

Record Number: 1250
Author, Monographic: Fortin, J. P.//Morin, G.//Dupont, L.//Lacroix, J.//Leblanc, D.//Plante, B.
Author Role:
Title, Monographic: Étude du réseau météorologique du Québec en vue de sa rationalisation.
Tome 1 : cadre général de rationalisation du réseau météorologique du Québec
Translated Title:
Reprint Status:
Edition:
Author, Subsidiary:
Author Role:
Place of Publication: Québec
Publisher Name: INRS-Eau
Date of Publication: 1980
Original Publication Date:
Volume Identification:
Extent of Work: viii, 100
Packaging Method: pages
Series Editor:
Series Editor Role:
Series Title: INRS-Eau, Rapport de recherche
Series Volume ID: 124
Location/URL:
ISBN: 2-89146-121-5
Notes: Rapport annuel 1979-1980
Abstract: Rapport rédigé pour Environnement Québec
15.00\$
Call Number: R000124
Keywords: rapport/ ok/ dl

ETUDE DU RESEAU METEOROLOGIQUE
DU QUEBEC EN VUE DE SA RATIONALISATION
Tome 1: Cadre général de rationalisation
du réseau météorologique du
Québec.

RAPPORT SCIENTIFIQUE No 124
1980

Rapport rédigé pour
Environnement Québec

par

J.-P. Fortin, G. Morin, L. Dupont,
J. Lacroix, D. Leblanc et B. Plante

INRS-Eau
Université du Québec
C.P. 7500, Sainte-Foy
Québec G1V 4C7

ETUDE DU RESEAU METEOROLOGIQUE
DU QUEBEC EN VUE DE SA RATIONALISATION
Tome 1: Cadre général de rationalisation
du réseau météorologique du
Québec.

RAPPORT SCIENTIFIQUE No 124
1980

Rapport rédigé pour
Environnement Québec

par

J.-P. Fortin, G. Morin, L. Dupont,
J. Lacroix, D. Leblanc et B. Plante

INRS-Eau
Université du Québec
C.P. 7500, Sainte-Foy
Québec GIV 4C7

CADRE GENERAL DE RATIONALISATION
DU RESEAU METEOROLOGIQUE DU QUEBEC

TABLE DES MATIERES

	<u>PAGE</u>
REMERCIEMENTS	i
TABLE DES MATIERES	ii
LISTE DES TABLEAUX	v
LISTE DES FIGURES	vii
CHAPITRE 1 - INTRODUCTION	1
CHAPITRE 2 - OBJECTIFS GENERAUX DE L'ETUDE	4
CHAPITRE 3 - LA RATIONALISATION DES RESEAUX: EVENTAIL DES PRINCIPAUX CHOIX	6
3.1 L'application de normes générales	7
3.2 La conception d'un réseau de mesures	8
3.3 Les techniques de rationalisation	13
3.3.1 La cartographie	14
3.3.2 La régionalisation	15
3.3.3 L'analyse de système	16
3.4 Les critères de contrôle d'atteinte des objectifs	16
3.4.1 L'atteinte d'une précision requise	17
3.4.2 L'optimisation bénéfices-coût	18
3.4.3 L'optimisation précision-coût	18

	<u>PAGE</u>
3.5 L'intégration des réseaux entre eux	19
3.6 La prise en compte des méthodes d'analyse	20
3.7 La durée des observations	21
3.8 L'intégration des données au sol et des données acquises par de nouvelles techniques	22
CHAPITRE 4 - LA RATIONALISATION DES RESEAUX: DEMARCHE PROPOSEE	25
CHAPITRE 5 - METHODES ANALYTIQUES	29
5.1 Estimation des valeurs ponctuelles	30
5.1.1 Revue de la littérature	30
5.1.2 Choix et conditions d'application d'une méthode d'analyse	34
5.1.3 Analyse en composantes principales	36
5.1.3.1 Objectif	36
5.1.3.2 Théorie	37
5.1.3.3 Application au réseau météorologi- que du Québec pour la définition de "régions météorologiques sta- tistiquement homogènes"	48
5.1.3.4 Interprétation des résultats	51
5.1.4 Interpolation optimale	55
5.1.4.1 Objectif	55
5.1.4.2 Structure des champs météorologi- ques	55
5.1.4.3 Théorie générale	67
5.1.4.4 Résolution pratique du système d'équations	73
5.1.4.5 Erreur d'interpolation en fonc- tion de la distance entre les stations	74
5.1.4.6 Cartographie des écarts types d'interpolation du réseau en opé- ration le 1er janvier 1978	79
5.1.4.7 Interprétation des résultats	82

	<u>PAGE</u>
5.2 Ecart type d'estimation de la moyenne des précipitations sur une surface	84
5.2.1 Objectif	84
5.2.2 Théorie	84
5.2.3 Condition d'application de la méthode de Kagan	86
5.2.4 Interprétation des résultats	87
REFERENCES BIBLIOGRAPHIQUES	95

LISTE DES TABLEAUX

		<u>PAGE</u>
3.1	Densité minimale d'un réseau de stations de mesure des précipitations	9
3.2	Distances inter-stations admissibles pour l'estimation de certaines variables météorologiques à la surface du sol (d'après Gandin, 1970)	10
5.1	Coefficients de corrélation entre les composantes principales et les stations, calculés à partir de 90 valeurs de pluie cumulée sur 10 jours, en été, pour un réseau de 14 stations situées sur le bassin de la rivière Eaton (avant rotation des axes) (d'après Morin, et al. , 1978)	39
5.2	Coefficients de corrélation entre les composantes principales et les stations, calculés à partir de 90 valeurs de pluie cumulée sur 10 jours, en été, pour un réseau de 14 stations situées sur le bassin de la rivière Eaton (d'après rotation par la méthode VARIMAX)	45
5.3	Variation des paramètres $\rho(o)$ et d_o pour l'analyse de sensibilité de ces paramètres	88
5.4	Ecart type d'estimation (mm) de la moyenne des précipitations liquides journalières au printemps en fonction de la superficie attribuée à chaque station et du nombre de station: région Saguenay - Lac-St-Jean	89
5.5	Ecart type d'estimation (mm) de la moyenne des précipitations liquides journalières au printemps en fonction de la superficie attribuée à chaque station et du nombre de station: région Saguenay - Lac-St-Jean. Paramètre $\rho(o) = 0.8$	91
5.6	Ecart type d'estimation (mm) de la moyenne des précipitations liquides journalières au printemps en fonction de la superficie attribuée à chaque station et du nombre de station: région Saguenay - Lac-St-Jean. Paramètre (ρ) = 0.95	92

5.7	Ecart type d'estimation (mm) de la moyenne des précipitations liquides journalières au printemps en fonction de la superficie attribuée à chaque station et du nombre de station: région Saguenay - Lac-St-Jean. Paramètre $d_0 = 50$ km	93
5.8	Ecart type d'estimation (mm) de la moyenne des précipitations liquides journalières au printemps en fonction de la superficie attribuée à chaque station et du nombre de station: région Saguenay - Lac-St-Jean. Paramètre $d_0 = 150$ km	94

LISTE DES FIGURES

		<u>PAGE</u>
5.1	Représentation graphique des coefficients de corrélation entre la station j et les q premières composantes principales, dans le plan des composantes principales 2 (axe X) et 3 (axe Y). (d'après Morin et al. , 1978)	41
5.2	Représentation dans le plan des coefficients de corrélation entre le réseau de 14 stations et les trois premières composantes principales pour les pluies cumulées sur 10 jours en été selon le tableau 5.1. Les intervalles de confiance au niveau de la probabilité de 80% sont tracés pour les groupes de stations définis après la rotation effectuée par la méthode VARIMAX (d'après Morin et al. , 1978)	43
5.3	Identification numérique des régions météorologiques du Québec	50
5.4	Limites initiales et finales des régions statistiquement homogènes pour les chutes de neige mensuelles	52
5.5	Variation spatiale des coefficients de corrélation entre les deux premières composantes principales et les stations retenues dans la première liste pour les chutes de neige mensuelles	53
5.6	Allure générale d'une fonction de structure	60
5.7	Points initiaux, points moyens par classe de distance et courbe ajustée	64
5.8	Points initiaux, points moyens par classe de distance et courbe ajustée	65
5.9	Position des stations pour le calcul de l'erreur d'interpolation au centre de deux triangles équilatéraux emboîtés	75

	<u>PAGE</u>
5.10 Ecart type d'interpolation en fonction de la distance entre les stations pour les pluies journalières au printemps	78
5.11 Répartition spatiale des écarts types	81

CHAPITRE I

INTRODUCTION

Mandat du ministère des Richesses naturelles

Le 3 avril 1962, un arrêté en conseil accordait au ministère des Richesses naturelles **"la responsabilité exclusive de l'observation, la compilation, l'analyse et la publication de tous les phénomènes météorologiques et climatologiques pour les besoins de tous les ministères du gouvernement de la province de Québec"**.

Créé par suite de ce mandat, le Service de météorologie du ministère des Richesses naturelles s'est efforcé depuis ce temps de remplir les fonctions qui lui ont été confiées.

Problématique de l'étude

Depuis la création du Service de météorologie du ministère des Richesses naturelles, un certain nombre de stations climatologiques sont venues s'ajouter à celles qui existaient déjà depuis plus ou moins longtemps. Concomitamment à l'ouverture de ces nouvelles stations, d'autres ont cessé leurs opérations pour différentes raisons, allant du désintéressement ou du départ de l'observateur à l'arrêt d'un programme de mesures.

Au cours de la même période, les divers paliers de gouvernements, l'industrie et le public en général prenaient de plus en plus conscience de l'importance et de l'influence des conditions météorologiques sur leurs activités propres. Quoique bien souvent ils ne savaient pas exactement comment préciser leurs besoins, ils ont présenté de plus en plus de demandes pour des données météorologiques, mais pas toujours celles dont ils auraient réellement eu besoin.

Face à l'évolution et de son réseau et des besoins, le Service de météorologie a considéré que le temps était venu d'analyser plus en détails son réseau d'acquisition de données météorologiques en vue de planifier plus adéquatement son développement futur.

C'est dans ce cadre qu'il a voulu savoir comment le réseau québécois d'acquisition de données météorologiques devait évoluer pour répondre le plus adéquatement possible aux besoins réels (exprimés ou non jusqu'ici) actuels et futurs du Québec.

Organismes impliqués dans l'étude

Tout en débutant lui-même les études visant à faire le point d'une part sur l'état des réseaux et d'autre part sur les besoins, le Service de météorologie, ne pouvant réaliser seul les études nécessaires à la rationalisation de son réseau climatologique, compte tenu de ses activités régulières et de ses effectifs, a confié à l'INRS-Eau, en 1977, l'orientation et la réalisation de l'ensemble des études portant sur:

- le réseau d'acquisition des précipitations de neige (chutes de neige et neige au sol);
- le réseau d'acquisition des pluies journalières;
- le réseau d'acquisition de la température de l'air sous abri.

Le ministère souhaitant de plus que d'autres universités soient associées à l'étude, des sous-contrats ont été octroyés par l'INRS-Eau à Monsieur Reynald Naud de l'Université Laval, à Monsieur Izstar Zawadzki de l'Université du Québec à Montréal et au département de Physique de l'UQAM où travaille Monsieur Zawadzki.

CHAPITRE II

OBJECTIFS GENERAUX DE L'ETUDE

Les objectifs généraux définis pour l'étude des trois réseaux climatologiques mentionnés plus haut visent tout d'abord à cerner les besoins à satisfaire, puis à étudier dans quelle mesure ces réseaux peuvent déjà ou pourront fournir les données désirées, compte tenu des contraintes financières, humaines, techniques et logistiques inhérentes à la création et l'opération de nouvelles stations là où il sera jugé nécessaire. Ces objectifs sont:

- la détermination des besoins en données climatologiques sur l'ensemble du territoire québécois;
- la définition de zones caractérisées par ces besoins;
- la détermination des caractéristiques de la variable à mesurer pour satisfaire ces besoins dans chaque zone (valeur ponctuelle ou moyenne spatiale, pas de temps, précision);
- la détermination des caractéristiques climatologiques et statistiques des variables désirées;
- la détermination de la densité de stations susceptible de satisfaire la précision requise;
- l'analyse des réseaux existants afin d'estimer dans quelle mesure ils répondent aux besoins;
- la recommandation de la démarche à suivre pour que les réseaux existants puissent répondre davantage aux besoins.

CHAPITRE III

LA RATIONALIATION DES RESEAUX:
EVENTAIL DES PRINCIPAUX CHOIX

Il est impensable de vouloir estimer une variable météorologique en tout point d'un territoire grâce à un très grand nombre de points de mesure au sol. Compte tenu de plusieurs contraintes tant physiques qu'économiques, on ne peut espérer compter sur un nombre limité de points de mesure constitués en réseau. Il importe alors, à partir de ce nombre limité de points de mesure, de pouvoir estimer les valeurs prises par une variable météorologique aux points où cette variable n'est pas mesurée avec une précision satisfaisante, compte tenu des différentes contraintes influençant à la fois le nombre, la distribution spatiale et la durée des stations d'un réseau. La réalisation des objectifs mentionnés plus haut passe donc par une stratégie d'intervention comportant un certain nombre de choix en ce qui a trait à:

- l'application de normes générales;
- la conception d'un réseau de mesures;
- les techniques de rationalisation;
- les critères de contrôle d'atteinte des objectifs;
- l'intégration des réseaux entre eux;
- la prise en compte des méthodes d'analyse;
- la durée d'observation;
- l'intégration des données au sol et des données acquises par de nouvelles techniques (radar, télédétection, etc...).

3.1 L'APPLICATION DE NORMES GENERALES

La rationalisation des réseaux peut être abordée de façon relativement simple en visant à appliquer des normes minimales générales, telles celles suggérées pour les précipitations par l'Organisation météorologique mondiale

(OMM) dans son "Guide to Hydrometeorological Practices" datant de 1965 et présentées au tableau 3.1.

L'application de ces normes ne constitue pas en soi une rationalisation de réseau. Elle peut même s'avérer irréaliste (Ferguson, 1972). Dans un autre rapport de l'OMM, Gandin (1970) suggère des normes plus détaillées concernant diverses variables météorologiques observées au niveau du sol. Selon Gandin, ces variables peuvent être divisées en trois groupes correspondant à la densité du réseau de mesure requis (tableau 3.2).

Gandin (1970) précise par ailleurs que la densité de réseau change substantiellement d'une région physico-géographique à l'autre selon les modifications et variations de la structure statistique des champs météorologiques étudiés. Ses estimés doivent donc être considérés uniquement comme une première approximation, surtout valable en terrain plat. Ils donnent, toujours selon Gandin (1970), seulement une idée générale de la planification d'un réseau, idée qui doit être complétée en l'appliquant spécifiquement à des conditions physiques et géographiques variées sur la base d'une information détaillée et sûre concernant la structure des champs météorologiques.

D'autres normes sont suggérées dans la littérature. Elles ont le plus souvent été établies pour des régions particulières et leur présentation n'apportant pas d'informations essentielles, elles ne sont pas rapportées ici.

3.2 LA CONCEPTION D'UN RESEAU DE MESURES

Il importe tout d'abord de se demander ce qui motive la création ou le maintien d'un réseau. De la réponse à cette question devrait découler la conception d'un réseau. Ainsi, la connaissance générale de la variation spatiale d'une variable météorologique à l'échelle d'un pays ou d'une région peut être l'objectif souhaité. Dans d'autres cas, les données recueillies par un réseau peuvent l'être afin de répondre à des besoins généraux ou

Tableau 3.1 Densité minimale d'un réseau de stations de mesure des précipitations¹.

Type de région	Gamme de densités minimales Superficie échantillonnée par station, en km ²	Gamme de densités provisoires tolérées dans les cas difficiles Superficie échantillonnée par station, en km ²
I- Régions plates des zones tempérées, méditerranéennes et tropicales	600 - 900	900 - 3,000
II- Régions montagneuses des zones tempérées méditerranéennes et tropicales Petites îles montagneuses avec des précipitations très irrégulières et un réseau hydrographique très dense.	100 - 250 25	250 - 1,000
III- Zones arides et polaires	1,500 - 10,000	

¹ D'après "Guide to Hydrometeorological Practices".

Tableau 3.2 Distances inter-stations admissibles pour l'estimation de certaines variables météorologiques à la surface du sol (d'après Gandin, 1970).

GROUPE	VARIABLE METEOROLOGIQUE	GAMME DE DISTANCES INTER-STATIONS ADMISSIBLES (en km)	CARACTERISTIQUE CONSIDEREE
I	pression de l'air température du sol durée d'ensoleillement	150 - 200	observations instantanées moyennes mensuelles total mensuel
II	température de l'air humidité de l'air vitesse du vent nébulosité	50 - 60	moyenne journalière pour toutes les variables; et (approximativement) ob- servations instantanées
III	précipitation caractéristiques du manteau nival phénomènes météo- rologiques (orages, brouillard, tempête de neige, etc.)	< 30	total mensuel observations instantanées* observations instantanées*

* Caractéristique non précisée par Gandin et suggérée par les auteurs du présent rapport.

spécifiques de gestion des ressources en eau au niveau d'un bassin ou d'une région. La conception d'un réseau peut donc être abordée en fonction du niveau de l'information requise. C'est l'approche choisie par le U.S. National Water Data System et rapportée par Rodda (1969).

Considérant que des différences existent dans le niveau (fréquence et précision) des informations requises pour des opérations menées à l'échelle nationale, régionale et locale, les auteurs du rapport cité par Rodda (1969) préconisent de concevoir les réseaux de manière à répondre adéquatement à chacun de ces besoins. Cela suppose, selon eux, la répartition des stations en trois catégories en fonction des niveaux d'information désirés.

La première catégorie de stations vise à fournir une information de base apte à répondre aux besoins de planification à l'échelle nationale ou supra-régionale, à permettre la surveillance des orages majeurs et à constituer une banque de données nationales.

La seconde catégorie de stations a pour but de répondre aux besoins exprimés à l'échelle régionale ou au niveau d'un bassin. Les stations de ce groupe s'ajoutent alors à celles du premier groupe de manière à fournir les informations supplémentaires requises.

Une troisième catégorie de stations vise à recueillir les informations nécessaires à la poursuite d'objectifs opérationnels, légaux et administratifs à l'échelle locale. Cette catégorie de stations est toutefois moins sujette à une planification.

Conçu selon ces propositions, un réseau météorologique peut être très souple, sa densité variant dans l'espace et dans le temps en fonction des divers objectifs poursuivis à quelque échelle que ce soit. Ceci est valable pour n'importe quel réseau météorologique. Il importe de plus, toutefois, de faire en sorte de planifier les divers réseaux météorologiques de manière à favoriser la mesure de plusieurs variables météorologiques à une même station.

Gandin (1970) propose une conception de réseau relativement similaire à celle rapportée par Rodda (1969) et découlant des normes générales qu'il suggère (tableau 3.2).

Selon lui, l'ensemble des stations peut être divisé en trois groupes, le premier comprenant celles où sont observés les phénomènes atmosphériques, la visibilité, la nébulosité, la précipitation, le couvert de neige, la température de l'air et du sol, le vent, la pression et la durée d'ensoleillement. La distance rationnelle entre les stations de ce groupe devrait être de 150 à 200 km, cette distance étant adaptée aux conditions physiques et géographiques de la région. De plus, parmi ces stations, il y aurait lieu de sélectionner un certain nombre de stations de référence desservant essentiellement deux objectifs: l'étude des changements climatiques à long terme et la correction des données des autres stations. Naturellement, les stations sélectionnées devraient déjà posséder de longues séries de données, leur appareillage être de première qualité et une attention spéciale devrait être portée à ces stations. Enfin, la probabilité d'un transfert, de la fermeture ou de la modification de l'environnement de la station devrait être aussi éloignée que possible. L'estimation de la densité rationnelle de ces stations de référence devrait être faite à partir de données concernant la structure statistique de valeurs moyennes mensuelles des variables météorologiques étudiées. De façon très approximative, ces stations devraient être situées à environ 300 à 400 km l'une de l'autre, de telle sorte qu'environ 20 à 25% des stations du groupe 1 soient des stations de référence.

Toujours d'après Gandin (1970), le second groupe de stations serait composé de stations où toutes les variables mesurées par les stations du groupe 1 le sont aussi, sauf la durée d'ensoleillement et la température du sol. De plus, la pression de l'air n'est généralement pas mesurée à ces stations, sauf en montagne. La distance rationnelle entre les stations du second groupe devrait être de 50 à 60 km, ce qui signifie qu'environ 10% des stations du second groupe sont aussi des stations du premier groupe.

Quant au troisième groupe de stations, que Gandin qualifie de postes, seules les observations portant sur les phénomènes atmosphériques, la préci-

pitation et le couvert de neige y sont prises. La distance rationnelle entre ces stations devrait être de 25 à 30 km en terrain plat et moitié moindre en zone montagneuse. Gandin note toutefois qu'il est très probable qu'avec le développement des radars météorologiques, ces densités de réseau pourront être diminuées.

Suite à ces différentes remarques, il apparaît nettement qu'il n'est pas souhaitable d'adopter une conception de réseau qui se traduise par une distribution régulière sur un territoire donné, de stations où toutes les variables météorologiques sont observées. Un réseau doit plutôt être conçu de façon souple, de manière à ce que la densité des points de mesure varie en fonction des objectifs poursuivis et des caractéristiques des variables observées.

3.3 LES TECHNIQUES DE RATIONALISATION

Dans le cas où aucune donnée représentative du territoire considéré ne serait disponible, seule l'application de normes générales devrait être considérée dans un premier temps. Si, au contraire, il existe au moins un nombre minimal de données, on peut penser à faire son choix parmi diverses approches ou techniques d'étude des données existantes, à moins que l'on ait déjà décidé d'appliquer des normes générales. L'Organisation météorologique mondiale a décrit brièvement ces techniques, en particulier dans deux rapports¹ édités respectivement en 1967 et 1972, le second reprenant le schéma suggéré par le premier, mais avec certaines différences.

¹ Rodda (1969). Hydrological Network Design - Needs, Problems and Approaches. WMO ref. no. 12 on WMO/IHD projects, 57 p.

Rodda (1972). Casebook on Hydrological Network Design Practice. WMO ref. no. 324.

Selon Rodda (1969), il existe deux problèmes scientifiques de base dans la planification des réseaux. Le premier est de déterminer combien de points d'acquisition de données sont nécessaires et le second est de savoir où les situer. Il ajoute plus loin qu'un troisième problème s'ajoute aux deux premiers, soit celui de déterminer la durée d'observation requise.

Les diverses approches ou techniques proposées ont pour but d'aider à répondre à ces questions. Elles peuvent se diviser en trois types principaux selon les rapports cités précédemment:

- la cartographie;
- la régionalisation;
- l'analyse de système.

Une brève description de ces techniques permettra de saisir certaines de leurs ressemblances et leurs différences.

3.3.1 La cartographie

La première technique consiste à cartographier la variable étudiée, c'est-à-dire à construire une carte décrivant la variation spatiale des valeurs prises par cette variable et ce, en traçant des isolignes basées sur l'interpolation des valeurs observées en certains points. Sous sa forme la plus simple, la cartographie a trait à une seule variable et uniquement les informations provenant du réseau de mesure de cette variable sont utilisées. On peut, par exemple, vérifier la précision de la carte résultante en comparant les écarts obtenus, en préparant la carte à partir d'un nombre variable de points de mesure. La cartographie d'une variable peut toutefois être améliorée par la prise en compte de caractéristiques physiographiques (à l'aide de régressions) ou de la structure statistique spatiale de la variable.

Une autre amélioration de la précision de la cartographie d'une variable particulière peut être obtenue par **cartographie conjointe** de plusieurs variables, de manière à respecter les liens entre ces

variables, par exemples, la précipitation, l'écoulement et l'évapotranspiration (Solomon, 1972; Nordenson, 1972). Encore là, les caractéristiques physiographiques peuvent intervenir dans la méthode pratique de cartographie.

Notons enfin que l'emploi d'une méthode basée sur une grille carrée (Solomon, 1972) peut faciliter la préparation de ces cartes.

3.3.2 La régionalisation

D'après Chapman (1972), la régionalisation peut être considérée comme étant n'importe quelle approche systématique de modélisation du cycle hydrologique pour une surface donnée en fonction de paramètres significatifs décrivant sa topographie, sa végétation, ses sols et sa lithologie. Benson (1972) considère que si l'échantillonnage a été bien conduit et adéquat, les équations (résultant dans son cas de régressions multiples) représentent des relations générales entre la variable considérée et l'environnement, qui sont applicables partout à l'intérieur de la région analysée et peuvent être utilisées pour estimer une caractéristique voulue aux points mesurés ou non, à l'intérieur de limites de précision connues.

La démarche suivie dans une étude de régionalisation de variables peut aussi amener à vouloir développer de façon coordonnée des régressions entre plusieurs caractéristiques hydrologiques, de même qu'avec le terrain, dans le but de délimiter des surfaces à l'intérieur desquelles il est possible d'utiliser les mêmes corrélations (régionales) pour estimer ces caractéristiques hydrologiques en plusieurs points (Solomon, 1972). On parle alors de **multi-régionalisation** et Solomon (1972) précise qu'il a appelé ces régions des "régions hydrologiques statistiques" dans les études qu'il a réalisées au Canada, afin de différencier les régions ainsi définies de celles qui le sont à partir de la physiographie.

3.3.3 L'analyse de système

L'étude et la planification rationnelle d'un réseau de mesure selon une approche basée sur l'analyse de système se traduisent par l'optimisation des moyens mis en oeuvre pour atteindre un but quelconque, sujet aux contraintes imposées au système. Au niveau des données rentrant dans l'analyse, cette approche se différencie des précédentes, en ce sens qu'en plus des variables hydrologiques et physiographiques déjà prises en compte par ces approches, elle nécessite aussi des données concernant les facteurs sociaux, politiques et économiques qui affectent le processus de décision.

Dans le cadre de cette approche, un réseau n'est pas uniquement constitué de stations où des données sont recueillies, mais aussi des modèles analytiques destinés à extrapoler et interpoler les données dans le temps et dans l'espace (Dawdy *et al.*, 1972). De plus, ces modèles analytiques servent à relier les réseaux entre eux et à identifier des structures équivalentes.

Ce type d'approche se prête donc davantage à des études de planification de réseaux spécifiques dans lesquelles les différents éléments nécessaires à l'optimisation du système sont connus ou peuvent être estimés de façon satisfaisante. Dans le cadre précis de la rationalisation de l'ensemble du réseau météorologique du Québec, il serait très difficile pour ne pas dire impossible présentement de rationaliser le réseau à partir de cette approche, les données nécessaires n'étant généralement pas connues ou ne pouvant être estimées avec assez de précision.

3.4 LES CRITERES DE CONTROLE D'ATTEINTE DES OBJECTIFS

L'application de techniques visant à planifier les réseaux nécessite le recours à des critères permettant de contrôler dans quelle mesure les objectifs fixés *a priori* sont ou peuvent être atteints et de comparer diverses structures de réseaux entre elles afin de juger laquelle est la plus sus-

ceptible de rencontrer ces objectifs. En fonction de la technique de rationalisation adoptée, mais surtout en fonction des données disponibles, les critères de contrôle peuvent se diviser en trois groupes, selon qu'ils ont trait à :

- l'atteinte d'une précision requise;
- l'optimisation bénéfices-coûts;
- l'optimisation précision-coûts.

3.4.1 L'atteinte d'une précision requise

Un réseau peut être planifié essentiellement sur la base de la précision requise. On fait alors abstraction, en première analyse, des implications économiques, logistiques ou autres, inhérentes à l'implantation et l'opération du réseau, pour ne s'attacher qu'à l'atteinte d'une précision requise déterminée **a priori** en fonction de divers besoins à satisfaire. Le critère de précision peut alors s'exprimer par rapport au réseau étudié. Ainsi, on peut désirer connaître les hauteurs de pluie journalière en tout point d'une région avec une précision de x mm. Naturellement, la précision requise initialement peut être revisée si son atteinte s'avère difficile, voire impossible.

Les données recueillies par un réseau particulier peuvent aussi servir à estimer d'autres variables. Par exemple, les hauteurs de pluie observées par un réseau de pluviomètres peuvent servir d'entrée à un modèle de simulation (déterministe ou statistique) utilisé pour préparer des prévisions de débits en certains points d'un bassin versant. L'objectif final est alors de réaliser une prévision de débits avec une précision donnée et non uniquement de connaître la pluie avec une précision donnée, ces précisions n'étant généralement pas les mêmes. Un critère exprimant la précision requise pour une variable dépendante peut, par conséquent, être préféré dans ce cas, comme dans d'autres. Le réseau de mesure de la variable indépendante est alors planifié en fonction de la précision voulue pour la variable dépendante et non en fonction de celle de la variable indépendante.

3.4.2 L'optimisation bénéfices-coût

Lorsque les données nécessaires sont disponibles, la planification d'un réseau peut être réalisée grâce à une approche permettant de prendre en compte les contraintes économiques, politiques, sociales et logistiques intervenant dans l'implantation et l'opération d'un réseau. Le nombre et la distribution géographique des stations sont alors optimisés de manière à maximiser les bénéfices nets. En d'autres termes, il s'agit de déterminer la densité optimale de stations résultant en des coûts et des pertes minimales (Gandin, 1970). Les bénéfices découlant de l'obtention de données additionnelles dans l'espace ou dans le temps peuvent aussi être comparés aux coûts nécessaires à leur obtention en requérant que les bénéfices ne soient pas inférieurs aux coûts (Rodriguez-Iturbe et Méjia, 1974).

Notons que ce type de critère permet de considérer un réseau comme essentiellement dynamique et susceptible de modifications en fonction de la variation des contraintes indiquées plus haut et de l'information déjà fournie.

3.4.3 L'optimisation précision-coût

Il est très souvent impossible d'appliquer un critère basé sur la maximisation des bénéfices nets, aussi doit-on penser à trouver d'autres façons d'exprimer les contraintes d'ordre économique ou autre, inhérentes à l'implantation et l'opération d'un réseau. Dans le type de critère décrit en 3.4.2, le problème est d'évaluer les bénéfices bruts résultant de l'opération du réseau et non de calculer les coûts qui lui sont associés. C'est la raison pour laquelle un critère liant la précision atteinte et le coût d'implantation et d'opération d'un réseau semble plus réaliste à certains auteurs (Rodriguez-Iturbe et Méjia, 1974; Rodda, 1969; Bras et Rodriguez-Iturbe, 1976). Ce critère précision-coût peut s'exprimer de deux façons selon que le coût ou la précision constitue la contrainte majeure:

- étant donné un budget déterminé, il s'agit de rechercher l'erreur minimale qui puisse être obtenue en modifiant le nombre et la distribution spatiale des stations d'un réseau;
- étant donné une précision requise, il s'agit de rechercher le coût minimal d'implantation et d'opération d'un réseau de stations permettant d'assurer cette précision.

3.5 L'INTEGRATION DES RESEAUX ENTRE EUX

Tout réseau météorologique et hydrologique peut être planifié de façon complètement indépendante. On peut alors trouver, pour ce réseau et sans égard aux autres, le nombre et la distribution spatiale de stations susceptibles de produire, par exemple, une erreur minimale pour un budget donné. A la limite, il pourrait toutefois y avoir un nombre total de stations égal à la somme du nombre de stations dans chacun des réseaux. Evidemment, tant qu'il s'agit de stations météorologiques, une intégration de fait des mesures de diverses variables météorologiques existe déjà et permet, en particulier, de limiter les coûts d'opération et d'étudier les relations entre diverses variables météorologiques en un point donné. Une telle intégration est effectivement proposée par Gandin (1970) au tableau 3.2.

L'intégration des réseaux météorologiques avec des réseaux hydrologiques (débits, qualité de l'eau) devrait aussi être envisagée (Ferguson, 1972). Ceci ne veut pas nécessairement dire de regrouper toutes les mesures à des stations données, mais plutôt de planifier chacun des réseaux en fonction de sa nature propre bien sûr, mais aussi en fonction de son interaction avec les autres réseaux, tant pour l'échantillonnage spatial que temporel. La planification intégrée des réseaux peut alors reposer sur la définition d'une hiérarchie entre les réseaux, selon laquelle plus un réseau est élevé dans la hiérarchie, plus les données qui y sont recueillies sont susceptibles d'utilisation à des fins multiples. Ainsi, si l'on considère des réseaux de mesure des précipitations, des débits et du transport solide en rivière, le réseau de précipitation sera le plus haut dans la hiérarchie, sui-

vi par le réseau hydrométrique. Le réseau de mesure des concentrations de solides en suspension sera le plus bas, l'interprétation des données qui y sont recueillies nécessitant des informations d'au moins un des deux premiers réseaux. En d'autres termes, il est possible de planifier un réseau hydrométrique sans se préoccuper du réseau de mesure des sédiments, mais l'inverse n'est pas vrai. De même, des réseaux pluviométriques et hydrométriques peuvent être planifiés de façon complètement indépendante, mais les planificateurs n'atteindraient pas ainsi une situation optimale. En effet, il est possible, grâce à des modèles mathématiques, de simuler les débits en divers points d'un bassin avec les précipitations et l'évapotranspiration comme données d'entrée. Dès lors, une planification optimale d'un réseau hydrométrique devrait tenir compte de ce fait. De même, à un pas de temps suffisamment long, l'écoulement mesuré sur un bassin peut fournir des renseignements sur les précipitations qui y sont tombées.

En conséquence, il y a lieu de s'interroger, lors d'une étude de planification rationnelle de réseau, sur l'opportunité de planifier le réseau selon une optique sectorielle ou plus ou moins intégrée.

3.6 LA PRISE EN COMPTE DES METHODES D'ANALYSE

Il est évident que si la "planification" d'un réseau consiste uniquement à vérifier si la densité de ce réseau respecte des normes générales, point n'est besoin de penser à prendre en compte les différentes méthodes d'analyse ou de simulation qui permettront de mieux valoriser l'information contenue dans les données observées. D'un autre côté, tenir compte de toutes les utilisations possibles de ces données pourraient vraisemblablement dans beaucoup de cas conduire à un échantillonnage spatial et temporel irréalisable en pratique. Entre ces deux extrêmes existe une solution qui peut consister à tenir compte dans la mesure du possible, en fonction des objectifs principaux du réseau exprimés région par région, des méthodes d'analyse ou de simulation les plus susceptibles d'être appliquées, en particulier celles qui peuvent exprimer des relations entre les données recueillies par différents réseaux. Ainsi, dans une région consistant essentiellement

en un bassin aménagé pour des fins hydro-électriques, il faudrait planifier un réseau de pluviomètres et pluviographes de telle sorte qu'un modèle hydrologique déterministe puisse simuler les débits et servir à préparer des prévisions avec une précision satisfaisante, dans le temps et dans l'espace, compte tenu des objectifs spécifiques reliés à la gestion du complexe hydro-électrique.

3.7 LA DUREE DES OBSERVATIONS

Une fois ouverte, une station doit-elle demeurer indéfiniment en opération ou peut-elle être fermée sans qu'une perte d'information importante et essentielle ne survienne? La réponse dépend évidemment du but poursuivi en ouvrant une station en un lieu donné ou en établissant un réseau dans une région donnée. Toute modification apportée à un réseau devrait résulter d'une décision prise après analyse des besoins et conséquemment du type de données nécessaires à leur satisfaction.

Ainsi, l'opération d'un système de prévisions peut requérir le maintien intégral d'un réseau de pluviomètres, aussi longtemps que les prévisions sont préparées. Il serait toutefois possible, si l'on constate que les données recueillies en certaines stations sont suffisamment corrélées entre elles, de tenir compte de la structure statistique de la variable pour supprimer des stations, sans affecter la précision de la prévision de façon importante.

Si le but visé est par exemple la connaissance de la hauteur moyenne de pluie sur une surface à un pas de temps de l'ordre du mois, de la saison ou de l'année, il est possible d'obtenir ces renseignements en implantant un petit nombre de stations pendant plusieurs années ou un nombre de stations plus grand pendant moins d'années. Rodriguez-Iturbe et Méjia (1974) présentent une méthode permettant d'évaluer le nombre m de stations en opération pendant n années nécessaires pour remplacer M ($M < m$) stations en opération pendant N ($N > n$) années, de façon à conserver la même précision. Ils font toutefois remarquer que ce remplacement peut s'avérer onéreux et n'est pos-

sible que si l'on ne réduit pas trop la durée d'observation, une diminution trop grande de cette durée conduisant à un réseau excessivement dense ($n \rightarrow \infty$).

Le concept d'équivalence temps-densité devrait toutefois être considéré pour certaines études importantes pour lesquelles il s'avère impossible d'attendre trop longtemps pour obtenir les données nécessaires.

3.8 L'INTEGRATION DES DONNEES AU SOL ET DES DONNEES ACQUISES PAR DE NOUVELLES TECHNIQUES

Ferguson (1972) considère qu'une attention spéciale doit être portée à l'impact des développements prévisibles en télédétection sur la planification des réseaux. Il précise qu'un système d'observation intégré reposant sur des mesures conventionnelles ou acquises par télédétection est une caractéristique importante de la Veille météorologique mondiale. Toujours selon lui, notre conception traditionnelle des réseaux devra être révisée à la suite des développements rapides qui se produisent en télédétection. Ce que Ferguson affirmait en 1972 est encore plus vrai aujourd'hui. Les différents satellites en orbite actuellement permettent déjà d'acquérir une foule de renseignements sur les variables météorologiques. Ainsi, les satellites géostationnaires, aidés en cela par les satellites à déplacement de type NOAA, TIROS, NIMBUS et autres, peuvent suivre le déplacement des masses nuageuses et même permettre d'estimer certaines des caractéristiques de l'atmosphère et de ces nuages, fournissant par le fait même des indications importantes sur les précipitations. Des informations précieuses sont aussi recueillies par ces satellites sur la couverture de neige au sol et les températures de surface. Bien sûr, ces satellites ne peuvent actuellement remplacer complètement les réseaux conventionnels de mesures au sol, et il serait nécessaire de les implanter s'ils n'existaient pas, de manière à fournir des données servant au contrôle et à l'interprétation des observations réalisées à bord de ces satellites.

D'autre part, il faut aussi être conscient et au courant des programmes de mise en orbite de nouveaux satellites portant un plus grand nombre de capteurs capables de fournir des informations plus précises sur des surfaces plus petites.

L'acquisition et le traitement de ces données doivent aussi être envisagés. Ces satellites fournissent déjà une masse de données énorme dont on ne tire déjà pas tout le potentiel. La situation risque donc de se détériorer dans l'avenir, à moins que des mesures ne soient prises pour tirer le meilleur parti possible de ces données. Sur ce point, il faut toutefois être réaliste en considérant que le contrôle se retrouve essentiellement aux USA, quoique des aménagements soient possibles pour la réception directe des images et leur traitement.

Un autre point à considérer est la possibilité que les satellites prévus n'aient qu'une existence éphémère par suite d'un trouble majeur ou que certains de leurs capteurs ne fonctionnent pas.

En ce qui a trait essentiellement aux précipitations, il faut ajouter une autre source d'informations très importante, la présence d'installations de radar. Le radar permet en effet de connaître la distribution spatiale des précipitations de même que leur intensité avec une précision relativement satisfaisante. Le couplage d'installations de radar, dans les régions couvertes par ces appareils, et de stations conventionnelles de mesures des précipitations devrait conduire à une réduction du nombre de stations nécessaires pour atteindre la précision désirée pour l'estimation des pluies ponctuelles ou moyennes sur une surface donnée. Cette opinion est partagée tant aux USA qu'en URSS, au Royaume Uni et au Canada (Eddy, 1976; Brady, 1978; Rumyantsev et Shanochkin, 1973; O'Connell **et al.**, 1978; Ferguson, 1972).

En résumé, compte tenu de la disponibilité accrue de données de plus en plus valables acquises par télédétection (satellites ou installations de radar), il importe de planifier les réseaux de manière à mieux intégrer les

renseignements fournis par la télédétection à ceux qui proviennent des mesures conventionnelles. Il faut alors considérer que si cette intégration peut amener une réduction du nombre de stations classiques nécessaires à l'atteinte d'un objectif donné, il ne saurait être question de se fier essentiellement sur la télédétection, au moins dans un avenir prévisible.

Il s'agit en fait de se définir une approche rationnelle à la planification des réseaux permettant une intégration optimale des diverses sources de données.

CHAPITRE IV

LA RATIONALISATION DES RESEAUX: DEMARCHE PROPOSEE

Compte tenu des objectifs définis pour l'étude du réseau météorologique du Québec en vue de sa rationalisation et des choix qui s'offrent pour atteindre ces objectifs, la démarche suivante est proposée.

1. L'application de normes générales n'est pas retenue, parce que ces normes ne constituent pas à proprement parler une rationalisation de réseau et que leur application ne devrait être considérée qu'en dernier ressort, lorsque l'emploi d'autres méthodes conduisant à des décisions plus adaptées aux besoins s'avère impossible.
2. L'évolution du réseau météorologique du Québec dépendant du mandat confié au ministère des Richesses naturelles et ce mandat spécifiant que ce ministère aurait "la responsabilité exclusive de l'observation, la compilation, l'analyse et la publication de tous les phénomènes météorologiques et climatologiques pour les besoins de tous les ministères du gouvernement de la province de Québec", il a paru nécessaire de réaliser tout d'abord une enquête visant à mettre à jour l'ensemble de besoins québécois non seulement des ministères mais aussi des organismes publics et para-publics. D'autre part, ces besoins étant bien souvent spécifiques à certaines régions du Québec, un réseau météorologique très souple, dont la densité puisse varier dans l'espace et dans le temps, est préconisé pour répondre à ces besoins. Sa conception devrait s'inspirer largement des recommandations rapportées en 3.2.
3. L'approche idéale consisterait à utiliser l'analyse de système, cette approche permettant de tenir compte des diverses contraintes inhérentes à l'évolution du réseau météorologique du Québec. Cependant, compte tenu des objectifs multiples poursuivis à l'intérieur du mandat

du ministère des Richesses naturelles et de l'impossibilité pratique d'évaluer quantitativement avec une précision satisfaisante le poids de ces diverses contraintes dans le processus de décision, cette approche n'est pas retenue.

D'autre part, l'étude des approches possibles ayant été menée en parallèle avec celle des méthodes pratiques d'analyse des données, la multi-régionalisation est considérée comme étant l'approche la plus appropriée, dans le cas présent, pour appliquer la conception de réseau décrite précédemment et satisfaire les hypothèses de base sur lesquelles sont fondées les diverses méthodes d'analyse des données.

4. Dans le cadre de la présente étude, des critères basés essentiellement sur la précision pour un réseau donné sont utilisés pour analyser les réseaux. Conséquemment, les réseaux sont étudiés essentiellement d'un point de vue météorologique, sans tenir compte de façon explicite des coûts ou des bénéfices nets rapportés. Il est toutefois suggéré au Service de météorologie d'appliquer éventuellement d'autres critères, en fonction de la disponibilité des données permettant de faire appel à ces critères. A cet égard, un critère de type précision-coûts devrait être approprié pour rechercher la configuration de réseau susceptible de conduire à une erreur minimale pour un budget déterminé et par suite d'évaluer les modifications possibles au réseau en fonction des budgets alloués à chaque année.
5. L'intégration des réseaux météorologiques entre eux et de ceux-ci avec d'autres réseaux d'observation de l'environnement est souhaitable et devrait être mise en

pratique lorsque possible selon l'optique exposée en 3.5. En tout état de cause, le choix du site d'une station et de la fréquence des observations qui y seront effectuées, devrait tenir compte au moins implicitement de la valeur potentielle de ces observations pour l'interprétation d'autres variables environnementales.

6. Dans le processus de modification des réseaux découlant de ce rapport, une attention particulière devrait être portée dans la mesure du possible aux méthodes d'analyse ou de simulation, selon le point de vue exprimé en 3.6. Cet effort devrait permettre de tirer le maximum des observations réalisées, grâce à une meilleure cohésion de la prise des données et des méthodes d'analyse.
7. En ce qui concerne la durée des observations, une conception dynamique des réseaux est suggérée de manière à atteindre puis maintenir un réseau optimal pour chaque variable météorologique, tenant compte de tous les éléments de la démarche proposée pour la rationalisation du réseau météorologique du Québec.
8. L'intégration des données fournies par la télédétection aux données acquises par les réseaux traditionnels, est fortement conseillée dans le respect des remarques apparaissant en 3.8. Plus particulièrement, dans le sud du Québec, on devrait mettre à profit le plus rapidement possible la présence d'installations de radar à Montréal et à Québec.

CHAPITRE V

METHODES ANALYTIQUES

5.1 ESTIMATION DES VALEURS PONCTUELLES

Il n'est pas inutile de rappeler ici que l'objectif poursuivi par la collecte de données météorologiques à un certain nombre de stations formant un réseau, est l'estimation des valeurs prises par une variable météorologique en tout point du champ considéré. D'où la nécessité d'interpoler les valeurs observées aux stations. On est dès lors amené à se demander quelle est la précision de cette interpolation. Plusieurs méthodes d'interpolation ont été suggérées et appliquées jusqu'à maintenant, mais si toutes permettent d'estimer une valeur prise par une variable en un point donné, toutes ne permettent pas d'estimer aussi l'erreur d'interpolation commise. C'est toutefois précisément ce qu'il faut pour analyser dans quelle mesure un réseau permet de répondre aux objectifs fixés et pour déterminer les modifications susceptibles d'améliorer la précision de ce réseau. L'objectif de ce chapitre est donc de présenter les méthodes d'analyse qui ont été retenues dans le cadre de l'étude du réseau météorologique du Québec, tant pour l'estimation des valeurs ponctuelles que pour celle des valeurs moyennes sur une surface d'étendue donnée. Cette présentation est précédée d'une brève revue de la littérature visant à situer ces méthodes par rapport à d'autres.

5.1.1 Revue de la littérature

Lenton et Rodriguez-Iturbe (1977) écrivent que ce serait fou, selon eux, de tenter de passer en revue tout ce qui s'est publié sur l'analyse des réseaux depuis que l'on s'intéresse à ce problème. Aussi, préfèrent-ils s'en tenir à rappeler essentiellement les travaux antérieurs importants effectués dans le cadre spécifique de leurs propres travaux. C'est en gros l'attitude adoptée pour le présent rapport, avec une nuance, cependant: l'addition (en fin de rapport) d'une bibliographie comprenant un certain nombre de publications non citées dans le cours du rapport, mais portant sur divers aspects de l'analyse des réseaux.

L'une des premières méthodes d'analyse à laquelle on puisse penser est l'étude des corrélations entre stations. Ces corrélations peuvent être exprimées de différentes manières, mais le résultat est le même à peu de choses près. Aussi, on peut cartographier les coefficients de corrélation entre diverses stations et une station centrale. Les lignes d'isocorrélation prennent alors généralement la forme d'ellipses allongées dans le sens du vent dominant. Si cette opération est répétée à plusieurs reprises en changeant de station centrale à chaque fois, il devient alors possible de planifier un réseau tel que tout point considéré soit, par rapport à une station donnée, à l'intérieur d'une ligne d'isocorrélation déterminée. Hershfield (1965) et Hendrick et Comer (1970) choisissent $r = 0.9$, désirant ainsi que 81% de la variance des valeurs prises par une variable en un point soient expliqués. Le tracé des ellipses correspondant à $r = 0.9$ permet alors facilement de planifier un réseau (en veillant à ce que les ellipses aient un recouvrement optimal) et même de tenir compte de la variabilité saisonnière des phénomènes qui se traduit par une variation plus ou moins rapide des coefficients de corrélation avec la distance et donc de la dimension de ellipses, selon les saisons.

On peut aussi produire un graphique exprimant la variation moyenne, en fonction de la distance inter-stations, des coefficients de corrélation entre toutes les paires de stations d'un réseau (Hutchinson, 1970; Haragan, 1976; Rodriguez-Iturbe et Méjia, 1974; Yefremova, 1970; Guschina, **et al.**, 1967; Gandin, 1970; Huff et Schickedanz, 1972; Sharon, 1972 a et b; O'Connell, 1977). On obtient alors des courbes qui varient principalement en fonction de la région étudiée et de la saison. Dans le cas des précipitations, leur intensité et leur nature (orage, pluie continue...) ont aussi une influence sur ces courbes. Par ailleurs, la description de la structure statistique spatiale d'une variable par des relations corrélation-distance ou covariance-distance sert souvent de base à des méthodes plus ou moins complexes d'analyse des réseaux (Rodriguez et Méjia, 1974; Guschina, **et al.** 1967; Gandin, 1965 et 1970...).

L'analyse des corrélations n'est cependant pas la seule méthode à laquelle les météorologistes se sont arrêtés afin de connaître la structure du champ d'une variable. En prévision, particulièrement, se sont développées diverses méthodes dites d'analyse objective des champs météorologiques. En pratique, l'analyse objective vise à reconstruire la structure d'un champ à partir d'un ensemble fini de mesures. Royer (1975) et Delfiner (1973) considèrent donc que le problème de l'analyse objective se réduit essentiellement à un problème d'interpolation et précisent que plusieurs méthodes mathématiques, empiriques et statistiques sont disponibles pour ce faire. En accord avec Gandin (1965) et Kruizinga et Yperlaan (1978), Royer (1975) écrit que lorsqu'un réseau est suffisamment dense, la méthode d'analyse importe peu et que toutes les méthodes donnent sensiblement les mêmes résultats. Il ajoute toutefois qu'en théorie, la méthode d'interpolation optimale (Gandin, 1965) est supérieure aux autres méthodes linéaires car elle fait intervenir de façon explicite les propriétés statistiques du champ et n'introduit aucun paramètre arbitraire. Toujours selon Royer (1975), cette méthode possède en outre un certain nombre d'avantages pratiques, dont la possibilité de calculer la variance de l'erreur d'interpolation en chaque point. Gandin (1970) explique, par ailleurs, que dans la plupart des méthodes, les poids attribués aux observations le sont de façon plus ou moins arbitraire, en supposant, par exemple, des poids variant en fonction de l'inverse de la distance au point calculé. Il précise que dans sa méthode, au contraire, les poids attribués aux observations sont déterminés dans sa méthode, sur la base d'informations concernant la structure statistique et la précision des observations, en posant comme condition que l'erreur-type d'interpolation, avec ce choix de poids, soit moindre qu'avec n'importe quel autre choix de poids.

Une revue des articles parus sur l'analyse objective des champs météorologiques au cours des douze dernières années indique, par ailleurs, une tendance très nette vers les méthodes statistiques (Eddy, 1967; Kruger, 1969; Rutherford, 1972; Petersen, 1973; Thiebaut, 1974; Schlatter, 1975; Brady, 1978...). De plus, sous une forme ou sous une

autre, ces articles traitent d'interpolation optimale selon l'optique de Gandin. Nous avons donc été nécessairement amenés à considérer sérieusement l'application de cette méthode pour étudier le réseau météorologique québécois, d'une part, parce que tous les auteurs s'entendent pour considérer ses bases théoriques sûres et son aptitude à déterminer la variance de l'erreur d'interpolation en chaque point et, d'autre part, parce que cette méthode sert de base aux systèmes de prévisions numériques actuels.

Une autre raison nous a aussi incités à adopter la méthode de l'interpolation optimale: le développement de la théorie du krigeage en géostatistique par Matheron (1965). Il est intéressant de noter ici que Matheron (1965) définit la géostatistique comme étant l'application de la théorie des variables régionalisées à l'estimation des gisements miniers. Il précise, par ailleurs, qu'il qualifie de **régionalisé**, un phénomène qui se déploie dans l'espace et y manifeste une certaine structure. En ce qui a trait au krigeage, il ajoute qu'il consiste à trouver la meilleure estimation possible de la teneur d'un panneau, compte tenu de l'information disponible. Toujours selon Matheron (1965), le krigeage revient donc à attribuer un poids à la teneur de chaque échantillon, ces poids étant calculés de manière à rendre minimale la variance d'estimation résultante, ce qui a pour effet de tirer le meilleur parti possible des informations disponibles ou, si l'on préfère, d'obtenir l'estimation la plus précise possible du panneau en cause. On retrouve donc là, appliquée à l'estimation des gisements miniers, une méthode qui possède les mêmes caractéristiques que l'interpolation optimale pour l'analyse des réseaux météorologiques. Mais il y a plus. Delfiner (1973) et Delhomme et Delfiner (1973) ont appliqué le krigeage pour l'estimation de variables météorologiques. Par ailleurs, Delfiner (1973) et Chauvet **et al.**, (1976) considèrent que le krigeage universel est une généralisation de l'interpolation optimale selon Gandin (1965). En d'autres termes, l'interpolation optimale est identique au krigeage simple, krigeage dans lequel la moyenne de la variable est considérée constante à l'intérieur des limites du champ

étudié. L'emploi de l'une ou l'autre méthode revient donc essentiellement au même, à condition que l'hypothèse d'une moyenne constante soit respectée.

En résumé, deux méthodes développées à l'origine pour répondre à des besoins d'estimation dans des domaines de recherche différents, l'interpolation optimale en météorologie et le krigeage en géostatistique, sont, pour l'essentiel, fondées sur les mêmes hypothèses et possèdent les mêmes propriétés à toutes fins pratiques. Cette convergence de pensée pour résoudre les problèmes d'estimation suggère fortement que ces méthodes, sans être nécessairement supérieures aux autres sur tous les points, reposent sur des bases théoriques relativement sûres et peuvent être utilisées avec confiance pour une analyse de réseau.

5.1.2 Choix et conditions d'application d'une méthode d'analyse

La méthode d'interpolation optimale développée par Gandin (1965), servant déjà de base à plusieurs systèmes de prévisions numériques à travers le monde, a été retenue, de préférence au krigeage, pour réaliser l'essentiel des analyses de réseaux prévues. Le krigeage aurait cependant pu être utilisé tout aussi bien.

Comme toute méthode de calcul, l'interpolation optimale repose sur des hypothèses. En particulier, les calculs sont simplifiés de beaucoup si le champ étudié s'avère homogène et isotrope par rapport à la fonction employée. Idéalement donc, l'interpolation optimale doit être appliquée à des champs homogènes et isotropes. Sur ce point, Gandin (1965) souligne que si les écarts par rapport à la moyenne sont étudiés plutôt que les valeurs originales, les champs sont davantage homogènes et isotropes, du moins en altitude. La surface du sol introduit en effet divers effets qui tendent à rendre les champs moins homogènes et isotropes en surface.

Compte tenu des dimensions du Québec et des variations de phy-

siographie rencontrées, de même que de la proximité plus ou moins grandes d'étendues d'eau importantes par rapport à certaines stations, il a paru évident dès le début que les données recueillies par le réseau météorologique québécois n'étaient pas homogènes spatialement. D'où le désir de travailler non pas à la grandeur du Québec, mais de diviser l'ensemble du Québec en régions à l'intérieur desquelles les données pourraient être considérées comme homogènes, cette division en régions ayant l'avantage espéré de mieux faire ressortir les caractéristiques régionales qui auraient risqué d'être masquées dans une étude globale de l'ensemble du réseau.

Définir des régions à l'intérieur desquelles les stations sont homogènes par rapport à celles qui sont situées à l'extérieur, demeure toutefois un problème. On peut tenter de le faire plus ou moins arbitrairement, en se basant, par exemple, sur une carte de pluies totales annuelles, ou sur la physiographie. Il existe toutefois une méthode d'analyse des données dont nous n'avons pas parlé jusqu'ici explicitement et qui peut solutionner ce problème de façon relativement satisfaisante: l'analyse en composantes principales (Dyer, 1975). Par ailleurs, Charbonneau **et al.**, (1978 et 1979) ont vérifié avec succès la possibilité d'utiliser ce type d'analyse afin de définir des groupes de stations homogènes pour l'application de l'interpolation optimale. On démontre dans les sections 5.1.3 et 5.1.4 comment ces deux méthodes se complètent pour fournir une analyse des caractéristiques statistiques régionales du réseau météorologique du Québec.

Quant à l'isotropie, ou plutôt à l'anisotropie, il est possible d'en tenir compte en calculant les fonctions désirées selon l'orientation du segment joignant deux stations. On obtient alors des valeurs différentes des paramètres des fonctions, selon l'orientation choisie. En pratique, on choisit deux classes d'orientations, l'une parallèle à l'orientation du vent dominant et l'autre perpendiculaire. Si suffisamment de stations sont disponibles et que cela soit justifié, on peut cependant prendre plus d'orientations. Toutefois, comme on le verra

dans ce rapport, beaucoup de régions déterminées à l'aide des composantes principales ne regroupent qu'un petit nombre de stations. Dans les circonstances, la structure statistique de la variable ne serait pas définie avec assez de précision si l'anisotropie était prise en compte. Dans le cadre de l'étude du réseau météorologique du Québec, on suppose donc que l'hypothèse d'isotropie s'applique au moins de façon approximative. Par contre, la forme des régions de même que celle des isocorrélations présentées dans les parties de l'étude qui traitent de la climatologie statistique des variables, permettent d'avoir une idée qualitative de l'importance de l'anisotropie pour une région donnée et d'interpréter les résultats présentés en conséquence.

5.1.3 Analyse en composantes principales

5.1.3.1 Objectif

L'objectif essentiel de l'analyse des données en composantes principales, dans le cadre de l'étude actuelle du réseau météorologique du Québec, est de déterminer des groupes de stations relativement homogènes, caractérisant des "régions météorologiques statistiquement homogènes" dont la structure statistique sera étudiée plus en détails par interpolation optimale dans le but de planifier le réseau régional.

Cependant, cette étude d'un réseau régional par interpolation optimale peut être complétée par une seconde analyse des données en composantes principales, mais cette fois au niveau régional et non au niveau de l'ensemble du Québec, qui permettra d'étudier le comportement relatif des stations du réseau régional. Il faut comprendre ici que les stations qui ont été considérées comme homogènes lors de l'analyse simultanée de l'ensemble des stations du Québec et conséquemment regroupées, ont, les unes par rapport aux autres, des comportements plus ou moins différents ou similaires. C'est ce comportement qu'une seconde analyse en composantes principales permettra de faire ressortir.

Cette analyse vient donc compléter, lorsque désiré, les informations fournies par l'interpolation optimale, notamment sur le comportement similaire de deux ou plusieurs stations autorisant la fermeture d'une ou plusieurs stations, si besoin est. Il est à noter que cette analyse régionale des données à l'aide des composantes principales n'a pas été réalisée dans le cadre de la présente étude, mais pourrait l'être au moment de mettre en application les résultats de l'étude en cours.

La section 5.1.3.2 présentera donc la théorie de l'analyse en composantes principales, tant en ce qui regarde la formation de groupes de stations relativement homogènes que l'étude des comportements relatifs des stations à l'intérieur de ces groupes.

5.1.3.2 Théorie

Ce n'est pas notre intention de rappeler ici en détail la théorie de l'analyse en composantes principales; ces renseignements peuvent facilement être obtenus dans la littérature (Hotelling, 1933; Kendall, 1957; Anderson, 1958; Matalas et Reither, 1967). Nous nous limiterons à une description des caractéristiques principales de cette théorie, utile dans le cadre de notre étude.

Un groupe de p variables, comptant chacune n observations, peut être représenté par une matrice de dimensions $p \times n$. Chaque observation peut aussi être située dans un espace à p dimensions. Cependant, comme les variables peuvent être corréllées, les p axes de cet espace ne sont pas toujours orthogonaux. L'objectif de l'analyse en composantes principales est de transformer les variables de telle sorte que les axes deviennent orthogonaux, ce qui permet alors la définition de nouvelles variables qui, elles, sont indépendantes. Pour ce faire, le premier axe est choisi de manière à expliquer le plus possible la variance totale des observations, le second de manière à expliquer le plus possible la variance résiduelle, et ainsi de suite. En

d'autres mots, le premier axe est défini de telle sorte que la somme des carrés des projections des points sur cet axe soit maximale. Ensuite, les points sont projetés sur un plan orthogonal au premier axe et le second axe est lui aussi obtenu en exigeant que la somme des carrés des projections des points sur cet axe soit maximale et ainsi de suite pour les autres axes, chacun de ces axes expliquant un pourcentage de plus en plus faible de la variance totale.

Une fois cela fait, on peut déterminer les coefficients de corrélation r_{jk} entre la station (ou variable) j et la composante (ou axe) principale d'ordre k :

$$r_{jk} = C_{jk} \sqrt{\lambda_k} \quad \text{Eq. 5.1}$$

C_{jk} est le vecteur propre entre la station j et la composante principale k et λ_k est la valeur propre de la composante k .

L'analyse de la variation de r_{jk} fournit des informations précieuses sur le comportement relatif des stations. Plus précisément, si les coefficients de corrélation entre chacune des stations et la première composante principale sont comparés, on observe généralement qu'ils diffèrent de station en station. La même chose est vraie par rapport aux autres composantes et peut être vérifiée à l'aide de l'exemple présenté au tableau 5.1, tiré d'une application antérieure des composantes principales à des précipitations par Charbonneau **et al.**, (1978). Il pourrait se trouver que les r_{jk} soient ou identiques pour deux plusieurs stations ce qui impliquerait des corrélations parfaites entre ces stations. C'est toutefois rarement le cas par suite d'erreurs d'observation et d'irrégularités microclimatiques (Gandin, 1970). La solution est alors de supposer que les stations dont les coefficients de corrélation avec les q premières composantes principales ont des comportements similaires et peuvent être regroupées ensemble. Le nombre

Tableau 5.1 Coefficients de corrélation entre les composantes principales et les stations, calculés à partir de 90 valeurs de pluie cumulée sur 10 jours, en été, pour un réseau de 14 stations situées sur le bassin de la rivière Eaton¹ (avant rotation des axes) (d'après Morin et al., 1978).

Numéro de la station	composante principale			coefficient de corrélation multiple
	No. 1	No.2	No.3	
1	0.786	0.488	0.054	0.927
2	0.925	- 0.275	0.065	0.968
3	0.859	0.165	0.371	0.950
4	0.846	- 0.396	0.187	0.953
5	0.856	0.329	0.275	0.957
6	0.903	- 0.353	0.137	0.980
7	0.919	0.121	- 0.248	0.960
8	0.908	0.272	- 0.066	0.951
9	0.920	- 0.068	- 0.240	0.953
10	0.917	- 0.196	- 0.121	0.946
11	0.912	- 0.191	0.039	0.933
12	0.950	0.074	0.092	0.957
13	0.942	0.097	- 0.158	0.960
14	0.921	- 0.000	- 0.127	0.930
Valeur propre	11.244	0.916	0.465	
Variance expliquée	80.8%	6.5%	3.3%	
Totale variance expliquée	80.8%	87.4%	90.7%	

de composantes à considérer n'est pas défini de façon rigoureuse. En général, ce choix est basé sur le pourcentage de la variance totale expliqué après considération des q premières composantes.

Comme il est difficile d'identifier des comportements similaires uniquement à partir de tableaux, une représentation graphique de ces r_{jk} est souhaitable. Habituellement, ces r_{jk} sont pointés sur un graphique tri-dimensionnel, chaque axe étant attribué à une composante, ce qui oblige à ne considérer que trois composantes. Il existe toutefois une autre présentation graphique de ces résultats (Duband, 1973 et 1974). Selon cette méthode (figure 5.1), les coordonnées X_{j1} et Y_{j1} du point A du vecteur \vec{AB} (représentant la station j) dans le plan des composantes 2 et 3, sont déterminées comme étant égales aux valeurs respectives des coefficients de corrélation entre la station j et les composantes 2 (axe 2) et 3 (axe 3). On a donc, à partir de l'équation 5.1:

$$X_{j1} = C_{j2} \sqrt{\lambda_2}$$

Eq. 5.2

$$Y_{j1} = C_{j3} \sqrt{\lambda_3}$$

Le vecteur \vec{AB} doit alors être tracé le long d'une ligne orthogonale au vecteur \vec{OA} , de manière à respecter l'orthogonalité des composantes principales. Si l'on considère que la longueur du vecteur \vec{AB} est égale à la valeur prise par le coefficient de corrélation r_{j1} entre la station j et la première composante principale, les coordonnées X_{j2} et Y_{j2} du point B sur cette ligne sont déterminées par:

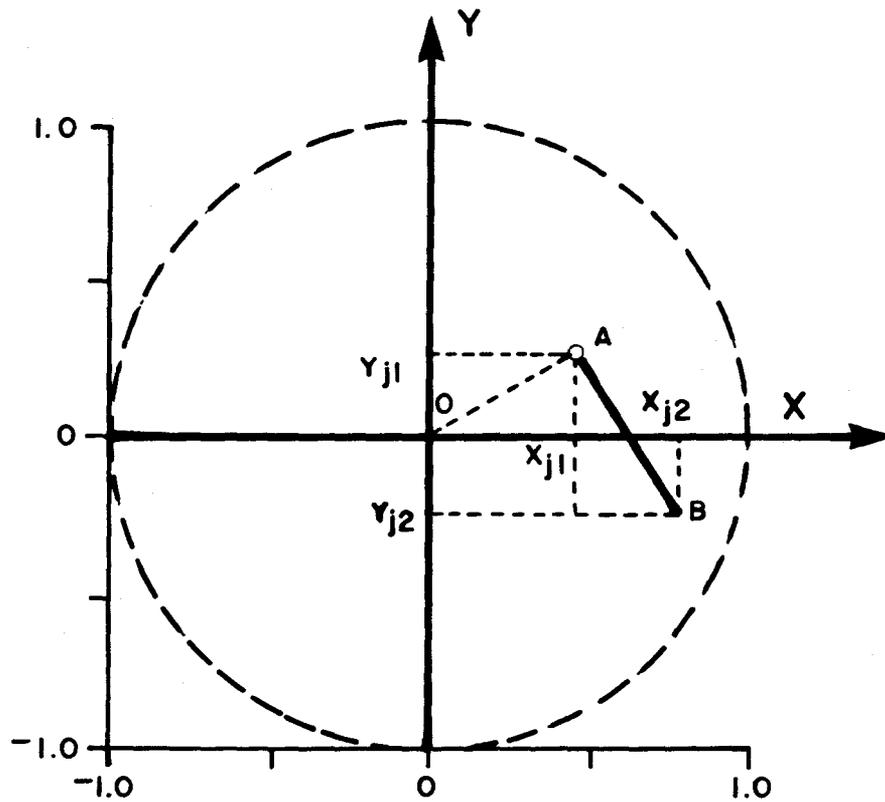


Figure 5.1 . Représentation graphique des coefficients de corrélation entre la station j et les q premières composantes principales, dans le plan des composantes principales 2 (axe X) et 3 (axe Y) . (d'après Morin et al., 1978) .

$$X_{j2} = X_{j1} \pm \frac{C_{j1} \sqrt{\lambda_1}}{1 + \left(\frac{X_{j1}}{Y_{j1}}\right)^2} \quad \text{Eq. 5.3}$$

$$Y_{j2} = Y_{j1} - \frac{X_{j1}}{Y_{j1}} (X_{j2} - X_{j1})$$

L'équation 5.3 peut être déduite facilement de la figure 5.1 et les signes plus ou moins proviennent de coefficients de corrélation positifs ou négatifs. Le vecteur \vec{OB} représente le coefficient de corrélation multiple entre la station j et les trois premières composantes. En procédant de la même façon, on peut tracer le nombre de composantes désiré. Finalement, le cercle centré sur l'origine des axes est le lieu des coefficients de corrélation multiples égaux à un.

Suite au traçage sur le même graphique de tous les vecteurs représentant les p stations, il devient possible d'identifier des stations de comportement similaire par la proximité relative de certains vecteurs.

Si tous les vecteurs se situent près les uns des autres, les stations peuvent être considérées comme appartenant à un groupe de stations homogènes. Plus souvent, toutefois, les vecteurs sont distribués dans les quatre quadrants, formant des groupes plus ou moins distincts. La figure 5.2, tirée du rapport cité précédemment pour le tableau 5.1, présente un exemple d'une telle distribution.

Dès lors, il est possible de poser l'hypothèse que les stations dont les vecteurs sont relativement près les uns des autres, comparés aux autres vecteurs, sont relativement homogènes et peuvent être regroupés entre elles. Il est important de noter ici que cette première manière de former des groupes de stations est nécessairement plus ou

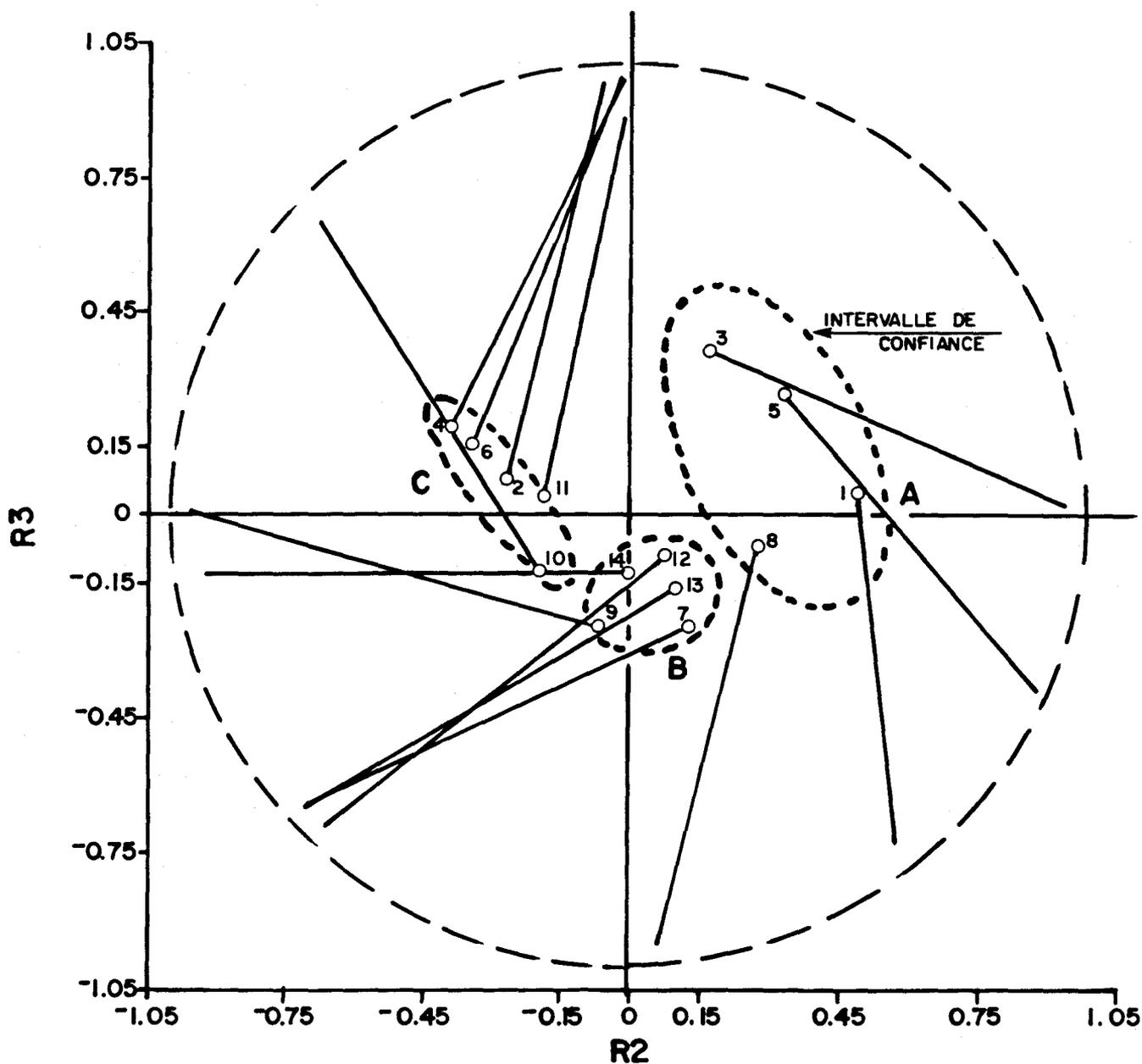


Figure 5.2 . Représentation dans le plan des coefficients de corrélation entre le réseau de 14 stations et les trois premières composantes principales pour les pluies cumulées sur 10 jours en été selon le tableau 5.1 . Les intervalles de confiance au niveau de la probabilité de 80% sont tracés pour les groupes de stations définis après la rotation effectuée par la méthode VARIMAX (d'après Morin et al., 1978).

moins arbitraire.

Une façon de définir ces groupes plus objectivement consiste à faire subir une rotation aux axes principaux de manière à redistribuer le pourcentage total de variance expliquée plus également entre les q composantes (axes). C'est la méthode VARIMAX qui a été retenue pour ce faire. Comme précédemment, les coefficients de corrélation entre les nouveaux axes principaux et les stations peuvent être calculés. Suite à cette rotation, pour chaque station, on identifie le coefficient de corrélation le plus élevé entre cette station et l'un des axes. Les stations sont alors regroupées selon l'axe avec lequel leur corrélation est la plus élevée, formant autant de groupes qu'il n'y a d'axes (Dyer, 1975). On peut, de plus, détecter des sous-groupes par l'analyse des coefficients de corrélation entre les stations et les autres axes. Le tableau 5.2 présente un exemple de coefficients de corrélation entre des stations et des composantes principales, obtenu après rotation des axes (Charbonneau *et al.*, 1978).

Une autre étape dans l'analyse des groupes est la détermination d'intervalles de confiance, représentés par des ellipses, confirmant l'appartenance ou la non-appartenance d'une station quelconque à un groupe défini objectivement. Pour ce faire, ayant posé l'hypothèse que m stations spécifiques font partie d'un groupe de stations comptant au moins trois stations, on trace les intervalles de confiance dans le plan des axes X et Y , tels qu'estimés par une loi normale à deux variables:

$$f(x,y) = \frac{\exp \left\{ \frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\bar{x}}{\sigma_x} \right)^2 - 2\rho \left(\frac{x-\bar{x}}{\sigma_x} \right) \left(\frac{y-\bar{y}}{\sigma_y} \right) + \left(\frac{y-\bar{y}}{\sigma_y} \right)^2 \right] \right\}}{2 \pi \sigma_x \sigma_y \sqrt{1-\rho^2}}$$

Eq. 5.4

Tableau 5.2 Coefficients de corrélation entre les composantes principales et les stations, calculés à partir de 90 valeurs de pluie cumulée sur 10 jours, en été, pour un réseau de 14 stations situées sur le bassin de la rivière Eaton* (après rotation par la méthode VARIMAX) (d'après Morin et al., 1978)

Numéro de la station	Composantes principales		
	No 1	No 2	No 3
1	0.156	0.480	<u>0.778*</u>
2	<u>0.796*</u>	0.433	0.341
3	0.530	0.215	<u>0.759</u>
4	<u>0.871</u>	0.271	0.279
5	0.381	0.317	<u>0.820</u>
6	<u>0.860</u>	0.350	0.313
7	0.411	<u>0.744*</u>	0.447
8	0.350	0.614	<u>0.634</u>
9	0.551	<u>0.708</u>	0.324
10	<u>0.677</u>	0.591	0.296
11	<u>0.719</u>	0.459	0.378
12	0.511	<u>0.628</u>	0.511
13	0.470	<u>0.680</u>	0.489
14	0.536	<u>0.629</u>	0.427

* Les coefficients de corrélation soulignés identifient les stations appartenant au groupe 1, 2 ou 3, selon la composante principale avec laquelle leur coefficient de corrélation est le plus élevé.

où:

x , \bar{x} et σ_x sont respectivement:

les valeurs individuelles, les valeurs moyennes et l'écart type des coefficients de corrélation entre les m stations, supposées appartenir à un groupe, et la composante principale 2;

y , \bar{y} et σ_y ont les mêmes définitions par rapport à la composante principale 3;

ρ est le coefficient de corrélation entre les valeurs de x et les valeurs de y .

Pour chaque groupe, le centre de l'ellipse a les coordonnées \bar{x} et \bar{y} . Une transformation d'axe est alors réalisée de telle sorte que:

$$\begin{aligned} u &= (x - \bar{x}) \cos \theta + (y - \bar{y}) \sin \theta \\ v &= (x - \bar{x}) \sin \theta + (y - \bar{y}) \cos \theta \end{aligned} \quad \text{Eq. 5.5}$$

Les nouvelles variables u et v sont alors orthogonales, pourvu que l'angle θ entre le système d'axe initial et le nouveau système d'axes soit égal à:

$$\theta = \text{arc tg} \left(\frac{2 \rho \sigma_x \sigma_y}{\sigma_x^2 - \sigma_y^2} \right) \quad \text{Eq. 5.6}$$

lorsque l'équation 5.4 est réécrite dans le nouveau système d'axes.

En tenant compte de l'équation 5.6, l'équation 5.4 devient:

$$f(u,v) = \frac{\exp \left[-\frac{1}{2} \left(\frac{u^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2} \right) \right]}{2 \pi \sigma_u \sigma_v} \quad \text{Eq. 5.7}$$

L'équation de l'ellipse est alors:

$$\frac{u^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2} = d^2 \quad \text{Eq. 5.8}$$

où:

$d^2 = 3.219$, par intégration de l'équation 5.7 pour une probabilité de 0.80

$$\sigma_u^2 = \frac{1}{2} (\sigma_x^2 + \sigma_y^2 + \sqrt{\delta})$$

$$\sigma_v^2 = \frac{1}{2} (\sigma_x^2 + \sigma_y^2 - \sqrt{\delta})$$

$$\delta = (\sigma_x^2 + \sigma_y^2)^2 - 4 \sigma_x^2 \sigma_y^2 (1 - \rho^2)$$

$$u = \sigma_u \sqrt{3.219}$$

$$v = \sigma_v \sqrt{3.219}$$

Le grand axe et le petit axe de l'ellipse sont u et v respectivement.

L'interprétation à donner à ces ellipses est la suivante. Si une station est située à l'intérieur de l'intervalle de confiance, l'hypothèse selon laquelle cette station fait partie du groupe ne peut être rejetée au niveau d'une probabilité de 80%. Alternativement, si une station est située en dehors de l'intervalle de confiance, son appartenance au groupe peut être rejetée avec la même probabilité. Les formes des ellipses représentant les intervalles de confiance peuvent être interprétées de la façon suivante: plus les caractéristiques climatologiques de stations sont similaires, plus les axes tendent à être petits et d'égale longueur. Une ellipse allongée implique que le phénomène ne peut être bien représenté par une seule station. Des ellipses plus ou moins superposées suggèrent que l'information provenant des groupes correspondants est plus ou moins redondante, selon l'importance de la superposition. Diverses formes et dimensions d'ellipses sont présentées dans l'exemple de la figure 5.2.

5.1.3.3 Application au réseau météorologique du Québec pour la définition de "régions météorologiques statistiquement homogènes".

On trouvera, dans les parties de ce rapport traitant des divers réseaux étudiés, des précisions concernant les données utilisées lors de l'analyse en composantes principales. L'objectif de cette section est plutôt d'expliquer comment cette analyse a été appliquée aux données disponibles et quels sont les principaux problèmes rencontrés.

L'interprétation à donner aux résultats est présentée à la section 5.1.3.4.

L'analyse des données en composantes principales a été réalisée à l'aide du programme STATIS (version 3.0) développé à l'Université du Québec¹. Tout comme le programme SPSS d'analyse en composantes principales disponible sur le système informatique de l'Université du Québec, le programme STATIS permet d'effectuer la rotation des axes et d'entrer une centaine de variables dans l'analyse, plus précisément 108. Chacun des réseaux étudiés comportent toutefois beaucoup plus que 108 stations. Il devient alors impossible de procéder à l'analyse simultanée de données provenant de toutes les stations. Deux procédures ont été imaginées et appliquées afin de solutionner ce problème, lorsque nécessaire.

Dans le cas des pluies journalières, les stations ont tout d'abord été divisées arbitrairement en deux groupes suprarégionaux, comportant chacun environ 108 stations, le premier associé aux régions météorologiques 701-702, 703, auxquelles s'ajoutait la région 707 lorsque le nombre total de stations ne dépassait pas 108, et le second groupe associé aux régions 704 à 711, incluant la région 707, même lorsque les stations situées dans cette région étaient retenues dans le premier groupe (figure 5.3). Un troisième regroupement de stations, en fonction de leurs coordonnées géographiques cette fois, visait à retenir les stations situées de part et d'autre de la frontière arbitraire entre les deux groupes suprarégionaux précédents, de manière à éliminer l'effet de la frontière sur le nombre et l'étendue des **régions statistiques** définies par l'analyse des données en composantes principales, avec rotation des axes.

¹ Le programme STATIS a été développé à l'INRS-Urbanisation par Normand Daigle.

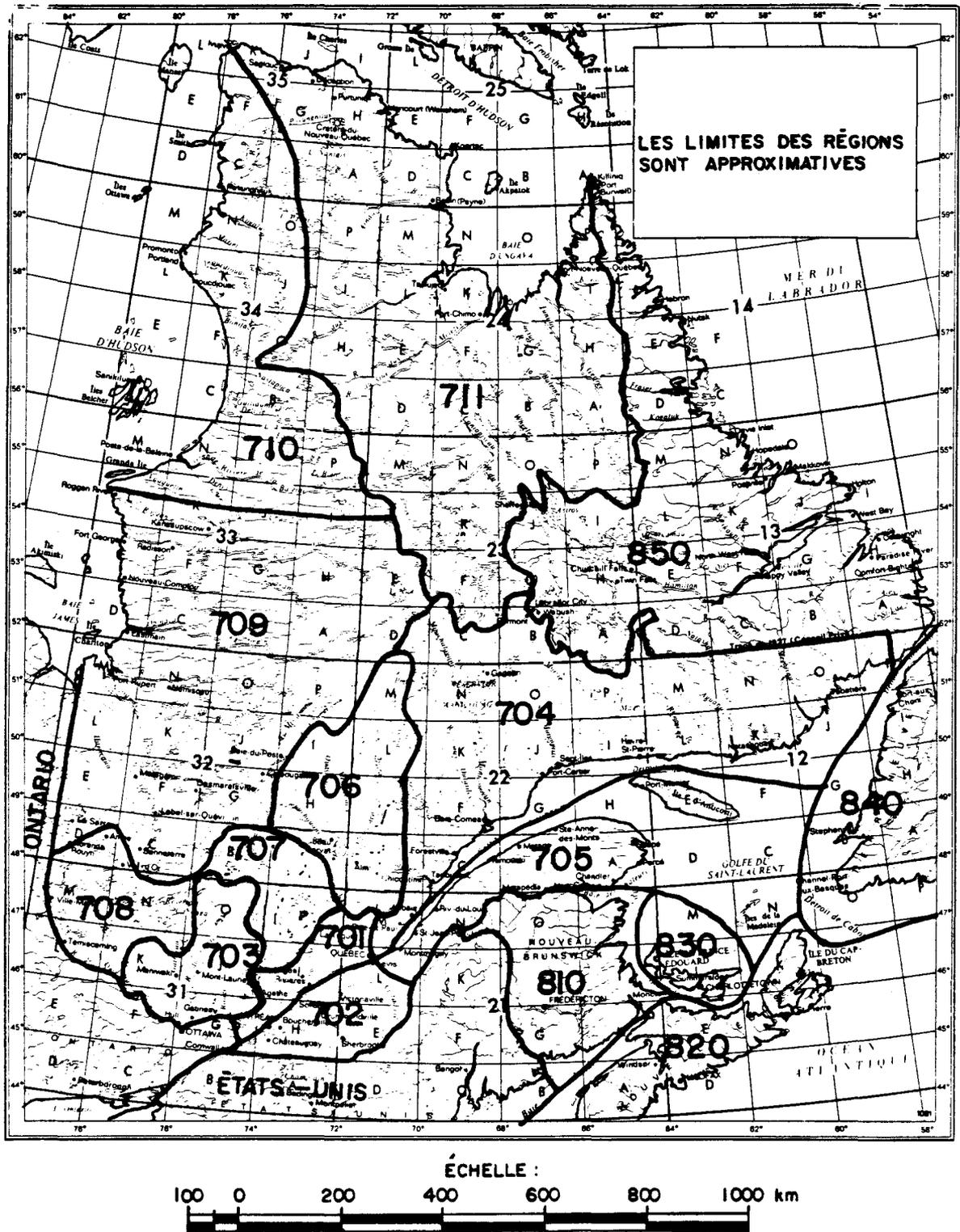


Figure 5.3. Identification numérique des régions météorologiques du Québec.

L'analyse des chutes de neige, des lignes de neige et des températures a été entreprise autrement. L'ensemble des stations a de nouveau été divisé en deux groupes, mais cette fois dans chaque groupe on a conservé des stations sur tout le territoire québécois. En pratique, les stations situées dans le nord du Québec ont été retenues pour les deux groupes, mais ont été associées à des stations différentes dans le sud, choisies de manière à garantir une bonne distribution spatiale. **Les limites des régions statistiquement homogènes** résultant des deux groupes ont alors été comparées et une nouvelle carte a été tracée en tenant compte des informations contenues sur les deux cartes initiales (figure 5.4).

5.1.3.4 Interprétation des résultats

Rappelons que les stations situées dans les régions ainsi définies sont regroupées selon la composante avec laquelle leur coefficient de corrélation est le plus élevé. Il est possible, d'autre part, pour chaque composante principale, de tracer la variation spatiale des coefficients de corrélation entre cette composante et les différentes stations (figure 5.5). On observe alors que la limite d'une région correspond en général à l'isoligne 0.6, c'est-à-dire que toutes les stations attribuées à une composante ont au moins un coefficient de corrélation de 0.6 avec cette composante. De même, les coefficients de corrélation entre cette composante, à laquelle est associée une région, et les stations environnantes situées dans une région contigüe ne tombent pas immédiatement à zéro, ce qui indique que l'influence de cette composante se fait sentir sur une certaine distance au-delà des limites des régions auxquelles elles sont reliées. On a donc défini une zone d'influence autour de chaque région en prenant comme limite de cette zone d'influence l'isoligne 0.3. Le choix de cette valeur est arbitraire mais exprime le fait que l'influence d'une région n'est plus tellement perceptible si les coefficients de corrélation tombent en bas de cette valeur. Il y a d'ailleurs bien souvent désorganisation du

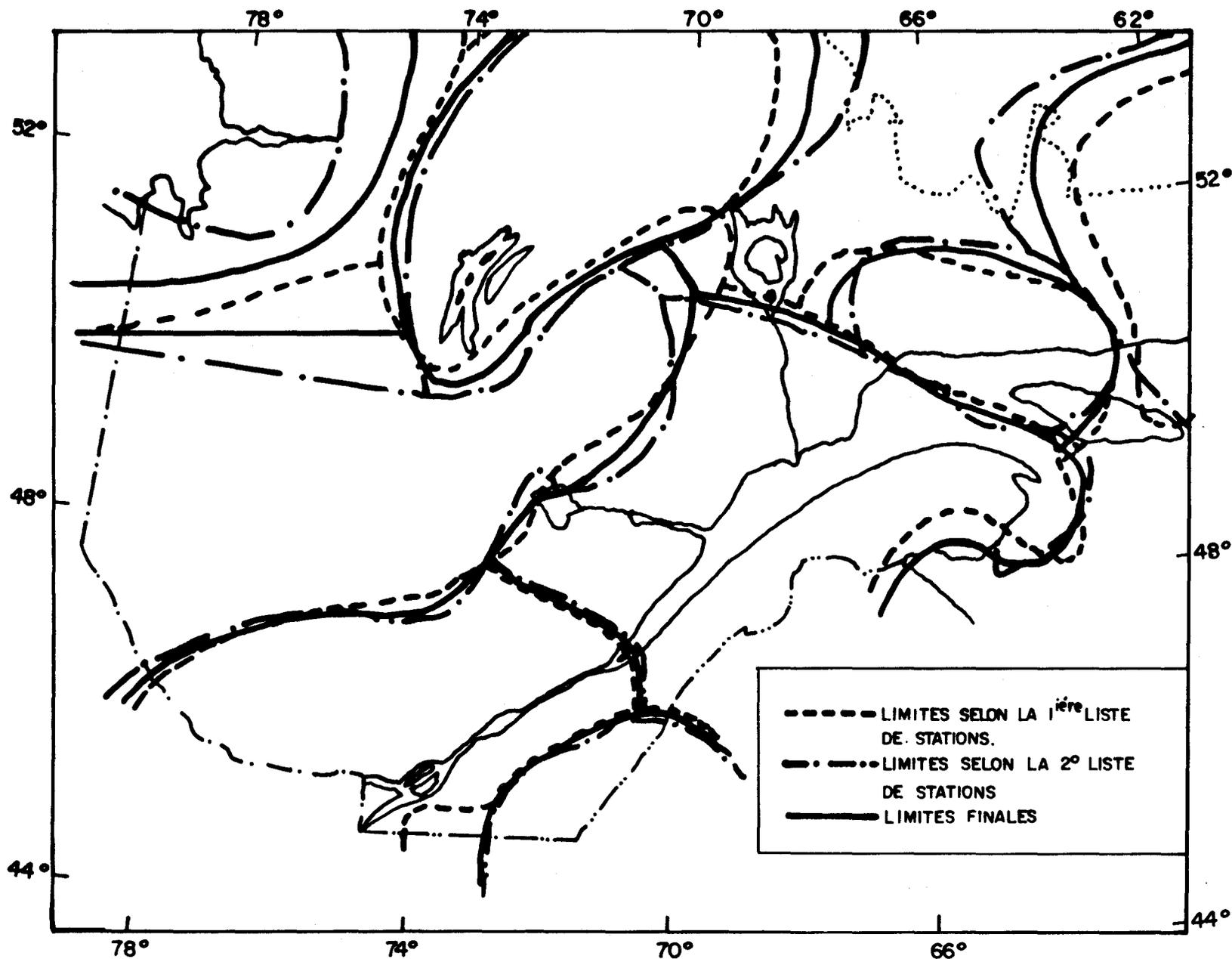


Figure 5.4. Limites initiales et finales des régions statistiquement homogènes pour les chutes de neige mensuelles.

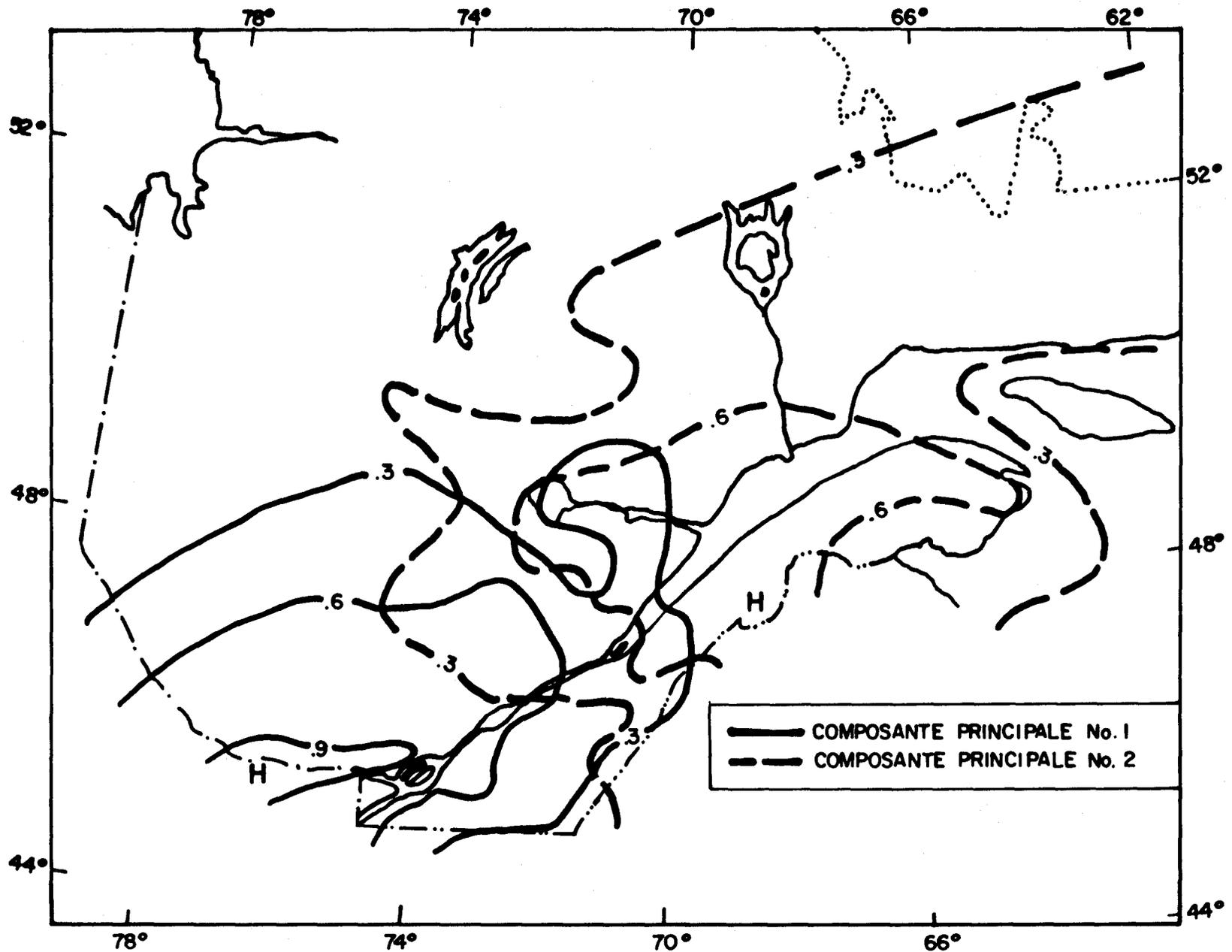


Figure 5.5. Variation spatiale des coefficients de corrélation entre les deux premières composantes principales et les stations retenues dans la première liste pour les chutes de neige mensuelles.

patron d'isolignes entre 0.3 et 0. En pratique, il y aura donc lieu de tenir compte du fait que l'on est ou non dans la zone d'influence d'une ou deux régions contigues, lorsque l'on s'interrogera sur les modifications à apporter à un réseau dans l'une des parties d'une région.

Dans le sud du Québec, où le nombre de stations est relativement élevé, les "limites" des régions sont relativement stables, lorsque l'on compare les cartes initiales. cependant, il faut comprendre que ces régions ne font qu'exprimer une influence prépondérante d'une composante principale. En d'autres mots, une station située près des limites d'une région est reliée de façon prépondérante à une composante principale, mais l'est aussi à un degré moindre avec les composantes principales caractérisant les régions environnantes.

Surtout dans le nord du Québec et sur la Côte-Nord, le nombre restreint de stations empêche de définir les régions de façon vraiment satisfaisante. On ne peut alors compter que sur quelques stations pour ce faire. Les limites deviennent alors plus imprécises et il est même possible d'ignorer des régions, faute de données. Ainsi, la région de Sherbrooke conserve toujours des dimensions relativement restreintes. Si cette région était située dans le nord du Québec, où la densité de stations est faible, il aurait été possible de ne pas la détecter, ou de considérer que la seule station qui indique sa présence est une station non homogène par rapport à une région plus grande. Il se glisse effectivement de telles stations dans diverses régions, par ailleurs assez bien définies. Comme ce sont les caractéristiques générales des régions qui sont recherchées, ces stations sont généralement oubliées par la suite, sauf en ce qui a trait à l'étude du réseau en place au premier janvier 1978.

5.1.4 Interpolation optimale

5.1.4.1 Objectif

Dans le cadre de l'étude du réseau météorologique du Québec en vue de sa rationalisation, l'objectif visé en appliquant la méthode de l'interpolation optimale pour l'analyse des données disponibles pour chaque "région météorologique statistiquement homogène" est d'obtenir les informations nécessaires à la détermination de la configuration optimale d'un réseau apte à répondre à des besoins en données météorologiques exprimés sous forme de précision souhaitée. Cette analyse des données par interpolation optimale a donc pour objectifs spécifiques de déterminer:

- l'erreur d'interpolation en fonction de la distance entre les stations dans le cas d'un réseau régulier;
- l'erreur réelle d'interpolation en tout point d'une région effectivement obtenue à partir du réseau en place à une date donnée.

5.1.4.2 Structure des champs météorologiques

Compte tenu des phénomènes d'échelle planétaire ou hémisphérique qui sont à l'origine des variations météorologiques à la surface de la terre, les variables météorologiques (pluie, neige, température,...) présentent une certaine structure spatiale et temporelle. En d'autres mots, ce qui se passe en un endroit donné en un temps donné n'est pas indépendant de ce qui se passe en d'autres points plus ou moins rapprochés au même moment. D'autre part, à cette structure supra-régionale ou régionale, se superposent, particulièrement dans le cas des champs de surface, des variations d'échelle beaucoup plus courtes, induites par l'inhomogénéité de la surface du sol (relief, végétation,...).

Si, selon Gandin (1965), i et j sont deux points d'un champ météorologique et f_i et f_j , les valeurs prises par une variable météorologique en ces points, il est possible de définir une fonction de structure B_{ij} , telle que:

$$B_{ij} = \frac{1}{2} [f_i - f_j]^2 \quad \text{Eq. 5.9}$$

Notons immédiatement que cette définition de la fonction de structure est équivalente à celle de Gandin (1965), la seule différence étant le facteur $1/2$. Cette définition évite l'apparition de facteurs $1/2$ dans les équations de la section 5.1.4.3 et correspond de plus à celle du variogramme à l'équation 5.13.

La fonction de structure B_{ij} dépend en général de la position des points i et j de la variable f . Plutôt que les valeurs prises par une variable, on peut considérer les anomalies par rapport à la normale.

$$f'_i = f_i - \bar{f}_i \quad \text{Eq. 5.10}$$

La fonction de structure de ces anomalies peut alors s'écrire:

$$b_{ij} = \frac{1}{2} [f'_i - f'_j]^2 \quad \text{Eq. 5.11}$$

En remplaçant f_i et f_j dans l'équation 5.9 par leur équivalent tiré de l'équation 5.10, on observe que:

$$B_{ij} = b_{ij} + [\bar{f}_i - \bar{f}_j]^2 \quad \text{Eq. 5.12}$$

Tant que les fonctions de structure B_{ij} et b_{ij} dépendent des points de mesure i et j , la structure statistique spatiale de la variable est pratiquement impossible à décrire analytiquement. Il est donc intéressant de rechercher quelles hypothèses faites sur le champ de la variable pourraient se réaliser en pratique et permettre une formulation plus simple des fonctions de structure. En particulier, la formulation de ces fonctions serait considérablement simplifiée si ces dernières ne dépendaient que de la distance d entre les points i et j . Cette condition se réalise si le champ est homogène et isotrope.

Un champ est considéré homogène si la valeur prise par une fonction dépend uniquement de la distance entre les points d'observation et ne change pas lorsque ces points sont déplacés dans le champ, en gardant leur orientation inchangée. D'autre part, un champ est isotrope si la valeur prise par une fonction ne change pas lorsque l'orientation du segment liant les deux points varie, le centre du segment ne bougeant pas. Gandin (1965) précise que les fonctions de structure des valeurs observées de variables météorologiques ne sont en général ni homogènes ni isotropes, par suite de l'inhomogénéité des normales de ces variables. Il s'en suit, toujours d'après Gandin (1965), que les anomalies des variables par rapport à la normale sont étudiées plutôt que les observations originales. Il précise que les données disponibles indiquent que les fonctions de structure des anomalies peuvent, pour la majorité des variables météorologiques, être considérées comme homogènes et isotropes sur des distances relativement grandes de sorte que dans l'équation 5.11, on pourrait écrire $b(d)$ plutôt que b_{ij} , d étant la distance entre les points i et j .

L'assertion de Gandin (1965) sur l'homogénéité des anomalies est

par ailleurs confirmée par Delfiner (1973). Ce dernier écrit, en effet, que deux hypothèses peuvent être faites sur le champ des anomalies, une première impliquant la stationnarité locale d'ordre deux des anomalies, la seconde étant une hypothèse intrinsèque. La première de ces hypothèses suppose l'existence de moments d'ordre 1 et 2 d'une fonction aléatoire. Elle implique que la moyenne et la variance des observations sont constantes, donc homogènes, dans le champ (en pratique, dans un voisinage de travail convenable). La réalisation de cette hypothèse permet d'utiliser des fonctions de covariance et de variance pour exprimer la structure statistique d'une variable. Dans ce cas, le champ des anomalies est stationnaire d'ordre 2. L'hypothèse intrinsèque est moins restrictive et implique que seuls les accroissements $(f_i - f_j)$ sont stationnaires d'ordre 2. Cette condition s'écrit, selon Delfiner (1973), si l'on pose $j = x$ et $i = x + d$:

$$E [f(x + d) - f(x)] = m(d)$$

Eq. 5.13

$$D^2 [f(x + d) - f(x)] = 2\gamma(d)$$

Elle signifie que la fonction aléatoire f_i est intrinsèque si pour tout vecteur d la moyenne et la variance des accroissements $[f(x + d) - f(x)]$ ne dépendent pas du point d'appui x , ou, en d'autres termes, si ces accroissements sont stationnaires d'ordre 2. La variance des accroissements est ce qu'on appelle le variogramme $\gamma(d)$, dans la théorie du krigeage. Delfiner (1973) ajoute alors qu'on suppose, en général, que les anomalies sont intrinsèques.

L'utilisation de la fonction de structure b_{ij} basée sur les anomalies permet donc de satisfaire les diverses hypothèses formulées plus haut. De plus, l'équation 5.12 indique que si le champ des valeurs observées est lui aussi homogène, c'est-à-dire que si les normales aux différentes stations sont identiques, les fonctions de structure

B_{ij} et b_{ij} sont elles-mêmes identiques.

Quant à l'hypothèse d'isotropie, sa non-vérification dans un champ implique la nécessité de tenir compte de l'orientation des vecteurs joignant deux points de mesure. L'estimation des fonctions de structure doit donc se faire séparément pour diverses classes d'orientations, tel que précisé en 5.1.2.

Dans le cas d'une variable météorologique et à l'intérieur des distances considérées, les fonctions de structure ont l'allure générale de la courbe présentée à la figure 5.6. Une remarque découle immédiatement de l'observation de cette figure. La courbe ne passe pas par l'origine. C'est ce qu'on appelle **l'effet de pépité C** en krigeage (Matheron, 1965; Delfiner, 1973; Delhomme, 1978). Delfiner (1973) écrit que cet effet de pépité représente la variance d'un phénomène qui se développe à une échelle très inférieure à celle des distances entre les observations. Cet effet signifie que les observations réalisées en deux points relativement près l'un de l'autre montreront des différences ayant une variance au moins égale à l'effet de pépité (Delhomme, 1978). On peut donc écrire, toujours d'après Delfiner (1973):

$$b_{ij} = 0 \quad \text{pour } i = j \quad \text{i.e. } d = 0$$

$$\tilde{b}_{ij} = b_{ij} + C \quad \text{pour } i \neq j \quad \text{i.e. } d \neq 0$$

En météorologie, on attribue cet effet à des irrégularités microclimatiques (Gandin, 1970). Ainsi, dans le cas de la pluie, cet effet prend sa source en particulier dans des perturbations locales, l'instabilité de l'air près du sol et les bourrasques irrégulières qui accompagnent les chutes de pluie au-dessus des pluviomètres.

Les irrégularités microclimatiques ne sont pas les seules causes de l'effet de pépité. A celles-ci s'ajoutent les erreurs de mesure.

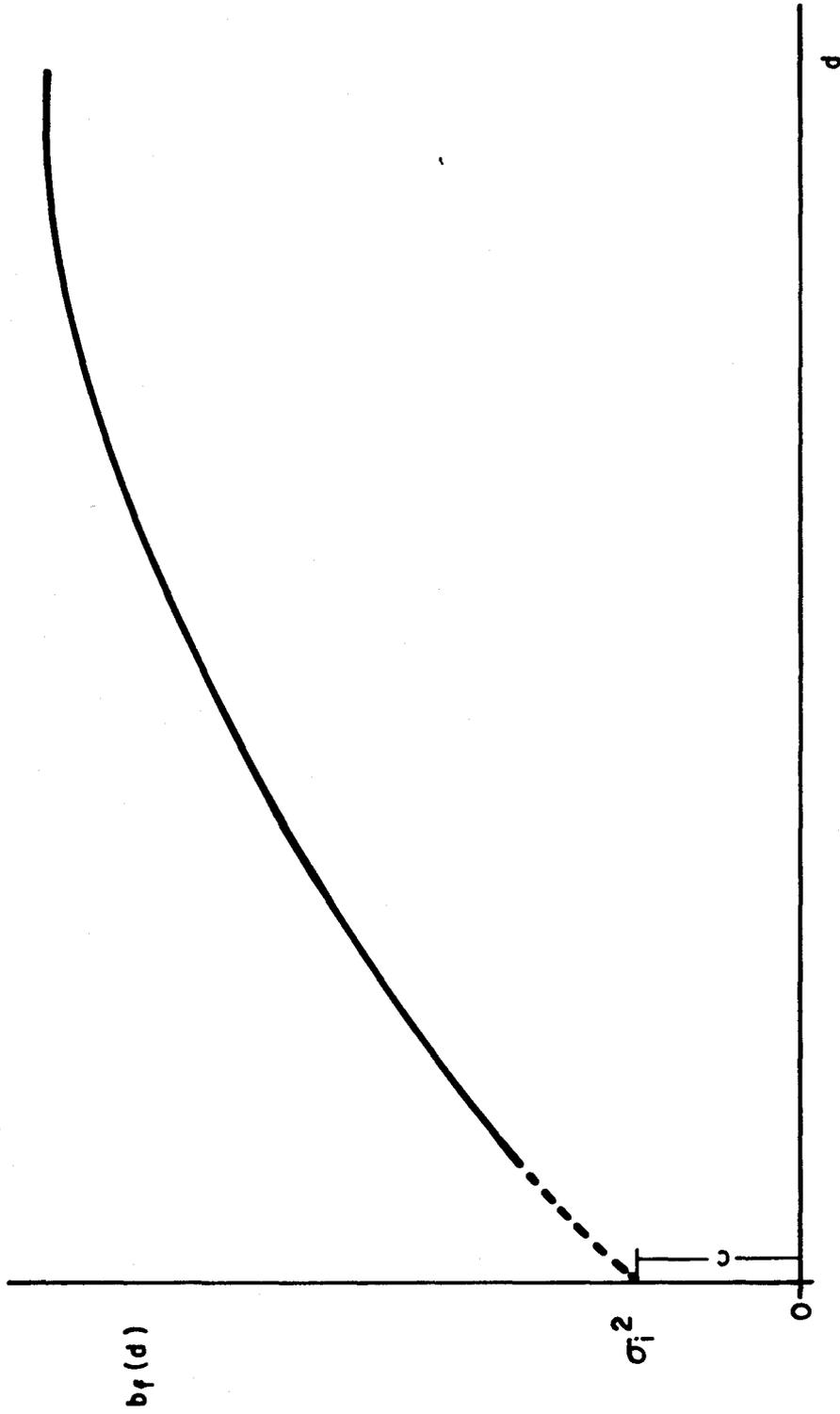


Figure 5.6 . Allure générale d'une fonction de structure .

Ces deux causes sont difficilement séparables et seront indissociées dans ce qui suit. Gandin (1965) et Delfiner (1973) supposent que ces sources d'erreurs sont aléatoires, c'est-à-dire que:

- leur moyenne est nulle: $\overline{\delta_i} = 0$ pour tout i ;
- elles ne sont pas corrélées avec le champ:

$$\overline{\delta_i \cdot f_j} = 0 \text{ pour tout } i \text{ et } j$$

- elles ne sont corrélées entre elles:

$$\overline{\delta_i \cdot \delta_j} = 0 \text{ pour tout } i \neq j$$

De plus, l'écart quadratique moyen de données de la variable peut s'écrire:

$$\overline{\delta_i^2} = \sigma^2 . \text{ Il représente la variance des erreurs de mesure en un point.}$$

Gandin (1965) ajoute que le cas le plus intéressant, en pratique, est celui d'un champ d'erreur homogène, c'est-à-dire un champ où l'erreur ne dépend pas du point i . Il démontre, par ailleurs, que des erreurs purement aléatoires conduisent à des valeurs exagérées de la fonction de structure d'une quantité égale ou double de l'écart quadratique moyen. Soit \tilde{f}_i l'anomalie observée pour une variable météorologique et f_i l'anomalie vraie de cette même variable.

On peut alors écrire:

$$\tilde{f}_i = f_i + \delta_i$$

Eq. 5.14

La fonction de structure expérimentale \tilde{b}_{ij} est alors reliée à la fonction de structure théorique b_{ij} :

$$\tilde{b}_{ij} = \overline{[\tilde{f}'_i - \tilde{f}'_j]^2} = \overline{[f'_i - f'_j]^2} + \overline{[\delta_i - \delta_j]^2} \quad \text{Eq. 5.15}$$

D'où, pour un champ homogène et isotrope (Gandin, 1965; Hutchinson et Cislérova, 1974):

$$\tilde{b}_{ij} = b_{ij} + \sigma^2 \quad \text{Eq. 5.16}$$

Sachant que $b_{ij} = 0$, pour $i=j$ (i.e. $d=0$), on peut déduire de l'équation 5.16 que $\tilde{b}_{d=0} = \sigma^2$. Retournant à la figure 5.6, on voit que cette valeur peut effectivement être estimée comme étant l'ordonnée à l'origine lorsque l'on extrapole la courbe \tilde{b}_{ij} jusqu'à $d=0$.

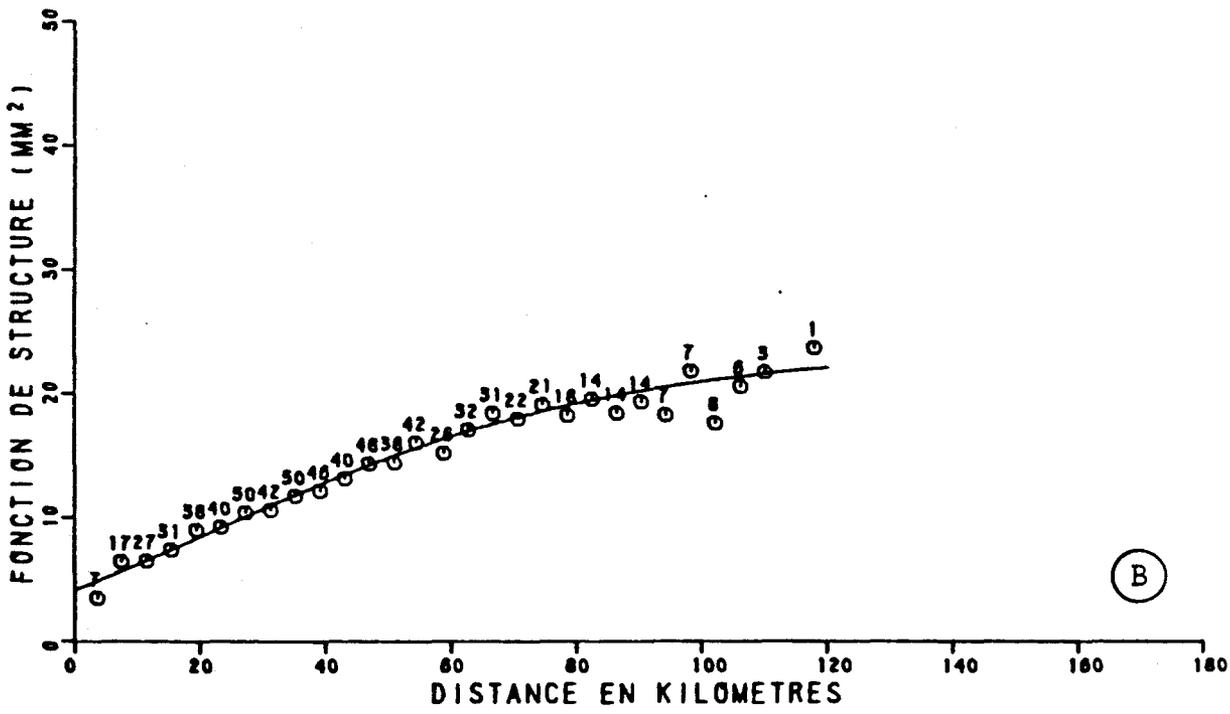
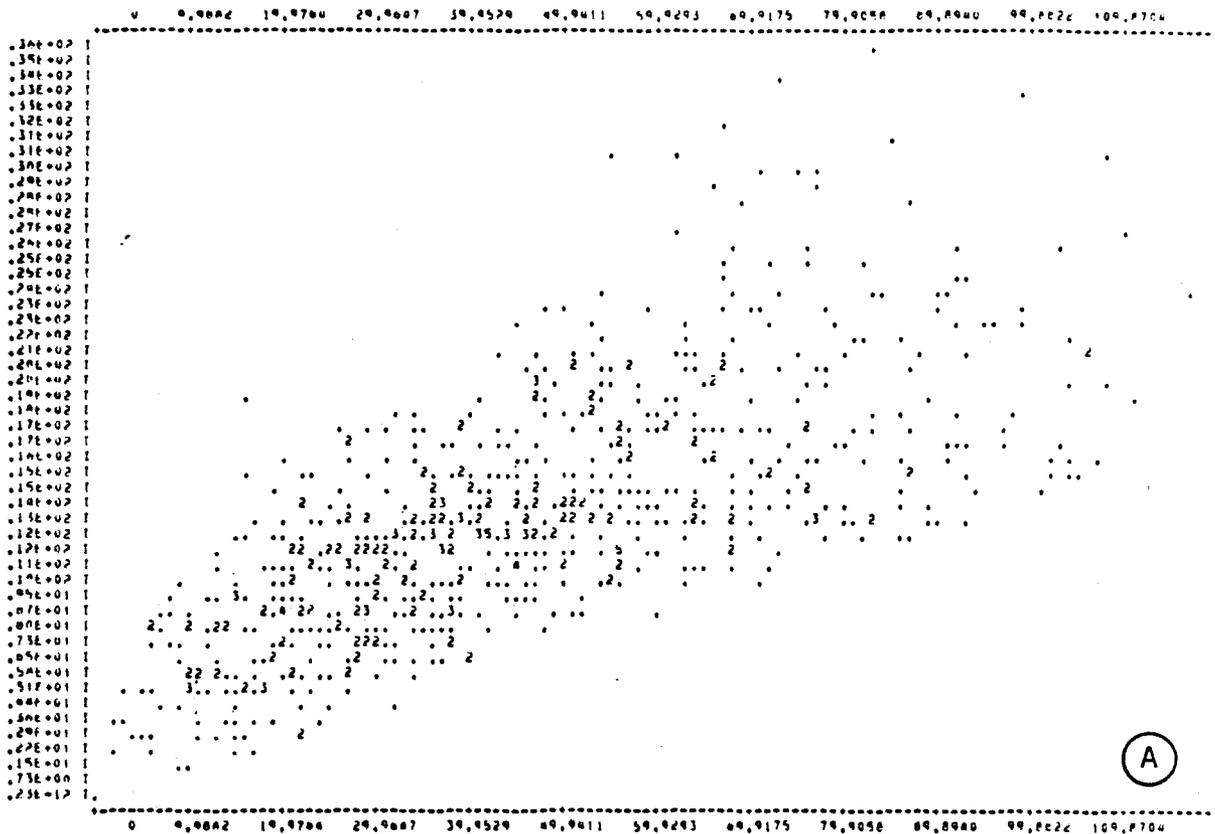
Gandin (1965) précise que pour obtenir la fonction de structure théorique b_{ij} , il est nécessaire de soustraire la quantité σ^2 de toutes les valeurs de \tilde{b}_{ij} , c'est-à-dire de tracer une courbe parallèle à la courbe initiale, mais passant par l'origine. Il note cependant, qu'en pratique, la détermination de σ^2 peut s'avérer difficile, compte tenu du manque de points définissant la courbe près de l'origine.

Enfin, la violation des hypothèses d'homogénéité et d'isotropie conduit normalement à des fonctions de structure expérimentale \tilde{b}_{ij} plus élevées. L'importance de cette erreur est toutefois difficile à

évaluer (Gandin, 1965). Nous avons effectivement observé cet effet en tentant d'obtenir des fonctions de structure communes pour des régions préalablement déterminées par l'analyse des données par les composantes principales, et dont les fonctions de structure régionales étaient à peu près identiques.

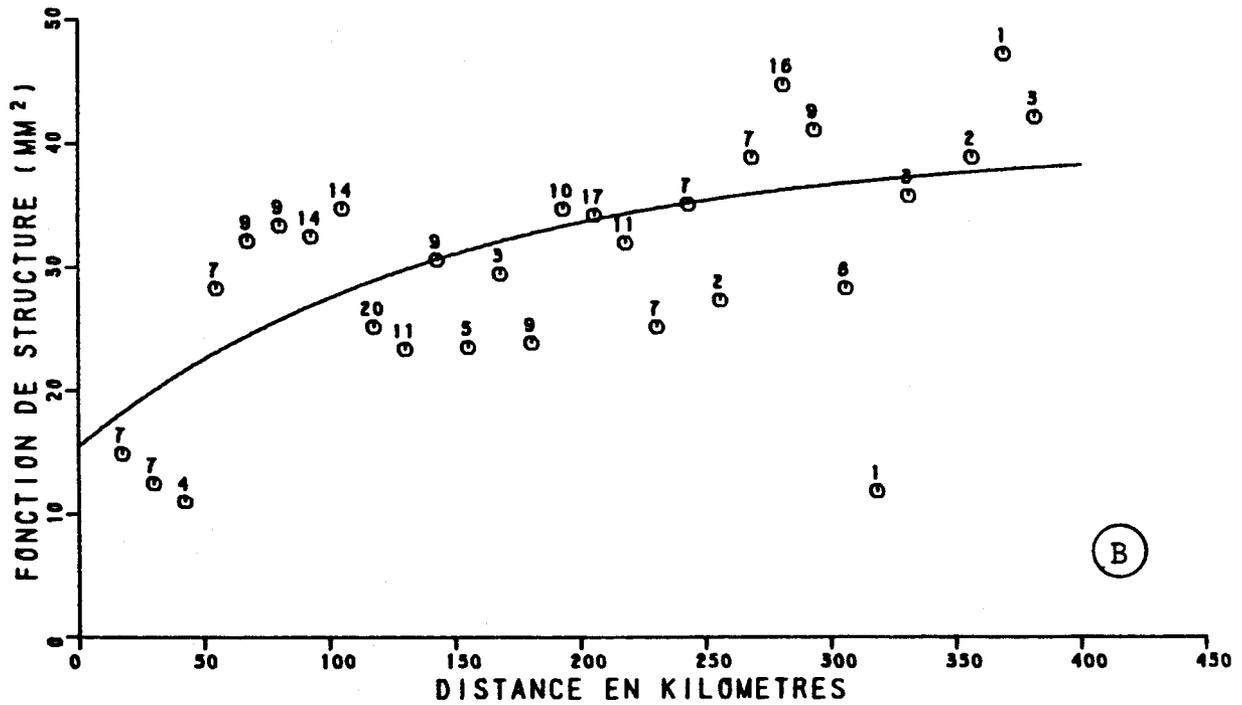
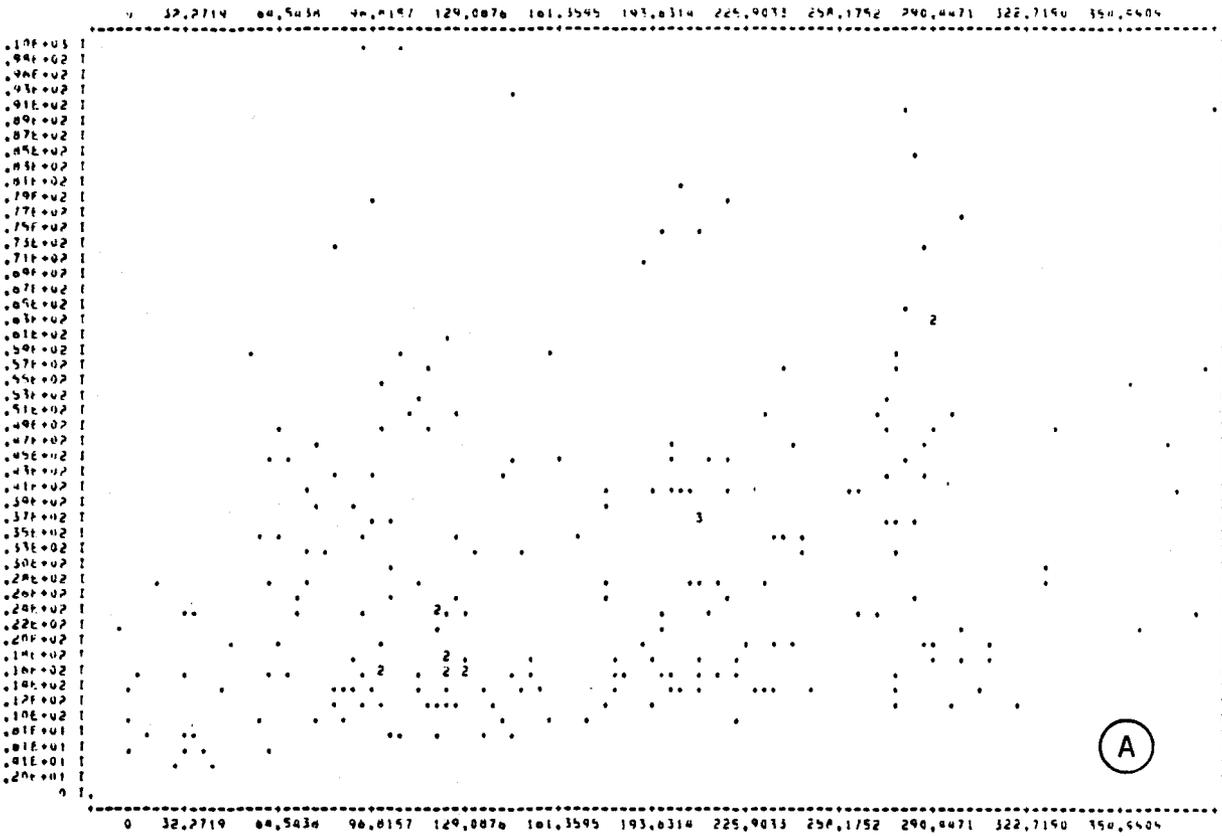
Le calcul des points expérimentaux d'une fonction de structure doit normalement, d'après Chauvet **et al.** (1976) et Gandin (1965), être réalisé à partir d'une longue série de réalisations de la variable météorologique, considérée comme fonction aléatoire. Cependant, Gandin (1970) considère qu'il est suffisant, en général, de disposer de soixante (60) événements à soixante (60) stations. Il précise, par ailleurs, d'autres conditions à respecter. Ainsi, les stations doivent être situées dans la région étudiée et de préférence distribuées irrégulièrement, de façon à ce qu'un nombre approximativement égal de paires de stations (donc de points sur la courbe \tilde{f}_{ij}) soit obtenu aux petites, moyennes et grandes distances. Dans le cas présent, les stations retenues sont celles qui sont situées à l'intérieur de chacune des régions (ou plutôt des zones d'influence de ces régions) définies à l'aide des composantes principales. Un intervalle de temps suffisamment grand doit aussi être choisi (habituellement deux à trois jours) pour que les événements soient suffisamment indépendants. Enfin, surtout dans le cas où une composante cyclique annuelle existe, l'échantillon doit être choisi à l'intérieur d'une période de temps relativement courte: une saison ou même un mois.

Il serait également possible de supposer que la structure du champ est homogène dans tout l'espace et de faire des moyennes spatiales à la place de moyennes temporelles, en utilisant une seule réalisation de la fonction aléatoire, sous certaines conditions d'ergodicité (Matheron, 1965). Toutefois, outre que Delhomme (1978) et Gandin (1965) considèrent que les conditions d'ergodicité sont difficiles à vérifier avec une seule réalisation d'une fonction aléatoire, Chauvet **et al.** (1976) expliquent que procéder de cette façon permet de



FONCTION DE STRUCTURE DES PLUIES JOURNALIERES AU PRINTEMPS, REGION I, MONTREAL

Figure 5.7. Points initiaux, points moyens par classe de distance et courbe ajustée.



FONCTION DE STRUCTURE DES PLUIES JOURNALIERES AU PRINTEMPS, REGION : GASPE - BAIE-COMEAU

Figure 5.8. Points initiaux, points moyens par classe de distance et courbe ajustée.

faire ressortir les caractéristiques de la situation étudiée et non les caractéristiques saisonnières du champ. Comme c'est aux caractéristiques saisonnières du champ que nous sommes intéressés, il est donc nécessaire de faire appel à des séries chronologiques plutôt qu'à des situations uniques particulières.

La mise en graphique des valeurs \tilde{b}_{ij} calculées pour chaque couple de stations à partir des séries chronologiques observées à ces stations présente l'aspect d'un nuage plus ou moins serré, comme on peut le constater aux figures 5.7-a et 5.8-a. Cette mise en graphique peut permettre notamment de détecter une station non homogène dont les \tilde{b}_{ij} sont très différents de ceux des autres stations et par conséquent sont isolés par rapport aux autres points. Il y a alors avantage à recalculer les \tilde{b}_{ij} en enlevant cette station. De manière à faciliter l'emploi de la fonction b_{ij} dans le cadre de l'interpolation optimale, il est nécessaire d'ajuster un modèle analytique à ces points. Pour ce faire, on détermine d'abord des classes de distances à partir desquelles on calcule la distance moyenne des points qui s'y trouvent, de même que la valeur moyenne de \tilde{b}_{ij} . Il en résulte un certain nombre de points, autant qu'il y a de classes (figure 5.7-b et 5.8-b). Une fois ces points calculés, divers modèles analytiques peuvent leur être ajustés, en tenant compte du nombre de points originaux que représentent ces points. L'un de ces modèles est alors retenu sur la base de la précision et du réalisme de l'ajustement. Les modèles analytiques utilisés sont:

$$a) \quad \tilde{b}_{ij} = ab^{cd} \quad (\text{courbe de Gumpertz}) \quad \text{Eq. 5.17}$$

$$b) \quad \tilde{b}_{ij} = ad^b + c \quad (\text{courbe parabolique}) \quad \text{Eq. 5.18}$$

$$c) \quad \tilde{b}_{ij} = a - be^{-d/c} \quad (\text{courbe exponentielle}) \quad \text{Eq. 5.19}$$

$$d) \tilde{b}_{ij} = ad + b \quad (\text{droite}) \quad \text{Eq. 5.20}$$

où a , b et c sont des paramètres d'ajustement et d , la distance entre les points i et j . Il est à remarquer que toutes ces courbes ont une ordonnée à l'origine non nulle et positive permettant de tenir compte de l'effet de pēpīte. Leur utilisation permet d'estimer la valeur la plus probable de $\tilde{b}_f(d)$ pour n'importe quelle distance d , mais cette valeur doit être interprétée en tenant compte de la dispersion des points originaux.

5.1.4.3 Théorie générale

Le but de toute interpolation est d'estimer la valeur f_0 d'une variable en un point quelconque du champ considéré. Cependant, tel qu'indiqué en 5.1.4.2, il peut être préférable de rechercher l'anomalie f'_0 (ou écart par rapport à la normale) plutôt que la valeur f_0 . Cette anomalie f'_0 peut effectivement être estimée grâce à une combinaison linéaire des anomalies f'_i aux stations. Toutefois, comme l'indique l'équation 5.14, les anomalies observées \tilde{f}'_i ne sont pas égales aux anomalies vraies f'_i , ces dernières étant entachées d'erreurs d'observation δ_i . Donc, l'erreur quadratique moyenne E entre l'anomalie vraie f'_0 au point d'interpolation et l'estimation linéaire peut s'écrire:

$$E = \overline{\left[f'_0 - \sum_{i=1}^n p_i (f'_i + \delta_i) \right]^2} \quad \text{Eq. 5.21}$$

Notons immédiatement qu'en première approximation, il aurait été possible de développer la théorie de l'interpolation optimale en supposant que les données disponibles sont exactes. Donc, ce cas E serait

égal à:

$$E = \left[f'_0 - \sum_{i=1}^n p_i f'_i \right]^2 \quad \text{Eq. 5.22}$$

Gandin (1965) présente effectivement un tel développement, mais précise, par ailleurs, que la décision de ne pas tenir compte de la présence d'erreurs de mesure dans les données observées lors du calcul des poids, conduit à une erreur d'interpolation certainement plus grande que les erreurs d'observation. Il ajoute, de plus, que si, au contraire, ces erreurs d'observation sont prises en compte et si ces dernières ne sont pas trop faibles, il est possible d'atteindre même une situation dans laquelle l'erreur d'interpolation va être, en moyenne, considérablement plus faible que les erreurs d'observation. D'autres auteurs insistent aussi sur l'avantage, sinon la nécessité de tenir compte des erreurs de mesure (Delfiner, 1973; Chauvet *et al.*, 1976). La théorie développée dans cette section suppose donc la présence d'erreurs d'observations dans les données.

Rappelons, de plus, que si la structure statistique du champ est introduite dans les calculs et que l'erreur quadratique moyenne E , selon l'équation 5.21, est minimisée par rapport aux poids p_i , l'interpolation réalisée à l'aide de ces poids est dite interpolation optimale (Gandin, 1965 et 1970).

Le développement de l'équation 5.21 donne:

$$E = \overline{(f'_0)^2} + \sum_{i=1}^n \sum_{j=1}^n p_i p_j \overline{(f'_i + \delta_i)(f'_j + \delta_j)} - 2 \sum_{i=1}^n p_i \overline{f'_0 (f'_i + \delta_i)}$$

Eq. 5.23

Si maintenant:

$$m_{00} = \overline{(f'_0)^2} \quad : \text{ la variance de l'anomalie au point 0;}$$

$$m_{ij} = \overline{f'_i \cdot f'_j} \quad : \text{ la covariance entre les anomalies aux points } i \text{ et } j;$$

$$\sigma_i^2 = \overline{(\delta_i)^2} \quad : \text{ la variance des erreurs d'observation,}$$

alors, l'équation 5.23 peut être réécrite de la façon suivante:

$$E = m_{00} + \sum_{i=1}^n \sum_{j=1}^n p_i p_j m_{ij} + \sum_{i=1}^n p_i^2 \sigma_i^2 - 2 \sum_{i=1}^n p_i m_{0i}$$

Eq. 5.24

Tel que précisé en 5.1.4.2, l'expression de E en fonction de m_{00} et m_{ij} , suppose l'existence d'une stationnarité d'ordre deux, c'est-à-dire des moments d'ordre 1 et 2 d'une fonction aléatoire. Nous allons plutôt utiliser ici une fonction de structure qui suppose, selon

l'hypothèse intrinsèque, que seuls les accroissements de la variable sont stationnaires d'ordre deux.

Si la fonction de structure b_{ij} est définie par l'équation 5.11 et le membre de droite de cette équation est développé, alors:

$$b_{ij} = \frac{m_{ii} + m_{jj}}{2} - m_{ij} \quad \text{Eq. 5.25}$$

L'erreur quadratique moyenne E peut maintenant être exprimée en fonction de b_{ij} , en introduisant b_{ij} , tel que défini à l'équation 5.25, dans l'équation 5.24 et en posant l'hypothèse que $m_{ii} = m_{00}$ pour tout i :

$$E = m_{00} \left(1 - \sum_{i=1}^n p_i\right)^2 - \sum_{i=1}^n \sum_{j=1}^n p_i p_j b_{ij} + \sum_{i=1}^n p_i \sigma_i^2 + 2 \sum_{i=1}^n p_i b_{0i} \quad \text{Eq. 5.26}$$

Pour que l'interpolation soit optimale, il est maintenant nécessaire de minimiser E par rapport aux poids p_i . Cette condition est effectivement réalisée si $\frac{\partial E}{\partial p_i} = 0$ pour tout i . Si de plus:

$$\sum_{i=1}^n p_i = 1, \text{ ce qui requiert l'introduction du multiplicateur } \lambda \text{ Lagrange;}$$

$$\text{et } \sigma_i^2 = \sigma^2, \text{ pour tout } i,$$

il n'est alors pas nécessaire de connaître la variance m_{00} et le système de $n + 1$ équations permettant de déterminer les poids est:

$$\sum_{j=1}^n p_j b_{ij} - p_i \sigma^2 + \lambda = b_{0i} \quad \text{pour tout } i=1, \dots, n$$

Eq. 5.27

$$\sum_{j=1}^n p_j = 1$$

L'erreur quadratique moyenne E telle qu'exprimée dans l'équation 5.26 peut enfin être simplifiée à l'aide du système d'équation 5.27, de telle sorte que:

$$E = \sum_{i=1}^n p_i b_{0i} + \lambda \quad \text{Eq. 5.28}$$

L'équation 5.28 n'est valable que dans le cas où l'interpolation est optimale.

Il peut être enfin intéressant de comparer la valeur effectivement observée à une station à la valeur estimée par interpolation optimale des valeurs observées à d'autres stations. Dans ce cas, les deux valeurs comparées sont entachées d'erreurs et l'erreur quadratique moyenne E_{obs} n'est pas définie par l'équation 5.21, mais par:

$$E_{\text{obs}} = \left[(f'_0 + \delta_0) - \sum_{i=1}^n p_i (f'_i + \delta_i) \right]^2 \quad \text{Eq. 5.29}$$

Si un développement similaire à celui de l'équation 5.21 est réalisé, le système d'équation 5.27 permettant de déterminer les poids, ne change pas, mais l'erreur quadratique moyenne E_{obs} s'écrit:

$$E_{\text{obs}} = \sum_{i=1}^n p_i b_{oi} + \lambda + \sigma^2 \quad \text{Eq. 5.30}$$

Donc, l'erreur quadratique moyenne E_{obs} d'estimation d'une valeur observée est égale à l'erreur quadratique moyenne d'estimation de la valeur vraie augmentée de la variance σ^2 des erreurs d'observation:

$$E_{\text{obs}} = E + \sigma^2 \quad \text{Eq. 5.31}$$

L'écart type S_{obs} d'estimation des valeurs observées est donc:

$$S_{\text{obs}} = (E_{\text{obs}})^{\frac{1}{2}} = (E + \sigma^2)^{\frac{1}{2}} \quad \text{Eq. 5.32}$$

tandis que l'écart type S d'estimation des valeurs vraies est par définition:

$$S = E^{\frac{1}{2}} \quad \text{Eq. 5.33}$$

La différence entre S_{obs} et S est, par conséquent, d'autant plus importante que la variance σ^2 est grande. En d'autres termes, si on désire déterminer l'écart type moyen entre une valeur observée et une valeur estimée par interpolation, on obtient une valeur d'écart type plus élevée que si la comparaison s'effectue entre une valeur vraie et une valeur estimée. Bien sûr, la valeur vraie n'est pas connue, mais il n'est pas nécessaire de la connaître selon la théorie, pour estimer l'écart type moyen entre cette valeur vraie et la valeur estimée (Gandin, 1965 et 1970; O'Connell et al., 1978).

5.1.4.4 Résolution pratique du système d'équations

Il est important de noter, au départ, que le système d'équations doit être solutionné autant de fois qu'il y a de points où une interpolation est désirée, c'est-à-dire de points de grille.

Pour ce faire, on peut considérer que toutes les stations situées dans la région étudiée (ou plus précisément à l'intérieur de la zone d'influence de cette région), forment un voisinage unique, quel que soit le point où l'on interpole (Delfiner, 1973; Delhomme, 1978). Cette façon de procéder a l'avantage d'impliquer une seule inversion de la matrice des b_{ij} et par conséquent de diminuer le temps de calcul. Son emploi est toutefois conditionné par la dimension de la matrice des b_{ij} , la matrice inversée b_{ij}^{-1} étant d'autant moins précise que le nombre de stations est grand.

Compte tenu de l'imprécision résultant souvent d'un voisinage unique il est préférable de considérer un voisinage glissant autour du point de grille où se fait l'interpolation. Ce voisinage peut être défini par un rayon d'influence autour du point ou par le nombre de

stations retenu. Ces deux critères ont de l'importance, en pratique. D'une part, il est inutile de retenir des stations situées à une distance du point considéré telle que la corrélation est nulle (voir section 5.1.4.2). D'autre part, pour un rayon donné, il peut arriver que le nombre de stations contenues à l'intérieur du voisinage soit très faible et qu'il y ait lieu d'augmenter ce rayon. Il est donc préférable de voir à ce qu'au moins 10 à 20 stations soient retenues au départ.

La disposition des stations autour du point de grille a aussi son importance. Il vaut mieux avoir moins de stations entourant le point de grille de tous les côtés qu'un plus grand nombre situées toutes du même côté.

5.1.4.5 Ecart type d'interpolation en fonction de la distance entre les stations

La géométrie d'un réseau réel est le plus souvent tributaire de la physiographie, des axes routiers et des centres de population. Il est toutefois important, comme point de comparaison, de déterminer qu'elle serait l'écart type d'interpolation obtenue à partir d'un réseau de géométrie régulière. Gandin (1965 et 1970) indique qu'un réseau à mailles triangulaires est plus susceptible d'être réalisé de façon approximative qu'un réseau à mailles carrées. Le même auteur suggère donc de calculer l'écart type d'interpolation au centre d'un triangle équilatéral aux sommets duquel les stations sont situées. Toutefois, les considérations exprimées en 5.1.4.4 sur le nombre de stations souhaitables dans un voisinage et leur disposition autour du point de grille, de même que les essais de calculs que nous avons réalisés, indiquent qu'il est préférable de prendre plus de trois stations. Nous avons donc opté pour un réseau de six stations disposées selon deux triangles équilatéraux emboîtés centrés sur le point de grille (figure 5.9). La distance entre les stations du triangle intérieur est 1 et celle entre les stations du triangle extérieur est 21. Six stations situées tout autour du point de grille sont donc retenues. Nous avons

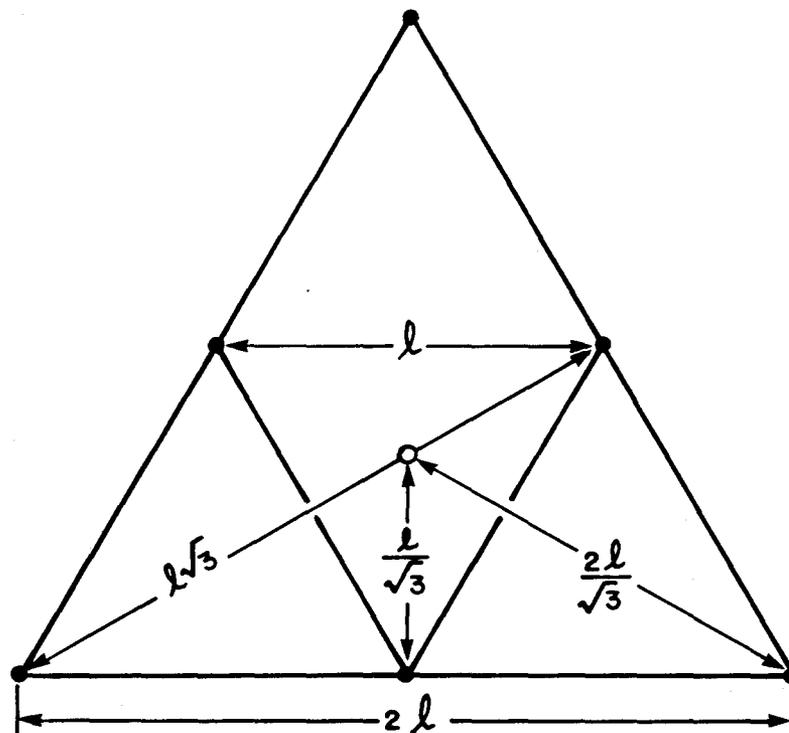


Figure 5.9. Position des stations pour le calcul de l'erreur d'interpolation au centre de deux triangles équilatéraux emboîtés.

constaté que l'addition d'une autre couronne de stations n'améliorait pas l'écart type de façon importante et avons décidé de nous en tenir à six stations de façon à compenser pour l'impossibilité pratique d'obtenir un réseau de géométrie régulière et par conséquent d'obtenir des écarts types aussi faibles.

Dans le cas particulier du calcul de l'écart type d'interpolation au centre de deux triangles équilatéraux emboîtés, les poids p attribués aux trois stations intérieures sont identiques. De même, les poids q attribués aux stations extérieures sont aussi identiques entre eux, mais différents des poids p . Si le système d'équations 5.27 est solutionné pour ce cas, alors:

$$p = \frac{b(21/\sqrt{3}) - \tilde{b}(1/\sqrt{3}) + \frac{1}{3} [2\tilde{b}(1) + \tilde{b}(1\sqrt{3}) - 2\tilde{b}(21)]}{2 [\tilde{b}(1) + \tilde{b}(1\sqrt{3}) - \tilde{b}(21)]} \quad \text{Eq. 5.34}$$

$$q = \frac{1}{3} - p \quad \text{Eq. 5.35}$$

$$\lambda = \tilde{b}(1/\sqrt{3}) - \frac{1}{3} [\tilde{b}(1\sqrt{3}) + 2\tilde{b}(1)] + \tilde{b}(1\sqrt{3}) p \quad \text{Eq. 5.36}$$

Comme on peut le constater à la figure 5.9, la distance l est la distance de base du réseau et correspond à la distance qui sépare une station de ses voisines immédiates. Les valeurs des fonctions de structure expérimentales \tilde{b}_{ij} entre les diverses stations doivent être évaluées pour des distances égales à $1, 2l$ et $l\sqrt{3}$, tandis que les \tilde{b}_{0j} , entre le point de grille et les stations correspondent à des distances de $l/\sqrt{3}$ et $2l/\sqrt{3}$.

L'équation 5.28 devient alors:

$$E = 3 [p \tilde{b}(1/\sqrt{3}) + q \tilde{b}(21/\sqrt{3})] + \lambda - \sigma^2 \quad \text{Eq. 5.37}$$

Comme précédemment, σ^2 est l'ordonnée à l'origine de la fonction de structure expérimentale \tilde{b}_{ij} . En estimant p , q et λ par rapport aux valeurs de la fonction de structure expérimentale \tilde{b}_{ij} , dont l'argument est exprimé sous forme de distance, il est donc possible, à l'aide de l'équation 5.37, d'exprimer l'erreur quadratique moyenne ou l'écart type moyen en fonction de la distance entre les stations d'un réseau à mailles triangulaires (figure 5.10).

Un point important doit être souligné ici. La variation de l'écart type en fonction de la distance l entre les stations n'est pas une simple transposition de la valeur prise par la fonction de structure à la même distance et donc de la forme de la fonction de structure. Tel qu'indiqué plus haut, l'écart type à une distance l donnée, résulte de relations mathématiques utilisant les valeurs de la fonction de structure à cinq (5) distances différentes (1 , 21 , $1/\sqrt{3}$ et $21/\sqrt{3}$) et non seulement à la distance l . Par conséquent, les valeurs prises par la fonction de structure à cinq distances et donc la forme de la courbe de $1/\sqrt{3}$ à 21 , influent sur les valeurs de l'écart type estimé pour une distance l entre les stations. Il en résulte qu'il ne suffit pas que la fonction de structure d'une région soit en dessous de la fonction de structure d'une autre région pour qu'il en soit automatiquement de même des écarts types. Ainsi, si l'on compare les fonctions de structure des figures 5.7-b et 5.8-b, on constate que la fonction de structure de la région de Montréal demeure en dessous de celle de Gaspé - Baie-Comeau, au moins jusqu'à 120 km, distance maximale tracée pour la région de Montréal. L'observation de la figure 5.10 permet toutefois de constater que ces courbes se croisent à une distance un peu plus faible

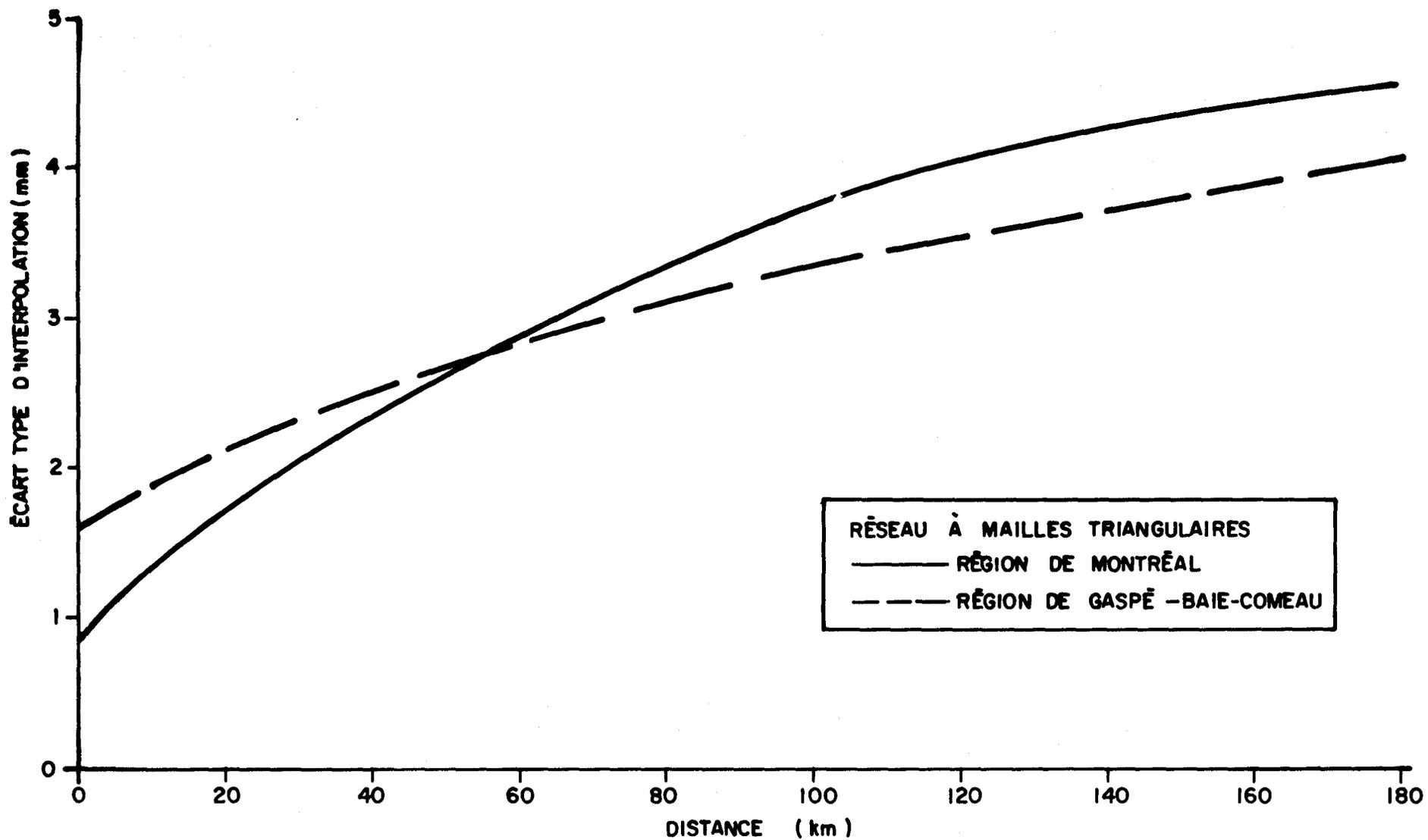


Figure 5.10 . Écart type d'interpolation en fonction de la distance entre les stations pour les pluies journalières au printemps.

que 60 km et que par la suite l'écart type est plus élevé dans la région de Montréal que dans la région de Gaspé - Baie-Comeau, même si la fonction de structure de la région de Montréal est plus faible que celle de la région de Gaspé - Baie-Comeau.

Il est à noter que cette courbe présente l'écart type moyen dans une région homogène et isotrope. Les écarts types obtenus pour une distance donnée sont donc à interpréter à la lumière des informations disponibles sur le degré de réalisation de ces propriétés du champ dans la région étudiée. L'étude de cette courbe permet aussi de déterminer la distance requise entre les stations pour atteindre une précision donnée et par suite d'évaluer le réalisme d'un tel critère, en fonction du nombre de stations que cela implique et des coûts associés. On peut aussi constater, par exemple, que l'erreur ne diminue pas tellement même si la distance entre les stations diminue de façon appréciable et que ce n'est que si les stations sont relativement près les unes des autres que l'erreur commence à vraiment diminuer. L'ordonnée à l'origine est aussi significative. Elle dénote qu'il est impossible d'obtenir, en moyenne, une erreur plus faible, quelle que soit la distance entre les stations. Naturellement, tout comme les fonctions de structure changent de région en région et avec la saison, cette erreur minimale fluctue dans l'espace et dans le temps (saison).

5.1.4.6 Cartographie des écarts types d'interpolation du réseau en opération le 1er janvier 1978

Ce n'est pas suffisant de connaître quel serait l'écart type d'interpolation d'un réseau de géométrie régulière dont la distance entre les stations est 1. Il importe de vérifier dans quelle mesure le réseau en opération à une date donnée permet d'atteindre la précision désirée. Un réseau réel possède, en effet, une géométrie plus ou moins irrégulière, la densité de stations variant à l'intérieur d'une région. Il peut donc arriver que dans certaines parties de cette région, la densité du réseau soit amplement suffisante pour atteindre et même

dépasser la précision souhaitée, alors qu'ailleurs, dans la même région, la densité et la géométrie du réseau soient telles que la précision souhaitée est loin d'être atteinte. La cartographie des écarts types d'interpolation permet donc de vérifier l'aptitude d'un réseau à répondre à des critères de précision déterminés en tout point d'une région et par suite de modifier ce réseau en conséquence. Des cartes d'iso-écarts types ont par conséquent été préparées pour chaque région définie par l'analyse en composantes principales des données représentatives d'une saison particulière, et ce pour chacune des variables étudiées, à condition qu'il y ait suffisamment de données pour assurer une précision raisonnable aux résultats.

Dans chaque cas, les stations utilisées pour l'interpolation optimale sont celles qui étaient en opération au 1er janvier 1978 et sont situées à l'intérieur de la zone d'influence d'une région donnée. En pratique, l'écart type d'interpolation à partir du réseau existant est calculé en chaque point d'une grille carrée en résolvant les équations 5.27 et 5.28 pour chaque point de grille et les iso-écarts types sont tracés par ordinateur en interpolant les valeurs obtenues aux points de grille. La figure 5.11 présente une telle carte.

La ligne continue sur laquelle s'arrête les iso-écarts types est la limite de la zone d'influence de la région. Quant à la limite même de la région, elle est représentée en traits discontinus. L'analyse doit porter principalement sur la distribution des écarts types à l'intérieur de la limite de la région. Le calcul a toutefois été étendu jusqu'aux limites de la zone d'influence pour diverses raisons.

D'une part, tel qu'expliqué en 5.1.3.4, la limite d'une région fait ressortir la distribution géographique des points où un facteur principal est prépondérant. Son influence ne s'arrête pas brusquement à la limite d'une région mais diminue progressivement à l'extérieur de cette limite. Il faut donc tenir compte de cette influence dans l'analyse des régions contigües. Le calcul des écarts types d'interpolation

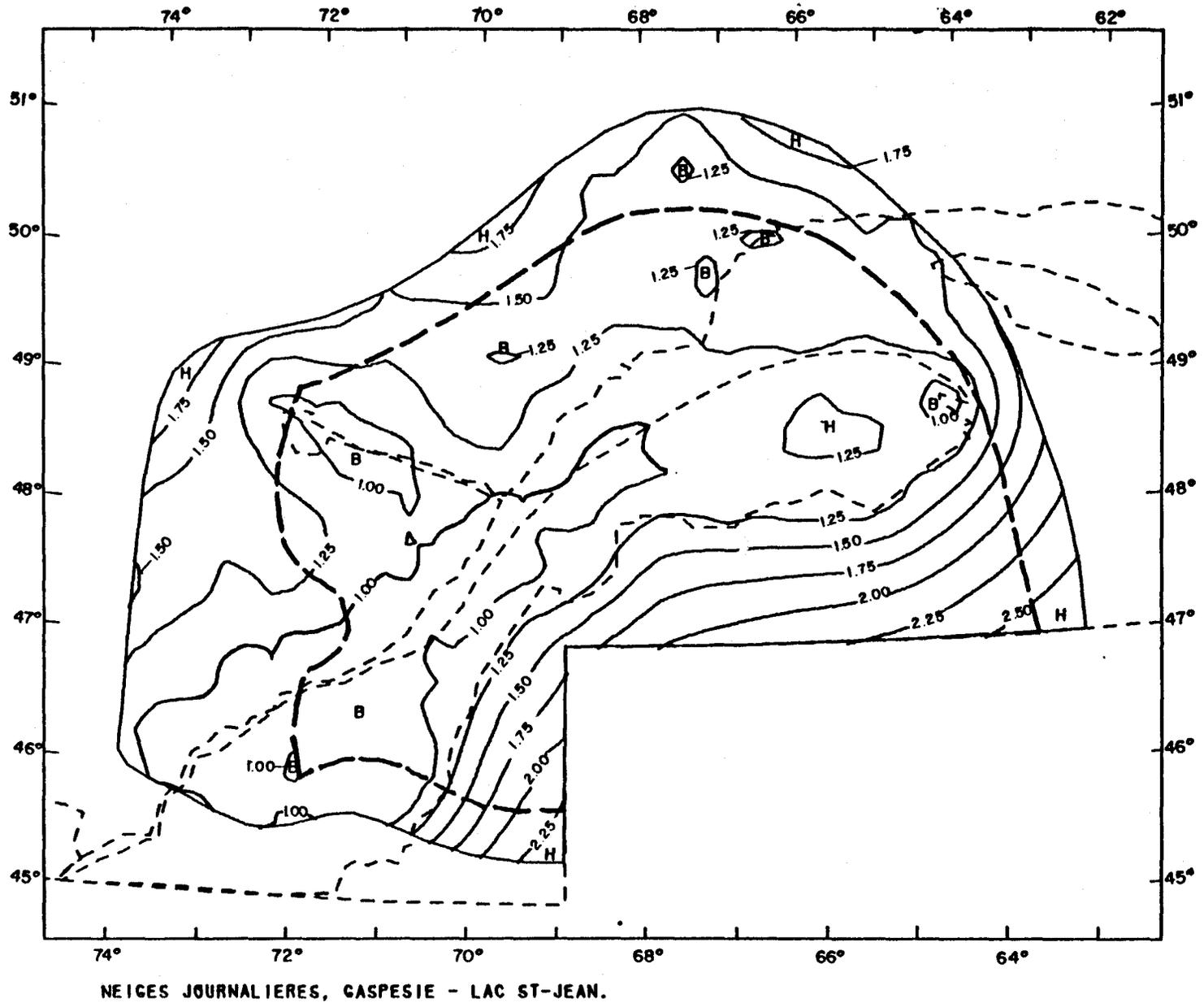


Figure 5.11. Répartition spatiale des écarts types.

est donc étendu jusqu'aux limites de la zone d'influence plutôt que jusqu'aux limites de la région, en considérant que la fonction de structure déterminée pour la région s'applique également entre ces limites. Ces valeurs peuvent être utilisées pour modifier à la hausse ou à la baisse les écarts types estimés dans chacune des régions contiguës sur lesquelles la zone d'influence de la région étudiée s'étend.

D'autre part, limiter le calcul aux points situés à l'intérieur de la région et surtout ne faire appel qu'aux stations situées à l'intérieur de cette même région, pourrait créer des effets de bord importants, c'est-à-dire une hausse artificielle et injustifiée des valeurs estimées le long des limites de la région. Ces effets de bord sont repoussés aux limites de la zone d'influence, dans le cas présent, ce qui permet d'assurer une meilleure continuité aux limites de la région.

5.1.4.7 Interprétation des résultats

Il importe tout d'abord de noter que les écarts types présentés aux figures 5.10 et 5.11 sont des valeurs moyennes et que l'écart réel d'une interpolation basée sur les observations aux points de mesure peut s'avérer fort différent de cet écart type moyen. Il faut donc interpréter ces valeurs dans un sens statistique.

L'examen d'une carte d'iso-écarts types permet rapidement de localiser les parties d'une région où les écarts sont les plus faibles de même que celles où ils sont le plus élevés. Logiquement, toute partie d'une région pour laquelle l'écart type d'interpolation est élevé est candidate à une augmentation de la densité de stations. Il faut alors vérifier si cette valeur élevée est supérieure à l'écart type accepté. Si oui, la courbe d'écarts types en fonction de la distance (voir figure 5.10) réalisée pour la même saison et la même région permet de déterminer la distance l requise entre les stations pour diminuer l'écart type et atteindre la précision requise. Si la distance requise est

trop faible et irréaliste, il y a lieu de reviser le critère de précision souhaité. Dans le cas contraire, la densité du réseau peut être augmentée sur cette base, en tenant compte des diverses contraintes logistiques, humaines et financières inhérentes au maintien d'un réseau. Mais il y a plus. Les propriétés de l'interpolation optimale permettent, avant même l'implantation de nouvelles stations en des lieux précis, d'analyser l'effet de leur implantation sur le patron d'iso-écarts types d'une région. Il suffit, pour cela, de reprendre le calcul des écarts types aux points de grille en ajoutant aux n stations déjà existantes les m stations projetées avec leurs positions. On peut aussi, en procédant à une série d'essais, trouver les positions optimales de ces stations (Delhomme et Delfiner, 1973).

Une procédure semblable permet d'éliminer des stations au besoin, lorsque la densité de stations est amplement suffisante pour respecter les critères de précision souhaités.

On verra, au cours de l'étude détaillée des diverses variables météorologiques, qu'un certain nombre de régions, principalement celles situées dans le nord du Québec, ne possédaient pas, au moment de réaliser l'étude, suffisamment de stations pour que la structure statistique des phénomènes soit connue avec suffisamment de précision. Dans ce cas, uniquement quelques renseignements préliminaires pourront être fournis sur la densité de stations nécessaire pour respecter une précision donnée.

Enfin, l'analyse ne tient pas compte de l'équivalence temps - (densité de stations). Par ailleurs, ce concept, qui se résume essentiellement à remplacer quelques stations opérant pendant longtemps par un plus grand nombre de stations en opération pendant un laps de temps plus court, s'applique uniquement pour l'estimation des valeurs moyennes à long terme des variables météorologiques.

5.2 ECART TYPE D'ESTIMATION DE LA MOYENNE DES PRECIPITATIONS SUR UNE SURFACE

5.2.1 Objectif

Jusqu'à maintenant, nous n'avons discuté que de l'évaluation des erreurs ponctuelles d'interpolation. Il est évident que l'intégration de ces erreurs sur une surface donnée ne correspond pas à l'erreur de l'estimé sur cette surface. Par ailleurs, les études hydrologiques, plus particulièrement les modèles paramétriques ou statistiques pluie - débit, ou encore les prévisions de débits sur un bassin versant, nécessitent l'estimé des erreurs sur une surface donnée.

5.2.2 Théorie

Kagan (1966) a développé une méthode simple pour estimer l'erreur commise dans l'évaluation de la précipitation moyenne sur une superficie donnée, si le réseau d'observation est uniformément distribué.

La base de la méthode proposée par Kagan est la fonction de corrélation calculée à l'aide des précipitations mesurées aux stations météorologiques d'une région homogène.

Cette fonction, qui dépend de la distance, peut être représentée par la forme exponentielle suivante:

$$\rho(d) = \rho(0) e^{-d/d_0} \quad \text{Eq. 5.38}$$

où:

$\rho(d)$ est la corrélation entre les précipitations observées à deux stations distantes de d ;

$\rho(0)$ est la corrélation correspondant à la distance zéro;

d_0 est le "rayon de corrélation" ou la distance pour laquelle la corrélation est égale à $\rho(0)/e$.

Théoriquement $\rho(0)$ doit être égal à l'unité, mais en pratique c'est rarement le cas. Ceci est dû aux erreurs de mesure des précipitations et au micro-climat, tout comme on l'a vu précédemment, pour la fonction de structure. On peut d'ailleurs estimer l'écart type de ces erreurs par:

$$\sigma = (1 - \rho(0)) \sigma_h^2 \quad \text{Eq. 5.39}$$

où:

σ = écart type relié aux erreurs de mesure et au micro-climat;

σ_h^2 = variance des précipitations en un point.

Kagan (1966) a montré que la variance de l'erreur d'estimation de la précipitation moyenne sur une surface (s) avec une station centrale est donnée par:

$$V = \sigma_h^2 (1 - \rho(0)) + 0.23 \sigma_h^2 \sqrt{\frac{s}{d_0}} \quad \text{Eq. 5.40}$$

où:

σ_h^2 est la variance des observations

d_0 est le "rayon de corrélation" introduit dans la formule 5.38.

Le premier terme de l'équation 5.40 est attribué à l'erreur associée aux mesures et au micro-climat et le second est attribué à la variance des champs de précipitations.

Pour une surface S avec n stations distribuées uniformément:

$$S = n s \quad \text{Eq. 5.41}$$

où:

n est le nombre de station;

s surface attribuée à chaque station.

La variance V_n de l'erreur sur la moyenne des précipitations sur la surface S est donnée par:

$$V_n = \frac{\sigma_h^2}{n} \left[1 - \rho(o) + \frac{0.23 \sqrt{S}}{d_o \sqrt{n}} \right] \quad \text{Eq. 5.42}$$

Et l'écart type s'écrit:

$$E_n = \sqrt{V_n} \quad \text{Eq. 5.43}$$

Si pour une région donnée on connaît σ_h , $\rho(o)$ et d_o , on peut calculer le nombre de stations requises pour obtenir un écart type spécifié E_n , ou, inversement, évaluer E_n pour un nombre n de stations.

5.2.3 Condition d'application de la méthode de Kagan.

Compte tenu des dimensions du Québec et des variations physiographiques, il est évident que le Québec doit être divisé en régions à l'intérieur desquelles les données peuvent être considérées comme homogènes dans l'espace.

Notons immédiatement que la méthode de Kagan (1966) sera appliquée aux mêmes régions que celles qui auront été définies par composantes principales pour la détermination des écarts types d'interpolation des valeurs ponctuelles.

La variation de la corrélation en fonction de la distance est déterminée en calculant, pour tous les couples de stations d'une région homogène, les coefficients de corrélation et les distances les séparant. Comme pour la fonction de structure, l'axe des distances est subdivisé en classes et on calcule à l'intérieur de chaque classe la corrélation moyenne. On ajuste alors la courbe exponentielle décrite par l'équation 5.38. Comme d'après Kagan (1966), la forme exponentielle décrit la plupart des fonctions de corrélation pour des distances de 10 à 150 km; l'ajustement a été fait en utilisant essentiellement les points dont la distance est inférieure à 150 km. Pour des distances inférieures à 10 ou 15 km, la corrélation décroît plus rapidement que dans l'équation 5.38, ce qui correspond, d'après Kagan (1966), aux irrégularités du micro-climat sur des petites distances et se traduit par un $\rho(0)$ plus élevé que celui qui est estimé par l'équation 5.38. Compte tenu du nombre de points disponibles, on ne peut ajuster de courbe pour ces distances. On peut, cependant, tenir compte de cette particularité dans l'analyse des résultats.

5.2.4 Interprétation des résultats

Le calcul des équations 5.42 et 5.43 pour différents n et s nous donne l'écart type d'estimation de la moyenne des précipitations sur la surface S ($s = n s$).

Cette équation dépend de trois paramètres soit σ_h^2 , $\rho(0)$ et d_0 . Les paramètres $\rho(0)$ et d_0 sont estimés à l'aide de l'ajustement de l'équation 5.38. Selon le nombre de points utilisés et la distribution de ces points, les paramètres $\rho(0)$ et d_0 ont un intervalle de confiance plus ou moins grand. Pour montrer l'influence de chaque paramètre sur le calcul de l'écart type de l'estimé de la moyenne on peut faire varier chacun de ces paramètres et analyser la variation de l'écart type.

A titre d'exemple, prenons les pluies journalières du printemps, région Saguenay - Lac-St-Jean. La première ligne du tableau 5.3 nous donne les paramètres estimés pour les précipitations journalières de

cette région. Les autres lignes nous montre les nouvelles valeurs imposées aux paramètres $\rho(o)$ et d_o dans le but d'analyser l'influence de chaque paramètre.

Tableau 5.3 Variation des paramètres $\rho(o)$ et d_o pour l'analyse de sensibilité de ces paramètres.

σ_h (mm)	$\rho(o)$ (mm)	d_o (km)	Résultats
6.1	.914	90	Tableau 5.4
6.1	.80	90	Tableau 5.5
6.1	.95	90	Tableau 5.6
6.1	.914	50	Tableau 5.7
6.1	.914	150	Tableau 5.8

Le tableau 5.4 nous donne l'écart type d'estimation de la moyenne des précipitations liquides journalières en printemps pour la région Saguenay - Lac-St-Jean.

Pour une station, l'écart type d'estimation varie de 2.02 à 4.05 mm pour des superficies variant de 100 à 20,000 km². Pour une superficie donnée, l'écart type d'estimation varie rapidement avec le nombre de stations. Prenons, par exemple, une superficie totale de 10,000 km², avec une station l'écart type d'estimation est de 3.54 mm. Deux stations réparties sur la même superficie, c'est-à-dire, 5,000 km² attri-

Tableau 5.4 Ecart type d'estimation (mm) de la moyenne des précipitations liquides journalières au printemps en fonction de la superficie attribuée à chaque station et du nombre de station: région Saguenay - Lac-St-Jean.

Nombre de stations	SUPERFICIE (km ²) PAR STATION							
	100	200	500	1000	2000	5000	10000	20000
1	2.02	2.12	2.29	2.47	2.71	3.13	3.54	4.05
2	1.43	1.50	1.62	1.75	1.92	2.21	2.50	2.87
3	1.17	1.22	1.32	1.43	1.57	1.81	2.04	2.34
4	1.01	1.06	1.15	1.24	1.36	1.56	1.77	2.03
5	0.90	0.95	1.03	1.11	1.21	1.40	1.58	1.81
10	0.64	0.67	0.72	0.78	0.86	0.99	1.12	1.28
20	0.45	0.47	0.51	0.55	0.61	0.70	0.79	0.91
50	0.29	0.30	0.32	0.35	0.38	0.44	0.50	0.57
100	0.20	0.21	0.23	0.25	0.27	0.31	0.35	0.40

1 Saguenay - Lac-St-Jean

bués à chaque station, nous donne un écart type d'estimation de 2.21 mm. De la même façon, nous obtenons 1.11, 0.72, 0.30 et 0.20 mm pour 5, 10, 50 et 100 stations respectivement, représentant chacune des superficies de 2000, 1000, 200 et 100 km².

Les données des tableaux 5.5 et 5.6 comparées aux données du tableau 5.4 nous montre l'influence du paramètre $\rho(o)$ sur l'écart type d'estimation. Le paramètre $\rho(o)$ a été modifié arbitrairement de 0.914 (tableau 5.4) à 0.80 et à 0.95 respectivement pour les calculs des données des tableaux 5.5 et 5.6. Sur le tableau 5.5, on voit que l'augmentation de l'écart type varie de 43% à 12% selon la superficie attribuée à chaque station. Le tableau 5.6 nous montre des diffluenes qui va de -18% à -4%.

La comparaison des données des tableaux 5.7 et 5.8 avec les données du tableau 5.4, nous renseigne sur l'influence du paramètre d_o pour le calcul de l'écart type d'estimation. Le paramètre d_o a été modifié arbitrairement de 90 km (tableau 5.4) à 50 km (tableau 5.7) et à 150 km (tableau 5.8). Pour une diminution de d_o de 90 km à 50 km, l'écart type de l'estimé augmente de 9% pour une superficie de 100 km² par station à 28% pour une superficie de 20,000 km² par station (tableau 5.7). Pour une augmentation de d_o de 90 km à 150 km, la diminution de l'écart type de l'estimé varie de -5% à -18% (tableau 5.8).

L'étude des tableaux 5.5 à 5.7 nous montre que si le rapport \sqrt{S}/\sqrt{n} reste constant, le pourcentage d'augmentation ou de diminution de l'écart type de l'estimé reste le même, peu importe le nombre de stations. On voit également que l'écart type de l'estimé baisse si $\rho(o)$ ou d_o augmentent.

Tableau 5.5 Ecart type d'estimation (mm) de la moyenne des précipitations liquides journalières au printemps en fonction de la superficie attribuée à chaque station et du nombre de station: région Saguenay - Lac-St-Jean. Paramètre $\rho(o) = 0.8$.

Nombre de stations	SUPERFICIE (km ²) PAR STATION							
	100	200	500	1000	2000	5000	10000	20000
1	2.88	2.94	3.07	3.21	3.40	3.74	4.09	4.54
2	2.03	2.08	2.17	2.27	2.40	2.64	2.89	3.21
3	1.66	1.70	1.77	1.85	1.96	2.16	2.36	2.62
4	1.44	1.47	1.54	1.61	1.70	1.87	2.04	2.27
5	1.29	1.32	1.37	1.44	1.52	1.67	1.83	2.03
10	0.91	0.93	0.97	1.02	1.07	1.18	1.29	1.44
20	0.64	0.66	0.69	0.72	0.76	0.84	0.91	1.01
50	0.41	0.42	0.43	0.45	0.48	0.53	0.58	0.64
100	0.29	0.29	0.31	0.32	0.34	0.37	0.41	0.45

Différence en % par rapport au tableau 5.4

43% 39% 34% 34% 25% 19% 16% 12%

Tableau 5.6 Ecart type d'estimation (mm) de la moyenne des précipitations liquides journalières au printemps en fonction de la superficie attribuée à chaque station et du nombre de station: région Saguenay - Lac-St-Jean. Paramètre (ρ) = 0.95.

Nombre de stations	SUPERFICIE (km ²) PAR STATION							
	100	200	500	1000	2000	5000	10000	20000
1	1.67	1.78	1.98	2.19	2.46	2.91	3.35	3.89
2	1.18	1.26	1.40	1.55	1.74	2.06	2.37	2.75
3	0.96	1.03	1.14	1.27	1.42	1.68	1.93	2.24
4	0.83	0.89	0.99	1.10	1.23	1.46	1.67	1.94
5	0.74	0.80	0.89	0.98	1.10	1.30	1.50	1.74
10	0.53	0.56	0.63	0.69	0.78	0.92	1.06	1.23
20	0.37	0.40	0.44	0.49	0.55	0.65	0.75	0.87
50	0.24	0.25	0.28	0.31	0.35	0.41	0.47	0.55
100	0.17	0.18	0.20	0.22	0.25	0.29	0.33	0.39

Différence en % par rapport au tableau 5.4

-18% -16% -13% -11% -10% -7% -5% -4%

Tableau 5.7 Ecart type d'estimation (mm) de la moyenne des précipitations liquides journalières au printemps en fonction de la superficie attribuée à chaque station et du nombre de station: région Saguenay - Lac-St-Jean.
Paramètre $d_0 = 50$ km.

Nombre de stations	SUPERFICIE (km ²) PAR STATION							
	100	200	500	1000	2000	5000	10000	20000
1	2.20	2.36	2.63	2.92	3.27	3.89	4.48	5.20
2	1.56	1.67	1.86	2.06	2.31	2.75	3.17	3.68
3	1.27	1.36	1.52	1.68	1.89	2.24	2.59	3.00
4	1.10	1.18	1.32	1.46	1.64	1.94	2.24	2.60
5	0.98	1.05	1.18	1.30	1.46	1.74	2.00	2.33
10	0.70	0.74	0.83	0.92	1.03	1.23	1.42	1.64
20	0.49	0.53	0.59	0.65	0.73	0.87	1.00	1.16
50	0.31	0.33	0.37	0.41	0.46	0.55	0.63	0.74
100	0.22	0.24	0.26	0.29	0.33	0.39	0.45	0.52

Différence en % par rapport au tableau 5.4

+9% 11% 15% 18% 21% 24% 26% +28%

Tableau 5.8 Ecart type d'estimation (mm) de la moyenne des précipitations liquides journalières au printemps en fonction de la superficie attribuée à chaque station et du nombre de station: région Saguenay - Lac-St-Jean. Paramètre $d_0 = 150$ km.

Nombre de stations	SUPERFICIE (km ²) PAR STATION							
	100	200	500	1000	2000	5000	10000	20000
1	1.93	1.99	2.10	2.22	2.38	2.67	2.96	3.33
2	1.36	1.41	1.49	1.57	1.68	1.89	2.10	2.36
3	1.11	1.15	1.21	1.28	1.38	1.54	1.71	1.93
4	0.96	0.99	1.05	1.11	1.19	1.34	1.48	1.67
5	0.86	0.89	0.94	0.99	1.07	1.19	1.33	1.49
10	0.61	0.63	0.66	0.70	0.75	0.84	0.94	1.05
20	0.43	0.44	0.47	0.50	0.53	0.60	0.66	0.75
50	0.27	0.28	0.30	0.31	0.34	0.38	0.42	0.47
100	0.19	0.20	0.21	0.22	0.24	0.27	0.30	0.33

Différence en % par rapport au tableau 5.4

-5% -6% -8% -10% -12% -15% -16% -18%

REFERENCES BIBLIOGRAPHIQUES

- ANDERSON, T.W. (1958).
An introduction to multivariate statistical analysis. John Wiley and Sons, New York, 374 p.
- BENSON, M.A. (1972).
Use of multiple-regression analysis in the design of a stream-gauging network - practice in the USA: Casebook on Hydrological Network Design Practice, chapter III, section 3.2, 4 p., WMO, Geneva, Switzerland.
- BRADY, P.J. (1978).
Optimal sampling and analysis using two variables and modeled cross-covariance functions. J. of Applied Meteorology, 17: 12-21, January.
- BRAS, R.L. and I. RODRIGUEZ-ITURBE. (1976).
Rainfall network design for runoff prediction. Water Resources Research, 12(6): 1197-1208, December.
- BRAS, R.L. and I. RODRIGUEZ-ITURBE. (1976).
Evaluation of mean square error involved in approximating the areal average of a rainfall event by a discrete summation. Water Resources Research, 12(2): 181-184, April.
- CHAPMAN, T.G. (1972).
The Australian representative basin program. Casebook on Hydrological Network Design Practice, chapter III, section 3.1, 5 p., WMO, Geneva Switzerland.
- CHARBONNEAU, R., J.P. FORTIN, J.P. LARDEAU, G. MORIN et W. SOCHANSKA. (1978).
Analyse des précipitations du bassin versant de la rivière Eaton. INRS-Eau, rapport scientifique No 88, 103 p., 2 annexes.
- CHAUVET, P., J. PAILLEUX and J.P. CHILES. (1976).
Analyse objective des champs météorologiques par cokrigeage. Sciences et Techniques, pp. 37-54.
- DAWDY, D.R., M.E. MOSS and N.C. MATALAS. (1972).
Application of systems analysis to network design. Casebook on Hydrological Network Design Practice, chapter III, section 4.1, 7 p., WMO, Geneva, Switzerland.

- DELFINER, P. (1973).
Analyse objective du géopotentiel et du vent géostrophique par Krigeage universel. La Météorologie, 25: 57.
- DELHOMME, J.P. et P. DELFINER. (1973).
Application du krigeage à l'optimisation d'une campagne pluviométrique en zone aride. Colloque sur l'élaboration des projets d'utilisation des ressources en eau sans données suffisantes. Madrid, juin 1973, 20 p.
- DELHOMME, J.P. (1978).
Kriging in the hydrosiences. Water Resources , 1(5): 251-266.
- DUBAND, D. (1974).
Analyse en composantes principales des séries pluviométriques des Alpes - Est Massif central-Cevennes, Rapport interne, Electricité de France, Service Energétique et Prévision, Grenoble, juillet.
- DUBAND, D. (1973).
Contrôle d'homogénéité des séries d'écoulement dans les Pyrénées. Reconstitution de données. Rapport interne, Electricité de France, Service Energétique et Prévision, Grenoble, juin.
- DYER, T.G.J. (1975).
The assignment of rainfall stations into homogeneous groups: an application of principal component analysis, Q.J. of the Royal Met. Soc., 101(430): 1005-1013.
- EDDY, A. (1967).
The statistical objective analysis of scalar data fields. Journal of Applied Meteorology, 6(4): 597-609, August.
- EDDY, A. (1976).
Optimal raingage densities and accumulation times: a decision-making procedure. Journal of Applied Meteorology, 15: 962-971, September.
- FERGUSON, H.L. (1972).
Precipitation network design for large mountainous areas. Association Internationale d'Hydrologie Scientifique. Pub. 106, 1972, pp. 85-110 et O.M.M. Pub. 326.
- GANDIN, L.S. (1970).
The planning of meteorological station network. World Meteorological Organization, Geneva, Switzerland, WMO-No. 265. T.P. 149. Technical Note No. 111, 35 p.

- GANDIN, L.S. (1965).
Objective analysis of meteorological fields. Israel Program for Scientific Translation, Jerusalem, 242 p.
- GANDIN, L.S. (1970).
The planning of meteorological station networks. WMO Technical Note 111, 1970.
- GUSCHINA, M.V., R.L. KAGAN and A.I. POLISHCHUK. (1967).
Accuracy in determining the mean precipitation depth over an area. Soviet Hydrology: Selected Papers, Issue No. 6, pp. 585-596.
- HARAGAN, D.R. (1976).
Spatial variation of precipitation on the Texas high plains. Water Resources Bulletin, 12(6): 1191-1204, December.
- HENDRICK, R.L. and G.H. COMER. (1970).
Space various of precipitation and implications for raingage network design. Journal of Hydrology, 10: 151-163.
- HERSHFIELD, D.M. (1965).
On the spacing of raingauges. Symposium Design of Hydrological Networks, Québec. IAHS Pub. No. 67, pp. 72-79.
- HOTELLING, H. (1933).
Analysis of a complex of statistical variables into principal components. J. Educ. Psych., 24: 417-441, 498-520.
- HUFF, F.A. and P.T. SCHICKEDANZ. (1972).
Space-time uncertainties in precipitation measurement. International Symposium on Uncertainties in Hydrologic and Water Resource Systems. 4.2 - 15 p.
- HUTCHINSON, P. (1970).
A contribution to the problem of spacing raingauges in rugged terrain. Journal of Hydrology, 12: 1-14.
- HUTCHINSON, P. and M. CISLEROVA. (1974).
The redesign of the raingauge network of Zambia. Hydrological Sciences, 19: 423-434.
- KAGAN, R.L. (1966).
An evaluation of the representativeness of precipitation data. Works of the Main Geophysical Observatory, vol. 191.

- KENDALL, M.G. (1957).
A course in multivariate analysis. Chas. Griffin and Co., London,
185 p.
- KRUGER, H.G. (1969).
General and special approaches to the problem of objective analysis of
meteorological variables. Quaterly Journal of Royal Meteorological
Society, 95: 21-39.
- KRUIZINGA, S. and G.J. YPERLAAN. (1978).
Spatial interpolation of daily totals of rainfall. Journal of
Hydrology, 36: 65-73.
- LENTON, R.L. and I. RODRIGUEZ-ITURBE. (1977).
Rainfall network systems analysis: The optimal estimation of total
areal storm depth. Water Resources Research, 13(5): 825-836, October.
- LENTON, R.L. and I. RODRIGUEZ-ITURBE. (1977).
A multidimensional model for the synthesis of processes of areal
rainfall averages. Water Resources Research, 13(3): 605-612, June.
- MATALAS, N. C. and B.J. REITHER. (1967).
Some comments on the use of factor analysis. Water Resources Research,
3(1): 213-223.
- MATHERON, G. (1965).
Les variables régionalisées et leur estimation, Paris, Masson et Cie.
306 p.
- MORIN, G., R. CHARBONNEAU, J.P. FORTIN et W. SOCHANSKA. (1979).
Etude du réseau météorologique et du réseau nivométrique de la région
du lac Saint-Jean. Rapport scientifique No 91, INRS-Eau, 268 p.
- NORDENSON, L. (1972).
Review of co-coordinated mapping. Casebook on Hydrological Network
Design Practice, chapter III, section 2.2., 4 p., WMO, Geneva,
Switzerland.
- O'CONNELL, P.E., R.S. GURNEY, D.A. JONES, J.B. MILLER, C.A. MICHOLASS and
M.R. SENIOR. (1978).
Rationalization of the Wessex Water Authority raingauge network.
Institute of Hydrology, Report No. 51, 179 p.

- O'CONNELL, P.E., M.A. BERAN, R.J. GURNEY, D.A. JONES and R.J. MOORE. (1977).
Methods for evaluating the U.K. raingauge network. Institute of Hydrology, Report No. 40, 262 p.
- PETERSEN, D.P. (1973).
A comparison of the performance of quasi-optimal and conventional objective analysis schemes. *Journal of Applied Meteorology*, 12(7): 1093-1101, October.
- RODDA, J.C. (1969).
Hydrological network design-needs, problems and approach. WMO ref. No 12 on WMO/IHD projects, 57 p.
- RODDA, J.C. (1972).
Casebook on Hydrological Network Design Practice, WMO ref. No. 324.
- RODRIGUEZ-ITURBE, I. and J.M. MEJIA. (1974).
The design of rainfall networks in time and space. *Water Resources Research*, 10(4): 713-728, August.
- ROYER, J.F. (1975).
Comparaison des méthodes d'analyse objective par interpolation optimale et par approximations successives. Secrétariat Général à l'Aviation Civile, Direction de la Météorologie Nationale, Notes internes de l'établissement d'études et de recherches météorologiques, No. 365, août, 25 p.
- RUMYANTSEV, V.A. and S.V. SHANOCHKIN. (1973).
Evaluation of the representativeness of the precipitation network in relation to the spatial interpolation of precipitation. *Soviet Hydrology: Selected Papers*, Issue No. 3, pp. 199-206.
- RUTHERFORD, I.D. (1972).
Data assimilation by statistical interpolation forecast error fields. *Journal of the Atmospheric Sciences*, American Meteorological Society, 29(5): 809-815, July.
- SHARON, D. (1972a).
The spottiness of rainfall in a desert area. *Journal of Hydrology*, 17: 161-175.

- SHARON, D. (1972b).
Spatial analysis of rainfall data from dense networks. Bull. Int. Assoc. scient. Hydrol., 17(3): 291-300.
- SCHLATTER, T.W. (1975).
Some experiments with a multivariate statistical objective analysis scheme. Monthly Weather Review, 103: 246-257, March.
- SOLOMON, S.I. (1972).
Multi-regionalization and network strategy. Casebook on Hydrological Network Design Practice, chapter III, section 3.3., 11 p., WMO, Geneva, Switzerland.
- SOLOMON, S.I. (1972).
Joint Mapping: Casebook on Hydrological Network Design Practice, chapter II, section 2.1, 16 p., WMO, Geneva, Switzerland.
- THIEBAUX, H.J. (1974).
Estimation of covariances of meteorological parameters using local-time averages. Journal of Applied Meteorology, 13: 592-600, August.
- YEFREMOVA, N.I. (1970).
Spatial statistical structure of atmospheric precipitation average over large areas. Soviet Hydrology: Selected Papers, Issue No. 3, pp. 236-251.
- ZAWADZKI, I.I. (1973).
Errors and fluctuations of raingauge estimates of areal rainfall. Journal of Hydrology, 18: 243-255.