

COLLOQUE SUR L'UTILISATION
DES TECHNIQUES INFORMATIQUES ET DE L'AUTOMATION
DANS LES SYSTEMES D'EAU

UN SYSTEME DE STOCKAGE INFORMATIQUE DES DONNEES HYDROLOGIQUES
par Jacques SIRCOULON.

1. Généralités.

Le stockage des données hydrologiques sur un support permettant leur traitement par ordinateur est de plus en plus largement employé dans un nombre croissant de pays et devient un mode d'utilisation presque classique.

La gestion d'une banque de données par l'informatique est très séduisante, mais pour qu'elle soit vraiment efficace, il ne faut pas perdre de vue que l'exploitation optimale des fichiers dépend non seulement de la puissance de l'ordinateur disponible, mais également de la façon dont le stockage a été conçu.

Les qualités demandées à un fichier sont la simplicité, la souplesse et la fiabilité. Elles doivent se traduire par une grande facilité de manipulations des données :

- insertion aisée de toute nouvelle information,
- correction de données anciennes déjà stockées,
- extraction ou recherche rapide de la valeur désirée,

et un volume de stockage minimal.

La mise au point d'un classement rationnel et opérationnel nécessite donc une préparation minutieuse des objectifs à atteindre, toute modification non prévue dans le système pouvant avoir des conséquences fâcheuses par les perturbations qu'elle entraîne.

Parmi les nombreux systèmes de classement, plus ou moins compliqués existant actuellement, nous présenterons un système de stockage à la fois simple et élaboré établi par le Service Hydrologique de l'ORSTOM. Cet organisme exploite une masse considérable d'informations provenant de la gestion d'une vingtaine de réseaux hydrométriques nationaux, africains pour la plupart.

71385

24 JUL 1971
ORSTOM Fonds Documentaire
N° : 33537, 001
Cote : B

2. Etablissement du fichier.

Quelle que soit la nature matérielle du support qui sera utilisé : cartes perforées, bandes magnétiques ou disques, il est nécessaire d'associer sans ambiguïté, à la valeur de la donnée à stocker :

- sa nature (type); définition du phénomène et de la variable,
- son lieu de provenance,
- sa position dans le temps,
- sa qualité éventuellement.

. Le type de donnée n'est pas toujours représenté explicitement par un chiffre; ainsi à l'ORSTOM, où l'on utilise comme support primaire la carte perforée pour ses facilités de manipulation, cette représentation est fournie implicitement par un modèle de cartes appropriée à l'information : hauteurs d'eau brutes, jaugeages, renseignements sur l'échelle, etc. Des contrôles de lecture dans les programmes de traitement rendent inutiles une numérotation préalable de chaque type de donnée.

. L'opération d'identification du lieu de provenance, soit ici la station hydrologique, est beaucoup plus délicate et fait intervenir ce qu'on appelle improprement le codage.

2.1. Le codage des stations hydrologiques.

En se plaçant logiquement à l'échelle du continent, l'identification d'une station hydrologique exige l'adoption d'un système tenant compte à la fois de la division politique de ce continent et de sa structure géographique.

Le codage retenu par l'ORSTOM permet de définir une station par une variable de 8 chiffres, pouvant se décomposer en 4 variables de 2 chiffres qui sont : ETAT, BASSIN, RIVIERE et STATION.

Ainsi en prenant l'exemple du continent africain :

. le code ETAT attribue un numéro de 01 à 49 aux 49 états existants, classés par ordre alphabétique (ALGERIE à ZAMBIE),

. le code BASSIN donne un numéro aux bassins hydrographiques de chaque état, ces bassins étant également classés par ordre alphabétique en laissant des créneaux suffisants pour le traitement futur des bassins non encore couverts par le réseau. On réserve les 29 premiers numéros aux bassins supérieurs à 70.000 km² susceptibles de s'étendre sur plusieurs pays, afin de ne pas leur affecter un numéro différent suivant l'état où ils se trouvent, soit par exemple : BASSIN : 18 pour le Nil et BASSIN : 29 pour le Zambeze.

Dans le cas des îles occupées par un seul état, la précaution des numéros réservés devient inutile, mais il est bon de distinguer entre les divers bassins côtiers suivant leur importance.

. le code RIVIERE est un reflet de la structure du réseau hydrographique. On réserve les numéros 01 à 09 au fleuve principal, celui-ci pouvant changer de nom le long de son cours. Les affluents sont classés par ordre alphabétique et une ventilation est introduite suivant l'ordre de chaque cours d'eau, soit en définitive :

n° 10 à 29	Affluent de 1er ordre ou petit fleuve côtier,
n° 30 à 59	Affluent de 2ème ordre,
n° 60 à 89	Affluent de 3ème ordre, ou plus,
n° 90 à 99	Lacs, affluents, lits majeurs.

. le code STATION affecte un numéro aux stations situées sur un même cours d'eau après classement par ordre alphabétique et aménagement de créneaux pour les stations futures possibles, mais cette disposition n'est pas impérative.

2.2. Le stockage de l'information brute.

L'information brute est représentée par les chroniques d'observations de hauteurs d'eau. Ces chroniques sont fournies, soit de façon discontinue par les relevés d'échelles limnimétriques, soit de façon continue par les enregistrements des limnigraphes.

. Le but du stockage est d'emmagasiner sous un volume très réduit, la partie de l'information disponible (réseau et bassins représentatifs) permettant une reconstitution aussi fidèle que possible de l'hydrogramme.

. La représentation chiffrée du chronogramme est pratiquée à l'ORSTOM suivant deux systèmes de pas de temps.

2.2.1. Système à pas de temps constant.

Dans ce système, le chronogramme est découpé suivant un nombre donné d'intervalles de temps égaux. Il s'applique principalement aux observations manuelles des stations de réseaux. Avec ce procédé, l'adresse dans le temps de chaque hauteur d'eau est connue implicitement et conventionnellement par la position de la valeur dans la carte.

Le pas de temps constant semble assez rigide, toutefois on peut l'adapter à la vitesse de variation du phénomène mais il doit rester constant au moins pendant une journée.

La hauteur de l'échelle étant généralement donnée au centimètre près, il suffit de quatre chiffres pour exprimer toutes ses valeurs possibles et par convention, les données manquantes sont représentées par 9999 (soit 99,99 m). Dans le cas d'études particulières exigeant la hauteur en millimètre, ce nombre de chiffres reste encore suffisant.

Le repérage dans le temps est fourni sur chaque carte par l'indication de la variable ANNEE et de la variable MOIS et la combinaison de deux autres variables. La variable CAS donne le nombre de lectures journalières et la variable GROUP la position des données de la carte dans le mois.

L'adressage de chaque donnée est ainsi déterminé sans qu'il soit nécessaire de numéroté les cartes.

Pour donner un exemple, la mise sur support des relevés d'une station ayant 2 lectures par jour aboutit à la perforation de 4 cartes par mois.

La variable CAS est égale à 2 et la variable GROUP prend les valeurs :

- GROUP : 1 = 1er au 8ème jour
- 2 = 9ème au 16ème jour
- 3 = 17ème au 24ème jour
- 4 = 25ème jours à la fin du mois.

chaque carte pouvant contenir jusqu'à seize hauteurs d'eau, on pourrait avec ce procédé perforer jusqu'à seize observations par jour. En fait, chaque fois que le nombre d'observations journalières est élevé, il est plus commode, pour avoir une meilleure connaissance des débits, de prendre le système à pas de temps variable. Une combinaison des deux systèmes, permettant de s'adapter à la forme de l'hydrogramme, est aussi prévue (il s'agit alors d'un pas de temps "mixte").

2.2.2. Système à pas de temps variable.

Ce système est surtout utilisé pour les petits bassins et notamment les bassins représentatifs ou expérimentaux. L'adresse dans le temps de chaque valeur est fournie ici par les variables JOUR, HEU (heures), MIN (Minutes) associés à la hauteur HAUT, ce qui porte à 10 le nombre de chiffres nécessaires pour chaque donnée.

Ce procédé ne permet pas de perforer plus de six hauteurs par cartes, mais il permet un choix beaucoup plus judicieux des données significatives en permettant, avec un minimum de points, de suivre parfaitement l'évolution de la hauteur d'eau. Signalons également que sur chaque carte figure une variable JUNITE indiquant si les hauteurs sont en cm ou en mm et que les quatre dernières colonnes sont réservées à la numérotation en séquence (NOCART) afin de permettre un contrôle rigoureux du chronogramme.

3. Le stockage des données traitées.

Il s'agit principalement des débits obtenus après traduction des hauteurs du fichier primaire. Suivant les programmes de traitement employés, les débits sont donnés, soit sous forme de débits instantanés, soit sous forme de moyennes, à une échelle de temps quelconque (débits moyens journaliers, mensuels et annuels). L'amplitude de l'échelle des débits est très grande puisqu'elle peut varier de moins d'un litre par seconde, à plusieurs dizaines de milliers de m³/s et elle pose donc le problème du volume du fichier. Afin de compacter l'information tout en gardant la précision souhaitable, l'ORSTOM a créé un système de notation exponentielle spéciale couvrant toute la gamme des débits possibles.

Cette notation, qui donne habituellement les débits en m^3/s (JUDEBI = 0), comprend seulement 4 chiffres :

3 chiffres significatifs XXX
1 chiffre exposant A

Tout débit représenté sous la forme $Q = 0,XXX 10^A$ s'écrira XXXA avec cette notation.

A titre d'exemple : $Q = 0,001 m^3/s$ donne 0010
 $Q = 0,152 m^3/s$ donne 1520
 $Q = 14,9 m^3/s$ donne 1492
 $Q = 12600 m^3/s$ donne 1265

Dans certaines études de débits d'étiage, où les dispositifs de mesures permettent de définir de très faibles débits, ceux-ci peuvent être donnés en l/s (JUDEBI = 1) en modifiant simplement la valeur du chiffre exposant A qui devient $A + 3$.

Quel que soit le système employé, les débits manquants sont toujours représentés par la valeur -10.