	1.995	1.843
m^0 cal.	2.04	1.963	1.926	1.860

Tableau 6. idem tableau 5 : état 0.

station	1	2	3	4
$m_{[2]}^1$ ex.	70.	62.648	50.965	42.192
$m_{[2]}^1$ emp.	61.890	55.148	50.084	35.330
$m_{[2]}^1$ cal.	61.890	49.447	43.744	34.020

Tableau 7. Moments exacts, empiriques et calculés d'ordre 2 des lois des temps de séjour dans l'état 1.

station	1	2	3	4
$m_{[2]}^0$ ex.	4.50	4.282	3.912	3.614
$m_{[2]}^0$ emp.	4.62	4.248	4.356	3.297
$m_{[2]}^0$ cal.	4.62	4.190	3.979	3.594

Tableau 8. idem tableau 7 : état 0.

$N_{t+1} \backslash N_t$	0	1	2	3
0	0.4561 0.4545 0.4506	0.4338 0.4630 0.4229	0.1547 0.1589 0.1609	0.1521 0.1534 0.1569
1	0.0438 0.0689 0.0591	0.0418 0.0556 0.0557	0.0164 0.0093 0.0238	0.0161 0.0246 0.0233
2	0.0439 0.0579 0.0570	0.0459 0.0185 0.0604	0.0713 0.1028 0.0923	0.0716 0.0728 0.0927
3	0.4562 0.4215 0.4333	0.4785 0.4630 0.4610	0.7576 0.7290 0.7231	0.7602 0.7483 0.7270

Tableau 9. Probabilités de transition du processus stochastique N_t : exactes, empiriques et calculées .

Longueur de la séquence	Nbre. de séquences (N=209)	Probabilités empiriques	Probabilités théoriques exactes	Probabilités théoriques calculées
1	59	1.0000	1.0000	1.0000
2	23	0.7177	0.7685	0.7460
3	23	0.6077	0.6235	0.5923
4	15	0.4976	0.5149	0.4794
5	21	0.4258	0.4291	0.3921
6	14	0.3254	0.3598	0.3227
7	6	0.2584	0.3030	0.2668
8	6	0.2297	0.2560	0.2214
9	7	0.2010	0.2168	0.1842
10	3	0.1675	0.1839	0.1536
11	3	0.1531	0.1564	0.1283
12	10	0.1388	0.1331	0.1073
13	4	0.0909	0.1134	0.0898
14	1	0.0718	0.0968	0.0753
15	0	0.0670	0.0826	0.0632
16	3	0.0670	0.0706	0.0531
17	1	0.0526	0.0604	0.0446
18	1	0.0478	0.0517	0.0376
19	3	0.0431	0.0442	0.0316
20	2	0.0287	0.0379	0.0266
21	1	0.0191	0.0325	0.0224
22	1	0.0144	0.0279	0.0189
23	0	0.0096	0.0239	0.0160
24	0	0.0096	0.0205	0.0135
25	0	0.0096	0.0176	0.0114
26	0	0.0096	0.0151	0.0096
27	0	0.0096	0.0130	0.0081
28	0	0.0096	0.0112	0.0068
29	0	0.0096	0.0096	0.0058
30	0	0.0096	0.0082	0.0049
31	0	0.0096	0.0071	0.0041
32	0	0.0096	0.0061	0.0035

Tableau 10. Lois des temps de séjour dans l'état 1 : empirique , théorique exacte , théorique calculée ; $\alpha_2 = .99$, $\alpha_2(n) = .978$ (voir tableau 4 pour les autres caractéristiques).

Longueur de la séquence	Nbre. de séquences (N= 213)	Probabilités empiriques	Probabilités théoriques exactes	Probabilités théoriques calculées
1	54	1.0000	1.0000	1.0000
2	24	0.7465	0.7580	0.7395
3	25	0.6338	0.6060	0.5811
4	16	0.5164	0.4925	0.4652
5	21	0.4413	0.4037	0.3761
6	20	0.3427	0.3328	0.3059
7	13	0.2488	0.2754	0.2499
8	5	0.1878	0.2286	0.2048
9	5	0.1643	0.1902	0.1683
10	2	0.1408	0.1585	0.1386
11	4	0.1315	0.1323	0.1143
12	7	0.1127	0.1106	0.0944
13	5	0.0798	0.0925	0.0781
14	2	0.0563	0.0775	0.0646
15	1	0.0469	0.0650	0.0536
16	1	0.0423	0.0545	0.0444
17	1	0.0376	0.0457	0.0369
18	0	0.0329	0.0384	0.0306
19	1	0.0329	0.0323	0.0254
20	1	0.0282	0.0272	0.0211
21	1	0.0235	0.0228	0.0176
22	1	0.0188	0.0192	0.0146
23	0	0.0141	0.0162	0.0122
24	0	0.0141	0.0136	0.0101
25	0	0.0141	0.0115	0.0085
26	0	0.0141	0.0097	0.0070
27	0	0.0141	0.0082	0.0059
28	0	0.0141	0.0069	0.0049
29	1	0.0141	0.0058	0.0041
30	0	0.0094	0.0049	0.0034
31	0	0.0094	0.0041	0.0028
32	0	0.0094	0.0035	0.0024

Tableau 11. idem tableau 10 :

$$\alpha_3 = .97 \quad , \quad \alpha_3(n) = .965 \quad .$$

Longueur de la séquence	Nbre. de séquences (N= 242)	Probabilités empiriques	Probabilités théoriques exactes	Probabilités théoriques calculées
1	68	1.0000	1.0000	1.0000
2	34	0.7190	0.7473	0.7254
3	31	0.5785	0.5881	0.5571
4	19	0.4504	0.4701	0.4353
5	27	0.3719	0.3788	0.3432
6	18	0.2603	0.3068	0.2721
7	5	0.1860	0.2493	0.2166
8	9	0.1653	0.2032	0.1729
9	5	0.1281	0.1660	0.1383
10	2	0.1074	0.1358	0.1109
11	4	0.0992	0.1113	0.0890
12	5	0.0826	0.0913	0.0716
13	3	0.0620	0.0750	0.0576
14	0	0.0496	0.0616	0.0464
15	2	0.0496	0.0507	0.0374
16	1	0.0413	0.0417	0.0302
17	2	0.0372	0.0344	0.0244
18	0	0.0289	0.0283	0.0197
19	3	0.0289	0.0234	0.0159
20	2	0.0165	0.0193	0.0129
21	1	0.0083	0.0159	0.0104
22	0	0.0041	0.0131	0.0084
23	0	0.0041	0.0108	0.0068
24	0	0.0041	0.0090	0.0055
25	0	0.0041	0.0074	0.0045
26	0	0.0041	0.0061	0.0036
27	0	0.0041	0.0051	0.0029
28	1	0.0041	0.0042	0.0024
29	0	0.0000	0.0035	0.0019
30	0	0.0000	0.0029	0.0016
31	0	0.0000	0.0024	0.0013
32	0	0.0000	0.0020	0.0010

Tableau 12. idem tableau 10 :

$$\alpha_4 = .95 , \alpha_4(n) = .936 .$$

	N_t	S_{30}^2	S_{30}^3	S_{30}^4
exact	2.455	22.350	22.050	21.750
empirique	2.153	21.902	21.705	21.294
calculé	2.412	21.724	21.532	21.128

Tableau 13. Moments d'ordre 1 des statistiques N_t et S_{30}^j ; $j=2,3,4$.

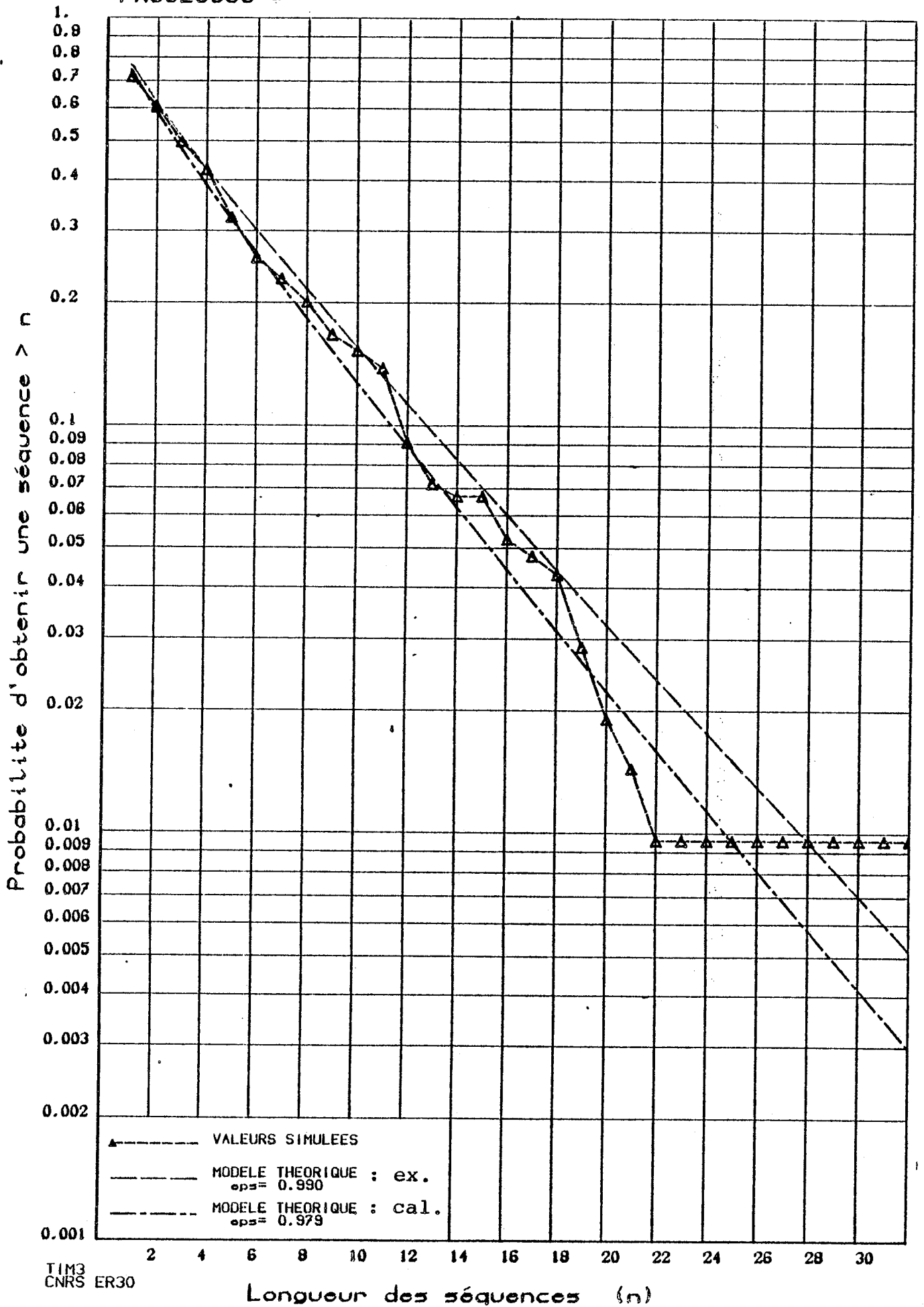
	N_t	S_{30}^2	S_{30}^3	S_{30}^4
exact	1.525	17.168	16.395	15.654
empirique	1.613	17.108	15.148	14.874
calculé	1.531	16.722	16.202	15.159

Tableau 14. idem tableau 13 : ordre 2.

ss-réseau	fréquences	prob. exactes	prob. emp.	prob. calculées
111	1017	0.6842	0.6600	0.6490
011	15	0.0073	0.0097	0.0148
101	31	0.0213	0.0201	0.0238
110	61	0.0361	0.0396	0.0445
001	26	0.0122	0.0169	0.0165
010	17	0.0074	0.0110	0.0095
100	11	0.0034	0.0071	0.0068
000	363	0.2281	0.2356	0.2350

Tableau 15. Probabilités pour qu'un sous-réseau soit sec : théoriques exactes, empiriques, théoriques calculées.

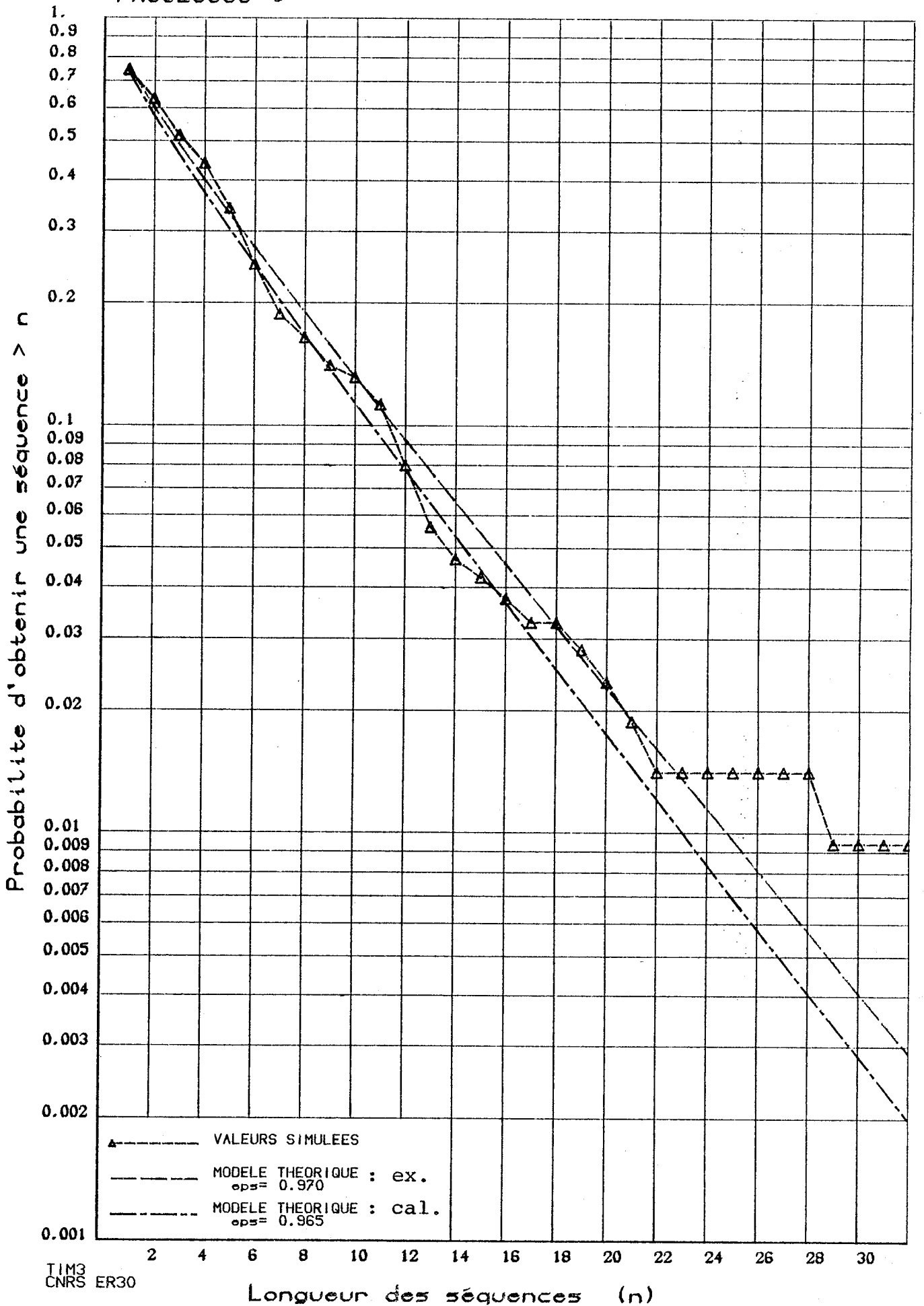
PROCESSUS 2



TIM3
CNRS ER30

Fig 1. Lois des temps de séjour des ...

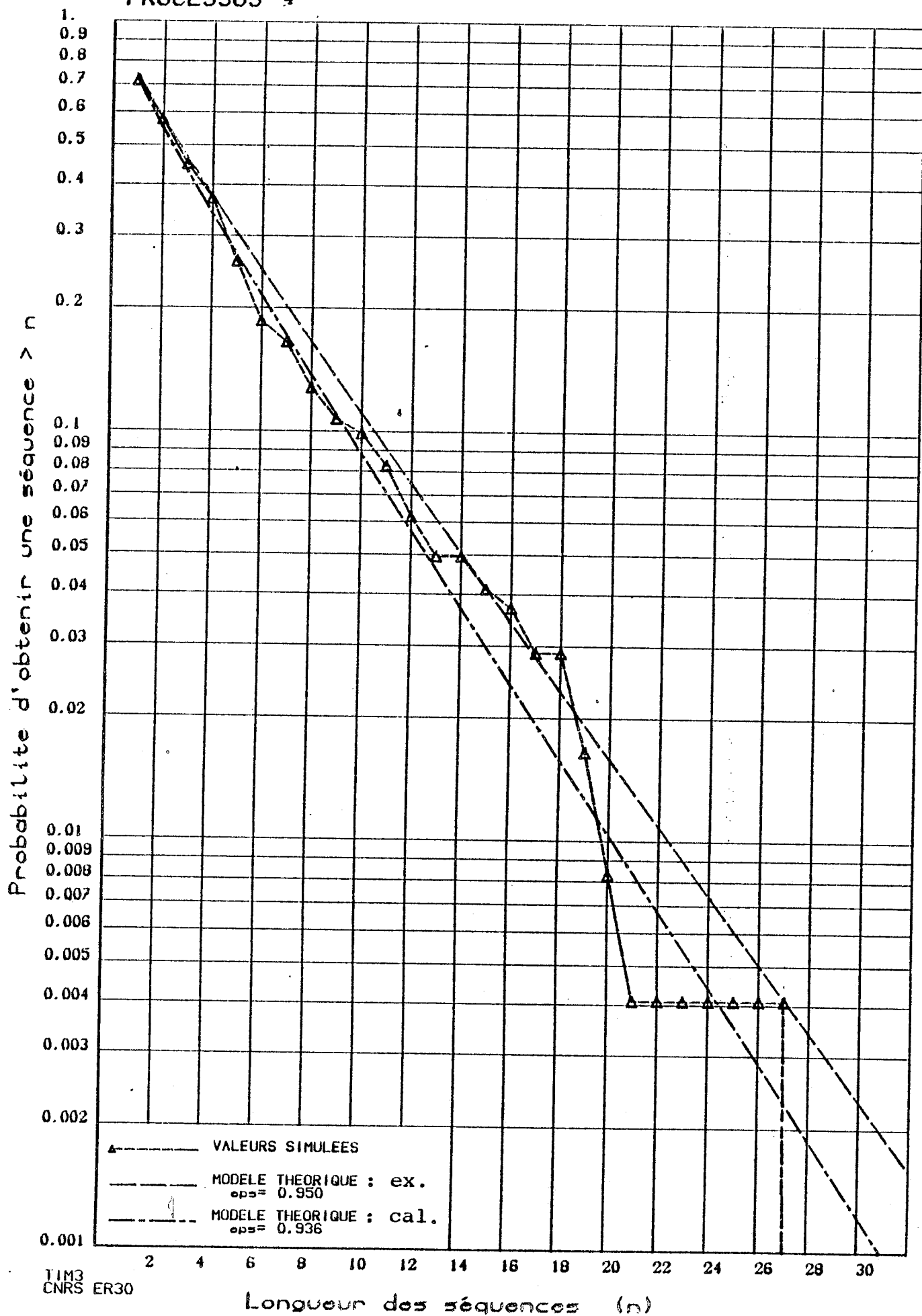
PROCESSUS 3



TIM3
CNRS ER30

Fig 2. idem figure 1.

PROCESSUS 4



TIM3
CNRS ER30

Fig 3. idem figure 1.

II.2. Cas d'une station test non observable.

On observe donc $\hat{X}_t = (X_t^1, \dots, X_t^q)'$ pour $t \in \{1, \dots, n\}$ où $X_t^j = \varepsilon_t^j X_t^* + (1 - \varepsilon_t^j)(1 - X_t^*)$; $j=1, \dots, q$, $(X_t^*; t \geq 1)$ et $(\varepsilon_t^j; t \geq 1)$, $j=1, \dots, q$ vérifient les conditions imposées dans le modèle (1).

On cherche à estimer au vu de cette seule observation les paramètres du processus $(X_t^*; t \geq 1)$ et $\alpha_1, \dots, \alpha_q$. Il convient de remarquer que le problème ainsi formulé est un problème mal posé; en effet il est facile de constater (cf. par exemple (10)) que la loi du processus $(X_t^*; t \geq 1)$ correspondant aux paramètres $(p^0(*), p^1(*), \alpha_1, \dots, \alpha_q)$ est la même que celle correspondant aux paramètres $(p^1(*), p^0(*), 1 - \alpha_1, \dots, 1 - \alpha_q)$.

On est donc amené à imposer certaines contraintes; compte tenu de l'objectif d'applications pour des réseaux climatologiques homogènes il est raisonnable de travailler sous l'hypothèse que les paramètres $\alpha_1, \dots, \alpha_q$ sont tous supérieurs à 1/2 (et même éventuellement tous voisins de 1). On est alors en mesure de proposer des procédures d'estimation de tous les paramètres.

(a) On peut d'abord envisager de faire l'estimation des paramètres $\alpha_1, \dots, \alpha_q$ et p^* et ce de deux manières différentes :

(a1) Comme pour tout $k \neq j$ on a

$$P[X_t^j = X_t^k] = \alpha_j \alpha_k + (1 - \alpha_j)(1 - \alpha_k)$$

et

$$\frac{\text{Cov}(X_t^j, X_t^k)}{\text{Cov}(X_{t+1}^j, X_t^j)} = \frac{2\alpha_k - 1}{2\alpha_j - 1}$$

on a aussi

$$\sum_{k \neq j} P[X_t^j = X_t^k] = \alpha_j \left(\sum_{k \neq j} \alpha_k \right) + (1 - \alpha_j) \left(\sum_{k \neq j} (1 - \alpha_k) \right)$$

et

$$\frac{\text{Cov}(X_{t+1}^j, N_t - X_t^j)}{\text{Cov}(X_{t+1}^j, X_t^j)} = \frac{2 \sum_{k \neq j} \alpha_k - (q-1)}{2\alpha_j - 1}$$

Alors, estimant $\sum_{k \neq j} P[X_t^j = X_t^k]$ par

$$r_j(n) = \frac{1}{n} \sum_{t=1}^n \sum_{k \neq j} [X_t^j X_t^k + (1-X_t^j)(1-X_t^k)] ,$$

$\text{Cov}(X_{t+1}^j, X_t^k)$ par

$$C_{j,k}(n) = \frac{1}{n-1} \sum_{t=1}^{n-1} X_{t+1}^j X_t^k - \frac{S_n^j}{n} \cdot \frac{S_n^k}{n} ,$$

et $\text{Cov}(X_{t+1}^j, N_t - X_t^j)$ par

$$C_{j\cdot}(n) = \sum_{k \neq j} C_{j,k}(n) ,$$

On est amené à construire un estimateur de α_j par résolution du système d'équations

$$(20) \quad \begin{cases} \alpha_j \sum_{k \neq j} \alpha_k + (1-\alpha_j) \left(\sum_{k \neq j} (1-\alpha_k) \right) = r_j(n) \\ \frac{2 \sum_{k \neq j} \alpha_k - (q-1)}{2\alpha_j - 1} = \frac{C_{j\cdot}(n)}{C_{jj}(n)} = \gamma_j(n) . \end{cases}$$

Posant $a_s = \alpha_s - 1/2$; $s=1, \dots, q$, le système (20) peut se réécrire sous la forme

$$\begin{cases} 2 a_j \sum_{k \neq j} a_k + (q-1)/2 = r_j(n) \\ \frac{\sum_{k \neq j} a_k}{a_j} = \gamma_j(n) \end{cases}$$

d'où l'on tire l'estimateur

$$a_j(n) = 1/2 \left(\frac{2r_j(n) - q + 1}{\gamma_j(n)} \right)^{1/2}$$

de a_j et l'estimateur

$$\alpha_j^1(n) = 1/2 \left\{ 1 + \left(\frac{2r_j(n) - q + 1}{\gamma_j(n)} \right)^{1/2} \right\}$$

de α_j .

Maintenant, comme $P[X_t^j = 1] = \alpha_j p^* + (1 - \alpha_j)(1 - p^*)$ pour tout $j=1, \dots, q$ on a aussi, posant $\pi^* = p^* - 1/2$,

$$P[X_t^j = 1] = 2a_j \pi^* + 1/2 \quad ; \quad j=1, \dots, q$$

puis

$$\pi^* = \frac{\sum_{j=1}^q (P[X_t^j = 1] - 1/2)}{\sum_{j=1}^q a_j}$$

Cela amène à proposer l'estimateur

$$\pi^*(n) = \frac{1}{2 \sum_{j=1}^q a_j(n)} \left(\frac{1}{n} \sum_{t=1}^n N_t - q/2 \right)$$

de π^* et donc l'estimateur

$$p_1^*(n) = 1/2 \left\{ 1 + \frac{1}{\sum_{j=1}^q a_j(n)} \left[\frac{1}{n} \sum_{t=1}^n N_t - q/2 \right] \right\}$$

de p^* .

(a2) On peut aussi procéder d'une autre façon. On constate que :

$$P[X_t^1 = X_t^2 = \dots = X_t^q = 1] = p^* \left[\prod_{j=1}^q \alpha_j \right] + (1 - p^*) \left[\prod_{j=1}^q (1 - \alpha_j) \right]$$

et

$$P[X_t^1 = X_t^2 = \dots = X_t^q = 0] = (1 - p^*) \left[\prod_{j=1}^q \alpha_j \right] + p^* \left[\prod_{j=1}^q (1 - \alpha_j) \right]$$

Alors, lorsque les paramètres α_j , $j=1, \dots, q$ sont voisins de 1, on a les approximations suivantes :

$$P[X_t^1 = \dots = X_t^q = 1] \approx p^* \left[\prod_{j=1}^q \alpha_j \right]$$

$$P[X_t^1 = \dots = X_t^q = 0] \approx (1-p^*) \left[\prod_{j=1}^q \alpha_j \right]$$

d'où l'on tire

$$p^* \approx \frac{P[X_t^1 = \dots = X_t^q = 1]}{P[X_t^1 = \dots = X_t^q]}$$

Utilisant d'autre part le fait que

$$E(X_t^j) = \alpha_j (2p^* - 1) - p^* + 1$$

et donc

$$\alpha_j = \frac{E(X_t^j) + p^* - 1}{2p^* - 1}$$

on est amené à proposer les estimateurs $p_2^*(n)$ de p^* et $\alpha_j^2(n)$ de α_j définis par :

$$p_2^*(n) = \frac{\sum_{t=1}^n \left(\prod_{k=1}^q X_t^k \right)}{\sum_{t=1}^n \left[\prod_{k=1}^q X_t^k + \prod_{k=1}^q (1-X_t^k) \right]}$$

et

$$\alpha_j^2(n) = \frac{\frac{1}{n} \sum_{t=1}^n X_t^j + p_2^*(n) - 1}{2 p_2^*(n) - 1}$$

(b) On peut maintenant procéder à l'estimation des paramètres du processus $(X_t^j; t \geq 1)$. Puisque d'après (15) ci-dessus on a, pour tout $j=1, \dots, q$,

$$m^i(j) = \frac{(2\alpha_j - 1) [\alpha_j m^1(j) - (1-\alpha_j) m^{1-1}(j)]}{1 - \alpha_j (1-\alpha_j) (m^0(j) + m^1(j))} ; \quad i \in \{0, 1\},$$

on a aussi

$$m^i(*) = \frac{\sum_{j=1}^q (2\alpha_j - 1) [\alpha_j m^i(j) - (1-\alpha_j) m^{1-i}(j)]}{q - \sum_{j=1}^q \alpha_j (1-\alpha_j) [m^0(j) + m^1(j)]} ; i \in \{0, 1\}$$

Cela suggère l'estimateur de $m^i(*)$ défini par

$$m^i(*, n) = \frac{\sum_{j=1}^q (2\alpha_j(n) - 1) [\alpha_j(n) m^i(j, n) - (1-\alpha_j(n)) m^{1-i}(j, n)]}{q - \sum_{j=1}^q \alpha_j(n) (1-\alpha_j(n)) [m^0(j, n) + m^1(j, n)]}$$

$i \in \{0, 1\}$,

où $m^i(j, n)$ est la moyenne empirique des séquences de type i débutant dans la station j sur la période $\{1, \dots, n\}$ et $\alpha_j(n)$ est l'un des deux estimateurs proposés ci-dessus de α_j . Cela fournit bien sûr un autre estimateur de p^* :

$$\frac{m^1(*, n)}{m^0(*, n) + m^1(*, n)}$$

Lorsque les lois $p^0(*)$ et $p^1(*)$ dépendent chacune de deux paramètres il reste à proposer des estimateurs des moments factoriels d'ordre 2 $m_{[2]}^i(*)$, $i \in \{0, 1\}$. Pour ce faire on procède comme au II.2.b. du chapitre IV, à partir d'approximations de ces moments en fonction des correspondants pour les processus $(X_t^j; t \geq 1)$ qui s'obtiennent comme celles du corollaire 3 ci-dessus. On est alors conduit aux estimateurs :

$$m_{[2]}^i(*, n) = \frac{1}{q} \sum_{j=1}^q \{ m_{[2]}^i(j, n) + (1-\alpha_j(n)) [\frac{1}{3} m_{[3]}^i(n) + m_{[2]}^i(j, n) (m^0(j, n) + m^1(j, n) - 2) - 2 m^i(j, n) (m^i(j, n) p^{1-i}(j, n) + 2)] \}$$

où $m_{[k]}^i(j, n)$ est le moment factoriel empirique d'ordre k des séquences de type i débutant sur la période $\{1, \dots, n\}$ dans la

station j et $p_1^i(j,n)$ est l'estimateur naturel de $p_1^i(j)$ défini par le rapport entre le nombre de séquences de type i de durée l jour observées dans la station j et le nombre total de séquences de ce type ayant débuté sur la période $\{1, \dots, n\}$ dans la dite station.

On envisage maintenant une étude expérimentale dans le même esprit qu'en I.1. On reprend les mêmes exemples.

Paramètres de la simulation : $X_t^* \sim \text{RA}(\text{BNT}(h_1^*, d_1^*), \text{BNT}(h_0^*, d_0^*))$

avec $\begin{cases} h_1^* = 5. \\ d_1^* = 7. \end{cases}$ $\begin{cases} h_0^* = 1. \\ d_0^* = 1.5 \end{cases}$

et $\alpha_1 = .99$ $\alpha_2 = .97$ $\alpha_3 = .95$ $n = 1541$.

	h_1^*	d_1^*	h_0^*	d_0^*	α_1	α_2	α_3
val. d'entrée	5.	7.	1.	1.5	.99	.97	.95
val. estimés	4.689	6.508	1.037	1.283	.978	.965	.936

Tableau 4. Paramètres d'entrée et estimés .

Lois des temps de séjour.

station	1	2	3	4
m^1 ex.	6.	5.732	5.266	4.873
m^1 emp.	5.635	5.358	5.211	4.500
m^1 cal.	5.689	5.192	4.921	4.459

Tableau 5. Moments exacts, empiriques et calculés d'ordre 1 des lois des temps de séjour dans l'état 1.

station	1	2	3	4
m^0 ex.	2.	1.962	1.898	1.848
m^0 emp.	2.04	1.985	1.995	1.843
m^0 cal.	2.037	1.960	1.920	1.856

Tableau 6. idem tableau 5 : état 0.

station	1	2	3	4
$m_{[2]}^1$ ex.	70.	62.648	50.965	42.192
$m_{[2]}^1$ emp.	61.89	55.148	50.084	35.330
$m_{[2]}^1$ cal.	61.88	49.504	43.420	34.107

Tableau 7. Moments exacts, empiriques et calculés d'ordre 2 des lois des temps de séjour dans l'état 1.

station	1	2	3	4
$m_{[2]}^0$ ex.	4.50	4.282	3.912	3.614
$m_{[2]}^0$ emp.	4.62	4.248	4.356	3.297
$m_{[2]}^0$ cal.	4.479	4.070	3.854	3.502

Tableau 8. idem tableau 7 : état 0.

$N_{t+1} \backslash N_t$	0	1	2	3
0	0.4561 0.4545 0.4500	0.4338 0.4630 0.4220	0.1547 0.1589 0.1593	0.1521 0.1534 0.1554
1	0.0438 0.0689 0.0590	0.0418 0.0556 0.0556	0.0164 0.0093 0.0236	0.0161 0.0246 0.0231
2	0.0439 0.0579 0.0571	0.0459 0.0185 0.0605	0.0713 0.1028 0.0925	0.0716 0.0728 0.0929
3	0.4562 0.4215 0.4339	0.4785 0.4630 0.4620	0.7576 0.7290 0.7246	0.7602 0.7483 0.7285

Tableau 9. Probabilités de transition du processus stochastique N_t : exactes, empiriques et calculées.

Longueur de la séquence	Nbre. de séquences (N=209)	Probabilités empiriques	Probabilités théoriques exactes	Probabilités théoriques calculées
1	59	1.0000	1.0000	1.0000
2	23	0.7177	0.7685	0.7550
3	23	0.6077	0.6235	0.6015
4	15	0.4976	0.5149	0.4875
5	21	0.4258	0.4291	0.3987
6	14	0.3254	0.3598	0.3280
7	6	0.2584	0.3030	0.2709
8	6	0.2297	0.2560	0.2245
9	7	0.2010	0.2168	0.1864
10	3	0.1675	0.1839	0.1551
11	3	0.1531	0.1564	0.1293
12	10	0.1388	0.1331	0.1079
13	4	0.0909	0.1134	0.0902
14	1	0.0718	0.0968	0.0754
15	0	0.0670	0.0826	0.0631
16	3	0.0670	0.0706	0.0529
17	1	0.0526	0.0604	0.0443
18	1	0.0478	0.0517	0.0372
19	3	0.0431	0.0442	0.0312
20	2	0.0287	0.0379	0.0262
21	1	0.0191	0.0325	0.0220
22	1	0.0144	0.0279	0.0185
23	0	0.0096	0.0239	0.0156
24	0	0.0096	0.0205	0.0131
25	0	0.0096	0.0176	0.0110
26	0	0.0096	0.0151	0.0093
27	0	0.0096	0.0130	0.0078
28	0	0.0096	0.0112	0.0066
29	0	0.0096	0.0096	0.0055
30	0	0.0096	0.0082	0.0047
31	0	0.0096	0.0071	0.0039
32	0	0.0096	0.0061	0.0033

Tableau 10. Lois des temps de séjour dans l'état 1 : empirique , théorique exacte théorique calculée ; $\alpha_1 = .99$, $\alpha_1(n) = .978$ (voir tableau 4 pour les autres caractéristiques).

Longueur de la séquence	Nbre. de séquences (N=213)	Probabilités empiriques	Probabilités théoriques exactes	Probabilités théoriques calculées
1	54	1.0000	1.0000	1.0000
2	24	0.7465	0.7580	0.7478
3	25	0.6338	0.6060	0.5895
4	16	0.5164	0.4925	0.4725
5	21	0.4413	0.4037	0.3819
6	20	0.3427	0.3328	0.3105
7	13	0.2488	0.2754	0.2534
8	5	0.1878	0.2286	0.2074
9	5	0.1643	0.1902	0.1701
10	2	0.1408	0.1585	0.1398
11	4	0.1315	0.1323	0.1151
12	7	0.1127	0.1106	0.0948
13	5	0.0798	0.0925	0.0782
14	2	0.0563	0.0775	0.0646
15	1	0.0469	0.0650	0.0534
16	1	0.0423	0.0545	0.0442
17	1	0.0376	0.0457	0.0366
18	0	0.0329	0.0384	0.0303
19	1	0.0329	0.0323	0.0251
20	1	0.0282	0.0272	0.0208
21	1	0.0235	0.0228	0.0173
22	1	0.0188	0.0192	0.0143
23	0	0.0141	0.0162	0.0119
24	0	0.0141	0.0136	0.0099
25	0	0.0141	0.0115	0.0082
26	0	0.0141	0.0097	0.0068
27	0	0.0141	0.0082	0.0057
28	0	0.0141	0.0069	0.0047
29	1	0.0141	0.0058	0.0039
30	0	0.0094	0.0049	0.0033
31	0	0.0094	0.0041	0.0027
32	0	0.0094	0.0035	0.0023

Tableau 11. idem tableau 10 :

$$\alpha_2 = .97 \quad , \quad \alpha_2(n) = .965 \quad .$$

Longueur de la séquence	Nbre. de séquences (N=242)	Probabilités empiriques	Probabilités théoriques exactes	Probabilités théoriques calculées
1	68	1.0000	1.0000	1.0000
2	34	0.7190	0.7473	0.7323
3	31	0.5785	0.5881	0.5640
4	19	0.4504	0.4701	0.4411
5	27	0.3719	0.3788	0.3477
6	18	0.2603	0.3068	0.2755
7	5	0.1860	0.2493	0.2190
8	9	0.1653	0.2032	0.1746
9	5	0.1281	0.1660	0.1394
10	2	0.1074	0.1358	0.1116
11	4	0.0992	0.1113	0.0894
12	5	0.0826	0.0913	0.0717
13	3	0.0620	0.0750	0.0576
14	0	0.0496	0.0616	0.0463
15	2	0.0496	0.0507	0.0372
16	1	0.0413	0.0417	0.0299
17	2	0.0372	0.0344	0.0241
18	0	0.0289	0.0283	0.0194
19	3	0.0289	0.0234	0.0157
20	2	0.0165	0.0193	0.0126
21	1	0.0083	0.0159	0.0102
22	0	0.0041	0.0131	0.0082
23	0	0.0041	0.0108	0.0066
24	0	0.0041	0.0090	0.0054
25	0	0.0041	0.0074	0.0043
26	0	0.0041	0.0061	0.0035
27	0	0.0041	0.0051	0.0028
28	1	0.0041	0.0042	0.0023
29	0	0.0000	0.0035	0.0018
30	0	0.0000	0.0029	0.0015
31	0	0.0000	0.0024	0.0012
32	0	0.0000	0.0020	0.0010

Tableau 12. idem tableau 10 : .

$$\alpha_3 = .95 \quad , \quad \alpha_3(n) = .936.$$

	N_t	S_{30}^1	S_{30}^2	S_{30}^3
exact	2.455	22.350	22.050	21.750
empirique	2.153	21.902	21.705	21.294
calculé	2.415	21.786	21.592	21.184

Tableau 13. Moments d'ordre 1 des statistiques N_t et S_{30}^j ; $j=1,2,3$.

	N_t	S_{30}^1	S_{30}^2	S_{30}^3
exact	1.525	17.171	16.397	15.656
empirique	1.613	17.108	15.148	14.874
calculé	1.523	16.725	16.204	15.161

Tableau 14. idem tableau 13 : ordre 2.

ss-réseau	fréquences	prob. exactes	prob. emp.	prob. calculées
111	1017	0.6842	0.6600	0.6509
011	15	0.0073	0.0097	0.0148
101	31	0.0213	0.0201	0.0239
110	61	0.0361	0.0396	0.0446
001	26	0.0122	0.0169	0.0164
010	17	0.0074	0.0110	0.0095
100	11	0.0034	0.0071	0.0067
000	363	0.2281	0.2356	0.2331

Tableau 15. Probabilités pour qu'un sous-réseau soit sec : théoriques exactes, empiriques, théoriques calculées

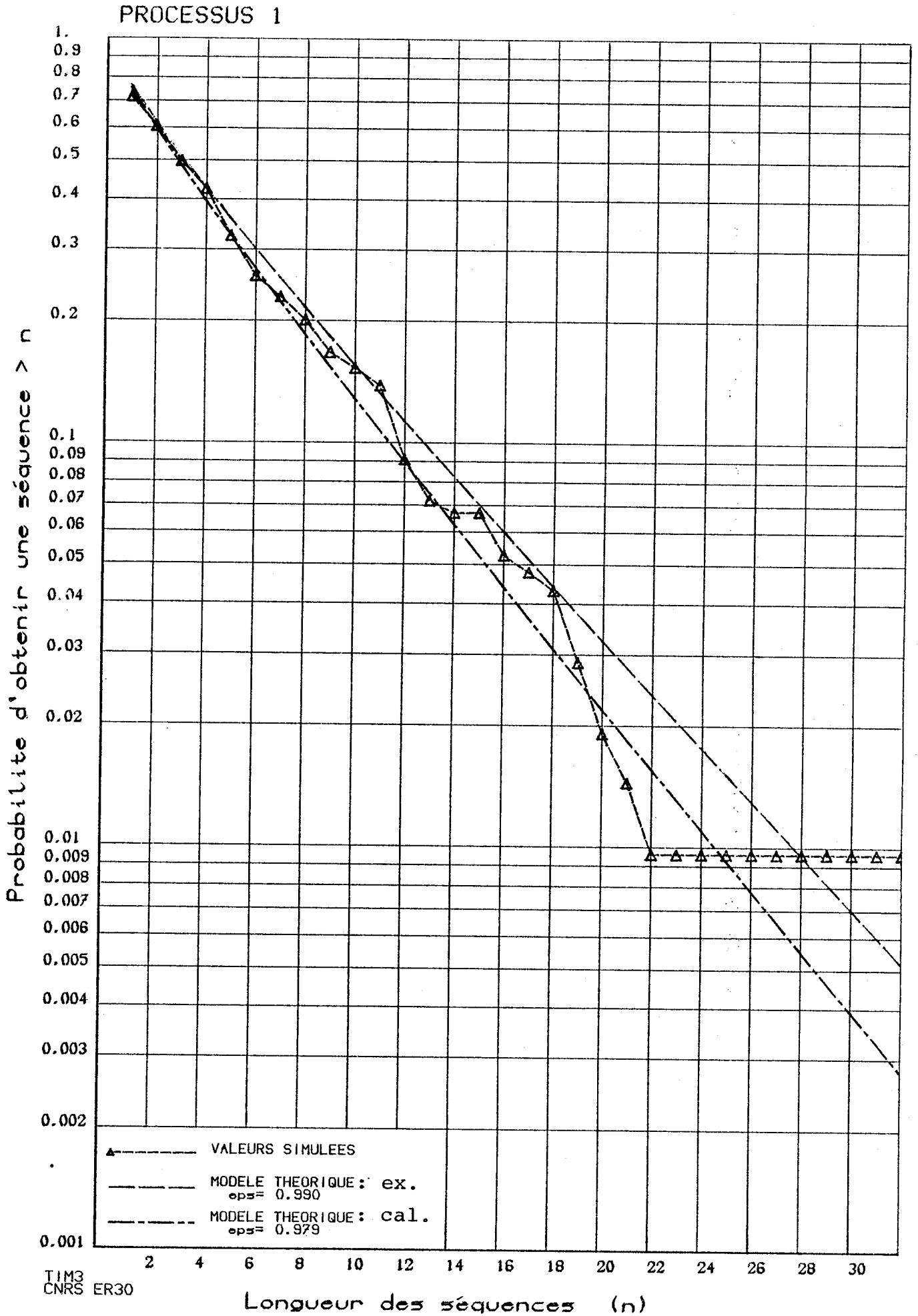
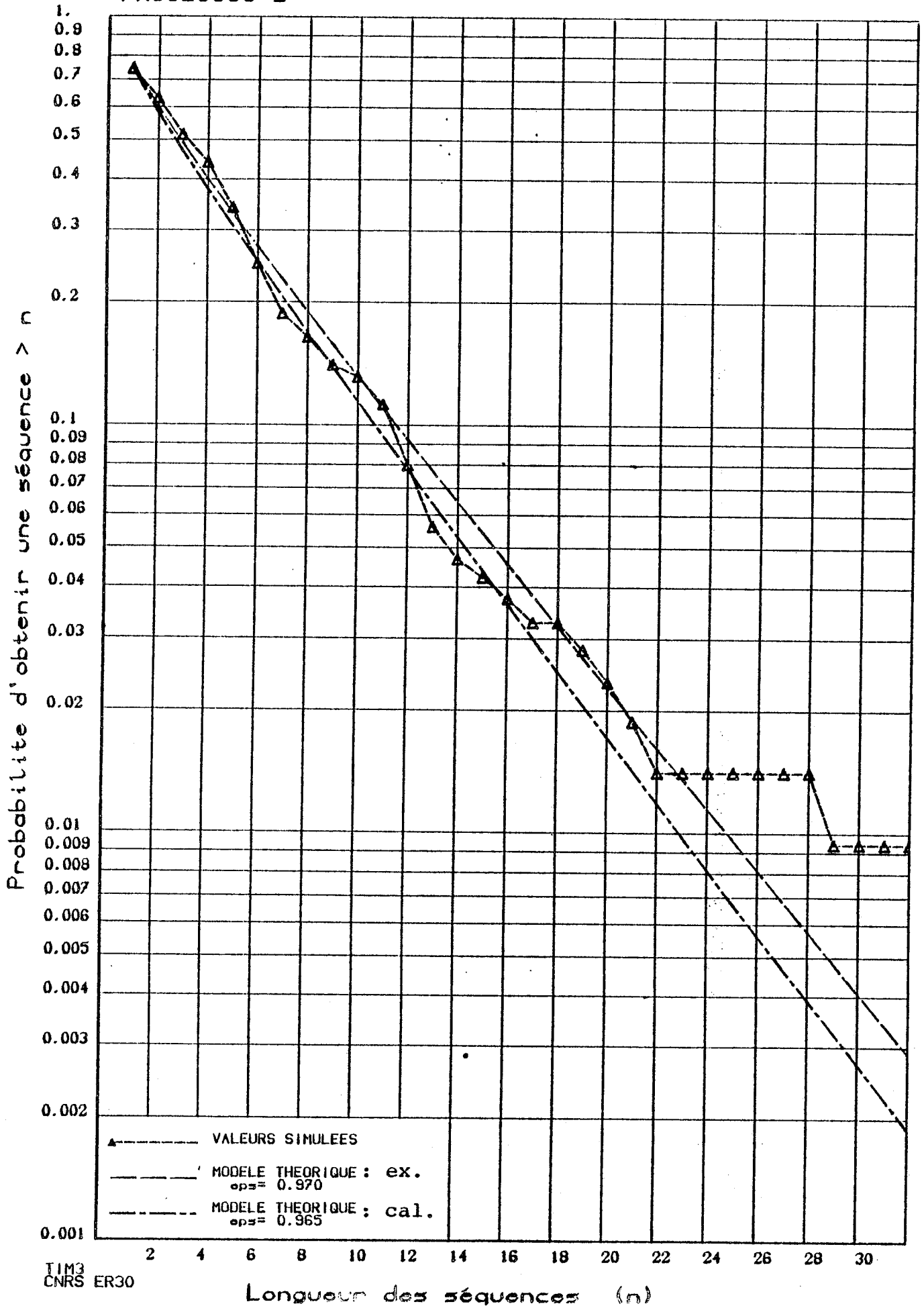


Fig 4. Lois des temps de séjour dans l'état 1

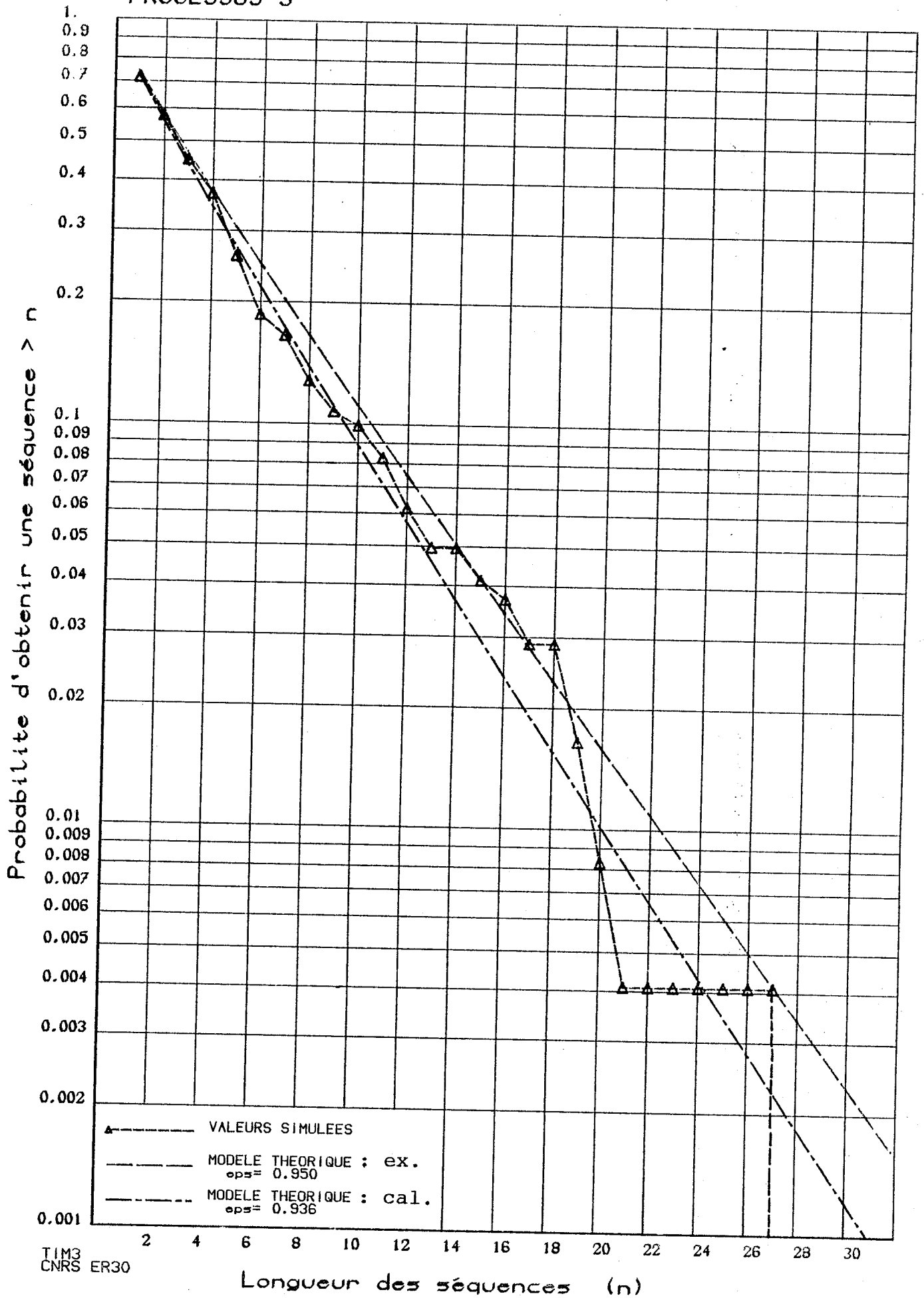
PROCESSUS 2



TIM3
CNRS ER30

Fig 5. idem figure 4.

PROCESSUS 3

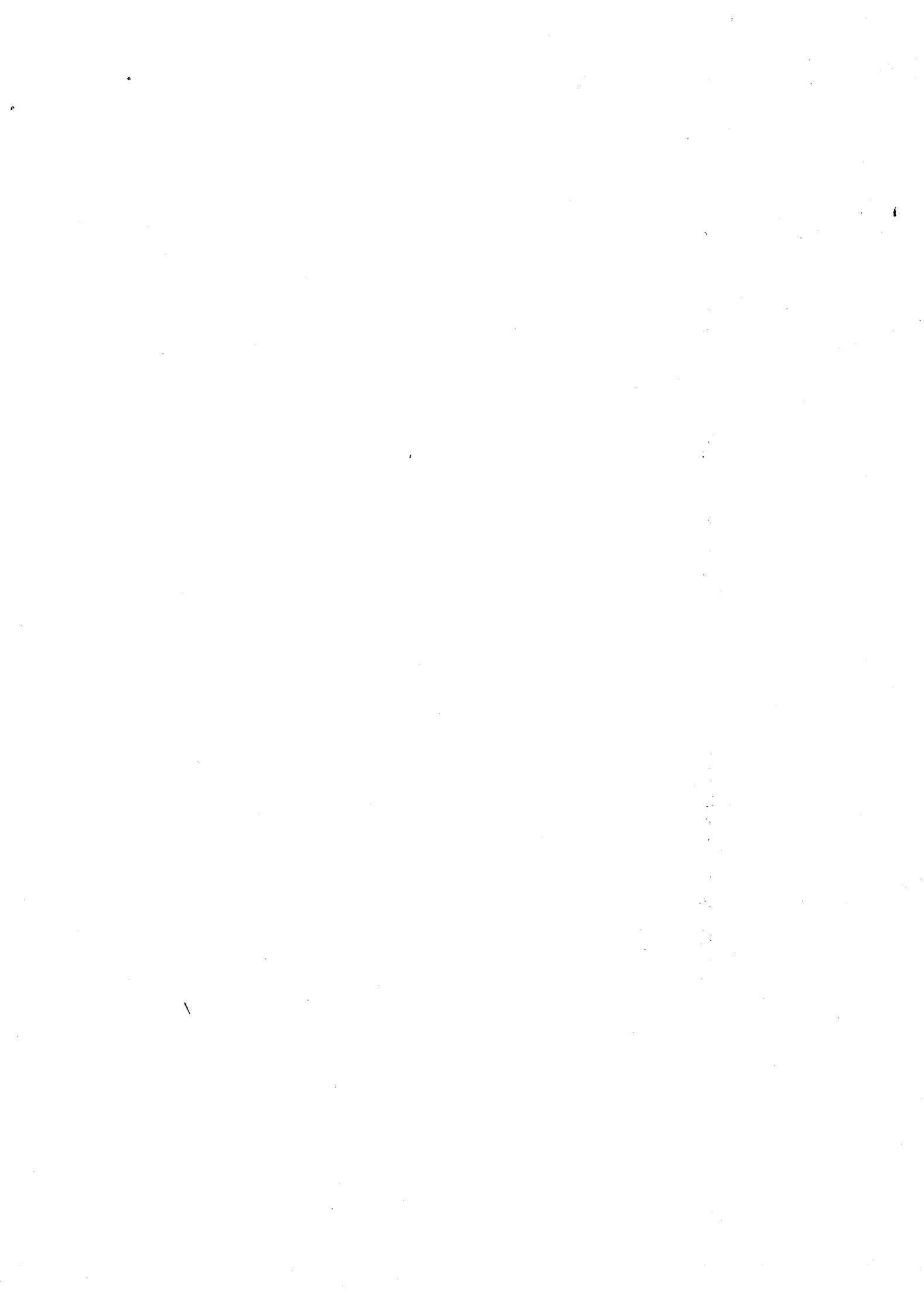


TIM3
CNRS ER30

Fig 6. idem figure 4.

CHAPITRE 6 :

APPLICATION A LA MODELISATION DU PHENOMENE CLIMATOLOGIQUE



Dans les applications nous cherchons à ajuster un modèle de séries chronologiques (resp. vectorielles à composantes) binaires d'une certaine classe de modèles à des données climatologiques concernant une station (resp. un réseau de stations) et une période $\{1, \dots, T\}$ (mois, ensemble de mois, ...) de l'année pour plusieurs années consécutives. Si on fait l'hypothèse que pour N années d'observation, le phénomène se reproduit, sur cette période, d'une année à l'autre, dans des conditions d'homogénéité et d'indépendance, on dispose de données $((x_1(u), \dots, x_T(u)); u=1, \dots, N)$ (resp. $(\vec{x}_1(u), \dots, \vec{x}_T(u)); u=1, \dots, N)$) dont l'étude s'inscrit dans le cadre de la statistique classique sur un échantillon indépendant de l'élément aléatoire (X_1, \dots, X_T) (resp. $(\vec{X}_1, \dots, \vec{X}_T)$) la famille des lois de probabilité définissant la structure statistique étant celle des lois T -dimensionnelles correspondant aux lois de processus associées aux différents modèles de la classe considérée. Nous nous plaçons ici plutôt dans la perspective où à partir des observations faites sur plusieurs années, on construit une suite (x_1, \dots, x_n) (resp. $(\vec{x}_1, \dots, \vec{x}_n)$), où n est de l'ordre de $N \times T$, dont on admet qu'elle est l'observation d'une trajectoire d'une série chronologique obéissant sur une longue trajectoire aux "règles" valables pour le phénomène sur la période $\{1, \dots, T\}$. Dans le cas unidimensionnel (i.e. d'une station) le procédé d'usage courant (qui se justifie pleinement sous l'hypothèse de renouvellement) pour construire une telle suite consiste à extraire de toutes les données concernant l'ensemble des années les séquences climatologiques sèches ou humides débutant dans la période considérée et à les mettre bout à bout. Dans le cas vectoriel où on s'oriente vers l'ajustement d'un des modèles définis dans les chapitres IV et V nous verrons comment bâtir une telle suite $(\vec{x}_1, \dots, \vec{x}_n)$. L'étude des données relève alors de l'analyse des séries chronologiques telle qu'elle a été envisagée dans les chapitres précédents.

Nous faisons maintenant quelques essais de modélisation. Nous renvoyons à notre conclusion pour quelques remarques concernant ces essais.

I. ESSAI DE MODELISATION 1.

Nous nous intéressons ici au réseau de stations de mesures du Sud-Est de la France constitué de :

MARSEILLE (1), MONTPELLIER (2), MONTELMAR (3), TOULON (4).

Nous nous proposons de modéliser le phénomène climatique pour la période de l'année Septembre-Octobre au vu des observations concernant les années 1953-1977.

Nous nous orientons vers l'utilisation du modèle défini dans le chapitre IV en considérant que MARSEILLE est une station test observée. Il est clair qu'en toute rigueur l'évolution du phénomène à MARSEILLE n'est pas un "majorant" pour les évolutions du réseau dans les trois autres stations du réseau; toutefois les écarts de la réalité climatique à la contrainte du modèle semblent pouvoir être négligés.

En ce qui concerne le calcul des caractéristiques des différentes lois empiriques de temps de séjour nous avons utilisé pour chaque station les observations des longueurs de séquences ayant débuté et s'étant terminées dans la période Septembre-Octobre pour la dite station (cf. Tableaux 2 et 3).

Pour la détermination des caractéristiques empiriques (cf. Tableaux 4 à 7) liées aux processus $((X_t^2, X_t^3, X_t^4); t \geq 1)$ et $(N_t; t \geq 1)$ ainsi qu'aux variables S_{30}^2 , S_{30}^3 et S_{30}^4 nous sommes partis des trajectoires vectorielles dont les marginales sont bâties par mise bout à bout, année après année, des trajectoires unidimensionnelles observées dans les stations pour la période Septembre-Octobre; ainsi $n = 1525$. Il aurait sans doute été préférable (cf. la discussion au début de ce chapitre) de retenir les données correspondant à des jours "appartenant" à des séquences ayant débuté dans la période pour la station 1 et s'étant éventuellement prolongées au delà d'octobre. L'organisation des fichiers de données devra être revue pour des essais en ce sens.

Les valeurs dites calculées (cf. Tableaux 2 à 7) ont été obtenues à l'aide des expressions exactes déterminées pour le modèle, à partir des valeurs estimées des paramètres. Les valeurs dites calculées BNT (cf. Tableaux 6 et 7) sont celles auxquelles conduit une modélisation station par station par un processus de renouvellement alterné basé sur des lois BNT.

paramètres	h_1	d_1	h_0	d_0	α_2	α_3	α_4
val. estimées	6.327	7.528	1.017	0.525	.985	.971	.981

Tableau 1. Valeurs estimées des paramètres du modèle.

station	MARSEILLE	MONTPELLIER	MONTELMAR	TOULON
m^1 emp.	7.327	6.688	6.197	6.539
m^1 cal.	7.327	6.691	6.191	6.540
m^0 emp.	2.017	1.807	2.014	1.873
m^0 cal.	2.017	1.972	1.940	1.962

Tableau 2. Moments d'ordre 1 : empiriques et calculés .

station	MARSEILLE	MONTPELLIER	MONTELMAR	TOULON
$m_{[2]}^1$ emp.	100.327	85.392	60.831	72.793
$m_{[2]}^1$ cal.	100.327	81.568	68.281	77.421
$m_{[2]}^0$ emp.	3.603	3.155	4.056	3.142
$m_{[2]}^0$ cal.	3.603	3.476	3.390	3.448

Tableau 3. idem tableau 2 : ordre 2.

$N_{t+1} \backslash N_t$	0	1	2	3
0	0.3016 0.5041	0.1349 0.1366	0.1313 0.1366	0.0438 0.1366
1	0.1825 0.0047	0.2222 0.0082	0.1313 0.0082	0.0467 0.0082
2	0.1667 0.0779	0.2460 0.1357	0.2222 0.1357	0.0971 0.1357
3	0.3492 0.4133	0.3968 0.7195	0.5152 0.7195	0.8124 0.7195

Tableau 4. Probabilités de transition du processus N_t : empiriques et calculées.

ss-réseau	fréquences	prob. empiriques	prob. calculées
111	1050	0.7000	0.6534
011	63	0.0420	0.0392
101	91	0.0607	0.0566
110	44	0.0293	0.0274
001	49	0.0327	0.0034
010	24	0.0160	0.0016
100	53	0.0353	0.0024
000	126	0.0840	0.2160

Tableau 5. Probabilités pour qu'un sous-réseau soit sec : empiriques et calculées.

variable	N_t	S_{30}^2	S_{30}^3	S_{30}^4
empirique	2.448	24.76	23.62	25.06
calculé BNT	2.319	23.618	22.641	23.320
calculé	2.303	23.169	22.859	23.076

Tableau 6. Moments d'ordre 1 des statistiques N_t et S_{30}^j ; $j=2,3,4$: empiriques, calculés théoriques et calculés BNT pour S_{30}^j .

variable	N_t	S_{30}^2	S_{30}^3	S_{30}^4
empirique	0.919	9.142	13.195	10.336
calculé BNT	0.525	11.276	10.482	9.712
calculé	1.460	12.736	12.702	12.726

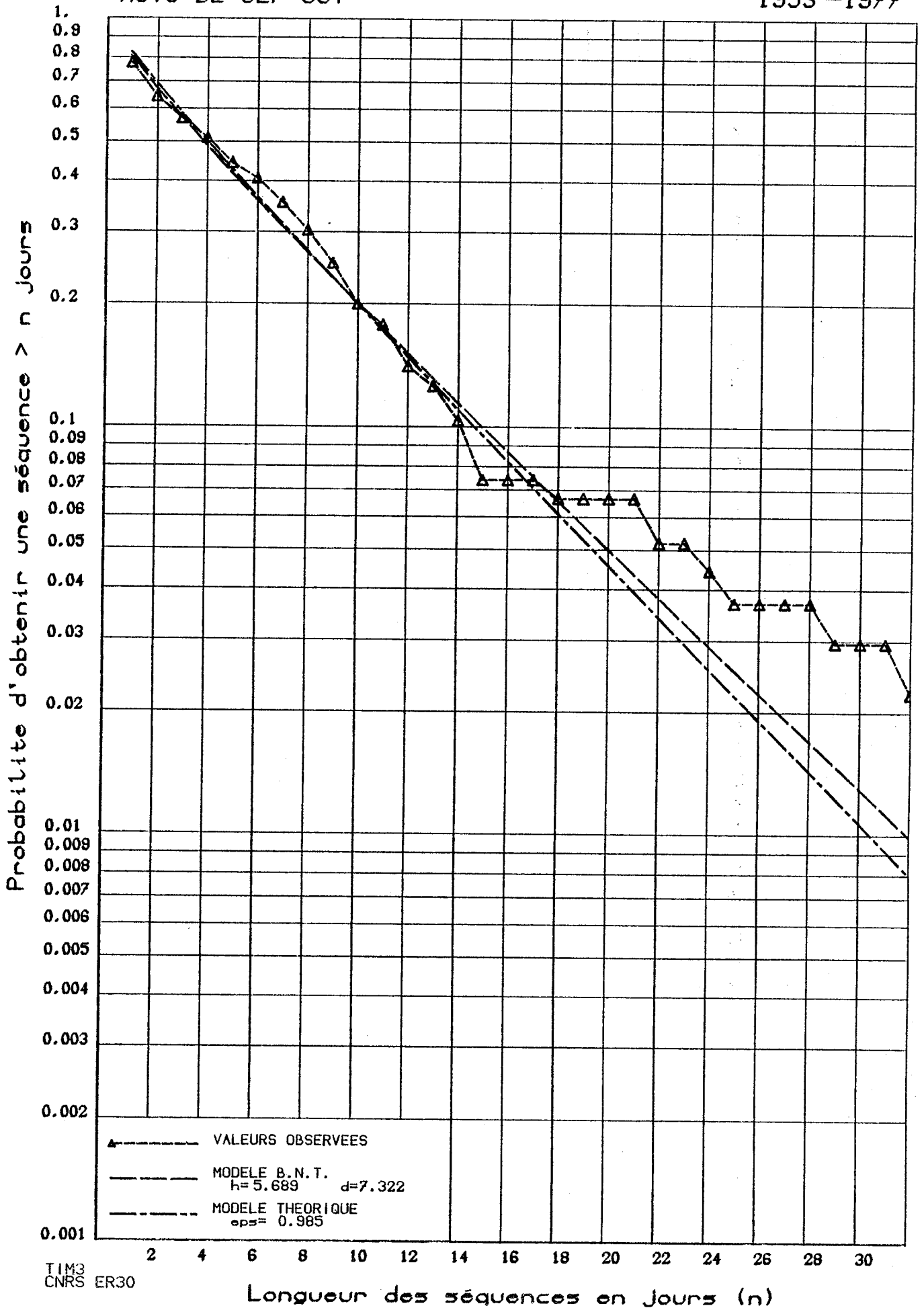
Tableau 7. idem tableau 6 : ordre 2.

MONTPELLIER

- 162 -

MOIS DE SEP-OCT

1953-1977



TIM3
CNRS ER30

Fig 1. Probabilités d'obtenir une séquence sèche > n jours.

MONTELMAR

- 163 -

MOIS DE SEP-OCT

1953-1977

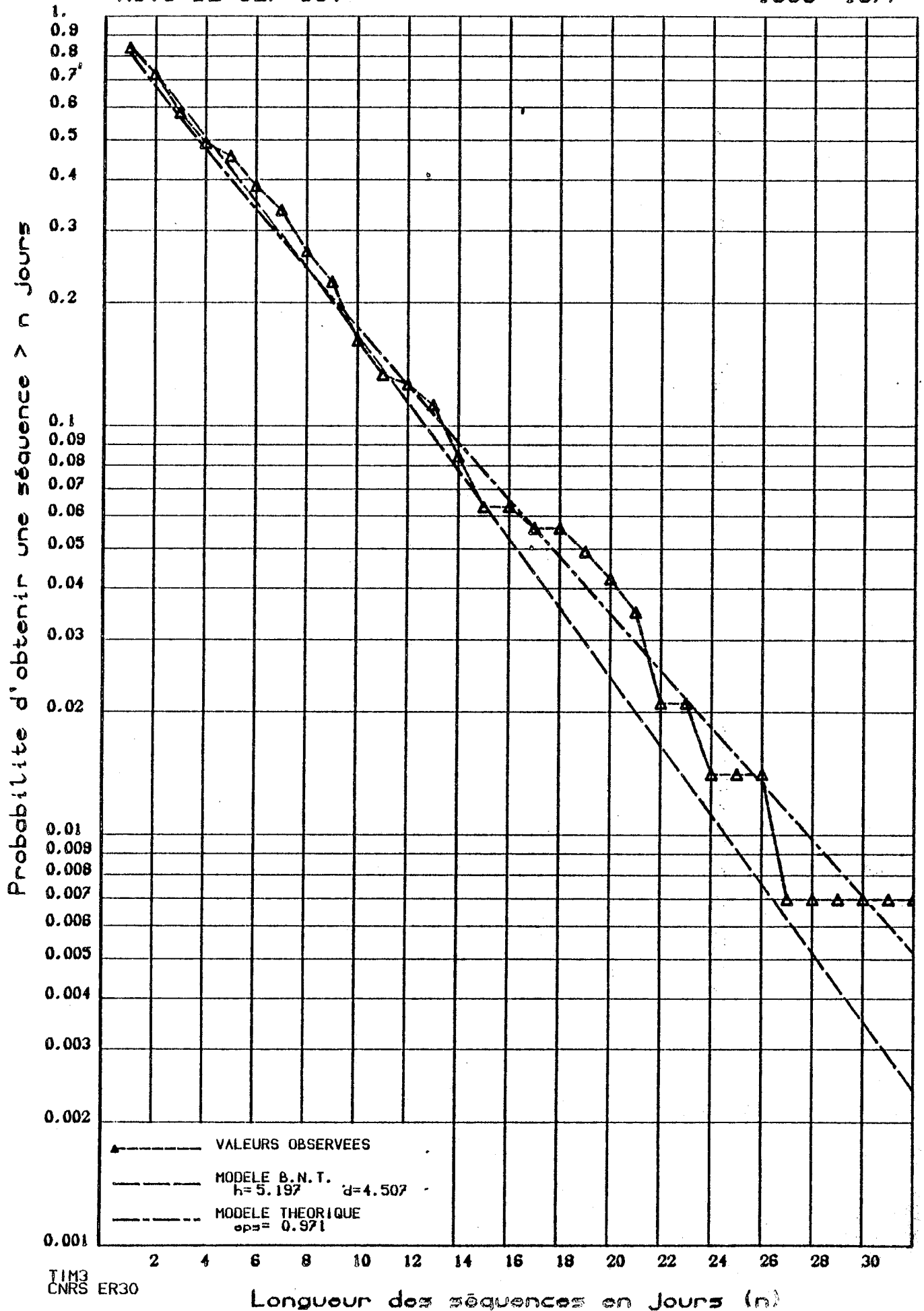


Fig 2. Probabilités d'obtenir une séquence sèche > n jours

TOULON

- 164 -

MOIS DE SEP-OCT

1953-1977

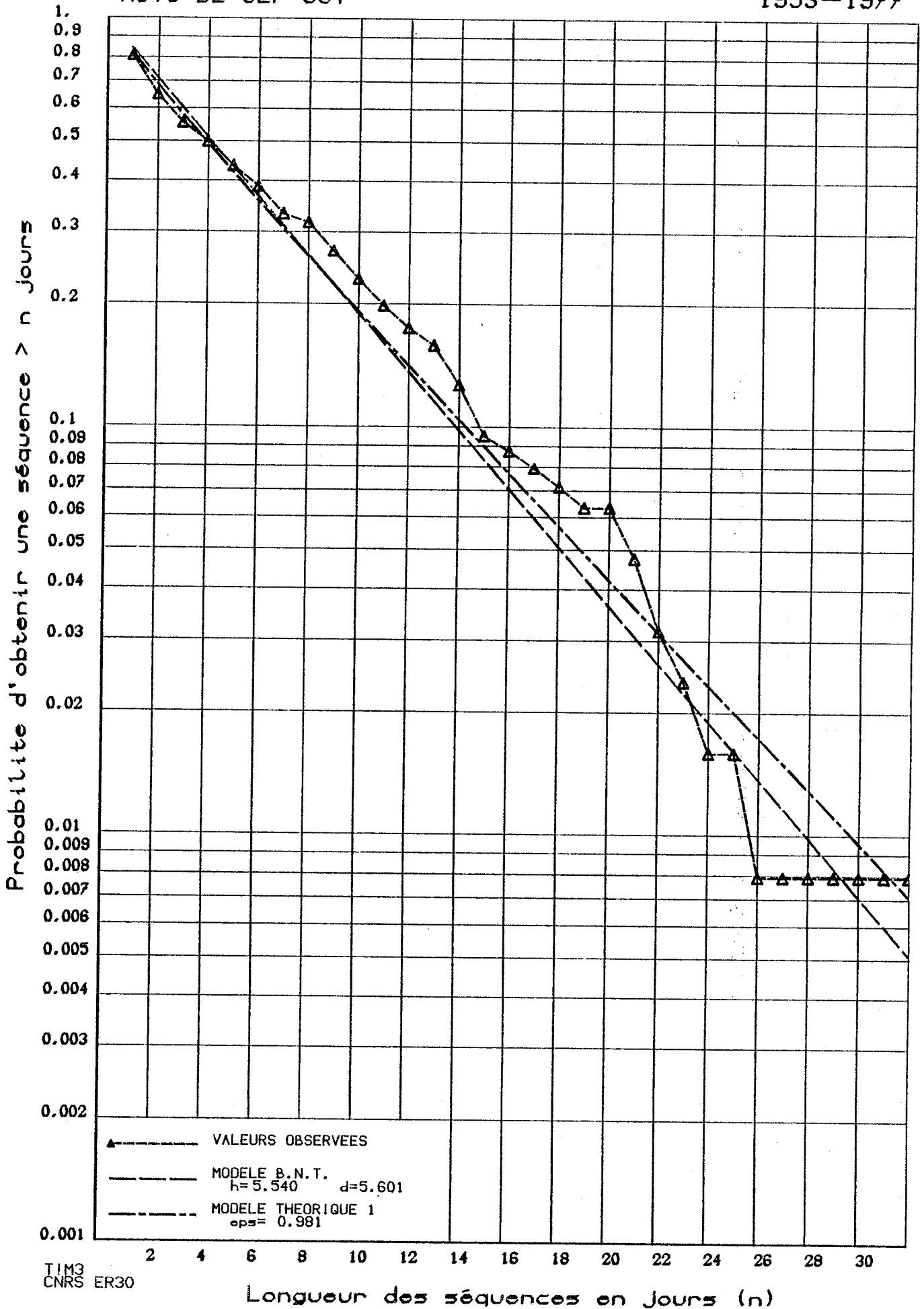


Fig 3. Probabilités d'obtenir une séquence sèche > n jours.

II. ESSAI DE MODELISATION 2.

Nous reprenons le même réseau de stations que celui considéré dans le précédent essai mais pour la période Novembre-Décembre. Nous envisageons encore l'utilisation du modèle défini dans le chapitre IV, MARSEILLE étant toujours considérée comme station privilégiée.

paramètres	h_1	d_1	h_0	d_0	α_2	α_3	α_4
val. estimées	5.819	7.918	0.881	1.322	.995	.948	.977

Tableau 8. Valeurs estimées des paramètres du modèle.

station	MARSEILLE	MONTPELLIER	MONTELMAR	TOULON
m^1 emp.	6.819	6.654	5.241	6.012
m^1 cal.	6.819	6.626	5.234	5.983
m^0 emp.	1.881	1.753	1.903	1.888
m^0 cal.	1.881	1.870	1.810	1.838

Tableau 9. Moments d'ordre 1 : empiriques et calculés.

station	MARSEILLE	MONTPELLIER	MONTELMAR	TOULON
$m_{[2]}^1$ emp.	91.583	84.985	53.903	71.254
$m_{[2]}^1$ cal.	91.583	85.476	48.649	66.946
$m_{[2]}^0$ emp.	3.708	2.774	3.505	4.00
$m_{[2]}^0$ cal.	3.708	3.660	3.426	3.537

Tableau 10. idem tableau 9 : ordre 2.

$N_{t+1} \backslash N_t$	0	1	2	3
0	0.2905 0.4684	0.1688 0.1469	0.1096 0.1469	0.0567 0.1469
1	0.1892 0.0074	0.1818 0.0118	0.1491 0.0118	0.0649 0.0118
2	0.1892 0.0998	0.2403 0.1601	0.2018 0.1601	0.1206 0.1601
3	0.3311 0.4244	0.4091 0.6812	0.5395 0.6812	0.7577 0.6812

Tableau 11. Probabilités de transition du processus N_t : empiriques et calculées.

ss-réseau	fréquences	prob. empiriques	prob. calculées
111	970	0.6467	0.6256
011	50	0.0333	0.0322
101	111	0.0740	0.0716
110	67	0.0447	0.0432
001	43	0.0287	0.0037
010	30	0.0200	0.0022
100	81	0.0540	0.0049
000	148	0.0987	0.2165

Tableau 12. Probabilités pour qu'un sous-réseau soit sec : empiriques et calculées.

variable	N_t	S_{30}^2	S_{30}^3	S_{30}^4
empirique	2.347	24.58	22.34	23.48
calculé BNT	2.286	23.743	22.009	22.830
calculé	2.289	23.411	22.295	22.969

Tableau 13. Moments d'ordre 1 des statistiques N_t et S_{30}^j ; $j=2,3,4$: empiriques, calculés théoriques et calculés BNT pour S_{30}^j .

variable	N_t	S_{30}^2	S_{30}^3	S_{30}^4
empirique	1.023	6.0436	8.584	7.049
calculé BNT	0.542	10.595	13.470	14.026
calculé	1.446	5.965	6.472	6.175

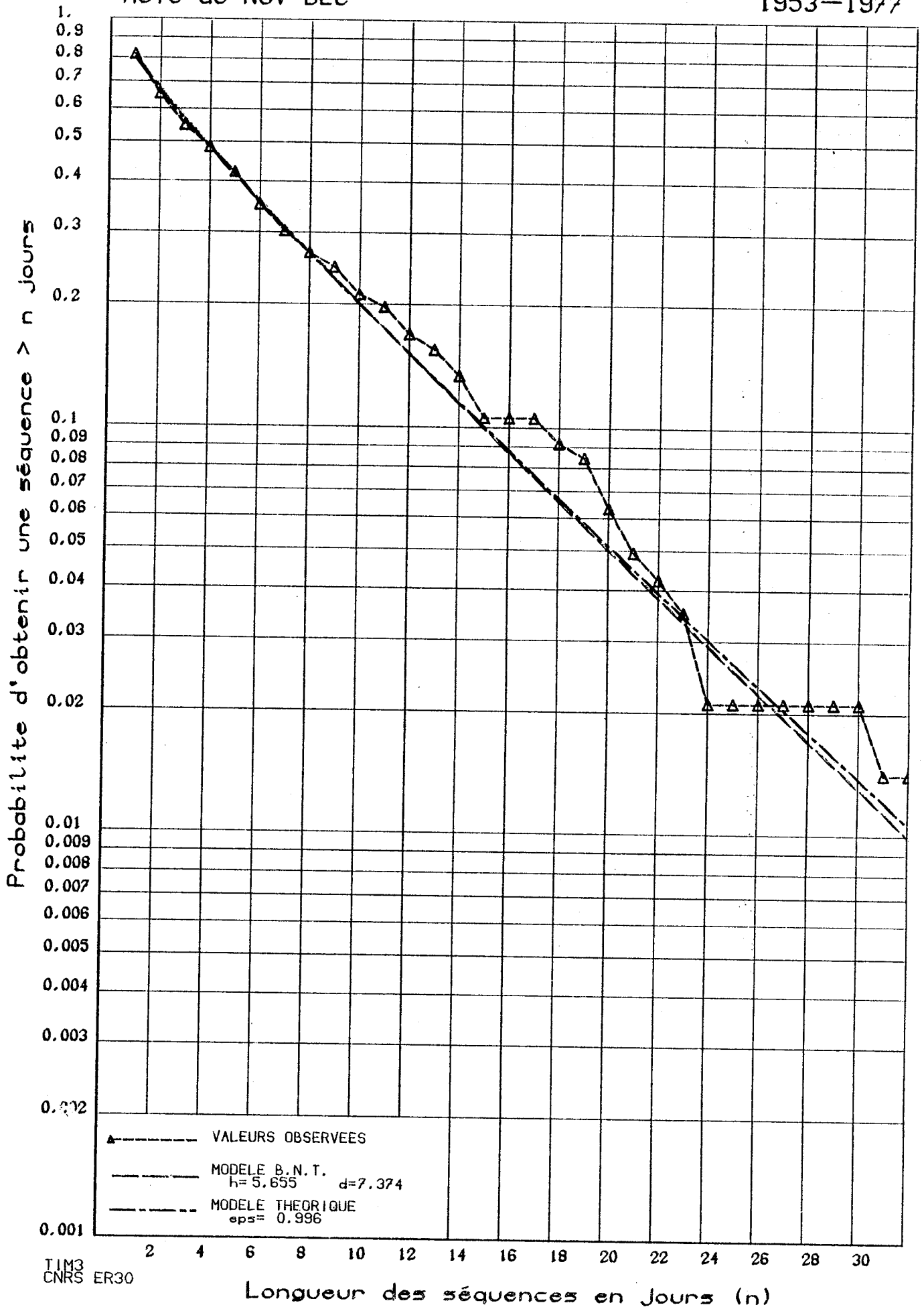
Tableau 14. idem tableau 13 : ordre 2.

MONTPELLIER

- 168 -

MOIS de NOV-DEC

1953-1977



TIM3
CNRS ER30

Fig 4. Probabilités d'obtenir une séquence sèche > n jours.

MONTELIMAR

- 169 -

MOIS DE NOV-DEC

1953-1977

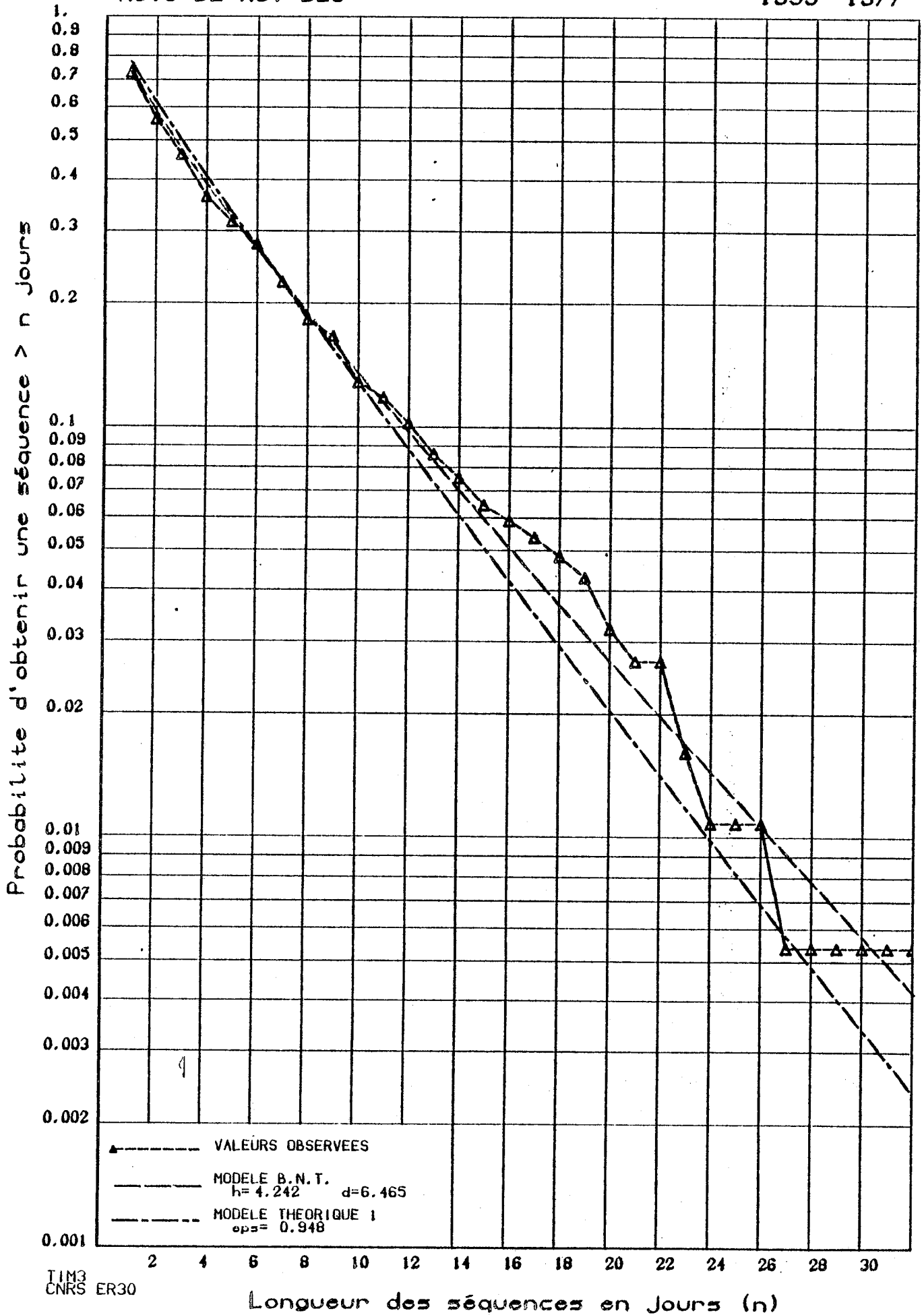


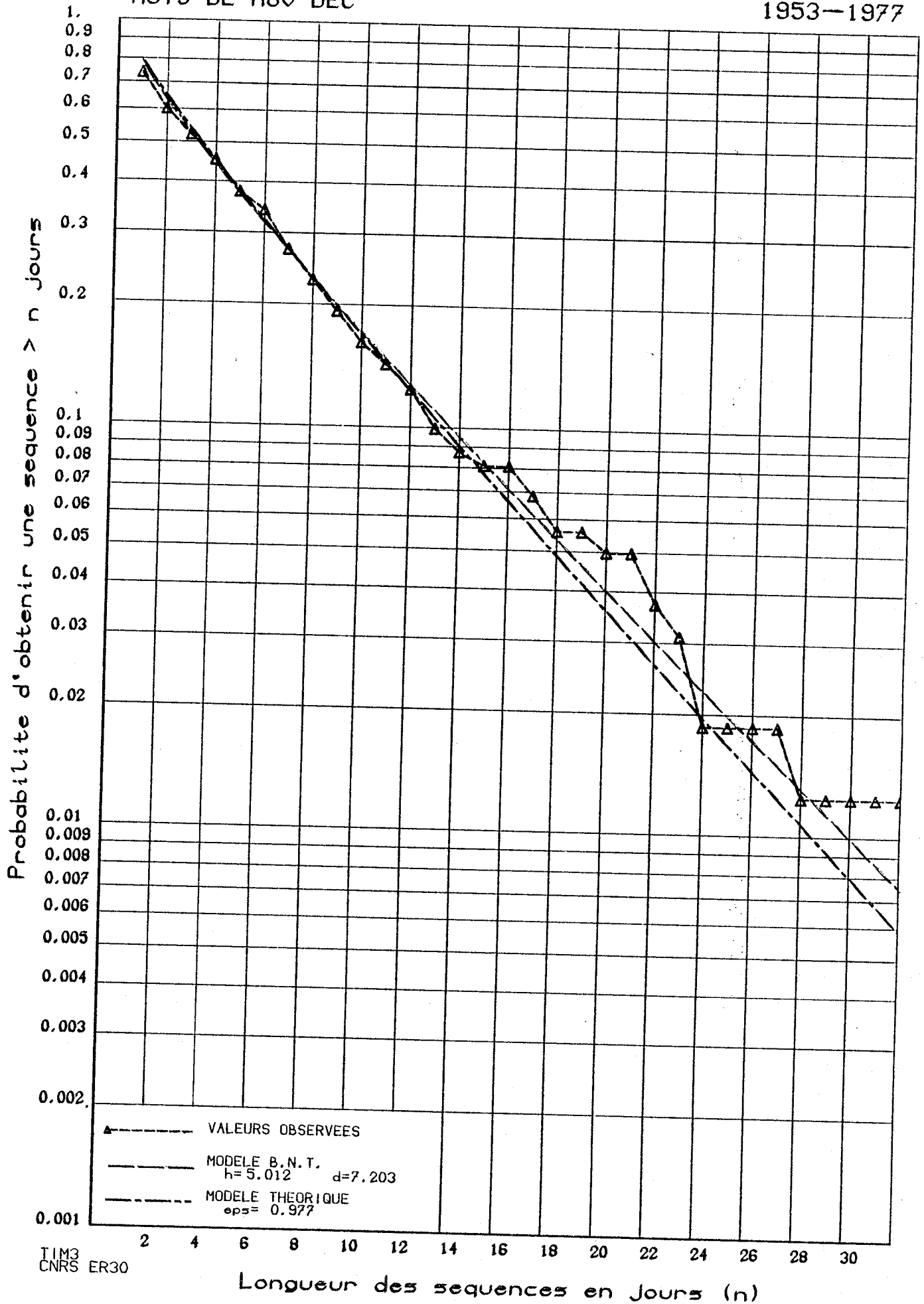
Fig 5. idem figure 4.

TOULON

- 170 -

MOIS DE nov-DEC

1953-1977



TIM3
CNRS ER30

Fig 6. idem figure 4.

III. ESSAI DE MODELISATION 3.

Nous considérons ici le réseau de trois stations MONTPELLIER(1), MONTELMAR(2), TOULON(3), pour la période Septembre Octobre. Nous envisageons une modélisation par un modèle du type de celui du chapitre IV avec une station fictive inobservable (*).

paramètres	h_1	d_1	h_0	d_0	α_1	α_2	α_3
val. estimées	7.535	7.496	1.018	0.785	.943	.920	.959

Tableau 15. Valeurs estimées des paramètres du modèle.

station	(*)	MONTPELLIER	MONTELMAR	TOULON
m^1 emp.		6.688	6.197	6.539
m^1 cal.	8.535	5.970	5.325	6.520
m^0 emp.		1.807	2.014	1.873
m^0 cal.	2.018	1.857	1.831	1.886

Tableau 16. Moments d'ordre 1 : empiriques et calculés

station	(*)	MONTPELLIER	MONTELMAR	TOULON
$m^1_{[2]}$ emp.		85.392	60.831	72.793
$m^1_{[2]}$ cal.	128.328,	59.276	46.011	71.881
$m^0_{[2]}$ emp.		3.155	4.056	3.142
$m^0_{[2]}$ cal.	3.871	3.320	3.237	3.415

Tableau 17. idem tableau 16 : ordre 2.

$N_{t+1} \backslash N_t$	0	1	2	3
0	0.3016 0.5043	0.1349 0.1173	0.1313 0.1173	0.0438 0.1173
1	0.1825 0.0047	0.2222 0.0083	0.1313 0.0083	0.0467 0.0083
2	0.1667 0.0779	0.2460 0.1387	0.2222 0.1387	0.0971 0.1387
3	0.3492 0.4131	0.3968 0.7356	0.5152 0.7356	0.8124 0.7356

Tableau 17. Probabilités de transition du processus N_t : empiriques et calculées.

ss-réseau	fréquences	prob. empiriques	prob. calculées
111	1050	0.7000	0.6739
011	63	0.0420	0.0404
101	91	0.0607	0.0584
110	44	0.0293	0.0282
001	49	0.0327	0.0035
010	24	0.0160	0.0017
100	53	0.0353	0.0024
000	126	0.0840	0.1914

Tableau 18. Probabilités pour qu'un sous-réseau soit sec : empiriques et calculées.

variable	N_t	S_{30}^1	S_{30}^2	S_{30}^3
empirique	2.448	24.76	23.62	25.06
calculé BNT	2.319	23.618	22.641	23.320
calculé	2.303	22.889	22.327	23.287

Tableau 19. Moments d'ordre 1 des statistiques N_t et S_{30}^j ; $j=1,2,3$: empiriques et calculés théoriques et calculés BNT pour S_{30}^j .

variable	N_t	S_{30}^1	S_{30}^2	S_{30}^3
empirique	0.919	9.142	13.195	10.336
calculé BNT	0.525	11.276	10.482	9.712
calculé	1.232	10.941	10.959	10.920

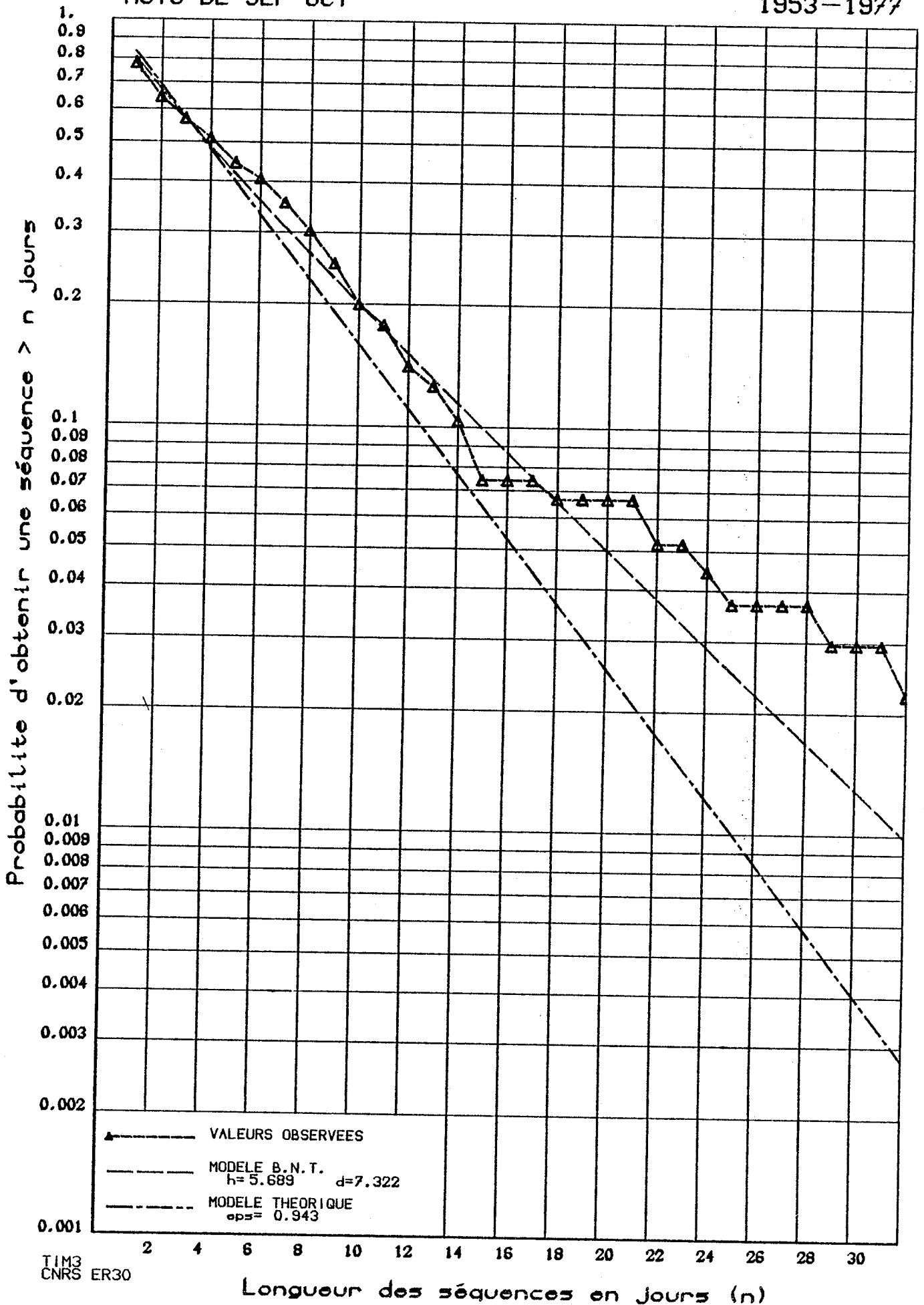
Tableau 20. idem tableau 19 : ordre 2.

MONTPELLIER

- 174 -

MOIS DE SEP-OCT

1953-1977



TIM3
CNRS ER30

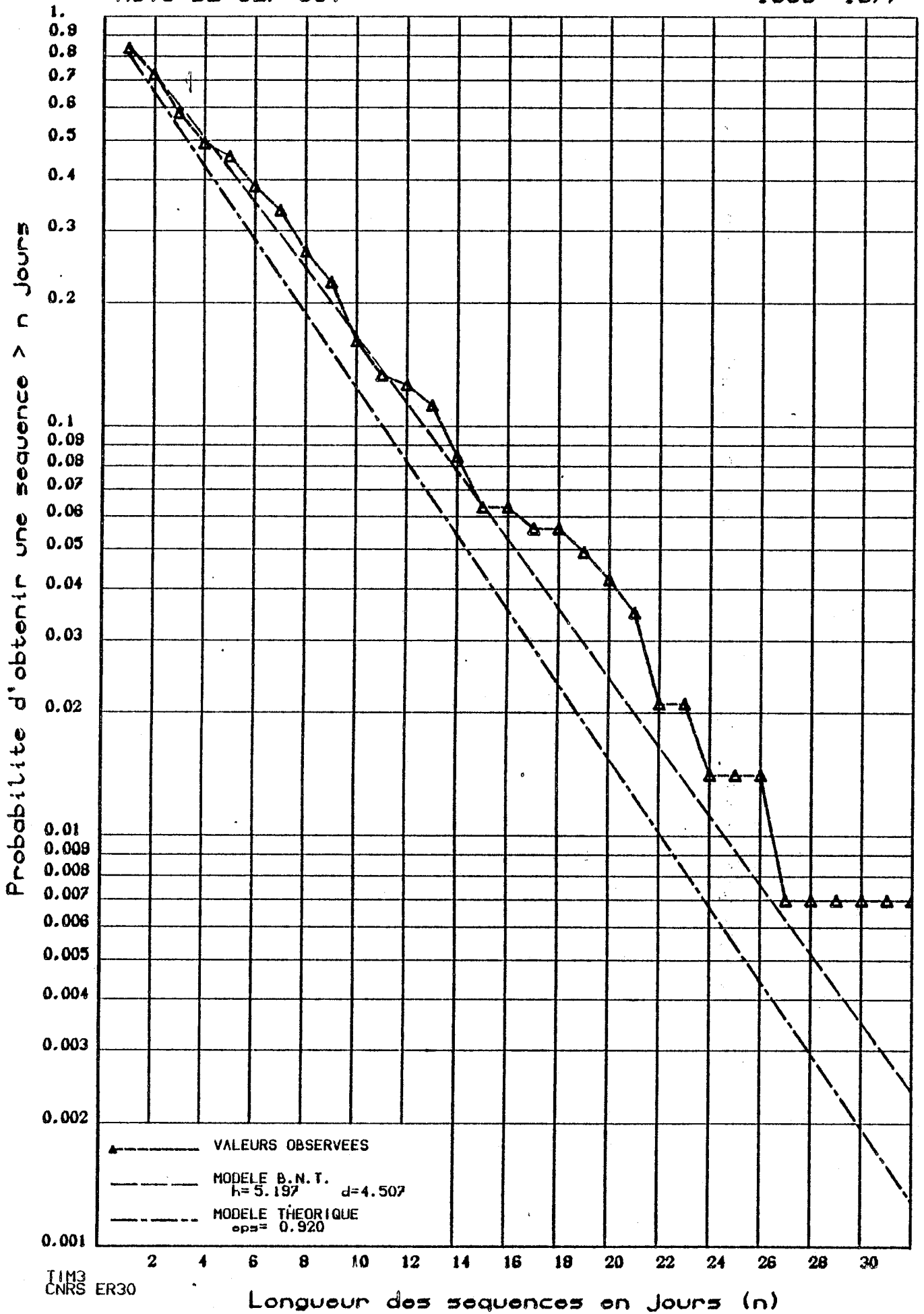
Fig 7. Probabilités d'obtenir une séquence sèche > n jours.

MONTELIMAR

MOIS DE SEP-OCT

- 175 -

1953-1977

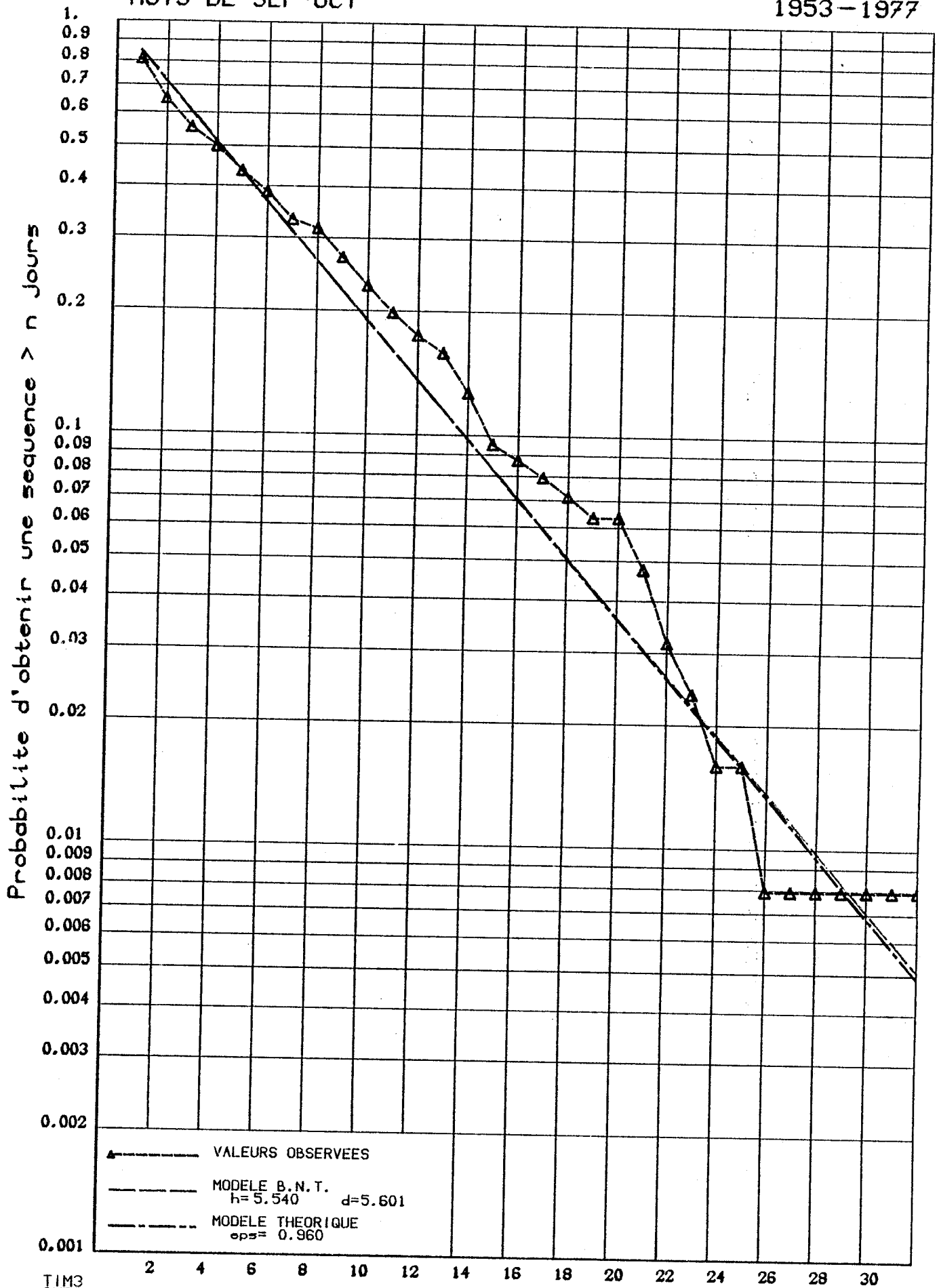


TOULON

- 176 -

MOIS DE SEP-OCT

1953-1977



TIM3
CNRS ER30

Fig 9. idem figure 7.

IV. ESSAI DE MODELISATION 4.

Pour ce dernier essai nous reprenons le réseau et la période qui ont été étudié dans l'essai du paragraphe I. Nous envisageons ici d'utiliser le modèle défini dans le chapitre V, la station de MARSEILLE étant considérée comme station test observée.

paramètres	h_1	d_1	h_0	d_0	α_2	α_3	α_4
val. estimées	6.327	7.528	1.017	0.918	.982	.937	.994

Tableau 20. Valeurs estimées des paramètres du modèle

station	MARSEILLE	MONTPELLIER	MONTELMAR	TOULON
m^1 emp.	7.327	6.688	6.197	6.539
m^1 cal.	7.327	6.607	5.846	5.013
m^0 emp.	2.017	1.807	2.014	1.873
m^0 cal.	2.017	1.930	1.843	1.758

Tableau 21. Moments d'ordre 1 : empiriques et calculées.

station	MARSEILLE	MONTPELLIER	MONTELMAR	TOULON
$m_{[2]}^1$ emp.	100.327	85.392	60.831	72.793
$m_{[2]}^1$ cal.	100.327	79.422	60.158	42.264
$m_{[2]}^0$ emp.	3.603	3.155	4.056	3.142
$m_{[2]}^0$ cal.	3.603	3.288	2.970	2.651

Tableau 22. idem tableau 21 : ordre 2.

$N_{t+1} \backslash N_t$	0	1	2	3
0	0.3016 0.4618	0.1349 0.4407	0.1313 0.1267	0.0438 0.1250
1	0.1825 0.0424	0.2222 0.0406	0.1313 0.0127	0.0467 0.0126
2	0.1667 0.0417	0.2460 0.0436	0.2222 0.0714	0.0971 0.0716
3	0.3492 0.4540	0.3968 0.4751	0.5152 0.7891	0.8124 0.7908

Tableau 23. Probabilités de transition du processus N_t : empiriques et calculées .

ss-réseau	fréquences	prob. empiriques	prob. calculées
111	1050	0.7000	0.7181
011	63	0.0420	0.0128
101	91	0.0607	0.0481
110	44	0.0293	0.0042
001	49	0.0327	0.0020
010	24	0.0160	0.0133
100	53	0.0353	0.0038
000	126	0.0840	0.1977

Tableau 24. Probabilités pour qu'un sous-réseau soit sec : empiriques et calculées.

variable	N_t	S_{30}^2	S_{30}^3	S_{30}^4
empirique	2.448	24.76	23.62	25.06
calculé BNT	2.319	23.618	22.641	23.320
calculé	2.417	23.226	22.455	23.429

Tableau 25. Moments d'ordre 1 des statistiques N_t et S_{30}^j ; $j=2,3,4$: empiriques et calculés théoriques et calculés BNT pour S_{30}^j .

variable	N_t	S_{30}^2	S_{30}^3	S_{30}^4
empirique	1.017	9.142	13.195	10.336
calculé BNT	0.525	11.276	10.482	9.712
calculé	1.136	12.407	11.530	12.652

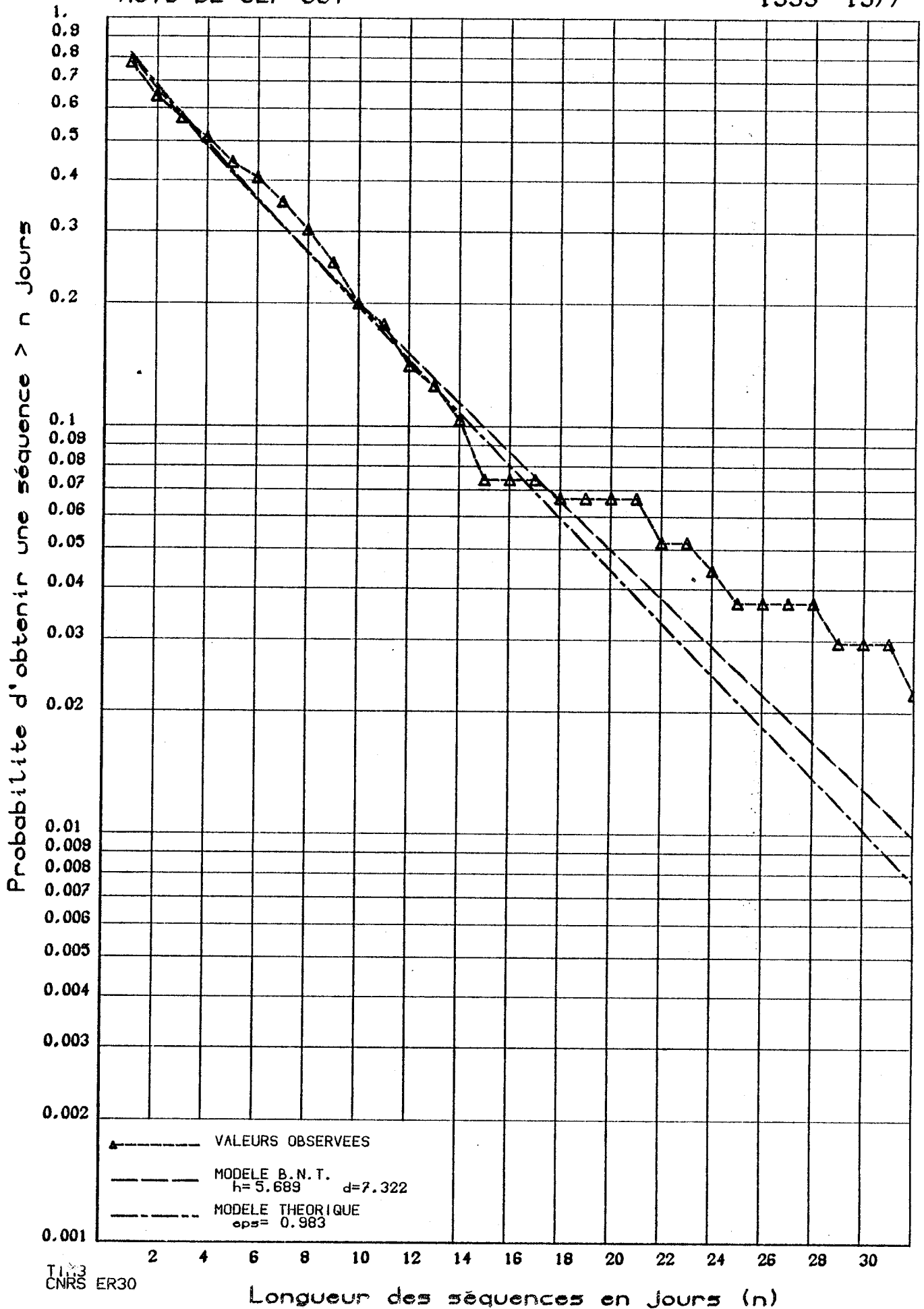
Tableau 26. idem tableau 25 : ordre 2.

MONTPELLIER

MOIS DE SEP-OCT

- 180 -

1953-1977



T113
CNRS ER30

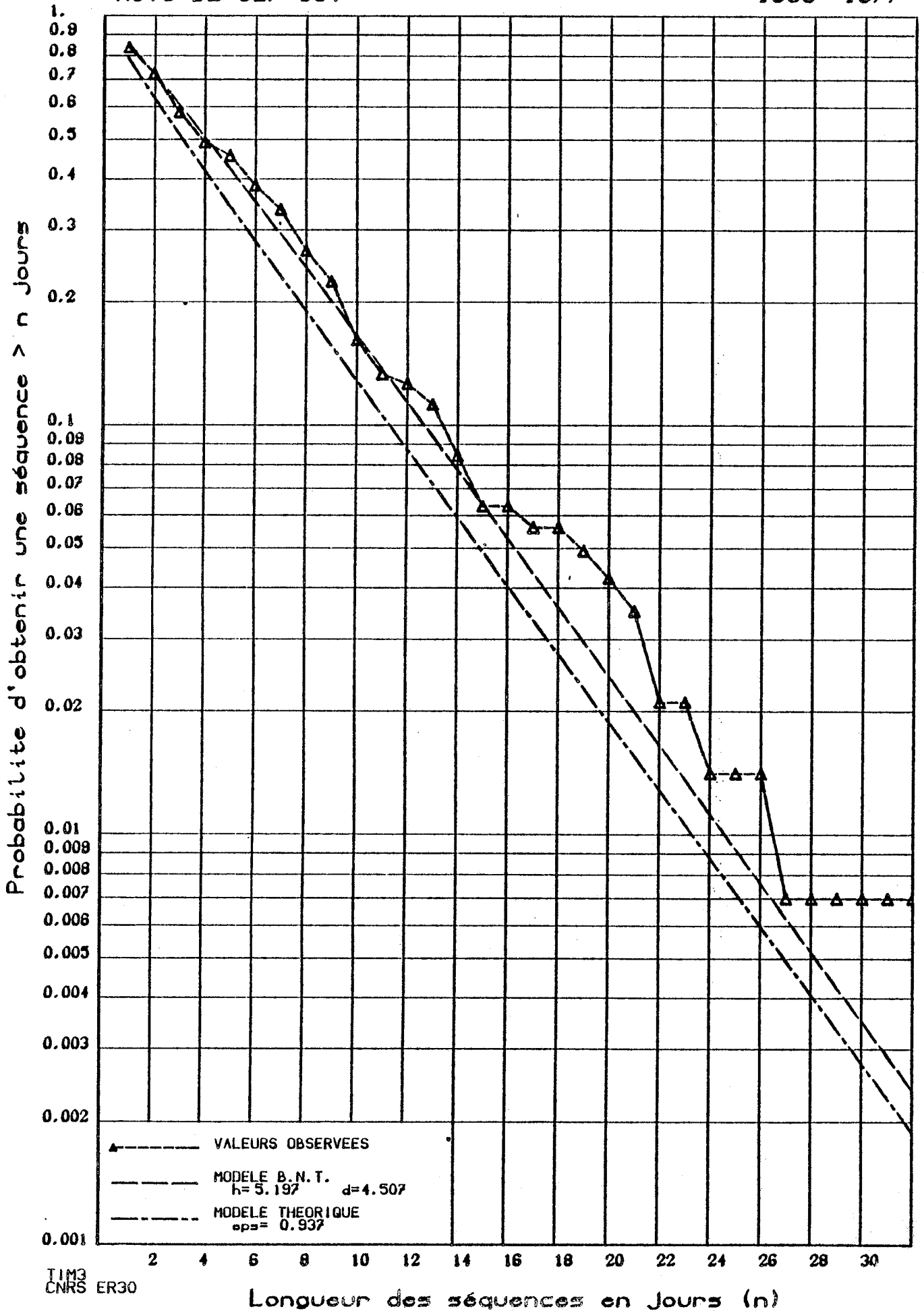
Fig 10. Probabilités d'obtenir une séquence sèche > n jours.

MONTELMAR

- 181 -

MOIS DE SEP-OCT

1953-1977

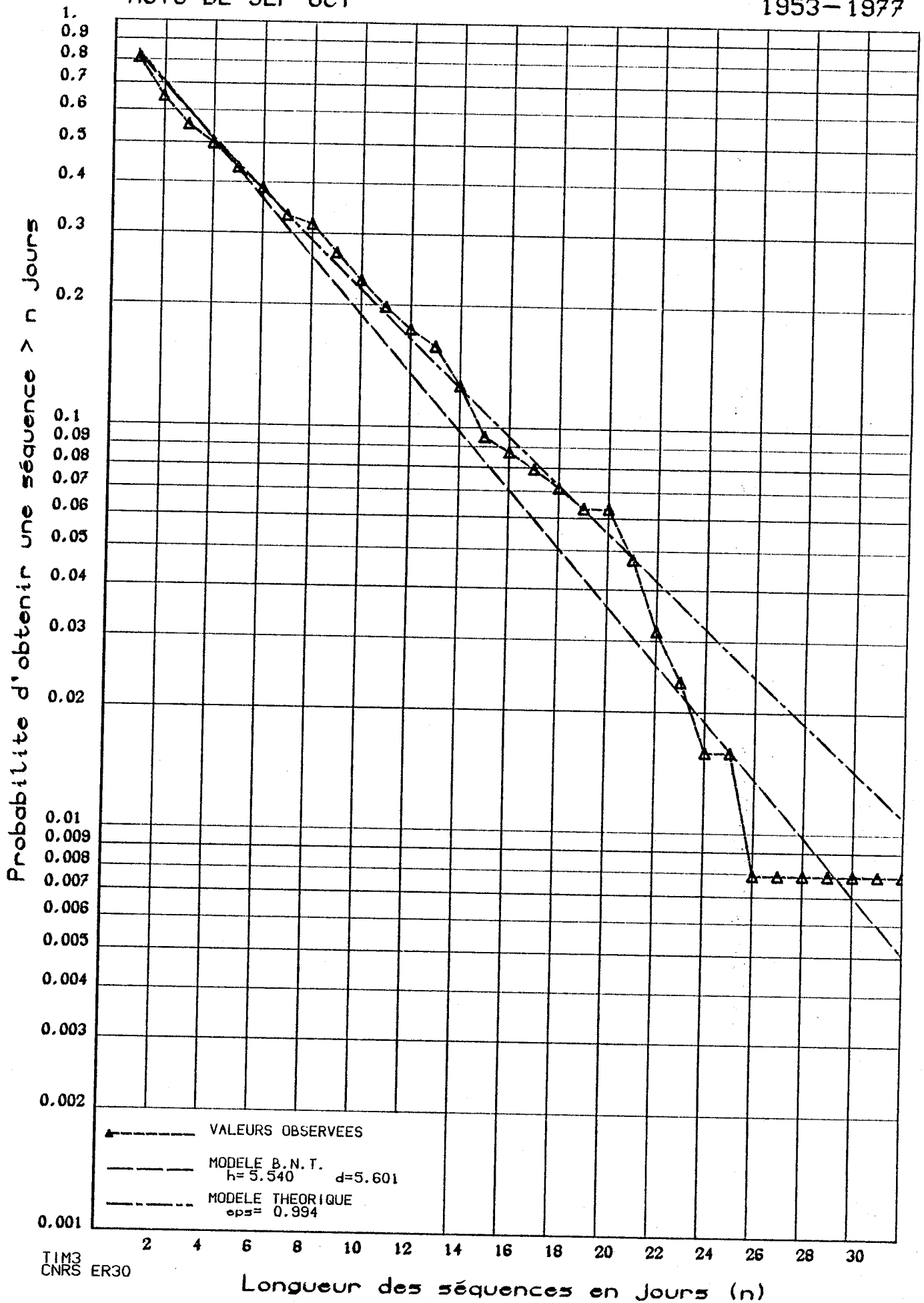


TOULON

- 182 -

MOIS DE SEP-OCT

1953-1977



TIM3
CNRS ER30

Fig 12. idem figure 10.

CONCLUSION

Nous venons donc d'étudier mathématiquement et expérimentalement des modèles de séries chronologiques vectorielles à composantes binaires susceptibles d'être utilisés en climatologie. Si au plan des objectifs que nous nous étions fixés, l'analyse montre la grande facilité d'utilisation de ces modèles compte tenu de la calculabilité qu'ils offrent, l'expérimentation faite conduit à certaines remarques.

Tout d'abord l'étude par simulations donne à penser que les procédures naïves d'estimation des paramètres qui ont été proposées devraient être améliorées. En particulier il est clair que le fait d'utiliser certaines approximations des moments pour $\alpha_j = 1$ conduit à des écarts assez sensibles entre la réalité qu'on est sensé simuler et la réalité calculée après estimation d'autant plus que l'on s'écarte de la dite condition. Il faut aussi noter que ces écarts sont du même ordre que ceux entre la réalité simulée et le résultat de la simulation ce qui peut faire douter de la qualité des simulateurs que nous avons utilisés.

Quant aux essais de modélisation sur données réelles on remarque que les utilisations du modèle du chapitre IV avec station test observée (essais 1 et 2) conduisent à des représentations du phénomène dans les différentes stations qui concurrencent valablement celles fournies par une modélisation station par station et ce à un "coût" nettement inférieur en nombre de paramètres. La représentation de la structure d'interdépendance est moins satisfaisante bien que les probabilités d'observation des différentes sous-réseaux secs soient assez bien reconstituées. La comparaison des essais 1 et 3 avec un même modèle sur un même réseau et pour une même période avec station test observée d'une part et station test fictive d'autre part montre une nette dégradation pour MONTPELLIER et MONTELLIMAR lorsqu'on passe du premier cas au deuxième. La comparaison des essais 1 et 4 sur un même

réseau et une même période avec deux modèles différents dans lesquels MARSEILLE joue le rôle d'une station test met en évidence une meilleure qualité des représentations station par station et conjointes dans le premier cas. Il faut remarquer que les couples de mois consécutifs agrégés pour ces essais n'offrent pas vraiment les garanties d'homogénéité souhaitables. D'autres essais, pour des réseaux de stations et des périodes de l'année choisis après une phase d'analyse préliminaire soigneusement menée, devraient permettre de mieux juger des possibilités d'utilisation des modèles considérés; en particulier il conviendrait de s'intéresser à des périodes plus homogènes (d'un mois par exemple) observées au cours d'un plus grand nombre d'années.

Les modèles de séries chronologiques vectorielles que nous avons envisagés semblent donc présenter un certain intérêt pour les applications en climatologie bien qu'il convienne d'améliorer les conditions de leur utilisation pratique. Il devrait être possible d'intégrer ces modèles dans une modélisation plus globale du phénomène climatologique où des variables du type quantité de précipitation ou rapport d'insolation seraient prises en compte. Notons enfin que ces modèles sont peut être susceptibles d'être utilisés dans d'autres domaines d'applications tels que la fiabilité.

BIBLIOGRAPHIE

- BERGER,A.,GOOSSENS,C.H.R.-(1982) .Persistence of wet and dry spells in Belgium, Scientific report 1983/3, Université Catholique de Louvain, Belgique.
- BERGER,A.,GOOSSENS,C.H.R.-(1983) .Persistence of wet and dry spells at UCCLE (Belgium). Journal of Climatology, vol. 3, 21-34.
- BOUSSEBOUA,M.-(1983) .Etude de quelques modèles de séries chronologiques binaires. Thèse de 3-eme cycle Mathématiques Appliquées Université de Grenoble 1.
- BUISSHAND,T.A.-(1977) .Stochastic modelling of daily rainfall sequences. Commun. Agric. Univ. Wageningen. 77-3, the Netherlands.
- BUISSHAND,T.A.-(1978) .The binary DARMA(1,1) process as a model for wet-dry sequences. Technical Note 78-01. Agric. Univ. Wageningen, the Netherlands.
- CASKEY,I.E.J.(1963) .A Markov chain model for the probability occurrence in intervals of various lengths, Month. Weather Rev. 91, 289-301.
- CHANG,T.J.,DELLEUR,J.W.,KAWVAS,M.L.-(1982) .Stochastic daily precipitations modeling and daily streamflow transfer processes. Technical report n° 146, Purdue University Water Resources Research Center.
- CHANG,T.J.,DELLEUR,J.W.,KAWVAS,M.L.-(1983a) .Daily precipitations modeling by the mixture of discrete autoregressive moving-average processes and exponential distributions. Preprint, Purdue University Water Resources Research Center.
- CHANG,T.J.,DELLEUR,J.W.,KAWVAS,M.L.-(1983b) .Multi-state discrete ARMA models and multi-state run lengths for the daily precipitation sequences. Preprint, Purdue University Water Resources Research Center.

- COX, D.R.-(1972) .The analysis of multivariate binary data. Jour. Royal Statist. Soc. C 21, p. 113-120.
- GABRIEL, K.R.-(1959) .The distribution of the number of successes in a sequence of dependent trials. Biometrika, 46, p. 454-460.
- GABRIEL, K.R., NEUMAN, J.-(1962) .A Markov chain model for daily rainfall occurrence at Tel-aviv. Quart. J. Roy. Meteor. Soc. 88 p. 90-99.
- GALLOY, E.-(1982) .Contribution à l'étude de la sécheresse. Thèse de 3-eme cycle, Laboratoire de géographie physique. Grenoble.
- GALLOY, E., LE BRETON, A., MARTIN, S.-(1982) .Analyse des séquences de jours secs consécutifs, Application à 31 postes du réseau météorologiques français, La météorologie, VI^e série, n° 28, 5-24.
- GALLOY, E., LE BRETON, A., MARTIN, S.-(1983) .A model for weather cycles based on daily rainfall occurrence, Lectures Notes in Biomathematics n° 49, 303-318.
- GREEN, J.R.-(1964) .A model for rainfall occurrence. J. of the Roy. Stat. Soc. B. 26, p. 345-353.
- GREGOIRE, G.-(1982) .Processus de renouvellement et généralisations. Une présentation en vue des applications. Rapport de recherche, Laboratoire IMAG.
- GRISOLET, M., GUILMET, B., ARLERY, R.-(1962) .Climatologie, méthodes et pratiques, Gauthier Villars et C-ie.
- HELGRET, H.J.-(1970) .On sums of random variables defined on a two states Markov chain. J. Appl. Prob. 7, 761-765.
- HOPKINS, J., ROBILLARD, P.-(1964) .Some statistics of daily rainfall occurrence for the canadian Prairie Provinces, J. App. Meteorology, 3, 600.

- JACOBS,P.A.,LEWIS,P.A.W.--(1978a) .Discrete time series generated by mixture I :
correlational and runs properties. J. R. Statist. Soc. B. 40, n° 1,
p. 94-105.
- JACOBS,P.A.,LEWIS,P.A.W.--(1978b) .Discrete time series generated by mixture II :
asymptotic properties. J. R. Statist. Soc. B. 40, n° 2, p. 222-228.
- KATZ,R.W.--(1974) .Computing probabilities associated with the Markov chain
model for precipitation, J. App. Meteorology, 13, 953.
- KEDEM,B.--(1977) .Counting the number of 0-1 stationary series having the same
likelihood. Discrete mathematics, 18, p. 285-289.
- KEDEM,B.--(1980) .Binary time series. Lecture Notes in pure and applied
mathematics. Vol. 52.
- KLOTZ,J.--(1972) .Markov chain clustering of births by sex. Proc. Sixth.
Berkeley Symp. Mathematics Statist. Prob.
- LE BRETON,A.,MARTIN,S.--(1979) .Un modèle pour l'étude des séquences climatologi-
ques. Séminaire de statistique, Laboratoire IMAG, p. 3-34.
- TODOROVIC,L.L.,WOOLHISER,D.A.--(1975) .A stochastic model of n-day precipitation,
J. App. Meteorology, 14,17.
- WEISS,L.L.--(1964) .Sequences of wet and dry days described by a Markov Chain
Probability model, Mont. Wea. Rev., 92, 160.

DERNIERE PAGE D'UNE THESE

3^E CYCLE, DOCTEUR INGÉNIEUR OU UNIVERSITÉ

Vu les dispositions de l'arrêté du 16 avril 1974,

Vu les rapports de M. A. LEBRETAN.....

M.

M. ESSERAR Belkeir..... est autorisé
à présenter une thèse en vue de l'obtention du grade de DOCTEUR ...3^e cycle
Mathématiques Appliquées

Grenoble, le 26-4-86

Le Président de l'Université Scientifique
et Médicale

M. TANCHE



[Handwritten signature]