

Sciences

TD

380

A1

C132

1,3

les cahiers de centreau



CENTREAU
Centre de recherches sur l'eau



UNIVERSITÉ LAVAL
Québec, Canada

LA TEMPÉRATURE DE L'AIR ET LE CYCLE HYDROLOGIQUE: STRUCTURE STATISTIQUE

José LLAMAS
Jean-Pierre TRIBOULET

Vol. 1, n° 3, mai 1976

LES CAHIERS DE CENTREAU

LES CAHIERS DE CENTREAU constituent une collection d'études qui fait connaître les travaux des chercheurs de l'Université Laval dans le domaine de l'eau et qui s'adresse à la fois aux scientifiques et aux praticiens intéressés aux richesses naturelles et à leur conservation. Ils traitent chacun de sujets particuliers et apportent des conclusions ou des recommandations susceptibles non seulement d'augmenter les connaissances, mais de fournir aussi des solutions à de nombreux problèmes concernant la gestion et l'aménagement des ressources.

Au moyen de cette revue scientifique bimestrielle, le CENTRE DE RECHERCHES SUR L'EAU de l'Université Laval désire mettre en évidence l'importance des sciences hydriques à cette institution et promouvoir des échanges à l'échelle nationale et internationale.

Ces cahiers sont adressés gratuitement sur demande. Les textes publiés sont sous la responsabilité de leurs auteurs et demeurent leur propriété. Leur reproduction est permise à condition d'en obtenir l'autorisation expresse du directeur de CENTREAU.

Dépôt légal - Bibliothèque nationale du Canada

Dépôt légal - Bibliothèque nationale du Québec

Pour tout renseignement, s'adresser à:

José Llamas, Ph.D., directeur

CENTREAU

Pavillon Pouliot

Université Laval, Québec,

Canada, G1K 7P4

COMITÉ DE RÉDACTION

Analyse: Guy Moreau, D.Sc.

Documentation: Andrée Tardif, L.L.

Édition: G.-Oscar Villeneuve, Ph.D.

TD
380
A1
C132
113

LA TEMPÉRATURE DE L'AIR
ET LE CYCLE HYDROLOGIQUE:
STRUCTURE STATISTIQUE

José LLAMAS

Jean-Pierre TRIBOULET



RÉSUMÉ

Cette étude a pu être réalisée grâce à l'aide financière du Conseil National de Recherches du Canada, projet CNR-A-B-540 dont M. José Llamas est le responsable, et du ministère Environnement Canada, projet 5043-4 dont M. Jean-Louis Verrette est le coordonnateur pour le Centre de Recherches sur l'Eau de l'Université Laval, et grâce à la collaboration du Service de l'Environnement Atmosphérique du Canada et du Service de la Météorologie du Québec qui ont fourni les données.

LLAMAS, José: Professeur agrégé à la faculté des Sciences et de Génie de l'Université Laval depuis 1972, José Llamas a d'abord travaillé au ministère des Richesses naturelles du Québec comme ingénieur et comme directeur du Service de l'Aménagement hydraulique. Diplômé en 1958 de l'École Supérieure des Travaux Publics de Madrid, il a obtenu son doctorat (Ph.D.) de l'Université d'État du Colorado en 1968. Il est actuellement responsable du programme des cours d'hydrologie de l'Université Laval et du groupe des chercheurs en cette discipline à CENTREAU où, en juin 1975, il a succédé à Alain Soucy comme directeur.

TRIBOULET, Jean-Pierre: Actuellement à l'emploi de la Division de l'Hydrologie, Service Hydraulique, Direction Projets de Centrales de l'Hydro-Québec, l'auteur est ingénieur diplômé de l'École Catholique d'Arts et Métiers de Lyon (France) en 1969 et ingénieur hydraulicien diplômé de l'Université de Grenoble en 1970. Après deux années à la Direction Générale des Eaux du ministère des Richesses naturelles du Québec, Service de l'Aménagement hydraulique, il a collaboré aux travaux en hydrologie du Centre de Recherches sur l'Eau de l'Université Laval et obtenu en 1975 une maîtrise en sciences de cette même institution.

Les méthodes classiques d'analyse des séries chronologiques permettent d'identifier les différentes composantes d'une série constituée par les valeurs des températures moyennes journalières de l'air enregistrées depuis plusieurs années à une station météorologique. La tendance à long terme reflétée par les fluctuations de la moyenne annuelle, la composante saisonnière plus ou moins sinusoidale, la déviation journalière partiellement persistante d'un jour à l'autre et enfin la composante résiduelle considérée comme aléatoire sont les quatre termes qui peuvent représenter la structure mathématique et séquentielle.

Un exemple d'utilisation du modèle est réalisé avec la série des valeurs de la température moyenne journalière calculées à partir des températures quotidiennes maximale et minimale enregistrées durant une période homogène de 28 années (1944-1971) à la station de l'aéroport de Québec.

L'étude fait ressortir la nécessité de définir la composante saisonnière moyenne interannuelle à l'aide de deux harmoniques en plus de la sinusoidale fondamentale afin d'admettre que le processus de la déviation journalière soit stationnaire en moyenne. Le comportement saisonnier de l'écart-type afférent à la distribution de chaque jour de l'année est aussi mis en évidence. Par normalisation, le processus de la déviation est rendu stationnaire au moins en variance, mais pas davantage car, après cette transformation, les coefficients d'asymétrie et d'aplatissement de la fonction de distribution d'ensemble fluctuent saisonnièrement.

La validité du modèle est testée par génération de séquences synthétiques. Les statistiques établies sur ces séquences sont très proches des statistiques observées en valeur moyenne alors que les valeurs extrêmes paraissent accentuées sans toutefois sortir des limites physiques possibles.

En suivant les variations des estimateurs des paramètres saisonniers en fonction de la taille de l'échantillon, on fixe la durée minimum d'observation pour obtenir des estimateurs représentatifs des valeurs à long terme. Cette méthode permet donc de limiter la durée d'implantation des stations secondaires.

La structure mathématique du modèle est suffisamment générale pour que l'utilisation d'un tel modèle puisse être envisagée à l'échelle régionale. En effet, par cartographie des coefficients déterminés par la méthode suggérée pour les stations de base d'un réseau en place depuis longtemps, il doit être possible d'interpoler la valeur des coefficients en

un site intermédiaire et de construire pour ce point des séries synthétiques de valeurs de température journalière de l'air pour les besoins d'un projet particulier.

ABSTRACT

Using classical methods of time series analysis, it is possible to identify the different components of a long-term daily mean air temperature record. The structure of the series may be constituted by superposition of four principal components: the long term trend, the seasonal cycle which is almost periodic, the stochastic dependent and the residual independent components.

The record used for analysis is the mean series of the daily maximum and minimum temperatures observed during a homogeneous period of 28 years (1944-1971) at the Quebec Airport meteorological station.

To assure first order stationarity of the stochastic component, the cyclic component must be described at least by the sum of the annual periodicity and the first two harmonics. The stationarity in the variance is obtained by normalization using the associated daily standard deviation having a seasonal pattern described by the first three terms of the Fourier series. The standardized process obtained in this way is second order stationary because its skewness and kurtosis coefficients also follow seasonal pattern.

The validity of the model is tested by generating synthetic series. The agreement between the statistic structure of the generated series and the observed values is really acceptable. However, the generated extreme values are lightly greater than the observed ones; but they remain within physical limits.

An attempt is made to analyse the variability of the model seasonal parameters in regard with the length of the initial record in order to determine the required minimum length of record given stable estimates at a significant level. So the second order stations can be operated only for the required minimum length.

The structure of the model is general enough for its application on a regional scale by mapping the coefficients of the model calculated for the long-term stations. Then, the values at any site may be interpolated and used to generate synthetic series for the needs of a specific project.

TABLE DES MATIÈRES

I.	<u>INTRODUCTION</u>	
A.	Cadre général	1
B.	Bibliographie et méthodes d'approche	2
C.	Contenu de la recherche	4
II.	<u>MODÈLE DE TEMPÉRATURE DE L'AIR À QUÉBEC</u>	
A.	Données	6
B.	Tendance à long terme	9
C.	Cycle saisonnier	15
D.	Température résiduelle ou déviation	24
	1. Identification et stationnarité	24
	2. Autocorrélation et résidu	29
E.	Validation du modèle de génération	34
	1. Comparaison des paramètres des séries observées et générées	35
	2. Statistiques sur les séries observées et générées	40
	3. Variation des estimateurs des paramètres en fonction de la longueur de la série initiale d'observation	44
III.	<u>CONCLUSIONS</u>	50
	<u>APPENDICE</u>	55
A.	Définitions	57
B.	Conception générale des modèles	60
C.	Composante déterministe	62
	1. Tendance	64
	2. Composante périodique	67
D.	Composante stochastique	70
	1. Fonction d'autocorrélation	70
	2. Analyse spectrale	72
	3. Processus stochastiques linéaires	73
	<u>RÉFÉRENCES</u>	79

I. INTRODUCTION

A. Cadre général

La recherche présentée dans les pages qui suivent est une analyse de la structure mathématique d'une série de valeurs de température journalière de l'air en un lieu donné et s'inscrit dans le cadre de l'étude de la relation entre la température de l'eau et celle de l'air ambiant.

L'intérêt porté par les chercheurs à l'évolution de la température de l'eau des cours d'eau s'est en effet accru au cours des quinze dernières années. La recherche dans ce domaine coïncide naturellement avec un début d'utilisation intensive des cours d'eau et des lacs à des fins motrices, industrielles ou récréatives. Les changements artificiels dans les réseaux hydrographiques, la régularisation des débits, l'érection des réservoirs et enfin les rejets d'eau chaude se multiplient à un tel point que le régime thermique des cours d'eau s'en trouve plus ou moins modifié. De la biologie à la formation des glaces, nombreux sont les domaines où l'on peut enregistrer des changements. L'am-

pleur des modifications devant être évaluée dans le cadre des études d'impact avant la réalisation de l'aménagement, le projeteur s'en remet à l'utilisation des modèles pour lesquels la température de l'air est toujours l'intrant principal.

Qu'il soit basé sur les lois des transferts thermiques ou sur les liaisons statistiques, le modèle utilisé nécessite toujours une bonne connaissance de la série des données de température de l'air. Or, les mesures de cet élément climatique ne sont pas toujours disponibles au site même d'un projet. Au Canada particulièrement, de nombreux projets sont réalisés dans des territoires éloignés des centres traditionnels d'activité. Le réseau météorologique est très peu dense dans ces secteurs. A partir des données à quelques stations de base, on peut estimer les paramètres d'un type de modèle valable pour toutes les stations. L'interpolation des coefficients permet alors de générer des températures "synthétiques" en des points intermédiaires sans mesures. La conception du type de modèle présenté ici repose sur les principes classiques de l'analyse des séries chronologiques.

B. Bibliographie et méthodes d'approche

Les techniques d'analyse des séries chronologiques sont utilisées pour identifier les différentes composantes du signal constitué par la série des températures moyennes

journalières. Une ou deux composantes déterministes peuvent être séparées: la tendance à très long terme, si elle existe, qui se reflète dans la température moyenne annuelle, et la variation saisonnière ou le cycle saisonnier. En raison de la courte période d'observation disponible, la tendance est rarement bien identifiable. Par contre, la variation cyclique annuelle est une caractéristique fondamentale bien établie.

D'après les analyses harmoniques et spectrales déjà effectuées sur des séries de températures de l'air (Kothandaraman, 1971; Griffith *et al.*, 1956; Panofsky, 1955; Julian, 1967), le cycle annuel est le seul qui soit vraiment significatif. Velz et Gannon (1960) mettent en évidence le cycle annuel de la température de l'air et son intervalle de confiance parallèlement aux cycles des autres éléments atmosphériques (vent, pression, rayonnement) afin d'arriver au cycle de la température de l'eau qui est une synthèse des précédents. Cluis (1972), après avoir défini la caractéristique climatique sinusoidale, introduit la température résiduelle de l'air dans un modèle qui génère la température résiduelle de l'eau par rapport à sa propre caractéristique annuelle.

La technique des modèles linéaires de dépendance pour représenter la persistance et celle des nombres pseudo-aléatoires générés suivant une fonction de répartition connue,

trouvent ici un champ d'application. Song *et al.* (1973) développent un modèle en retenant la fondamentale de la série de Fourier comme composante saisonnière, à laquelle s'ajoute une composante résiduelle stochastique. Cette dernière est reconnue persistante sur deux jours consécutifs et est modélisée par un processus autorégressif à l'ordre 1. Un dernier terme aléatoire pur est introduit pour tenir compte de la variance non expliquée.

C. Contenu de la recherche

A la lumière des études précédemment citées, il est apparu nécessaire d'insister sur la vérification de la stationnarité des processus mis en jeu. En effet, les auteurs semblent ne considérer que le cycle saisonnier plus ou moins sinusoïdal et le résidu obtenu par simple soustraction du cycle à la série observée. Si la série résiduelle ne suit pas un processus stationnaire, elle ne peut pas être représentée par un modèle à coefficients constants. En hydrologie et en météorologie, l'expérience prouve en effet que le résidu a généralement un comportement saisonnier (Quimpo, 1967) (Triboulet, 1975). Il faut alors fractionner le domaine d'observation et fournir un modèle pour chaque saison, ce qui n'est pas très "économique", ou rendre au préalable le processus stationnaire dans son ensemble avant toute analyse séquentielle.

La méthodologie présentée ici consiste à identifier

les composantes déterministes, tendance et cycle saisonnier, obtenir une série résiduelle, tester la stationnarité du processus et, si nécessaire, transformer le résidu afin qu'il soit possible de modéliser par des techniques simples. Dans une deuxième phase certains tests portent sur la stabilité des paramètres saisonniers et autres retenus dans le modèle de génération en fonction de la taille de l'échantillon. Le cycle saisonnier n'est en réalité pas identique d'une année à l'autre; son amplitude, son déphasage et le pourcentage de variance expliquée fluctuent. Les paramètres qui sont introduits pour simuler le comportement de la série résiduelle (les coefficients d'autorégression par exemple) peuvent également fluctuer. Un modèle de génération étant utilisé pour générer de longues séquences de températures dans le but d'en connaître les moyennes et extrêmes caractéristiques, il est opportun de savoir quelle est la longueur minimale de l'échantillon initial qui permet d'atteindre ce but.

Les divers tests évoqués ci-dessus ont été appliqués dans le cas de l'analyse de la structure de la température journalière de l'air enregistrée à la station de l'aéroport de l'Ancienne Lorette, près de la ville de Québec, durant 28 années (Triboulet, 1975). Les résultats obtenus au cours de ce traitement viennent illustrer les quelques rappels théoriques présentés en appendice.

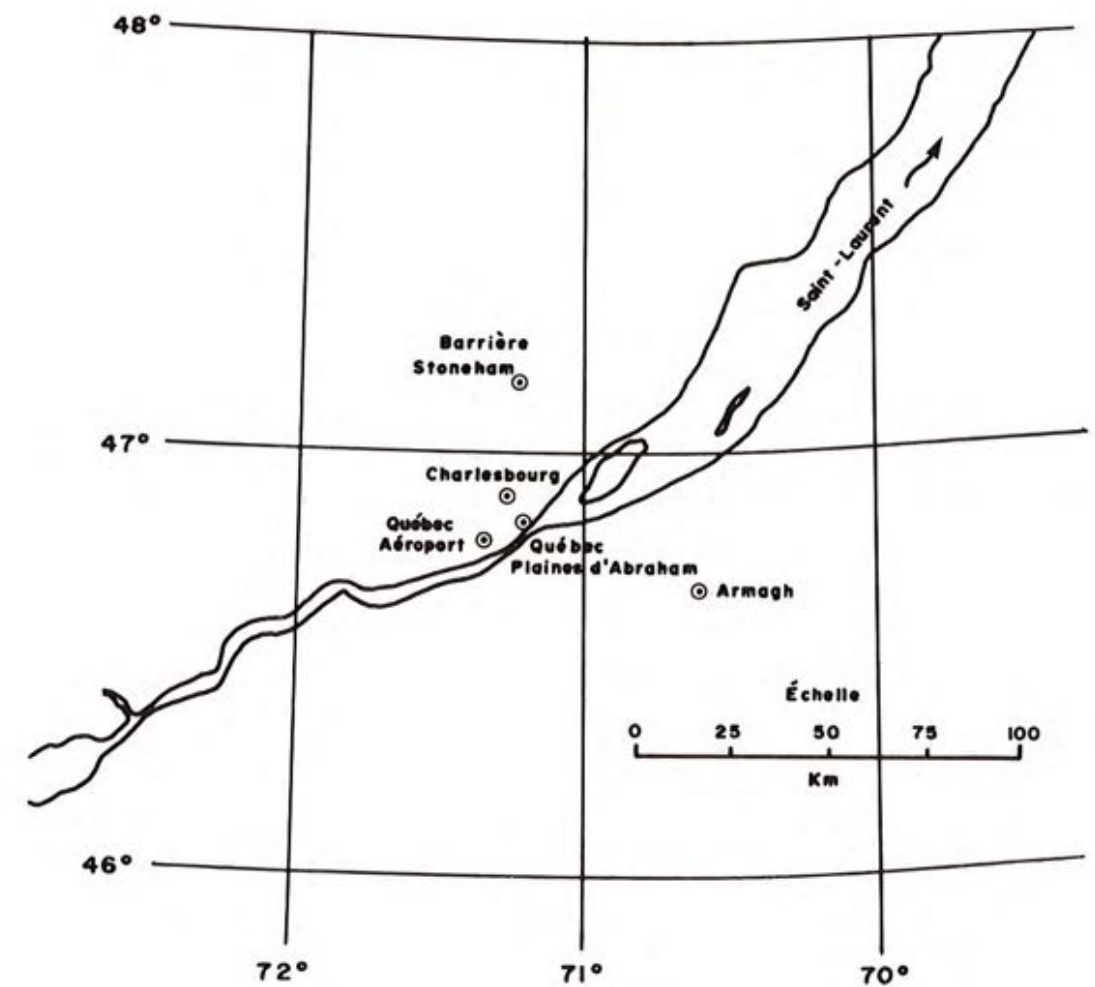
II. MODÈLE DE TEMPÉRATURE DE L'AIR
À QUÉBEC

A. Données

La série des données de température de l'air dont la structure est analysée est celle qui est enregistrée en permanence depuis 1943 à la station de l'aéroport de l'Ancienne Lorette (latitude Nord $46^{\circ}48'$, longitude Ouest $71^{\circ}23'$, altitude 245 pieds), près de Québec (figure 1). L'aéroport est situé en rive gauche du Saint-Laurent, entre les premiers contreforts du massif des Laurentides et le fleuve lui-même.

Dans la région de Québec, les enregistrements ont débuté bien avant l'implantation de l'aéroport. La station de

Fig. 1 Situation des stations météorologiques



l'Observatoire de Québec est restée de 1874 à 1953 dans le secteur du Parc des Plaines d'Abraham, au centre de la haute ville et au bord des falaises dominant le Saint-Laurent [Service Météorologique du Canada, 1971]. Jusqu'en 1958, la station aurait été exploitée dans d'autres secteurs de la ville [Villeneuve, 1970]. Cette longue série additionnelle est très précieuse pour étudier l'évolution à long terme de la température entre 1874 et 1973.

Quelques stations voisines ayant des enregistrements couvrant plusieurs des dernières décennies ont fourni des informations supplémentaires utiles pour évaluer la tendance générale dans la région de Québec (voir figure 1).

Stations	Latitude nord	Longitude ouest	Altitude	Période
Armagh	46°45'	70°35'	1150 pi	1916-1970
Barrière Stoneham	47°10'	71°15'	1960 pi	1933-1973
Charlesbourg (Jardin Zoologique)	46°54'	71°17'	500 pi	1946-1973
Québec Aéroport	46°48'	71°23'	245 pi	1944-1973
Québec Plaines	46°48'	71°13'	296 pi	1874-1953

Le Service de l'Environnement Atmosphérique du Canada, qui exploite les stations météorologiques attenantes aux aéroports, fournit les températures minimum et maximum quotidien-

nes sur bande magnétique. La température moyenne journalière, définie par la demi-somme du minimum et du maximum, a été calculée et analysée entre le 1er janvier 1944 et le 31 décembre 1971 (28 années). Pour simplifier les calculs, le 29ième jour du mois de février des années bissextiles a été volontairement omis. Les tableaux 1 et 2 présentent les données à long terme des deux stations successives de Québec, soit les moyennes annuelles (de 1874 à 1975) et les moyennes mensuelles (de 1944 à 1975). Les températures de 1972-73 et de 1974-75 n'ayant été disponibles qu'en cours d'étude ou que dernièrement, n'ont été utilisées que partiellement.

B. Tendance à long terme

La recherche d'une tendance à long terme dans les fluctuations de la température moyenne annuelle dans la région de Québec a déjà été abordée par Villeneuve (1970). Cette étude montre que, entre 1901 et 1968, l'effet de l'urbanisation de la région n'est pas la seule cause au réchauffement progressif, durant les mois d'hiver en particulier. Les mesures à Barrière Stoneham, station située environ à 30 km au nord de la ville, accusent en effet aussi une tendance ascendante, moins forte que celle observée en ville, mais assez sensible, alors que l'environnement de cette station n'a subi aucune modification depuis sa création. L'ordre de grandeur du gradient est de 1°C par 58 ans à Barrière Stoneham tandis qu'il atteint 1°C par 32 ans à Québec.

Tableau 1 - Température moyenne annuelle (°C)
à Québec-Plaines (1874-1953) et à
l'aéroport de l'Ancienne Lorette
(1944-1973)

année	Québec Plaines	année	Québec Plaines	année	Québec Plaines	Aéroport
1874	3.2			1944	5.3	4.9
1875	1.8	1910	4.4	1945	5.4	4.9
1876	3.6	1911	4.3	1946	5.4	4.7
1877	5.0	1912	3.2	1947	5.2	4.6
1878	5.3	1913	4.9	1948	6.5	4.8
1879	3.2	1914	3.4	1949	6.3	5.6
1880	4.1	1915	5.3	1950	5.1	4.5
1881	3.9	1916	4.1	1951	5.3	4.7
1882	3.7	1917	2.7	1952	6.2	5.3
1883	2.2	1918	3.6	1953	6.7	6.1
1884	3.0	1919	4.3	1954		4.7
1885	2.3	1920	4.4	1955		4.7
1886	3.1	1921	5.1	1956		4.1
1887	2.8	1922	4.7	1957		4.8
1888	2.3	1923	3.7	1958		4.2
1889	3.2	1924	3.9	1959		4.7
1890	3.1	1925	4.1	1960		5.3
1891	4.6	1926	3.3	1961		4.7
1892	4.1	1927	5.1	1962		3.9
1893	3.0	1928	4.0	1963		3.7
1894	4.2	1929	4.0	1964		4.3
1895	4.2	1930	4.9	1965		3.7
1896	3.7	1931	5.7	1966		4.9
1897	4.0	1932	4.7	1967		3.6
1898	4.9	1933	4.3	1968		3.8
1899	4.4	1934	3.9	1969		4.9
1900	4.4	1935	4.1	1970		4.2
1901	4.7	1936	4.3	1971		4.0
1902	4.3	1937	5.9	1972		2.8
1903	4.2	1938	5.4	1973		4.9
1904	2.7	1939	4.4	1974		3.3
1905	3.6	1940	4.3	1975		4.4
1906	4.2	1941	5.3			
1907	3.1	1942	5.4			
1908	4.1	1943	4.1			
1909	4.1					

Les résultats énoncés par l'auteur semblent en contradiction avec les observations de plusieurs autres auteurs qui admettent généralement que la tendance au réchauffement à la surface de la terre, amorcée vers 1890, s'est poursuivie jusqu'aux années 1945-50 pour se changer en tendance au refroidissement depuis. Lamb (1974) rapporte la courbe illustrant la variation de la température moyenne de l'air à la surface de la terre sous forme d'écarts par rapport à la moyenne entre 1880 et 1969 (d'après les calculs de Mitchell révisés au National Center for Atmospheric Research de Boulder, Colorado). Vers 1945 la température a subi une inversion de tendance et s'abaisse au moins jusqu'en 1970 à raison de -0.3°C par 30 ans. Lamb reproduit également les courbes calculées par Manley pour l'Angleterre et qui ont le même comportement au cours du dernier siècle.

Sur la courbe de Québec (figure 2) on remarque d'abord la non homogénéité de la série due au déplacement de station. Durant la période commune l'écart est sensiblement constant, la température étant plus chaude de 0.8°C environ en ville (tableau 1). Ce décalage ne masque pas le changement de pente générale vers 1950 qui marque le début du refroidissement. Dans la suite on considère une tendance uniforme à partir de 1951. Sur la figure 2 on peut également voir que la température annuelle aux autres stations de la région est stationnaire, lorsqu'elle n'est pas à la baisse, à partir de 1950.

	JAN.	FÉV.	MARS	AVRIL	MAI	JUIN	JUIL.	AOÛT	SEPT.	OCT.	NOV.	DÉC.	MOY.
1944	- 9.4	-11.8	-6.6	1.8	13.7	17.1	19.9	20.4	15.1	5.4	2.0	-10.1	4.9
45	-13.7	- 8.7	-0.4	7.1	10.2	15.8	19.8	18.3	14.1	5.6	-0.6	- 9.5	4.9
46	-11.8	-13.0	0.1	2.2	10.4	15.8	18.6	17.2	15.3	8.8	-0.4	- 8.7	4.7
47	-11.7	-10.6	-2.8	-0.5	8.1	15.9	20.7	20.4	13.6	10.5	-0.1	- 9.9	4.6
48	-13.3	-13.7	-6.2	3.3	11.1	15.8	19.6	19.6	14.8	7.2	3.4	- 5.4	4.8
49	- 7.6	- 9.7	-4.6	5.6	11.2	18.1	20.5	18.9	13.1	9.3	-2.3	- 6.3	5.6
1950	- 8.9	-12.4	-7.3	2.8	11.9	16.0	19.2	17.3	11.4	7.4	2.2	- 6.7	4.5
51	-10.3	- 8.8	-3.4	4.6	11.7	16.1	19.0	16.9	12.9	7.5	-2.3	- 8.8	4.7
52	-11.1	- 8.5	-3.0	3.9	10.4	17.2	21.8	18.6	13.3	5.0	0.2	- 5.8	5.3
53	- 9.2	- 8.0	-3.5	5.7	12.0	17.5	19.4	18.1	13.9	8.1	3.2	- 4.3	6.1
54	-13.6	- 5.5	-4.5	3.1	10.0	16.2	17.0	16.9	12.2	7.8	1.3	- 6.6	4.7
55	-11.6	- 9.7	-5.8	4.0	12.2	17.7	20.7	19.8	11.9	7.4	-0.1	-11.7	4.7
56	- 6.7	- 8.4	-7.1	2.9	7.9	15.4	17.3	17.1	11.1	7.6	0.1	- 8.5	4.1
57	-15.5	- 8.1	-2.8	4.2	10.8	16.3	18.6	16.1	13.6	8.1	1.9	- 5.9	4.8
58	- 7.9	-10.9	0.1	4.8	9.8	13.2	18.4	17.5	13.6	6.6	0.3	-14.7	4.2
59	-11.9	-14.1	-6.5	3.6	13.6	15.9	22.1	18.5	15.0	6.5	-0.7	- 7.1	4.7
1960	-11.6	- 5.8	-5.2	3.1	13.8	17.3	18.7	18.8	13.9	6.9	2.2	- 9.3	5.3
61	-15.2	- 9.8	-4.8	3.5	10.0	15.9	19.2	18.1	15.6	8.2	1.6	- 6.6	4.7
62	-13.6	-14.0	-1.9	3.4	11.1	17.1	16.7	17.8	12.0	6.9	-0.7	- 9.4	3.9
63	-11.5	-14.6	-6.3	3.2	10.3	17.4	19.8	15.4	11.1	9.3	1.8	-13.1	3.7
64	-10.1	-10.2	-4.2	3.7	11.8	16.3	19.2	16.0	11.8	5.4	-1.3	- 6.9	4.3
65	-13.6	-12.2	-3.9	3.5	10.7	16.6	17.1	16.7	12.8	5.1	-2.7	- 7.2	3.7
66	-10.2	- 9.0	-3.1	3.9	9.7	17.1	19.5	17.3	11.5	6.6	1.8	- 6.9	4.9

Tableau 2 - Températures mensuelles à l'aéroport de Québec (°C)

	JAN.	FÉV.	MARS	AVRIL	MAI	JUIN	JUIL.	AOÛT	SEPT.	OCT.	NOV.	DÉC.	MOY.
1967	- 9.6	-15.8	-8.3	1.6	7.5	17.3	20.0	18.2	13.0	6.7	-1.3	- 7.8	3.6
68	-15.6	-14.5	-4.1	4.8	10.4	14.9	18.7	15.4	15.4	9.0	-1.6	- 9.0	3.8
69	- 8.5	- 7.0	-4.2	1.8	4.9	16.3	18.2	19.2	12.1	6.2	1.3	- 7.2	4.9
1970	-15.4	-11.6	-4.2	3.4	10.4	16.2	20.5	18.6	11.9	8.7	1.2	-10.9	4.2
71	-15.0	-10.0	-4.3	2.4	11.1	16.1	18.2	16.9	14.6	8.8	-1.3	-10.4	4.0
72	-12.5	-13.8	-7.5	1.1	11.3	15.5	18.8	16.1	12.7	3.9	-1.9	-10.9	2.8
73	-11.1	-11.8	-0.9	3.7	9.7	17.8	20.7	19.7	11.8	7.5	-1.3	- 6.4	4.9
74	-12.9	-12.8	-6.5	1.7	7.7	16.8	19.2	17.9	11.2	3.6	-0.4	- 5.8	3.3
75	-11.4	-10.5	-5.4	1.7	12.4	17.8	21.1	18.2	12.3	7.1	0.8	-11.5	4.4

Tableau 2 (suite) - Températures mensuelles à l'aéroport de Québec (°C)

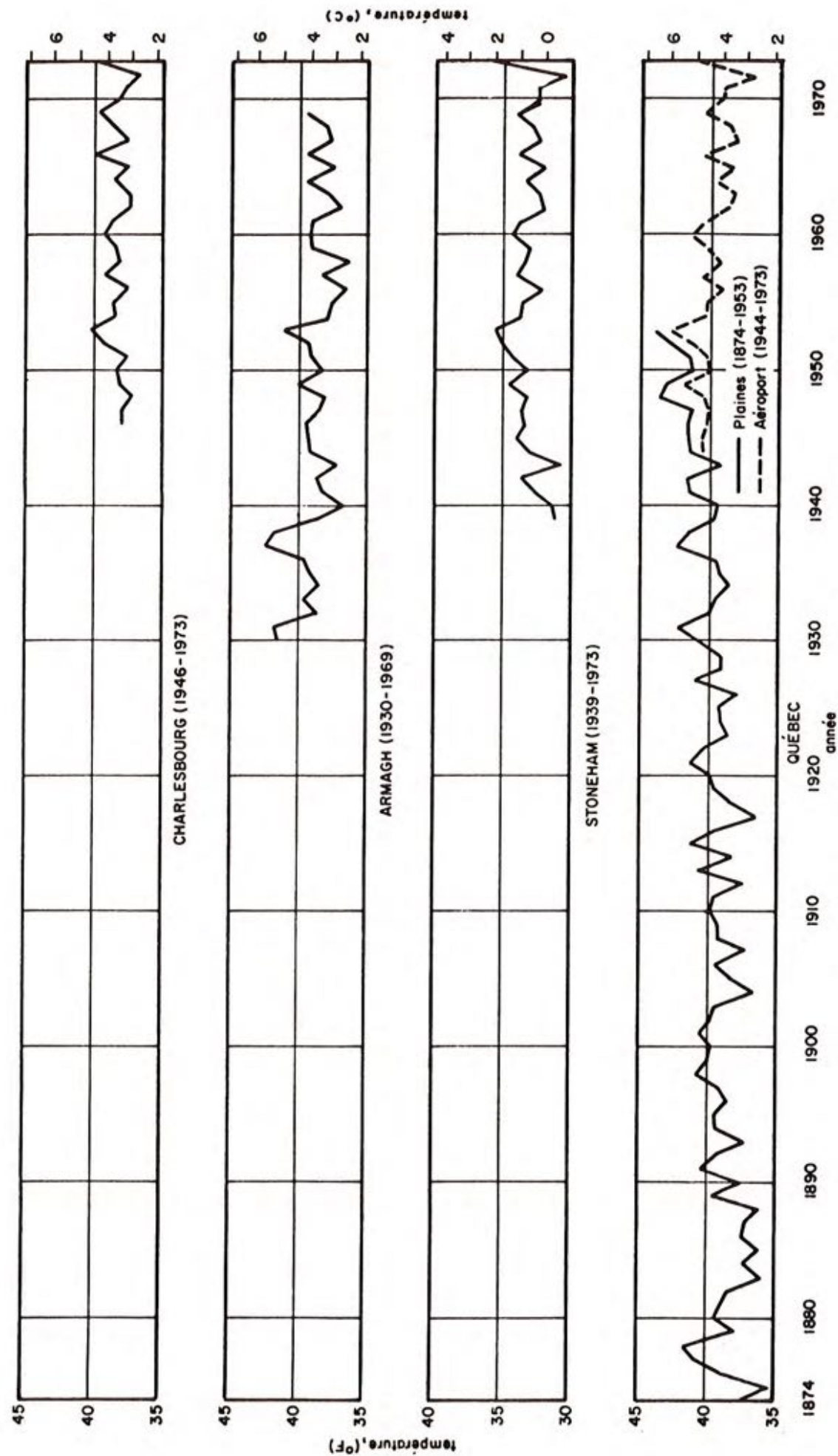


Fig. 2 — TEMPÉRATURE MOYENNE ANNUELLE DANS LA RÉGION DE QUÉBEC.

La période s'étendant sur 23 années entre 1951 et 1973 est relativement courte pour estimer une tendance. Cependant, d'après le test de Spearman, l'hypothèse de la tendance n'est pas à rejeter au niveau de 94%. Un premier lissage par moyenne mobile de 5 années met en évidence cette tendance qui est finalement approximée par une droite:

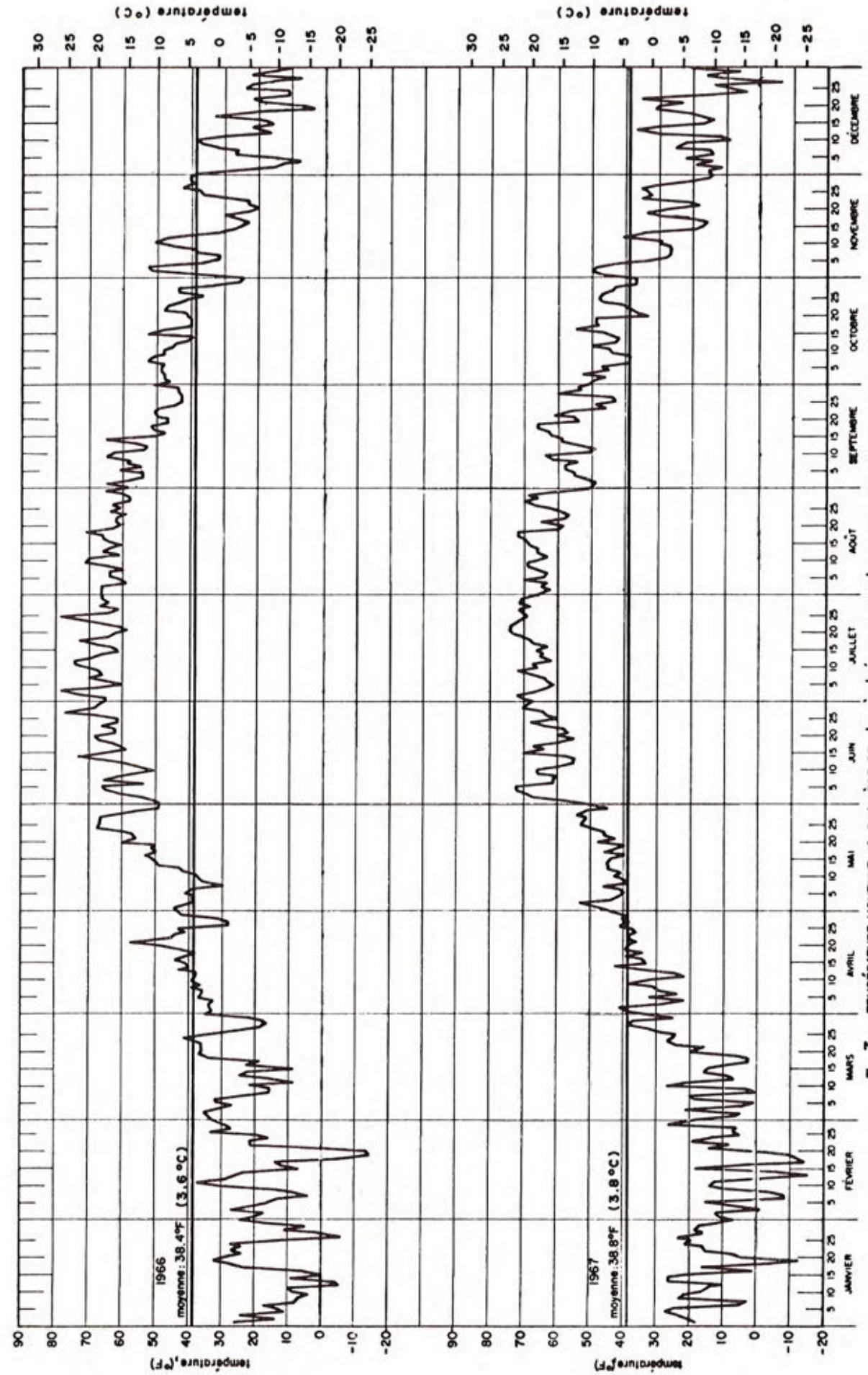
$$Td_A = 5.11 - 0.0586 (A-1950) \quad (1)$$

où A est l'année civile et Td_A est exprimée en °C. Le gradient de refroidissement est égal à -1°C par 18 ans. Pour les autres stations il est très variable, du simple à Stoneham (-1°C par 20 ans) au tiers à Charlesbourg (-1°C par 56 ans) et intermédiaire à Armagh (-1°C par 43 ans).

On peut vérifier aussi qu'en première approximation, compte tenu de la taille de l'échantillon, l'écart de la moyenne annuelle par rapport à la tendance linéaire est aléatoire et normalement distribué.

C. Cycle saisonnier

La figure 3 illustre bien la variation saisonnière de la température journalière sur laquelle se greffent les oscillations irrégulières dues aux épisodes météorologiques courts. L'allure sinusoïdale du cycle saisonnier est assez marquée.



Les 28 séries annuelles de 365 valeurs ont été soumises à l'analyse de Fourier. Le tableau 3 récapitule les caractéristiques de la composante fondamentale et deux premières harmoniques. L'amplitude, la phase de la fondamentale ainsi que le pourcentage de variance expliquée par cette dernière présentent une certaine stabilité autour de leurs moyennes qui sont respectivement 15.4°C, 114 jours et 85.8%. Les écarts maxima ne dépassent pas 10%. En ce qui concerne les harmoniques, les variations des paramètres sont plus fortes, mais l'ordre de grandeur par rapport à ceux de la fondamentale est conservé au cours des années. On peut donc conclure qu'un cycle saisonnier se répète d'année en année presque identique à lui-même. Pour éliminer les fluctuations aléatoires des paramètres on peut calculer le cycle moyen interannuel qui apparaît en figure 4. Il est centré sur zéro puisque à chaque série annuelle a d'abord été soustraite sa moyenne. L'analyse harmonique du cycle expérimental T_{sm} montre que la fondamentale de la série de Fourier explique 98.84% de la variance. Cependant, afin que la déviation par rapport au cycle saisonnier présente un caractère stationnaire acceptable (figure 5), il faut incorporer à la définition du cycle saisonnier lissé \tilde{T}_{sm_j} (°C) les deux premières harmoniques:

$$\begin{aligned} \tilde{T}_{sm_j} = & 12.4 \sin \left[\frac{2\pi}{365} (j - 113.7) \right] \\ & + 0.4 \sin \left[\frac{2\pi}{365} (2j - 143.2) \right] \\ & + 0.6 \sin \left[\frac{2\pi}{365} (3j - 139.3) \right] \end{aligned} \quad (2)$$

$j = 1, \dots, 365$

Avec cette représentation, la variance totale est expliquée à 99.50%.

Conjointement au calcul du cycle saisonnier, il est intéressant de calculer l'écart-type associé à chaque échantillon journalier (figure 6). Ce paramètre varie aussi de façon saisonnière. L'analyse de Fourier montre que la fondamentale explique 64% de la variance, et que les trois premières composantes réunies permettent d'atteindre 80.6%. La formulation de la représentation lissée $\tilde{s}m$ est la suivante:

$$\begin{aligned} \tilde{s}m = & 7.8 + 1.5 \sin \left[\frac{2\pi}{365} (j - 107.2) \right] \\ & + 0.6 \sin \left[\frac{2\pi}{365} (2j - 132.1) \right] \\ & + 0.5 \sin \left[\frac{2\pi}{365} (3j - 161.0) \right] \end{aligned} \quad (3)$$

$j = 1, \dots, 365$

Durant la période chaude, soit d'avril à octobre, l'écart-type est sensiblement constant, tandis qu'en hiver il s'accroît jusqu'au double du palier estival. Cette constatation est très conforme à la réalité puisque, pour un jour donné du calendrier, il est probable que la température soit élevée en été et sensiblement la même d'une année à l'autre, alors qu'en hiver on peut avoir des écarts considérables. Par exemple le 1er février 1970, la température journalière a été de -3°C et le 1er février 1971 de -24°C .

L'estimateur de l'écart-type, même issu d'un estimateur

Année	Fondamentale			1er Harmonique			2è Harmonique		
	Amplitude (°C)	Phase (jours)	Variance expliquée %	Ampli. (°C)	Phase (jours)	Variance expliquée %	Ampli. (°C)	Phase (jours)	Variance expliquée %
1944	16.7	114.1	86.3	0.9	182.8	0.3	0.1	271.2	0.0
1945	15.6	107.8	85.5	2.1	272.2	1.6	1.9	156.6	1.3
1946	15.3	115.1	83.6	1.8	122.9	1.2	1.1	118.5	0.4
1947	16.2	118.7	86.3	1.3	222.9	0.6	2.3	96.8	1.7
1948	16.2	119.3	87.2	2.1	163.2	1.4	1.9	200.7	1.3
1949	15.2	112.3	85.5	0.6	158.0	0.1	0.6	251.5	0.1
1950	15.1	116.4	84.0	2.3	202.1	2.0	0.7	236.4	0.2
1951	14.8	110.3	83.0	1.2	123.5	0.5	0.3	96.1	0.0
1952	15.1	111.4	89.3	0.5	243.9	0.1	1.1	167.1	0.5
1953	14.1	114.2	83.5	1.5	169.3	0.9	1.3	178.1	0.7
1954	14.2	114.2	83.7	0.9	142.0	0.4	1.2	99.6	0.6
1955	16.7	111.1	89.0	0.9	142.7	0.3	0.9	109.0	0.3
1956	13.7	116.5	83.9	0.5	91.8	0.1	0.7	198.7	0.2
1957	14.7	113.2	83.2	1.8	135.8	1.3	1.1	113.6	0.4
1958	14.6	109.2	82.7	2.4	252.0	2.3	0.9	93.9	0.3
1959	16.9	114.6	86.6	1.3	177.7	0.5	0.9	239.5	0.3
1960	15.3	112.4	87.2	0.9	139.8	0.3	1.1	234.3	0.4
1961	15.8	117.5	88.2	1.8	128.6	1.1	1.2	175.2	0.5
1962	15.9	111.4	86.1	2.2	137.5	1.6	1.1	175.5	0.4
1963	16.4	114.8	86.4	2.5	169.8	2.0	1.2	118.2	0.5
1964	14.5	109.8	85.6	0.8	199.1	0.3	0.5	145.9	0.1
1965	15.2	110.8	87.4	0.9	157.9	0.4	0.3	193.8	0.0
1966	14.3	112.7	84.5	0.6	207.7	0.1	1.1	122.0	0.5
1967	16.3	118.1	88.0	1.3	219.6	0.5	0.8	238.4	0.2
1968	16.3	114.0	84.8	3.2	128.3	3.0	0.4	146.8	0.1
1969	14.2	114.5	85.9	0.3	178.4	0.0	0.4	266.2	0.1
1970	16.6	113.7	88.3	1.8	132.3	1.0	1.3	94.2	0.6
1971	16.1	113.6	87.2	2.1	117.9	1.4	1.1	97.9	0.4

Tableau 3 - Caractéristiques des composantes fondamentales et harmoniques

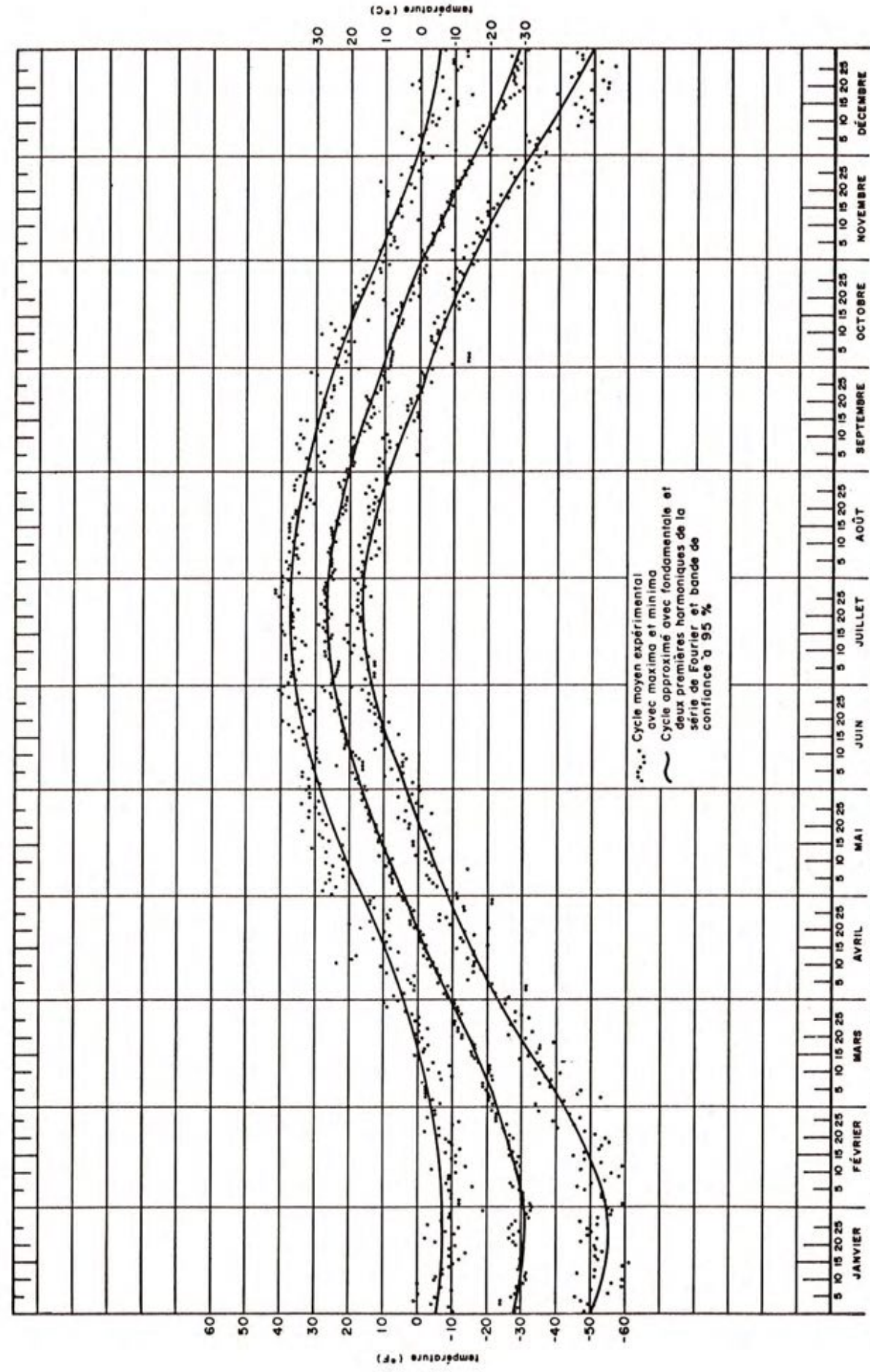


Fig. 4 — CYCLE SAISONNIER MOYEN INTERANNUEL (28 ANNÉES) DE LA DÉVIATION DE LA TEMPÉRATURE JOURNALIÈRE PAR RAPPORT À LA MOYENNE ANNUELLE.

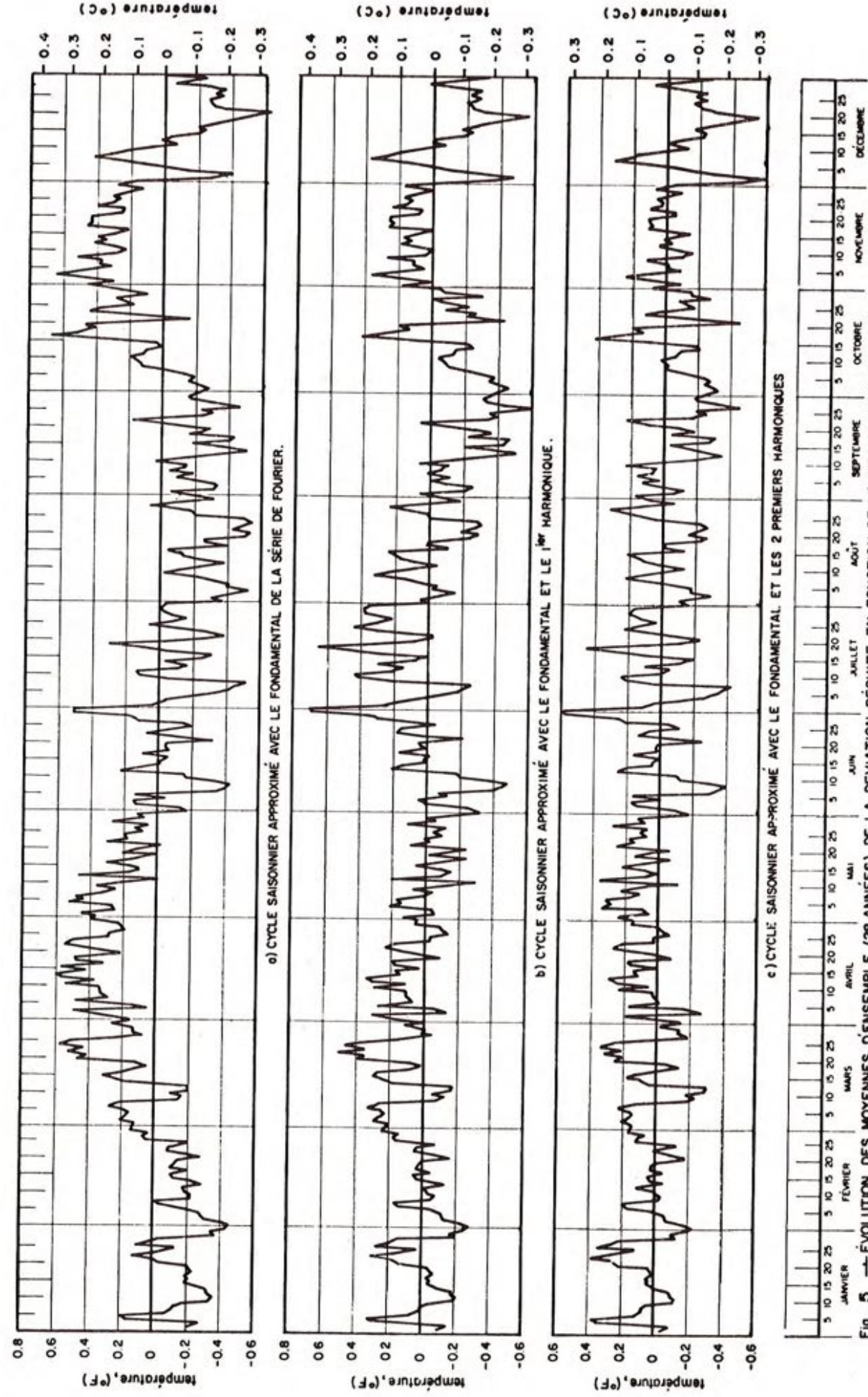


Fig. 5 — ÉVOLUTION DES MOYENNES D'ENSEMBLE (28 ANNÉES) DE LA DÉVIATION RÉDUITE EN FONCTION DE LA REPRÉSENTATION DU CYCLE SAISONNIER.

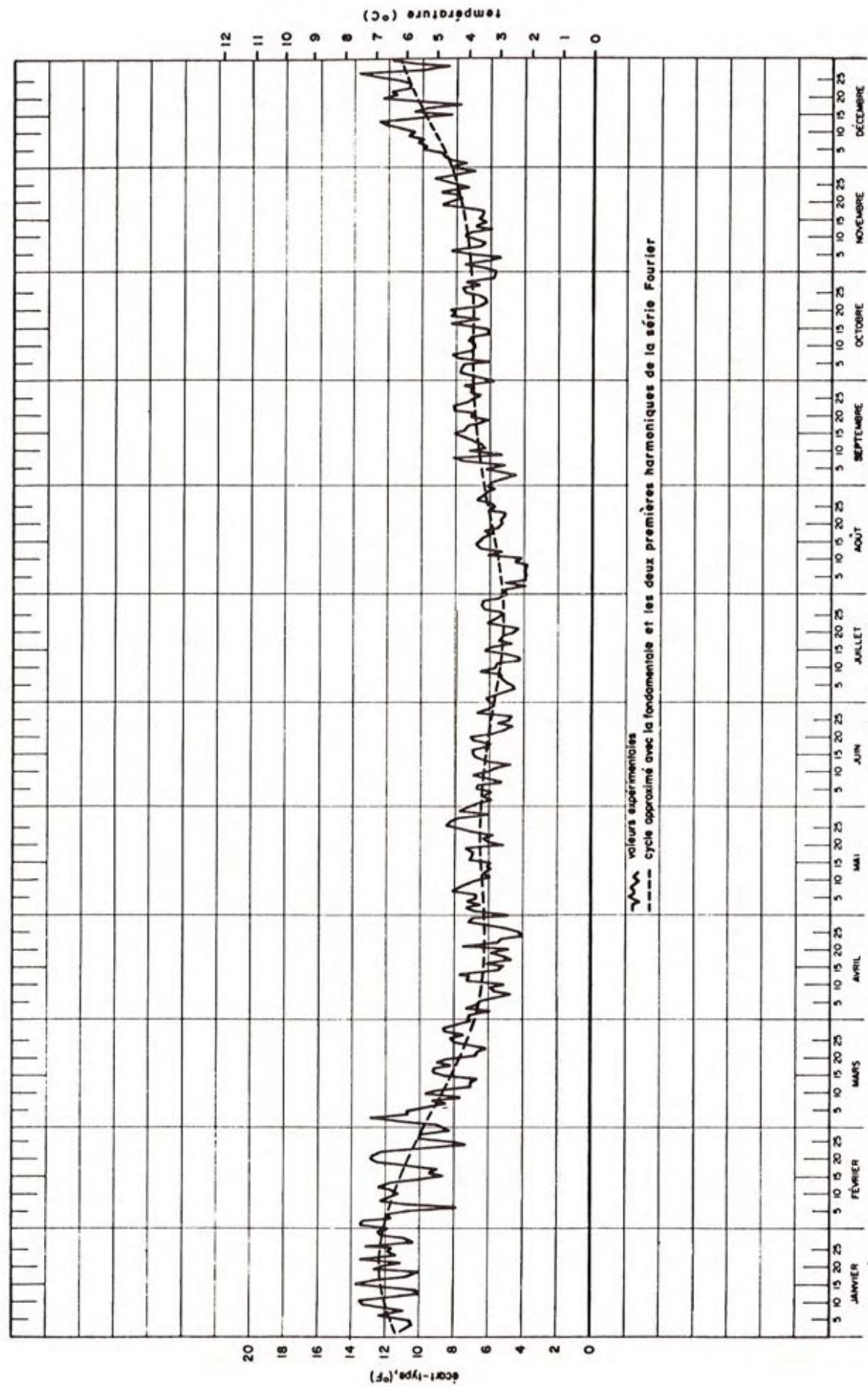


Fig. 6 — ÉCART-TYPE PAR RAPPORT AU CYCLE SAISONNIER MOYEN INTERANNUEL (28 ANNÉES) DE LA DÉVIATION DE LA TEMPÉRATURE, JOURNALIÈRE PAR RAPPORT À LA MOYENNE ANNUELLE.

non biaisé de la variance, n'est pas non biaisé. Pour être consistant dans le calcul des estimateurs, il faut adopter un estimateur non biaisé de l'écart-type. Cela a surtout une importance lorsque la taille de l'échantillon est petite. Gurland et Tripathi (1971) donnent l'expression du coefficient correcteur C_N , qui tend vers 1 par valeur supérieure lorsque la taille N de l'échantillon croît, à appliquer à l'écart-type s de sorte que:

$$s = C_N \sqrt{\frac{1}{N-1} (x_i - \bar{x})^2} \quad (4)$$

où \bar{x} est la moyenne des observations x_i , $i = 1, N$.

La distribution statistique de la déviation par rapport à la moyenne annuelle, pour un jour calendaire donné, est une caractéristique importante à considérer. En première approximation et en dépit de la petite taille (28) des 365 échantillons, les tests sur les coefficients d'asymétrie et d'aplatissement (Pearson et Hartley, 1954) ne permettent pas de rejeter l'hypothèse de normalité dans la plupart des cas.

Connaissant l'écart-type afférent à chaque jour, les bandes de confiance à 95% peuvent être tracées à une distance de 1.96 ssm (figure 4). On peut constater que, pour 28 années, les minima observés en hiver et les maxima en été sont souvent à l'extérieur de la bande. Ceci est en accord avec la figure 9 qui révèle une légère asymétrie, soit à gauche, soit à droite, mais qui n'est généralement pas significative.

D. Température résiduelle ou déviation

1. Identification et stationnarité

La température résiduelle ou déviation e_{ij} est définie ainsi:

$$e_{ij} = (T_{ij} - T_{ma_i}) - \tilde{T}_{sm_j} \quad (5)$$

$$i = 1, \dots, Na, j = 1, \dots, 365$$

où T_{ij} = la température observée le jour j de l'année i

T_{ma_i} = la moyenne annuelle de l'année i

\tilde{T}_{sm_j} = la moyenne du jour j (dite température "normale") issue du cycle saisonnier lissé.

Par définition de \tilde{T}_{sm} , la moyenne d'ensemble \bar{e}_j reste toujours voisine de zéro. Le processus $\{e\}$ est au moins stationnaire en moyenne (figure 7). Par contre en raison du comportement saisonnier de ssm , le processus n'est pas stationnaire au second ordre. La variance d'ensemble varie avec la saison. Pour réaliser la stationnarité au second ordre, laquelle permet la poursuite de l'analyse par les méthodes simples, la déviation e_{ij} est réduite par son écart-type associé \tilde{ssm}_j (Roesner et Yevjevich, 1966). La déviation centrée réduite est alors:

$$Y_{ij} = \frac{(T_{ij} - T_{ma_i}) - T_{sm_j}}{\tilde{ssm}_j} \quad (6)$$

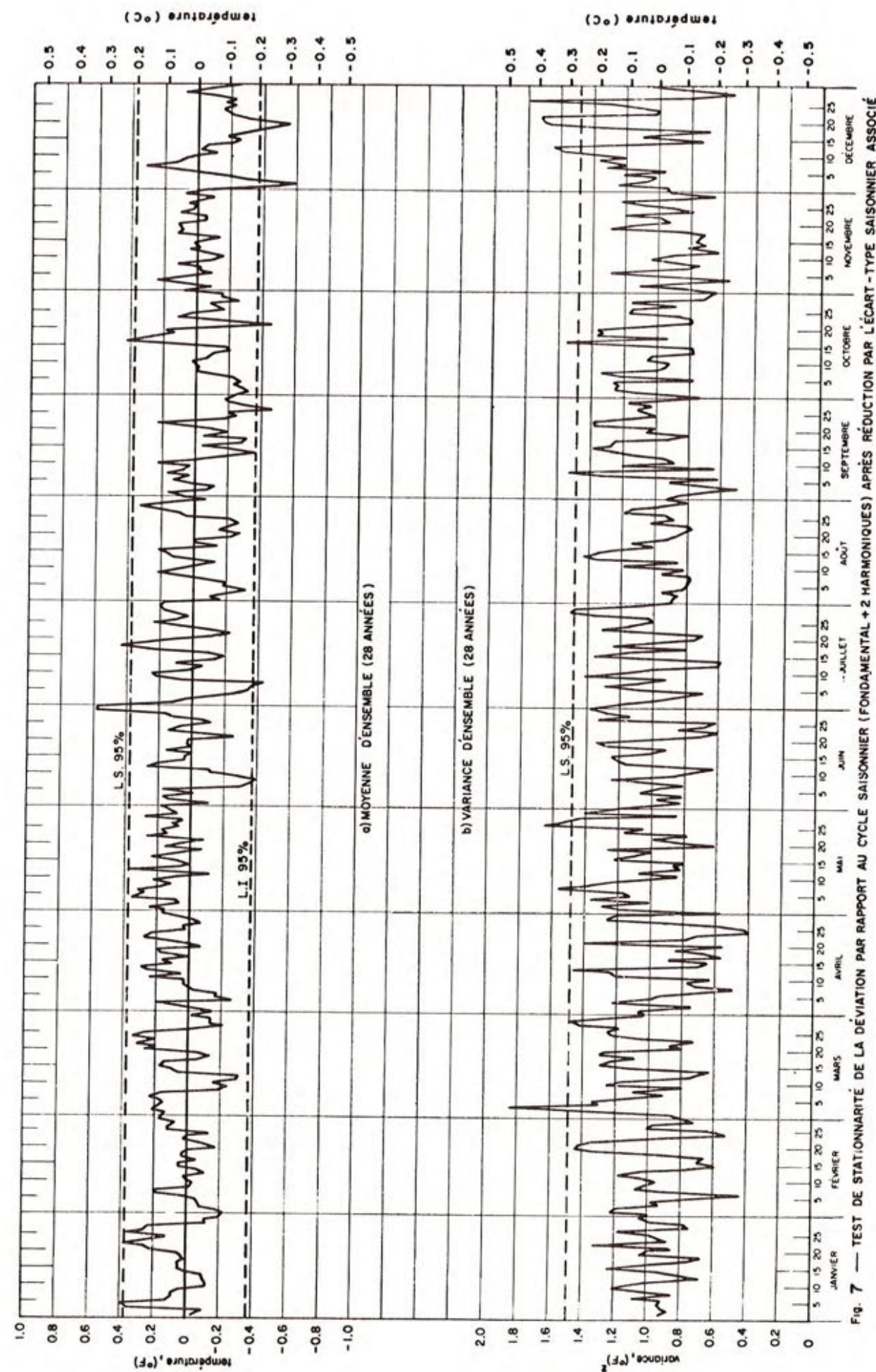


Fig. 7 — TEST DE STATIONNARITÉ DE LA DÉVIATION PAR RAPPORT AU CYCLE SAISONNIER (FONDAMENTAL + 2 HARMONIQUES) APRÈS RÉDUCTION PAR L'ÉCART - TYPE SAISONNIER ASSOCIÉ

Le test de stationnarité consiste alors à vérifier que moyenne et variance d'ensemble, estimées pour chaque jour, ne diffèrent pas simultanément de façon significative des valeurs respectives 0 et 1 (voir figure 7). La technique de ce test qui définit une région de confiance pour moyenne et variance est exposée dans Mood et Graybill (1963). Le test de stationnarité exige de prendre en compte, pour définir le cycle saisonnier, deux harmoniques en plus de la sinusoïde fondamentale. La figure 5 illustre le fait que l'oscillation plus ou moins périodique qui apparaît dans les deux premiers cas disparaît suffisamment avec le troisième schéma. Afin d'éviter un modèle trop lourd qui engloberait plus d'harmoniques on peut se limiter à cette approximation.

Le test de stationnarité peut être plus poussé si on tient compte des coefficients d'autocorrélation d'ensemble (figure 8) ou des moments d'ordre supérieur à deux de la distribution d'ensemble de la déviation centrée réduite. En raison de la petite taille des échantillons, il est un peu illusoire de calculer les coefficients d'asymétrie et d'aplatissement d'ensemble, même à l'aide d'estimateurs non biaisés. Cependant, ce n'est pas tant les valeurs absolues que les valeurs relatives des coefficients de forme qui sont intéressantes. La figure 9 permet en effet de remarquer que ces coefficients présentent de légères variations saisonnières qui traduisent la non stationnarité du processus à un ordre supérieur à 2. La variabilité au cours des saisons de

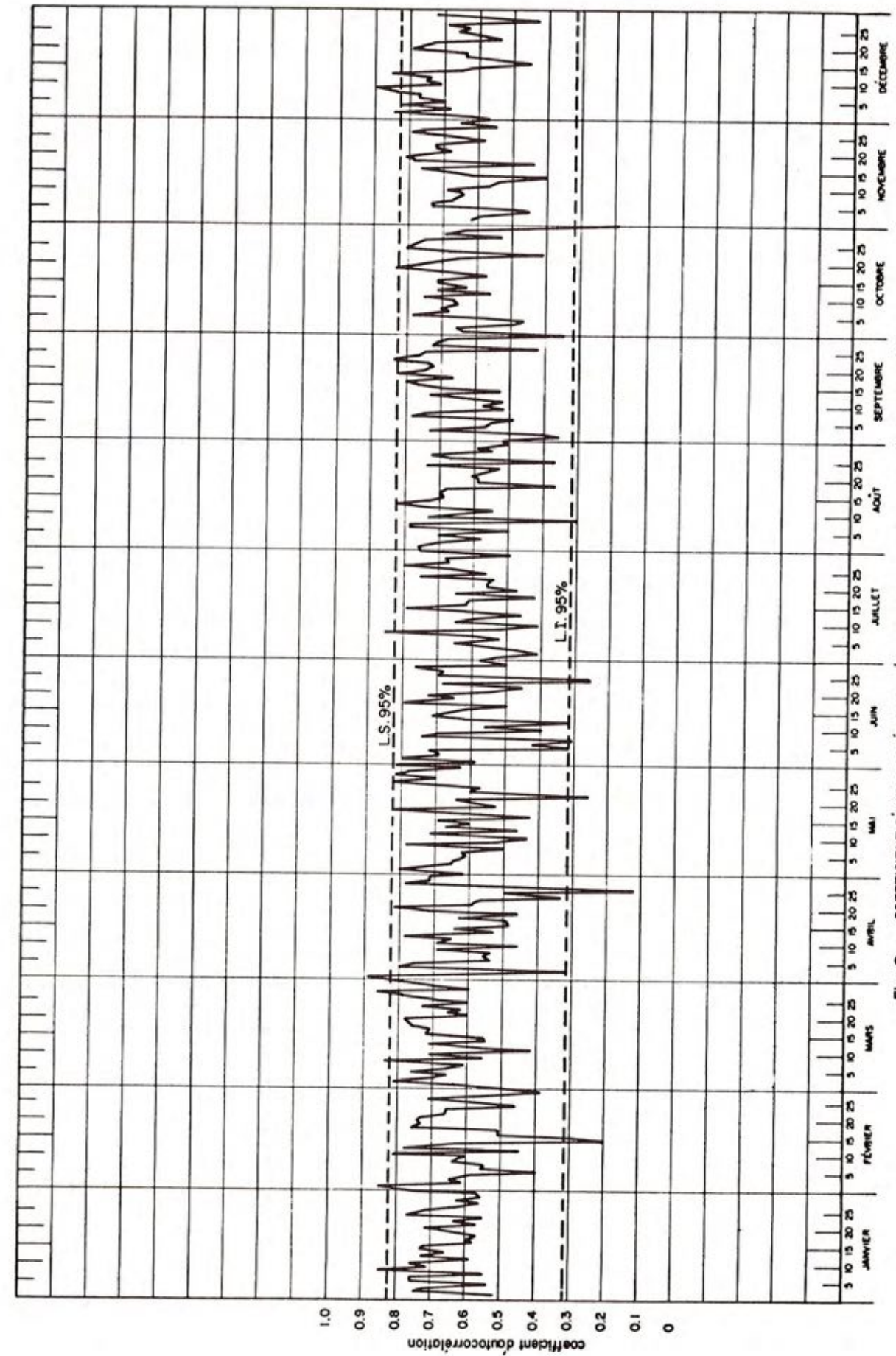


Fig. 8 — COEFFICIENT D'AUTOCORRÉLATION D'ENSEMBLE (log 1) DE LA DÉVIATION RÉDUITE.

transition, l'automne et le printemps surtout, apparaissent nettement en figure 9.

A l'aide des bandes de confiance, tracées pour 95%, on peut admettre en première approximation la normalité des distributions, à l'exception des périodes de transition.

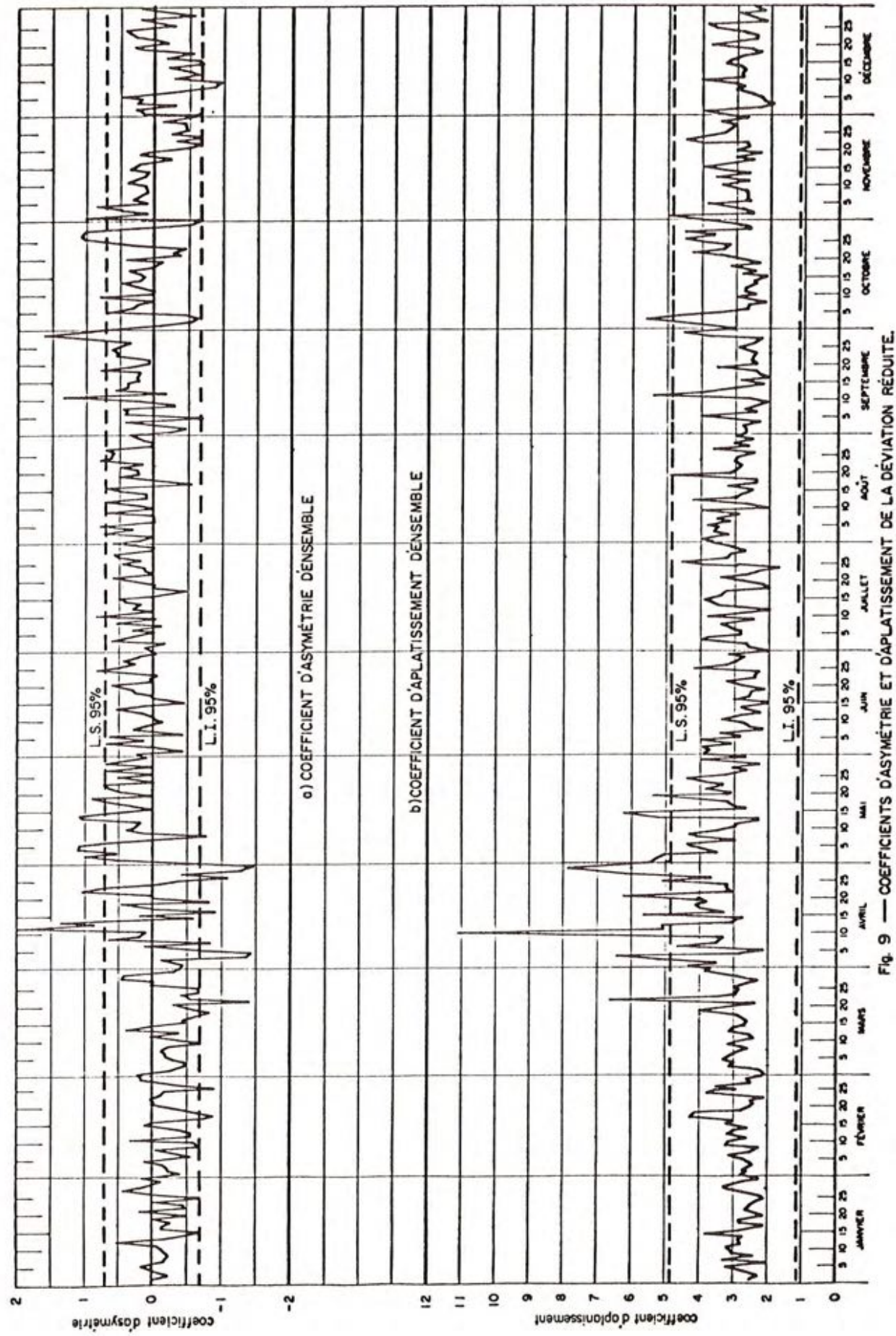
L'hypothèse de processus stationnaire permet de ne considérer qu'une seule population et non plus 365 populations. Pour obtenir une variable parfaitement centrée réduite de moyenne nulle et variance unitaire on peut effectuer la transformation:

$$z_{ij} = \frac{y_{ij} - \bar{y}}{s_y} \quad (7)$$

où \bar{y} et s_y sont respectivement la moyenne et l'écart-type de y .

2. Autocorrélation et résidu

Le degré de dépendance séquentielle de la déviation est évalué à l'aide de la fonction d'autocorrélation, surtout des premiers coefficients d'autocorrélation. La forme du corrélogramme est comparée à celle des corrélogrammes théoriques liés aux divers types de modèles autorégressifs, ou à moyenne mobile, ou mixtes. En figure 10 on peut voir que l'allure est à décroissance régulière, ce qui confirme l'existence d'une dépendance séquentielle à l'ordre 1 ou 2. De la même



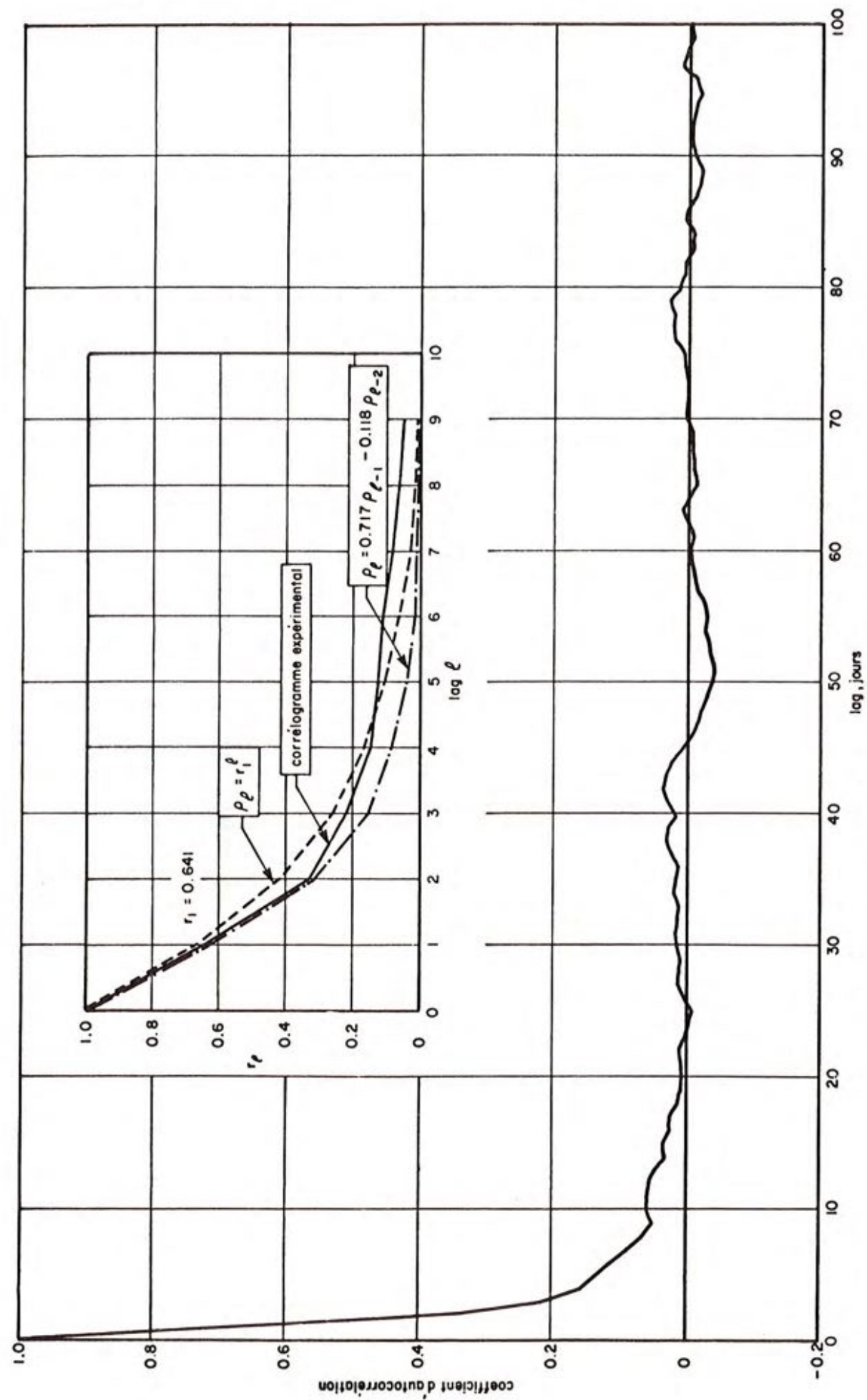


Fig. 10 — CORRÉLOGRAMME DES DÉVIATIONS JOURNALIÈRES STANDARDISÉES. (21 ANNÉES, N=7665).

façon, la fonction densité spectrale comparée aux densités théoriques des processus connus permet d'identifier le processus expérimental (figure 11). Là encore, la densité spectrale épouse bien les formes afférentes aux modèles autorégressifs des premier et deuxième ordres.

Pour un échantillon constitué de 7665 valeurs représentant les déviations journalières sur 21 années, les estimateurs des coefficients d'autocorrélation sont $r_1 = 0.641$ et $r_2 = 0.341$. La déviation journalière e_j peut alors s'exprimer par:

$$e_j = 0.641 e_{j-1} + \epsilon_j \quad (8)$$

ou
$$e_j = 0.717 e_{j-1} - 0.118 e_{j-2} + \epsilon_j \quad (9)$$

avec $j = 1, \dots, Na \times 365$

pour les modèles autorégressifs au premier et au deuxième ordre respectivement. Les coefficients de détermination sont respectivement 0.411 et 0.419. Ces deux valeurs mettent en évidence le fait que le modèle au deuxième ordre ne présente pratiquement pas d'avantage au sens de la variance expliquée. Pour le troisième ordre, le coefficient de détermination n'est que 0.422. Ceci illustre bien la notion de palier atteint par R^2 lorsque l'ordre du modèle s'élève (Fiering et Jackson, 1971). Ces résultats permettent d'en déduire aussi que la liaison en chaîne n'est pas très forte: seulement 42% de la variance expliquée contre 58% attribuée au résidu aléatoire ϵ .

Du résidu centré (moyenne nulle) ε , on peut passer à un résidu centré réduit η tel que:

$$\eta = \frac{\varepsilon}{\sqrt{1-R^2}} \quad (10)$$

Pour tester l'indépendance séquentielle de η , le coefficient d'autocorrélation de pas 1 est un bon indice. Si ε est représenté par le modèle au deuxième ordre (équation 45), le coefficient d'autocorrélation $r_{1\eta}$, pour un échantillon η de 7665 valeurs, est égal à 0.0085, valeur très faible qui ne peut mettre en doute le caractère aléatoire.

La fonction de répartition du résidu η est enfin l'élément essentiel à déterminer. La déviation e ayant été admise comme normalement distribuée, il devrait en être de même de ε et de η . Le coefficient d'asymétrie expérimental révèle une légère asymétrie négative (-0.285), mais le test de Kolmogorov-Smirnov indique que la population à laquelle appartient l'échantillon ne diffère pas significativement d'une population de référence normalement distribuée. Le résidu centré réduit η est donc considéré normalement distribué. Il peut être généré au moyen des techniques connues de génération des nombres pseudo-aléatoires sur les ordinateurs (Moshman, 1967) (Kite et Pentland, 1971).

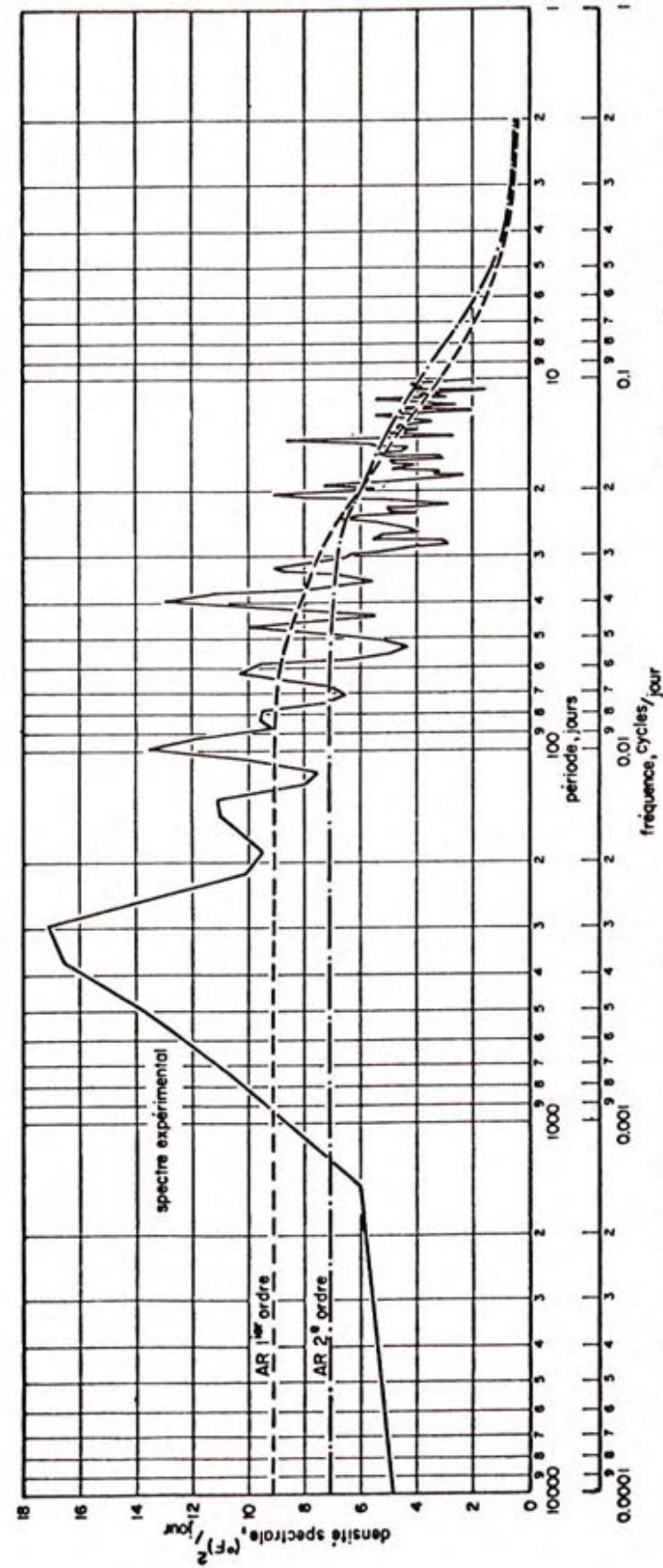


Fig. 11 — DENSITÉ SPECTRALE EXPÉRIMENTALE DE LA DÉVIATION STANDARDISÉE (N=7665, n=739) COMPARÉE À LA DENSITÉ SPECTRALE THÉORIQUE POUR LES PROCESSUS AUTORÉGRESSIFS. (du 1^{er} ordre: $Z_i = 0.641 Z_{i-1} + \varepsilon_i$, du 2^e ordre: $Z_i = 0.717 Z_{i-1} - 0.118 Z_{i-2} + \varepsilon_i$).

E. Validation du modèle de génération

L'analyse de la structure de la série a permis d'identifier quatre composantes: une tendance, un cycle saisonnier, une composante autorégressive et un résidu aléatoire. La génération et la superposition de ces composantes conduisent à une série "synthétique" qui n'est pas une restitution fidèle de la série observée, mais une série qui présente les mêmes caractéristiques statistiques et séquentielles. La série observée est une réalisation parmi toute la gamme des possibilités. Par l'utilisation du modèle qui permet de générer toute cette gamme, ou du moins un certain nombre de réalisations, on peut préciser les propriétés statistiques d'ensemble et en particulier la fonction de répartition de chaque variable: température moyenne mensuelle pour un mois donné, température journalière pour un jour donné, etc...

Dans l'application de ce type de modèle, il est fréquent de générer 100 séquences de 100 années chacune. En raison de la présence d'une tendance sur les 21 dernières années d'observation, la longueur de chaque séquence a été limitée à 21 années pour fin de validation. Dans le cadre d'une prévision, on pourrait faire une hypothèse quant à la forme future de la tendance. Pour diminuer le volume des calculs subséquents, le nombre de séquences a été limité à 30.

1. Comparaison des paramètres des séries observées et générées.

Les paramètres de la tendance, du cycle saisonnier, de la déviation et du résidu de chaque séquence générée ont été estimés en suivant la même procédure que pour ceux de la série observée. Le tableau 4 récapitule les résultats en présentant les valeurs caractéristiques d'un échantillon de 30 valeurs pour chaque paramètre. Le modèle introduit une composante aléatoire à deux niveaux: celui de la moyenne annuelle d'abord et celui de la déviation ensuite. On peut voir que cela n'entraîne pas une grande dispersion des paramètres du cycle saisonnier des moyennes. Cette dispersion est même plus petite que celle observée en réalité entre 1951 et 1971. Par contre le cycle des écarts-types est dans l'ensemble moins bien reproduit. Le pourcentage de variance expliquée à l'aide de la fondamentale et de deux harmoniques est égal à 58% en moyenne contre 64% d'après la série observée. Le modèle tend donc à engendrer une plus grande variance que la réalité. Les paramètres séquentiels de la déviation sont assez bien restitués.

Pour illustrer la représentativité et évaluer le degré de vraisemblance des séquences générées, on a comparé l'une d'entre elles à la série observée au point de vue des moyennes annuelles (figure 12), des moyennes mensuelles (figure 13), des températures journalières (figure 14). Dans ce dernier cas la position chronologique de l'année générée

Tableau 4 - Comparaison des paramètres de la série observée avec les paramètres des séquences générées

Paramètres	Série observée*	30 séquences générées de 21 années					
		Min.	Moy.	Max.	Ec. type	Coef. Var.	
A. Composante saisonnière							
1. Cycle des moyennes (approx. sinus.)							
- fondamental	: amplitude (°C)	15.4	15.0	15.0	15.9	0.2	0.011
	: déphasage (jours)	113.7	113.7	114.9	115.7	0.4	0.004
	: variance expliquée	98.84	98.13	98.56	98.78	0.16	0.002
- 1er harmonique	: amplitude	1.1	0.9	1.2	1.4	0.1	0.105
	: déphasage	143.2	130.1	144.8	155.9	6.4	0.044
	: variance expliquée	0.50	0.32	0.57	0.77	0.11	0.192
- 2ème harmonique	: amplitude	0.6	0.4	0.7	0.9	0.1	0.202
	: déphasage	139.3	115.0	141.1	168.0	12.4	0.088
	: variance expliquée	0.16	0.07	0.18	0.36	0.07	0.400
2. Cycle des écarts-types (approx. sinus.)							
- fondamental	: amplitude	1.5	1.3	1.4	1.6	0.1	0.058
	: déphasage	107.1	103.8	106.8	112.2	2.3	0.022
	: variance expliquée	64.08	50.92	58.15	62.88	2.97	0.051
- 1er harmonique	: amplitude	0.6	0.4	0.6	0.8	0.1	0.012
	: déphasage	132.0	121.8	134.9	149.7	7.1	0.053
	: variance expliquée	10.40	6.60	9.75	13.68	1.75	0.179
- 2ème harmonique	: amplitude	0.5	0.3	0.5	0.6	0.1	0.144
	: déphasage	160.9	143.1	162.6	188.3	9.3	0.057
	: variance expliquée	6.58	3.21	6.80	9.83	1.6	0.244
B. Déviation journalière par rapport à la composante saisonnière							
1. Coefficients d'autocorrélation							
- lag 1 (r_1)		0.641	0.625	0.641	0.653	0.008	0.012
- lag 2 (r_2)		0.341	0.314	0.331	0.355	0.011	0.033
2. Coefficients du modèle autorégressif (2ème ordre)							
- a_1		0.717	0.702	0.727	0.750	0.012	0.016
- a_2		-0.118	-0.115	-0.134	-0.154	0.009	0.067
- coefficient de détermination		0.419	0.400	0.421	0.437	0.014	0.025
3. Composante résiduelle aléatoire							
- coefficient d'autocorrélation (lag 1)		0.0085	-0.0048	-0.0006	-0.0031	0.0015	25.347

* Paramètres de la composante saisonnière estimés sur 28 années (1944-71)
 Paramètres de la déviation journalière estimés sur 21 années (1951-71)

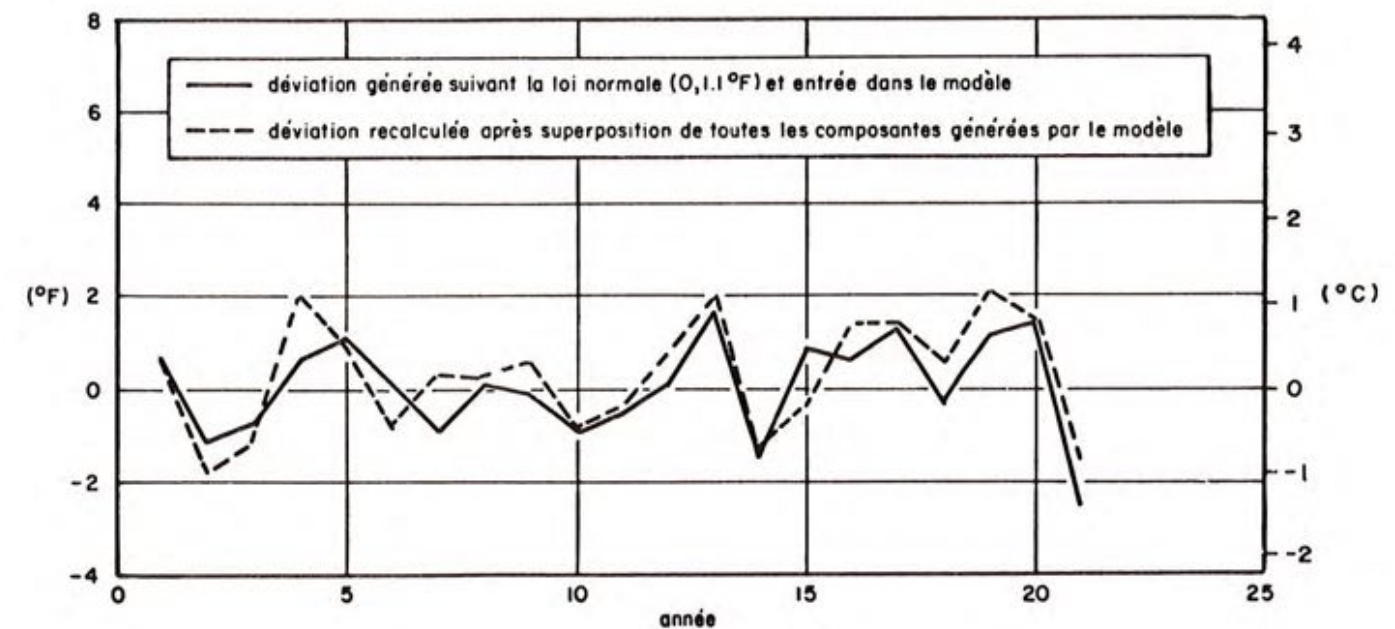
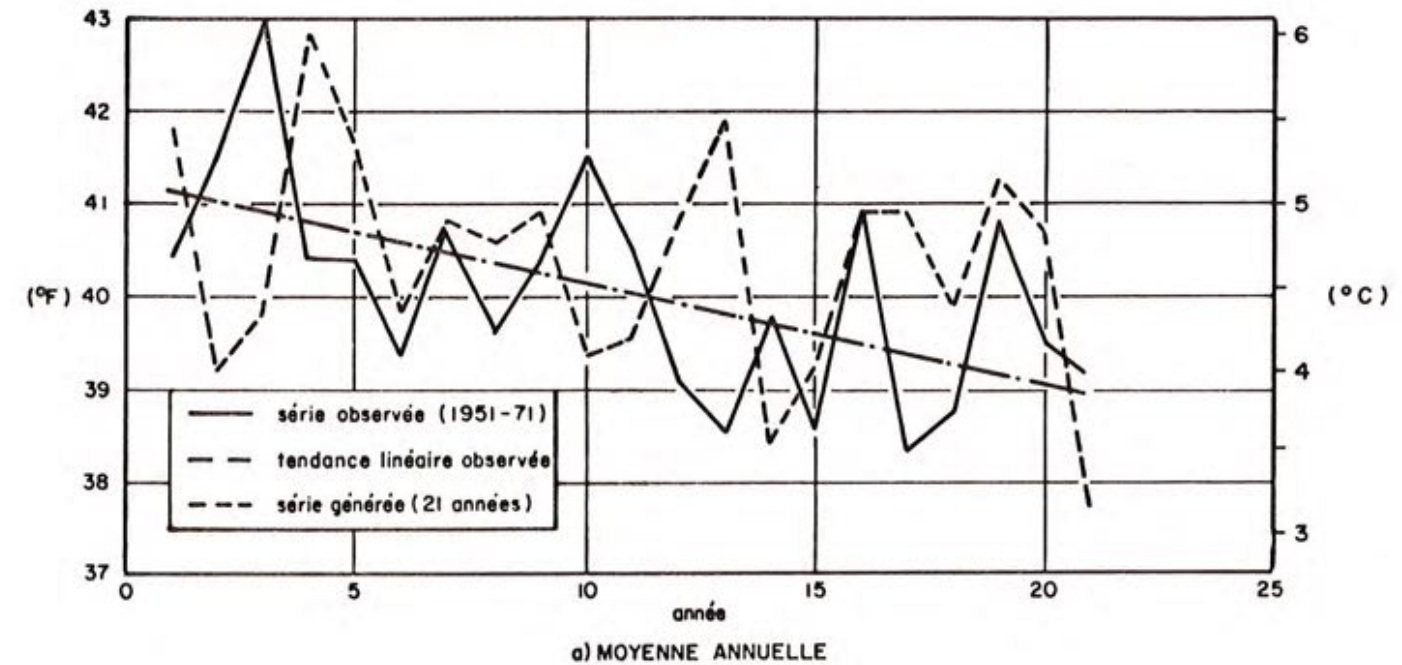


Fig. 12 — COMPARAISON ENTRE LES TEMPÉRATURES MOYENNES ANNUELLES OBSERVÉES ET GÉNÉRÉES POUR UNE SÉQUENCE.

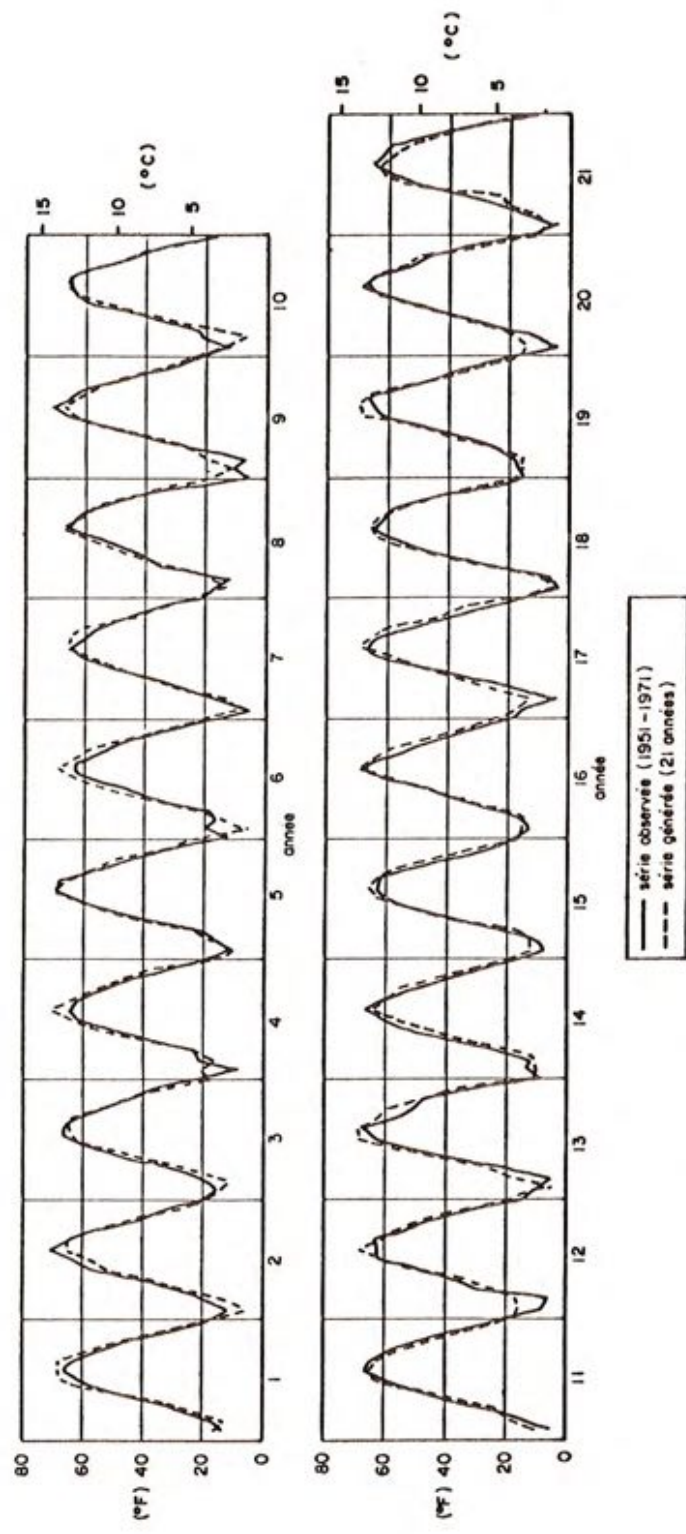


Fig. 13 — COMPARAISON ENTRE LES TEMPÉRATURES MOYENNES MENSUELLES OBSERVÉES ET GÉNÉRÉES POUR UNE SÉQUENCE.

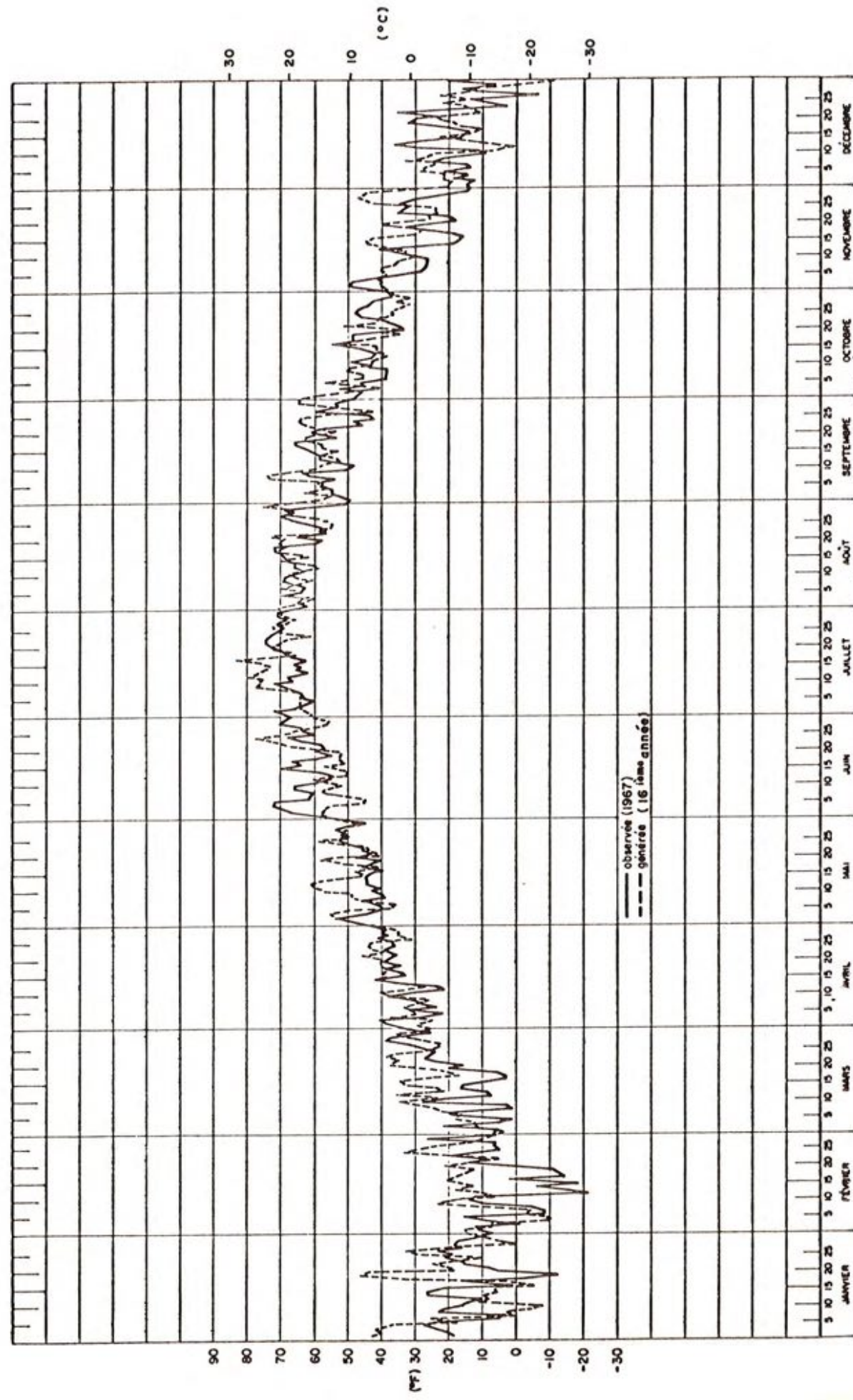


Fig. 14 — COMPARAISON ENTRE LES TEMPÉRATURES MOYENNES JOURNALIÈRES OBSERVÉES ET GÉNÉRÉES.

correspond à celle de l'année observée 1967 pour respecter l'effet de la tendance. Sur la figure 12 on remarquera l'écart entre moyenne annuelle générée au départ, et moyenne recalculée à la fin après l'addition des autres composantes dont le résidu stochastique. La fluctuation aléatoire se répercute d'abord sur les paramètres saisonniers qui, de fixes qu'ils étaient au départ, ont évolué d'année en année, et ensuite sur la moyenne annuelle. Ce phénomène a été observé par Quimpo *et al.* (1974) qui montrent, dans le cas d'une séquence de débits mensuels obtenus en superposant un cycle saisonnier sinusoidal et une composante aléatoire normalement distribuée, qu'il faut générer 100 années de valeurs avant de retrouver des paramètres moyens égaux à ceux introduits au départ. Dans le cas présenté ici les écarts sont minimes et non significatifs.

2. Statistiques sur les séries observées et générées.

L'introduction de la tendance à long terme ne facilite pas la comparaison entre valeurs observées et générées car il en résulte que chaque valeur de la chronologie appartient à une population différente. C'est d'ailleurs le but premier du modèle, la génération d'un grand nombre de séquences permettant d'obtenir directement la fonction de répartition pour chaque variable étudiée. A titre d'exemple on donne les statistiques comparées de l'année observée 1971 et de la 21^{ème} année de chaque séquence générée.

Variable	Observée (1971)	Générée (30 séquences) 21 ^{ème} année			Écart type
	(°C)	Min	Moy	Max	
Moyenne annuelle	4.0	2.6	3.7	4.8	0.4
Moyenne de février	-10.0	-15.6	-11.3	- 6.6	2.3
Moyenne de juillet	18.2	16.0	18.6	20.6	1.1
1 ^{er} février	-24.4	-33.2	-14.3	4.1	8.7
1 ^{er} juillet	20.0	11.7	19.1	25.9	3.6

Les valeurs observées en 1971, chacune représentant une réalisation particulière de la variable correspondante, s'intègrent bien dans leur distribution respective. Si l'on ignore la présence d'une tendance à long terme, c'est-à-dire si la moyenne annuelle est considérée comme une série stationnaire et fluctuante autour d'une moyenne interannuelle fixe, on peut établir les statistiques mensuelles sans tenir compte de la position chronologique de l'année. Ainsi toutes les moyennes de janvier sont supposées appartenir à la même population, celles de février à une autre, etc... Pour la série observée (1951-1971) et les 30 séquences générées de 21 années chacune, les résultats sont les suivants:

Mois (°C)	Température observée (1951-71)			Température générée (30 séquences de 21 années)		
	Min	Moy	Max	Min	Moy	Max
Janvier	-15.6	-11.8	-6.7	-20.9	-12.3	- 6.2
Février	-15.8	-10.3	-5.5	-18.4	-10.6	- 3.9
Mars	- 8.3	- 4.3	0.1	- 9.7	- 4.6	0.4
Avril	- 0.5	3.6	5.7	- 0.2	3.1	6.9
Mai	7.6	10.7	13.8	6.6	10.3	14.7
Juin	13.2	16.4	17.7	11.8	16.2	20.2
Juillet	16.7	19.1	22.1	15.7	19.2	22.9
Août	15.4	17.5	18.8	14.0	17.8	22.3
Septembre	11.1	13.0	15.6	8.8	13.6	18.2
Octobre	5.0	7.2	9.3	3.3	7.6	11.6
Novembre	- 2.8	0.2	3.2	- 4.2	0.2	5.2
Décembre	-14.7	- 8.5	-4.3	-12.9	- 7.6	- 1.3

En ce qui concerne les températures journalières, on donne ci-dessous deux exemples particuliers, celui du 1er février et celui du 1er juillet. Les valeurs maximales, moyennes et minimales observées sont comparées à leurs homologues obtenues par génération, mais pour la 21^{ème} année de chacune des 30 séquences seulement:

Jour (°C)	Température observée			Température générée		
	Min	Moy	Max	Min	Moy	Max
1er février	-26.7	-13.6	- 2.2	-33.2	-14.3	4.1
1er juillet	12.8	17.7	25.0	11.7	19.1	25.9

Les deux derniers tableaux montrent que le modèle tend à générer des températures dont les valeurs moyennes sont près des moyennes observées, mais dont les extrêmes peuvent être plus accentuées, en hiver surtout, que les observations minimales et maximales faites dans le passé. Cela tient au fait que la composante aléatoire de la déviation a une part importante et que l'on a admis pour elle une distribution normale. L'occurrence de valeurs extrêmes est donc rendue mathématiquement possible alors qu'en réalité une limite physique existe. Un type de loi asymétrique et bornée pour la distribution de la déviation pourrait améliorer le modèle sur ce point particulier.

En effet, d'après la figure 7, on a admis que les coefficients d'asymétrie et d'aplatissement d'ensemble demeureraient suffisamment entre leurs limites de confiance respectives pour retenir l'hypothèse de normalité de la déviation. Or, on a constaté (figure 9) que ces coefficients ont cependant une structure saisonnière marquée. Le coefficient d'asymétrie est en général négatif en hiver, positif en été, tandis que durant les périodes transitoires, le printemps et l'automne, il prend des valeurs élevées positives ou négatives. Le coefficient d'aplatissement est stable au voisinage de 3 en plein hiver et en plein été, tandis qu'en automne et au printemps surtout, il peut s'élever au-delà de 10. Ces constatations sont en parfait accord avec la réalité physique du phénomène et illustrent très bien le fait que la stationna-

rité du processus ne dépasse pas le deuxième moment de la distribution d'ensemble.

C'est pourquoi un modèle plus raffiné devrait tenir compte d'abord de l'asymétrie. En hiver une distribution asymétrique négative tendrait à générer des valeurs en majorité supérieures à la moyenne avec des extrêmes négatifs plus grands mais plus rares, tandis qu'en été, avec une asymétrie positive l'inverse se produirait, soit des valeurs en majorité inférieures à la moyenne avec des extrêmes positifs plus grands mais plus rares.

3. Variation des estimateurs des paramètres en fonction de la longueur de la série initiale d'observation.

L'examen de cette question est essentiel. En pratique on n'a pas souvent l'occasion de l'aborder car les séries d'observation sont relativement courtes. Aussi dans le cas de la série étudiée, qui couvre 28 années, il est possible, et opportun, d'évaluer l'influence de la durée d'échantillonnage sur les estimateurs des paramètres.

Les paramètres de la composante saisonnière sont les plus sensibles. En se reportant au tableau 3, on peut évaluer la dispersion des paramètres de la composante saisonnière approximée individuellement pour chacune des 28 années. Pour la fondamentale de la série de Fourier, on a par exemple

la dispersion rapportée dans le tableau ci-dessous.

	Min.	Moy.	Max.	Ec. type	Coef. Var.
Amplitude (°C)	13.7	15.4	16.9	0.9	0.061
Phase (jour)	107.8	113.6	119.3	2.9	0.025
% de variance expliquée	82.7	85.8	89.3	1.9	0.022

Ces résultats montrent donc une certaine variabilité d'année en année. Si l'on possède un échantillon court (2, 3, 5, 10 années par exemple) le cycle saisonnier moyen approximé peut donc n'être pas représentatif des caractéristiques à plus long terme. Pour évaluer le nombre minimal optimal d'années d'observation nécessaire pour que les paramètres du cycle saisonnier moyen soient aussi proches que possible de ceux du cycle saisonnier approximé sur le nombre maximum d'années disponibles, on a calculé les paramètres de tous les échantillons possibles d'une longueur donnée. On ne considère que des échantillons constitués d'années consécutives. On a ainsi pour 28 années:

Longueur de l'échantillon (années)	Nombre d'échantillons disponibles
2	27
3	26
5	24
10	19
20	9
28	1

Les résultats sont portés en figures 15 et 16. En ce qui concerne les paramètres du cycle saisonnier approximé par une série de Fourier, notamment ceux de la sinusoïde fondamentale, un échantillon de 5 années est suffisant pour que le cycle moyen représente à 5% près le cycle calculé sur la plus longue période (28 années). Ceci est valable aussi bien pour l'amplitude et pour la phase de la fondamentale que pour la variance expliquée par cette dernière. Par contre, pour le cycle de l'écart-type associé, approximé par les trois premières harmoniques y compris la fondamentale, la définition ne cesse de croître significativement avec le nombre d'années d'échantillonnage. Cependant à partir de 10 années, on peut admettre que l'approximation atteint une précision suffisante. On rappelle que pour une taille d'échantillon aussi petite que 10 pour estimer un écart-type, on peut avoir un estimateur non-biaisé par la correction qui a été exposée précédemment.

Les conclusions sont identiques dans le cas des paramètres séquentiels de la déviation standardisée par rapport au cycle saisonnier. La précision dans la définition de ce dernier et du cycle de l'écart-type associé se répercute d'ailleurs sur le calcul des coefficients d'autocorrélation. C'est pourquoi pour $Na < 5$ on peut trouver des valeurs de coefficient qui diffèrent quelque peu des valeurs trouvées pour une plus longue période. Le second coefficient d'autorégression est le plus sensible à la longueur d'échantillonnage, mais puisqu'il reste petit devant le premier, l'erreur

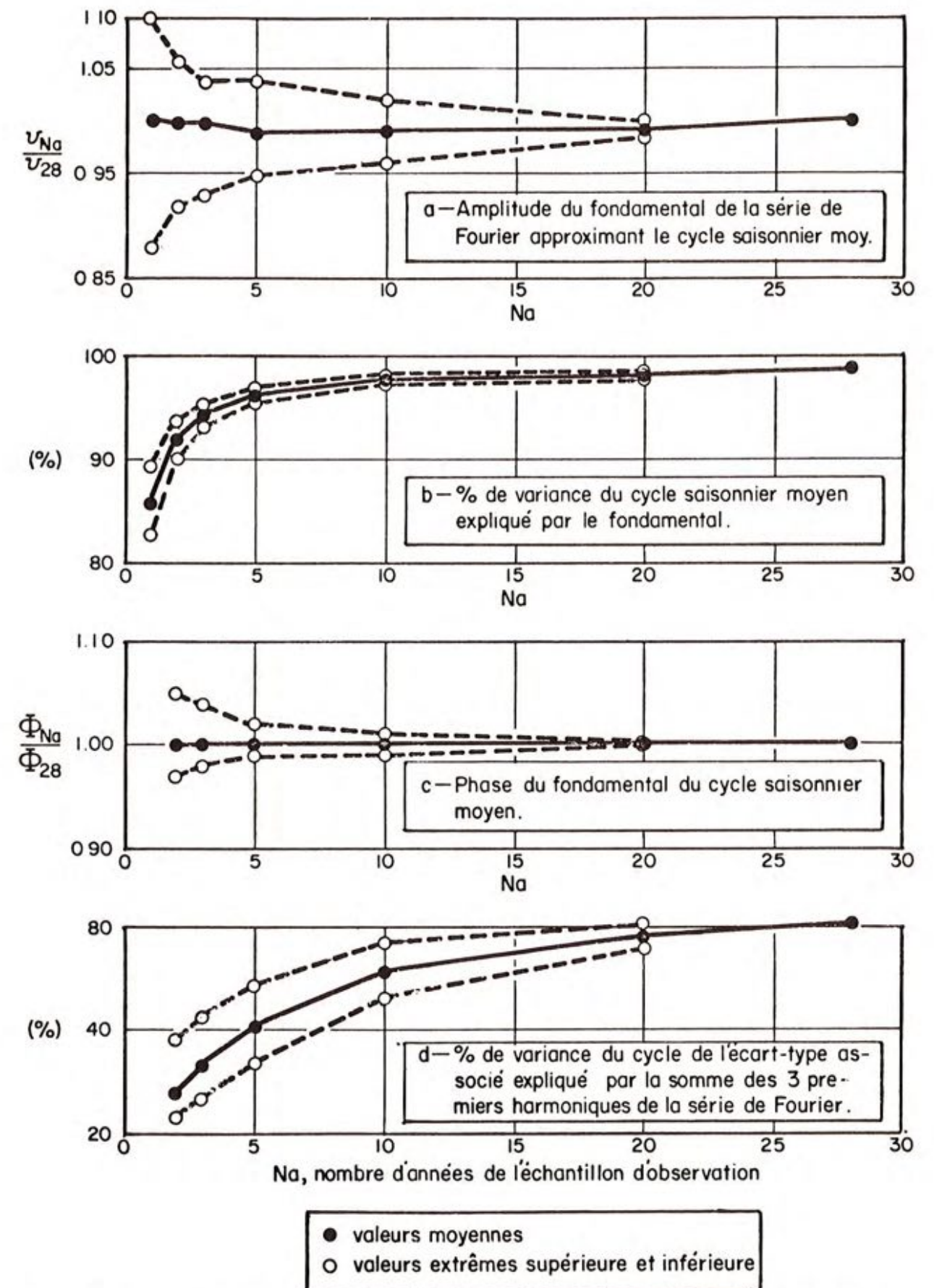


Fig. 15 — ÉVOLUTION DES PRINCIPAUX PARAMÈTRES SAISONNIERS EN FONCTION DE LA TAILLE DE L'ÉCHANTILLON D'OBSERVATION.

sur sa valeur a moins de conséquence. En conclusion, on peut affirmer qu'un échantillon de 5 années conduit à une précision acceptable pour les paramètres de la composante saisonnière et du coefficient d'autocorrélation de pas 1 de la déviation. A partir de 10 années, la fourchette du coefficient d'autocorrélation de pas 2 se resserre à $\pm 5\%$ par rapport à la valeur calculée sur long terme.

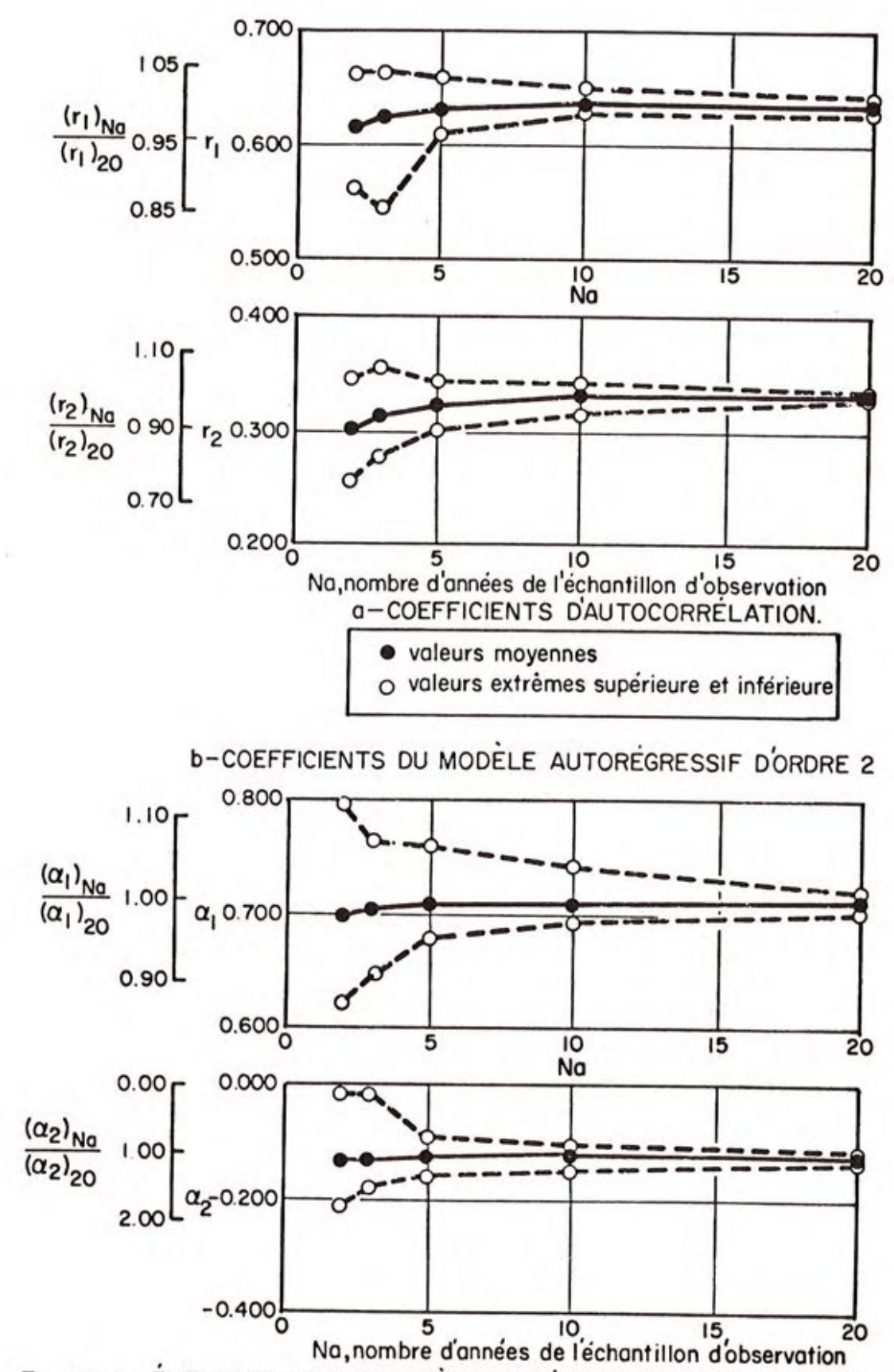


Fig. 16 — ÉVOLUTION DES PARAMÈTRES SÉQUENTIELS DE LA DÉVIATION STANDARDISÉE PAR RAPPORT AU CYCLE SAISONNIER.

III. CONCLUSIONS

Les méthodes classiques d'analyse des séries chronologiques s'appliquent avec succès dans le cas d'une série constituée par les observations journalières de la température de l'air. Les résultats exposés précédemment conduisent aux conclusions suivantes en ce qui concerne la structure de la série:

- a. Les quatre composantes identifiables dans la série sont:
 - la tendance à long terme (sur plusieurs décennies),
 - la composante saisonnière annuelle,
 - la déviation persistante,
 - la composante résiduelle aléatoire.
- b. Compte tenu de la taille de l'échantillon observé, l'allure de la tendance ne peut être définie que partiellement ou approximativement. Cependant, il est nécessaire d'en tenir compte car cette composante joue à l'encontre de la stationnarité du processus et parfois de façon non négligeable. Dans le cas de Québec, une approximation

linéaire de la diminution de la température moyenne annuelle de l'air conduit à un taux de -1°C en 18 ans. Le modèle de génération laisse la possibilité d'envisager n'importe quel type de tendance à long terme.

- c. L'influence saisonnière répétée chaque année confère à la température, après élimination de la moyenne annuelle, une variation cyclique. Son estimation est faite à l'aide de la sinusoïde fondamentale et des premières harmoniques de la série de Fourier, deux dans le cas de Québec. Le rôle de ces dernières est de rendre stationnaire tout le long de l'année la déviation par rapport au cycle saisonnier. Ceci est valable pour les températures de Québec. Dans le cas d'une autre station, le nombre et le numéro des harmoniques à prendre en compte peuvent être différents.
- d. La dispersion journalière, par rapport au cycle saisonnier moyen au cours des années, est elle-même saisonnière. L'estimation de ce cycle est suffisante avec les trois premières composantes de la série de Fourier. Ce nombre est également déduit des conditions particulières rencontrées à Québec, mais le principe reste valable pour d'autres stations.
- e. La déviation par rapport au cycle saisonnier, transformée afin que le processus soit stationnaire au moins en covariance est considérée comme persistante au moins sur deux jours consécutifs. Le processus autorégressif au premier ordre semble adéquat. Le processus au deuxième ordre of-

APPENDICE

RAPPEL DE MÉTHODOLOGIEA. Définitions

La composition et les propriétés séquentielles d'une série chronologique peuvent être analysées au moyen de diverses techniques relativement bien connues. Yevjevich (1972) expose clairement ces différentes techniques appliquées au domaine de l'hydrologie. Un certain nombre de définitions, qui sont à la base de l'analyse des séries chronologiques sont rappelées ci-dessous.

La série, discrète dans la plupart des cas étudiés à partir d'échantillons, suit un *processus*. Celui-ci est défini par une famille de variables. Le processus est *déterministe* si la valeur de chaque variable peut être calculée par une relation exacte, *stochastique* si les variables de la famille sont en totalité ou en partie aléatoires. A chaque valeur prise par une variable dans la séquence est associée une probabilité d'occurrence.

Dans le cas général des processus stochastiques, une séquence de valeurs observées est un *échantillon* ou une *réalisation* parmi toute la gamme des réalisations possibles. Un *ensemble* est constitué par n échantillons de même taille, mais pris au cours de périodes différentes. La figure 17 illustre le concept d'ensemble (Yevjevich, 1972).

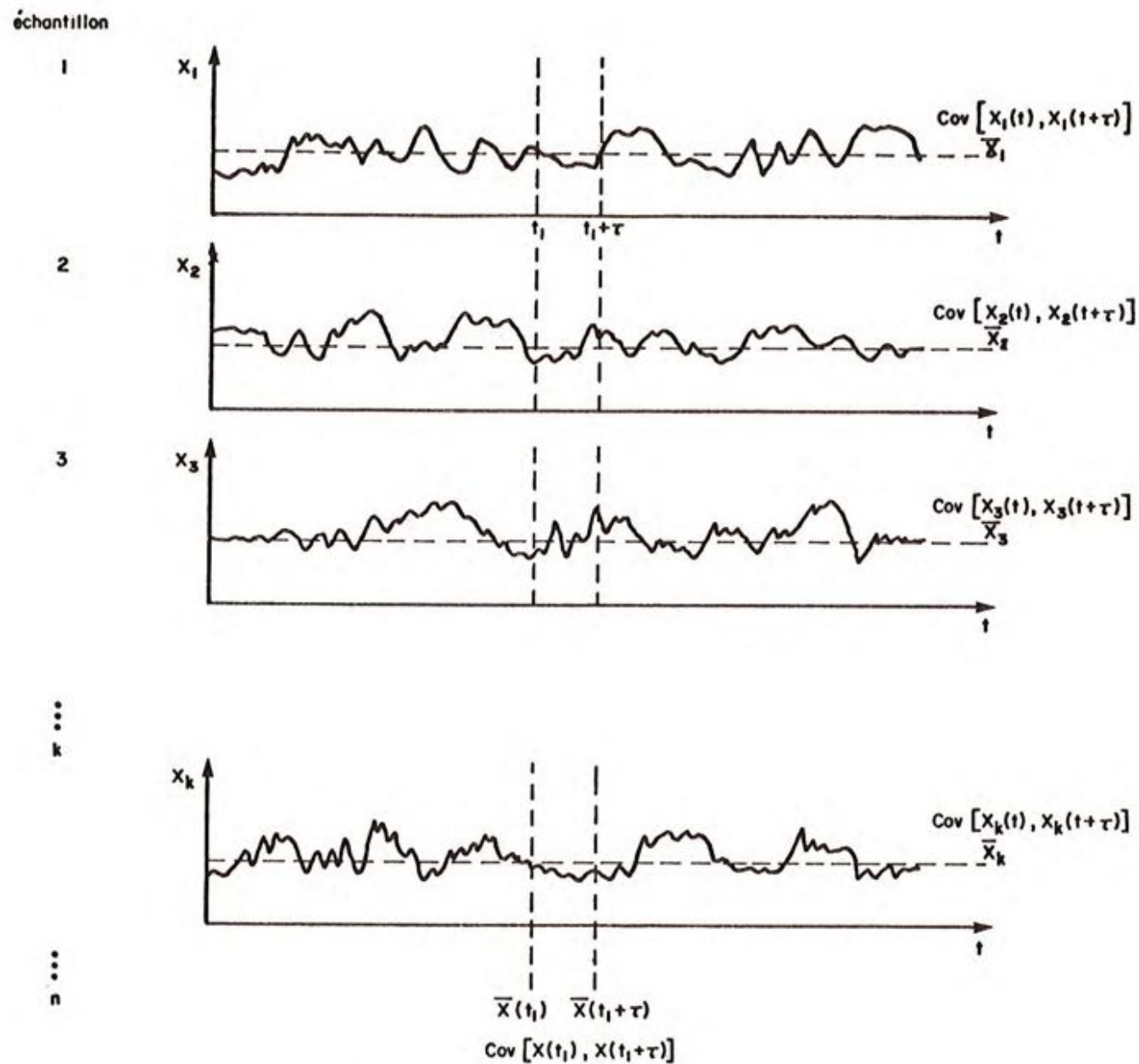


Fig. 17 — ENSEMBLE DE SÉRIES CHRONOLOGIQUES.

Le processus $\{X(t)\}$, où t représente le temps, est essentiellement caractérisé par la famille des fonctions de répartition $F(x(t_1))$, $F(x(t_2))$, ..., $F(x(t_N))$ des variables aléatoires $X(t_1)$, $X(t_2)$, ..., $X(t_N)$.

Le concept de *stationnarité* est alors le suivant: si les fonctions de répartition sont identiques quel que soit t , le processus est *stationnaire au sens strict*. Dans le cas contraire, le processus est non stationnaire ou évolutif. La stationnarité implique donc que toutes les variables de la famille appartiennent à la même population, soit que les moments de la fonction de répartition soient indépendants de t . D'une façon générale, si $\mu_i^{(p)}$ est le i ème moment de la population à laquelle appartient $X(t_i)$ et $m_i^{(p)}$ le moment de l'échantillon observé de taille n , le processus est stationnaire à l'ordre p s'il y a convergence en probabilité quand n tend vers l'infini de $m_i^{(p)}$ vers $\mu_i^{(p)}$ quel que soit i :

$$P[m_i^{(p)} - \mu_i^{(p)} \leq \varepsilon] = 1 \quad \forall \varepsilon > 0 \quad \text{lorsque } n \rightarrow \infty$$

En pratique on limite souvent l'étude de la stationnarité à l'ordre deux. Ainsi le processus est classé *faiblement stationnaire* ou *stationnaire au sens large*, si les deux premiers moments, moyenne et covariance, sont indépendants du temps. Quel que soit t , moyenne et covariance d'ensemble sont alors telles que:

$$E[X(t)] = \mu = \text{cste} \quad (11)$$

$$\text{cov}[X(t), X(t+\tau)] = E[(X(t) - \mu)(X(t+\tau) - \mu)] = f(\tau) \quad (12)$$

où τ est un pas de temps dont est fonction la covariance, et $E[.]$ est le symbole de l'espérance mathématique.

Il faut noter qu'un processus peut être séparément stationnaire en moyenne ou en covariance. Si le processus est stationnaire, et si les moyennes et autocovariances d'échantillons convergent également en probabilité vers les moments de la population, le processus est dit *ergodique*.

Les définitions rappelées précédemment permettent d'établir une classification des séries chronologiques. La figure 18 représente une schématisation de cette classification (Yevjevich, 1972).

B. Conception générale des modèles

Toute réalisation du processus $\{X(t)\}$ résulte de la superposition de plusieurs composantes identifiables aux différents types de la classification mentionnée précédemment. Le schéma général adopté est le suivant:

$$x(t) = d(t) + e(t) \quad (13)$$

où

$x(t)$ est la grandeur caractérisant le phénomène au temps t

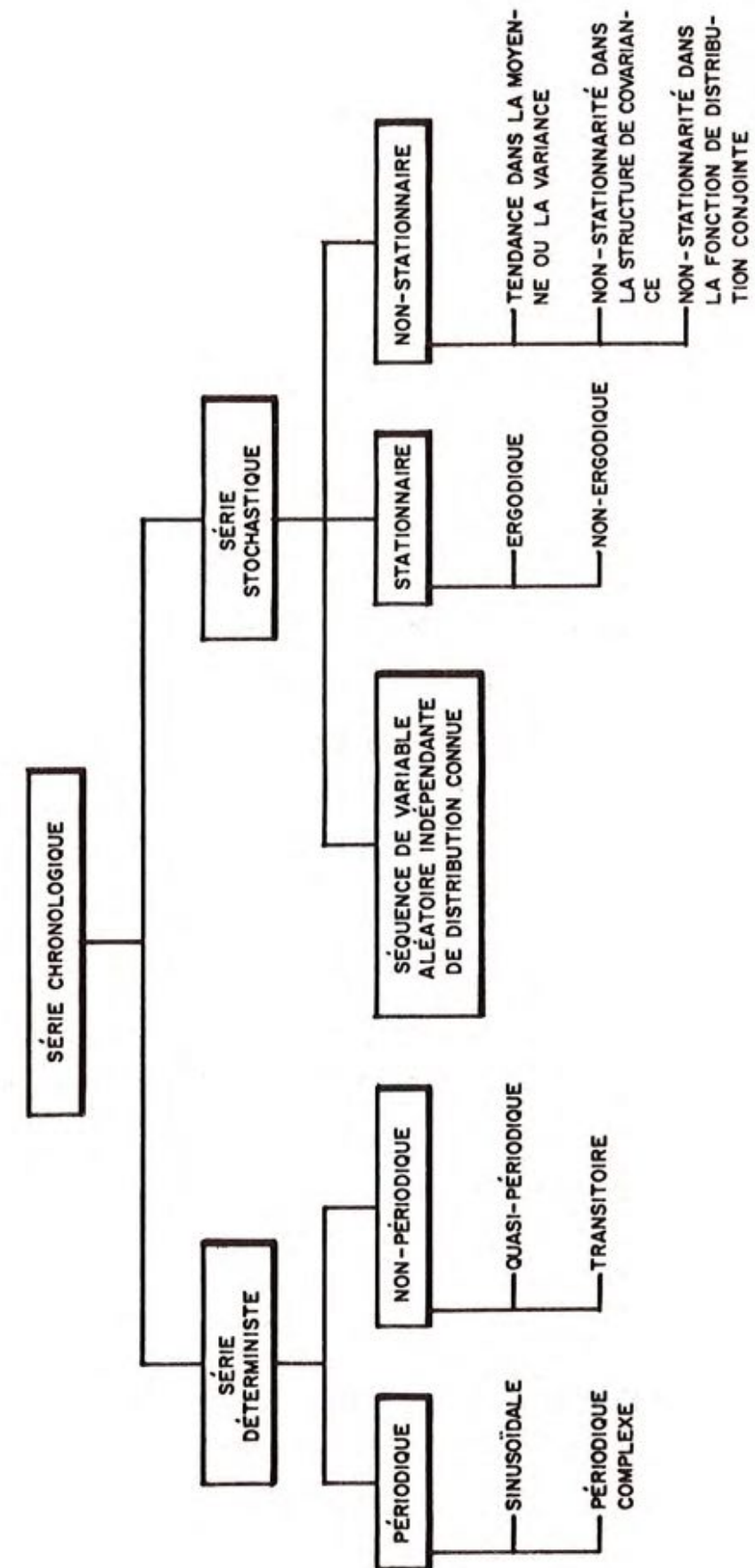


Fig. 18 — CLASSIFICATION DES SÉRIES CHRONOLOGIQUES.

$d(t)$ est la composante déterministe qui, si elle existe, peut résulter d'une tendance ou d'une somme de périodicités ou des deux simultanément

$e(t)$ est la composante résiduelle stochastique, pure ou non.

Les méthodes mises en oeuvre pour détecter d sont relativement classiques. Il s'agit de la technique de la moyenne mobile, de l'analyse de Fourier, du corrélogramme et de l'analyse spectrale. Quant aux formes que peut prendre e , elles sont diverses suivant sa structure séquentielle, résidu purement aléatoire dans le cas d'indépendance, résidu autocorrélé linéairement ou non dans le cas de dépendance. Les trois principaux types exploités sont le processus à moyenne mobile, le processus autorégressif et le processus mixte qui est une combinaison des deux précédents.

C. Composante déterministe

La définition de la composante déterministe permet d'inclure dans cette catégorie un éventail assez large de phénomènes, des *transitoires* aux *périodiques*. Les phénomènes transitoires se présentent sous la forme d'une *tendance* résultant d'une variation graduelle à long terme par exemple, ou d'une *discontinuité* qui a lieu sous l'effet d'un changement brusque. Les fluctuations périodiques sont en fait souvent *quasi-périodiques*. A part le cycle hydroclimatologique an-

nuel, les cycles à très long terme ou à court terme sont en effet de période irrégulière de telle sorte que le sens même de périodicité est presque impropre dans la plupart des cas.

La vérification de l'*homogénéité* d'une série de mesures est une étape qui doit nécessairement précéder toute analyse afin d'éviter les erreurs d'interprétation sur les tendances et autres fluctuations (Mitchell *et al.*, 1966). Par exemple il est manifeste que l'accroissement d'une agglomération urbaine entraîne une hausse graduelle de la température de l'air. Pour une station installée depuis longtemps au centre d'une ville, la série d'observations n'est donc pas homogène. En ce qui concerne les discontinuités, ou sauts, un déplacement de station même minime en est souvent la cause. Les comparaisons graphiques ou la courbe double de masse permettent de détecter ces phénomènes parasites.

Selon le principe exposé par Durbin (1963), la composante déterministe $d(t)$ incluse dans la mesure de $x(t)$ peut s'écrire ainsi:

$$d(t) = Td(t) + Cs(t) \quad (14)$$

où $Td(t)$ est la tendance, considérée comme une fonction déterministe du temps t , et $Cs(t)$ est la composante saisonnière de type cyclique. L'identification de ces termes est décrite ci-dessous.

1. Tendance

Pour confirmer ou infirmer les conclusions de l'examen graphique quant à la présence d'une tendance on peut utiliser le test de Spearman (Mitchell *et al.*, 1966). La statistique d'ordre permet de tester l'hypothèse de caractère aléatoire d'une série contre l'hypothèse de tendance.

Dans la série de N valeurs, chaque observation a un ordre séquentiel i et un numéro d'ordre croissant k_i .

En posant $\delta_i = k_i - i$, la statistique r_s est définie par:

$$r_s = 1 - \frac{6 \sum_{i=1}^N \delta_i^2}{N(N^2-1)} \quad (15)$$

Pour $N > 8$ la statistique

$$st = r_s \sqrt{\frac{N-2}{1-r_s^2}} \quad (16)$$

suit une distribution de Student à $(N-2)$ degrés de liberté. La valeur de st calculée est comparée à la valeur théorique correspondant à un niveau de signification α , le test étant à deux queues. Si $|st| \leq st(1-\frac{\alpha}{2})$ l'hypothèse de série aléatoire est acceptable, et en conséquence on ne peut pas supposer la présence d'une tendance dans la série.

L'approximation de Td , pour une série de N valeurs, peut toujours être réalisée à l'aide d'un polynôme de degré p tel que:

$$\hat{Td}(t) = \lambda_0 + \lambda_1 t + \lambda_2 t^2 + \dots + \lambda_p t^p \quad (17)$$

où $1 \leq p \leq N-1$. Pour p relativement petit devant $(N-1)$, les coefficients $\lambda_0, \dots, \lambda_p$ peuvent être déterminés par la méthode des moindres carrés. Si un polynôme de premier degré est suffisant, on est en présence d'une *tendance linéaire* et ce cas est le plus fréquent. Dans d'autres cas, l'approximation par une parabole ou une courbe du troisième degré donne de bons résultats [Chakravarti *et al.*, 1967]. Mais lorsque l'allure de la tendance devient plus complexe, il n'est pas toujours justifié d'élever indéfiniment le degré du polynôme [Kendall et Stuart, 1966]. L'estimation de la tendance par la moyenne mobile se présente alors comme la technique la plus intéressante.

Le calcul de la *moyenne mobile* consiste à substituer à chaque valeur de la série originale une nouvelle valeur obtenue par moyenne simple ou pondérée de h valeurs consécutives encadrant la valeur de départ. Ainsi, avec une série de N valeurs et une période de moyenne mobile entière et impaire $h = 2h'+1$, on obtient une nouvelle série de $(N-h+1)$ termes telle que:

$$\hat{Td}(t) = \frac{1}{h} [x(t-h') + x(t-h'+1) + \dots + x(t) + \dots + x(t+h'-1) + x(t+h')] \quad (18)$$

Les poids des observations $x(t-h')$, ..., $x(t+h')$ sont considérés ici tous égaux à l'unité. Il s'agit de la *moyenne mobile simple*. Matalas [1967a] présente des relations plus générales avec un poids attribué à chaque observation et mentionne que ces poids peuvent être considérés comme des coefficients binomiaux. Une telle *moyenne mobile pondérée* tend à produire un meilleur lissage de la tendance.

La période h impaire permet d'associer la moyenne calculée avec l'observation centrale de la période. Lorsque h est paire par nécessité, et c'est le cas si on considère les 12 moyennes mensuelles d'une variable hydrologique qui fluctue de façon saisonnière, on applique une moyenne mobile de période 2 à la série obtenue par moyenne mobile d'ordre $h = 2h'$, de sorte que la série finale de $(N-h)$ termes peut être calculée directement par:

$$\hat{T}d(t) = \frac{1}{2h} \left[x(t-h') + 2x(t-h'+1) + \dots + 2x(t) + \dots + 2x(t+h'-1) + x(t+h') \right] \quad (19)$$

Le nombre d'estimés perdus à chaque extrémité de la série est $h' = (h-1)/2$, que h soit paire ou impaire. La période h doit donc rester aussi petite que possible devant N pour atténuer cette perte néfaste, surtout à la fin de la série puisque l'extrapolation dans le futur est alors moins précise [Kendall et Stuart, 1966].

Le choix de h , comme celui du degré p d'un poly-

nôme, n'est pas soumis à des critères particuliers, mais dépend du but de l'estimation de la tendance. Cependant, si la série présente au départ une oscillation régulière surimposée à la tendance, ce mouvement cyclique n'est effacé que si la période de la moyenne mobile est égale ou multiple de la période de l'oscillation.

L'inconvénient majeur mentionné par les auteurs [Durbin, 1963], [Kendall et Stuart, 1966], [Mitchell *et al.*, 1966] à propos de l'emploi de la moyenne mobile est que cette technique de lissage introduit une corrélation positive dans une série, qui, au départ, n'en avait pas. Il s'agit de l'*effet Slutzky-Yule* qui se traduit par l'apparition d'une fluctuation artificielle de longue période plus ou moins régulière.

2. Composante périodique

L'analyse de Fourier est la méthode simple qui permet de mettre en évidence les périodicités présentes dans la série traitée. Son usage est fréquent dans l'analyse de nombreuses années climatologiques et météorologiques (Panofsky et Brier, 1963). Mais elle a l'inconvénient de ne faire intervenir qu'un certain nombre d'harmoniques. Pour une série discrète x_t constituée d'un nombre pair $N = 2n$ d'observations régulièrement espacées de Δt , la série de Fourier qui passe par les N points est constituée d'une constante qui est la moyenne, de $N/2$ harmoniques du type cosinus, de

$(N/2)-1$ harmoniques de type sinus:

$$x(t) = A_0 + \sum_{q=1}^{N/2} A_q \cos \frac{2\pi q t}{N} + \sum_{q=1}^{N/2-1} B_q \sin \frac{2\pi q t}{N} \quad (20)$$

où les coefficients A_0 , A_q , B_q sont définis par:

$$A_0 = \frac{1}{N} \sum_{t=1}^N x_t = \bar{x}$$

$$A_q = \frac{2}{N} \sum_{t=1}^N x_t \cos \frac{2\pi q t}{N} \quad q = 1, (\frac{N}{2}-1)$$

$$B_q = \frac{2}{N} \sum_{t=1}^N x_t \sin \frac{2\pi q t}{N} \quad q = 1, (\frac{N}{2}-1)$$

$$A_{N/2} = \frac{1}{N} \sum_{t=1}^N x_t \cos \pi t$$

$$B_{N/2} = 0$$
(21)

La fréquence fondamentale est $f_0 = 1/N\Delta t$ tandis que la plus grande fréquence décelable est $f_{\max} = 1/2\Delta t$. L'équation (20) peut être présentée sous la forme

$$x(t) = \bar{x} + \sum_{q=1}^{N/2} U_q \cos \left(\frac{2\pi q t}{N} - \phi_q \right) \quad (22)$$

$$\text{avec } U_q = \sqrt{A_q^2 + B_q^2} \quad \text{et} \quad \phi_q = \arctan \frac{B_q}{A_q} \quad (23)$$

U_q est l'amplitude de l'harmonique q , ϕ_q est son déphasage. D'après le théorème de Parseval (Jenkins et Watts, 1968), on peut montrer que la variance totale du signal σ_x^2 est la somme des contributions des harmoniques:

$$\sigma_x^2 = \sum_{q=1}^{N/2} \sigma_q^2 \quad (24)$$

$$\text{avec } \sigma_q^2 = \frac{1}{2} (A_q^2 + B_q^2), \quad q = 1, 2, \dots, (\frac{N}{2}-1)$$
(25)

$$\sigma_{N/2}^2 = A_{N/2}^2 + B_{N/2}^2, \quad q = \frac{N}{2}$$

La représentation graphique de σ_q^2 en fonction de la fréquence $f_q = q/N\Delta t$ est le spectre de variance ou *périodogramme*. La contribution de chaque harmonique, souvent exprimée en pourcentage de σ_x^2 , est alors immédiatement appréciable. Finalement, la sommation de m harmoniques peut suffire à approximer x_t .

La technique du corrélogramme et celle de l'analyse spectrale, qui est un développement plus élaboré du corrélogramme, sont aussi utilisables pour mettre en évidence les périodicités. Mais dans le cadre de l'étude de la composante saisonnière dont la période est connue à priori, l'information fournie par ces techniques est moins intéressante que celle obtenue par analyse harmonique classique. Par contre, ces techniques se révèlent très utiles dans l'analyse de la composante résiduelle.

D. Composante stochastique

La composante résiduelle est obtenue par soustraction de la composante déterministe des observations:

$$e(t) = x(t) - Td(t) - Cs(t) \quad (26)$$

Les caractéristiques séquentielles sont fondamentales. Leur connaissance permet d'identifier le processus générateur. Les notions d'autocovariance et de coefficient d'autocorrélation sont à la base de l'identification.

1. Fonction d'autocorrélation

En adoptant la formulation discrète x_i au lieu de $x(t)$, la forme générale de la fonction d'autocorrélation est la suivante:

$$\rho_\ell = \frac{\text{cov}[X_i, X_{i+\ell}]}{\text{var}[X_i]} = \frac{E[X_i X_{i+\ell}] - E[X_i]E[X_{i+\ell}]}{E[X_i^2] - (E[X_i])^2} \quad (27)$$

où ℓ est le pas de temps séparant les observations x_i et $x_{i+\ell}$. Pour une série ouverte, par opposition à une série circulaire, de N points, l'estimateur r_ℓ de ρ_ℓ est:

$$r_\ell = \frac{\frac{1}{N-\ell} \sum_{i=1}^{N-\ell} x_i x_{i+\ell} - \frac{1}{(N-\ell)^2} \left(\sum_{i=1}^{N-\ell} x_i \right) \left(\sum_{i=1}^{N-\ell} x_{i+\ell} \right)}{\left[\frac{1}{N-\ell} \sum_{i=1}^{N-\ell} x_i^2 - \frac{1}{(N-\ell)^2} \left(\sum_{i=1}^{N-\ell} x_i \right)^2 \right]^{\frac{1}{2}} \left[\frac{1}{N-\ell} \sum_{i=1}^{N-\ell} x_{i+\ell}^2 - \frac{1}{(N-\ell)^2} \left(\sum_{i=1}^{N-\ell} x_{i+\ell} \right)^2 \right]^{\frac{1}{2}}} \quad (28)$$

La représentation graphique de la fonction d'autocorrélation est le *corrélogramme*. En raison de la forme adimensionnelle de ρ , on lui donne généralement le nom de coeffi-

cient d'autocorrélation. L'allure du corrélogramme renseigne immédiatement sur les propriétés de dépendance ou d'indépendance entre les valeurs de la série. Plus la dépendance pour un pas ℓ donné est forte, plus r_ℓ est près de ± 1 . Le signe négatif exprime une liaison entre valeurs de signes opposés.

Pour $\ell = 0$, ρ_ℓ vaut toujours 1. Pour les pas successifs la valeur $\rho_\ell = 0$ indique une indépendance totale entre les valeurs séparées du pas ℓ . Si pour tout ℓ $\rho_\ell = 0$, le signal est purement aléatoire; il s'agit d'un *bruit blanc*. Cependant, une valeur $\rho_\ell \neq 0$ ne signifie pas nécessairement qu'il existe une relation de dépendance. L'écart entre la valeur calculée et l'espérance mathématique peut ne pas être significatif. La variance de l'estimateur r_ℓ croît avec ℓ . Pour cette raison il est recommandé d'imposer une limite à ℓ par rapport à la taille N .

Pour vérifier l'indépendance ($\rho_\ell = 0$), le test d'Anderson (1941) est un des moyens les plus utilisés. Bien que de nombreuses restrictions soient imposées, les limites d'acceptation de l'indépendance, au niveau α sont, après généralisation pour un pas ℓ quelconque:

$$L_{r_\ell}(\alpha) = \frac{-1 \pm u_\alpha \sqrt{N - \ell - 1}}{N - \ell} \quad (29)$$

où $u(\alpha)$ est la variable normale centrée réduite correspon-

dant au niveau α . Si r_ρ est à l'extérieur de la bande, l'hypothèse d'indépendance n'est pas retenue.

2. Analyse spectrale

A la différence de l'analyse harmonique qui repose sur le principe qu'un signal discret ne se décompose qu'en un nombre fini d'harmoniques de fréquence discrète, l'analyse spectrale permet de considérer l'existence d'une distribution continue des fréquences sur un domaine qui s'étend de l'infini à la plus petite valeur que l'on peut déceler. Mais l'emploi de l'analyse spectrale ne se résume pas à la recherche des périodicités. Il conduit aussi à l'identification des processus stochastiques qu'est susceptible de suivre un signal délesté de sa composante déterministe. Les spectres théoriques des processus sont connus. L'identification se fait par comparaison et test. Jenkins et Watts (1968), Box et Jenkins (1970) exposent en détails la théorie de l'analyse spectrale en relation avec l'étude des processus. Des applications à l'hydro-météorologie sont présentées par Mitchell *et al.* (1966).

L'usage classique de l'analyse spectrale dans la recherche de fréquences particulières est plus courant. En météorologie on le retrouve dans les travaux de Panofsky (1955), Griffith *et al.* (1956) et de Julian (1967), en hydrologie dans ceux de Quimpo (1967), Delleur (1971), et en qualité de l'eau dans ceux de Wastler (1963), Gunnerson (1966), Thomann (1967) et Demayo (1969). Pour un exposé théorique et pratique l'ouvrage de Bendat et Piersol (1971) est une excellente référence.

3. Processus stochastiques linéaires

La forme la plus générale du modèle utilisé pour représenter un processus stochastique linéaire stationnaire est la suivante:

$$Z_i = \sum_{j=1}^p \alpha_j Z_{i-j} + \sum_{j=1}^q \beta_j \varepsilon_{i-j} + \varepsilon_i \quad (30)$$

où Z_i est la valeur de la variable au temps i , α_j , $j = 1, p$ et β_j , $j = 1, q$ sont les coefficients et ε_i est le résidu distribué aléatoirement suivant une loi donnée. Le processus est dit mixte à l'ordre (p, q) puisqu'il inclut deux types séparés: le processus à moyenne mobile et le processus autorégressif.

Un processus stationnaire $\{Z_i\}$ est du type à *moyenne mobile* s'il peut être représenté par l'équation:

$$Z_i = \sum_{j=1}^q \beta_j \varepsilon_{i-j} + \varepsilon_i \quad (31)$$

où β_j , $j = 1, q$ sont des coefficients et où $\{\varepsilon_i\}$ est un processus stochastique pur, stationnaire de moyenne μ_ε et variance σ_ε^2 . Le nombre de termes q peut varier de 1 à ∞ . Le corrélogramme est tronqué à l'ordre q . La détermination de l'ordre q optimal à retenir et l'estimation des coefficients β_j sont relativement complexes (Jenkins et Watts, 1968), (Box et Jenkins, 1970). Comme exemple d'utilisation en hydrologie, Matalas (1967a) montre que le processus

à moyenne mobile peut représenter la relation entre le ruissellement Z_i et la précipitation effective ϵ_i sur une base annuelle.

Le modèle autorégressif est d'un emploi beaucoup plus simple et courant. Un processus stationnaire $\{Z_i\}$ est dit autorégressif s'il prend la forme:

$$Z_i = \sum_{j=1}^p \alpha_j Z_{i-j} + \epsilon_i \quad (32)$$

où α_j , $j = 1, p$ sont les coefficients de régression et $\{\epsilon_i\}$ est un processus stochastique pur stationnaire. L'observation i est donc linéairement dépendante des p observations précédentes.

Le coefficient d'autocorrélation suit la relation:

$$\rho_\ell = \sum_{j=1}^p \alpha_j \rho_{|j-\ell|} \quad \ell = 0, 1, \dots, p \quad (33)$$

Les coefficients α_j sont estimés, soit par la méthode de régression multiple à partir de l'équation (32), soit par la résolution du système de Yule-Walker (34) après remplacement des ρ_ℓ par les estimateurs r_ℓ .

$$\begin{aligned} \rho_0 &= 1 \\ \rho_1 &= \alpha_1 + \alpha_2 \rho_1 + \dots + \alpha_p \rho_{p-1} \\ \vdots & \\ \rho_p &= \alpha_1 \rho_{p-1} + \alpha_2 \rho_{p-2} + \dots + \alpha_p \end{aligned} \quad (34)$$

Seuls les modèles à l'ordre premier ou deuxième sont couramment utilisés.

Le modèle autorégressif du premier ordre est de la forme:

$$Z_i = \alpha_1 Z_{i-1} + \epsilon_i \quad (35)$$

On démontre aisément que, lorsque $p = 1$, $\alpha_1 = \rho_1$ et que l'expression mathématique du corrélogramme est:

$$\rho_\ell = \rho_1^\ell \quad (36)$$

$|\rho|$ décroît asymptotiquement vers zéro lorsque ℓ tend vers l'infini. Entre les variances de $\{Z\}$ et de $\{\epsilon\}$ on a la relation:

$$\sigma_Z^2 = \frac{\sigma_\epsilon^2}{1 - \rho_1^2} \quad (37)$$

Pour Z variable centrée réduite et ϵ remplacée par une composante résiduelle centrée réduite η , l'équation (35) se transforme ainsi:

$$Z_i = \rho_1 Z_{i-1} + \sqrt{1 - \rho_1^2} \eta_i \quad (38)$$

Dans le cas où Z est remplacée par une variable y de moyenne μ_y et variance σ_y^2 , le schéma général du modèle

est alors:

$$y_i = \mu_y + \rho_1(y_{i-1} - \mu_y) + \sigma_y \sqrt{1 - \rho_1^2} \eta_i \quad (39)$$

La fonction densité spectrale pour le premier ordre est:

$$S(f) = \frac{2 \sigma_\epsilon^2}{1 + \alpha_1^2 - 2 \alpha_1 \cos 2\pi f} \quad (40)$$

$$0 \leq f \leq \frac{1}{2}$$

Le modèle autorégressif au deuxième ordre présente un terme de plus:

$$z_i = \alpha_1 z_{i-1} + \alpha_2 z_{i-2} + \epsilon_i \quad (41)$$

Les relations entre coefficients α et coefficients d'autocorrélation ρ sont les suivantes:

$$\alpha_1 = \frac{\rho_1 - \rho_1 \rho_2}{1 - \rho_1^2} \quad \text{et} \quad \alpha_2 = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \quad (42)$$

En remplaçant ρ_1 et ρ_2 par leur estimateur respectif r_1 et r_2 on obtient les estimateurs de α_1 et α_2 . Le corrélogramme peut prendre diverses formes. La plus caractéristique, rencontrée lorsque $(\alpha_1^2 + 4 \alpha_2) < 0$, est pseudo-périodique amortie.

Lorsque toutes les variables sont centrées réduites,

la forme générale est:

$$z_i = \alpha_1 z_{i-1} + \alpha_2 z_{i-2} + \sqrt{1 - R^2} \eta_i \quad (43)$$

où R^2 , dit coefficient de détermination, joue le rôle de ρ_1^2 de l'équation (39).

$$R^2 = \frac{\rho_1^2 + \rho_2^2 - 2 \rho_1 \rho_2}{1 - \rho_1^2} \quad (44)$$

Le coefficient de détermination est le critère qui permet d'apprécier la capacité du modèle à expliquer la variance du phénomène. A partir de la relation:

$$R^2 = \frac{\sigma_z^2 - \sigma_\epsilon^2}{\sigma_z^2} \quad (45)$$

on peut voir que plus R^2 est grand en tendant vers 1, plus la variance résiduelle σ_ϵ^2 est petite. Fiering et Jackson (1971) basent le choix de l'ordre du modèle sur la valeur de R^2 qui atteint souvent rapidement un palier, ce qui montre que les modèles "multilag" ne sont pas toujours avantageux ou "explicatifs".

La fonction densité spectrale pour le modèle au deuxième ordre est de la forme:

$$S(f) = \frac{2 \sigma_\epsilon^2}{1 + \alpha_1^2 + \alpha_2^2 - 2\alpha_1(1-\alpha_2)\cos 2\pi f - 2\alpha_2 \cos 4\pi f} \quad (46)$$

Les modèles des types précédents sont souvent construits de telle sorte qu'ils conservent moyenne et variance des variables Z_i quel que soit i . La stationnarité au sens large est ainsi respectée.

Si, de plus, la ressemblance entre la série synthétique et la série historique doit s'étendre au coefficient d'asymétrie qui dépend du troisième moment, alors le schéma général du modèle doit être modifié [Matalas, 1967b]. Fiering et Jackson [1971] indiquent également la procédure à suivre dans le cas du modèle de Markov, autorégressif au premier ordre, lorsque Z n'est pas normalement distribué mais d'un type gamma par exemple.

RÉFÉRENCES

- Anderson, R.L., 1941: Distribution of the Serial Correlation Coefficient, *Ann. Math. Stat.*, 13 : 1-13.
- Bendat, J.S. et A.G. Piersol, 1971: Random Data: Analysis and Measurement Procedures, *Wiley-Interscience*, John Wiley & Sons Inc., 407 pp.
- Box, G.E.P. et G.M. Jenkins, 1970: *Time Series Analysis Forecasting and Control*, Holden-Day, 553 pp.
- Chakravarti, I.M., R.G. Laha et R. Roy, 1967: *Handbook of Methods of Applied Statistics*, 2 vol., John Wiley & Sons, Inc., 1, 560 pp.
- Cluis, D., 1972: Relation entre la température de l'eau d'une rivière et la température de l'air ambiant, *Rap. Tech., INRS-Eau*, Québec, 3, 11 p.
- Delleur, J., 1971: Développements récents en matière d'analyse des chroniques d'évènements hydrologiques: liaison en chaîne, analyse spectrale, *La Houille Blanche*, No. 2: 147-157.
- Demayo, A., 1969: The Computation and Interpretation of the Power Spectra of Water Quality Data, *Tech. Bull. Inland Waters Branch, Canada Dep. Energy, Mines and Resources*, 30 pp.
- Durbin, J., 1963: Trend Elimination for the Purpose of Estimating Seasonal and Periodic Components of Time Series, *Proc. of the Symp. on Time Series* held at Brown University, June 11-14, 1962, edited by M. Rosenblatt, John Wiley & Sons, Inc., New York, N.Y., 3-16.
- Fiering, M.B. et B. Jackson, 1971: Synthetic Streamflows, *Water Resources Monograph 1*, *Am. Geophysical Union*, 98 pp.
- Griffith, H.L., H.A. Panofsky et J. Van Der Hoven, 1956: Power Spectrum Analysis over Large Ranges of Frequency, *J. Meteorol.*, 13(3) : 279-282.
- Gunnerson, C.G., 1966: Optimizing Sampling Intervals in Tidal Estuaries, *Proc. Am. Soc. Civ. Eng., J. Sanit. Eng. Div.*, SA2 : 103-123.

- Gurland, J. et R.C. Tripathi, 1971: A Simple Approximation for unbiased Estimation of the Standard Deviation, *The American Statistician*, October: 30-32.
- Jenkins, G.M. et D.G. Watts, 1968: *Spectral Analysis and its Applications*, Holden-Day, 525 pp.
- Julian, P.R., 1967: Variance Spectrum Analysis, *Water Resour. Res.*, 3(3) : 831-845.
- Kendall, M.G. et A. Stuart, 1966: *The Advanced Theory of Statistics*, 3 vol., 3, Design and Analysis of Time Series, Hafner Publishing Co., New York, N.Y., 552 pp.
- Kite, G.W. et R.L. Pentland, 1971: Data Generating Methods in Hydrology, *Tech. Bul. Inland Waters Branch, Canada Dep. Energy, Mines and Resources*, 36, 22 pp.
- Kothandaraman, V., 1971: Analysis of Water Temperature Variations in Large Rivers, *Proc. Am. Soc. Civ. Eng., J. Sanit. Eng. Div.*, 97(SA1) : 19-31.
- Lamb, H.H., 1974: Reconstitution des zones climatiques dans le passé historique, *Endeavour*, 33(118) : 40-47.
- Matalas, N.C., 1967a: Time Series Analysis, *Water Resour. Res.*, 3(3) : 817-829.
- Matalas, N.C., 1967b: Mathematical Assessment of Synthetic Hydrology, *Water Resour. Res.*, 3(4) : 937-945.
- Mitchell, J.M. et al., 1966: Climatic Change, *Tech. Note, World Meteorol. Org.*, Genève, 79, WMO-No. 195. Tp. 100, 79 pp.
- Mood, A.M. et F.A. Graybill, 1963: *Introduction to the Theory of Statistics*, McGraw-Hill Book Co., Inc., 443 pp.
- Moshman, J., 1967: Random Number Generation, *Mathematical Methods for Digital Computers*, 2 vol., edited by A. Ralston and H.S. Wilf, John Wiley & Sons, Inc., 2 : 249-263.
- Panofsky, H.A., 1955: Meteorological Application of Power-Spectrum Analysis, *Bul. Amer. Meteor. Soc.*, 36(4) : 163-166.
- Panofsky, H.A. et G.W. Brier, 1963: Some Applications of Statistics to Meteorology, *The Pennsylvania State University*, University Park, Pa., 224 pp.

- Pearson, E.S. et H.O. Hartley, 1954: *Biometrika Tables for Statisticians*, University Press, Cambridge, 2 vol., 2, 270 pp.
- Quimpo, R.G., 1967: Stochastic Model of Daily River Flow, Sequences, *Hydrol. Paper, Colorado State University*, Fort Collins, Colo., 18, 30 pp.
- Quimpo, R.G. and Cheng, Mow-Soung, 1974: On the Variability of Seasonal Parameters in Hydrology Time Series, *J. Hydrology*, 23(3/4) : 279-287.
- Roesner, L.A. et V.M. Yevjevich, 1966: Mathematical Models for Time Series of Monthly Precipitation and Monthly Runoff, *Hydrol. Paper, Colorado State University*, Fort Collins, Colo., 15, 50 pp.
- Service Météorologique du Canada, 1971: *Daily Climatological Data, Quebec, P.Q.*, Observatory June 1972. February 1959, Airport April 1943 - Dec. 1969, Toronto.
- Song, C.C.S., A.F. Pabst et C.E. Bowers, 1973: Stochastic Analysis of Air and Water Temperatures, *Proc. Am. Soc. Civ. Eng., J. Environ. Div.*, EE6 : 785-800.
- Thomann, R.V., 1967: Time-series Analysis of Water - Quality Data, *Proc. Am. Soc. Civ. Eng., J. Sanit. Eng. Div.*, SA1 : 1-33.
- Triboulet, J.P., 1975: *Modèle stochastique de la température moyenne journalière de l'air*. Thèse de maîtrise ès sciences. Département de Génie civil, Université Laval, sous la direction de José Llamas, jan. 1975, 257 p.
- Velz, C.J. et J.J. Gannon, 1960: Forecasting Heat Loss in Ponds and Streams, *J. WPCF*, 32(4) : 392-417.
- Villeneuve, G.O., 1970: Fait-il plus chaud qu'à l'époque de nos grands-parents? *Feuillelet Météorologique, Québec, Service de la Météo., Min. Rich. Nat.*, (3) : 95-106.
- Wastler, T.A., 1963: Application of Spectra Analysis to Stream and Estuary Field Surveys, I., Individual Power Spectra, *Environ. Health Series, U.S. Dep. Health, Education and Welfare*, Cincinnati, Ohio.
- Yevjevich, V.M., 1972: *Stochastic Processes in Hydrology*, Water Resources Publications, Fort Collins, Colo., 246 pp.