

L'ELABORATION DES FICHIERS DE DONNEES :
TÂCHE D'HYDROLOGUE ou D'ANALYSTE ?

JACCON Gilbert
Directeur de Recherches de l'ORSTOM

INTRODUCTION

C'est dans la seconde moitié de la décennie 1960-70 que de nombreux Services Hydrologiques Nationaux ou Régionaux ont senti la nécessité de faire appel à une nouvelle discipline, l'informatique : soit par besoin d'organiser une volumineuse masse de données d'observation, soit pour moderniser des méthodes de calcul et de gestion, soit pour développer des techniques d'analyse, soit plus simplement pour ne pas "rater un train déjà lancé à vive allure". Il était d'ailleurs difficile en 1970 de résister à l'extraordinaire essor de l'informatique, marqué par l'apparition sans cesse renouvelée d'un matériel séduisant et par l'intense activité d'une multitude de spécialistes.

L'informatisation de l'hydrologie ne s'est pas faite sans douleur pour les hydrologues, très souvent effarouchés par une technique nouvelle au vocabulaire presque exclusivement anglais et incompréhensible pour un non-initié. Selon le cas, les uns ont choisi d'apprendre les langages de programmation puis les procédures de gestion des fichiers, et les autres ont dû accepter un partage des responsabilités avec les informaticiens. Quelle que soit la démarche, des conflits trop fréquents ont surgi entre hydrologues et analystes à l'instant des choix décisifs.

Quels sont ces choix ? Le conflit était-il inévitable ? A qui doit revenir la responsabilité d'établir les fichiers ? A l'hydrologue ? A l'informaticien ?

Cette communication apporte le témoignage, qui se voudrait impartial, d'un hydrologue de l'ORSTOM amené entre 1970 et 1980 à contribuer à la création de bases de données d'abord en ALGERIE, puis au BRESIL.

L'INFORMATISATION DE L'HYDROLOGIE

Les activités hydrologiques classiques se rapportent à deux thèmes principaux :

- l'acquisition et le traitement des données,
- l'exploitation des données observées.

Le premier thème correspond généralement à plus de 75 %, sinon à la totalité, des activités et du budget : installation, gestion et opération des réseaux de mesures, collecte, vérification et analyse des documents, classement, archivage et publication des données, constituent une chaîne de traitement de l'information conduite manuellement depuis plus d'un siècle dès la mise en place des premiers pluviomètres. Ce thème apparaît idéal pour une application de l'informatique qui est, par définition, la science du traitement automatique de l'information.

C'est pourtant par le second thème d'activités que l'hydrologie s'est ouverte à l'informatique : l'analyse hydrologique a trouvé dans l'ordinateur un outil nécessaire à son développement. En quelques années ont été lancés de très nombreux processus, algorithmes, méthodes et modèles mathématiques, généralement disponibles sur le marché et donc facilement accessibles, moyennant finances, à tout Service Hydrologique.

La caractéristique commune à tous ces modules d'analyse est une très grande consommation de données de base. D'où la nécessité urgente de constituer des fichiers de données directement accessibles, donc d'informatiser, c'est-à-dire de rendre automatique la chaîne de traitement de l'information hydrologique.

Mais dans ce domaine les "modules-gadgets" n'existent plus et les transferts de technologie d'un Service à un autre, ou a fortiori d'un pays à un autre, sont très difficiles en raison de divergences entre les méthodes de mesures et l'organisation du traitement.

C'est pour le Service Hydrologique le début de la longue et fastidieuse préparation de sa base de données. C'est aussi la nécessité d'une étroite collaboration entre l'hydrologue et l'informaticien, et d'une stricte répartition des tâches et des compétences.

LA BASE DE DONNEES HYDROLOGIQUES

Pour un Service Hydrologique bien organisé, gestionnaire d'un réseau hydropluviométrique, toutes les données recueillies sont classées, organisées en fichiers (cf. tableau joint), microfilmées pour leur sauvegarde et rendues facilement accessibles à un utilisateur.

Les fichiers s'accroissent régulièrement par l'entrée des observations les plus récentes :

- . certaines vont directement aux fichiers après vérification : par exemple les bulletins de pluviométrie ou les fiches limnimétriques,
- . d'autres doivent être élaborées : par exemple les barèmes de tarage ou les débits.

Créer une base de données c'est mémoriser toute cette information sur un support compatible avec l'ordinateur. Il faut pour cela :

- choisir une organisation générale,
- définir la structure de chaque fichier,
- opter pour une hiérarchisation des données.

Puis mener parallèlement :

- un travail de saisie, d'analyse et de validation des données,
- une tâche de mise au point des chaînes de programmes de création, maintenance, interrogation et exploitation des fichiers (logiciel d'application).

La première action est une *fonction de production* qui incombe essentiellement à l'hydrologue, la seconde est une *fonction de conception méthodologique* qui revient totalement à l'informaticien. Il existe néanmoins une forte interaction d'une fonction sur l'autre et les choix d'organisation, de structure des fichiers ou de procédure d'exploitation ne peuvent être que des choix collectifs.

LES CRITERES DE CHOIX

A - Point de vue de l'Hydrologue

L'hydrologue attend de l'informatique :

- . une amélioration de la gestion de ses fichiers (classement, accès aux données),
- . une réduction des travaux de routine (pointage, vérification des totaux),
- . une possibilité d'automatisation de la critique et de l'homogénéisation des données en vue de la création de fichiers plus sûrs et plus complets (fichiers opérationnels).

Pour cela, trois critères doivent être respectés :

- 1°/ - l'informatisation du traitement des données ne doit *jamaïs* entraîner une perte d'information ni en quantité ni en qualité;
- 2°/ - le logiciel d'application doit être construit pour simplifier au maximum la tâche des hydrologues : il doit par exemple s'adapter aux protocoles de mesures utilisés sur le terrain;
- 3°/ - l'accès aux fichiers doit être facile et possible en toutes circonstances : un fichier ne doit jamais être un placard fermé dont seul l'informaticien possède la clef.

Ceci donne toute compétence à l'hydrologue pour imposer son choix :

- . dans l'organisation générale de la base : par exemple pour le choix du nombre de fichiers, le regroupement éventuel de données de plusieurs types dans un

- même fichier, le découpage en systèmes indépendants (voir tableau);
- . dans la structure des fichiers : par exemple pour le choix du découpage temporel : l'information (pas de temps fixe ou variable, nombre d'observations par jour);
- . dans la hiérarchisation de l'information, presque obligatoirement à base géographique puisque l'hydrologue travaille le plus souvent par bassin versant ou région climatique;
- dans la nature des testes nécessaires pour la critique de la saisie (recherche des erreurs) ou pour la validation des données;
- dans la présentation des listages de données.

B - Point de vue de l'Informaticien

Les choix de l'analyste dans l'organisation de la base de données sont orientés par :

- 1°/ - une recherche de la simplicité générale du logiciel d'application (nombre de programmes minimal),
- 2°/ - une optimisation des performances par :
 - . une condensation de l'information (diminution des "entrées-sorties"),
 - . une structure de fichier adéquate (accès direct pour les fichiers volumineux très utilisés),
 - . le choix du langage de programmation le mieux adapté;
- 3°/ - l'équipement disponible (mémoire centrale, périphériques, compilateurs, moyens de saisie),
- 4°/ - le volume d'information à traiter,
- 5°/ - les exigences de l'hydrologue.

Cette liste n'est pas ordonnée mais il paraît normal de considérer qu'une priorité sera donnée par l'analyste aux critères purement informatiques.

Ces critères donnent toute compétence à l'informaticien :

- . en matière d'équipement : pour le choix du support des fichiers par exemple;
- . dans l'organisation interne : mode d'accès aux données par exemple;
- . dans le langage de programmation utilisé.

Il faut d'ailleurs remarquer que ces choix n'ont aucune influence sur la qualité et la quantité d'information ni sur son libre accès qui sont les deux exigences fondamentales de l'hydrologue. De ces choix ne dépendent que la plus ou moins grande complexité du logiciel d'exploitation et les temps de réponse à toute sollicitation. Un système d'exploitation d'une base de données doit être conçu

dès le départ pour subir périodiquement des modifications de structure, d'organisation ou d'implantation, sans autres inconvénients qu'une courte fermeture des fichiers. Nous ne connaissons pas, pour notre part, un seul Centre de calcul qui n'apporte pas périodiquement des modifications dans son équipement ou son système d'opérations.

ELABORATION DE LA BASE DE DONNEES

Les critères de choix étant définis et les responsabilités partagées entre l'hydrologue, d'une part, et l'analyste, d'autre part, rien ne doit venir en principe troubler l'harmonie nécessaire à la construction de base de données, qui est une tâche difficile et de très longue haleine.

Plus de cinq années de travail acharné auront été nécessaires à une équipe franco-brésilienne d'une quinzaine de personnes, dont trois analystes-programmeurs, pour élaborer une base de données hydrologiques dont les chiffres suivants permettent d'évaluer l'importance :

- . en pluviométrie, les données journalières de 2 295 stations ont été saisies et constituent une information de plus de 20 millions de valeurs (environ 700 000 mois);
- . en hydrométrie, les fichiers de hauteurs d'eau, de jaugeages et de débits contiennent les données se référant à 309 stations limnimétriques et hydrométriques; le seul fichier de débits moyens journaliers contient 33 500 registres mensuels;
- . le logiciel d'exploitation comprend plus de 150 programmes.

Les trois caractéristiques essentielles de cette base de données hydrologiques du NORDESTE du BRESIL, dont la description est faite dans un article à paraître prochainement dans la série des Cahiers HYDROLOGIE de l'ORSTOM (1), sont les suivantes :

- 1°/ - existence de deux niveaux de fichiers (voir tableau) :
- les fichiers du premier niveau contiennent des données originales non corrigées, soit directement mesurées, soit élaborées par un traitement automatique (calcul des débits) ou non (tracé des courbes de tarage);
 - les fichiers du second niveau, dits fichiers opérationnels, résultent des précédents après critique, complètement et homogénéisation des séries par des procédures partiellement automatisées;

(1) *BASE DE DONNEES HYDROLOGIQUES DU NORD-EST BRESILIEN*, par G. JACCÓN et P. SECHET, à paraître dans *Cahiers ORSTOM - Service Hydrologique*, n°3, 1980.

- 2°/ - codification de l'origine (donnée directement observée, ou calculée, ou estimée) et de la qualité (information sûre ou douteuse) de chaque donnée (débit journalier, par exemple) ou ensemble de données (mois pluviométrique, par exemple);
- 3°/ - existence d'un système de manutention, d'interrogation et d'exploitation de la base, totalement exempt de contraintes informatiques telles que l'utilisation d'une perforatrice de cartes ou d'une console : toute opération est mise en oeuvre moyennant une codification alphabétique, transmise au Centre de calcul par une simple sollicitation de service.

Si ces trois caractéristiques constituent toute l'originalité de cette base et lui confèrent un très haut niveau de performance, on ne peut ignorer :

- . d'une part, l'énorme travail que constitue la validation des données : d'abord pour l'élimination totale des erreurs de saisie fort nombreuses, lorsque celle-ci est faite à partir de microfilms, puis pour le passage aux fichiers opérationnels,
- . d'autre part, la difficulté que représente la mise au point du système d'utilisation de la base, tel qu'il a été conçu.

CONCLUSION

La construction d'une base de données ne peut être la tâche d'un homme seul, fût-il aussi bon hydrologue qu'informaticien. C'est une oeuvre collective qui ne peut être menée à son terme que par une équipe parfaitement organisée et confiante dans la réussite de sa tâche.

Aucun choix en informatique n'est totalement décisif pour l'avenir : aussi les conflits ne sont-ils pas nécessaires. Ils suffisent pour cela :

- . que l'informaticien accepte de prendre en considération les exigences de l'hydrologue et surtout son désir de rester le seul responsable du contenu de ses fichiers;
- . que l'hydrologue accorde sa confiance à l'analyste, assurément plus compétent que lui-même en matière d'analyse de systèmes, et réserve son temps disponible à la réflexion hydrologique et à la programmation de nouveaux modèles d'analyse.

BASE DE DONNEES HYDROLOGIQUES : Inventaire des fichiers

NOM DU FICHIER	SYSTEME	CONTENU	ELABORATION	TYPE	REMARQUES	
1. Identification	-	Code, nom, carte géologique des stations	Par saisie directe	Original	Fichier "maître"	
2. Précipitations	PLUVIOMET.	Pluies journalières	Par saisie directe des bulletins	Original		
3. Pluviographie		Pluies partielles / durée variable	Après pré-traitement des pluviogrammes	Original		
4. Pluviométrie		Pluies journalières ou mensuelles suivant le niveau d'homogénéisation	Fichiers 2 + 3 et homogénéisation	<u>OPERATIONNEL</u>		
5. Limnimétrie	HYDROMETRIE	Relevés limnimétriques	Par saisie directe des bulletins	Original		
6. Limnigraphie		Relevés limnigraphiques	Après pré-traitement des limniogrammes	Original		Facultatif
7. Hauteurs d'eau		Relevés limnimétriques + limnigraphiques	Fichiers 5 + 6	Original		Facultatif (si 6 absent)
8. Jaugeages		Résultats du mesurage des débits	Après dépouillement des jaugeages	Original		
9. Tarage		Points caractéristiques ou barèmes	Après tracés de la courbe de tarage	Original		
10. Débits instantanés		Débit instantané pour chaque hauteur d'eau	Fichiers 7 + 9	Original		Facultatif
11. Débits moyens		Débits moyens journaliers	Fichier 7 + intégration	Original		Facultatif
12. Débits moyens		Débits moyens journaliers	+ Homogénéisation	<u>OPERATIONNEL</u>		
13. Climatologie	CLIMATOLO.	Températures, humidité, vent, insolation, évaporation, pluie au sol	Par saisie directe des fiches de relevés	Original		
14. Chimie des eaux	QUALITE DES EAUX	Conductivité, résidu sec, bilan ionique éléments rares	Par saisie directe des fiches d'analyse	Original	} peuvent être associés (FACULTATIF)	
15. Matières en solution		Mesures de salinité totale par conductimétrie	Par saisie directe des fiches d'analyse	Original		
16. Matières en suspension		Poids de matières en suspension	Par saisie directe des fiches d'analyse	Original		
17. Salinité et charge		Salinité moyenne journalière et charge en suspension moyenne journalière	Fichiers 10+15+16 + homogénéisation	Original <u>OPERATIONNEL</u>		