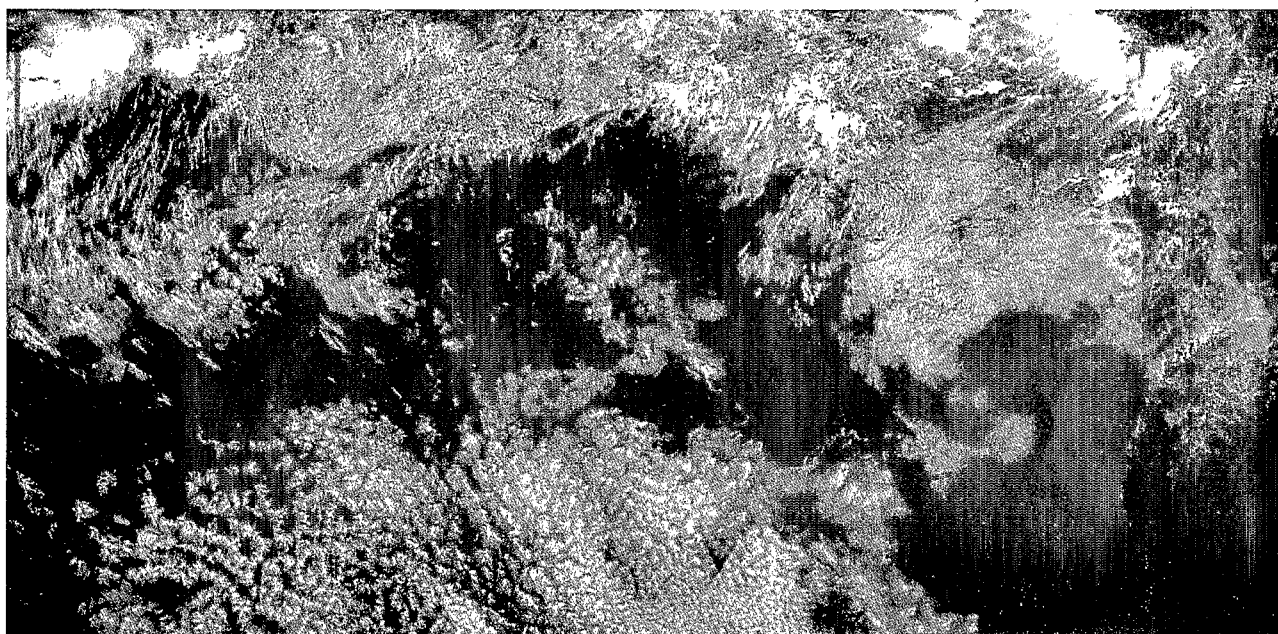
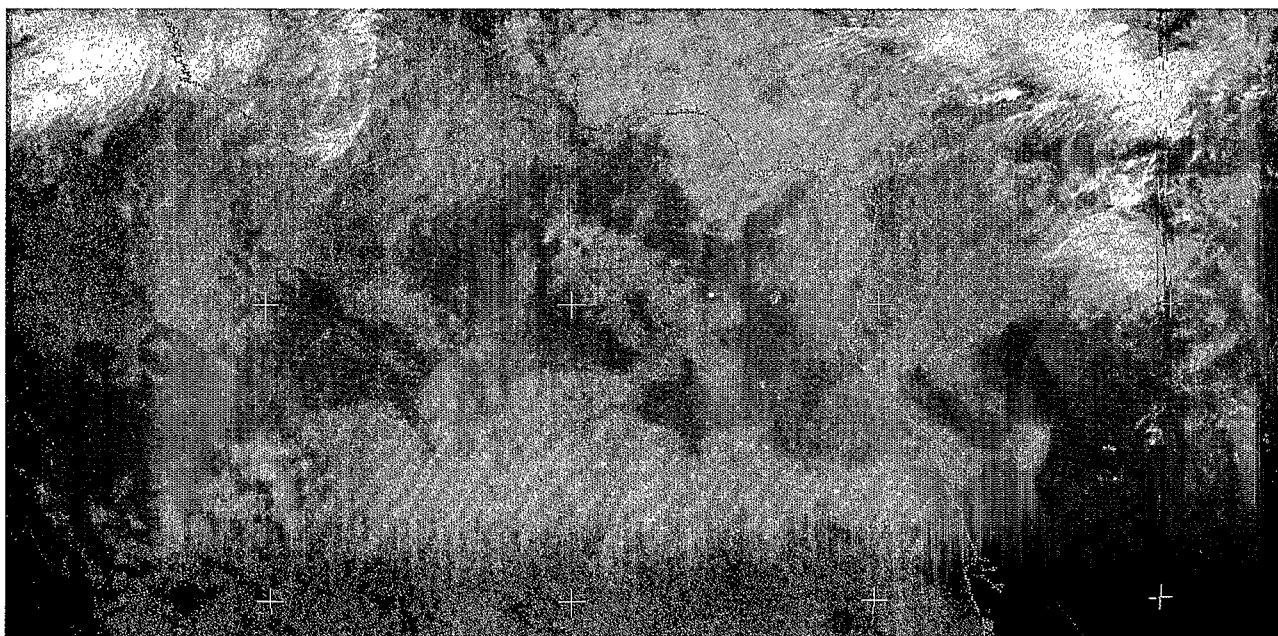


Ministère de la Coopération

# VEILLE CLIMATIQUE SATELLITAIRE



METEOFRANCE CMS LANNION - METEOSAT 4 - 01/08/92 - 12 H 00 UTC - CANAL VISIBLE



METEOFRANCE CMS LANNION - METEOSAT 4 - 01/08/92 - 12 H 00 UTC - CANAL INFRAROUGE

B\* 14441; 43; 45; 47; 49; 51



# APPLICATION DE LA MÉTHODE DU VECTEUR RÉGIONAL A L'ANALYSE DE LA PLUVIOMÉTRIE ANNUELLE DU BASSIN AMAZONIEN

## APPLICATION OF THE REGIONAL VECTOR METHOD TO THE ANALYSIS OF ANNUAL PLUVIOMETRY OF THE AMAZON BASIN

par G. HIEZ<sup>1</sup>, G. COCHONNEAU<sup>2</sup>, P. SECHET<sup>2</sup>, U.M. FERNANDES<sup>3</sup>

(traduit du portugais<sup>5</sup> et révisé par G. HIEZ)

### RÉSUMÉ

La Méthode du Vecteur Régional est un modèle simple qui permet de représenter les informations pluviométriques d'une région par une séquence d'indices annuels, représentatifs des précipitations dans cette région, et par des coefficients caractéristiques de chaque poste d'observation. La méthode a été conçue pour répondre à deux objectifs : d'une part la critique de l'information et d'autre part sa synthèse. En vue de son application systématique, un logiciel professionnel convivial a été développé. Le modèle a été appliqué aux données pluviométriques annuelles du bassin amazonien, fournies par le "Système d'Informations Hydrologiques" (SIH), géré par le DNAEE<sup>4</sup>. 46 vecteurs régionaux ont ainsi été élaborés. Ils permettent une représentation synthétique de la pluviométrie de la partie brésilienne du bassin amazonien et fournissent un outil efficace d'étude des tendances climatiques, entre autres, par l'Analyse en Composantes Principales.

### ABSTRACT

The Regional Vector Method is a simple model which permits pluviometric data from a region to be represented by a sequence of annual indexes, representative of the precipitations in this region, and by the characteristic coefficients of each observation post. The method was devised to satisfy two objectives : on the one hand the critical examination of data and on the other hand its synthesis. With its systematic application in mind, a convivial professional software has been developed. The model has been applied to annual pluviometric data from the Amazon Basin, supplied by the "Hydrological Data System", managed by the DNAEE<sup>4</sup>. 46 regional vectors have been elaborated in this manner. They permit a synthetic representation of the pluviometry of the Brazilian part of the Amazonian basin and provide a useful study tool for climatic trends, for example, through the Principal Component Analysis.

### INTRODUCTION

L'analyse de la pluviométrie d'une grande région, comme le bassin de l'Amazone, avec ses 6,1 millions de km<sup>2</sup>, dont 3,9 millions en territoire brésilien, implique en un premier temps la prise en considération d'innombrables problèmes de données. Cela nécessite l'établissement d'une méthodologie de traitement informatisé pour la vérification, l'homogénéisation et la synthèse de l'information qu'il est possible de réunir.

A cette fin, deux outils fondamentaux sont indispensables :

– une base de données contenant les séries historiques des observations pluviométriques de la région considérée ;

– un logiciel pour l'application systématique d'un modèle de synthèse aux informations recueillies.

### INTRODUCTION

The analysis of the pluviometry of a vast region, like the Amazonian Basin, with its 6.1 million km<sup>2</sup>, of which 3.9 million km<sup>2</sup> is Brazilian, first of all, involves innumerable data problems being taken into account. This requires the setting up of a computerized processing method for the validation, homogenization and the synthesis of data which it was possible to collect.

With this aim in mind, two basic tools are indispensable :

– a data base with a historical series of pluviometric observations from the region under study ;

– software for the systematic application of a synthesis model to collected data.

1 Directeur de Recherche de l'ORSTOM.

2 Ingénieur de Recherche de l'ORSTOM.

3 Ingénieur du DNAEE.

4 Departamanto Nacional de Aguas e Energia Elétrica.

5 Aplicação do método do vetor regional a análise da pluviometria anual da bacia amazônica.



Ces deux conditions ont été rencontrées au DNAEE et plus précisément dans son département de "Coordination Générale des Ressources Hydriques" (CGRH) à Brasilia, responsable de la maintenance du "Système d'Informations Hydrologiques", le SIH (Brésil 1982). La technologie de la Méthode du Vecteur Régional, (MVR), quant à elle, développée à l'Institut Français de Recherche Scientifique pour le Développement en Coopération, ou ORSTOM (Hiez, 1977) a été transférée et adaptée au contexte brésilien (Hiez et Rancan, 1983), dans le cadre de la convention signée entre les deux institutions. Elle a donné lieu à la création du logiciel MVR, publié dans la collection "Logorstom".

La méthode repose sur l'extraction de la tendance régionale prépondérante formée des indices représentatifs de la région – le vecteur régional – et des coefficients propres à chaque poste d'observation. Cette double extraction est réalisée à partir de la matrice des données grâce à un procédé de normalisation successive des lignes et des colonnes, utilisant un algorithme de maximum de probabilité. Outre le vecteur régional, MVR fournit une aide précieuse à l'interprétation des résultats et suggère les corrections susceptibles d'être apportées par l'utilisateur aux totaux annuels.

En un deuxième temps, et bien que la méthode précédente ait été résolument ancrée sur des principes non euclidiens, il s'est avéré intéressant de soumettre le *champ des vecteurs régionaux* ainsi obtenus à la méthode des composantes principales, avec l'objectif de dégager, si elles existent, les grandes tendances indépendantes (mathématiquement) des régimes pluviométriques.

Cet article montre les facilités offertes par le puissant outil que constitue MVR pour synthétiser l'information fournie par les quelque 1.250 postes de la région amazonienne. Le traitement porte principalement sur les données de la période 1961-1988, mais pour certains postes, il inclut des séries plus longues. Il a été possible d'évaluer la qualité des observations afin d'identifier des postes de référence et détecter les séries les plus douteuses. La deuxième méthode, quant à elle, a permis de mettre en évidence des zones homogènes et de caractériser objectivement la pluviométrie dans la partie brésilienne de l'Amazonie.

## PRÉSENTATION DU LOGICIEL "MVR"

La méthode du vecteur régional (MVR) repose sur un modèle mathématique conçu pour répondre à une double préoccupation : la critique des informations d'une région déterminée et leur synthèse. Le logiciel MVR a été développé pour rendre possible son exploitation systématique sur des informations contenues dans les grandes banques de données pluviométriques.

These two conditions were met at the DNAEE and more precisely in its "General Co-ordination of Hydric Resources" department at Brasilia, responsible for the maintenance of the "Hydrological Data System" (SIH) (Brasil 1982). The Regional Vector Method (RVM) technology, developed by the French Institute of Scientific Research for the Development in Cooperation, or ORSTOM, (Hiez, 1977), has been transferred and adapted to the Brazilian context (Hiez and Rancan, 1983), within the framework of the convention signed between the two institutions. It gave rise to the creation of the RVM software, published in the "Logorstom" collection.

The method is based on the extraction of the principal regional trend formed from indexes which are representative of the region – the regional vector – and coefficients characteristic of each observation post. This double extraction is carried out from the data matrix thanks to a successive normalization procedure of lines and columns, using a maximum probability algorithm. Besides the regional vector, the RVM provides a precious aid in interpreting the results and gives an indication of possible corrections that could be made on the annual totals.

In the second phase, although the method mentioned above has been resolutely anchored on non-euclidean principles, it proved to be interesting to submit the regional vector field which was obtained to the principal component analysis method, with the intention of extracting, if they exist, the significant independent trends (mathematically) of the pluviometric regime.

This article shows the facilities offered by the powerful tool which makes up the RVM to synthesize data supplied by the 1250 posts in the Amazonian region. The data processing has been mainly carried out for the 1961-1988 period, but for some posts, it includes a longer series. It has been possible to evaluate the quality of observations in order to identify the reference posts and to detect the most doubtful series. The second method has enabled homogeneous zones to be emphasized and to objectively characterise the pluviometry in the Brazilian part of the Amazon.

## PRESENTATION OF THE "RVM" SOFTWARE

The regional vector method (RVM) is based on a mathematical model elaborated to satisfy a double concern : the criticism of data from a defined region and its synthesis. The RVM software has been developed to make its systematic operation possible on data contained in the large pluviometric Data Base.

## Méthode du vecteur régional

La pluviométrie d'une région est mesurée à intervalles de temps prédéterminés sur un réseau de postes d'observation. L'information de base peut donc être représentée par une matrice de données où chaque colonne représente un *lieu* et chaque ligne une *date* d'observation.

Malheureusement, les données collectées, sans compter les manques d'observation, souffrent généralement d'erreurs en tout genre. Toutefois, dans la mesure où le réseau de postes est suffisamment dense pour garantir un certain degré de redondance et en supposant qu'il existe une tendance climatique régionale prédominante, il est raisonnable de supposer que la matrice des données, considérée dans son ensemble, est susceptible de fournir les éléments de critique et de synthèse recherchés, mieux que n'importe quelle station prise isolément.

Pour ce faire, il est justifié par l'expérience de formuler l'hypothèse de "quasi-linéarité" des totaux pluviométriques entre stations voisines. La matrice des données peut alors être considérée comme une matrice dégénérée, c'est-à-dire comme le produit matriciel d'un vecteur-colonne (L) – série chronologique d'indices pluviométriques – ou vecteur régional, par un vecteur-ligne (C), constitué des coefficients de proportionnalité de chaque station.

Ces vecteurs sont extraits de la matrice originale au moyen d'un processus itératif de recherche des valeurs normalisées les plus fréquentes, appliqué successivement aux lignes et aux colonnes de cette dernière.

La méthode comprend également une étape facultative de comparaison de chaque station avec les données du vecteur régional ainsi élaboré, ce qui permet éventuellement d'identifier les périodes durant lesquelles la séquence des observations se comporte différemment de la série des indices qui forment le vecteur.

## Caractéristiques de MVR

Dans le but de transformer la méthode décrite succinctement ci-dessus et mise au point sur un équipement lourd, en un outil d'utilisation facile, sans qu'il soit nécessaire de faire appel à des spécialistes, un logiciel a été développé sur micro-ordinateur. Ce logiciel est suffisamment complet et cohérent pour pouvoir être utilisé de manière autonome, c'est-à-dire indépendante d'une quelconque base de données et de l'environnement d'un ordinateur de grande puissance.

A partir d'un lot de données correspondant normalement à la zone que l'on doit étudier (pays, grande région,...), MVR fournit aux utilisateurs les fonctionnalités suivantes :

## Regional vector method

The pluviometry of a region is measured at predetermined time intervals on a network of observation posts. The basic data can therefore be represented by a data matrix where each column represents a place and each line an observation date.

Unfortunately, the collected data, without counting the lack of observations, generally suffers from all sorts of errors. However, in so far as the network of posts is sufficiently dense to guarantee a certain degree of redundancy and supposing that there exists a predominant regional climatic trend, it is reasonable to assume that the data matrix, taken as a whole, can provide elements of criticism and synthesis which we wanted better than any single station.

In order to do this, from experience, it is justified to formulate the hypothesis of "quasi-linearity" of pluviometric totals between neighbouring stations. The data matrix can therefore be considered as a degenerate matrix, that is as a matricial product of a column-vector (L) – chronological series of pluviometric indexes – or regional vector, by a line-vector (C), made up of the proportionality coefficients from each station.

These vectors are extracted from the original matrix by means of an iterative process of research of the most frequent normalized values, successively applied to the lines and columns of this value.

The method also consists of a facultative step of comparing each station with the data from the elaborated regional vector, which may permit the identification of periods during which the observation sequence behaves differently from the series of indexes which form the vector.

## RVM characteristics

With the intent of transforming the method briefly described above and perfected on heavy equipment, into a easy to use tool, without the need for specialists' help, software has been developed on a micro-computer. This software is sufficiently complete and coherent as to be used in an autonomous manner, that is to say independently from any Data Base and from any heavy computer environment.

From a set of data normally corresponding to the zone which has to be studied (country, large region, ...), RVM provides users with the following functions :

– réception d'un lot de données à partir de la lecture de fichiers externes de formats prédéterminés (format standard de la base de données considérée). Cette opération constitue le chargement de la base de données de MVR ;

– émission de divers rapports de caractérisation des informations contenues dans la base (liste des caractéristiques, inventaire, matrice des données...) ;

– définition des régions, rendant possible la division de l'espace d'étude en aires élémentaires ;

– vérification préliminaire des hypothèses de "quasi linéarité" du modèle sur les données de la région élémentaire considérée ;

– création du vecteur régional de la région étudiée, avec visualisation et enregistrement éventuel du résultat ;

– critique des observations par comparaison d'une station avec le vecteur considéré ;

– gestion des corrections avec possibilité d'actualisation et d'émission ;

– extraction des résultats à savoir : vecteur, corrections proposées, observations déjà corrigées.

Quelques fonctions opérationnelles complètent le logiciel, telles que la possibilité de décrire les formats des fichiers externes pour la réception des données, ou d'en créer de nouveaux pour l'extraction des résultats.

## MISE EN ŒUVRE DU LOGICIEL "MVR"

Bien que la plus grande partie des informations pluviométriques nécessaires pour cette étude se trouvent disponibles dans la base de données du SIH, diverses tâches préliminaires ont dû être réalisées afin de les adapter aux propos de ce travail et en particulier de les rendre accessibles au traitement par MVR.

### Partition de l'espace géographique

Dans une première étape, en vue d'une application systématique de la méthode et pour garantir un échantillonnage objectif de l'espace, celui-ci a été divisé en régions "isomorphes". Un réseau d'hexagones réguliers de 12.500 km<sup>2</sup> de superficie, identifiés par un code numérique tenant compte de leur position géographique (Brésil 1986), a ainsi été élaboré.

Toutefois, la densité relativement basse du réseau d'observation en Amazonie et la répartition hétérogène des stations ont exigé leur regroupement en "macro-régions" de trois, sept ou dix-neuf cellules élémentaires, comme illustré sur la *figure 1*.

– reception of a data set from the reading of files which comply with the predetermined format (considered standard format of the Data Base). This operation constitutes the loading of the RVM Data Base ;

– transmission of various characterisation relations of data contained in the base (list of characteristics, inventory, data matrix,...) ;

– definition of regions, making it possible to divide the study space into elementary areas ;

– preliminary validation of the hypotheses of "quasi-linearity" from the model based on the data from the elementary region under consideration ;

– creation of the regional vector of the region under study, with visualisation and possible recording of the result ;

– criticism of the observations by comparing a station with a vector under consideration ;

– management of corrections with the possibility of up-dating and transmission ;

– extraction of results that is to say : vector, proposed corrections, observations already corrected.

Some operational functions complete the software, such as the possibility to describe external file formats for data reception, or to create new ones for the extraction of results.

## RVM DATA PREPARATION

Although the major part of the pluviometric data needed for this study is available in the SIH, various preliminary tasks had to be carried out in order to adapt the data with respect to this study and in particular to make it accessible to RVM processing.

### Partition of the geographical space

During the first stage, with the view to the systematic application of the method and to guarantee an objective sampling of the space, this has been divided into "isomorphic" regions. A network of regular hexagons with a surface area of 12,500 km<sup>2</sup>, identified by a numerical code taking into account their geographical position (Brazil 1986), has been elaborated.

However, the relatively low density of the observation network in the Amazon and the heterogeneous distribution of stations meant that they had to be grouped into "macro-regions" of three, seven or nineteen elementary cells, as is shown in *figure 1*.



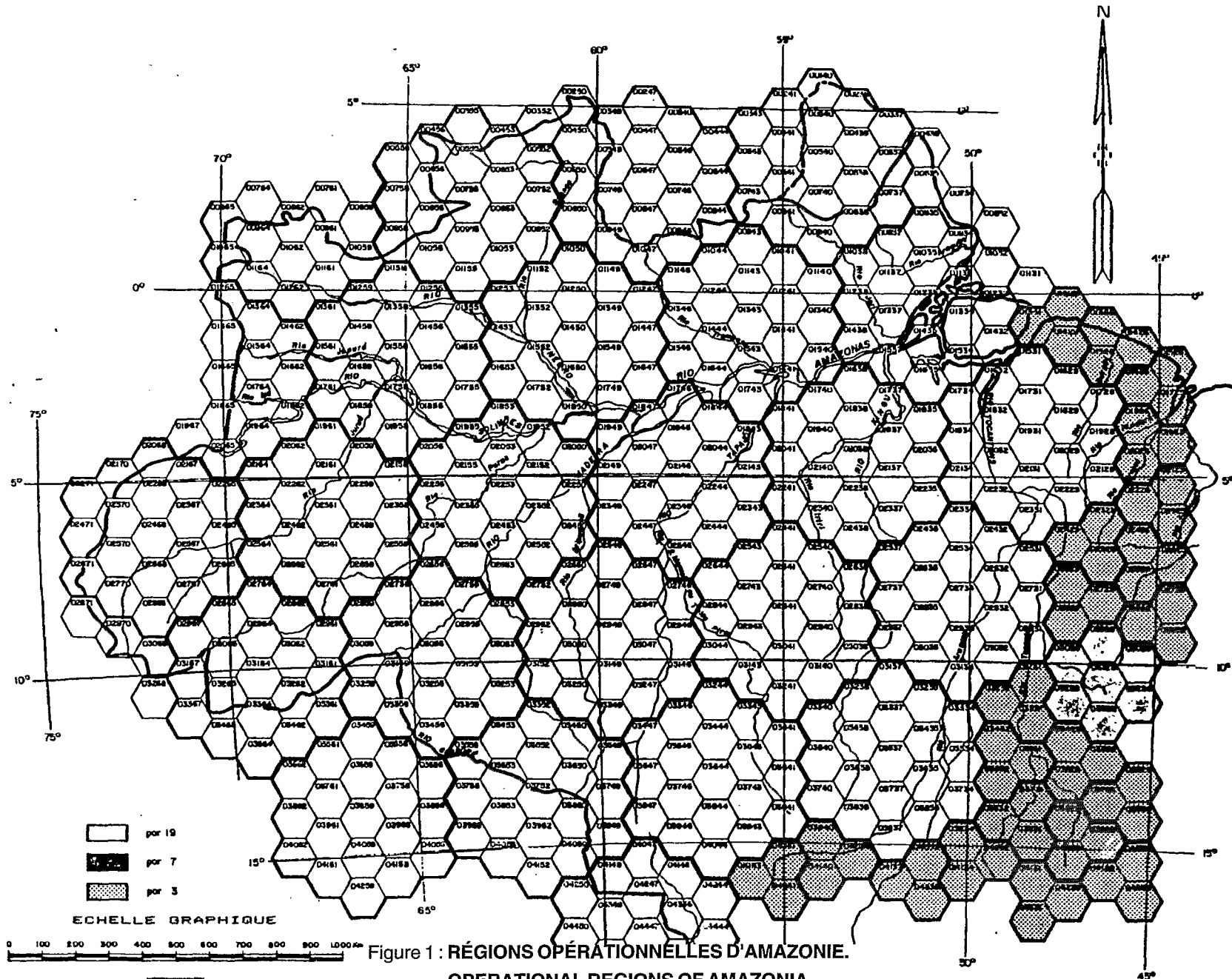


Figure 1 : RÉGIONS OPÉRATIONNELLES D'AMAZONIE.  
 OPERATIONAL REGIONS OF AMAZONIA

## Obtention des fichiers de données

MVR autorise l'entrée des données sous divers formats, pourvu qu'elles soient contenues dans deux fichiers. Le premier définit les caractéristiques des postes (code, nom, coordonnées géographiques, etc). Le second enregistre les données observées proprement dites, c'est-à-dire les totaux pluviométriques de chaque année à chaque station. Les deux fichiers ont été extraits de la banque de données du SIH (sur ordinateur central) et transférés sur des disquettes au format approprié pour micro-ordinateur de type PC, sous système d'exploitation MS-DOS.

Dans une première approche, le traitement a porté sur les données journalières, ce qui a permis d'obtenir une base de 1.496 totaux annuels, pour 290 stations sur la période 1961-1988. Il est apparu que cet ensemble de données était insuffisant, soit par le nombre de stations disponibles (moins de sept par région en moyenne), soit par la période couverte quand on sait qu'il existe quelques postes observés depuis 1910.

Les données historiques ont alors été emmagasinées sous forme mensuelle. Environ 22.000 totaux annuels ont été obtenus pour les 1.248 stations de l'aire d'étude et quelques postes des pays limitrophes comme le Venezuela, la Bolivie et la Guyane française, ceci pour la période 1901-1988.

Diverses opérations ont dû être ensuite réalisées pour obtenir un fichier opérationnel : élimination des stations sans données et des données non identifiées, exclusion des données répétées, correction des informations visiblement erronées (coordonnées géographiques par exemple)... Le complément avec des données obtenues plus récemment (1988 et 1989), permet d'arriver à un total de presque 16.000 stations/années.

La dernière étape de la préparation a finalement consisté à exclure tous les postes qui présentaient moins de trois totaux annuels significatifs. Après cette ultime épuration, 933 stations présentant 10.902 valeurs annuelles utiles ont été retenues pour l'application de MVR.

## Obtention des vecteurs régionaux élémentaires

Le lot de 10.902 observations finalement disponibles pour les 933 stations utilisées dépasse la capacité maximum de traitement simultané d'un lot qui est de 500 stations. Il a donc été nécessaire, à partir de la fonction de réception, de procéder à un découpage, purement technique, en deux zones d'étude. La première, appelée "Amazonie orientale", comprend 24 macro-régions et 474 stations fournissant 4.970 totaux annuels, groupés dans une matrice de 78 lignes (période 1911/1988) et 474 colonnes. La seconde "Amazonie occidentale" regroupe 22 macro-régions et 459 stations, formant une matrice de 88 lignes (1901/1988) et 459 colonnes comprenant 5.932 valeurs.

## Obtainment of data files

RVM permits data input in various forms, provided that it is contained in two files. The first one defines the post characteristics (code, name, geographical coordinates, etc.). The second one records the actual observed data, that is the pluviometric totals for each station. The two files have been extracted from the SIH data bank (on the central computer) and transferred on floppy disks of the appropriate format for PC type microcomputers, using the MS-DOS operating system.

Firstly, the daily data was processed, which enabled a Data Base of 1496 annual totals, for 290 stations over the 1961-1988 period. It became apparent that this set of data was insufficient, either due to the number of available stations (generally less than seven per region), or due to the lengths of the observation periods when one knows that there exists a few posts which have been observed since 1910.

The historical data has therefore been stored in a monthly form. Approximately 22,000 annual totals were obtained for 1248 stations in the area under study and for a few posts in the bordering countries such as Venezuela, Bolivia, French Guiana, for the 1901-1988 period.

Various operations then had to be carried out to obtain an operational file : elimination of stations without data and with unidentified data, exclusion of repeated data, correction of clearly errored data (for example geographical co-ordinates)... The addition of recently obtained data (1988 and 1989) has enabled a total of 16,000 stations/years to be almost reached.

The last step in the preparation finally consisted of excluding all the posts which supplied less than three significant annual totals. After this ultimate purge, 933 stations supplying 10,902 useful annual values were kept for RVM application.

## Obtainment of elementary regional vectors

The set of 10,902 observations which were finally available for the 933 stations exceeded the maximum capacity of simultaneous processing of a lot of 500 stations. It was therefore necessary, from the reception function, to proceed to a partitioning, purely technical, into two study zones. The first zone, named "eastern Amazon", was made up of 24 macro-regions and 474 stations supplying 4,970 annual totals, grouped into a 78-line (1911/1988 period) and 474-column matrix. The second zone "western Amazon" grouped 22 macro-regions and 459 stations, forming a 88-line (1901/1988 period) and 459-column matrix made up of 5,932 values.



Les deux matrices ont un taux de remplissage voisin de 14 %.

L'étape suivante a consisté à redéfinir les régions dans l'environnement de MVR, simple opération de tri facilitée par les ressources de sélection du logiciel, permettant l'utilisation d'un champ spécifique pour le code de la région.

En conservant les valeurs par défaut des paramètres de traitement, les 46 vecteurs régionaux ont alors été élaborés, imprimés et stockés dans un fichier "vecteur". Cette opération a pris de une à cinq minutes par vecteur sur un PC/AT compatible (microprocesseur 80286 + coprocesseur arithmétique 80287, cadencés à 12 MHz).

Enfin, la dernière étape a consisté à extraire les vecteurs régionaux, par la fonction "extraction des résultats" du logiciel, dans le format défini pour MVR, et à les joindre en un fichier unique.

### **Commentaires sur les vecteurs obtenus**

Pour la plus grande part, les 46 vecteurs ont été générés sur la période allant de 1966 à 1987 (ou 1988), soit sur 22 ou 23 ans. Le plus long couvre 41 ans (région 91235, période 1931-1988) ; les plus petits n'ont pas plus de 10 ans.

Malgré la grande hétérogénéité des données d'une part, et la diversité régionale mise en évidence par la représentation graphique des vecteurs d'autre part, une grande parenté dans l'évolution temporelle des vecteurs voisins a été constatée tout au long du traitement. Par exemple, dans la partie la plus orientale de la zone d'étude, on observe clairement que les années 1974 et 1987 sont systématiquement excédentaires.

A des fins de comparaison visuelle des vecteurs élémentaires régionaux entre eux, divers vecteurs de portée géographique plus grande (vecteurs de vecteurs) ont été élaborés, assimilant alors chaque vecteur régional élémentaire à une station. De tels groupements sont issus de l'application de l'analyse multivariée (voir plus loin).

### **ANALYSE DES RÉSULTATS**

Deux directions ont été suivies pour analyser les produits obtenus. La première conduit au tracé des isohyètes sur l'ensemble de la région avec des données homogénéisées. La seconde, qui s'appuie sur les techniques d'analyse multivariée, autorise une interprétation des tendances climatiques (méthode des composantes principales).

The two matrix had a filling ratio around 14 %.

The next step consisted of re-defining the regions in the RVM environment, simple sorting facilitated by software resource selection, permitting the use of a specific field for the region code.

By keeping the values by default of processing parameters, the 46 regional vectors were therefore elaborated, printed and stocked in a "vector" file. This operation took from one to five minutes per vector on a compatible PC/AT (80286 microprocessor + 80287 arithmetic co-processor, cadence at 12 MHz).

Lastly, the final step consisted of extracting regional vectors, by the "result extraction" function of the software, in the defined format for RVM, and to join them together into a single file.

### **Comments on the obtained vectors**

Most of the 46 vectors were created over the period going from 1966 to 1987 (or 1988), that is over a period of 22 or 23 years. The longest covers 41 years (91235 region, 1931-1988 period) ; the shortest has no more than 10 years.

In spite of the great heterogeneity of the data on the one hand, and on the other hand the regional diversity emphasized by the graphical representation of the vectors, a great resemblance in the temporal evolution of neighbouring vectors has been noted during all the processing. For example, in the most eastern part of the zone under study, one can clearly see that 1974 and 1987 systematically showed excess rainfall.

With the aim of visually comparing regional elementary vectors with each other, various vectors of a greater geographical range (vectors of vectors) have been worked out, assimilating each regional vector to a station. Such groups stem from the application or multivariate analysis (see further on).

### **ANALYSIS OF THE RESULTS**

The products were analysed in two different ways. The first method involved the plotting of the isohyets for the whole region using homogenized data. The second, was based on multivariate analysis techniques, permitting an interpretation of climatic trends (principal component analysis).

## A) Tracé des isohyètes

L'estimation de la moyenne pluviométrique de chaque poste du bassin de l'Amazonie, sur une période commune, est obtenue en multipliant la valeur moyenne des indices annuels constituant le vecteur de la sous-région considérée par le coefficient propre à chaque station. La valeur synthétique obtenue est la meilleure selon le modèle.

La dimension de l'aire étudiée est telle que pour rendre les résultats exploitables, d'un point de vue planimétrique, il s'est avéré indispensable d'utiliser un système de projection qui conserve les superficies pour transformer les coordonnées géographiques en coordonnées cartésiennes. La projection azimutale, dont l'origine est située au point 5°S et 60°W, a été choisie pour effectuer la transformation. Elle a été appliquée aux coordonnées des 933 stations ainsi qu'aux points d'application virtuels des vecteurs. 13 stations ont dû être exclues faute d'observation sur la période correspondante du vecteur (coefficient de station indéterminé).

La *figure 2*, élaborée à l'aide d'un logiciel graphique (SURFER), représente le réseau d'isohyètes obtenu d'après cette méthode. En abscisse et en ordonnées ont été portées les distances en unités de projection ; parallèles et méridiens sont gradués de 5 en 5 degrés. Les stations n'y sont pas représentées, mais les vecteurs le sont par un triangle surmonté de leur numéro d'ordre. Les isohyètes couvrent la période 1972-1987 (16 ans). Leur examen permet de vérifier la variabilité spatiale de la pluviométrie annuelle dans la région amazonienne : de 1.400 mm (Roraima, sud du Mato Grosso et à l'est du rio Xingu) à 3.400 mm et même 3.600 mm sur le littoral nord et sur l'Amazonie colombienne.

La comparaison de ces résultats avec ceux des travaux antérieurs met en évidence une grande ressemblance dans les tracés, corroborant l'intérêt de l'automatisation inhérente à la méthode du vecteur régional.

## B) Analyse multivariée

Le MVR a fourni un champ de *vecteurs régionaux* dont chacun caractérise l'évolution dans le temps de l'indice des précipitations régionales. Il a paru intéressant de soumettre ce champ à l'analyse en composantes principales pour en extraire les facteurs communs éventuels et permettre, si possible, une représentation spatiale de l'influence de ces facteurs dans les différentes régions. La méthode a été appliquée à la matrice formée par les indices des 46 vecteurs régionaux sur 45 années de période commune, mais non continue ; après extraction, on a pu tracer les courbes d'isovaleurs des coefficients de corrélation linéaires (ou coefficients régionaux) de chaque vecteur régional avec chacune des principales composantes.

## A) Plotting of the isohyets

The estimation of the average pluviometry of each post in the Amazonian Basin, over a known period, was obtained by multiplying the average value of annual indexes constituting the sub-region vector considered by the coefficient of each own station. The synthetic value obtained is the best one according to the model used.

The dimension of the area studied is such that in order to render the results usable, from a planimetric point of view, it check spelling proved to be indispensable to use a projection system which preserved the areas in order to transform the geographical co-ordinates into cartesian co-ordinates. The azimuthal projection, whose origin is situated at the 5°S and 60°W point, has been chosen to carry out the transformation. It was applied to the co-ordinates of 933 stations as well as the virtual application points of the vectors. 13 stations had to be excluded due to a gap in the observations over the corresponding period of the vector (indetermined station coefficient).

*Figure 2*, worked out with the help of graphical software (SURFER), represents the isohyetal network obtained using this method. The distances in projection units were put on the X and Y axes ; parallels and meridians were graduated 5 by 5 degrees. The stations are not shown, but the vectors are shown by a triangle with their order number above it. The isohyets cover the 1972-1987 (16 years) period. Their examination enabled the spatial variability of the annual pluviometry in the Amazonian region to be verified : from 1400 mm (Roraima, southern Mato Grosso and to the east of River Xingu) at 3400 mm and even 3600 mm over the northern littoral and over the Colombian Amazon.

The comparison of these results with previous studies shows a great similarity in the plotting, emphasizing the interest of automatisation inherent in the regional vector method.

## B) Multivariate analysis

The RVM provides a field of *regional vectors* ; each one characterizes the evolution in time of the regional precipitation index. It appeared to be interesting to submit this field to principal components analysis to extract possible common factors and to permit, if possible, a spatial representation of the influence of these factors in the different regions. The method was applied to the matrix formed by the indexes from 46 regional vectors over a 45 year common but not continuous period ; after extraction, one could plot the linear correlation isovalue curves (or regional coefficients) for each regional vector, with each one of the principal components.

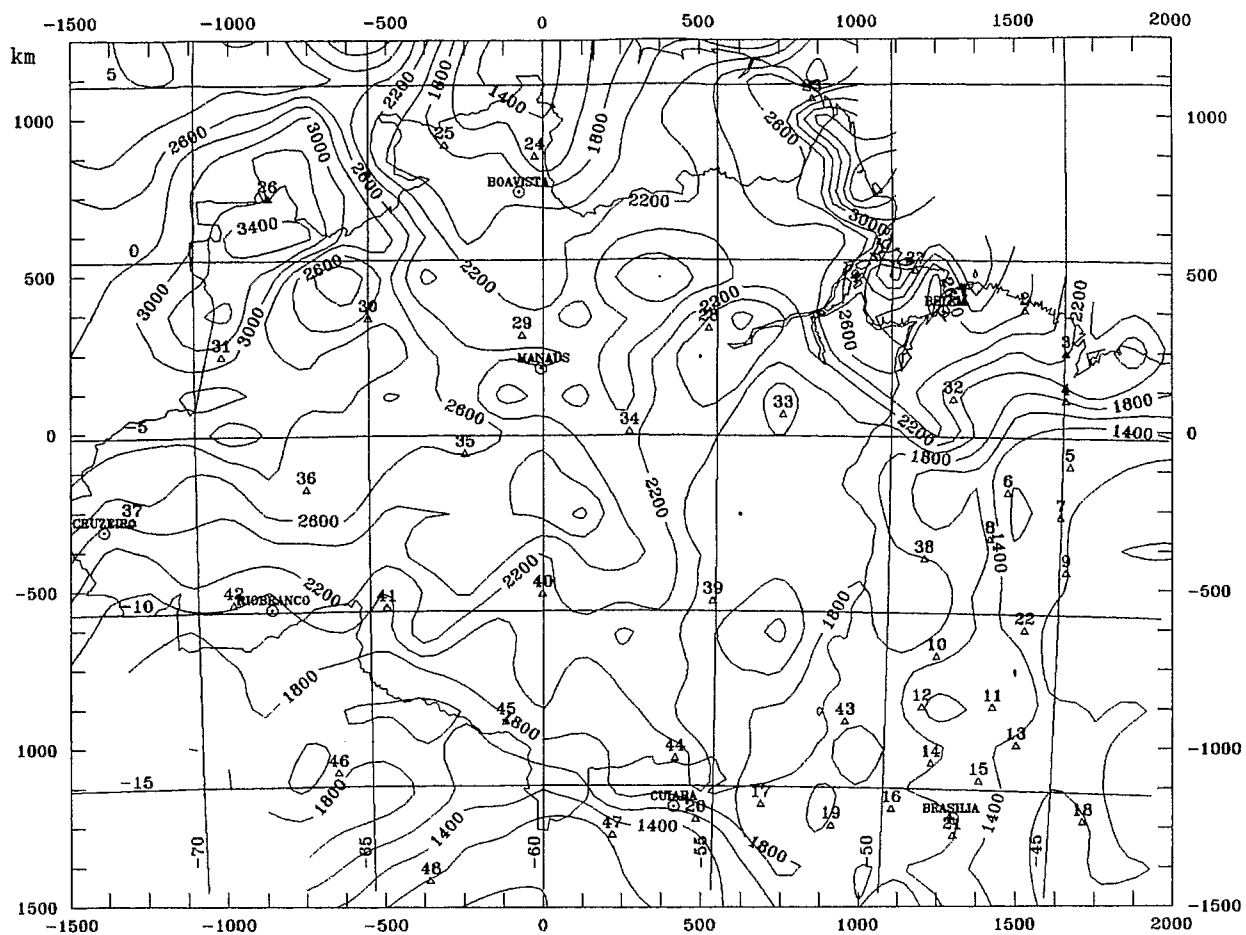


Figure 2 : ISOHYÈTES ANNUELLES EN AMAZONIE (Période : 1972-1987).  
 ANNUAL ISOHYETALS IN AMAZONIA (Period : 1972-1987).

### Préparation de la matrice de travail

La matrice ainsi formée comporte malheureusement 52 % de valeurs manquantes alors que la méthode exige de travailler sur une matrice complète. Il a donc fallu prendre les mesures suivantes :

- limiter la dimension de la matrice en éliminant les années les plus anciennes pour lesquelles les vecteurs élémentaires contiennent peu d'information significative. Cela a conduit à ne garder que la période 1972-1987, soit 16 années. Malgré tout, la matrice comporte 48 lacunes sur 736 données, soit un déchet de 6,5 % ;

- éliminer les lacunes restantes par estimation des valeurs manquantes. Pour cela, dans chaque région élémentaire qui comportait une ou plusieurs lacunes, on a élaboré un nouveau vecteur à partir des valeurs du vecteur de cette région, considérées comme les valeurs d'une station ordinaire, et de celles de tous les vecteurs des régions directement adjacentes. La valeur manquante a alors été remplacée par le produit de l'indice correspondant du nouveau vecteur et du "coefficient de station" du vecteur de la région considérée. On a ainsi limité au maximum les risques d'une estimation biaisée, mais au prix de la génération de 34 vecteurs auxiliaires.

### Preparation of the study matrix

The matrix which was formed unfortunately included 52 % of the missing values but this method has to be used on a complete matrix. Therefore the following measures had to be taken :

- a limitation of the matrix size by eliminating the most ancient years for which the elementary vectors contained little significant data. This led to only the 1972-1987 period being kept, that is 16 years. In spite of everything, the matrix still had 48 gaps in the 736 data, that is 6.5 % waste ;

- the elimination of the remaining gaps by estimating the missing values. For this, in each elementary region which had one or several gaps, we worked out a new vector from the values of the vector of this region, esteemed to be the values from an ordinary station, and from those of all the vectors in the directly adjacent regions. Therefore the missing value was replaced by the product of the corresponding value of the new vector and the "station coefficient" of the vector from the region under study. We have therefore limited to a maximum the risks of a biased estimation, but at the price of generating 34 auxiliary vectors.

### Application de l'analyse en composantes principales

C'est sur cette matrice complète qu'a été appliquée la méthode des composantes principales (programmation G. Hiez). Elle a fourni :

- les valeurs moyennes de chaque vecteur (utilisées pour l'estimation des moyennes à chaque station sur la période commune, à partir desquelles ont été tracées les isohyètes), ainsi que leur variance ;
- les valeurs et vecteurs propres de la matrice ;
- les coefficients régionaux.

La première composante est montrée sur la figure 3 avec le vecteur qui lui est le plus fortement

### Application of PCA

The Principal Component Analysis Application method was applied to this complete matrix. It provided (programmation G. Hiez) : Principal Component Analysis.

- the average values of each vector (used for the estimation of the averages of each station over the common period of time, from which the isohyets had been plotted), as well as their variance ;
- the eigen values and eigen vectors of the matrix ;
- the regional coefficients.

The first component is shown on figure 3 with the vector with which it is the most strongly correlated ;

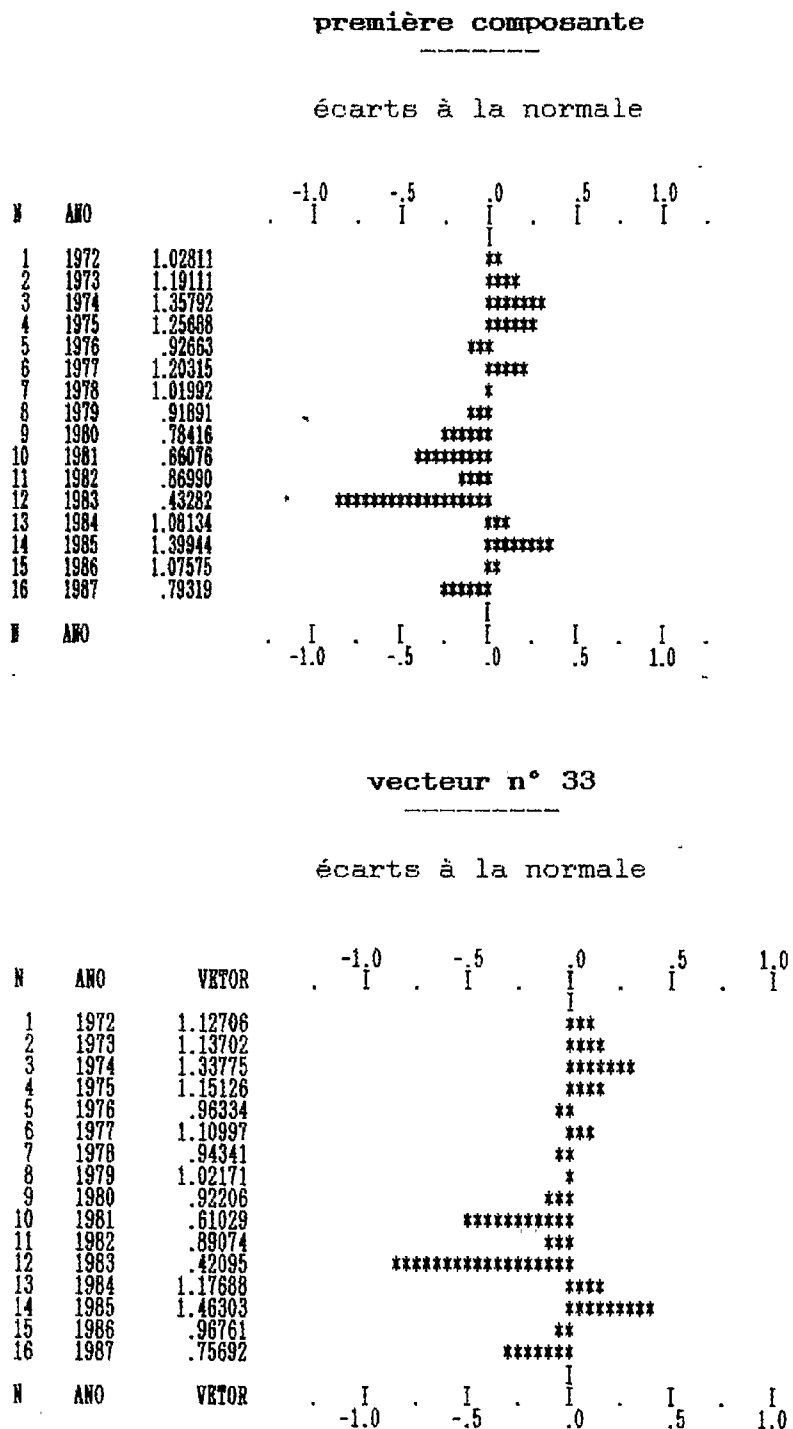


Figure 3

corrélé ; dans la troisième colonne sont portées respectivement la valeur de la composante en fonction de l'année rapportée à la moyenne 1 et la valeur de l'indice annuel correspondant. Les graphiques sont ceux des écarts logarithmiques (logarithmes des valeurs).

Elle n'explique, seule, que 30,4 % de la variance, tandis que la deuxième, en *figure 4*, en explique encore 22,5 %. Il faudrait 11 composantes pour arriver à 95 % ! Cette dispersion est probablement due à la diversité des situations climatiques liée à la grande extension de l'aire d'étude, mais aussi, en partie peut-être, à la faible extension dans le temps des séries d'indices.

the value of the component as a function of the year in relation to the average 1 and the value of the corresponding annual index are respectively shown in the third column. The graphs are those with a logarithmic deviation (logarithms of values).

The first component only explains, by itself, 30.4 % of the variance, whereas the second component, in *figure 4*, still only explains 22.5 %. Therefore, 11 components will have to be used to reach 95 % ! This dispersion is probably due to the diversity of climatical situations linked to the great extension of the area studied, but perhaps also partly to the short time span of the series of indexes.

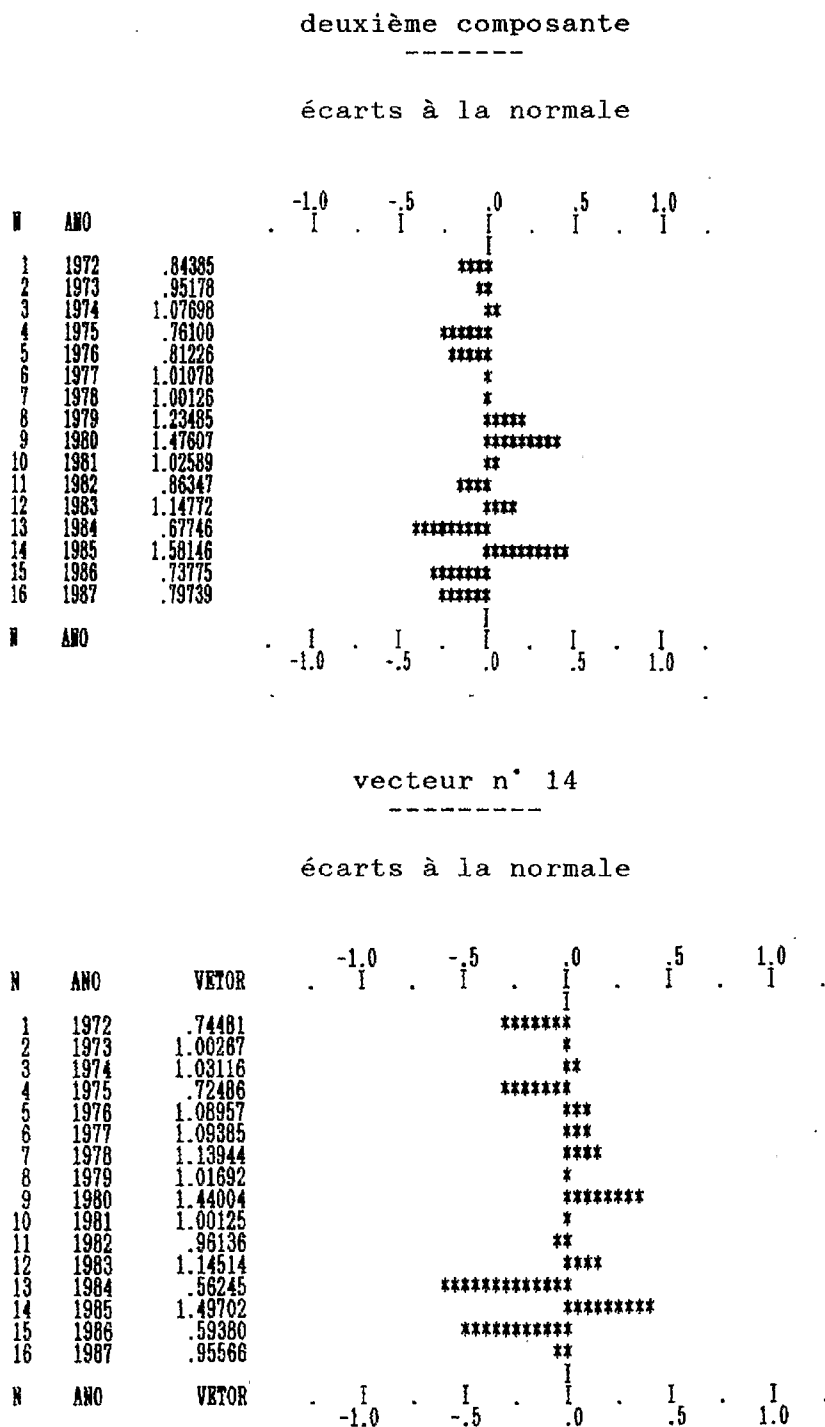


Figure 4

Ces résultats ont été directement exploités, avec le même logiciel graphique, pour visualiser les réseaux d'isovaleurs des coefficients de corrélation, ou coefficients régionaux, de chacun des vecteurs avec ces deux premières composantes. Ils sont représentés sur la *figure 5*. Comme précédemment, la projection isoplane azimutale a été utilisée, sur la partie brésilienne du bassin de l'Amazone, dont le contour a été tracé.

Les deux cartes présentent une régularité notable. Il est vrai que, pour ce tracé, seul le réseau des 46 vecteurs est pris en compte et non pas les 933 stations qui ont servi de base au tracé des isohyètes. Par elle-même, cette régularité renforce l'intérêt d'une telle méthode pour la régionalisation.

Sur le premier graphique, sur la partie supérieure de la figure, apparaît un véritable couloir, orienté ENE-WSW, et balisé par la courbe d'isovaleur 0.5. De plus de 1.000 km de largeur, il s'étend sur plus de 3.000 km de l'Atlantique à la frontière Brésil/Pérou. Dans cette bande, les coefficients régionaux, tous supérieurs à 0.5, croissent de la partie occidentale à la partie orientale, où ils peuvent atteindre 0.94. Ces observations ne laissent aucune doute quant à l'importance de l'influence océanique révélée clairement par la première composante. De plus, ce couloir correspond très sensiblement à la position moyenne de la Zone InterTropicale de Convergence (Nimer 1979).

Une deuxième constatation importante concerne la forte corrélation négative observée sur le versant sud-est du bassin entre Cuiaba et le fleuve São Francisco, signe d'une inversion climatique par rapport à la partie principale du bassin : l'isovaleur zéro forme une véritable frontière climatique. Le même type de phénomène, moins accusé toutefois, se rencontre dans la partie nord-ouest du bassin.

L'examen de la carte correspondant à la deuxième composante, montre en premier lieu que les isolignes négatives ont une tendance à l'orientation nord-sud, formant un couloir entre deux isolignes zéro. Par contre, la partie basse du bassin n'est pratiquement pas soumise à l'effet de cette deuxième composante. Il reste une zone à forte corrélation positive centrée à peu près sur la région dite du "Planalto" brésilien, au nord de Brasília, et deux autres zones faiblement positives respectivement au nord du bassin, sur les contreforts du Roraima, et au sud-ouest, sur le piémont andin en Bolivie.

De toute évidence, cette deuxième composante est d'interprétation plus difficile : elle reflète peut-être une influence continentale qui aurait son origine dans la circulation atmosphérique venue de l'ouest, tandis que le grand couloir d'anomalies négatives serait la manifestation de la remontée des masses d'air polaire venant du sud par le "Gran chaco". Il serait souhaitable que des spécialistes en climatologie se penchent avec attention sur ces premiers résultats pour en faire une interprétation plus affinée.

These results have been directly used, with a graph software, to visualize the isovalue networks of the correlation coefficients, or regional coefficients, of each vectors with these two first components. They are shown on *figure 5*. Like before, the azimuthal projection has been used, over the Brazilian part the Amazonian Basin, whose contour has been plotted.

The two charts show a notable regularity. It is true that, for this plotting, only the network of the 46 vectors has been taken into consideration and not the 933 stations which were used as a base for the plotting of the isohyets. By itself, this regularity emphasizes the interest of such a method for the regionalization.

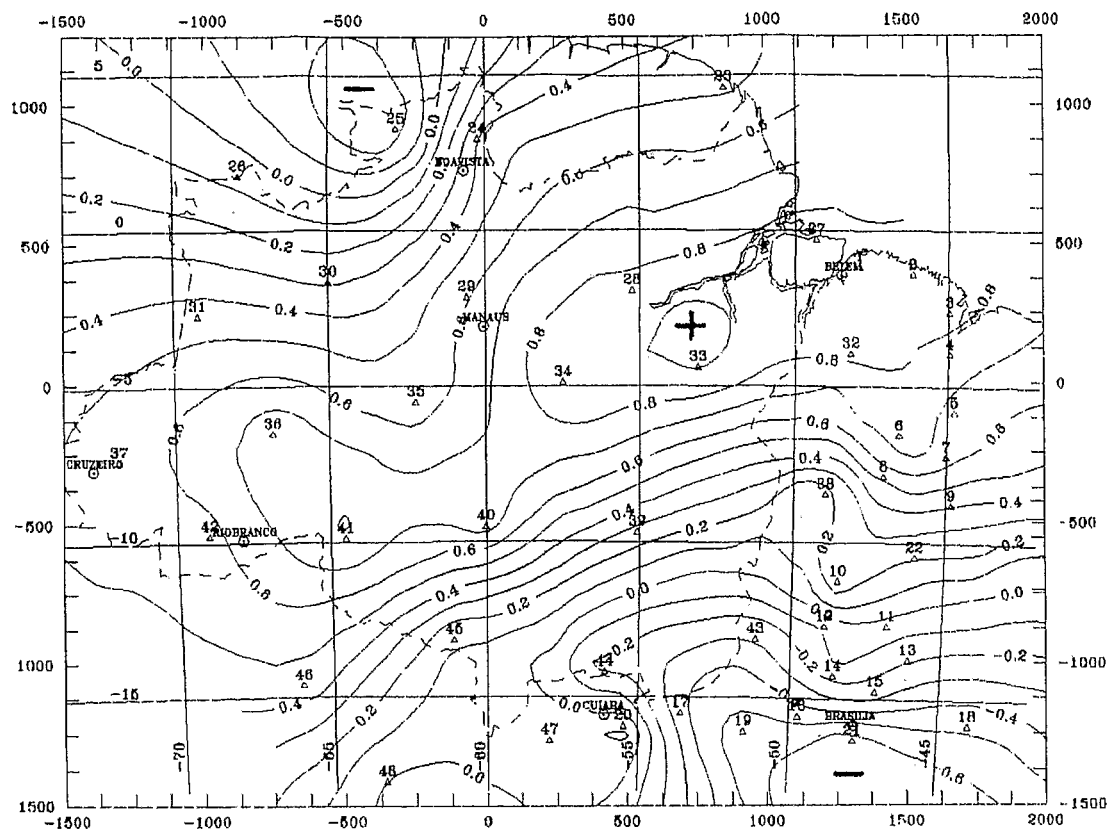
On the first chart, over the top part of the figure, appeared a true corridor, orientated ENE-WSW, and marked by the 0.5 isovalue curve. More than 1000 km in width, it covered more than 3000 km of the Atlantic at the Brazil/Peru border. In this band, the regional coefficients, all more than 0.5, crossed from the western side to the eastern side, where they could reach 0.94. These observations do not leave any doubt as to the importance of the oceanic influence, clearly revealed by the first component. Moreover, this corridor very clearly corresponds to the average position of the InterTropical Convergence Zone (Nimer, 1979).

A second important finding concerns the strong negative correlation observed over the south-eastern slope of the Basin, between Cuiaba and the River San Francisco, a sign of a climatic reversal in relation to the principal part of the basin : the zero isovalue forms a real climatic border. The same type of phenomenon, but however less marked, is found in the north-west part of the Basin.

The examination of the chart corresponding to the second component firstly shows that the negative isolines have a tendency to be orientated north-south, forming a corridor between two zero isolines. On the other hand, the lower part of the Basin is almost not subjected to the effect of this second component. There remains a strong positive correlation zone centred approximately over the region said to be the Brazilian "Planalto", in the north of Brasília, and two other zones slightly positive in the north of the Basin, over the lesser chain buffering the mountain range of Roraima, and over the south-west, over the Andes piedmont in Bolivia respectively.

Evidently, this second component is more difficult to interpret, it could reflect the continental influence which would have originated in the atmospheric circulation coming from the west, whereas the large corridor of negative anomalies would be the manifestation of the rise in polar air masses coming from the south by the "Gran Chaco". It is desirable that climatology specialists carefully study these first results in order to make a more detailed interpretation.

a) Première composante



b) Deuxième composante

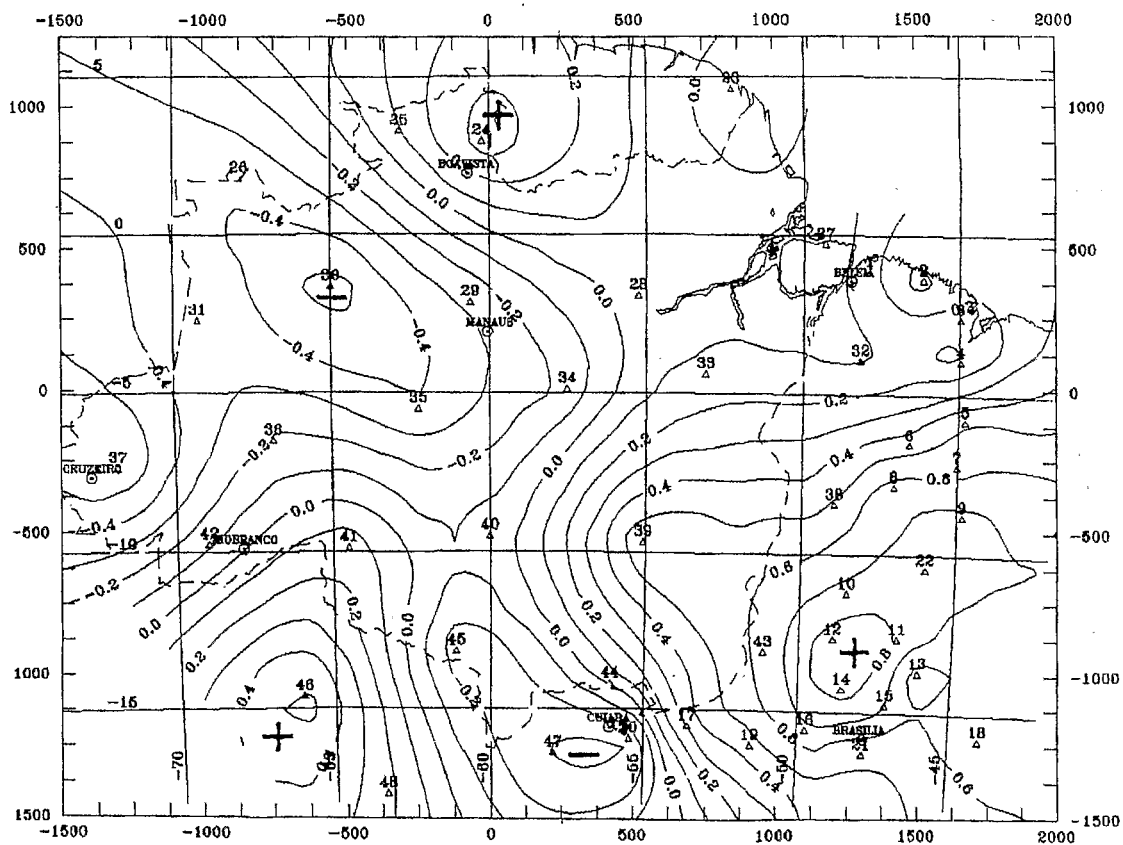


Figure 5 : CORRÉLATION DES VECTEURS RÉGIONAUX AVEC LES DEUX PREMIÈRES COMPOSANTES.  
CORRELATION OF REGIONAL VECTORS WITH THE TWO FIRST COMPONENTS.



## CONCLUSION

Le bilan positif résultant de ce travail est essentiellement dû à l'utilisation de la méthode du vecteur régional. Fondée, on le rappelle, sur l'extraction de l'information la plus fréquente – donc la plus représentative – des événements pluviométriques intervenus dans une région, elle utilise la totalité des données disponibles observées dans cette région.

En raison de la grande étendue de la zone étudiée (plus de mille postes pluviométriques), l'application systématique de la méthode n'a été rendue possible que grâce au développement d'un logiciel approprié autorisant la manipulation rapide d'un grand volume d'informations.

La combinaison des deux méthodes – Vecteur Régional et Composantes principales – offre un nouveau potentiel pour l'interprétation de grandes masses de données pluviométriques, même hétérogènes, et sur une grande extension géographique.

La disponibilité d'informations hydrologiques en banques de données permettrait d'envisager une systématisation de ce type d'étude à d'autres régions du Brésil, voire à d'autres parties du monde, en particulier dans la bande intertropicale.

## CONCLUSION

The positive evaluation resulting from this study is mainly due to the use of the regional vector method. Based, it must be remembered, on the extraction of the most frequent data – therefore on the most representative –, it uses all the available data from the observations in this region.

Due to the vast study zone (more than 1000 pluviometric posts) the systematic application of the method was only possible thanks to the development of an appropriate software permitting the rapid handling of a great quantity of data.

The combining of two methods – Regional Vector and Principal Component – brings a new potential for the interpretation of great masses of pluviometric data, even heterogenous, and over a wide geographical area.

The availability of hydrological data from data banks would enable a systematization of this type of study to be envisaged in other parts of Brazil, even in other parts of the world, in particular in the intertropical zone.

## RÉFÉRENCES

- DNAEE, 1992. – SIH, Subsistema de dados mensais : projeto. CPD/CAEEBB, Brasília.
- DNAEE/Divisão de Controle de Recursos Hídricos, 1983. – Vetor Regional. Informativo Técnico n° 3, 34 p., Brasília.
- DNAEE/Divisão de Controle de Recursos Hídricos, 1986. – O Vetor Regional e as flutuações climáticas. Informativo Técnico n° 7, 16 p., Brasília.
- HIEZ G., 1977. – L'homogénéisation des données pluviométriques. Cahiers Orstom, série hydrol., 14(2), pp. 129-172, Paris.
- HIEZ G. et RANCAN L., 1983. – Aplicação do método do vetor regional no Brasil. III Simposio brasileiro de hidrologia e recursos hídricos, 5, Anais vol. 3, pp. 205-227, Blumenau.
- NIMER E., 1979. – Climatologia do Brasil, IBGE, 422 p., Rio de Janeiro.
- SUDAM/Projeto de Hidrologia e Climatologia da Amazônia, 1984. – Atlas climatológico da Amazônia Brasileira. Publicação 39, Belem.
- HIEZ G., COCHONNEAU G., SECHET P., FERNANDES U. M., 1991. – Aplicação do método do vetor regional a análise da pluviometria anual da bacia amazônica. IX Simpósio brasileiro de hidrologia e recursos hídricos, Rio de Janeiro, Anais vol. I, pp. 367-377.
- COCHONNEAU G., HIEZ G., SECHET P. – MVR, Logiciel pour la critique, l'homogénéisation et la synthèse d'observations pluviométriques. Coll. LogOrstom, 1 vol. + 3 disq. (en prép.).
- DONGUY P. et LASNIER J., 1972. – Application des composantes principales à un champ pluviométrique. Mémoire de fin de stage, Bureau de l'Eau.

