

L'ORGANISATION ET LE CONTROLE DES DONNEES

L'organisation des données

La compréhension des processus intervenant dans le cycle de l'eau ainsi que l'étude de leurs variations spatiales et temporelles nécessitent de disposer de données. Celles-ci sont essentielles et constituent un préalable à toute analyse hydrologique, que ce soit dans le but de procéder à une étude du cycle de l'eau, d'impacts environnementaux ou pour procéder au dimensionnement d'ouvrages hydrauliques.

De façon générale, pour permettre le passage de l'acquisition des données à leur utilisation effective dans le cadre d'une analyse hydrologique on distingue les étapes suivantes :

acquisition, traitement, contrôle et validation, organisation, diffusion et publication.

1- Acquisition des données

L'acquisition de données consiste à procéder, par le biais d'un instrument de mesure, à acquérir de l'information (par exemple : hauteur d'eau d'une station limnimétrique, comptage des basculements d'un pluviographe à augets, vitesse du vent etc...).. Le procédé peut être automatisé ou non. Ceci aura une influence sur le type d'erreurs que l'on peut commettre.

2- Traitement primaire des données

La donnée acquise précédemment nécessite souvent un traitement préalable - ou *traitement primaire* - afin de la rendre pertinente et exploitable. Il s'agit pour l'essentiel de la conversion de la mesure effectuée en une grandeur qui soit hydrologiquement significative (par exemple : transformation d'impulsions électriques en intensités de précipitations, de hauteurs d'eau en débits, génération de données à pas de temps constant à partir de mesures effectuées à pas de temps variables etc.

Le traitement des données inclut aussi le *contrôle primaire* des données qui comprend les contrôles de cohérence à l'exclusion de tous traitements statistiques. Il s'agit par exemple, dans le cas d'une acquisition manuelle des données, de les convertir en fichiers informatiques. Dans ce cas, on procède généralement à une double saisie des données puis les fichiers sont comparés afin de déceler d'éventuelles erreurs de saisie. Dans la situation où l'on procède à l'acquisition de données de précipitations et de débits, on vérifie encore la cohérence temporelles des données acquises, à savoir par exemple qu'une crue est bien la conséquence d'un épisode pluvieux.

3- Contrôle des données

Avant de pouvoir exploiter les données et bien qu'elles soient dans un format adéquat, il importe de contrôler la fiabilité et la précision de ces dernières. Le contrôle permet de valider les données avant leur organisation au sein d'une banque de données pour leur mise à disposition à des fins opérationnelles. Lors de cette opération, on introduit des indices de qualité de la donnée ainsi que des indices indiquant que celle-ci est reconstituée, calculée voire manquante.

4- Organisation des données

Au vu de l'importance quantitative et qualitative des données, il importe de les organiser avec soin. Ceci se fait à partir d'un corpus de documents originaux (formulaires de terrain, diagrammes, unité de stockage électronique) constituant les *archives* qui sont en règle générale accessibles uniquement à un personnel spécifique (responsable du centre de collecte, archiviste...).

La traduction des archives sous la forme de fichiers de base génère les "*fichiers en l'état*" et fournit une indication sur la provenance de la donnée (mesure, calcul, copie etc.) ainsi que sur sa qualité (fiable, complète ou non) et sa précision.

Enfin, on constitue *un fichier de travail* provisoire permettant une visualisation des données et permettant de procéder aux différents tests de qualité et de précision des données qui seront développés tout au long de ce chapitre.

L'élément ultime de cette chaîne opératoire est la constitution des *fichiers opérationnels* avec indices de qualité pouvant être publiés et distribués auprès des utilisateurs.

5- Diffusion et publication

La diffusion des données en vue de leur utilisation opérationnelle peut-être faite selon deux modes :

Publication papier : Il s'agit par exemple de la publication d'annuaires comme l'annuaire hydrologique de l'Algérie ou de bulletins journaliers tels ceux émis par Météo-Algérie accessible par une ligne téléphonique (fax).

Diffusion informatique : Actuellement et grâce aux développements des interfaces liées au réseau Internet, les données sont de plus en plus accessibles de manière conviviale et en tout temps. Le site de Météo-Algérie présente notamment des cartes de situation météorologique.

Le contrôle des données

La constitution d'une série de valeurs, constituant un échantillon au sens statistique du terme, est un processus long, parsemé d'embûches, et au cours duquel de nombreuses [erreurs](#), de nature fort différentes, sont susceptibles d'être commises.

Des erreurs peuvent en effet être perpétrées lors de l'une ou de l'autre des quatre phases du déroulement classique des opérations, à savoir : la mesure ; la transmission de l'information ; le stockage de l'information ; le traitement de l'information (prétraitement et analyse). Il est donc indispensable, avant d'utiliser des séries de données, de se préoccuper de leur qualité et de leur représentativité en utilisant diverses techniques en général de type statistique ou graphiques.

Une erreur de mesure est définie comme étant la différence entre la vraie valeur (qui est l'idéal recherché, mais qui n'est en principe et malheureusement jamais connue) et la valeur mesurée. Il est commode, tant pour les présenter que pour différencier la façon de les aborder, de considérer deux types d'erreur : les **erreurs aléatoires** et les **erreurs systématiques**.

•Les erreurs aléatoires (ou accidentelles) - Elles affectent la précision des données et sont non corrélées. Ce type d'erreur est dû à des raisons nombreuses et variées, généralement inconnues, affectant différemment chaque mesure individuelle. Ces erreurs étant inévitables, il faut en estimer l'importance afin de pouvoir en tenir compte lors de l'évaluation de l'incertitude finale. Dans la mesure du possible, la technique de mesure induisant les erreurs aléatoires les plus faibles devrait être préférée.

•les erreurs systématiques - Elles affectent la fiabilité des données et sont totalement corrélées. On parle aussi d'inconsistance. Supposons qu'aucune erreur aléatoire n'affecte les mesures. La différence entre la vraie valeur et la valeur mesurée, si elle existe, est alors due à une erreur systématique. L'origine des erreurs systématiques est le plus souvent liée à la calibration de l'appareil de mesure qui n'est pas parfaite ou à un phénomène extérieur qui perturbe la mesure (erreur d'appareillage, changement d'observateur...).

Recherche des erreurs et corrections des mesures

Selon la nature des erreurs constatées ou supposées la recherche de ces dernières fait appel à différentes techniques et méthodes

« **in situ** » qui consiste à vérifier sur place la manière dont les données ont été organisées, traitées et/ou transformées.

Investigation de bureau qui consiste à vérifier la chaîne de traitement de la mesure/donnée à chaque étape de son élaboration, tout comme la manière dont on a constitué les séries de données soumises à contrôle et/ou publication.

Investigation statistique qui, à l'aide d'outils spécifiques, permet de mettre en évidence certaines erreurs ou inconsistance. Ces techniques efficaces ont largement été utilisées dans la pratique professionnelle et se basent sur des hypothèses spécifiques qu'il convient de bien connaître.

Hypothèses de l'analyse statistique :

Les calculs statistiques sont basés sur un certain nombre d'hypothèses qui doivent en principe être vérifiées. Parmi celles-ci, citons :

Les mesures reflètent les vraies valeurs - Cette hypothèse n'est malheureusement jamais réalisée en pratique, du fait des erreurs systématiques ou aléatoires.

Les données sont consistantes - Aucune modification dans les conditions internes du système n'intervient durant la période d'observation (position du pluviomètre, procédures d'observation, observateur unique).

La série de données est stationnaire - Les propriétés de la loi statistique qui régit le phénomène (moyenne, variance ou moments d'ordre supérieur) sont invariantes au cours du temps.

Les données sont homogènes - Une série de données est réputée non homogène lorsque:

- elle provient de la mesure d'un phénomène dont les caractéristiques évoluent durant la période de mesure; le phénomène est alors dit non-stationnaire (par exemple: variations climatiques, variations du régime des débits dues à une déforestation ou un reboisement).
- Il est également possible d'observer des signes d'une non stationnarité apparente lorsque l'électronique intégrée à l'équipement de mesure présente une dérive temporelle ou lors du changement de l'observateur.
- elle reflète deux ou plusieurs phénomènes différents. Le régime d'une rivière à l'aval de la confluence de deux sous bassins dont le comportement hydrologique est très contrasté constitue un bon exemple de ce défaut d'homogénéité.

La série de données est aléatoire et simple - Le caractère aléatoire et simple d'une série d'observations est une hypothèse fondamentale pour l'analyse statistique.

- *Un échantillon aléatoire* signifie que tous les individus de la population ont la même probabilité d'être prélevés.
- *Un échantillon simple* signifie que le prélèvement d'un individu n'influe pas la probabilité d'apparition des individus suivants. Autrement dit, si toutes les observations de la série sont issues de la même population et qu'elles sont indépendantes entre elles, la série est alors aléatoire et simple.
- La non vérification du caractère aléatoire et simple peut avoir plusieurs causes, parfois simultanément. Ces causes se groupent en deux catégories, les défauts d'autocorrélation d'une part (caractère non aléatoire des séries) et les défauts de stationnarité du processus d'autre part (dérive à long terme et dérive cyclique).

La série doit être suffisamment longue - La longueur de la série influe sur les erreurs d'échantillonnage, notamment sur le calcul des moments d'ordre supérieurs donc sur les tests inhérents à leur fiabilité.

Quelques notions très importantes pour le contrôle et l'organisation des données hydropluviométriques

Les Précipitations

- 1mm de pluie = épaisseur de la couche d'eau (1mm) sur une surface de 1m^2 = volume d'1L d'eau versé sur la surface d' 1m^2
- Pluie journalière (P_j) = totale de la pluie tombée dans 24h
- Intensité de la pluie $i(\text{mm/h})$ = quantité de la pluie / temps de l'averse ramenée à 1h
- Pluviogramme = enregistrement à l'aide d'un pluviographe $P = f(\text{temps})$
- Hyétogramme = graphique donnant l'intensité i de la pluie par tranche de temps (1h, 2h, 30mn)
- Pluie mensuelle (P_m) = totale de la pluie enregistrée pendant tous les jours du mois considéré de l'année considérée = Somme des pluies journalières de tous le mois ($P_m = \sum P_j$)
- Pluie annuelle P_a = total de la pluie enregistrée pendant l'année considérée = somme de toutes les pluies mensuelles des mois de l'année considérée ($P_a = \sum P_m = \sum P_j$)

- Série de pluies annuelles = Pluies annuelles enregistrées sur une période donnée de (10, 20, 30 et plus) d'années)
- Pluie (précipitation) moyenne annuelle ou interannuelle notée P_{moy} ou \bar{P} = moyenne arithmétique des pluies annuelles de la série considérée représentant une période donnée. P_{moy} ou $\bar{P} = \sum P_a / \text{nombre d'année de la période}$
- Pluie (précipitation) moyenne mensuelle interannuelle = moyenne arithmétique des pluies mensuelles de chaque mois enregistrée pendant le mois considéré de toutes les années de la série. Pour chaque série d'années on obtient une pluie mensuelle moyenne pour chaque mois de l'année et donc 12 valeurs (une valeur par mois) qui vont constituer une année moyenne de la pluie (précipitations) mensuelle qui va représenter la période considérée,
- (exemple : Pluie (précipitation) moyenne mensuelle interannuelle du mois de janvier sur une période donnée de (10, 20, 30 et plus) d'années) notée $\overline{P_{janvier}} = \sum P_{m \text{ janvier}} / \text{nombre d'année de la période}$,

- lame d'eau précipitée moyenne sur un bassin est déterminée par extrapolation spatiale (méthodes des polygones de Thiessen, des Isohyètes, altimétrique, des deux axes, éléments finis,.....) des valeurs des précipitations enregistrées dans les différentes stations pluviométriques situées dans le territoire du bassin et celles avoisinantes.
- Le volume d'eau précipité sur un bassin est calculé par

$$V_{\text{précipitations}} (\text{m}^3) = \text{lame précipitée } L_p (\text{mm}) \times 10^{-3} \times \text{Surface du bassin } (\text{m}^2)$$

NB : Ce calcul peut être réalisé en pondérant les précipitations au niveau de chaque station par le % de la surface correspondante des polygones si on utilise par exemple la méthode de Thiessen).

Les Débits

- Le débit d'un cours d'eau est donnée en m^3/s ou en L/s (pour les cours d'eau à faible débit)
- Le débit moyen journalier Q_j = moyenne arithmétique de l'ensemble des débits Q_i mesurés pendant la journée $Q_j = \sum_1^n Q_i/n$ (nombre de mesures)
- Le débits moyen mensuel Q_m = moyenne arithmétique des débits journaliers du mois considéré $Q_m = \sum_1^n Q_j/n$ (nombre de jours du mois considéré)
- Le débit moyen annuel Q_a = moyenne arithmétique des débits moyens mensuels des mois de l'année considérée $Q_a = \sum Q_m/12$

- Série des Débits annuels = Débits moyens annuels enregistrés sur une période donnée de (10, 20, 30 et plus) d'années)
- débit moyen annuel ou interannuel (Module) notée Q_{moy} ou \bar{Q} = moyenne arithmétique des Débits annuels de la série considérée représentant une période donnée. Q_{moy} ou $\bar{Q} = \sum Q_a / \text{nombre d'année de la période}$
- Débit moyen mensuel interannuel = moyenne arithmétique des débits moyens mensuels de chaque mois enregistré pendant le mois considéré de toutes les années de la série. Pour chaque série d'années on obtient un débit mensuel moyen pour chaque mois de l'année et donc 12 valeurs (une valeur par mois) qui vont constituer une année moyenne des débits moyens mensuels qui va représenter la période considérée,
- (exemple : débit moyen mensuel interannuel du mois de janvier sur une période donnée de (10, 20, 30 et plus) d'années) notée $\overline{Q_{\text{janvier}}} = \sum Q_{\text{m janvier}} / \text{nombre d'année de la période ,}$

➤ Le volume d'eau écoulé sur un bassin est calculé par

$$V \text{ écoulé (m}^3\text{)} = \text{Débit m}^3\text{/s) x temps (s)}$$

Exemple Volume annuel = $Q_a(\text{m}^3/\text{s}) \times 365(\text{jours}) \times 24(\text{heures}) \times 3600(\text{s})$

➤ $\text{Lame d'eau écoulée (mm)} = [V_{\text{écoulé}} (\text{m}^3) / \text{Surface du bassin contrôlée (m}^2\text{)}] \times 10^3$
 $= [\text{Débit m}^3\text{/s) x temps (s) / Surface du bassin contrôlée (m}^2\text{)}] \times 10^3$

La lame d'eau écoulée est calculée en mm pour pouvoir la comparer avec la lame d'eau précipitée (calcul du bilan hydrologique par exemple), on l'utilise aussi dans les modèles hydrologiques)

Courbes intéressantes à tracer

- Courbes de variation temporelle des pluies ou des débits : courbes qui représentent l'évolution des précipitations ou des débits en fonction du temps. Ces courbes peuvent être tracées pour tous les pas de temps (journalier, mensuel, saisonnier et annuel), pour chaque année ou pour une période donnée). Ces courbes peuvent faire l'objet d'une analyse descriptive quantitative qui permet à prime à bord d'avoir une idée assez claire sur l'évolution du paramètre étudié et le régime hydrologique du cours d'eau en question. Elles permettent par exemple de déceler les mois, les saisons, les années et les périodes sèches et humides, de voir aussi leurs importance leurs fréquences sur une année ou une série d'années.
- Courbe des débits classés : cette courbe est tracée à partir des débits moyens journaliers interannuels relatifs à une série de plusieurs années de mesures. Cette courbe permet de déterminer les débits caractéristiques de la série qui sont:

- a- Débit caractéristique maximum (DCM) ou Débit caractéristique de crue (DCC10) : Représente le débit maximum (pointe) de l'année (débit atteint ou dépassé 10 jours par an).
- b- Débit caractéristique de 6 mois (DC6) : Représente le débit de fréquence 1/2 ou débit dépassé 6 mois par an.
- c- Débit caractéristique d'un an, trois ou neuf mois (DC1, DC3 et DC9) : Représente les débits dépassés respectivement 1,3 ou 9 mois par an.
- d- Débit spécifique d'étiage (DCE): Représente le débit dépassé 355 jours par

