



# Evaluation des tendances d'évolution des concentrations en polluants dans les eaux souterraines.

#### Guide méthodologique

B. LOPEZ (BRGM),

Avec la participation de A. BLUM (BRGM) et N. BARAN (BRGM)

**Avril 2011** 





# Evaluation des tendances d'évolution des concentrations en polluants dans les eaux souterraines.

Guide méthodologique.

BRGM/RP-59931-FR Avril 2011

Étude réalisée dans le cadre des projets de Service public du BRGM 2010

B. Lopez,

Avec la participation de A. Blum et N. Baran

| Vérificateur : | Approbateur : |
|----------------|---------------|
| Nom :          | Nom :         |
| Date :         | Date :        |
| Signature :    | Signature :   |

En l'absence de signature, notamment pour les rapports diffusés en version numérique, l'original signé est disponible aux Archives du BRGM.









Contexte de programmation et de réalisation

La Directive 2000/60/CE et la directive eaux souterraines 2006/118/CE impose aux Etatsmembres d'identifier les tendances et les inversions des tendances d'évolution de la qualité des eaux souterraines. Les efforts de collecte et de bancarisation des données brutes sur la qualité des eaux souterraines offrent des bases de données de taille croissante qu'il est possible de valoriser. Des outils mathématiques de détermination de tendances existent et doivent être testés pour connaître leurs limites d'application dans le domaine de l'eau souterraine.

Le projet se décompose en trois phases sur une durée de un an : La première phase est une revue bibliographique des outils mathématiques d'étude de chroniques et des outils déjà utilisés dans le domaine de l'eau. La deuxième phase se propose d'identifier les spécificités du domaine de l'eau souterraine afin de tester la faisabilité des traitements identifiés en première phase. Un rapport scientifique final et un guide méthodologique pour la détermination des tendances avec recommandations aux utilisateurs sont délivrés en fin de projet et leur rédaction constitue la phase 3 du projet.

Les auteurs

Benjamin Lopez
Ingénieur hydrogéologue
b.lopez@brgm.fr
BRGM service EAU
Bureau de Recherches Géologiques et Minières
3, avenue Claude-Guillemin
B.P. 36009 45060 Orléans cedex 2

Les correspondants

Onema: Pierre-François Staub, DAST, pierre-francois.staub@onema.fr

Référence du document : BRGM/RP-59931-FR





Droits d'usage : Accès libre

Couverture géographique : Nationale

Niveau géographique : National

Niveau de lecture : **Professionnels et experts** 

Nature de la ressource : **Document** 





## Evaluation des tendances d'évolution des concentrations en polluants dans les eaux souterraines.

#### Guide méthodologique

B. Lopez

#### **SOMMAIRE**

| 1. Introduction   | 11        |
|---|-----------|
| 2. Procédure opérationnelle pour l'analyse des tendances        | 15        |
| A. Caractérisation des données                                  | 15        |
| B. Analyse graphique et exploratoire des données                | 17        |
| C. Tests statistiques   | 21        |
| D. Détermination des facteurs explicatifs des tendances observe | ées<br>27 |
| 3. Les points essentiels à retenir                              | 31        |
| 4. Conclusion   | 35        |
| 5. Bibliographie  | 37        |





#### LISTE DES ILLUSTRATIONS

| Illustration 1 : Schéma de caractérisation des données brutes (avec N : nombre de données ;      |
|--|
| L : longueur de chronique)16   |
| Illustration 2 : Les différentes méthodes visuelles d'identification des tendances d'évolution   |
| des concentrations en contaminant dans les eaux naturelles17                                     |
| Illustration 3 : Schéma d'analyse exploratoire des données. A gauche, questions successives      |
| avec indication de la méthode à appliquer ; à droite, orientations préliminaires sur les actions |
| à réaliser ou les méthodes pressenties pour l'analyse statistique19                              |
| Illustration 4 : Arbre décisionnel pour le choix de la (ou des) méthode(s) statistique(s) à      |
| appliquer sur les chroniques d'évolution de la qualité des eaux souterraines23                   |
| Illustration 5 : Exemple de recherche de facteurs explicateurs par covariance croisée entre      |
| les évolutions des concentrations en NO3 et l'évolution des niveaux piézométriques dans un       |
| aquifère carbonaté à nappe libre du Jurassique moyen (Bathonien) au Sud-Est de Caen.             |
| L'exemple illustre le cas d'une corrélation positive entre les signaux avec un déphasage         |
| moyen de réponse des concentrations en NO <sub>3</sub> de 10 mois                                |





### Evaluation des tendances d'évolution des concentrations en polluants dans les eaux souterraines.

#### Guide méthodologique

B. Lopez

#### Résumé

Le suivi de la qualité des eaux souterraines en France est réalisé selon un cadre réglementaire défini dans La Directive 2000/60/CE et la directive eaux souterraines 2006/118/CE. Les données de qualité des eaux s'accumulent dans les bases de données de sorte que la question de leur valorisation devient une problématique centrale. Les textes demandent ainsi aux Etats membres d'identifier les tendances d'évolution des concentrations en polluants afin d'inverser celles orientées vers une hausse durable. Des méthodes statistiques doivent être employées à cet effet mais aucune étude spécifique au domaine de la qualité des eaux souterraines et orientée vers les gestionnaires n'a été menée pour évaluer l'aptitude des tests disponibles à répondre à ces objectifs.

Le BRGM, en partenariat avec l'ONEMA (convention ONEAM-BRGM 2010), propose dans un premier temps de faire l'inventaire des méthodes statistiques existantes qui peuvent, sous certaines conditions détaillées dans l'étude, être appliquées sur des données de qualité des eaux souterraines. La lecture de la littérature scientifique sur le sujet permet de retracer l'historique du développement et de l'amélioration de ces méthodes et révèle certaines limites pour leur application sur des données de qualité des eaux. L'autocorrélation des données apparaît par exemple comme un phénomène très perturbateur des résultats des tests statistiques classiques. Une revue complémentaire des travaux réalisés en France et à l'international pour évaluer les tendances d'évolution de la qualité des eaux superficielles et souterraines rend compte des biais qui existent entre l'évolution de la recherche scientifique et les applications faites par les gestionnaires.

Afin d'accompagner les gestionnaires (les agences de l'eau en particulier), une procédure de traitement est proposée. Le choix final des tests statistiques à utiliser est argumenté par une analyse préalable de leur robustesse et de leur puissance face aux caractéristiques spécifiques des données de qualité des eaux souterraines. Ces caractéristiques sont issues de l'analyse des données bancarisées dans la base de données nationale ADES sur la qualité des eaux souterraines. L'outil global d'aide à la décision qui conclut l'étude est ainsi nourri des dernières avancées de la Recherche sur le domaine, propose l'utilisation des méthodes statistiques les plus robustes en fonction des caractéristiques des données à traiter et permet aux opérateurs d'être accompagnés pas à pas pour mettre en œuvre une analyse statistique des chroniques d'évolution des concentrations en polluants dans les eaux souterraines.





Mots clés : Tendance, DCE, tests statistiques, qualité, eaux souterraines





# Guide méthodologique pour l'évaluation des tendances d'évolution des concentrations en polluants dans les eaux souterraines

B. Lopez

#### 1 Introduction

La collecte, la bancarisation ainsi que la valorisation des données de qualité des eaux souterraines sont des questions scientifiques centrales reprises et traduites de façon réglementaire par l'intermédiaire de la Directive Cadre sur l'Eau 2000/60/CE et la Directive eaux souterraines 2006/118/CE notamment. L'identification et la quantification des tendances d'évolution des concentrations en contaminants dans les eaux souterraines fait l'objet d'un volet spécifique, le législateur recommandant l'emploi d'une méthode statistique sans toutefois préciser et cadrer les modalités de la procédure à suivre. Or, l'analyse des chroniques d'évolution de la qualité des eaux souterraines est très contrainte par les caractéristiques particulières de ces données, impliquant la mise en œuvre d'une procédure rigoureuse et l'application de méthodes statistiques adaptées. Ainsi, pour faciliter l'analyse des chroniques d'évolution des concentrations en contaminants dans les eaux souterraines, il est proposé de mettre à disposition des gestionnaires un guide méthodologique permettant d'évaluer de manière robuste et adaptée les tendances d'évolution.

Le document s'appuie à la fois sur une revue bibliographique des méthodes statistiques existantes pour l'évaluation des tendances d'évolution des séries chronologiques, sur des retours d'expériences de travaux réalisés en France et à l'étranger, ainsi que sur des tests de robustesse et de puissance des méthodes statistiques sur des chroniques réelles ou simulées possédant les caractéristiques spécifiques des données de qualité des eaux





souterraines. Il sera ainsi possible, pour l'utilisateur de ce quide méthodologique, de recourir en cas de besoin au rapport BRGM/RP-59515-FR « Evaluation des tendances d'évolution des concentrations en polluants dans les eaux souterraines. Revue des méthodes statistiques existantes et recommandations pour la mise en œuvre de la DCE ». Ce dernier propose en effet une description détaillée des méthodes statistiques potentiellement applicables au domaine de la qualité des eaux souterraines ainsi qu'une revue bibliographique des travaux d'identification des tendances d'évolutions des concentrations en contaminants dans le domaine des eaux superficielles et souterraines. Des tests de robustesses des méthodes statistiques face aux caractéristiques des données de qualité des eaux souterraines ont été réalisés afin d'identifier les méthodes les plus pertinentes en fonction des données disponibles. Pour affiner le choix des méthodes à appliquer, des tests de puissance ont aussi été réalisés dans le but de montrer, dans des conditions initiales favorables à l'application des différentes méthodes, quelles sont celles qui détectent les tendances les plus ténues. Le détail des méthodes statistiques existantes, la revue bibliographique des études d'identification des tendances réalisées en France et à l'étranger ainsi que les tests de robustesse et de puissance ont permis de faire des préconisations à la fois sur une démarche à adopter, sur le choix des tests statistiques à appliquer et sur les données à prendre en compte pour l'analyse.

Le guide méthodologique est rédigé de façon à aider les gestionnaires de la qualité des eaux souterraines dans la démarche d'identification et de quantification des tendances d'évolution des concentrations en polluants dans les eaux souterraines. L'identification des tendances doit en effet s'intégrer dans une procédure plus globale d'analyse des séries chronologiques disponibles dont les tests statistiques ne constituent qu'une étape, parfois facultative. Un certain nombre de caractéristiques des chroniques d'évolution de la qualité des eaux souterraines peuvent en effet influencer de façon significative la robustesse et la puissance des tests statistiques. Ces caractéristiques – nombre de données dans la série, distribution des données, présence d'autocorrélation – demandent à être identifiées avant la mise en œuvre des tests statistiques. De plus, de la longueur et de la fréquence d'analyses dépendent les conclusions qu'il est possible de tirer de l'analyse de tendance. Enfin, l'analyse approfondie du comportement de la substance suivie au point de mesure (tendance visuelle, cycles, « saut » des valeurs de concentrations) est doublement nécessaire : elle permet d'émettre la (ou les) hypothèse(s) à tester avec le choix des tests statistiques en corolaire et peut amener à réaliser des traitements préalables sur les données brutes.





La procédure systématique d'analyse et de traitement des données de qualité des eaux proposée est élaborée dans le but de bien contraindre la détermination des tendances. Cette méthode « pas à pas » permet de minimiser l'identification de fausses tendances et d'éliminer les points fallacieux. Elle est composée de quatre étapes :

- Caractérisation des données
- Analyse graphique et exploratoire des données
- Analyse statistique
- Détermination des facteurs explicatifs

Ces étapes, à réaliser dans l'ordre proposé, constituent le déroulement de l'analyse des chroniques d'évolution des concentrations en contaminants dans les eaux souterraines. Les travaux à mettre en œuvre peuvent différer suivant si l'analyse concerne une ou plusieurs séries de données, si les tendances recherchées le sont à une échelle ponctuelle ou spatiale et si des données du type « inférieur à » (une limite de détection ou de quantification) sont présentes dans les chroniques à analyser. Les différentes étapes de l'analyse des séries chronologiques sont présentées sous forme d'arbres décisionnels dont chaque branche est détaillée dans le corps du document.





## 2 Procédure opérationnelle pour l'analyse des tendances

#### A. Caractérisation des données

Après avoir sélectionné le jeu de données pour la détermination de tendance, la première étape consiste à étudier leurs caractéristiques. Cette évaluation débute généralement par la représentation graphique des données (y = f(t)) suivie par le calcul des statistiques de base qui caractérisent la chronique (moyenne, médiane, écart type, variance, minimum, maximum, quartiles et quantiles). Cette première étape permet d'appréhender la « structure » des données, de déterminer les meilleures approches et les limites des tests statistiques à mettre en œuvre.

La longueur de la chronique et le nombre de données sont des éléments à prendre en compte lors de cette phase. Si le nombre de point est inférieur à 10, alors il est impossible de mettre en œuvre un autre test statistique que la simple régression linéaire (dont les résultats sur des chroniques de moins de 10 données doivent être considérés avec la plus grande prudence). De même, si la longueur est inférieure à 1 an, il faut faire très attention à la présence possible de fluctuations périodiques, qui empêchent toute analyse de tendances si la période est saisonnière. Enfin, la régularité des pas de temps et la fréquence d'échantillonnage sont analysées. Ces caractéristiques influent en effet le choix voire la possibilité d'appliquer un test statistique de stationnarité dans le cas où les chroniques présentent de l'autocorrélation.

Les données inférieures aux limites de détections/quantifications sont identifiées. Dans le cas de la présence de données de ce type, il est important de noter, pour chaque valeur, la limite de quantification du laboratoire lors de l'analyse. Les méthodes spécifiques de prise en compte des valeurs inférieures aux limites de quantifications (réglementaire <LQ=LQ/2 ou autres) sont alors mises en œuvre afin de reconstituer les chroniques et de poursuivre la démarche d'analyse des tendances. Il est de plus important de noter les périodes où les données sont manquantes.

C'est aussi lors de cette phase que des « outliers » (données extrêmes) peuvent être supprimés lorsqu'ils correspondent à des erreurs de mesures ou d'analyses. Si des données





extrêmes sont repérées, l'opérateur peut choisir de les ôter de la chronique à traiter ou bien, dans le cas d'un jeu de données important, de s'orienter vers un test peut sensible aux « outliers » comme les tests non paramétriques de Mann-Kendall ou le test de Spearman.

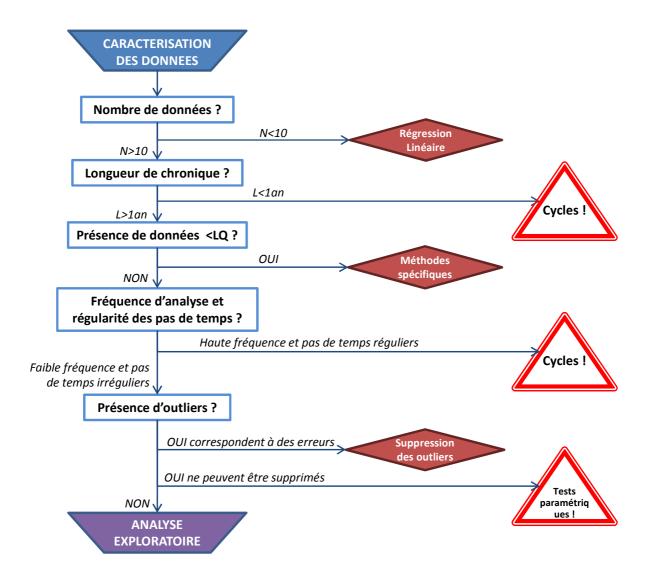


Illustration 1 : Schéma de caractérisation des données brutes (avec N : nombre de données ; L : longueur de chronique).

A gauche, questions successives pour la validation ; au centre, orientations préliminaires ; à droite, alertes sur des processus ou l'utilisation de certaines méthodes (attention aux [...]).

L'Illustration 1 résume la première phase de validation des données. Toutes les étapes, rédigées sous forme de questions, sont à réaliser quelques soient les réponses données. A l'issue de la phase de caractérisation des données, l'opérateur maitrise les différents paramètres statistiques de la série chronologique à analyser, est alerté sur quelques points critiques des données et peut être orienté vers des méthodes très spécifiques.





#### B. Analyse graphique et exploratoire des données

L'analyse visuelle (graphique) et exploratoire des données à traiter est une étape essentielle pour la détermination des tendances. Elle consiste à tracer la ou les variables concernées en fonction du temps et à les ajuster par une courbe de lissage.

Plusieurs méthodes peuvent être utilisées pour représenter graphiquement l'évolution des concentrations en contaminant dans les eaux souterraines comme présenté en Illustration 2.

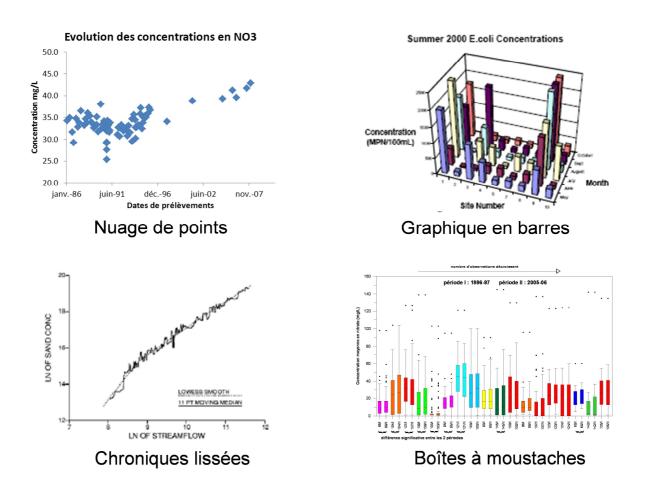


Illustration 2 : Les différentes méthodes visuelles d'identification des tendances d'évolution des concentrations en contaminant dans les eaux naturelles.

De plus, il existe de nombreuses techniques permettent d'ajuster les données brutes comme la moyenne mobile, la régression linaire, la régression quadratique ou encore le lissage de LOWESS (Helsel et Hirsch, 1992). Il est aussi possible de tracer l'autocorrélogramme des chroniques de qualité des eaux afin de rendre compte des possibles cyclicités dans les





signaux. L'inspection visuelle des données brutes couplée à l'analyse des données ajustées peut indiquer clairement le type, la direction et l'amplitude de la ou des tendances. Cette analyse peut aussi révéler les cycles à plus ou moins long terme et d'autres structures caractéristiques des données comme la linéarité, l'aspect monotonique des variations, les paliers éventuels et les inversions de tendances. Si des cyclicités saisonnières sont par exemple mises en évidence, et que le calendrier de prélèvement est très régulier, alors il est possible d'envisager un test qui prenne en compte ce type de variations : le test Kendall Saisonnier.

Si l'analyse qualitative des données est bien menée, elle permet, dans certains cas et suivant des objectifs non quantitatifs, de s'affranchir de la mise en œuvre de tests statistiques de détermination des tendances. Dans la plupart des cas néanmoins, il est nécessaire de tester les hypothèses sur la structure des données émises lors de l'analyse exploratoire (stationnarité, tendance, rupture de pente). Le choix du test statistique est en effet dépendant à la fois du type d'hypothèse à tester et des caractéristiques des séries temporelles.

La distribution normale des données doit ainsi être évaluée durant cette phase. Plusieurs méthodes permettent de réaliser cet exercice, le test de Shapiro-Wilk apparaissant bien adapté pour les données de qualité des eaux souterraines. Dans le cas où les données sont non normalement distribuées, il est nécessaire de s'orienter vers un test non paramétrique. Si les données montrent une distribution normale, il est possible d'appliquer, au choix, les tests paramétriques mais aussi les tests non paramétriques dont la robustesse et la puissance est tout aussi bonne dans ce cas de figure.

L'autocorrélation des données est calculée lors de l'analyse exploratoire et sa significativité doit aussi être testée. Si une autocorrélation est mise en évidence, il peut s'agir soit d'un biais lié à une tendance effective ou une certaine cyclicité, soit d'une autocorrélation vraie, induite par un phénomène d'inertie du système. Or, le problème de distinguer ce qui relève de l'autocorrélation ou d'une tendance ne peut se faire qu'en étudiant le comportement du polluant au point de mesure, c'est-à-dire en contraignant les facteurs qui régissent l'évolution de ses concentrations dans l'eau. Cette connaissance passe par des fréquences de prélèvements élevées et par des durées d'acquisition longues.

Seule cette approche permet de choisir un test qui prenne en compte ou non l'autocorrélation des données.





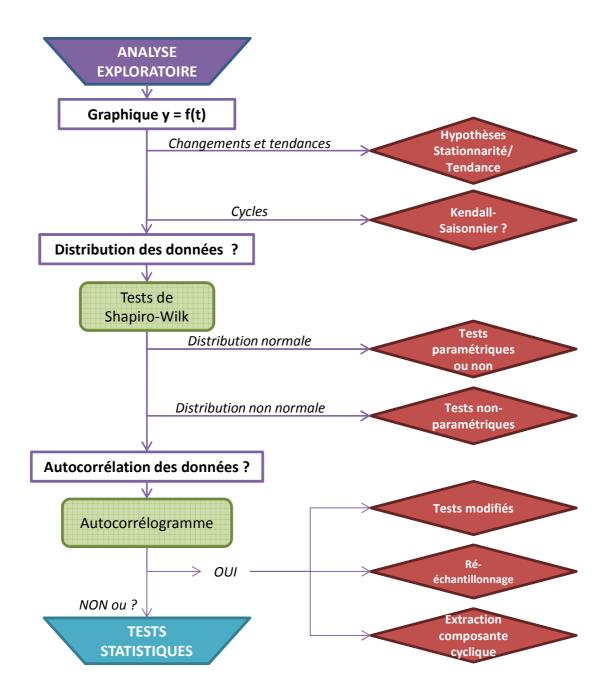


Illustration 3 : Schéma d'analyse exploratoire des données. A gauche, questions successives avec indication de la méthode à appliquer ; à droite, orientations préliminaires sur les actions à réaliser ou les méthodes pressenties pour l'analyse statistique.

La deuxième phase d'analyse des chroniques d'évolution de la qualité des eaux souterraines est l'analyse exploratoire des données dont le schéma est présenté dans l'Illustration 3. Comme précédemment, toutes les étapes de l'analyse exploratoire doivent être réalisées et ce, quelques soient les réponses données après chaque étape.





A la fin de la phase exploratoire des données l'opérateur sait quelles hypothèses sont à tester, est capable de dire si les données sont normalement distribuées et il a identifié la présence d'autocorrélation dans les données brutes.

Remarque: si le jeu de données est très important et que seule une faible part des chroniques est affectée par de l'autocorrélation, alors il peut être choisi de ne pas la prendre en compte lors de l'analyse des tendances. Il faut juste avoir à l'esprit que, pour les chroniques autocorrélées, ce phénomène a tendance à surestimer les tendances, notamment celles identifiées comme significativement croissantes.





#### C. Tests statistiques

Après avoir identifié un changement potentiel ou une possible tendance lors de l'analyse exploratoire des séries temporelles, la réalisation d'un test statistique permet de confirmer ou d'infirmer ces observations visuelles. Il a été montré que de nombreuses méthodes statistiques permettent de tester la présence ou non de tendance ainsi que les changements dans les séries chronologiques.

Le choix du test à appliquer dépend de la longueur des chroniques, des effectifs dans les séries, de la distribution des données, de l'occurrence des valeurs extrêmes (« outliers »), des données inférieures à la limite de quantification, des données manquantes, de la présence possible d'autocorrélation dans la série de données et du jugement de l'opérateur (fondé sur une connaissance a priori des mécanismes en jeu).

Lorsqu'aucun élément a priori ne permet à l'opérateur de choisir telle ou telle méthode, l'arbre de décision présenté en Illustration 4 permet de l'aider à orienter son choix en fonction des caractéristiques des données à traiter et de l'objectif de l'analyse. Le schéma de sélection du test statistique à appliquer fait suite, et se nourri, des connaissances acquises sur la série lors des deux précédentes phases de la procédure d'analyse. Contrairement à ces dernières, chaque réponse aux questions guide l'opérateur vers une branche de l'arbre décisionnel de sorte que toutes les étapes ne sont pas à réaliser.

Afin de limiter au maximum les sources d'erreurs lors du traitement statistique, certains points sont à considérer pour juger de la pertinence et de la capacité à utiliser les données disponibles en vue d'une analyse statistique :

- Qualité des données constantes : il est préférable d'utiliser des données acquises selon des protocoles de terrain et des méthodes analytiques similaires.
- Longueur du suivi et nombre de prélèvements : Selon Lettenmaier et al. 1978, 5 années de suivi mensuel sont nécessaire pour l'identification de tendances monotoniques. Pour des changements par palier, 2 ans de suivis mensuels avant et après le changement sont généralement préconisés (Hirsch, 1982).
- Pour tous les tests statistiques, <u>au minimum 10 données dans la série</u>. De plus, la longueur du suivi détermine l'échelle temporelle de la détermination de tendance.





- <u>Saisonnalité</u> : les paramètres qui varient selon des cycles saisonniers nécessitent soit des tests particuliers soit de « dé-saisonnaliser » la chronique.
- <u>Distribution des données</u> : si les données ne suivent pas une loi normale alors l'utilisation d'un test non paramétrique s'impose.
- Variance similaire pour tout le jeu de données : les méthodes statistiques sont plus robustes lorsque la <u>fréquence de prélèvement est constante</u> (minimisation de la variance).
- <u>Données manquantes minimum</u> : les données manquantes tendent à augmenter la variance et donc à diminuer la robustesse des tests.
- <u>Données extrêmes</u> : les tests paramétriques sont très sensibles aux données extrêmes, à l'inverse des tests non paramétriques.
- Paramètres externes : les paramètres comme le débit, la recharge, les pompages peuvent impacter l'évolution de certains paramètres de la qualité des eaux. Leur suivi concomitant permet de les prendre en compte dans le calcul de tendance par le calcul du test sur les résidus de la régression entre le paramètre étudié et le paramètre externe.
- Changements des valeurs de <u>limites de détection</u>: de tels changements peuvent engendrer de fausses tendances.

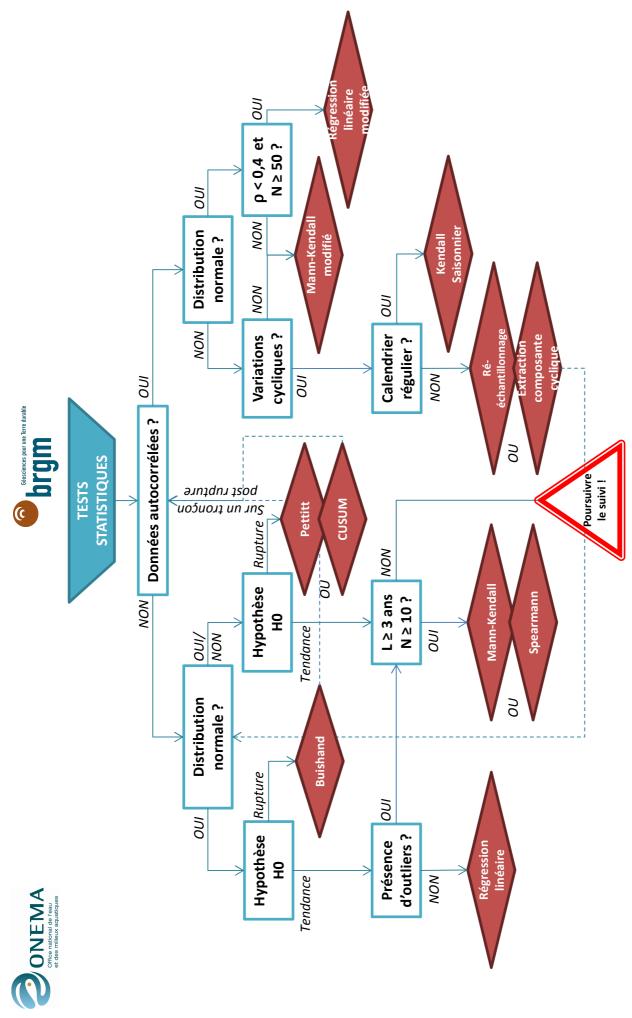


Illustration 4 : Arbre décisionnel pour le choix de la (ou des) méthode(s) statistique(s) à appliquer sur les chroniques d'évolution de la qualité des eaux souterraines.





La présence d'autocorrélation dans les données est la première caractéristique qui oriente le choix du test à appliquer. S'il y a lieu, la régression linéaire modifiée (Renard, 2006) sera préconisée pour identifier la tendance des séries normalement distribuées à faible écart type (< 0,4) et nombre de données supérieur à 50. Si ces deux critères ne sont pas respectés, le test de Mann-Kendall modifié est préconisé (Lopez et Leynet, 2011 §2.2.2; Yue et Wang, 2004; Hamed et Rao, 1998), ainsi que dans le cas d'une distribution des données non normale et d'une série non cyclique. Si des variations cycliques sont observées et que le calendrier de prélèvement est régulier alors le test Kendall-Saisonnier (Lopez et Leynet, 2011 §2.2.2; Helsel et Hirsch, 2002) est recommandé tant il a prouvé sa robustesse et sa puissance face au test de Mann-Kendall classique. Si le calendrier n'est pas régulier, alors l'autocorrélation peut être « cassée » de deux manières différentes: soit en rééchantillonnant la série, soit en extrayant la composante cyclique au signal brut. Dans les deux cas, le schéma de sélection du test à appliquer sur la série transformée sera le même que pour des chroniques dont les données ne sont pas autocorrélées.

Lorsque les données ne sont pas autocorrélées, les tests paramétriques sont préconisés pour les séries à distribution normale. Si l'hypothèse à tester est l'homogénéité des paramètres statistiques de la chronique, le test de Buishand (Lopez et Leynet, 2011 §2.2.3) est le plus approprié. Si l'hypothèse à tester est la stationnarité, alors la régression linéaire est recommandée seulement s'il n'y a pas d'outliers dans la série. Dans le cas contraire, les tests non paramétriques de Mann-Kendall ou de Spearman (Lopez et Leynet, 2011 R §2.2.2) sont préférés seulement si les chroniques comportent plus de 10 données et sont longues d'au moins 3 ans. Dans le cas contraire, on ne peut que recommander de continuer le suivi de la qualité de l'eau en ce point et d'attendre l'acquisition d'autres données pour pouvoir appliquer un test statistique de tendance.

Pour les données non autocorrélées à distribution non normale des données, les tests non-paramétriques sont recommandés. Pour la détection de rupture dans les séries, les tests de Pettitt ou de CUSUM apparaissent appropriés (Lopez et Leynet, 2011 §2.2.4). Pour l'identification des tendances d'évolution, les tests de Mann-Kendall ou de Spearman peuvent être utilisés seulement sur des chroniques de 10 données et de 3 ans minimum. Comme précédemment, le cas contraire ne permet pas aujourd'hui d'évaluer statistiquement la présence de tendance.

Ce schéma de sélection n'est pas strictement restrictif et correspond aux choix idéaux en fonction des données à traiter. Ainsi, il a été montré que les tests non-paramétriques sont





presque aussi robustes et puissants que les tests paramétriques dans le cas de données normalement distribuées. De plus, même si le test de Mann-Kendall est moins pertinent que la version modifiée dans le cas de données autocorrélées, les résultats sont tout de même acceptables. Ainsi, dans le cas où un nombre important de chroniques doit être testé, et que le temps vient à manquer pour décrire de façon individuelle chacune des séries temporelles, alors il est préconisé la mise en œuvre du test de Mann-Kendall qui représente le meilleur compromis en terme de robustesse. Il n'est en effet pas besoin de vérifier la distribution des données, ce test est le moins sensible aux outliers et les résultats dans le cas de données fortement autocorrélées sont faiblement biaisés. Il sera juste nécessaire de préciser que les processus d'autocorrélation n'ont pas été vérifiés avant le traitement statistique, impliquant une possible surestimation des tendances identifiées à la hausse.





#### D. Détermination des facteurs explicatifs des tendances observées

L'identification des tendances et des changements dans les séries chronologiques est une première étape de l'exploitation des données de qualité des eaux mais les tests statistiques n'expliquent en rien les causes des variations observées. L'étape qui consiste à expliquer les tendances d'évolutions ou les changements de la qualité des eaux est donc primordiale dans la procédure globale de traitement des données. Etre capable d'expliquer les tendances signifie qu'il est possible d'en déterminer les causes et donc de mettre en œuvre les actions appropriées pour les inverser si nécessaire.

Les facteurs externes qui peuvent expliquer les évolutions constatées des concentrations en contaminants diffus dans les eaux souterraines peuvent être classés en deux groupes :

- Les facteurs naturels hydroclimatiques,
- Les facteurs anthropiques.

Les facteurs naturels hydroclimatiques guident les comportements hydrodynamiques des aquifères par l'intermédiaire des pluies efficaces qui s'infiltrent vers les nappes. Or, dans certains contextes climatiques et hydrogéologiques, les fluctuations des niveaux des nappes peuvent impacter les variations des concentrations en polluants stockés dans la zone non saturée et mobilisables lors des remontées de nappes. De plus, dans des contextes à infiltrations rapides d'eaux récentes, des fortes variations de la chimie des eaux souterraines peuvent apparaître lors d'évènements pluvieux intenses. Ces fluctuations peuvent être la conséquence d'une dilution des eaux souterraines par les eaux de pluies peu concentrées en polluants ou bien, à l'inverse, une augmentation rapide des concentrations dans les eaux souterraines si les eaux récentes dans le système ont rencontré une source importante de pollution lors de leur infiltration.

Les facteurs externes anthropiques correspondent quant à eux aux pressions polluantes en entrée des systèmes hydrogéologiques liées aux activités humaines. Il s'agit à la fois de l'occupation des sols sur les zones de recharge des aquifères, des quantités de substances polluantes appliquées sur ces sols, ainsi que de l'évolution des usages et des pratiques d'utilisation de ces substances. Il est aisé d'admettre que l'augmentation graduelle de l'utilisation d'une substance en zone de recharge d'un aquifère engendre, selon un délai plus



Partenariat 2010

Domaine: Savoirs

Action n8

Géosciences pour une Terre durable

ou moins important et sous réserve d'une forte persistance de la molécule dans les sols, une augmentation de la concentration de cette substance dans les eaux de l'aquifère.

Toutefois, les évolutions des concentrations en contaminants dans les eaux souterraines résultent, dans la plupart des cas, d'une combinaison de différents facteurs externes, hydroclimatiques et anthropiques. L'exercice consiste alors à évaluer la part de chacun des paramètres externes expliquant les tendances observées, c'est-à-dire comprendre leur rôle sur le comportement des substances dans le milieu naturel. La gestion intégrée de la qualité des eaux souterraines demande ainsi à la fois une bonne connaissance des contextes géologiques et environnementaux, des évolutions historiques des paramètres climatiques, ainsi que de l'historique d'évolution des pratiques de gestion et d'usages du bassin versant hydrogéologique considéré. Plusieurs approches permettent alors de travailler sur, et d'identifier, les facteurs explicatifs des tendances observées :

La datation des eaux souterraines est un moyen de rendre compte de possibles corrélations entre l'âge apparent des eaux et l'évolution constatée des contaminations. Des traceurs peuvent ainsi être recherchés en fonction de l'âge probable des eaux investiguées (CFC et tritium pour les eaux d'âge inférieur à 60 ans, SF6 pour les eaux très récentes ou, à l'inverse, carbone 14 pour les eaux très anciennes). Il est aussi possible de réaliser des traçages dans les systèmes à réponse rapide et dont on connaît les zones d'infiltrations, ce qui est généralement le cas des systèmes karstiques.

Pour explorer le champ des facteurs potentiellement responsables des tendances observées, il est aussi nécessaire de retracer l'historique de tous les phénomènes potentiellement impliqués dans l'acquisition de la chimie par les eaux souterraines. Ce sont bien sûr les phénomènes climatiques avec l'évolution des pluies efficaces en zones de recharge des aquifères mais aussi l'évolution des pressions responsables des pollutions. Lorsque les pollutions sont d'origine agricole, l'historique des pratiques culturales doit être reconstitué. Si les pollutions sont d'origine industrielle, la connaissance de l'évolution des volumes de polluants potentiels utilisés (ou de leurs précurseurs) est un gros plus pour travailler sur les facteurs explicatifs.

Une fois toutes les données potentiellement responsables des tendances observées mises à disposition, l'exercice consiste à rechercher les corrélations possibles entre les facteurs explicatifs potentiels et les chroniques de qualité des eaux analysées. L'analyse par corrélation croisée permet par exemple de faire cette interprétation par couples de





chroniques. Cet outil permet de dégager à la fois l'importance de la corrélation entre les signaux, le sens de la corrélation, ainsi que le déphasage entre les deux chroniques s'il existe. L'Illustration 5 montre un exemple de la recherche de facteurs explicatifs par covariance croisée entre lune chronique d'évolution des niveaux piézométriques et un chronique d'évolution des concentrations en NO<sub>3</sub> dans un aquifère carbonaté à nappe libre du Jurassique moyen au Sud-Est de Caen. D'autres méthodes d'analyses des corrélations existent, dont l'Analyse en Composante Principale (ACP) qui permet de réaliser une analyse multicritère et de déterminer, au sein d'un nombre important de variables aléatoires, celles qui montrent des comportements pouvant être rapprochés.

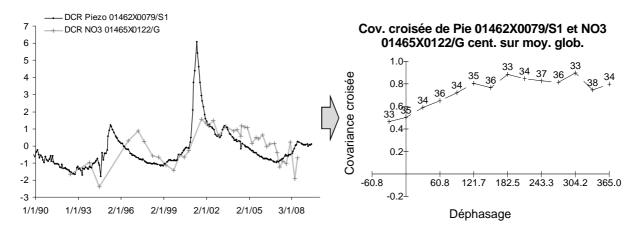


Illustration 5 : Exemple de recherche de facteurs explicateurs par covariance croisée entre les évolutions des concentrations en NO<sub>3</sub> et l'évolution des niveaux piézométriques dans un aquifère carbonaté à nappe libre du Jurassique moyen (Bathonien) au Sud-Est de Caen. L'exemple illustre le cas d'une corrélation positive entre les signaux avec un déphasage moyen de réponse des concentrations en NO<sub>3</sub> de 10 mois.

Rappelons que la variabilité totale d'une série chronologique d'évolution des concentrations en un contaminant dans les eaux souterraines est généralement le résultat de la somme d'une variabilité environnementale, d'une variabilité des pressions anthropiques, d'une variabilité analytique et d'une variabilité due à un bruit de mesure. L'objectif idéal à atteindre est de pouvoir quantifier quelle est la part du signal d'évolution de la qualité des eaux souterraines expliquée par tel phénomène, quelle est l'autre part de ce même signal expliquée par un autre phénomène et ainsi de suite. Ce type de résultats, très complexe à obtenir, permet pourtant d'identifier les facteurs sur lesquels il est possible d'intervenir pour inverser les tendances observées, et d'évaluer les impacts futurs de ces interventions en fonction de l'importance qu'a chacun des phénomènes dans l'explication de l'orientation de la tendance.





#### 3 Les points essentiels à retenir

#### Comment analyser et présenter une tendance ?

Cette question en sous-tend plusieurs autres qu'il est nécessaire d'avoir à l'esprit pour mener à bien l'analyse de tendance et valoriser de façon appropriée les résultats de l'analyse :

- La qualité des données disponibles est-elle cohérente avec les objectifs de l'analyse ? Le suivi de la qualité des eaux pour la mise en œuvre de tests de tendance demande en effet de suivre un protocole strict.
- A qui s'adressent les résultats ? Quels types d'analyse et de présentation sont alors appropriés ?
- Si un test statistique est envisagé, les données disponibles remplissent-elles les conditions d'applications du test ?
- Comment interpréter les tendances ? Déterminer les causes des tendances identifiées est souvent plus difficile que déterminer les tendances elles-mêmes.
- A-t-on pris en compte tous **les paramètres externes** susceptibles d'influences les tendances (débits, pressions, périodes d'analyses ...) ?

#### Comment identifier une tendance?

#### Visuellement

<u>Avantages</u> : L'élaboration de graphiques et/ou de cartes est un moyen facile de communiquer sur l'existence de tendances, surtout pour les non-initiés.

<u>Inconvénients</u> : Cette approche ne permet pas d'estimer la « véracité » de la tendance ou de quantifier la pente de celle-ci.

Par l'application d'une méthode statistique



Partenariat 2010

Domaine: Savoirs

Action n8

66osciences pour une Terre durable

<u>Avantages</u>: Cette approche permet d'identifier des tendances difficiles à voir, propose un résultat répétable et permet d'estimer les probabilités de se tromper.

<u>Inconvénients</u>: De nombreuses contraintes sont à respecter pour l'application des méthodes statistiques et des erreurs sont possibles lorsque les conditions initiales d'application des tests ne sont pas respectées.

#### Les méthodes visuelles d'identification des tendances

Si ces méthodes apparaissent faciles à mettre en œuvre et d'un accès aisé pour les noninitiés, leur choix ne peut être justifié que dans le cas de tendances très prononcées, sans aucune ambiguïté sur leur sens et leur amplitude. Il est en effet reconnu que :

- Les personnes se focalisent souvent sur les valeurs extrêmes et moins sur les changements subtils.
- Les faibles tendances graduelles sont difficiles à détecter visuellement.
- Les variations saisonnières et d'autres paramètres extérieurs peuvent « masquer » une tendance.
- Le public peut parfois « voir ce qui veut voir ».

Dans de nombreux cas, les méthodes statistiques d'analyse des tendances permettent de lever des ambiguïtés et de révéler des tendances difficiles à identifier par la seule analyse visuelle des données brutes.

#### Les méthodes statistiques d'analyse des tendances

Choisir d'appliquer une méthode statistique peut être très délicat de sorte qu'une certaine rigueur dans l'analyse doit être observée.

- En premier lieu, il est nécessaire d'étudier certaines caractéristiques des données :
- Distribution normale
- Changements brusques
- Cycles



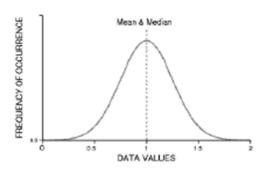


- Valeurs extrêmes
- Données manquantes
- Données < LQ (LD)
- Autocorrélation

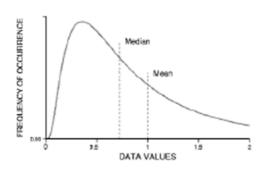
L'objectif est de sélectionner un test statistique approprié aux caractéristiques des données à traiter.

Le type de test statistique dépend des caractéristiques des données.

Pour savoir si les données d'une chronique sont normalement distribuées il est possible (i) de tracer l'histogramme des fréquences d'occurrence (ii) d'appliquer un test de normalité comme le test de Shapiro-Wilk W test.



Les tests paramétriques sont appropriés aux données à distribution normale. Ceci inclus la régression de paramètres vis-à-vis du temps ou d'une mesure spatiale. Toutefois, les tests paramétriques sont rarement appropriés aux données environnementales dont les distributions s'écartent souvent de la normalité.



Les tests non paramétriques sont indépendants de la distribution des données. Ce sont généralement les tests les plus sûrs, mais pas les plus faciles à mettre en œuvre. Ceci Inclus les tests de Kendall couplés au test de pente de Sen ainsi que les tests de CUSUM ou de Pettitt pour identifier les ruptures de pente par exemple.

• Les tests non paramétriques permettent de s'affranchir du « nettoyage » des chroniques étudiées (distribution, données extrêmes, valeurs manquantes ...).

Ces tests seront de ce fait préférés lorsque le jeu de chroniques à traiter est très important (travail à l'échelle des grands bassins hydrographiques par exemple).





#### L'autocorrélation des données

<u>Définition</u>: dépendance d'une donnée observée à un instant t à des évènements observés antérieurement.

#### Problèmes:

- Ce phénomène biaise la plupart des tests statistiques de tendance ;
- il est très présent dans les chroniques d'évolution de la qualité des eaux ;
- l'autocorrélation a plusieurs causes possibles (cycles, tendances ...) qui peuvent se superposer.

#### Solutions:

- Des tests modifiés existent pour la prise en compte de l'autocorrélation (Régression linéaire modifiée, Mann-Kendall modifié, Kendall Saisonnier)
- Possibilité de désaisonnaliser le signal (décomposition du signal)

MAIS il est impératif d'identifier les causes de l'autocorrélation pour savoir s'il est nécessaire de la « traiter ». Si la cause est la présence d'une tendance alors il ne faut pas la retirer du signal puisque c'est l'information à mettre en évidence!

Seule une connaissance fine de l'évolution du paramètre permet de statuer sur les causes de l'autocorrélation (multiplication des analyses).

#### • Quelques recommandations pour la mise en œuvre d'une analyse statistique :

- Vérifier et documenter les conditions initiales d'application des tests. Si ces conditions ne sont pas satisfaites alors les résultats peuvent être fortement biaisés.
- Le test le plus facile à mettre en œuvre, la régression linéaire par moindres carrés (OLS régression), est très souvent inapproprié aux caractéristiques des données environnementales.
- Les tests non paramétriques plus robustes, comme le test Kendall saisonnier par exemple, sont souvent non disponibles et demandent à être programmés.
- Ne pas trouver de tendance signifie seulement que les données sont insuffisantes pour en détecter une.
- L'utilisation trop rare des méthodes statistiques demande de réapprendre des procédures souvent complexes à chaque fois.





#### 4 Conclusion

S'il n'était qu'une recommandation à retenir de cette étude, elle serait l'intérêt primordial de disposer de chroniques d'évolution des concentrations en polluants dans les eaux souterraines les plus longues possibles avec des fréquences de prélèvements les plus élevées. Il est en effet nécessaire, pour bien comprendre le comportement d'un polluant dans le milieu naturel, d'avoir un recul important sur ses évolutions historiques. De plus, la connaissance de sa variabilité à différentes échelles de temps ne peut être appréhendée qu'en analysant sa concentration le plus fréquemment possible. Le choix du test statistique est ainsi grandement amélioré lorsque les processus qui guident l'évolution des concentrations du polluant étudié dans les eaux souterraines sont connus. Ceci est particulièrement vrai dans le cas de données autocorrélées. Il est en effet montré la grande difficulté de distinguer ce qui relève d'une chronique stationnaire avec des dépendances entre les observations successives, d'une séquence de données indépendantes non stationnaires affectées par une tendance. Dans le premier cas, un test statistique prenant en compte l'autocorrélation doit être mise en œuvre alors que dans le second cas, un test de stationnarité non modifié permet de mettre en évidence la tendance d'évolution.

Ce choix ne peut se faire que grâce à des considérations physiques du phénomène étudié et donc seulement si la connaissance du comportement du contaminant au point de mesure est suffisante. Cette connaissance est aussi primordiale pour l'interprétation des tendances identifiées. La recherche des facteurs explicatifs passe en effet par l'analyse couplée de tous les paramètres externes qui peuvent influencer l'évolution des concentrations du contaminant étudié. Ces paramètres peuvent être environnementaux (évolution des pluies efficaces, relations nappes rivières...), anthropiques (évolutions des pratiques culturales, pompages...) ou analytiques (changement de méthodes d'analyses...). De leur facilité d'accès dépendra la facilité et la qualité d'interprétation des tendances d'évolution des concentrations en contaminants.

Enfin, dans l'optique de faciliter et de coordonner les actions qui pourraient être menées en France pour l'analyse des évolutions des concentrations en contaminants dans les eaux souterraines, il serait judicieux que l'ensemble des acteurs nationaux appliquent une méthode unique. Il est ainsi proposé de créer un outil intégré d'analyse des chroniques d'évolution des concentrations en contaminants dans les eaux souterraines reprenant l'approche présentée, afin de le mettre à disposition des gestionnaires de la qualité des eaux souterraines après le suivi d'une formation sur l'outil et, plus globalement, sur la thématique.





#### 5 Bibliographie

DIRECTIVE 2000/60/CE du parlement européen et du conseil du 23 octobre 2000 établissant un cadre pour une politique communautaire dans le domaine de l'eau

DIRECTIVE 2006/118/CE du parlement européen et du conseil du 12 décembre 2006 sur la protection des eaux souterraines contre la pollution et la détérioration

Hamed K. H. et Rao A. R., (1998) - A modified Mann-Kendall trend test for autocorrelated data. Journal of Hydrology 204: 182-196.

Helsel D.R., Hirsch R.M., (1992) - Statistical method in water resources, Studies in Environmental Science 49, Elsevier, Amsterdam

Helsel D.R., Hirsch R.M., (2002) - Statistical Methods in Water Resources. Techniques of Water Resources Investigations of the United States Geological Survey. Book 4, Hydrologic Analysis and Interpretation, Chapter A3.

Hirsch, R.M., Slack J. R., Smith R. A., (1982) - Techniques of trend analysis for monthly water quality data. Water Resources Research 18: 107-121.

Lettenmaier, M. (1978) - Design and considerations for ambient stream quality monitoring. Water resources Bulletin, Vol. 14 (no. 4), 804-902.

Lettenmaier, D. O. (1988) - Multivariate nonparametric tests for trend in water quality, Water Resour. Bull., 24(3), 505-512.

Lopez B., Leynet A., (2011) - Evaluation des tendances d'évolution des concentrations en polluants dans les eaux souterraines. Revue des méthodes statistiques existantes et recommandations pour la mise en œuvre de la DCE. Rapport BRGM/RP-59515-FR.

Renard B. (2006) - Détection et prise en compte d'éventuels impacts du changement climatique sur les extrêmes hydrologiques en France. Thèse de l'Institut National Polytechnique de Grenoble. Unité de Recherche Hydrologie-Hydraulique, Cemagref (Lyon).





Yue S. et Wang C. Y. (2004) - The Mann-Kendall test modified by effective sample size to detect trend in serially correlated hydrological series. Water Resources Management 18: 201-218.