

MVR - AUTOMATISATION D'UN MODELE D'INFORMATION PLUVIOMETRIQUE REGIONALE

Gérard COCHONNEAU, Gérard HIEZ, Patrick SÉCHET

I RESUME

La méthode du vecteur régional modélise l'information pluviométrique régionale (ensemble de séries chronologiques d'observations collectées en des stations voisines), sous la forme d'une suite d'indices représentatifs de la pluviométrie annuelle sur la région, tout en permettant de quantifier la contribution relative de chaque poste. Elle s'appuie sur un processus itératif convergent de recherche des valeurs modales des distributions expérimentales des valeurs observées, des indices représentatifs et des coefficients de proportionnalité spécifiques de chaque station.

La méthode a été mise au point dans un double souci de critique (les observations sont traditionnellement entachées d'erreurs), mais aussi de synthèse d'information. Le *vecteur régional* obtenu présente en effet la tendance pluviométrique de la zone considérée sous une forme compacte et facile à prendre en compte dans les études d'inventaire ou de mécanismes climatiques de grande extension spatiale.

Un progiciel d'automatisation permettant la génération du modèle est en cours de réalisation, de façon à en rendre possible l'exploitation systématique à l'aval des grandes bases de données pluviométriques

(PLUVIOM, en particulier). L'utilisation d'une méthodologie professionnelle de développement de logiciel tend à rendre celui-ci convivial et très souple, de mode à disséminer l'utilisation du modèle et surtout d'en étendre considérablement le champ d'application. Des applications deviendront donc possibles en climatologie, hydrologie, océanographie, géophysique, etc..

II INTRODUCTION

La connaissance des précipitations pluviométriques, de leur évolution dans le temps et de leur répartition dans l'espace, est essentielle pour de nombreuses activités de l'homme dans son milieu (production agricole, alimentation en eau des concentrations urbaines et industrielles, protection contre les crues, etc.).

La variabilité qui caractérise les phénomènes atmosphériques, oblige à en appréhender l'intensité par le biais d'un réseau d'observations ponctuelles, qui fournit un échantillonnage spatio-temporel. Ainsi, lorsque l'on individualise une région pour une étude de la pluviométrie, l'information de base collectée peut être représentée par une matrice de données (A) où chaque colonne représente un point d'observation (station ou poste) et chaque ligne une date d'observation déterminée (figure 01).

Malheureusement, comme pour toute acquisition de données par sondage, les données recueillies sont entachées d'erreurs de toute nature, qui tiennent autant au caractère individuel de la mesure qu'aux défauts qui peuvent intervenir dans le système (site, capteur, lecture ou enregistrement, transmission, etc.) de collecte de l'information. De plus, toute absence d'observation en une station et à une date déterminée provoque une lacune dans la matrice des données.

Dans la mesure où le réseau d'observation considéré est suffisamment dense pour fournir un certain degré de redondance dans les informations recueillies, non seulement une critique systématique des données est possible (et nécessaire), mais encore il est envisageable de présenter les mesures de la pluviométrie régionale sous une forme plus synthétique et plus facile à prendre en compte dans les applications.

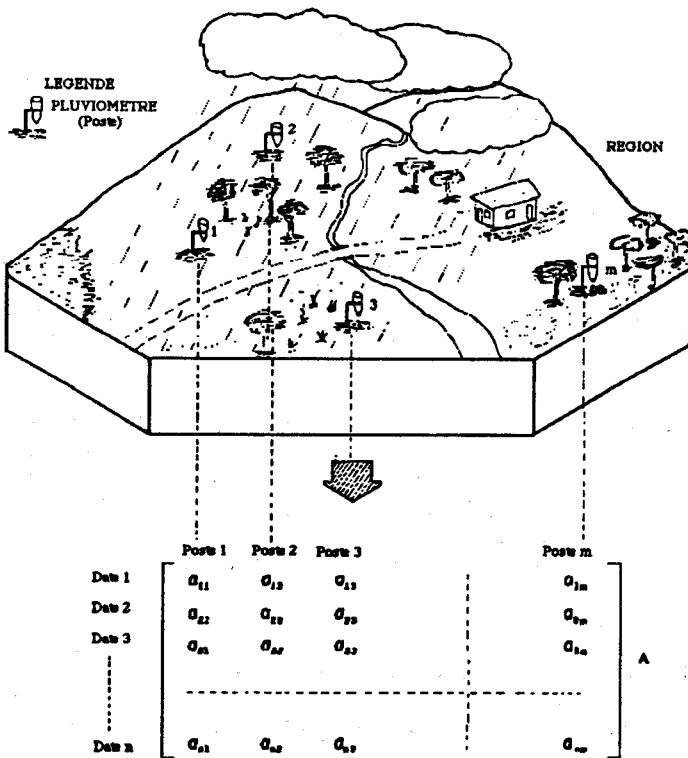


Figure 1 - Pour une région déterminée, les observations pluviométriques collectées à chaque date et en chaque poste fournissent la matrice initiale A

C'est dans ce double objectif que Gérard Hiez, hydrologue à l'ORSTOM, a développé un modèle conceptuel : la méthode du vecteur régional (Gérard Hiez 1977).

III. HYPOTHESES

L'hypothèse fondamentale consiste à admettre que l'ensemble des stations sélectionnées obéit à une tendance climatique régionale unique, chacune d'elles n'étant par conséquent qu'une représentation de cette tendance, déformée à la fois par de possibles anomalies locales et par d'éventuelles erreurs d'observation. C'est le principe de *pseudo-proportionnalité* des totaux pluviométriques entre postes voisins.

Dans ces conditions, on conçoit que l'information pluviométrique pour une date i déterminée, qui est fournie globalement par l'ensemble des stations, contient une valeur estimative L_i plus représentative de celle donnée par l'une quelconque des stations, prise isolément.

IV. LE MODELE

La méthode du vecteur régional (MVR) fournit une estimation du vecteur de référence formé par l'ensemble des indices pluviométriques L_i . Sur la matrice des données, le principe de la *pseudo-proportionnalité* des postes se traduit par la *quasi-colinéarité* des vecteur-colonnes qui la constituent.

Autrement dit, et toujours en faisant abstraction des fluctuations aléatoires propres à chaque poste, la matrice initiale est *peu différente* d'une matrice dégénérée, que l'on peut considérer comme le produit d'un vecteur-colonne L par un vecteur-ligne C (figure 02). Il est clair que les coefficients C_j sont propres à chaque station, tandis que les L_i sont les indices pluviométriques (inhérents à chaque date) recherchés.

1) Modèle : $A = B + E$, où

A est la matrice des données observées;

B est la matrice des données "théoriques" ;

E est la matrice de erreurs (ou anomalies)

2) Application du principe de "pseudo-proportionnalité"

$$B = L \times C$$

ou

$$\begin{array}{|c} L_1 \\ L_2 \\ \vdots \\ L_1 \\ \vdots \\ L_n \end{array} * \begin{bmatrix} C_1 & C_2 & \dots & C_1 & \dots & C_n \end{bmatrix} = B$$

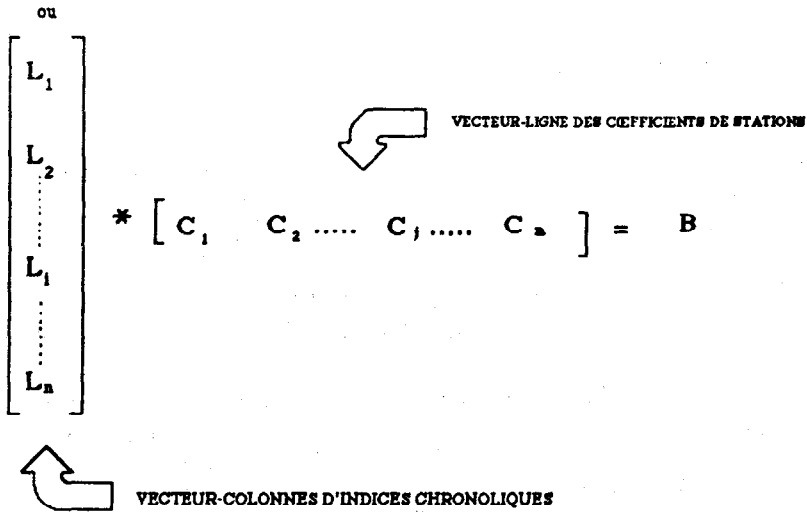


Figure 2- La matrice initiale (A) est "peu différente" (aux anomalies près) d'une matrice dégénérée (B), obtenue à partir du produit d'un vecteur colonne (L) par un vecteur-ligne (C)

Le modèle théorique constitué par la matrice dégénérée, résultat du produit $L \cdot C$, représente l'information à caractère linéaire qu'il est possible d'extraire des données de base. La différence entre la matrice initiale et la matrice dégénérée est appelée matrice des résidus E ($E = A - L \cdot C$).

V. CALAGE DU MODELE

La détermination des coefficients C_j , $j=1..m$ et des indices L_i , $i=1..n$ correspond au calage du modèle proposé : puisqu'il s'agit d'un système de $n \cdot m$ équations (correspondant à chaque élément de la matrice initiale) à $n+m$ inconnues, celui-ci doit être réalisé en cherchant à minimiser la matrice des résidus.

Plutôt que de réduire quantitativement la masse des résidus, en introduisant le critère de distance défini à partir de la norme euclidienne (méthode des moindres carrés), MVR choisit d'en réduire le nombre. Ceci revient à dire que les vecteurs L et C sont correctement estimés lorsque la valeur la plus probable des résidus est nulle.

Il s'avère en fait plus intéressant d'utiliser les valeurs relatives des résidus, de telle sorte que la valeur la plus fréquente de celles-ci sera zéro,

si la valeur modale de $\frac{A_{ij}}{L_i \cdot C_j}$ est l'"unité", ce que l'on écrira $M(\frac{A_{ij}}{L_i \cdot C_j}) = 1$. Ainsi, en fixant une colonne (j) de la matrice A, on peut écrire :

$\frac{M(\frac{A_{ij}}{L_i})}{C_j} = 1$, soit $C_j = M(\frac{A_{ij}}{L_i})$, étant donné que $M(\frac{1}{C_j}) = \frac{1}{M(C_j)} = \frac{1}{C_j}$. De même, pour toute ligne (i) de la matrice : $L_i = M(\frac{A_{ij}}{C_j})$.

Autrement dit, si l'on suppose les C_j connus, il suffit de normaliser la suite des valeurs $\frac{A_{ij}}{C_j}$ pour obtenir L_i et, inversement, si l'on suppose les L_i déjà calculés, le mode de l'échantillon des $\frac{A_{ij}}{L_i}$ fournira C_j . Ainsi, si l'on dispose d'un moyen de calcul du mode, une solution peut être trouvée par un processus itératif : on initialisera (par exemple) tous les L_i à 1 pour calculer une première approximation des C_j , puis on calculera les L_i à l'aide de ces valeurs et on utilisera les nouveaux L_i pour commencer une deuxième itération.

L'auteur ayant mis au point une méthode non paramétrique de détermination de la valeur modale d'un échantillon quelconque, la démarche d'élaboration du vecteur régional repose finalement sur un procédé itératif d'estimation du mode des lignes et des colonnes de la matrice initiale (*traitement LC*). Le processus se poursuit tant que la variation des L_i et C_j , d'une itération à l'autre, est supérieure à un seuil prédéterminé.

VI VALIDATION DU MODELE

Il est facile d'éprouver la fiabilité et la sensibilité de l'algorithme mis en oeuvre. En partant d'une matrice dégénérée artificielle dans laquelle on introduit des anomalies locales, on peut constater que le procédé permet de retrouver, sans aucune erreur, le vecteur-ligne et le vecteur-colonne d'origine. Tel test, réalisé sur une matrice 9x5, est entièrement décrit dans l'article de 1977.

Par ailleurs, un exemple portant sur une série composée de 7 postes, permet de montrer que si l'on applique une déviation systématique sur trois d'entre eux, le vecteur régional reste inchangé et les graphiques de double-cumul effectués entre le vecteur et ces mêmes postes mettent en évidence une *cassure* pendant la période affectée par la déviation. Si l'on ajoute la même déviation sur un quatrième poste, le vecteur régional *bascule* et les cassures observées se transportent sur les trois postes inchangés.

Ce dernier test montre bien la sensibilité de la méthode : il en fournit aussi les limites, dans la mesure où il indique que les résultats obtenus pourront être systématiquement faussés si, par exemple, les postes sélectionnés n'obéissent pas tous à la même tendance climatique (mauvais choix du groupement régional, ne permettant pas de respecter l'hypothèse fondamentale).

VII. MISE EN OEUVRE

La méthode du vecteur régional, comme outil d'analyse ponctuelle (critique des données observées en chaque station) et comme outil de synthèse, prend toute sa valeur si elle est appliquée sur de grandes régions, pour l'homogénéisation de la pluviométrie annuelle (par exemple). La rendre disponible à l'aval des bases de données pluviométriques existantes, est donc un préalable à une utilisation systématique.

La phase d'opérationnalisation du modèle a franchi un certain nombre d'étapes, qui se traduisent à chaque fois par la mise au point d'un, ou plusieurs, programmes adaptés à l'environnement informatique de la structure d'accueil.

Un premier programme, expérimental, a été développé en 1973 sur IBM 370-168, pour le bureau central hydrologique à Paris. A la suite de l'affectation de G.Hiez au Brésil, deux versions ont été adaptées, l'un sur IBM 1130 à l'université de Fortaleza, l'autre sur IBM 360-30 (sous DOS) à la Sudene (superintendance pour le développement du nordeste) à Recife. L'une comme l'autre ont dû migrer, respectivement pour un environnement DEC 10 à Fortaleza, et sur IBM 370-145 (sous DOS-VSE) à Recife.

Ces deux dernières versions ont donné lieu à l'amélioration des algorithmes et des sorties imprimées d'une part, à des applications de grande envergure, pour le tracé des isohyètes de l'état du Cearà et l'homogénéisation des données de l'état de la Paraíba (Jacon 1982), d'autre part.

Depuis 1983, une adaptation est en cours de réalisation au département national (brésilien) des eaux et de l'énergie électrique (DNAEE), administration chargée de centraliser les données hydropluviométriques collectées sur le territoire brésilien. Celle-ci, intégrée au système national d'informations hydrologiques (SIH) qui catalogue plus de 10 000 postes, permettrait une exploitation de la méthode à grande échelle.

Un travail important a déjà été réalisé dans le but d'appliquer systématiquement la méthode : les grandes lignes du travail à réaliser ont été tracées et les produits attendus ont été précisés (Brasil 1983). En particulier, on a défini un échantillonnage objectif de l'espace en régions iso-

morphes basé sur un réseau de mailles hexagonales de 12 500 km² de superficie (elles-mêmes parfois regroupées en macrorégions, ou au contraire subdivisées en microrégions, de façon à prendre en compte les grandes variations du nombre de stations par région). Il a ainsi été possible de fournir quelques résultats intéressants sur les fluctuations climatiques, au niveau de grandes régions (Brasil 1986)

VIII. PROBLEMES RENCONTRES

L'implantation se heurte toutefois à un certain nombre de difficultés qui, jusqu'à présent, n'ont pu être totalement surmontées. Au-delà des contraintes entraînées par l'adaptation à une nouvelle configuration matérielle (IBM 3081 et 4381) et du système d'exploitation (VM/CMS sur MVS), gênantes mais rapidement contournables, la mise en place effective affronte deux obstacles importants : difficulté d'interfaçage avec le SIH, géré sous ADABAS et incapacité du site de traitement à assimiler les programmes développés en Fortran, dans un environnement axé sur l'utilisation du Cobol et du PL/1, et d'outils plus spécifiques (SAS, Natural, etc.).

Plusieurs tentatives infructueuses d'adaptation partielle, ou de développement d'un système fondé sur les programmes développés par l'auteur du modèle, ont conduit à proposer un schéma de travail fondamentalement différent, susceptible d'aboutir à une version totalement opérationnelle.

IX. SOLUTION

L'idée est venue de constituer une équipe de développement pour assister l'auteur de la méthode à la mise au point d'un logiciel, suffisamment complet et cohérent pour être utilisé de manière autonome, indépendamment de l'environnement informatique de la base de données. Le travail à réaliser prend alors une toute autre dimension, dans laquelle la génération du vecteur régional se réduit à une fonction parmi d'autres, mais occupe toutefois une position centrale.

La finalité de cette nouvelle opération est évidemment de valoriser les résultats obtenus par un chercheur, tout au long d'une partie significative de sa carrière, en transformant la méthode mise au point en un outil facilement utilisable et pouvant être mis à profit par plusieurs équipes. Dans la mesure où le modèle sous-jacent faisait déjà l'objet d'un large consensus auprès de la communauté hydrologique, ce projet a été facilement approuvé.

X CONCEPTION DU LOGICIEL

Un dossier de conception détaillé a été établi (Séchet & Hiez, 1988), selon une méthode d'analyse (Gane & Sarson, 1979) déjà éprouvée et largement utilisée par les analystes de l'équipe dans d'autres programmes.

Le diagramme de flot de données du système à développer, présenté figure 3, en montre les principales fonctionnalités :

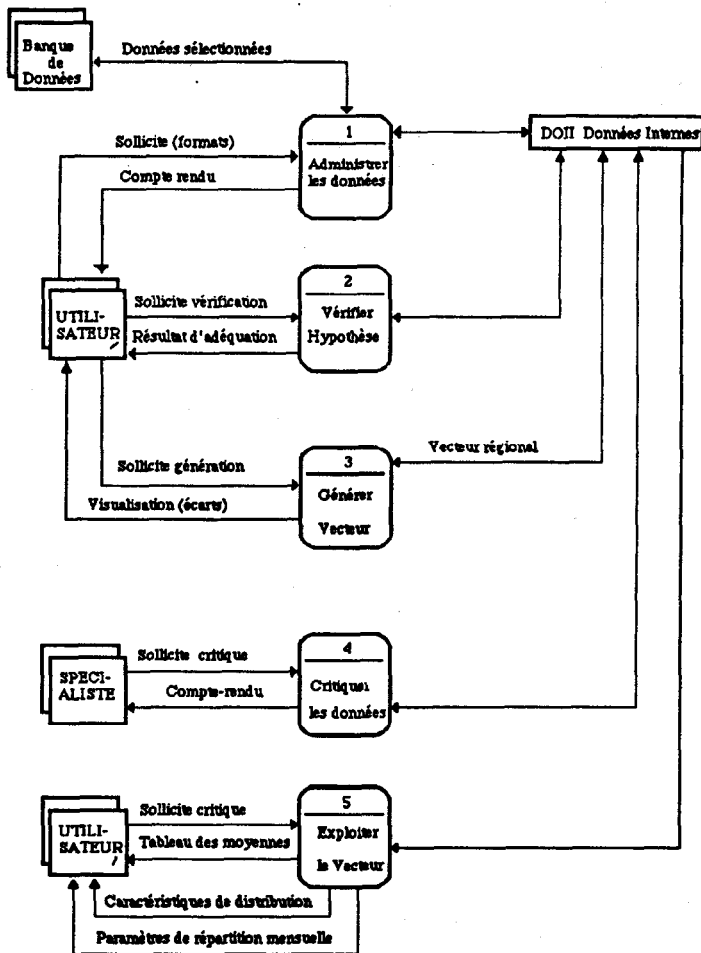


Figure 3 - Le diagramme de flux de données présente les fonctionnalités du logiciel MVR

- la fonction *administrer les données* prévoit l'entrée d'un lot de données dans le système MVR, soit par saisie directe, soit par lecture dans un fichier de format prédéfini, et toute mise à jour qui s'avérerait nécessaire. Elle permettra de la même façon l'extraction des résultats et l'émission automatisée d'un certain nombre de documents de travail. Elle inclut également un processus fondamental de sélection, dans le lot initial, de la région et de la période à traiter (fournissant par conséquent la matrice de données A) ;

- *vérifier les hypothèses* est un processus qui s'appuie sur des tests opérés sur la matrice d'autocorrélation des séries sélectionnées pour déterminer si l'hypothèse de *pseudo-proportionnalité* des stations est respectée. Dans le cas contraire, il interdit l'utilisation de la méthode sur l'échantillon choisi ;

- la fonction *générer le vecteur* constitue le noyau central du système. Elle prévoit le calcul du vecteur régional, tant au niveau annuel que mensuel, et autorise la visualisation graphique des écarts cumulés (ou simples cumuls), méthode utilisée pour mettre en évidence les fluctuations pluviométriques régionales. Elle fournira également une matrice de *données reconstituées*, obtenue par estimation de la valeur de la pluie en chaque poste, à partir du vecteur régional et du coefficient C_j propre à chaque poste ;

- les deux dernières fonctions permettront la mise en oeuvre des applications classiques du vecteur régional : critique des données observées à une station et production d'un certain nombre de paramètres régionaux, par exploitation du pouvoir de synthèse du vecteur régional.

Chaque fonction fait l'objet d'une décomposition en processus, chacun d'entre eux dûment documentés, à la fois en ce qui concerne la logique à codifier et les flux d'entrée et de sortie qui interviennent. Par exemple, la figure 4 montre le détail de la critique des données et fait apparaître un processus interactif, le spécialiste (hydrologue) s'appuyant sur un tracé graphique de doubles cumuls entre le vecteur régional et un poste déterminé, pour définir les corrections à appliquer sur ce dernier.

4. CRITIQUES DES DONNEES

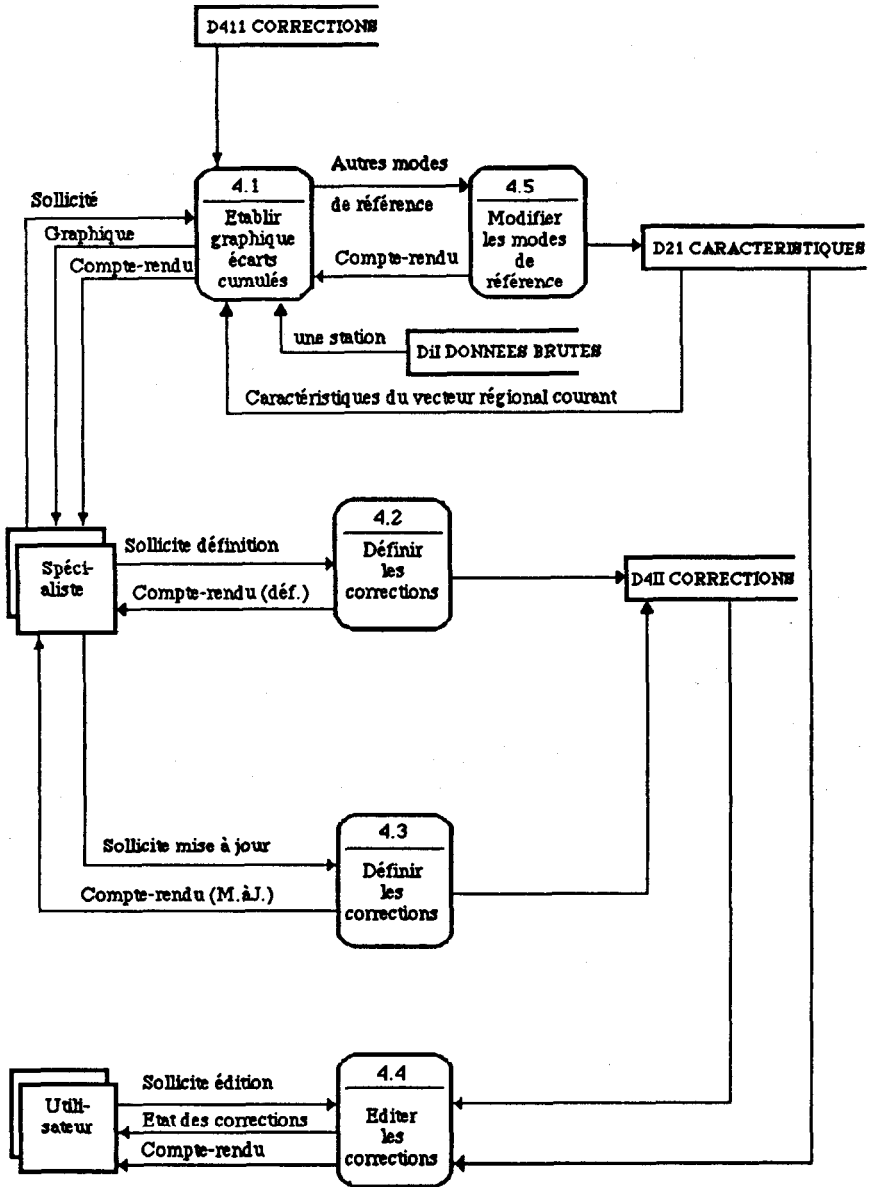


Figure4 - La correction des données, poste par poste, fait appel à une fonction interactive

XII. PROGRAMMATION DU LOGICIEL

Le dossier de conception élaboré ne compromet aucun élément de la définition du logiciel par rapport à une configuration particulière, de telle sorte qu'il a été décidé de développer une première version sur un micro-ordinateur au standard IBM PC-XT, permettant d'en percevoir une ample dissémination. D'ailleurs, une version expérimentale rudimentaire, développée par Gérard Hiez courant 1987, en a largement montré l'intérêt et la faisabilité, même si l'addition d'un coprocesseur arithmétique apparaît comme quasiment indispensable pour rendre les temps de traitement acceptables.

La programmation proprement dite de MVR est réalisée en deux étapes. Dans une première phase, dite de spécification, le choix de toutes les structures de données est arrêté, ainsi que toutes les modalités de l'interface système-utilisateur (écrans, formats des sorties imprimées, documentation d'utilisation et aide en ligne) et les programmes sont définis, dans un environnement d'atelier-logiciel. La codification est alors réalisable par plusieurs programmeurs différents, la cohérence de l'ensemble étant garantie, à la fois par les outils communs et par l'utilisation du dictionnaire de variables produit au cours de l'étape de conception. La conclusion de la première version opérationnelle est prévue pour fin 1988.

XII. CONCLUSION

Bien qu'il soit prématuré de compter dès maintenant sur un produit encore en cours de mise au point, les bénéfices que l'on peut tirer, d'ores et déjà, de cette expérience ne sont pas négligeables.

En premier lieu, il convient de remarquer que les principes de base de la méthode ont été énoncés il y a plus d'une quinzaine d'années. Les années suivantes ont été surtout mises à profit pour perfectionner les algorithmes et diffuser la méthodologie auprès de la communauté scientifique : même si les quelques adeptes de la première heure ont à leur actif des applications intéressantes, il faut bien reconnaître que l'on a pas exploité tout le potentiel de MVR. En ce sens, la production d'un outil logiciel, facilitant la mise en oeuvre et l'exploitation de la méthode, crée une expectative prometteuse de diffusion et d'assimilation de la technologie correspondante.

Une importance toute particulière revient au travail de conception effectué, qui aura été l'occasion d'une totale remise en question des aspects liés à la mise en oeuvre de la méthode. A l'occasion de l'analyse fonctionnelle correspondante, des modules autonomes ont été définis, ainsi que l'interfaçage de cette application avec les bases de données, en précisant

clairement un protocole de communication et les fonctions propres à l'une et l'autre. De plus, il a été jugé nécessaire de mieux distinguer certaines fonctions, dont l'accès sera réservé au spécialiste, par exemple en ce qui concerne l'application des corrections suggérées par la méthode.

Cette étape s'est accompagnée d'une réflexion sur les possibilités de généralisation de MVR, susceptible d'être utilisé dans plusieurs domaines scientifiques, où l'on manipule des ensembles d'observations spatio-temporelles de phénomènes continus (climatologie, océanographie, géophysique, etc.). L'intégration de la fonction de vérification des hypothèses, d'une part, la possibilité de saisie directe de la matrice initiale d'autre part, vont dans ce sens.

On notera pour conclure que cette opération s'insère parfaitement dans une politique de valorisation des acquis scientifiques de l'Institut, en venant couronner un travail de longue haleine mené individuellement par un chercheur, par la réalisation, avec l'aide d'ingénieurs informaticiens, d'un logiciel susceptible de devenir l'un des grands produits de l'ORSTOM dans ce domaine.

XIII. BIBLIOGRAPHIE

Brasil, DNAEE. Divisão de Controle de Recursos Hídricos. Vektor regional. Brasília, 1983. 34p. (Informativo técnico,3).

Brasil, DNAEE. Divisão de Controle de Recursos Hídricos. O vetor regional e as flutuações climáticas. Brasília, 1986. 16p. (Informativo técnico,7).

GANE, C. & SARSON, T. Structured system analysis: tools and techniques. Prentice-Hall, Inc., New York, 1979.

HIEZ, G. L'homogénéité des données pluviométriques. In: Cahiers ORSTOM, sér. hydrol. Paris, 14(2), pp.129-72, 1977.

HIEZ, G. Bases théoriques du "vecteur régional". Les premières applications et leur mise en oeuvre informatique. In: Premières journées hydrologiques de l'ORSTOM à Montpellier, Coll. & Sém., ORSTOM, Paris, 1986.

HIEZ, G. & RANCAN, L. Aplicação do método do vetor regional no Brasil. In: Simpósio brasileiro de hidrologia recursos hídricos, 5, Blumenau, 1983. Trabalhos apresentados. USP/EESC, Blumenau, SC, pp.242-66.

JACCON, G. As precipitações anuais da região paraibana: homogeneização e análise regional. Recife, SUDENE/DRN, 1982, 97p.

SÉCHET, P & HIEZ, G. Méthode du vecteur régional. Dossier de conception détaillé. Brasília, 1988, 140p., diff.rest.