

Partenariat 2010 – *Domaine : Savoirs – Action n°8*



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 final

***B. LOPEZ (BRGM), A. LEYNET (BRGM),
Avec la participation de A. BLUM (BRGM) et N. BARAN (BRGM)***

Janvier 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.

BRGM/RP-59515-FR
Février 2011

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

B. Lopez, A. Leynet

Avec la participation de **A. Blum** et **N. Baran**

<p>Vérificateur :</p> <p>Nom :</p> <p>Date :</p> <p>Signature :</p>
--

<p>Approbateur :</p> <p>Nom :</p> <p>Date :</p> <p>Signature :</p>

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 Etats-membres 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-59515-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.

Rapport final

B. Lopez, A. Leynet

SOMMAIRE

1	Introduction.....	21
1.1	Objectifs de l'étude.....	24
1.2	Règlementation européenne sur l'évaluation des tendances d'évolution de la qualité des eaux souterraines	25
2	Théorie sur les tests statistiques appliqués à l'évaluation des tendances et des ruptures de pentes des chroniques temporelles	29
2.1	Définitions et concepts de base sur les tests statistiques	29
2.2	Revue des tests statistiques potentiellement applicables au domaine de la qualité des eaux naturelles	33
2.2.1	Test paramétrique de tendance monotonique	34
2.2.2	Tests non paramétriques de tendance monotonique	35
2.2.3	Tests paramétriques de rupture	39
2.2.4	Tests non paramétriques de rupture.....	43
2.2.5	Test paramétrique de dépendance des données.....	45
2.2.6	Tests non paramétriques de dépendance des données	46
2.2.7	Tests de tendance spatialisée	47
2.2.8	Tests de normalité.....	50
2.3	Cas particulier des données inférieures à une limite de quantification (LQ) ou de détection (LD).....	51
2.3.1	Ce que dit la législation	51
2.3.2	Les solutions alternatives	52
3	Revue bibliographique des travaux de détermination des tendances d'évolution de la qualité des eaux en France et à l'international.....	55
3.1	La recherche scientifique internationale sur les méthodes statistiques d'évaluation des tendances dans le domaine de la qualité des eaux.....	55
3.1.1	Développement des méthodes statistiques	55
3.1.2	Comparaison de l'efficacité des méthodes statistiques.....	57
3.1.3	Les guides méthodologiques.....	61
3.1.4	Les recherches actuelles.....	62

3.2	Exemples de détermination des tendances d'évolution de la qualité des eaux souterraines à l'international.....	64
3.3	Les études réalisées en France par les organismes gestionnaires de la qualité des eaux	71
3.3.1	Les eaux de surface.....	71
3.3.2	Les eaux souterraines.....	75
4	Caractéristiques des données de qualité des eaux souterraines : implications pour l'identification des tendances d'évolution.....	81
4.1	Variété et variabilité des données à traiter.....	82
4.2	Nombre de données, de points et longueur des chroniques.....	84
4.3	Pas de temps des prélèvements.....	86
4.4	Données hors domaine de validité.....	87
4.5	Distribution des données.....	90
4.6	Autocorrélation des données.....	92
5	Tests de robustesse et de puissance des outils statistiques les plus couramment utilisés dans le domaine de la qualité des eaux.....	99
5.1	Processus des tests de robustesse et de puissance.....	100
5.2	Nomenclature des collections.....	101
5.2.1	Modèles de calendriers et de chroniques.....	101
5.2.2	Paramètres utilisés.....	103
5.2.3	Compatibilité entre calendriers et chroniques.....	105
5.2.4	Génération des collections.....	105
5.2.5	Méthodes statistiques d'évaluation des tendances testées.....	106
5.3	Résultats.....	107
5.3.1	Collections de chroniques gaussiennes (distribution normale des données).....	107
5.3.2	Collections de chroniques à distribution log-normale des données.....	114
5.3.3	Collections de chroniques autocorrélées et variogrammes.....	120
5.4	Comparaison des tests Mann-Kendall et Kendall Saisonnier pour la détection de tendances de séries à cycles marqués.....	134
6	Recommandations pour la détermination des tendances d'évolution de la qualité des eaux souterraines.....	141
6.1	Etapes principales pour l'analyse des tendances d'évolution de la qualité des eaux souterraines.....	141
6.1.1	Validation des données.....	141
6.1.2	Analyse visuelle et exploratoire des données.....	142
6.1.3	Analyse statistique.....	144

6.1.4	Détermination des facteurs explicatifs	144
6.2	Questions préliminaires pour le choix et la mise en œuvre d'un test de tendance et de rupture	146
6.3	Arbres décisionnels pour le choix d'un test statistique de stationnarité.....	151
7	Conclusions et perspectives	157

LISTE DES ILLUSTRATIONS

Illustration 1 : Aide à la lecture d'un diagramme théorique boîte à moustache.....	48
Illustration 2 : Exemple d'application de la méthode de la distribution des données pour la prise en compte des valeurs reportées inférieures à une limite de quantification a) courbe de répartition des concentrations et loi de distribution estimée b) moyenne et médiane des données estimées suivant la loi de distribution (extrait de Helsel et Hirsch, 2002).....	53
Illustration 3 : Exemple d'application de la méthode robuste pour la prise en compte des données reportées inférieures à une limite de quantification a) régression des logs des concentrations selon les quantiles d'une distribution normale pour l'extrapolation des données inférieures à une LQ b) après transformation des données extrapolées en vraies valeurs et combinaison avec les valeurs reportées supérieures à une LQ, estimations des statistiques de la série chronologique (extrait de Helsel et Hirsch, 2002).....	54
Illustration 4 : Concentration en phosphore total, rivière Klamath, Californie (tendance = - 0,005 mg/L/an, extrait de Hirsch et al., 1982)	57
Illustration 5 : Ratio Tr/σ_e pour les chroniques avec (Significant Trends) ou sans (Non-Significant Trends) tendances significatives (extrait de Bouchard et Haemmerli, 1991).....	60
Illustration 6 : Procédure à trois étapes pour la sélection de la méthode statistique d'analyse des tendances d'évolution des contaminants dans les eaux souterraines (1) distribution normale/non normale (2) détection de tendance (3) évaluation de la pente (dans Visser et al. 2009).....	65
Illustration 7 : Exemples de tests sur des données problématiques : (a) et (b) données irrégulière dans l'aquifère de la craie, avec groupement dans le milieu de la série ; (c) et (d) données irrégulières dans l'aquifère de la craie avec groupement en fin de série ; (e) données non linéaires dans l'aquifère sableux de Sherwood ; (f) série avec changement brutal par palier ; (g) données extrêmes en milieu de séries dans l'aquifère sableux de Sherwood ; (h) données extrêmes en fin de série durant l'hivers 2000-2001 dans l'aquifère de la craie (extrait de Stuart et al., 2007).....	67
Illustration 8 : Résumé des tendances d'évolution des concentrations en nitrate dans les aquifères majeurs d'Angleterre et du Pays de Galles (cercles vides, tendances positives ; cercles pleins, tendances négatives) extrait de Stuart et al., 2007.....	68
Illustration 9 : Distribution spatiale des tendances nitrates dans les eaux souterraines du bassin de Geer, Belgique, avec les valeurs de tendances à la hausse en mg/L NO_3^- par an. (Extrait de Aguilar et al., 2007).	69
Illustration 10 : Temps (en années) extrapolés avant lesquels le seuil de 50 mg/L NO_3^- sera dépassé par les eaux souterraines. (Extrait de Aguilar et al., 2007).	70

Illustration 11 : Evolution par point de mesure du respect de l'état chimique des cours d'eau du bassin Ile-de-France (DIREN, 2006).....	73
Illustration 12 : Evolution des teneurs moyennes en nitrates des eaux superficielles du bassin Ile-de-France de 1992 à 2004 (OIEau, 2006).	73
Illustration 13 : Evolution des teneurs moyennes en nitrates des eaux souterraines du bassin Ile-de-France de 2000 à 2004 (OIEau, 2006).	74
Illustration 14 : Evolution de la teneur moyenne en nitrates de 2000 à 2004 – Réseau eaux souterraines de l'agence de l'eau Seine-Normandie (DIREN, 2006).....	75
Illustration 15 : Evolution des teneurs moyennes en nitrates des eaux souterraines entre les campagnes 2000-2001 et 2004-2005.....	76
Illustration 16 : Tendances d'évolution des teneurs en nitrate sur le bassin Loire-Bretagne et pentes des tendances (en mg/L/an) étudiées par décennie sur la période 1970-2007 suivant le test statistique de Mann-Kendall régional (extrait de Baran et al., 2009).....	77
Illustration 17 : Estimation sur le bassin Loire-Bretagne des tendances d'évolution des concentrations en nitrate par zone sur la période 1975-1993 suivant le test Kendall-régional (extrait de Baran et al., 2009).	78
Illustration 18 : Estimation sur le bassin Loire-Bretagne des tendances d'évolution des concentrations en nitrate par zone sur la période 1995-2007 suivant le test Kendall-régional (extrait de Baran et al., 2009).	79
Illustration 19 : Evolution des concentrations en nitrate au point 02507X0505/P (Pre-En-Pail, Mayenne -53-) de 1975 à 2005.	86
Illustration 20 : Evolution de la concentration en Arsenic ($\mu\text{g/L}$) au point 05008X0132/AEP (à Poncey-Les-Athee, 21).....	90
Illustration 21 : Evolution des concentrations en nitrate au point 00721X0025/F1 dans la masse d'eau 3507 à l'Est de Cherbourg.....	92
Illustration 22 : Autocorrélogramme (fonction d'autocorrélation FAC) de la chronique d'évolution des concentrations en nitrate au point 00721X0025/F1.	93
Illustration 23 : Evolution des concentrations en nitrate au point 02552X0023/AEP à Jouy, en Eure-et-Loir.	94
Illustration 24 : Autocorrélogramme (fonction d'autocorrélation FAC) de la chronique d'évolution des concentrations en nitrate au point 02552X0023/AEP.	94
Illustration 25 : Evolution de la différence entre la valeur observée d'autocorrélation et la borne supérieure de l'intervalle de confiance 95% en fonction du nombre de points dans les chroniques pour les autocorrélations du 1 ^{er} ordre (en bleu) et du 2 nd ordre (en rouge) pour a) les 100 chroniques nitrates extraites d'ADES et b) les 100 chroniques sulfates extraites d'ADES.	96

Illustration 26 : Taux de rejet de H0 selon les longueurs de chroniques à distribution de données gaussiennes sans tendance (taux de rejet attendu = 5%).	108
Illustration 27 : Taux de rejet de H0 en fonction de l'augmentation de l'écart type des données de chroniques à distribution gaussienne sans tendance (taux de rejet attendu = 5%).	109
Illustration 28 : Taux de rejet de H0 en fonction du nombre de données (effectif) des chroniques à distribution gaussienne sans tendance (taux de rejet attendu = 5%).	110
Illustration 29 : Taux de rejet de H0 en fonction du type et du bruit des calendriers pour des chroniques à distribution gaussienne sans tendance (taux de rejet attendu = 5%).	111
Illustration 30 : Taux de rejet de H0 en fonction de la valeur de la pente a de la tendance pour des chroniques à distribution gaussienne + tendance (taux de rejet attendu = 100%).	112
Illustration 31 : Taux de rejet de H0 selon les longueurs de chroniques à distribution de données Log-normale + tendance (taux de rejet attendu = 100%).	115
Illustration 32 : Taux de rejet de H0 en fonction de l'augmentation de l'écart type des données de chroniques à distribution Log-normale + tendance (taux de rejet attendu = 100%).	117
Illustration 33 : Taux de rejet de H0 en fonction du nombre de données (effectif) des chroniques à distribution Log-normale + tendance (taux de rejet attendu = 100%).	118
Illustration 34 : Taux de rejet de H0 en fonction de la valeur de la pente a de la tendance pour des chroniques à distribution Log-normale + tendance (taux de rejet attendu = 100%).	119
Illustration 35 : Comparaison du taux de rejet de H0 sous l'effet de l'augmentation de la composante pépite (=diminution de l'autocorrélation) pour un calendrier régulier R et troué T dans le cas de chroniques cycliques annuelles sans tendance V0 (taux de rejet attendu = 5%).	122
Illustration 36 : Comparaison du taux de rejet de H0 en fonction de la longueur des chroniques cycliques annuelles sans tendance (modèle V0) pour un calendrier régulier R et troué T (taux de rejet attendu = 5%).	124
Illustration 37 : Comparaison du taux de rejet de H0 sous l'effet de l'augmentation de l'écart type des données de chroniques cycliques annuelles sans tendance (modèle V0) pour un calendrier régulier R et troué T (taux de rejet attendu = 5%).	126
Illustration 38 : Comparaison du taux de rejet de H0 sous l'effet de l'augmentation du nombre de données de chroniques cycliques annuelles sans tendance (modèle V0) pour un calendrier régulier R et troué T (taux de rejet attendu = 5%).	127

Illustration 39 : Comparaison du taux de rejet de H0 sous l'effet de l'augmentation de la pente des tendances de chroniques cycliques annuelles avec tendance (modèle Va) pour un calendrier régulier R et troué T (taux de rejet attendu = 100%).128

Illustration 40 : Comparaison de l'évolution du taux de rejet de H0 pour des chroniques à cycles annuels sans tendance (modèle V0) et à cycles de 10 ans sans tendance (modèle W0) en fonction de l'augmentation de la composante pépite des variogrammes (taux de rejet attendu = 5%).....130

Illustration 41 : Comparaison de l'évolution du taux de rejet de H0 pour des chroniques à cycles annuels sans tendance (modèle V0) et à cycles de 10 ans sans tendance (modèle W0) en fonction de la longueur des chroniques (taux de rejet attendu = 5%).131

Illustration 42 : Comparaison de l'évolution du taux de rejet de H0 pour des chroniques à cycles annuels sans tendance (modèle V0) et à cycles de 10 ans sans tendance (modèle W0) en fonction de la fréquence de mesure des données en calendrier régulier (en haut) et en calendrier troué (en bas, taux de rejet attendu = 5%)132

Illustration 43 : Comparaison du taux de rejet de H0 des méthodes Mann-Kendall et Kendall-Seasonal dans le cas de chroniques cycliques annuelles à calendriers réguliers journaliers sous l'effet de la diminution de la pente des tendances (taux de rejet attendu = 100%).136

Illustration 44 : Chroniques cycliques annuelles à tendance linéaire de 0,5/an simulées selon 12 fréquences de prélèvements différentes.....137

Illustration 45 : Comparaison du taux de rejet de H0 des tests de Mann-Kendall (MK) et Kendall-Seasonal (KS) sur des chroniques cycliques annuelles à tendances linéaires de pentes a en fonction de la fréquence de prélèvement (taux de rejet attendu = 100%).138

Illustration 46 : Schéma de validation des données brutes. 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 [...]).152

Illustration 47 : 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.153

Illustration 48 : 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.154

LISTE DES TABLEAUX

Tableau 1 : Valeurs critiques de Z du test de Buishand40

Tableau 2 : Valeurs critique de Q/\sqrt{n} pour le test de déviation cumulative.....41

Tableau 3 : Valeurs critiques de W du test du ratio de Worsley Likelihood	42
Tableau 4 : Approches recommandées pour la détection des tendances et leur extrapolation en contexte d'aquifère simple et complexe (dans Visser et al. 2009).....	66
Tableau 5 : Nombre de données, nombre de chroniques, nombre moyen de données par chronique et longueur moyenne des chroniques pour 7 paramètres extraits d'ADES en septembre 2010.	84
Tableau 6 : Nombre et pourcentage de données de concentration de l'atrazine, de l'isoproturon, du trichloroéthylène, de l'arsenic et de la carbamazépine bancarisées dans ADES (septembre 2010) dans le domaine de validité, inférieures à la limite de détection et en-dessous des seuils de quantification	88
Tableau 7 : Tests de normalité des données issues de 100 chroniques de concentration en nitrate et de 100 chroniques de concentration en sulfate prises aléatoirement dans ADES en septembre 2010.	91
Tableau 8 : Autocorrélation du 1 ^{er} et 2 nd ordre des données issues de 100 chroniques de concentration en nitrate et de 100 chroniques de concentration en sulfate extraites d'ADES	95
Tableau 9 : Modèles de calendrier	102
Tableau 10 : Modèles de chroniques.....	102
Tableau 11 : Paramètres des modèles de calendriers et de chroniques.....	104
Tableau 12 : Méthodes de génération des calendriers selon les lois de distribution des chroniques.....	105
Tableau 13 : Méthodes statistiques d'évaluation de tendance testées.....	106

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

Rapport final

B. Lopez, A. Leynet

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 : Tendances, DCE, tests statistiques, qualité, eaux souterraines

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

Rapport final

B. Lopez, A. Leynet.

Synthèse pour l'action opérationnelle

La Directive 2006/118/CE (GWD) sur la protection des eaux souterraines contre la pollution et la détérioration, adoptée le 12 décembre 2006, est composée de 3 grands axes :

- La définition du bon état des eaux souterraines avec l'article 3 qui définit les « Critères pour l'évaluation de l'état chimique des eaux souterraines », et l'article 4 sur la « Procédure d'évaluation de l'état chimique des eaux souterraines ».
- L'analyse de l'évolution de la chimie des eaux souterraines au travers de l'article 5 sur l' « Identification des tendances à la hausse significatives et durables et définition des points de départ des inversions de tendance »,
- L'identification des actions de réduction des pollutions mentionnées dans l'article 6 sur les « Mesures de prévention ou de limitation des introductions de polluants dans les eaux souterraines ».

L'étude s'inscrit dans le deuxième axe de la directive, c'est-à-dire l'article 17 de la DCE, et l'article 5 de la directive fille 2006/118/CE sur les eaux souterraines qui en découle. Dans le détail, la Directive exige que les « Etats membres identifient les tendances significatives à la hausse et durables des concentrations de polluants, groupes de polluants ou d'indicateurs de pollution observées dans les masses ou groupes de masses d'eau souterraine identifiés comme à risque et identifient le point de départ de l'inversion de ces tendances ».

L'annexe IV.A de la directive fille développe la méthode à utiliser pour l'identification d'une tendance à la hausse significative et durable sur un point de surveillance. Le point 2.c explique que l'évaluation doit « être basée sur une méthode statistique, par exemple la technique de la régression, pour l'analyse des tendances temporelles dans des séries chronologiques de points de surveillance distincts ».

Mais la Directive n'est pas plus explicite sur la méthode à utiliser. A défaut de cadrage, les districts sont donc libres de retenir la méthode statistique la plus appropriée à chaque masse d'eau (en fonction notamment des données disponibles et du modèle conceptuel de la masse d'eau). Par défaut, une régression linéaire sera appliquée.

Afin d'accompagner la mise en œuvre de la DCE, l'étude se propose de faire l'inventaire des tests statistiques de détection des tendances pouvant être facilement mis en œuvre sur des données de qualités des eaux souterraines et de préconiser la méthode la plus robuste en fonction des données

disponibles et de l'objectif à atteindre. Il est ainsi proposé aux gestionnaires de la qualité des eaux souterraines une revue des tests de tendance et de rupture existants et de définir une procédure pour choisir la (ou les) méthode(s) la (les) plus pertinente(s) en fonction de l'objectif à atteindre, du type et de la qualité des données à traiter.

De nombreux tests statistiques d'analyse des chroniques temporelles sont présentés et détaillés dans le rapport. Ils sont choisis en raison de leur potentialité à être appliqués (sous certaines conditions) sur des chroniques d'évolution de la qualité des eaux souterraines, et en raison de leur relative facilité de compréhension et de mise en œuvre. Une revue bibliographique détaillée retrace l'historique des développements et des applications des méthodes statistiques tant dans le domaine de la qualité des eaux superficielles que des eaux souterraines. Les travaux de recherche mettent en évidence les limites d'application de certaines méthodes et insistent sur les caractéristiques des chroniques qui peuvent influencer les résultats des tests statistiques. Les biais qui peuvent perturber le bon fonctionnement des tests sont engendrés par certaines caractéristiques des séries chronologiques :

- nombre de données disponibles dans la chronique,
- distribution non normale des données,
- cycles marqués dans l'évolution des données,
- changements abrupts,
- données aberrantes dans la série,
- données manquantes,
- autocorrélation des données d'une série.

Partant de ce constat les caractéristiques des données de qualité des eaux souterraines françaises sont analysées. Les données de 6 paramètres à risque de non atteinte du bon état chimique + 1 paramètre nouvellement recherché* sont extraites d'ADES :

- 2 ions majeurs (Nitrate NO_3^- et Sulfate SO_4^{2-})
- 2 pesticides, (Atrazine de la famille des triazines et Isoproturon de la famille des urées substituées)
- 1 substance industrielle (Trichloroéthylène de la famille des solvants organiques chlorés)
- 1 élément trace (Arsenic)
- *1 médicament (Carbamazépine de la famille des antiépileptiques)

Les données sont triées et analysées afin d'extraire leurs caractéristiques « critiques » face à l'application des méthodes statistiques pour l'étude des séries temporelles. Il est ainsi montré, en fonction du type de paramètre, une grande variabilité du nombre de données par chronique, des longueurs moyennes des séries, des fréquences de prélèvements et des fréquences d'apparition de données inférieures aux limites de quantification. De plus, une étude de la distribution des données ainsi que des phénomènes d'autocorrélation des données révèle qu'un nombre important de chroniques d'évolution de la qualité des eaux souterraines extraites d'ADES montre des données non

normalement distribuées (40 % environ des chroniques NO₃ par exemple) et que le processus d'autocorrélation n'est pas un phénomène marginal.

D'une manière générale, l'analyse montre qu'il est important de prendre le temps de bien analyser les chroniques mises à disposition avant l'application d'un test statistique. Toutes les caractéristiques critiques détaillées dans le rapport peuvent en effet influencer le choix des tests à appliquer et leur connaissance peut aussi aider à identifier les facteurs qui pourraient expliquer les tendances observées.

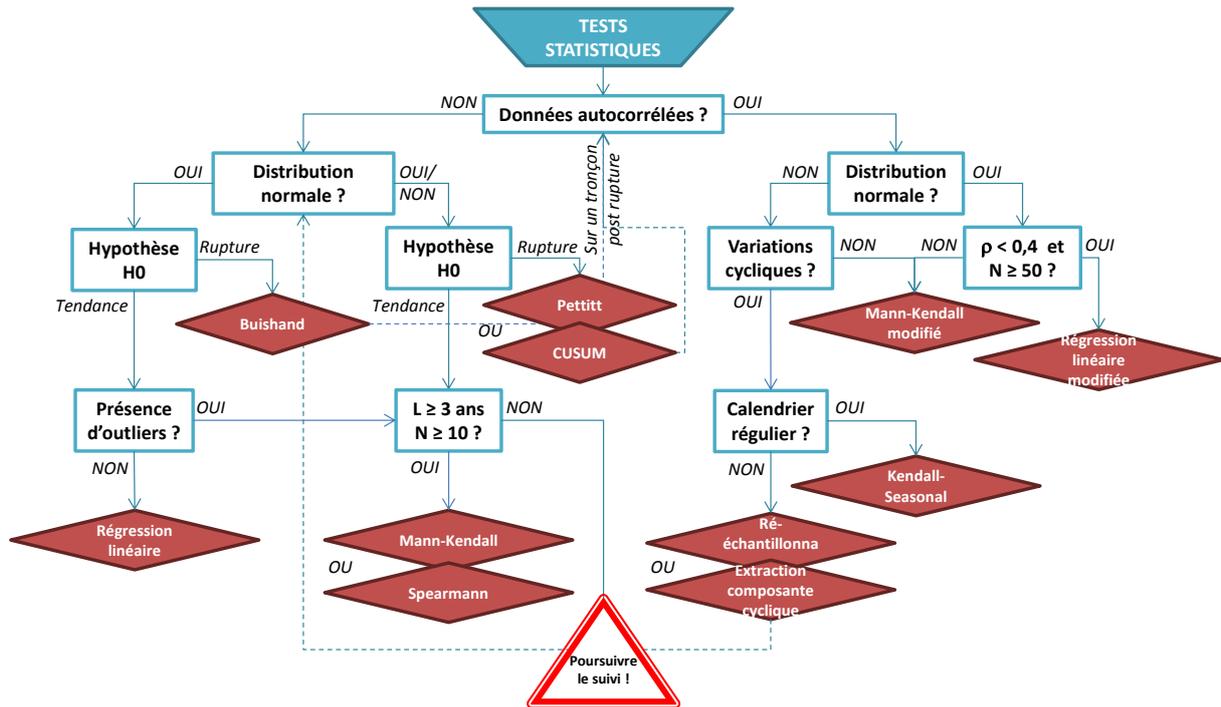
Par la suite, afin de rendre compte de la robustesse (risque de première espèce) des tests les plus couramment utilisés pour l'évaluation des tendances d'évolution de la qualité des eaux face aux caractéristiques des données, il est proposé d'identifier les méthodes qui ont la propriété de rester non biaisées lorsque les observations ne respectent plus les conditions initiales d'applicabilité du test. Parallèlement, les méthodes couramment employées sont testées afin de rendre compte de leur puissance à détecter à raison des tendances dans les séries à caractéristiques « critiques ».

Parmi les tests rencontrés dans la littérature, la régression linéaire par moindres carrés, les tests de Mann-Kendall et Mann-Kendall modifié, le test de Spearman, le test de Cox-Stuart et, dans un second temps le test Kendall Saisonnier, sont sélectionnés pour l'évaluation de leur robustesse et de leur puissance face aux différents caractères critiques des chroniques d'évolution de la qualité des eaux souterraines identifiés lors de l'analyse des données vraies extraites d'ADES. Il est impossible de sélectionner, parmi l'ensemble des données de qualité des eaux souterraines disponibles dans ADES, des jeux de chroniques suffisamment importants pour être significatifs d'un point de vue statistique et spécifiques chacun d'une caractéristique critique pouvant biaiser les tests statistiques. Les séries temporelles sur lesquelles les tests sont réalisés sont donc simulées en reprenant les caractéristiques réelles mises en évidence après analyse des données d'ADES.

Les résultats des tests de robustesse et de puissances des méthodes statistiques couplés aux résultats mentionnés dans la littérature scientifique permettent de faire des recommandations sur la marche à suivre pour l'analyse statistique des chroniques d'évolution des concentrations en polluants dans les eaux souterraines. Il est ainsi proposé de travailler en quatre phases :

- 1) validation des données
- 2) analyse exploratoire des séries chronologiques
- 3) mise en œuvre des tests statistiques
- 4) recherche des facteurs explicatifs

Les deux premières phases correspondent à la description des données et à l'étude du comportement du paramètre étudié en fonction du temps. La troisième phase correspond au choix du test statistique à appliquer en fonction de la qualité des données à traiter. Des arbres décisionnels guident les gestionnaires dans leurs travaux d'analyse statistique. A titre d'exemple, le schéma de sélection des tests statistiques à appliquer en fonction de la qualité des données est reporté dans cette synthèse.



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

Chaque « branche » du diagramme trouve une justification dans le rapport, soit issue de la littérature scientifique, soit des tests de robustesse et de puissance réalisés spécifiquement avec des caractéristiques de données de qualité des eaux souterraines en France. Des pistes sont aussi données pour identifier les facteurs qui expliquent les tendances observées.

Enfin, les conclusions de l'étude peuvent aider à la révision des programmes de surveillance de la qualité des eaux souterraines en indiquant les fréquences de prélèvements, les nombres de données et les longueurs optimums des chroniques pour une application robuste et puissance des méthodes statistiques.

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

Rapport final

B. Lopez, A. Leynet.

1 Introduction

L'un des objectifs environnementaux de la Directive cadre sur l'eau 2000/60/CE vis-à-vis des eaux souterraines est la mise en œuvre de mesures « pour inverser toute tendance à la hausse, significative et durable, de la concentration de tout polluant résultant de l'activité humaine » (article 4.1.b.iii).

Mais au préalable, la directive fille sur les eaux souterraines 2006/118/CE exige que les Etats membres « identifient les tendances à la hausse [...] et définissent le point de départ de l'inversion de ces tendances » (article 5.1). Bien que cette même directive recommande de baser l'évaluation de la tendance « sur une méthode statistique, par exemple la régression », il appartient aux Etats membres de définir la méthode la plus appropriée pour cette évaluation. Le guide¹ européen sur l'évaluation de l'état et des tendances le confirme. Aucune méthode particulière n'est imposée.

La préparation du premier plan de gestion rapporté à la Commission Européenne au 22 mars 2010 a mis en évidence des difficultés pour répondre aux exigences de l'article 5 de la directive fille sur les eaux souterraines (Blum et Chariot, 2011). La principale difficulté rencontrée est liée aux données. En effet, les chroniques disponibles d'une masse d'eau à l'autre ont des caractéristiques différentes (longueur, fréquence de mesures, performances analytiques variables, outliers...) en fonction du type de molécules analysées et de l'évolution des capacités analytiques des laboratoires. Il n'existe d'autre part pas d'approche méthodologique unique. La diversité des données et des contextes hydrogéologiques impose donc la mise en œuvre d'approches méthodologiques différentes suivant le type de

¹ European Commission (2009), Guidance on groundwater status and trends, Common Implementation Strategy for the Water Framework Directive 2000/60/CE, Guidance document n°18, 82p.

polluant avec un objectif commun d'identification des tendances d'évolution et des inversions si elles existent.

Il apparaît donc aujourd'hui nécessaire de :

- **faire le point sur les méthodes statistiques existantes pour la détermination des tendances d'évolution de la qualité des eaux souterraines,**
- **de mettre à disposition des gestionnaires des outils d'aide à la décision leur permettant de sélectionner la méthode la plus adéquate en fonction des données disponibles sur chaque masse d'eau, du type de polluant (élément majeur, micropolluants organiques, micropolluants minéraux), du contexte hydrogéologique et des objectifs de gestion.**

L'étude se déroule en trois phases qui correspondent chacune à des objectifs ciblés auxquels il est nécessaire de répondre chronologiquement pour fournir un outil robuste d'aide au traitement des données brutes :

La première phase est une revue bibliographique des outils statistiques d'analyse des chroniques temporelles. Les méthodes statistiques d'identification et de quantification des tendances et des ruptures de pentes potentiellement applicables au domaine de la qualité des eaux souterraines sont listées. Pour étayer la constitution de cette liste, une revue des recherches scientifiques menées sur le développement et l'évaluation de ces méthodes est réalisée. Par la suite, les approches de détermination des tendances d'évolution de la qualité des eaux déjà utilisées dans le domaine de l'eau superficielle et souterraine en France et à l'international sont présentées.

La deuxième phase de l'étude concerne les potentialités d'application des outils de détermination des tendances à la problématique de l'évolution de la qualité des eaux souterraines. Des molécules sur lesquelles une attention particulière doit être portée, identifiées par les bassins comme responsables d'un risque de non atteinte du bon état, sont sélectionnées et extraites de la base de données nationale sur l'eau souterraine ADES. Le domaine «qualité des eaux souterraines » présente des spécificités qu'il est nécessaire de bien identifier. Le choix final des tests à appliquer est en effet dépendant pour partie des caractéristiques des données d'entrée dont notamment :

- la distribution des données,
- les fluctuations saisonnières (spatiales et/ou temporelles) des concentrations,

- les phénomènes d'autocorrélation,
- les valeurs manquantes dans les séries,
- les données reportées inférieures aux seuils de quantification.

Afin de compléter la bibliographie sur l'évaluation de l'efficacité des méthodes statistiques à détecter des tendances, des tests de robustesse et de puissance vis-à-vis des caractéristiques spécifiques aux données de qualité des eaux souterraines de certaines méthodes les plus couramment utilisées sont effectués.

La troisième et dernière phase de l'étude consiste en l'établissement de recommandations pour mener à bien une analyse complète des tendances d'évolution des données de qualité des eaux souterraines. Sur la base des recherches scientifiques relevées dans la littérature spécialisée et en fonction des résultats des tests de robustesse et de puissance menés spécifiquement dans un contexte « qualité des eaux souterraines », des méthodes statistiques sont recommandées en fonction des caractéristiques des données à traiter et des objectifs de gestion. L'arbre décisionnel établi en fin de rapport constitue un outil d'aide à la décision qui intègre l'ensemble des travaux bibliographiques et d'évaluation des méthodes réalisés pour l'étude.

1.1 Objectifs de l'étude

L'étude est menée dans le but de faciliter l'application de la réglementation européenne sur l'analyse de l'évolution de la qualité des eaux souterraines. La Directive 2006/118/CE sur la protection des eaux souterraines contre la pollution et la détérioration, adoptée le 12 décembre 2006, prévoit en effet d'évaluer l'état chimique des eaux souterraines mais aussi d'identifier les tendances d'évolution des concentrations en polluants. Toutefois, la procédure d'identification des tendances et notamment le type de test à appliquer n'est pas clairement défini dans la Directive, la recommandation se limitant à l'utilisation d'une « méthode statistique pertinente ». Il est donc proposé aux gestionnaires de la qualité des eaux souterraines une revue des tests de tendance et de rupture existants et de définir une procédure pour choisir la (ou les) méthode(s) la (les) plus pertinente(s) en fonction de l'objectif à atteindre, du type et de la qualité des données à traiter.

La revue et l'analyse des travaux déjà réalisés pour déterminer les tendances d'évolution dans les domaines liés à l'eau (hydrologie, qualité des eaux superficielles et souterraines) doit permettre de lister et de choisir les outils les plus pertinents, c'est-à-dire à même de répondre le plus justement aux questions des gestionnaires. Les tests identifiés n'étant pas spécifiques du domaine « qualité des eaux souterraines », il est important d'analyser, en premier lieu, les spécificités de ce domaine d'étude et de relever les caractéristiques générales des chroniques qui peuvent influencer l'application des tests statistiques. La transposition de certains tests statistiques au domaine de la qualité des eaux souterraines nécessite alors de contrôler la puissance et la robustesse de ces méthodes face à certains « paramètres critiques » des chroniques d'évolution de la qualité des eaux souterraines (nombre de prélèvements, pas de temps de mesure, amplitude des variations...). Il est en effet important de comprendre les limites d'application des méthodes statistiques disponibles pour choisir le test le plus pertinent en fonction des données à traiter et de l'objectif à atteindre.

L'application d'un test statistique ne constituant qu'une étape dans l'analyse des chroniques d'évolution de la chimie des eaux souterraines, le champ de l'étude est élargi à l'ensemble de la procédure d'analyse des chroniques, de la validation des données collectées à l'identification des facteurs explicatifs, en passant par l'analyse visuelle et l'application des tests statistiques. L'étude a en effet pour but d'accompagner pas à pas les gestionnaires confrontés à l'analyse des données historiques d'évolution de la chimie des eaux

souterraines, en émettant un certain nombre de recommandations à la fois pour choisir les tests pertinents pour l'identification des tendances et des ruptures de pentes, mais aussi et plus largement, pour mener à bien l'analyse complète de l'évolution des données de qualité des eaux souterraines.

L'analyse des tendances d'évolution des chroniques de qualité des eaux souterraines est généralement conduite par point de prélèvement, cet objectif constituant une partie importante du projet. Toutefois, il est également demandé aux gestionnaires de compléter cette détermination ponctuelle de tendances par une analyse spatialisée dont l'échelle peut être la masse d'eau souterraine ou tout autre secteur à limites justifiées. Des travaux récents ont été menés en ce sens en France et à l'international par l'application de méthodes statistiques développées spécifiquement ou en cours de développement pour répondre à cet objectif. Ce volet est ainsi abordé tout au long de l'étude, soit par de exemples de travaux, soit par des recommandations pour la mise en œuvre d'études spatialisées des tendances.

1.2 Règlements européenne sur l'évaluation des tendances d'évolution de la qualité des eaux souterraines

La Directive 2006/118/CE (GWD) sur la protection des eaux souterraines contre la pollution et la détérioration, adoptée le 12 décembre 2006, est composée de 3 grands axes :

- La définition du bon état des eaux souterraines avec l'article 3 qui définit les « Critères pour l'évaluation de l'état chimique des eaux souterraines », et l'article 4 sur la « Procédure d'évaluation de l'état chimique des eaux souterraines ».
- L'analyse de l'évolution de la chimie des eaux souterraines au travers de l'article 5 sur l' « Identification des tendances à la hausse significatives et durables et définition des points de départ des inversions de tendance »,
- L'identification des actions de réduction des pollutions mentionnées dans l'article 6 sur les « Mesures de prévention ou de limitation des introductions de polluants dans les eaux souterraines ».

L'étude s'inscrit dans le deuxième axe de la directive, c'est-à-dire l'article 17 de la DCE, et l'article 5 de la directive fille 2006/118/CE sur les eaux souterraines qui en découle. Dans le détail, la Directive exige que les « Etats membres identifient les tendances à la hausse significatives et durables des concentrations de polluants, groupes de polluants ou d'indicateurs de pollution observées dans les masses ou groupes de masses d'eau

souterraine identifiés comme à risque et identifient le point de départ de l'inversion de ces tendances ».

L'identification des tendances telle qu'elle est demandée dans l'article 17 de la DCE est un outil visant à évaluer l'objectif fixé à l'article 1.b à savoir « inverser toute tendance à la hausse significative et durable, de la concentration de tout polluant résultant de l'impact de l'activité humaine afin de réduire progressivement la pollution des eaux souterraines ».

L'article 5.2 de la GWD le confirme : « les États membres inversent les tendances qui présentent un risque significatif d'atteinte à la qualité des écosystèmes aquatiques ou terrestres, à la santé humaine ou aux utilisations légitimes, qu'elles soient réelles ou potentielles, de l'environnement aquatique au moyen du programme de mesures visé à l'article 11 de la directive 2000/60/CE, afin de réduire progressivement la pollution des eaux souterraines et de prévenir la détérioration de l'état de celles-ci ».

D'après l'article 5.1 de la GWD, « les États membres identifient les tendances à la hausse significatives et durables des concentrations de polluants, groupes de polluants ou d'indicateurs de pollution observées dans les masses ou groupes de masses d'eau souterraine identifiés comme étant à risque et définissent le point de départ de l'inversion de ces tendances, conformément à l'annexe IV ».

Comme cela est évoqué dans l'article 2.6 et dans l'annexe IV.A.3, cette évaluation s'accompagne de l'évaluation du « point de départ de l'identification ».

Les Etats membres doivent donc évaluer 3 éléments :

- le point de départ de l'identification (ou « valeur initiale de l'identification »),
- l'identification de tendances à la hausse significatives et durables,
- le calcul des points d'inversion de tendances.

L'annexe IV.A de la directive fille développe la méthode à utiliser pour l'identification d'une tendance à la hausse significative et durable sur un point de surveillance. Le point 2.c explique que l'évaluation doit « être basée sur une **méthode statistique**, par exemple la technique de la régression, pour l'analyse des tendances temporelles dans des séries chronologiques de points de surveillance distincts ».

Mais la GWD n'est pas plus explicite sur la méthode à utiliser. Quant au guide européen, il ne donne pas non plus d'autre méthode si ce n'est celle déjà employée en 2001 dans le cadre du projet GWSTAT (Grath et al, 2001). Cette méthode, qui avait été testée en France sur la nappe de la Craie du Nord et sur la nappe d'Alsace (Larpin et al., 2003) n'est applicable que pour les macropolluants et avec un nombre de données suffisants (en

fréquence et en densité de points) et généralement incompatibles avec les caractéristiques des réseaux RCS (Réseau de Contrôle et de Surveillance) et RCO (Réseau de Contrôle Opérationnel) français.

A défaut de cadrage, les districts sont donc libres de retenir la méthode statistique la plus appropriée à chaque masse d'eau (en fonction notamment des données disponibles et du modèle conceptuel de la masse d'eau). Par défaut, une régression linéaire sera appliquée.

Afin d'accompagner la mise en œuvre de la DCE, l'étude se propose de faire l'inventaire des tests statistiques de détection des tendances facilement mis en œuvre sur des données de qualités des eaux souterraines et de préconiser la méthode la plus robuste en fonction des données disponibles. Les conclusions de l'étude doivent aussi aider à la révision des programmes de surveillance de la qualité des eaux souterraines en indiquant les pas de temps de prélèvements, les nombres de données et les longueurs optimums des chroniques pour une application robuste et puissance des méthodes statistiques.

2 Théorie sur les tests statistiques appliqués à l'évaluation des tendances et des ruptures de pentes des chroniques temporelles

Dans ce chapitre, les tests statistiques appliqués à l'étude des séries chronologiques sont présentés et détaillés. Il s'agit des tests les plus couramment utilisés dans le domaine. Le parti est pris de limiter la présentation aux seuls tests statistiques comme le préconise la Directive Cadre Européenne, les méthodes plus poussées d'étude du signal, d'ajustement théoriques des chroniques ou de modélisation ne constituant pas l'objet de ce rapport. Avant de détailler les tests statistiques, les concepts théoriques et les termes spécifiques au domaine sont rappelés.

2.1 Définitions et concepts de base sur les tests statistiques

La stationnarité (d'une chronique) : Une des grandes questions dans l'étude de séries temporelles (ou chronologiques) est de savoir si celles-ci suivent un processus stationnaire. On entend par là le fait que la structure du processus sous-jacent supposé évolue ou non avec le temps. Si la structure reste la même, le processus est alors stationnaire. Autrement dit, une chronique stationnaire aura toujours le même comportement que l'on regarde à la date t ou à la date $t+k$. Par exemple, une chronique dont la moyenne n'évolue pas avec le temps est stationnaire. Il en est de même si la variance et/ou l'autocorrélation n'évoluent pas avec le temps. Les causes de non stationnarité peuvent être :

- soit une modification graduelle de la série au cours du temps, qui se manifeste par une tendance à la baisse (ou à la hausse) des valeurs de la série,
- soit une (ou plus) d'une rupture survenant à une certaine date, les caractéristiques de la série n'étant plus les mêmes à partir de la date de rupture,
- soit un changement dans la distribution de probabilité d'une variable caractéristique à partir d'une date donnée.
-

Tester la stationnarité d'une chronique sera donc par exemple de rechercher la présence d'une tendance. En effet, par définition, **une tendance** (au sens mathématique) correspond à une certaine orientation prise par les valeurs d'une série de données en fonction du temps, c'est-à-dire à une évolution du processus observé en fonction du temps.

L'hypothèse nulle : Le point de départ pour la mise en œuvre d'un test statistique est la définition d'une hypothèse dite nulle (H_0) et de son alternative (H_a). Dans le cas d'un test de tendance d'une série chronologique, H_0 est l'absence de tendance dans la série de données, et H_a est la présence d'une tendance croissance ou décroissante. Dans le cas d'un test de rupture de pente, l'hypothèse H_0 est l'absence de rupture, H_a correspondant à la présence d'une rupture de pente dans la série de données.

Le test statistique : Le test statistique est le moyen de comparer les deux hypothèses H_0 et H_a . C'est une valeur numérique calculée à partir de la série de données testée (on calcule la valeur d'une certaine variable appelée la « statistique » du test). Le test permet de choisir, parmi les deux hypothèses émises, celle qui a le plus de chances d'être vraie. Toutefois, en statistique, les hypothèses ne sont jamais « acceptées ». Il est préféré l'expression « on ne peut rejeter l'hypothèse H_0 à un niveau de confiance α » ou bien, dans le cas de l'alternative, « on doit rejeter l'hypothèse H_0 et retenir l'hypothèse H_a à un niveau de confiance α ».

Niveau de confiance : Le niveau de confiance (α), ou niveau de signification, est un seuil de probabilité donné, comparé à la valeur calculée de la "statistique" du test pour savoir si l'écart observé est compatible avec l'hypothèse nulle ou non. On appelle **p-value** le niveau de signification le plus bas où l'hypothèse nulle peut être rejetée. Ainsi donc si $p\text{-value} > \alpha$ alors l'hypothèse nulle ne peut être rejetée.

Le niveau de confiance correspond à la probabilité qu'a la valeur du test statistique d'égaliser ou de dépasser la valeur observée lorsque l'hypothèse H_0 (pas de tendance) est vraie. Par exemple, pour $\alpha = 0,05$ la valeur critique du test statistique est celle qui serait dépassée par 5 % des valeurs des tests statistiques réalisés sur des données aléatoires. Si la valeur du test statistique est plus importante que la valeur critique, alors l'hypothèse H_0 est rejetée.

Le niveau de confiance permet de rendre compte de la probabilité pour un test de détecter une tendance (rejet de H_0) alors qu'aucune n'est effectivement présente (erreur de Type I ou erreur de première espèce).

Ainsi le niveau de confiance peut être interprété comme suit :

$\alpha > 0,1$	faible preuve contre l'hypothèse H_0
$0,05 < \alpha < 0,1$	preuve possible contre l'hypothèse H_0
$0,01 < \alpha < 0,05$	forte preuve contre l'hypothèse H_0

$\alpha < 0,01$ très forte preuve contre l'hypothèse H_0

Pour l'essentiel des méthodes statistiques classiques, les valeurs critiques des tests statistiques sont reportées dans des tables statistiques ou calculées pour toutes valeurs de niveaux de confiance. Lorsque les séries statistiques testées ne satisfont pas les conditions initiales des tests statistiques (autocorrélation, distribution normale...) des tests spécifiques doivent être utilisés. Il est par exemple possible d'utiliser des méthodes de ré-échantillonnage pour estimer le niveau de confiance du test statistique.

Le calcul de la "statistique" du test (c'est-à-dire la façon dont elle est construite) renvoie soit à une valeur dont la probabilité est $\alpha/2$ soit à une valeur dont la probabilité est α . Suivant ce qui est recherché, suivant l'hypothèse H_0 émise et suivant la loi de probabilité du test, il est alors possible de définir une région d'acceptation et 2 régions de rejet de part et d'autre de la région d'acceptation de H_0 ("two tailed test") ou bien une région d'acceptation et une seule région de rejet à l'une ou l'autre extrémité de la loi de distribution ("one tailed test"). Ce sont donc les valeurs de la loi de probabilité qui diffèrent mais non le niveau de signification, $+S_{\alpha/2}$ dans un cas, $+S'_{\alpha}$ ou $-S'_{\alpha}$ dans l'autre cas (si distribution symétrique).

Erreur du test statistique : Il existe deux types possibles d'erreurs. Une erreur de Type I, ou risque de première espèce, qui apparaît lorsque l'hypothèse H_0 est incorrectement rejetée. Une erreur de type II, ou risque de seconde espèce, exprime le fait que l'hypothèse H_0 n'est pas rejetée alors que l'hypothèse H_a est vraie. Un test statistique avec une faible erreur de type II est dit puissant.

Robustesse du test statistique : La première qualité d'un test statistique est de respecter le risque de première espèce (erreur de type 1). Or, les valeurs critiques des tests ne sont en général valables que sous certaines conditions d'application (nombre de données, distribution...). Si ces contraintes ne sont pas respectées, le test pourra avoir tendance à rejeter l'hypothèse H_0 trop souvent (test dit « libéral ») ou au contraire trop rarement (test dit « conservatif »). Les tests conservatifs sont généralement préférés aux tests libéraux car le risque de première espèce sera en réalité plus faible que celui spécifié (au détriment de la puissance définie ci-après). Si le risque reste stable même en dehors des conditions d'applications requises, le test sera dit robuste.

Puissance du test statistique : C'est la probabilité de rejeter à raison l'hypothèse H_0 . Dans le cadre de la détection de non-stationnarités il s'agit de la probabilité de détecter

correctement une tendance ou une rupture. Cette probabilité dépend d'un certain nombre de facteurs comme la taille de l'échantillon, l'importance du changement à détecter, la position de la rupture, la distribution de la série à traiter, etc.

Tests paramétriques et tests non-paramétriques.

Les tests paramétriques présument que les données des séries chronologiques ainsi que leurs erreurs (écarts à la tendance par exemple) suivent une distribution particulière (généralement une distribution normale). Cette condition doit être remplie pour pouvoir appliquer ces tests. Ils sont utiles car ils permettent de quantifier les changements dans les données (amplitude du changement de moyenne ou gradient de tendance) et sont généralement plus puissants que les tests non-paramétriques. Lorsque l'hypothèse de distribution normale des données n'est pas satisfaite, il est parfois possible de procéder à un ré-échantillonnage pour estimer le niveau de confiance ou les valeurs critiques du test statistique pour différents niveaux de confiance.

Les tests non-paramétriques ne nécessitent pas d'hypothèse distribution des données des séries chronologiques. Ces tests permettent de détecter des tendances ou des changements mais ne sont généralement pas appropriés pour quantifier les pentes des tendances ou l'amplitude des changements. Ils sont toutefois très utiles dans le domaine des sciences de l'eau car beaucoup de séries hydrologiques, hydrogéologiques ou hydrochimiques ne suivent pas une loi de distribution normale. Quelques tests non-paramétriques sont utilisés conjointement avec des méthodes permettant de quantifier les phénomènes (tendances ou ruptures) identifiés.

Autocorrélation des données : correspond à la dépendance d'une donnée observée à un instant t à des événements observés antérieurement. L'autocorrélation peut être estimée par le calcul de l'autocorrélogramme de la série temporelle qui correspond à la corrélation croisée d'un signal par lui-même pour différents décalages de temps. L'autocorrélation permet de détecter des régularités, des profils répétés dans un signal comme un signal périodique perturbé par beaucoup de bruit, ou bien une fréquence fondamentale d'un signal qui ne contient pas effectivement cette fondamentale, mais l'implique avec plusieurs de ses harmoniques.

2.2 Revue des tests statistiques potentiellement applicables au domaine de la qualité des eaux naturelles

Dans ce chapitre sont listés les tests statistiques les plus couramment rencontrés dans la littérature pour détecter des tendances, des changements (ruptures de pente) ou pour décrire le caractère plus ou moins aléatoire de séries chronologiques. En effet, toute série chronologique peut n'être qu'une « réalisation » particulière d'un phénomène complexe générant un nombre substantiel de séries différentes. Les méthodes statistiques dont il est fait état ici concernent essentiellement l'exploitation d'une série de données et une seule. Cette analyse est qualifiée de ponctuelle ou "par site" et s'impose bien sûr avant de procéder à des interprétations prenant en compte la dimension spatiale des phénomènes générateurs des séries dites ponctuelles. 3 méthodes d'évaluation des tendances de manière spatialisée sont néanmoins présentées en fin de paragraphe.

Les tests statistiques détaillés dans cette partie sont choisis en raison de leur potentialité à être appliqués (sous certaines conditions) sur des chroniques d'évolution de la qualité des eaux souterraines, et en raison de leur relative facilité de mise en œuvre. L'historique des développements et des applications de ces méthodes statistiques est présenté au chapitre 3.

En premier lieu, il est nécessaire de définir les hypothèses H_0 telles qu'elles sont testées par les méthodes statistiques dédiées à cette tâche :

- L'hypothèse nulle H_0 pour les tests de tendance (ou de stationnarité) est qu'il n'y a pas de tendance significative dans les données.
- L'hypothèse nulle H_0 pour les tests de saut de moyenne/médiane ou de différence dans les moyennes/médianes (tests de rupture de pente) est qu'il n'y a pas de saut ou de différence significative entre les moyennes/médianes de deux portions d'une même série chronologique de données.
- L'hypothèse nulle H_0 pour les tests de dépendance des données (test d'autocorrélation) est que les données résultent d'un processus aléatoire et qu'elles sont indépendantes entre elles.

2.2.1 Test paramétrique de tendance monotonique

• **La Régression linéaire** est un test paramétrique pour les données à distribution normale.

Il faut aussi que les "erreurs" soient :

- distribuées suivant une loi normale (ce qui découle de l'hypothèse de la distribution normale des données de la série),
- de moyenne nulle,
- non corrélées entre elles.

Elle permet de tester s'il existe une tendance linéaire par l'examen des relations entre le temps (x) et la variable considérée (y).

La pente a de la régression est estimée selon la relation :

$$a = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

et l'ordonnée à l'origine b est estimée comme suit :

$$b = \bar{y} - a \cdot \bar{x}$$

Le test statistique S est :

$$S = a / \sigma$$

où
$$\sigma = \sqrt{\frac{12 \cdot \sum_{i=1}^n (y_i - b - ax_i)^2}{n \cdot (n-2) \cdot (n^2-1)}}$$

Cette expression n'est vraie que si la séquence temps est régulière ($t, t+1, t+2, \dots$).

Sinon :

$$\sigma^2 = \frac{1}{n-2} \frac{\sum_{i=1}^n (y_i - (at_i + b))^2}{\sum_{i=1}^n (t_i - \bar{t})^2}$$

Le test statistique S suit une distribution de Student à $n-2$ degrés de liberté sous l'hypothèse nulle H_0 (pas de tendance significative). Les valeurs critiques du test statistique pour

différents niveaux de confiance peuvent être obtenus à partir des tables statistiques de Student-t.

Remarque : la régression linéaire présume une distribution normale des données et fait l'hypothèse que les erreurs (écarts à la tendance) sont indépendantes et suivent la même distribution normale avec une moyenne des erreurs égale à 0.

• **La régression linéaire modifiée** a été développée pour une application de la régression linéaire dans le cas où les données présentent de l'autocorrélation. La modification repose une estimation de la variance du test prenant en compte le coefficient d'autocorrélation. Le détail des étapes du calcul de la régression linéaire modifiées est présenté par Renard (2006).

2.2.2 Tests non paramétriques de tendance monotonique

• **La méthode de Mann-Kendall** permet de tester s'il existe une tendance monotonique dans une série de données chronologiques. Ce test est non paramétrique et ne demande donc pas de vérification de la distribution des données.

Le principe du test est le remplacement des n valeurs de la série temporelle ($X_1, X_2, X_3, \dots, X_{n-1}, X_n$) par leur rang relatif ($R_1, R_2, R_3, \dots, R_{n-1}, R_n$).

Le test statistique S est alors le suivant :

$$S = \sum_{i=1}^{n-1} \left[\sum_{j=i+1}^n \text{sgn}(R_j - R_i) \right]$$

où

$$\left| \begin{array}{l} \text{sgn}(x) = 1 \text{ pour } x > 0 \\ \text{sgn}(x) = 0 \text{ pour } x = 0 \\ \text{sgn}(x) = -1 \text{ pour } x < 0 \end{array} \right.$$

Si l'hypothèse nulle H_0 est vraie, S approche d'une distribution normale avec les valeur de moyenne μ et de variance σ^2 respectivement :

$$\mu = 0$$

$$\sigma^2 = n \cdot (n-1) \cdot (2n+5)/18$$

La statistique z est alors de la forme (les valeurs critiques du test statistique pour différents niveau de confiance sont reportées dans les tables de probabilités de distribution normale) :

$$z = |S| / \sigma^{0.5}$$

S'il y a des valeurs égales dans la série, la variance σ^2 doit être calculée comme suit :

$$\sigma^2 = \frac{1}{18} \left[n(n-1)(2n+5) - \sum_{p=1}^q E_p(E_p-1)(2E_p+5) \right]$$

où E_p désigne le nombre d'égalités identifiées dans la série impliquant p valeurs.

Si le nombre n de valeurs de la série est supérieur à 10, la mise en œuvre du test fait intervenir la variable aléatoire Z définie par :

$$Z = \frac{S-1}{\sqrt{\text{Var}(S)}} \quad \text{si } S > 0$$

$$Z = \frac{S+1}{\sqrt{\text{Var}(S)}} \quad \text{si } S < 0$$

$$Z = 0 \quad \text{si } S=0$$

Cette variable suit (approximativement) une loi normale centrée réduite. La probabilité d'observer une tendance pour laquelle la valeur de Z est au moins égale en valeur absolue à Z_{obs} est alors :

$$P(|Z| \geq |z_{\text{obs}}|) = 2(1 - \Phi(|z_{\text{obs}}|))$$

Une valeur positive de S indique une tendance à l'augmentation et vice versa.

La pente de la tendance et l'intervalle sont estimés en utilisant **la méthode de la ligne robuste de Kendall-Theil** (ou méthode de Sen ; Helsel and Hirsch, 1992). La pente β est calculée comme la médiane de toutes les $\frac{n(n-1)}{2}$ pentes entre chaque paire de données

(X, Y) :

$$\beta = \frac{(Y_j - Y_i)}{(X_j - X_i)} \quad \text{pour tout } i < j \text{ et } i = 1, 2, \dots, (n-1) \text{ et } j = 2, 3, \dots, n$$

Le coefficient de corrélation τ (Tau) du test Kendall est une mesure robuste de la corrélation avec une bonne résistance aux biais dus à quelques valeurs exceptionnelles ou incohérentes. Le coefficient de corrélation τ (Tau) égale :

$$\tau = \frac{S}{n(n-1)/2}$$

• **La méthode de Mann-Kendall modifiée.** Comme il est évoqué plus loin au paragraphe 4.6, la présence d'autocorrélation dans les données peut affecter sérieusement la puissance des tests statistiques en engendrant une surestimation de la significativité statistique des tendances. Dans les données de qualité des eaux naturelles, cette autocorrélation est généralement engendrée par une cyclicité périodique dans l'évolution des données traduisant une dépendance à des phénomènes externes (variation de la recharge, cycles d'applications). Une approche complémentaire au test de Mann-Kendall classique est ainsi proposée afin de prendre en compte ce phénomène d'autocorrélation. Le principe repose sur une modification du test S de Mann-Kendal plutôt que de modifier les données elles-mêmes :

$$Var_{\rho}(S) = \gamma \cdot Var_{\rho=0}(S)$$

où γ est un facteur correctif appliqué à la variance.

L'autocorrélation positive des données retire de l'information sur la série qui se comporte alors comme une série de taille inférieure. Deux méthodes sont relevées dans la littérature.

- Yue et Wang (2004) proposent de corriger le test de Mann-Kendall de la façon suivante :

$$\gamma = 1 + 2 \frac{\rho_1^{n+1} - n\rho_1^2 + (n-1)\rho_1}{n(\rho_1 - 1)^2}$$

où ρ_1 désigne l'autocorrélation d'ordre 1.

- Hamed et Rao (1998) proposent quant à eux une formule empirique spécifiquement calculée pour corriger la statistique de Mann-Kendall :

$$\gamma = 1 + \frac{2}{n(n-1)(n-2)} \sum_{k=1}^{n-1} (n-k)(n-k-1)(n-k-2)\rho_k$$

Dans le cas d'un processus autorégressif d'ordre 1 AR(1) il est possible de simplifier l'équation en supposant $\rho_k = \rho_1^k$, avec k ordre de l'autocorrélation.

Les tests réalisés par Renard (2006) sur la puissance de ces méthodes à rejeter H_0 dans le cas de chroniques cycliques autocorrélées avec tendances montrent que la modification proposée par Hamed et Rao est légèrement meilleure sous l'hypothèse AR(1) que la formule de Yue et Wang.

- **La méthode Kendall Saisonnier** (Kendall-Seasonal en anglais) est une autre variante du test Mann-Kendall utilisée pour l'estimation des tendances de séries cycliques saisonnières autocorrélées.

Le principe, identique à celui du test de Mann-Kendall, est basé sur la classification des données les unes par rapport aux autres en fonction de leur rang dans la série. Toutefois, dans le cas du test de Mann-Kendall avec saisonnalité, le caractère saisonnier de la série est pris en compte. Autrement dit pour des données mensuelles ayant une saisonnalité de 12 mois, on ne va pas chercher à savoir s'il y a une croissance au global sur la série, mais simplement si d'un mois de janvier à l'autre, d'un mois de février à l'autre, et ainsi de suite, il y a une tendance.

Pour ce test, l'ensemble des τ (Tau) de Kendall pour chaque saison est calculé en premier, permettant le calcul du τ de Kendall moyen pour la série. La variance de la statistique peut être calculée en faisant l'hypothèse que les séries sont indépendantes (par exemple les valeurs des mois de janvier et des mois de février sont indépendantes) ou dépendantes, ce qui requiert le calcul de covariances.

Pour le calcul de la p-value de ce test, une approximation normale pour la distribution de la moyenne des τ de Kendall est généralement utilisée.

- **Le test de Cox-Stuart** est un test non paramétrique de recherche de tendance. Il est souvent considéré comme peu puissant, mais robuste pour une recherche de tendance.

Le principe repose sur une division de la chronique de taille N en deux parts en sa moitié (négligeant la valeur centrale le cas échéant). Les paires de données $(X_i ; X_{i+N/2})$ sont créées et les termes constitutifs d'une paire sont comparés. Si $X_{i+N/2} > X_i$ le signe « + » est assigné à la paire. A l'inverse, si $X_{i+N/2} < X_i$ le signe « - » est affecté à la paire.

Sous l'hypothèse nulle H_0 d'absence de tendance, le nombre de paires « + », noté T, suit une distribution binomiale \mathcal{B} équilibrée ($p = 1/2$). Il devient donc possible de calculer la probabilité π d'avoir N succès pour $N^- + N^+$ épreuves. L'hypothèse nulle est rejetée (non-stationnarité) si π est inférieure au risque acceptable $\alpha = 5\%$.

• **Le test du Rho de Spearman** est basé sur le classement des valeurs afin de déterminer s'il les corrélations entre deux variables sont significatives. Pour l'analyse des tendances, la première variable correspond au temps, la seconde étant constituée par les données de la série temporelle.

A l'instar du test de Mann-Kendall, les n valeurs de la série temporelles sont remplacées par leur rang.

Le test statistique ρ_s correspond au coefficient de corrélation entre les deux variables. Il est calculé de la même manière qu'un coefficient de corrélation classique en substituant les valeurs brutes par leur rang :

$$\rho_s = S_{xy} / \sqrt{(S_x \cdot S_y)}$$

où

$$\left\{ \begin{array}{l} S_x = \sum_{i=1}^n (x_i - \bar{x})^2 \\ S_y = \sum_{i=1}^n (y_i - \bar{y})^2 \\ S_{xy} = \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y}) \end{array} \right.$$

et x_i (temps), y_i (donnée de qualité), \bar{x} et \bar{y} réfèrent aux rangs des données brutes.

Pour une série de grande taille, la quantité $\rho_s \sqrt{n-1}$ approche d'une distribution normale avec une moyenne de 0 et une variance égale à 1. Ainsi, les valeurs critiques du test statistique peuvent être obtenues par la consultation des tables de probabilité de distribution normale.

2.2.3 Tests paramétriques de rupture

Tous les tests paramétriques de ruptures supposent que les données de la série chronologique sont normalement distribuées

• **Le test de Buishand** permet de déterminer s'il existe une rupture, un saut significatif dans les moyennes.

Le test est construit à partir des écarts cumulés à la moyenne jusqu'à un certain rang k en définissant la variable suivante :

$$U(k) = \sum_{i=1}^k (x_i - \bar{x})$$

Sous l'hypothèse H_0 (pas de rupture) l'écart cumulé devrait osciller autour de 0, sans atteindre de trop grandes valeurs. A l'inverse, une valeur trop importante en un point de rang k peut être le signe d'une rupture à cet endroit.

La valeur critique Z du test statistique est :

$$Z = \max_{k=1, \dots, n} \left(\frac{|U(k)|}{\sqrt{n \cdot \sigma_x}} \right)$$

Les valeurs critiques Z pour différents niveaux de confiance et différentes tailles d'échantillon sont données dans le Tableau 1.

Tableau 1 : Valeurs critiques de Z du test de Buishand

N	Z au niveau de confiance		
	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
10	1.05	1.14	1.29
20	1.10	1.22	1.42
30	1.12	1.24	1.46
40	1.13	1.26	1.50
50	1.14	1.27	1.52
100	1.17	1.29	1.55
∞	1.22	1.36	1.63

L'instant de rupture \hat{k} peut être estimé par la relation :

$$\hat{k} = \text{Arg} \max_{k=1, \dots, n} \left(\frac{|U(k)|}{\sqrt{n \cdot \sigma_x}} \right)$$

Remarque : le test de Buishand peut, sous certaines conditions, être utilisé pour évaluer la stationnarité d'une série chronologique.

• **La Déviation cumulative** permet de tester si les moyennes de deux parties d'une série chronologique sont significativement différentes (pour une date de changement inconnue).

Le principe du test est de détecter un changement dans la moyenne de la série après m observations :

$$E(x_i) = \mu \quad i = 1, 2, 3, \dots, m$$

$$E(x_i) = \mu + \Delta \quad i = m+1, m+2, \dots, n$$

où μ est la moyenne des données antérieures au changement et Δ le changement dans la moyenne.

A partir des moyennes, les déviations cumulatives sont calculées selon :

$$S_0^* = 0 \quad S_k^* = \sum_{i=1}^k (x_i - \bar{x}) \quad k = 1, 2, 3, \dots, n$$

Et les sommes partielles ajustées recalculées sont obtenues par la division des valeurs S_k^* par la déviation standard :

$$S_k^{**} = S_k^* / D_x$$

$$D_x^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n}$$

Le test statistique Q est :

$$Q = \max |S_k^{**}|$$

Tableau 2 : Valeurs critique de Q/\sqrt{n} pour le test de déviation cumulative

N	Q/\sqrt{n} au niveau de confiance		
	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
10	1.05	1.14	1.29
20	1.10	1.22	1.42
30	1.12	1.24	1.46
40	1.13	1.26	1.50
50	1.14	1.27	1.52
100	1.17	1.29	1.55
∞	1.22	1.36	1.63

Il est calculé pour chaque année, la valeur maximale indiquant le point de changement.

Les valeurs critiques de Q/\sqrt{n} sont données par le Tableau 2. Une valeur négative de S_k^* indique que la moyenne des données de la partie la plus récente de la série chronologique est significativement plus élevée que celle de la partie ancienne et inversement.

• **Le Ratio de Worsley Likelihood** teste, comme précédemment, si les moyennes de deux parties d'une série chronologique sont significativement différentes (pour une date de changement inconnue).

Le principe du test est identique à la déviation cumulative mais les valeurs S_k^* sont pondérées suivant leur position dans la série chronologique.

$$Z_k^* = [k(n-k)]^{-0,5} \cdot S_k^*$$

$$Z_k^{**} = Z_k^* / D_x$$

Le test statistique W est :

$$W = \frac{(n-2)^{0,5} \cdot V}{(1-V^2)^{0,5}} \quad \text{où} \quad V = \max |Z_k^{**}|$$

Les valeurs critiques de W sont données dans le Tableau 3. Une valeur négative de W indique que la moyenne des données de la partie la plus récente de la série chronologique est significativement plus élevée que celle de la partie ancienne et inversement.

Tableau 3 : Valeurs critiques de W du test du ratio de Worsley Likelihood

N	W au niveau de confiance		
	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
10	3.14	3.66	4.93
15	2.97	3.36	4.32
20	2.90	3.28	4.13
25	2.89	3.23	3.94
30	2.86	3.19	3.86
35	2.88	3.21	3.87
40	2.88	3.17	3.77
45	2.86	3.18	3.79
50	2.87	3.16	3.79

• **Le t de Student** permet de tester si les moyennes de deux périodes d'une série chronologique sont significativement différentes.

Le test statistique t de Student est :

$$t = \frac{(\bar{x} - \bar{y})}{S \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

où \bar{x} et \bar{y} sont les moyennes de la première et de la seconde période de la série chronologique respectivement, m et n sont le nombre d'observations dans la première et la seconde période respectivement, et S est l'écart type (de la totalité des m et n observations).

Les valeurs critiques du test statistique pour différents niveaux de confiance peuvent être obtenues à partir d'une table statistique du t de Student.

2.2.4 Tests non paramétriques de rupture

- **La libre distribution CUSUM** (ou somme cumulée) permet de tester si les moyennes de deux parties d'une série chronologique sont différentes (pour une date de changement inconnue).

Etant donné une série chronologique $(x_1, x_2, x_3, \dots, x_n)$, le test statistique est défini par :

$$V_k = \sum_{i=1}^k \text{sgn}(x_i - x_{median}) \quad k = 1, 2, 3, \dots, n$$

où

$$\begin{cases} \text{sgn}(x) = 1 \text{ pour } x > 0 \\ \text{sgn}(x) = 0 \text{ pour } x = 0 \\ \text{sgn}(x) = -1 \text{ pour } x < 0 \end{cases}$$

x_{median} est la valeur médiane de la série x_i

La distribution de V_k suit une loi statistique de Kolmogorov-Smirnov à deux échantillons

($KS = (2/n) \cdot \max|V_k|$) avec les valeurs critiques de $\max|V_k|$ données par :

$\alpha = 0.10$	$1.22\sqrt{n}$
$\alpha = 0.05$	$1.36\sqrt{n}$
$\alpha = 0.01$	$1.63\sqrt{n}$

Une valeur négative de V_k indique que la dernière partie de la série chronologique (partie la plus récente) a une moyenne plus grande que la partie plus ancienne et inversement.

Remarque : le test CUSUM peut, sous certaines conditions, être utilisé pour évaluer la stationnarité d'une série chronologique.

• **Le test de Pettitt** est à rapprocher, dans sa construction, au test de Mann-Kendall mais il est plutôt dédié à la détection d'une rupture que d'une tendance. Le principe de ce test non paramétrique est d'évaluer si les couples de valeurs ont tendance à être rangés préférentiellement dans un sens, les couples étant formés par sélection d'une valeur de chaque côté d'un point de rupture hypothétique k . k est un entier compris entre 1 et $n-1$. Le test statistique de Pettitt $U(k)$ est de la forme :

$$U(k) = \sum_{i=1}^k \sum_{j=k+1}^n \text{signe}(X_j - X_i)$$

La statistique Z de ce test vaut :

$$Z = \max_{k=1, \dots, n-1} |U(k)|$$

La fonction de répartition de Z peut-être approchée asymptotiquement par :

$$F(z) = P(Z \leq z) = 1 - 2 \exp\left(-\frac{6z^2}{n^3 + n^2}\right)$$

Le test de Pettitt permet en outre de faire une estimation du point de rupture \hat{k} telle que :

$$\hat{k} = \underset{k=1, \dots, n-1}{\text{Arg max}} |U(k)|$$

• **Le test de Mann-Whitney-Wilcoxon** (aussi appelé test de la Somme de rang de Wilcoxon ou test de Mann-Whitney) est un test non paramétrique qui permet d'estimer si les médianes de périodes différentes d'une même série chronologique sont significativement différentes.

Ce test s'applique en suivant différentes étapes :

- Classement des données de la valeur la plus petite (rang 1) à la valeur la plus grande (rang N). Dans le cas de valeurs égales, la moyenne des rangs de ces valeurs est utilisée.

- Réalisation d'un test S (Cf. test de Mann-Kendall, 2.2.2) comme une somme des rangs des observations dans le plus petit groupe (le nombre d'observations dans le plus petit groupe est noté n , celui du plus grand groupe est noté m).

- Calcul de la moyenne théorique μ et de l'écart type σ de S sous l'hypothèse H_0 pour la totalité des données de la chronique tel que :

$$\mu = n(N + 1) / 2$$

$$\sigma = \sqrt{[n.m.(N + 1) / 12]}$$

- Calcul de la forme standardisée du test statistique Z_{rs} tel que :

$$Z_{rs} = (S - 0,5 - \mu) / \sigma \quad \text{si } S > \mu$$

$$Z_{rs} = 0 \quad \text{si } S = \mu$$

$$Z_{rs} = |S + 0,5 + \mu| / \sigma \quad \text{si } S < \mu$$

Z_{rs} suit approximativement une loi normale de sorte que les valeurs critiques du test statistique pour différents niveaux de confiance peuvent être obtenues en consultant une table de probabilités de loi normale.

2.2.5 Test paramétrique de dépendance des données

• **Le test d'Autocorrélation** permet d'estimer dans quelle mesure une valeur issue d'une série chronologique est dépendante des valeurs précédentes.

Un moyen d'estimer l'autocorrélation d'un signal est de calculer la corrélation croisée du signal par lui-même. Ainsi, le coefficient d'autocorrélation d'ordre 1 r_1 (qui exprime la dépendance d'une donnée avec la première donnée précédente) est :

$$r_1 = \frac{\left[\sum_{i=1}^{n-1} (x_i - \bar{x}) \cdot (x_{i+1} - \bar{x}) \right]}{\left[\sum_{i=1}^n (x_i - \bar{x})^2 \right]}$$

Si la série de donnée provient d'un processus aléatoire, les valeurs attendues de variance et de r_1 sont :

$$E(r_1) = -1 / n$$

$$Var(r_1) = (n^3 - 3n^2 + 4) / [n^2(n^2 - 1)]$$

La valeur critique du test statistique z est alors :

$$z = |r_1 - E(r_1)| / \text{var}(r_1)^{0,5}$$

Les valeurs critiques pour différents niveaux de confiance peuvent être obtenues à partir d'une table de probabilités de loi normale.

2.2.6 Tests non paramétriques de dépendance des données

• **Le Croisement de médianes** est un test où les n valeurs de la série chronologique sont remplacées par 0 si $x_i < x_{median}$ et par 1 si $x_i > x_{median}$.

Si les données de la chronique sont issues d'un processus aléatoire, alors m (le nombre de fois où une valeur 0 est suivie par 1 ou bien la valeur 1 est suivie par 0) est approximativement distribué selon une loi normale avec :

$$\mu = (n - 1) / 2$$

$$\sigma = (n - 1) / 4$$

La valeur critique du test statistique z est alors :

$$z = |m - \mu| / \sigma^{0,5}$$

Les valeurs critiques pour différents niveaux de confiance peuvent être obtenues à partir d'une table de probabilités de loi normale.

• **Le test des Points tournants** nécessite le remplacement des n valeurs de la série chronologique par 1 si $x_{i-1} < x_i > x_{i+1}$ ou $x_{i-1} > x_i < x_{i+1}$ et par 0 dans tous les autres cas.

Le nombre de fois où la valeur 1 apparaît (m^*) est approximativement normalement distribuée de sorte que :

$$\mu = 2(n - 2) / 3$$

$$\sigma = (16n - 29) / 90$$

La valeur critique du test statistique z est alors :

$$z = |m^* - \mu| / \sigma^{0,5}$$

Les valeurs critiques pour différents niveaux de confiance peuvent être obtenues à partir d'une table de probabilités de loi normale.

• **Le test de Rang de différence** est calculé en remplaçant les n valeurs de la série chronologique par leurs rangs respectifs qui varient de 1 pour la plus faible valeur à n pour la plus élevée.

Le test statistique U correspond à la somme des différences des rangs en valeur absolue entre les rangs successifs :

$$U = \sum_{i=2}^n |R_i - R_{i-1}|$$

Pour un échantillon de grande taille (n important), les valeurs de U suivent une loi normale telle que :

$$\mu = (n+1).(n-1)/3$$

$$\sigma = (n-2).(n+1).(4n-7)/90$$

La valeur critique du test statistique z est alors :

$$z = |U - \mu| / \sigma^{0,5}$$

Les valeurs critiques pour différents niveaux de confiance peuvent être obtenues à partir d'une table de probabilités de loi normale.

2.2.7 Tests de tendance spatialisée

• **La Boîte à moustache** est un moyen rapide de figurer le profil essentiel d'une série statistique quantitative. Elle a été inventée en 1977 par John Tukey, mais peut faire l'objet de certains aménagements selon les utilisateurs. Ce diagramme (Illustration 1) résume quelques caractéristiques de position du caractère étudié (médiane, quartiles, minimum, maximum et outliers). Il est utilisé principalement pour comparer un même caractère dans deux populations différentes. Il s'agit de tracer un rectangle allant du premier quartile (0,25) au troisième quartile (0,75) et coupé par la médiane. Ce rectangle suffit pour le diagramme en boîte. On ajoute alors des segments (les moustaches) aux extrémités de la boîte menant jusqu'à la valeur minimum de la série vers le bas et maximum vers le haut.

Les valeurs extrêmes (outliers en anglais) exclues de l'analyse statistique sont représentées en points individuels sur le même diagramme. Un outlier est un point dont la valeur est inférieure à $QL - 1,5 \cdot IQR$ ou supérieure à $QU + 1,5 \cdot IQR$ avec :

IQR : différence entre les quartiles

QL : valeur du premier quartile (bas de la boîte)

QU : valeur du dernier quartile (haut de la boîte)

« Boite à moustache » est le nom courant, mais ce type de représentation peut prendre le nom de diagramme à pattes ou bien de Box plot en anglais.

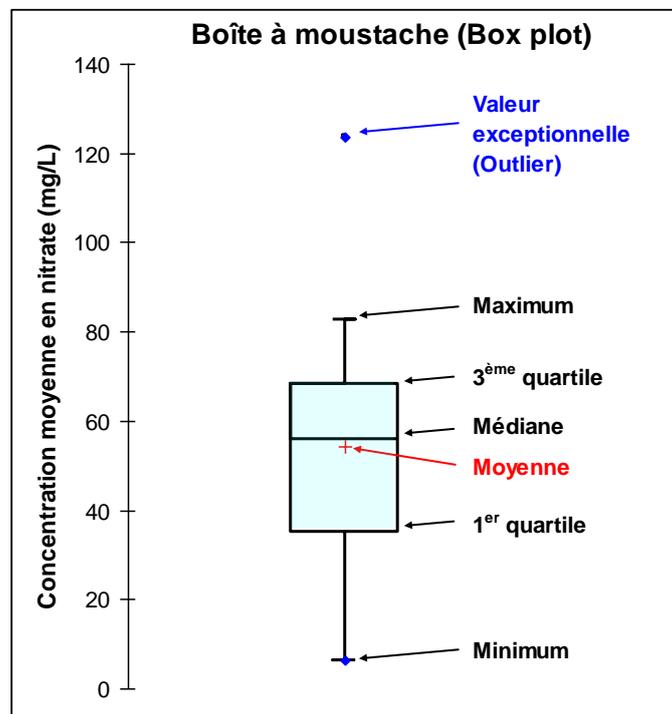


Illustration 1 : Aide à la lecture d'un diagramme théorique boîte à moustache

Dans l'exemple théorique présenté dans l'illustration 1 :

- Minimum = 6,2 mg/L
- Maximum = 82,7 mg/L
- Outlier = 123,5 mg/L
- Moyenne = 54,3 mg/L
- Médiane = 56 mg/L
- 1^{er} quartile = 35,2 mg/L
- 3^{ème} quartile = 68,5 mg/L

La boîte à moustache est un outil tout à fait approprié pour représenter l'évolution d'un paramètre de la qualité des eaux sur un secteur entre deux périodes choisies. Le caractère spatial de l'analyse de tendance est rendu possible par l'agrégation de plusieurs points d'un même secteur pour la constitution de la boîte à moustache. La comparaison de l'état qualitatif de la « population » pour deux périodes différentes permet d'estimer la tendance d'évolution de l'état qualitatif du secteur entre les deux périodes. Un test de Wilcoxon permet alors d'estimer si la différence entre les deux périodes est significative à un seuil de significativité α donné.

- **Le test Kendall-Régional** permet de déterminer des tendances spatialisées en calculant les mêmes paramètres que pour la méthode Mann-Kendall ou Kendall-Seasonal. Toutefois, les paramètres (Tau, S et z) sont calculés sur un nuage de points formé par le regroupement des données acquises en chacun des points de mesure de la zone à renseigner. Ce nuage de points constitue une chronique d'un « qualitomètre virtuel » caractéristique de la zone étudiée. Une fois la chronique virtuelle constituée, l'application du test Kendall-Regional est en tout point similaire à celle du Mann-Kendall classique ou du test Kendall-Seasonal.

- **La Méthode du Vecteur Régional** est un modèle simple proposé par Hiez G. (1977) qui permet de représenter les informations (pluviométriques à l'origine) d'une région par une séquence d'indices annuels, représentatifs des évènements dans cette région, et par des coefficients caractéristiques de chaque poste d'observation. La méthode a été conçue pour répondre à deux objectifs : d'une part la critique de l'information et d'autre part sa synthèse. Les tendances régionales peuvent être ainsi étudiées par l'Analyse en Composantes Principales.

Le principe de la méthode repose sur la représentation de l'information de base par une matrice de données où chaque colonne représente un lieu et chaque ligne une date d'observation. Les données collectées, sans compter les manques d'observation, souffrent généralement d'erreurs diverses. Toutefois, dans la mesure où le réseau de points de mesures est suffisamment dense pour garantir un certain degré de redondance et en supposant qu'il existe une tendance régionale prédominante, la matrice des données, considérée dans son ensemble, est susceptible de fournir de meilleures informations que n'importe quelle station prise isolément.

Pour ce faire, l'hypothèse de "quasi-linéarité" des totaux pluviométriques entre stations voisines est formulée. La matrice des données est considérée comme une matrice dégénérée, c'est-à-dire comme le produit matriciel d'un vecteur-colonne (L) ou vecteur

régional (série chronologique), par un vecteur-ligne (C), constitué des coefficients de proportionnalité de chaque station.

Ces vecteurs sont extraits de la matrice originale au moyen d'un processus itératif de recherche des valeurs normalisées les plus fréquentes, appliqué successivement aux lignes et aux colonnes de cette dernière. La méthode comprend également une étape facultative de comparaison de chaque station avec les données du vecteur régional ainsi élaboré, ce qui permet éventuellement d'identifier les périodes durant lesquelles la séquence des observations se comporte différemment de la série des indices qui forment le vecteur. Le champ de vecteur régionaux est ensuite analysé en Composante Principale (ACP) pour en extraire les facteurs communs éventuels et permettre, si possible, une représentation spatiale de l'influence de ces facteurs dans les différentes régions.

Cette méthode, historiquement utilisée uniquement sur des données de pluviométrie est actuellement en cours de test par l'Agence de l'eau Rhin-Meuse pour une application sur des données de qualité des eaux souterraines.

Remarque sur les tests de tendance spatialisée : La difficulté de mise en œuvre de toutes les méthodes statistiques d'évaluation des tendances par région provient de la définition de la zone à renseigner et donc du choix des différentes chroniques à regrouper. Dans le domaine de la qualité des eaux souterraines, l'exercice de sectorisation consiste à créer des unités spatiales homogènes d'un point de vue lithologique, climatique, du comportement hydrodynamique, du comportement hydrochimique ainsi qu'en termes de pression. Il est souvent difficile d'accéder à l'ensemble de ces données de sorte qu'il peut être judicieux de rechercher des indices intégrateurs de plusieurs paramètres qui influent sur l'homogénéité des secteurs à créer.

2.2.8 Tests de normalité

D'autres méthodes statistiques d'analyses des séries temporelles existent sans toutefois entrer dans le champ de la présente étude. Il est toutefois nécessaire de citer 4 tests de normalité utilisés spécifiquement dans le chapitre 4, paragraphe 4.5 du rapport :

- le **test de Shapiro-Wilk** bien adapté aux échantillons de moins de 5000 observations,
- le **test d'Anderson-Darling** proposé par Stephens (1974) qui est une modification du test de Kolmogorov-Smirnov adaptée à plusieurs lois dont la loi normale, pour le cas où les paramètres de la loi ne sont pas connus et doivent donc être estimés,

- le **test de Lilliefors** qui est aussi une modification du test de Kolmogorov-Smirnov adapté au cas de la normalité dans le cas où les paramètres de la loi, la moyenne et la variance, ne sont pas connus et doivent donc être estimés,
- le **test de Jarque-Bera** qui est d'autant plus performant que le nombre de données est important.

Afin de vérifier visuellement si un échantillon suit une loi normale, il est possible d'utiliser les graphiques P-P et les graphiques Q-Q : les graphiques P-P = Probabilité-Probabilité (P-P plots en anglais) permettent de comparer la fonction de répartition empirique d'un échantillon à celle d'un échantillon distribué suivant une loi normale de même moyenne et même variance. Si l'échantillon suit une loi normale, les points doivent être confondus avec la première bissectrice du plan. Les graphiques Q-Q = Quantile-Quantile (Q-Q plots en anglais) permettent de comparer les quantiles de l'échantillon à ceux d'un échantillon distribué suivant une loi normale de même moyenne et même variance. Si l'échantillon suit une loi normale, les points doivent être confondus avec la première bissectrice du plan.

2.3 Cas particulier des données inférieures à une limite de quantification (LQ) ou de détection (LD)

2.3.1 Ce que dit la législation

L'annexe IV.2.d de la GWD (Ground Water Directive) recommande, pour l'évaluation de la tendance, que les résultats <LQ soient remplacés par LQ/2 pour les substances individuelles et par zéro pour le total pesticides. Ce choix est d'autant plus justifié qu'il est cohérent avec les préconisations de la future directive de la Commission Européenne sur les spécifications techniques pour l'analyse chimique et la surveillance de l'état des eaux (Commission Européenne, 2009).

Dans le cas des **micropolluants**, conformément au point 2.d de l'annexe IV de la GWD, la valeur de LQ/2 « la plus élevée de toutes les séries temporelles est affectée à toutes les mesures inférieures à la limite de quantification, sauf pour le total des pesticides ».

L'article 5 de cette directive, qui concerne toutes les catégories d'eau couvertes par la DCE (eaux de surface, eaux souterraines, eaux marines), est consacré au calcul des valeurs moyennes. Le paragraphe 1 précise ainsi que « lorsque les quantités de mesures physico-chimiques ou chimiques dans un échantillon donné sont inférieures à la limite de

quantification, les résultats de mesures devront être établis comme la **moitié de la valeur de la limite de quantification concernée** pour le calcul des valeurs moyennes ».

Plus loin (paragraphe 3), il est écrit que cette règle « ne s'appliquera pas aux mesures qui correspondent aux **sommes totales de paramètres physico-chimiques** donnés, y compris leur métabolites, et produits de dégradation et réaction pertinents. Dans ces cas, les résultats inférieurs à la limite de quantification des substances individuelles devront être remplacés par **zéro** ».

Bien que cet article ne traite que du calcul des valeurs moyennes, il est raisonnable d'adopter la même règle pour l'évaluation des tendances. Par conséquent, pour les sommes de paramètres (exemples : pesticides, 4 HAP), les résultats < LQ seront remplacés par zéro.

2.3.2 Les solutions alternatives

Une première solution alternative relativement simple de prise en compte des données inférieures à une limite de quantification consiste à compter le nombre de ces données dans la série temporelle. Il ne s'agit alors plus de travailler en valeurs vraies mais en fréquence de dépassement de la limite de quantification. Ainsi, selon un pas de temps choisi en fonction du calendrier de prélèvement de la chronique (l'année par exemple si la fréquence de prélèvement est pluriannuelle), l'information reportée correspond à l'évolution temporelle de la fréquence annuelle de dépassement de la limite de quantification. Une augmentation de la fréquence de dépassement de la limite de quantification avec le temps est le signe d'une orientation à la hausse des concentrations et inversement.

La « tendance » ainsi obtenue n'est pas aussi robuste d'un point de vue statistique que les tendances déterminées avec les méthodes citées en 2.2, mais cette méthode permet de valoriser de façon satisfaisante des données initialement peu informatives. Aussi, cet exercice n'est possible que si les limites de quantification sont égales pour toute la longueur de la chronique. Dans le cas contraire, il est nécessaire de rechercher une méthode plus complexe.

Deux autres approches alternatives sont décrites en détail dans Helsel et Hirsch (2002), Chapitre 13 : la « *méthode de la distribution des données* » et la « *méthode robuste* ». Ces alternatives sont issues du constat qu'une substitution par 0 des données reportées inférieures à une limite de quantification (LQ) tend à sous-estimer grandement les vraies valeurs des tests, alors qu'une substitution par la valeur de la LQ tend à les surestimer. Il est

aussi montré par les auteurs qu'une substitution par $\frac{1}{2}$ de la valeur de LQ est très peu puissante face aux approches alternatives proposées.

Sans entrer dans le détail, le principe de la méthode de la distribution des données utilise les caractéristiques d'une distribution estimée des données afin d'évaluer les statistiques de la série chronologique. Dans l'exemple extrait de Helsel et Hirsch (2002) présenté en Illustration 2, la courbe de répartition des données montre une distribution qui peut être ajustée suivant une loi log-normale (Illustration 2 a).

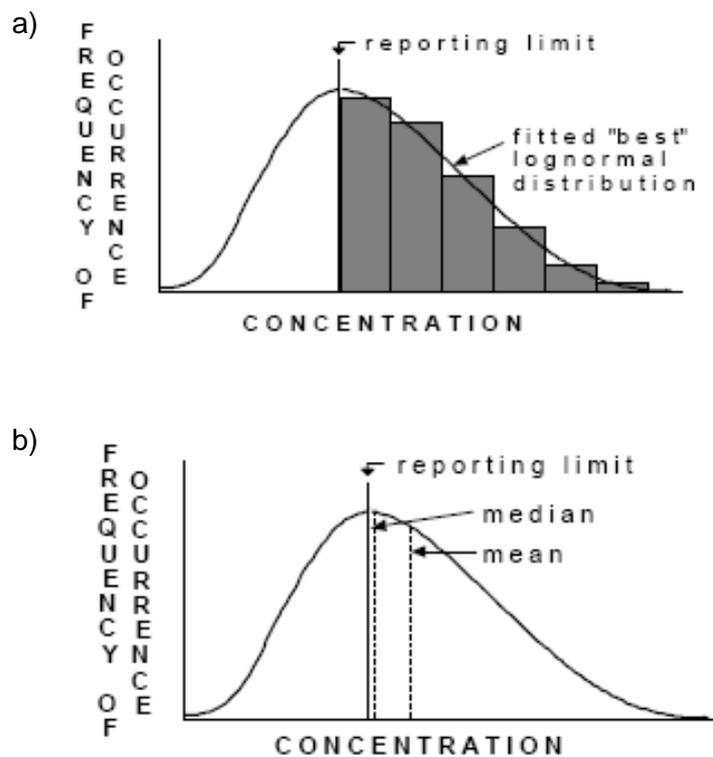


Illustration 2 : Exemple d'application de la méthode de la distribution des données pour la prise en compte des valeurs reportées inférieures à une limite de quantification a) courbe de répartition des concentrations et loi de distribution estimée b) moyenne et médiane des données estimées suivant la loi de distribution (extrait de Helsel et Hirsch, 2002).

Connaissant la loi de distribution des données, il est alors possible d'estimer les statistiques de la série chronologique et notamment sa moyenne et sa médiane (Illustration 2 b). Ces statistiques de la série entrent alors dans le calcul des tests de stationnarité.

La méthode robuste combine quant à elle les données observées reportées supérieures à une limite de quantification avec les données inférieures à une LQ extrapolées suivant une

forme de distribution (Illustration 3). Les données inférieures à une LQ, après extrapolation (Illustration 3 a) ne sont pas considérées comme une estimation des valeurs initiales. Il est en effet impossible d'attribuer les valeurs extrapolées aux bonnes dates de prélèvements. Ces valeurs ne sont utilisées que pour dessiner la courbe de répartition de la chronique (Illustration 3 b) et ainsi estimer les paramètres statistiques de la série. Comme précédemment, ce sont ces paramètres qui seront utilisés lors du calcul des tests statistiques.

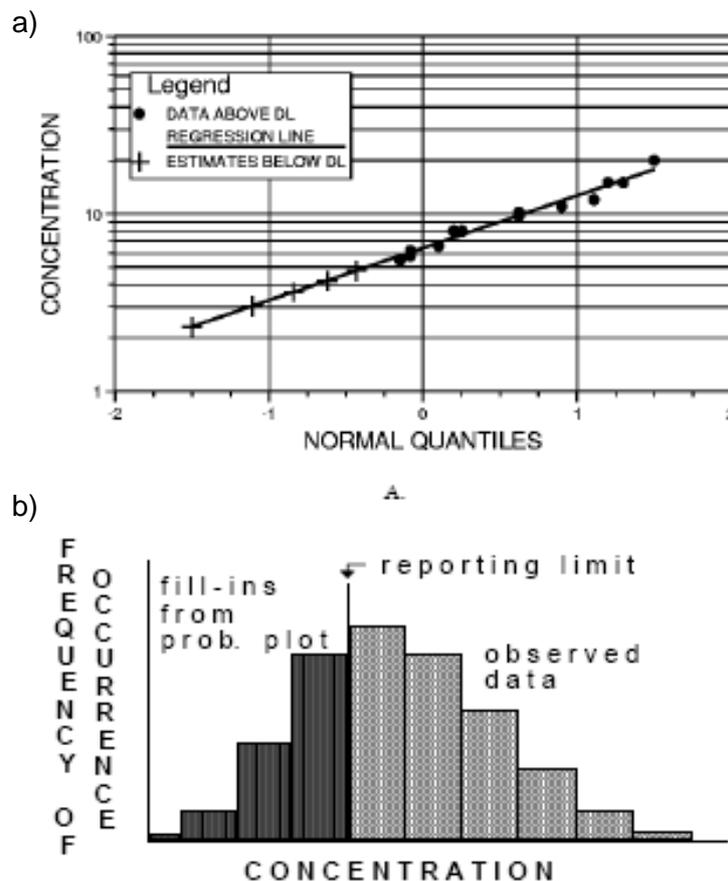


Illustration 3 : Exemple d'application de la méthode robuste pour la prise en compte des données reportées inférieures à une limite de quantification a) régression des logs des concentrations selon les quantiles d'une distribution normale pour l'extrapolation des données inférieures à une LQ b) après transformation des données extrapolées en vraies valeurs et combinaison avec les valeurs reportées supérieures à une LQ, estimations des statistiques de la série chronologique (extrait de Helsel et Hirsch, 2002).

3 Revue bibliographique des travaux de détermination des tendances d'évolution de la qualité des eaux en France et à l'international.

Dans le monde scientifique international, l'évaluation des tendances d'évolution des chroniques de données environnementales, et de la qualité des eaux souterraines en particulier, occupe une place importante et croissante depuis de nombreuses années. L'intérêt pour les méthodes statistiques de traitements des séries temporelles a émergé en même temps que la collecte des analyses de qualité des eaux est devenue plus importante, augmentant ainsi les jeux de données à traiter. Des réseaux de mesures de la qualité des eaux sont apparus dans les années 1970 sous l'impulsion d'une pollution croissante de l'environnement. Aujourd'hui, les connaissances sur les comportements des polluants dans les eaux souterraines ont été améliorées. Toutefois, toujours plus de substances recherchées. Les méthodes statistiques ont évoluées au fur et à mesure que les chroniques devenaient plus longues et nombreuses, bien que les bases et les problèmes spécifiques au domaine de la qualité des eaux aient été posés très tôt.

3.1 La recherche scientifique internationale sur les méthodes statistiques d'évaluation des tendances dans le domaine de la qualité des eaux

3.1.1 Développement des méthodes statistiques

Les premiers exemples d'évaluation des tendances d'évolutions temporelles à partir de données de qualité des eaux issues d'un réseau de mesure sont reportés dans Steele et al. (1974) et Lettenmaier (1978). Ces auteurs avaient déjà pressentis le problème que pouvait engendrer la corrélation des signaux de qualité avec les phénomènes climatiques mais les techniques employées pour décrire les tendances étaient encore essentiellement visuelles, par création de catégories de corrélation toutefois pour Steele et al. (1974).

Les méthodes visuelles ont été les premières utilisées et sont aujourd'hui encore largement employées pour décrire les évolutions des paramètres de la qualité des eaux. Boyer et al. (2000) démontrent ainsi l'efficacité de la représentation par boîte à moustache pour dégager

et valoriser les informations issues d'un réseau de suivi de la qualité des eaux. Cette méthode, relativement simple de mise en œuvre se révèle en effet très utile pour comparer les paramètres statistiques de plusieurs populations. L'analyse peut être temporelle et spatiale lorsque les populations sont créées par regroupement de points de mesure appartenant à un même secteur. Aujourd'hui, les représentations par boîtes à moustache sont utilisées, entre autres, pour rendre compte de l'évolution des concentrations en pesticides dans les eaux souterraines à l'instar de Bexfield (2008) qui compare, grâce à cette méthode, les concentrations des 80 pesticides recherchés dans les eaux souterraines aux USA entre deux périodes 1993-1995 et 2001-2003.

Hirsch et al. de l'US Geological Survey furent les premiers en 1982 à définir clairement la problématique de l'évaluation des tendances dans le domaine de la qualité des eaux et à proposer des méthodes statistiques adaptées. Les auteurs ont montré que les caractéristiques qui compliquent l'analyse des tendances de chroniques temporelles de qualité des eaux sont les distributions non normales des données, les fluctuations saisonnières des concentrations, les phénomènes d'autocorrélation, les valeurs manquantes dans les séries et les données reportées inférieures aux seuils de quantification. Les méthodes statistiques décrites dans l'article sont choisies et développées car elles répondent aux problèmes précédemment cités et sont adaptées aux données de qualité acquises mensuellement. Les auteurs démontrent la robustesse de la méthode Kendall Saisonnier face aux données manquantes dans les séries temporelles et la pertinence de son utilisation dans le cas de fluctuations saisonnières dans les données. L'application de la méthode est en effet réalisée sur une chronique temporelle d'évolution des concentrations en phosphore total dans la rivière Klamath (Californie) qui présente un cycle saisonnier bien marqué visible sur l'illustration 4.

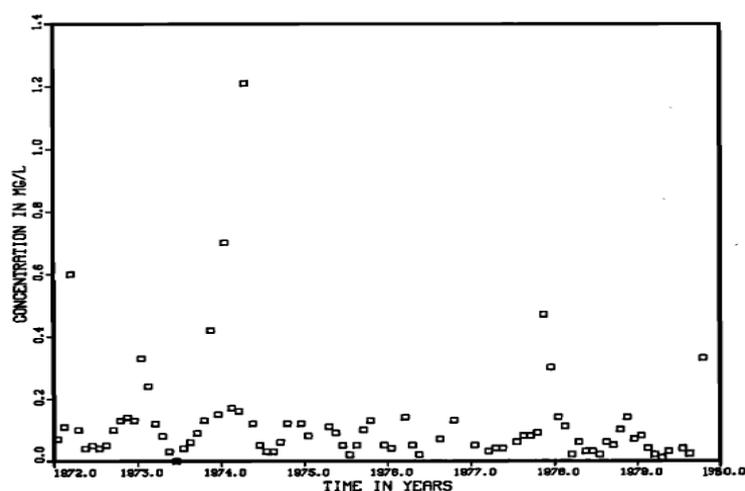


Illustration 4 : Concentration en phosphore total, rivière Klamath, Californie (tendance = -0,005 mg/L/an, extrait de Hirsch et al., 1982)

La puissance du test Kendall-Seasonal est comparée à celle de la régression linéaire classique dans le cas de chroniques autocorrélées afin de montrer le danger de l'utilisation de méthodes paramétriques dans un tel cas de figure. La seconde procédure décrite est l'estimation de la pente de la tendance Kendall-Seasonal. Cette procédure permet de pallier au manque de quantification de la pente de la tendance lorsque seul le test non paramétrique de Mann-Kendall est appliqué. Ces méthodes sont encore aujourd'hui couramment utilisées car elles apparaissent particulièrement adaptées à la qualité des chroniques disponibles sur les eaux souterraines. Toutefois, des améliorations et des alternatives ont depuis été proposées sans que les problèmes inhérents aux traitements des données temporelles de qualité des eaux exposés dans l'article n'aient été résolus.

Dans le but d'améliorer la robustesse du test Kendall-Seasonal lorsque les données temporelles sont autocorrélées, Hirsch et Slack ont proposé, en 1984, de modifier le test initial. Cette modification consiste en une estimation de la covariance des statistiques du test Kendall-Seasonal à partir des données, plutôt que de fixer une valeur de covariance à 0. Toutefois, cette modification ne permet d'augmenter la robustesse du test que si les chroniques sont de longues durées (supérieures à 40 ans) et acquises au pas de temps mensuel. Dans le cas contraire, l'application du test Kendall-Seasonal classique se révèle tout aussi efficace, voire plus robuste que le test modifié lorsque les données sont indépendantes. Cette modification sera donc pour l'instant réservée au domaine de l'hydrologie où les chroniques sont longues et les pas de temps d'acquisition des mesures assez fins. Les chroniques de qualité des eaux souterraines ne se prêtent pas encore à ce type de test mais il est intéressant d'avoir à l'esprit la possibilité d'améliorer l'efficacité des méthodes statistiques d'évaluation de la stationnarité des chroniques de qualité lorsque celles-ci seront plus longues et plus renseignées.

3.1.2 Comparaison de l'efficacité des méthodes statistiques

A la fin des années 80 et au début des années 90, de nombreux auteurs se sont penchés sur l'évaluation de l'efficacité des méthodes statistiques disponibles pour l'analyse des séries temporelles d'évolution de la qualité des eaux.

Yu et al. (1993) de l'USGS (United States Geological Survey) ont par exemple comparé la puissance de 4 tests non paramétriques (Mann-Kendall et Kendall-Seasonal pour tester la stationnarité, la méthode T de Sen pour l'estimation de la pente des tendances et le test de Van Belle et Hughes (1984) pour évaluer l'homogénéité des séries) sur des chroniques de qualité d'eaux de rivières de tailles variables. Ils ont montré que ces tests possèdent une puissance similaire pour des chroniques égales ou supérieures à 9 ans et notent que les tendances détectées entre les stations d'une même rivière ne sont pas homogènes. Cette hétérogénéité entre les différents points de mesure pose en effet de nombreux problèmes lors des exercices de détermination des tendances spatialisées, tant dans le domaine de la qualité des eaux de surface que des eaux souterraines.

Quelquefois, les conclusions des études apparaissent contradictoires de sorte qu'il devient difficile de choisir les méthodes les plus appropriées aux types de chroniques à analyser. Loftis et al. (1991) des départements d'agriculture et de statistiques de l'université Colorado State aux Etats-Unis, et Thas et al. (1998), du département Biomath de l'université de Gent en Belgique, ont ainsi comparé la puissance de 4 tests non paramétriques de détection des tendances : le Kendall-Seasonal, le Kendall-Seasonal modifié par Hirsch et Slack (1984), la « covariance eigenvalue » (Lettenmaier, 1988) et le test d'inversion de covariance. Les premiers auteurs basent leur étude sur des chroniques de qualité simulées tandis que les seconds appliquent ces méthodes à des chroniques vraies d'évolution des concentrations en nitrate dans la rivière Dender à Geraardsbergen. Les deux études confirment les conclusions précédemment citées sur l'efficacité du test Kendall-Seasonal modifié par rapport au Kendall-Seasonal. Par contre, Loftis et al. recommandent l'utilisation de la « covariance eigenvalue » proposée par Lettenmaier (1988) dans le cas de séries autocorrélées alors que Thas et al. notent le caractère trop conservatif (non rejet de H_0 en présence de tendance) des tests basés sur la covariance.

Au-delà de ces différences, une nouveauté apparaît dans l'esprit avec lequel sont menées ces études. Les auteurs ont en effet de plus en plus tendance à orienter l'objectif de leur recherche sur l'optimisation des fréquences de prélèvements en fonction de la puissance des méthodes statistiques à appliquer. Cette orientation nouvelle de la définition des réseaux de mesures devrait être reprise aujourd'hui pour la révision des programmes de surveillance des eaux souterraines. Les programmes doivent en effet être adaptés aux objectifs d'évaluation de l'état qualitatif des ressources et non l'inverse.

Tang et Macneill (1991) du ministère de l'environnement canadien et du département de statistiques de l'université d'Ontario respectivement se sont quant eux concentrés sur les

méthodes statistiques dédiées à la détection des changements de tendances sur des chroniques de qualité des eaux. Les méthodes comparées sont le test de CUSUM, le ratio de Worsley Likelihood et une modification de ce dernier dans le cas d'une régression polynomiale. Les auteurs démontrent par cette étude l'intérêt de l'application de telles méthodes en soulignant le fait que celles-ci peuvent être adaptées à tous types de données environnementales, hydrologiques, climatiques, et de qualité des eaux tant superficielles que souterraines.

La multiplication des tests d'efficacité des méthodes statistiques pour l'analyse des tendances a permis aux opérateurs d'adapter le choix des outils en fonction du type de chroniques de qualité des eaux à traiter, superficielles dans un premier temps. Au Canada par exemple, Bouchard et Haemmerli (1991) ont appliqué, lors de l'analyse de la qualité des eaux de 12 lacs, le test de Mann-Kendall seulement sur les chroniques qui ne présentaient ni autocorrélation, ni périodicité dans leur évolution. Les chroniques affectées par une saisonnalité (comme la plupart des chroniques d'évolution des concentrations en nitrate) ont été traitées avec le test Kendall Saisonnier. Pour les séries autocorrélées (comme celle d'évolution des concentrations en sulfate) les auteurs se sont orientés vers le test de Spearman. Enfin, les chroniques qui présentaient à la fois une périodicité et une forte autocorrélation des données ont été analysées avec le test Kendall-Seasonal modifié par Hirsch et Slack. L'illustration 5 extraite de Bouchard et Haemmerli (1991) permet de rendre compte de la puissance de chacun des tests par l'analyse du rapport Tr/σ_e entre la tendance Tr et l'écart type σ_e des données. Le test est considéré d'autant plus puissant que ce rapport est grand.

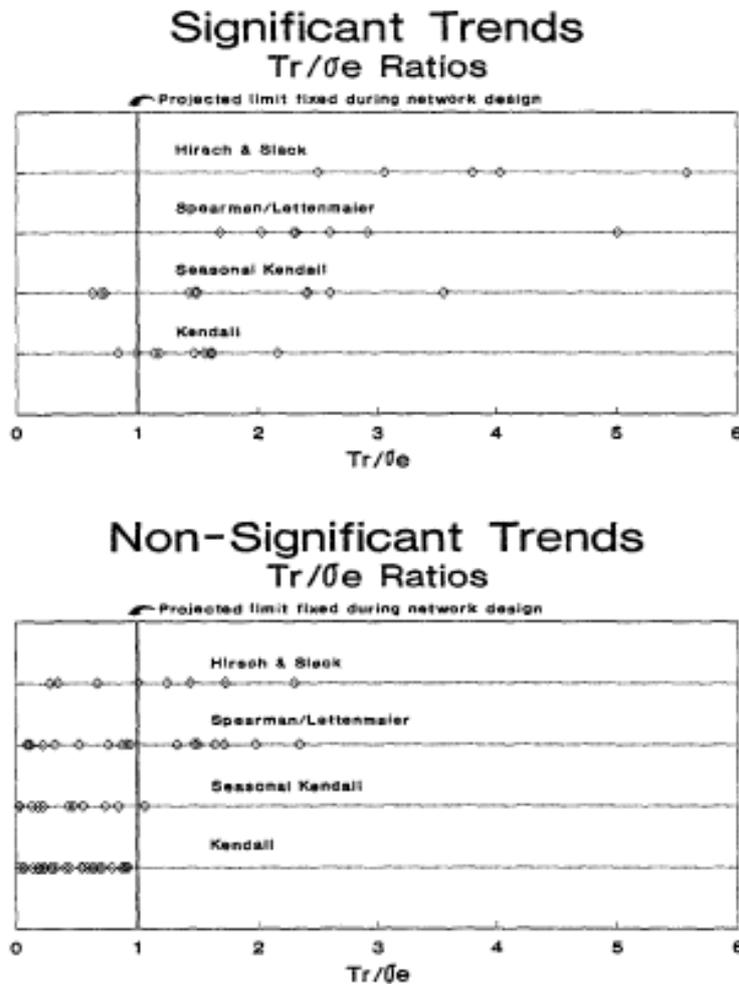


Illustration 5 : Ratio Tr/σ_e pour les chroniques avec (Significant Trends) ou sans (Non-Significant Trends) tendances significatives (extrait de Bouchard et Haemmerli, 1991).

Les auteurs montrent que pour des chroniques non autocorrélées et non périodiques, il est possible de détecter, par le test de Mann-Kendall, des tendances de faibles amplitudes ($Tr/\sigma_e < 1$). Il en est de même lorsque le test Kendall Saisonnier est appliqué sur des séries cycliques saisonnières. Par contre, lorsque les données sont autocorrélées (test de Spearman) ou bien autocorrélées et cycliques saisonnières (Kendall-Seasonal modifié), les tendances ne sont détectées que pour un rapport $Tr/\sigma_e > 1,5$ et $2,5$ respectivement. Les résultats amenant à des tendances non significatives sont surprenants pour les tests de Spearman et Kendall Saisonnier modifié. Des tendances avec un rapport $Tr/\sigma_e > 1$ ne sont en effet pas détectées par ces tests. Les auteurs incriminent alors des longueurs de chroniques trop faibles pour que les tests soient puissants et préconisent alors de toujours tenter d'acquérir le plus grand nombre de données possibles avant de faire une analyse de stationnarité, particulièrement lorsque les données présentent une autocorrélation forte. Cette conclusion sera largement discutée dans la suite du rapport car il est démontré que

l'autocorrélation peut être diminuée en réduisant les fréquences et en augmentant l'irrégularité des prélèvements.

3.1.3 Les guides méthodologiques

C'est à la même période qu'apparaissent au Canada et aux Etats-Unis les premiers guides techniques pour l'analyse des tendances des séries temporelles d'évolution de la qualité des eaux dont celui de McLeod-Hipel & Assoc. Ltd. pour le compte de l'Environment Ontario en 1991. Dans ce guide méthodologique, sont décrites à la fois des méthodes graphiques telles que la régression lissée ou la représentation par boîte à moustache et des méthodes statistiques non paramétriques (test de Mann-Kendall et test de Spearman). L'effet de la saisonnalité ainsi que la présence de tendances sont testés sur des données de qualité des eaux des rivières Grand et Saugeen du sud-ouest de l'Ontario.

Plus récemment Helsel et Hirsch, de l'U.S. Geological Survey, éditent en 2002 un guide méthodologique très complet sur l'application des méthodes statistiques dans le domaine des ressources en eaux. Cet ouvrage, qui fait aujourd'hui office de référence, reprend toutes les étapes de l'analyse et du traitement statistique des chroniques temporelles de qualité des eaux. De la collecte des données à l'évaluation des tendances, en passant par les modes de représentations graphiques, l'analyse des dépendances des données, des corrélations et des moyens de prendre en compte les données inférieures aux limites de quantification, de nombreux outils statistiques sont proposés en fonction du type de données à traiter et des objectifs à atteindre. Le public visé est celui du monde scientifique qui maîtrise les principes de base de la statistique. Certains principes et outils exposés dans l'ouvrage et en prise directe avec les objectifs de l'étude sont donc repris et explicités afin de pouvoir être utilisés par les organismes gestionnaires.

Peu de guides concernent spécifiquement l'étude des tendances de la qualité les eaux souterraines. Quelques recommandations sont néanmoins présentées dans Margane (2004) qui propose, dans le cadre de la gestion, de la protection et de l'utilisation raisonnées des eaux souterraines et des ressources du sol des régions du Maghreb un guide pour le suivi des eaux souterraines. En ce qui concerne les tendances, l'analyse recommandée est essentiellement visuelle, le guide ne spécifiant pas de méthodes statistiques particulières. Ceci est généralement le cas dans les guides de construction et de valorisation des réseaux de suivi où les méthodes statistiques d'exploitation et de valorisation des données ne sont que très rarement évoquées. Olmedo et Luque Espinar (2010) proposent toutefois un guide

spécifique sur l'analyse statistique des données hydrogéologiques en réponse aux exigences de la directive eaux souterraines. Des recommandations sont émises avant la mise en œuvre de techniques paramétriques, la régression linéaire en particulier, afin d'éviter les résultats erronés. Les auteurs recommandent ainsi le test de Shapiro-Wilk pour l'étude de la stationnarité des données et montrent, dans le cas de données non-normales, l'intérêt de l'utilisation du test de Mann-Kendall. Les biais engendrés par les variations saisonnières des chroniques sont aussi abordés, les auteurs renvoyant vers le guide de Helsel et Hirsch (2002) pour ce point particulier.

3.1.4 Les recherches actuelles

Aujourd'hui, la recherche est concentrée sur les moyens de prendre en compte l'autocorrélation des données qui apparaît comme un des phénomènes les plus perturbateurs des tests statistiques de stationnarité des données (Yue et al., 2002 ; Darken, 2002 ; Libiseller et Grimvall, 2002). Hamed et Rao (1998) et Yue et Wang (2004) proposent ainsi de corriger le test statistique de Mann-Kendall pour prendre en compte l'autocorrélation. Le détail de ces modifications est présenté dans le paragraphe 0. Par la suite, Hamed (2008) a testé la méthode Mann-Kendall modifiée sur un jeu de données composé de 57 séries d'évolution des débits de rivières provenant de toutes les zones du monde. Il a montré que les tests classiques ont tendance à donner plus de tendances significatives que le test modifié. Il conclut donc que la présence de tendance dans les séries hydrologiques est plus faible que ce qui est mentionné dans les études récentes. Il montre aussi que la prise en compte des phénomènes d'autocorrélations dans le test permet dans certain cas de lever les incohérences parfois rapportées entre les tendances déterminées sur un point et celles déterminées spatialement. Toutefois, cette méthode paraît aujourd'hui difficile à appliquer sur des données de qualité des eaux souterraines en raison des plus faibles fréquences de prélèvements et des dépendances aux phénomènes climatiques généralement moins marquées que lorsque l'on s'intéresse aux débits de rivières.

En France, C. Bernard-Michel (Ecole des Mines de Paris) publie en 2006 une thèse sur les indicateurs géostatistiques de la pollution des cours d'eau. Cet ouvrage propose, dans un premier temps, d'examiner les biais et les incertitudes des indicateurs actuels de la qualité des eaux superficielles. Ceux-ci reposent en effet sur deux hypothèses implicites mais erronées : l'indépendance des mesures et la stationnarité des concentrations durant l'année. Or l'évolution des concentrations des principaux polluants présente généralement des variations saisonnières et l'auteur met en évidence, par l'utilisation de variogrammes

expérimentaux, des corrélations temporelles dans les données. Une nouvelle méthode de krigeage qui prend en compte les irrégularités d'échantillonnages et les corrélations temporelles est proposée et testée sur les données vraies issues du réseau de mesure du bassin Loire-Bretagne. Dans un second temps, la question de l'interpolation des indicateurs le long des cours d'eau est posée. Une des solutions proposée est un modèle général de fonctions aléatoires le long d'un réseau hydrographique. Les résultats sont discutés à partir d'une application sur le bassin de la Moselle. Cette étude poussée peut lancer des pistes de réflexions sur les moyens de spatialiser les observations ponctuelles de la qualité des eaux souterraines. L'idée d'un krigeage prenant en compte les corrélations temporelles des données pourrait être développée et testée sur des réseaux de mesure de la qualité des eaux souterraines. Cette méthode n'en est toutefois qu'au stade du développement et donc pas suffisamment opérationnelle pour être proposée aux organismes gestionnaires de la qualité des eaux souterraines.

En complément de ces travaux, il est intéressant de citer la thèse de B. Renard (Institut National Polytechnique de Grenoble) publiée en 2006 sur la détection et la prise en compte d'éventuels impacts du changement climatique sur les extrêmes hydrologiques en France. En effet, si le domaine d'application de la thèse n'est pas celui de la qualité des eaux, de nombreux tests statistiques de stationnarité des chroniques temporelles sont détaillés dans l'ouvrage avec une partie importante dédiée à l'évaluation de leur robustesse et de leur puissance. La question des méthodes de détection des changements dans les séries temporelles est clairement posée avec un chapitre consacré aux moyens de réaliser une analyse à l'échelle régionale. L'argumentaire de la thèse est basé sur les crues des cours d'eau mais les problèmes de significativités régionales des stations de mesures pourraient être repris et adaptées au domaine de la qualité des eaux souterraines.

Un autre champ de recherche actuellement exploré est le domaine de l'extrapolation des évolutions actuellement constatées. Des méthodes complexes de modélisation et de déconvolution des chroniques temporelles sont expérimentées. Elles sont basées sur la connaissance des différentes composantes qui construisent les signaux, c'est-à-dire la composante cyclique, le bruit et la tendance. Néanmoins, ces méthodes sont aujourd'hui toujours en développement et leur application au domaine de la qualité des eaux souterraines pas toujours adaptée. De plus, il apparaît aujourd'hui prématuré de proposer aux gestionnaires de tels outils, tant leur mise en œuvre est complexe et les résultats non encore validés. Ces méthodes sortant du sujet de l'étude, elles ne sont pas détaillées dans ce rapport.

Enfin, il est important de citer les recherches en cours sur l'impact des incertitudes (de mesures et environnementales) sur l'évaluation des tendances d'évolution de la qualité des eaux souterraines. Ce volet fait toutefois l'objet d'un rapport spécifique ONEMA-BRGM (Ghestem, 2010) très complémentaire de l'étude présentée ici.

3.2 Exemples de détermination des tendances d'évolution de la qualité des eaux souterraines à l'international

Jusqu'au début des 90, les travaux de détermination des tendances des chroniques de qualité concernaient essentiellement les eaux superficielles. Les données sur les eaux souterraines devenant de plus en plus importantes et les problèmes rencontrés dans ce domaine se révélant à peu près les mêmes que pour les eaux superficielles, des travaux d'évaluation de la stationnarité des chroniques d'évolution de la qualité des eaux souterraines apparaissent.

Ainsi, en 1994, les néerlandais Frapporti et al. proposent d'utiliser, sur des données de qualité des eaux souterraines, la méthode de corrélation par rang de Spearman car il a été démontré que ce test non paramétrique est peu sensible aux outliers (données extrêmes ou aberrantes). Les auteurs insistent bien sur le fait que l'utilisation d'une méthode non paramétrique est nécessaire lorsque la distribution des données est inconnue, ce qui est le cas dans leur étude. De plus, la méthode de Spearman permet de traiter des groupes de points avec un résultat statistique unique pour le groupe. Un point avec peu de données et donc porteur de peu d'information peut ainsi être agrégé avec d'autres points appartenant à une zone identique d'un point de vue géochimique. Ainsi, les informations se somment et la connaissance des tendances d'évolution de la qualité des eaux souterraines peut être associée à une surface plutôt qu'à un point unique de mesure.

Plus récemment en 2009, dans le cadre du projet européen Aquaterra, un consortium européen constitué de scientifiques du Geological Survey Néerlandais, de l'université d'Utrecht au Pays-Bas, du BRGM en France, de l'université de Liège en Belgique et de l'institut pour l'écologie et les aires industrielles de Pologne a publié une étude (Visser et al., 2009) qui compare les méthodes de détection et d'extrapolation des tendances de la qualité des eaux souterraines. Visser et al. (2009) rappellent que pour la protection de la qualité des eaux souterraines il est nécessaire (1) d'appréhender le statut actuel de la qualité des eaux

souterraines, (2) de détecter les changements et les tendances dans l'évolution de la qualité, (3) d'évaluer la menace de détérioration et (4) de prédire les futurs changements de la qualité des eaux souterraines. Plusieurs méthodes sont possibles pour détecter et extrapoler les tendances et les auteurs ont choisi de comparer 4 approches globales : l'analyse statistique, la datation des eaux souterraines, les fonctions de transfert et la modélisation déterministe. Après avoir rappelé que l'approche choisie dépend à la fois de l'objectif spécifique de l'étude (évaluation des tendances uniquement ou extrapolation), du système étudié ainsi que des ressources disponibles, les méthodes statistiques paramétriques et non paramétriques de détection des tendances sont comparées. Un logigramme simple de sélection de la méthode statistique à appliquer en fonction de la qualité des données est repris de Helsel et Hirsch (2002) (Cf. Illustration 6). Ce type d'arbre de décision peut s'avérer très utile pour accompagner les opérateurs dans leur exercice de détermination des tendances. Cette idée est reprise et améliorée pour la présente étude en détaillant plus les caractéristiques des chroniques de qualité des eaux souterraines avec, pour chaque étape, une proposition de traitement.

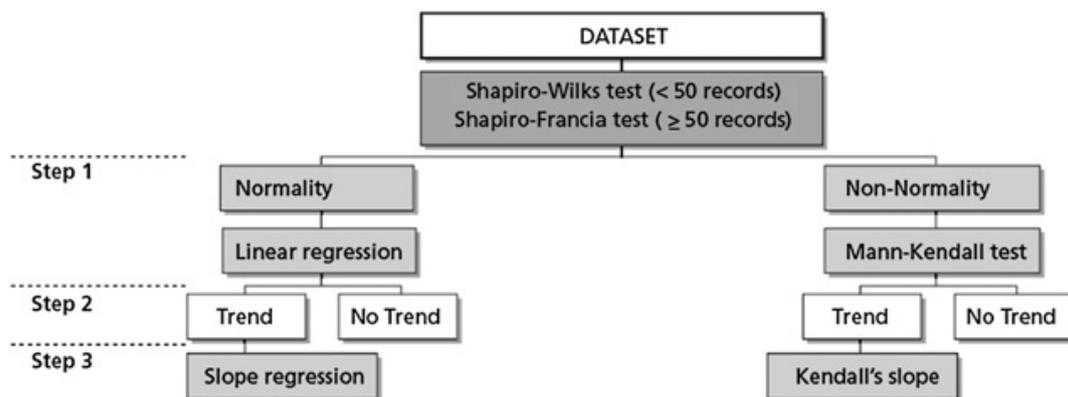


Illustration 6 : Procédure à trois étapes pour la sélection de la méthode statistique d'analyse des tendances d'évolution des contaminants dans les eaux souterraines (1) distribution normale/non normale (2) détection de tendance (3) évaluation de la pente (dans Visser et al. 2009).

Les auteurs concluent qu'il n'y a pas d'approche optimale unique pour répondre aux objectifs d'évaluation, de détection de tendance et d'extrapolation de la qualité des eaux souterraines en raison de la multiplicité des contextes existants. Un des aspects les plus importants est la définition du réseau et du programme de mesure qui doit couvrir, selon les auteurs, la plus grande variété possible de types d'eaux (récentes et anciennes). Toujours en conclusion, un tableau (Tableau 4) résume les approches qu'il est possible d'envisager en fonction du type de système aquifère (simple ou complexe) et de l'objectif visé.

L'approche statistique semble ainsi, d'après les auteurs, tout à fait appropriée à la détection des tendances d'évolution de la qualité des eaux souterraines que le système soit simple ou complexe. Selon ces mêmes auteurs, dans le cas de systèmes simples, une extrapolation des tendances observées peut même être envisagée.

Tableau 4 : *Approches recommandées pour la détection des tendances et leur extrapolation en contexte d'aquifère simple et complexe (dans Visser et al. 2009).*

		Groundwater system	
		Simple	Complex
Trend detection	Preliminary	Statistics	Statistics
	Elaborate	Groundwater dating	Transfer functions
Trend extrapolation	Preliminary (short-term)	Statistical methods	Transfer functions
	Elaborate (long-term)	Deterministic model	Deterministic model

Toujours dans le domaine de la qualité des eaux souterraines, Stuart et al., du British Geological Survey, publie en 2007 une étude sur les modalités d'analyse des tendances à long terme d'évolution des concentrations en nitrate dans les eaux souterraines. Cette étude est en prise directe avec les exigences de la directive cadre européenne sur la qualité des eaux souterraines, notamment avec les exigences de détection de tendance et d'inversion mentionnées dans l'Article 4. Les tests de tendances comparés dans l'article sont la régression linéaire par moindres carrés (OLS), la régression linéaire robuste utilisant un estimateur MM (méthode des moments) et la méthode Kendall-Theil d'estimation des pentes des tendances. La robustesse de ces tests est comparée pour un nombre important de cas de figures différents (variations saisonnières ou non, données régulières et irrégulières, changements brusques dans les données, données manquantes...) et ce toujours sur des chroniques vraies d'évolution des concentrations en nitrate dans les eaux souterraines de Grande Bretagne. L'illustration 7 extraite de l'article montre des exemples de données de concentrations en nitrate problématiques pour l'analyse des tendances.

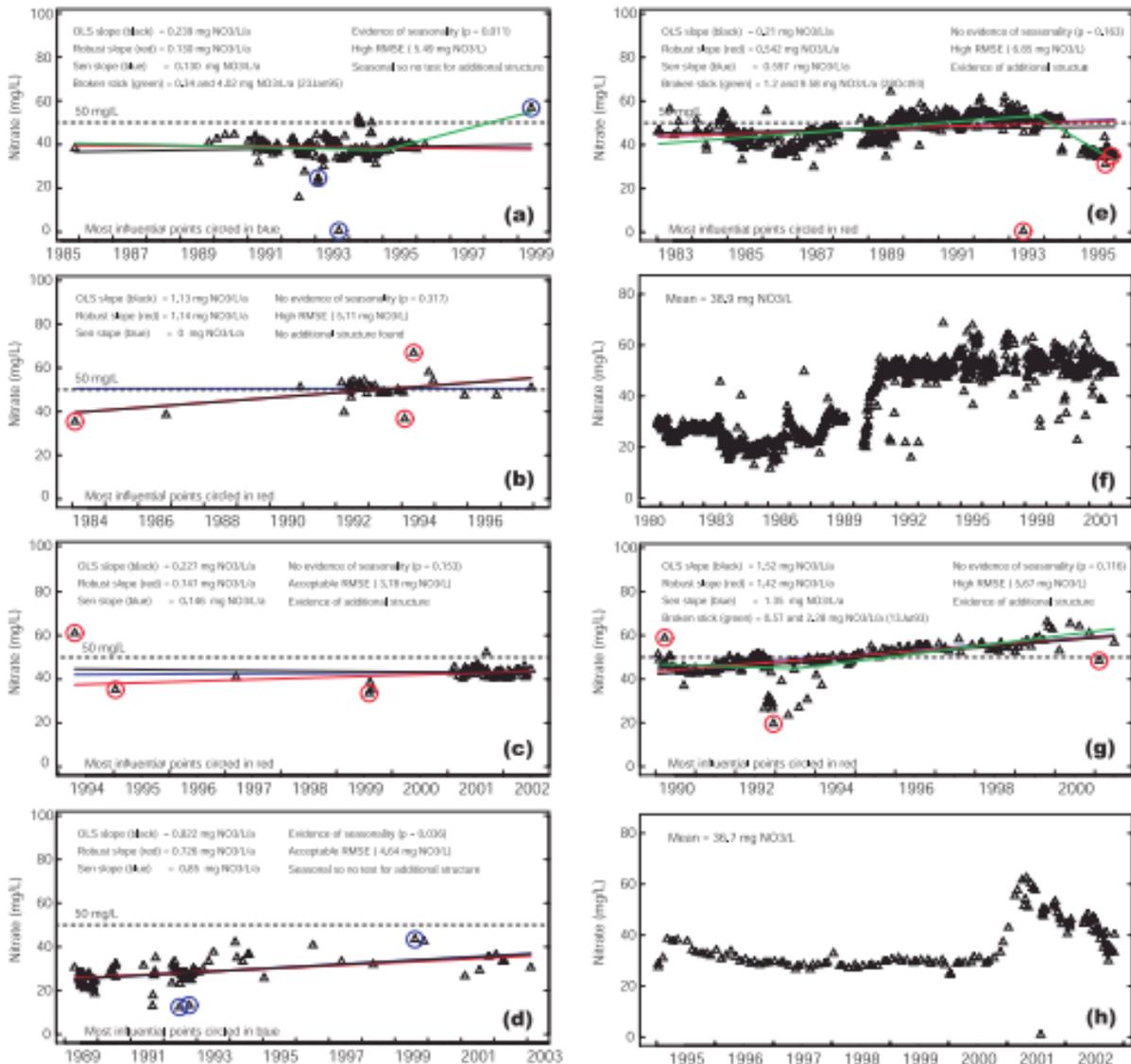


Illustration 7 : Exemples de tests sur des données problématiques : (a) et (b) données irrégulières dans l'aquifère de la craie, avec groupement dans le milieu de la série ; (c) et (d) données irrégulières dans l'aquifère de la craie avec groupement en fin de série ; (e) données non linéaires dans l'aquifère sableux de Sherwood ; (f) série avec changement brutal par palier ; (g) données extrêmes en milieu de séries dans l'aquifère sableux de Sherwood ; (h) données extrêmes en fin de série durant l'hivers 2000-2001 dans l'aquifère de la craie (extrait de Stuart et al., 2007).

Le screening effectué permet de donner une image des tendances d'évolution des concentrations en nitrate dans les eaux souterraines pour l'ensemble des aquifères majeurs d'Angleterre et du Pays de Galles comme reporté dans l'illustration 8.

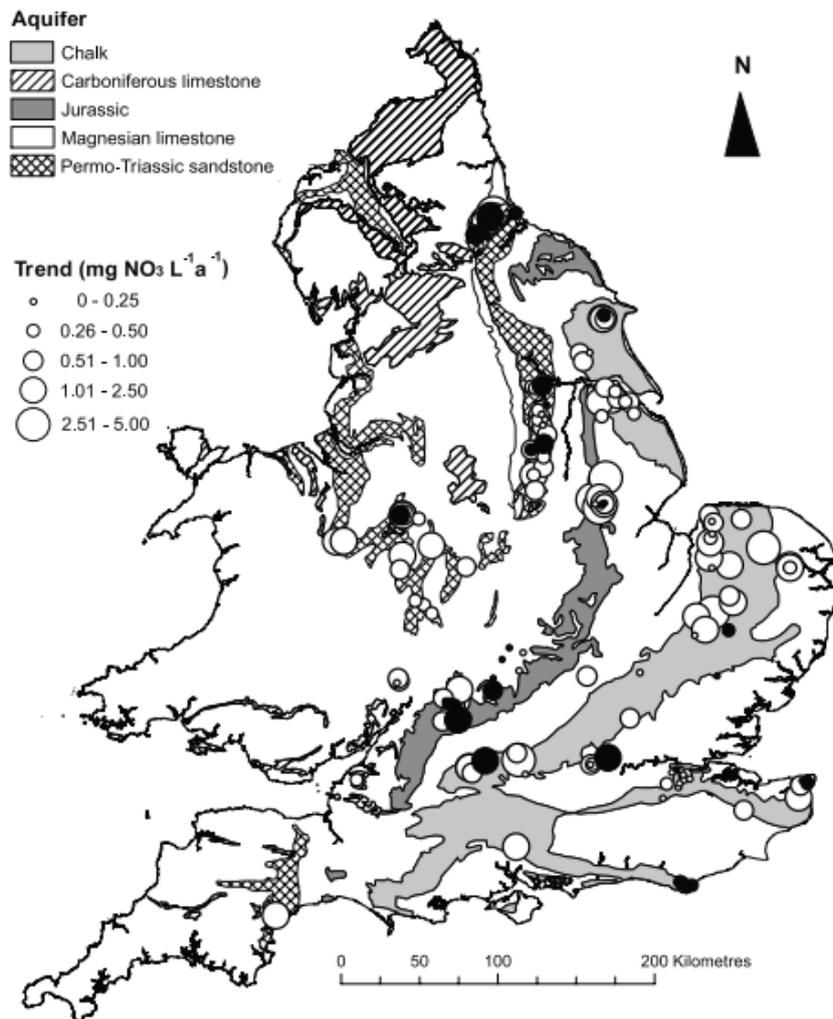


Illustration 8 : Résumé des tendances d'évolution des concentrations en nitrate dans les aquifères majeurs d'Angleterre et du Pays de Galles (cercles vides, tendances positives ; cercles pleins, tendances négatives) extrait de Stuart et al., 2007.

Les tendances sont ici déterminées par régression linéaire par moindres carrés (OLS : Ordinary Least Squares), méthode simple est facile à mettre en œuvre. Les auteurs reconnaissent ne pas avoir pris en compte ni l'autocorrélation des données, ni l'éloignement à la normalité des résidus. Toutefois, la saisonnalité des données, lorsque constatée, a été traitée comme une variable explicative supplémentaire et un effort a été consenti afin d'identifier les données aberrantes. D'une manière globale, les auteurs soulignent l'intérêt de travailler sur les facteurs explicatifs des variations observées, que ce soit sur les causes des périodicités observées ou bien sur les raisons des valeurs qui sortent manifestement de la moyenne de la série. Ce travail permet à la fois de comprendre le comportement géochimique des aquifères mais aussi, dans la plupart des cas, de s'affranchir de méthodes statistiques d'analyse complexes et difficiles à mettre en œuvre.

Aguilar et al. (2007), ont exploité les données très détaillées issues d'un réseau de suivi de la nappe de l'Hesbaye, circulant dans un aquifère crayeux majeur de Belgique. Afin de réaliser les objectifs environnementaux de la Directive cadre sur l'eau (DCE, 2000) les auteurs proposent une méthodologie efficace pour identifier et quantifier les tendances en nitrates dans les eaux souterraines. La procédure statistique est basée sur les travaux de Helsel et Hirsch (2002) reprise par Visser et al. (2009). Ainsi, respectant le schéma présenté en Illustration 6, les auteurs ont utilisé la régression linéaire lorsque les données étaient normalement distribuées et le test de Mann-Kendall pour les données non normales. Ils montrent que ces deux tests sont robustes mais précisent que le test de Mann-Kendall semble plus adéquat car il permet de ne pas vérifier la normalité des données et il produit des tendances très proches de celles obtenues avec la régression linéaire. D'un point de vue hydrogéologique, les résultats montrent une tendance à la hausse des concentrations en nitrate dans l'ensemble du bassin étudié (Cf. Illustration 9).

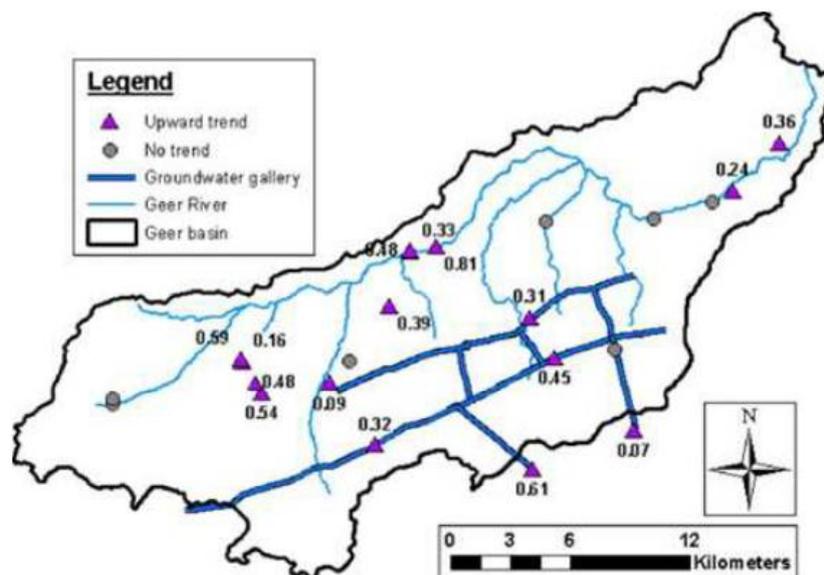


Illustration 9 : Distribution spatiale des tendances nitrates dans les eaux souterraines du bassin de Geer, Belgique, avec les valeurs de tendances à la hausse en mg/L NO₃⁻ par an. (Extrait de Aguilar et al., 2007).

Suivant les tendances calculées, les auteurs proposent de déterminer le temps qu'il reste avant que le seuil de 50 mg/L NO₃⁻ soit dépassé dans les eaux souterraines.

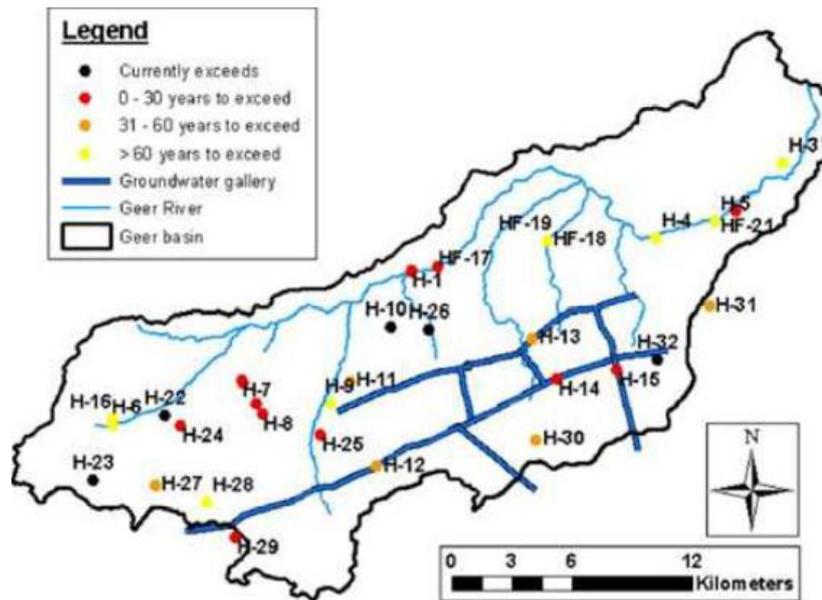


Illustration 10 : Temps (en années) extrapolés avant lesquels le seuil de 50 mg/L NO₃⁻ sera dépassé par les eaux souterraines. (Extrait de Aguilar et al., 2007).

L'illustration 10 extraite de l'article rend compte de ce calcul. L'extrapolation des résultats de l'analyse de tendance montre que des mesures doivent être prises sans tarder pour éviter une dégradation majeure de la qualité des eaux souterraines du bassin, le dépassement du seuil de qualité étant estimé entre 10 et 70 ans. Le test de Mann-Kendall montre son efficacité pour répondre aux objectifs de la DCE, cet outil qui peut être utilisé sur des données à distribution normale ou non normale permettant une estimation robuste des pentes des tendances. De plus, dans le cas de systèmes aquifères simples (hydrodynamique non prépondérante) et si les pressions en entrée du système restent stables, l'outil autorise une extrapolation des tendances observées.

Aux Etats-Unis, la connaissance des concentrations en pesticides dans les eaux souterraines devient suffisant pour entreprendre des études de tendances ponctuelles et spatialisées sur certaines substances de la famille de phytopharmaceutiques. L. Frans, de l'U.S. Geological Survey, publie ainsi en 2008 une étude sur les tendances d'évolution des concentrations en pesticides dans les eaux souterraines du plateau de Colombie centrale. Le test statistique utilisé est la méthode Kendall Regional appliquée sur des données de 72 puits suivi de 1993 à 1995 puis en 2002 et 2003. Le test indique que seule la déséthylatrazine (DEA) montre une tendance régionale significative entre ces deux périodes, à la hausse de surcroît. L'augmentation de DEA dans les eaux souterraines semble négativement corrélée avec le type de sol (estimation des corrélations par régressions logistiques) ainsi qu'avec la classe de drainage (corrélation de Spearman). En France, le

recul nécessaire à la mise en œuvre d'une telle étude n'est pas encore assez important à l'échelle d'un district hydrogéologique, mais les travaux réalisés aux Etats-Unis démontrent la faisabilité d'une telle analyse sur les pesticides. Ceci renforce l'idée que les programmes de suivis doivent être maintenus et améliorés en vue de l'application future de méthodes robustes d'analyse spatiale des tendances d'évolution de la qualité des eaux souterraines, notamment sur des paramètres dont le suivi encore insuffisant ne permet pas, aujourd'hui, l'application de méthodes statistiques robustes.

3.3 Les études réalisées en France par les organismes gestionnaires de la qualité des eaux

Les travaux d'évaluation et de quantification des tendances d'évolution des concentrations dans les eaux souterraines sont mis en œuvre depuis très peu de temps en France de sorte qu'aucune méthode harmonisée n'est disponible. Chaque organisme choisit la manière dont le problème des tendances est abordé, en retenant le plus souvent les méthodes d'analyse visuelle ou bien les méthodes simples de régressions linéaires. Néanmoins, quelques agences de bassin choisissent de participer au développement de méthodes plus robustes d'analyse des tendances et lancent la réflexion sur les outils de spatialisation des tendances observées.

3.3.1 Les eaux de surface

Les programmes de mesures instaurés par les nouveaux Schémas Directeurs d'Aménagement et de Gestion de l'Eau 2010-2015 (SDAGE 2010-2015) prennent en compte l'objectif d'inversion des tendances identifiées à la hausse mais aucune allusion n'est faite sur les moyens d'identification des tendances ou bien les fréquences de prélèvements à respecter pour la mise en œuvre de méthodes permettant l'identification de ces tendances. Le SDAGE Loire-Bretagne mentionne la nécessité de stopper ou d'inverser les tendances nitrates identifiées à la hausse dans les cours d'eau. De plus, dans le chapitre dédié à l'acquisition des connaissances sur la qualité des cours d'eau, le SDAGE Loire-Bretagne note que lorsque les résultats des impacts des mesures agroenvironnementales seront obtenus, il sera important de les découpler de la variabilité due à l'hydrologie, et ce afin d'estimer les tendances de fond. Cette recommandation est en effet tout à fait pertinente lors de l'analyse statistique des tendances de signaux affectés par une forte cyclicité.

Actuellement, certains organismes s'attachent tout de même à identifier les tendances d'évolution de la qualité des cours d'eau. Sur le bassin de l'Orne, par l'intermédiaire du SAGE Orne Moyenne, l'évaluation des tendances d'évolution de la qualité des eaux est faite par observation de l'évolution des concentrations de différents paramètres en des points de prélèvements donnés. La Commission Locale de l'Eau du bassin versant de l'Yerres, dans le cadre de l'élaboration du SAGE de ce bassin versant propose quant à elle d'étudier l'évolution des indicateurs de qualités sur des altérations définies par le SEQ-Eaux puis de faire la différence entre les valeurs de ces indicateurs prises entre deux dates données.

La DIREN est l'un des organismes qui participe le plus activement à l'identification des tendances dans les cours d'eau. La DRIEE Ile-de-France propose ainsi un mode de représentation du suivi de l'évolution de la qualité des cours d'eau du bassin tout à fait intéressant. Comme le montre l'illustration 11, la carte permet de suivre l'évolution d'un indicateur, ici le respect du bon état chimique des cours d'eau, dans le temps (de 2001 à 2004) et par point de prélèvements.

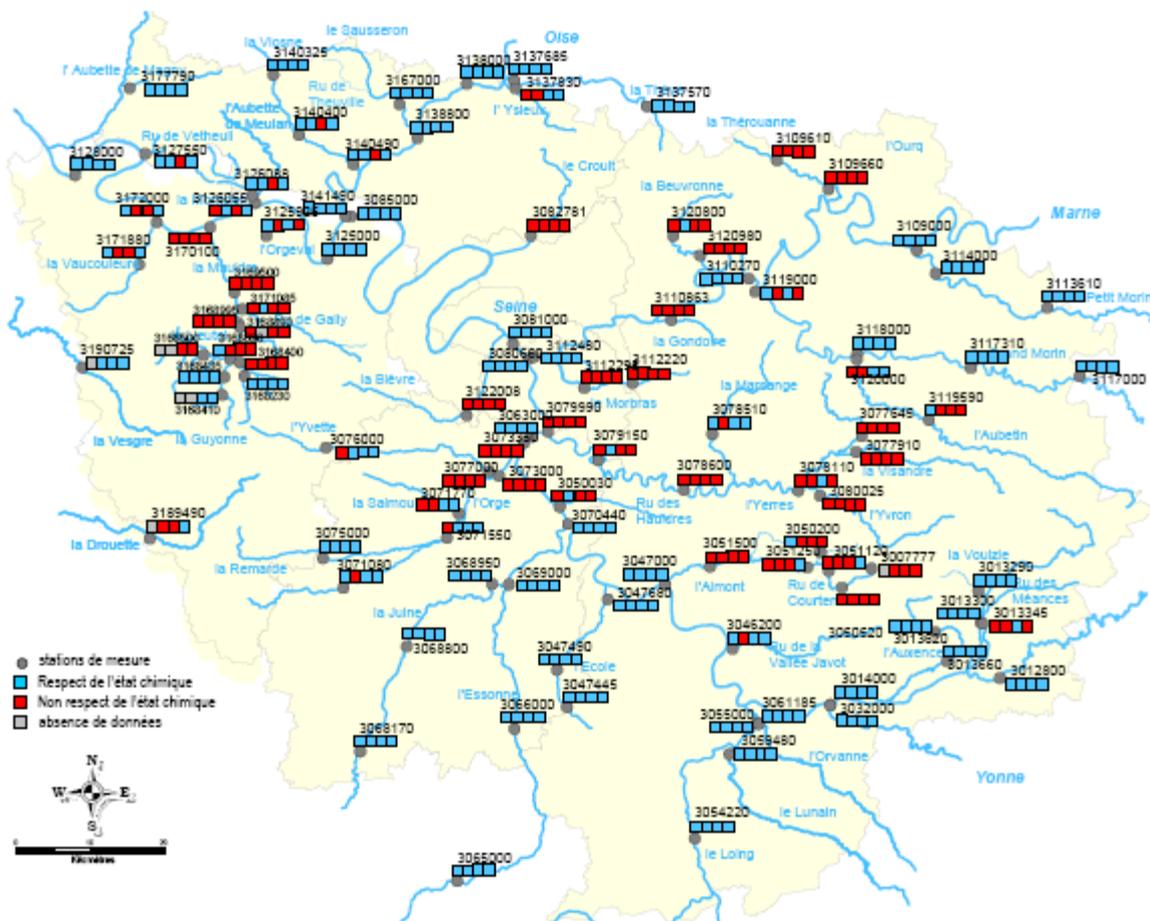


Illustration 11 : Evolution par point de mesure du respect de l'état chimique des cours d'eau du bassin Ile-de-France (DIREN, 2006).

Pour l'évaluation des tendances, les travaux de la DRIEE Ile-de-France consistent à calculer une tendance linéaire par régression linéaire de la qualité des cours en chacun des points de prélèvement. Des cartes d'évolution des concentrations en polluants sont ainsi éditées par l'OIEau à partir des données de la DIREN comme celles montrées en exemple dans l'illustration 12 sur l'évolution des concentrations en nitrate dans les cours d'eau du bassin Ile-de-France de 1992 à 2004. L'illustration 13 reprend le même calcul sur une portion plus restreinte des chroniques (de 2000 à 2004) en permettant ainsi d'évaluer la tendance la plus récente d'évolution des concentrations en nitrate dans les eaux superficielles du bassin.

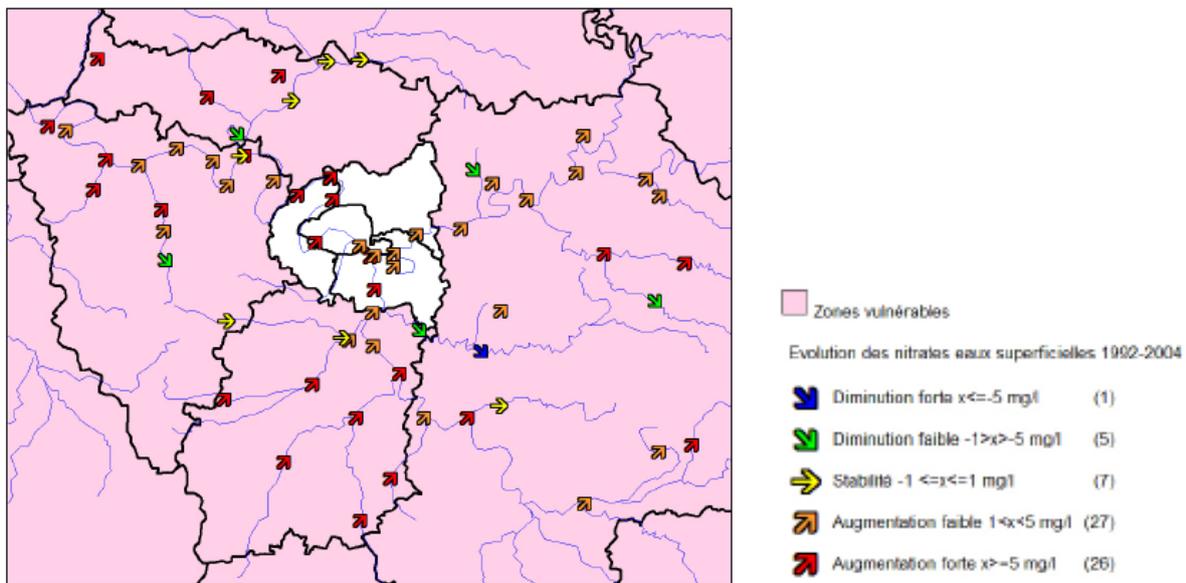


Illustration 12 : Evolution des teneurs moyennes en nitrates des eaux superficielles du bassin Ile-de-France de 1992 à 2004 (OIEau, 2006).

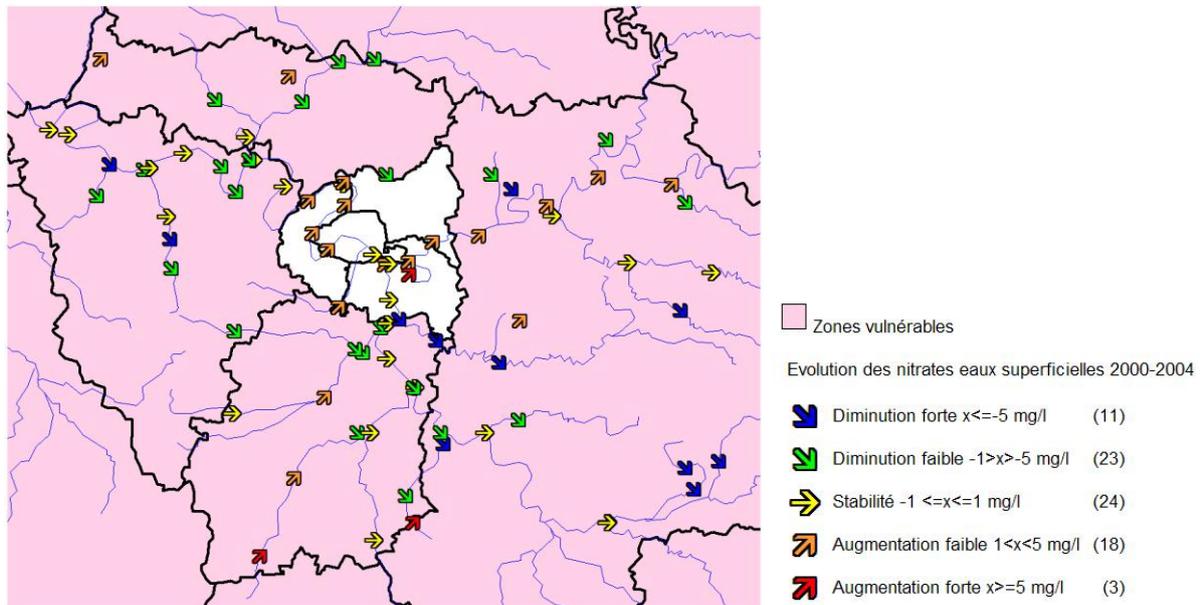


Illustration 13 : Evolution des teneurs moyennes en nitrates des eaux souterraines du bassin Ile-de-France de 2000 à 2004 (OIEau, 2006).

Pour aller plus loin, l'Agence Loire-Bretagne propose une interpolation linéaire des moyennes pondérée par les débits pour l'évaluation des tendances de flux d'azote et de phosphore en rivière (à partir des données DIREN).

Les organismes de bassin ont aussi à leur disposition l'outil PEGASE (Smits et al., 1997) développé par l'université de Liège en Belgique. Le modèle PEGASE (Planification Et Gestion de l'Assainissement des Eaux) a été développé afin d'orienter les choix en matière de gestion des eaux de surface par le calcul prévisionnel et déterministe de la qualité des eaux en fonction des apports et rejets polluants, dans des conditions hydrologiques diverses. PEGASE est un modèle intégré bassin versant / réseau hydrographique. Ce modèle peut traiter plusieurs centaines de rivières simultanément et la superficie des bassins hydrographiques considérés peut atteindre plusieurs dizaines de milliers de km^2 . PEGASE comprend un sous-modèle hydrologique et hydrodynamique, un sous-modèle thermique et un sous-modèle de la qualité de l'eau et du fonctionnement de l'écosystème aquatique. PEGASE représente de façon structurée les rejets urbains, les rejets industriels, le rôle des stations d'épuration, les rejets dus aux activités d'élevage et les apports diffus des sols. PEGASE calcule explicitement les mécanismes d'autoépuration dans les cours d'eau et calcule l'évolution de l'eutrophisation. D'une manière générale, le modèle permet de comparer des scénarios afin de dégager des politiques optimales. Les données d'entrée et les résultats calculés sont traités et visualisés par Système d'Informations Géographiques

(SIG). Ce modèle demande toutefois un nombre important de données d'entrée d'origines diverses (milieux et pressions) de sorte qu'il n'est pas utilisé en routine par les gestionnaires.

3.3.2 Les eaux souterraines

D'une manière générale, peu de travaux sur la détermination des tendances d'évolution de la qualité des eaux souterraines sont menés par les organismes publics français.

Comme pour les eaux superficielles, la DRIEE IDF édite des cartes d'évolution de différents paramètres de qualité dans les eaux souterraines du bassin Seine-Normandie. Le mode de représentation est le même que pour les cours d'eau avec, comme le montre en exemple l'illustration 14, une indication, pour chaque point de prélèvements, une indication de la gamme de concentration en nitrate moyenne par année de 2000 à 2004. A partir de ces données complétées par celles de la DRASS, l'OIEau publie les mêmes cartes de tendances que pour les eaux superficielles. L'illustration 15 est un exemple de ces représentations sous forme de flèches colorées dont l'orientation et la couleur indiquent le sens et la pente des tendances observées. Le mode de calcul présente toutefois des limites car il ne prend pas en compte, ni le caractère normal ou non des données, ni les valeurs extrêmes potentiellement présentes dans les séries, ni les valeurs manquantes, ni les autocorrélations probables dans les données.



Illustration 14 : Evolution de la teneur moyenne en nitrates de 2000 à 2004 – Réseau eaux souterraines de l'agence de l'eau Seine-Normandie (DIREN, 2006).

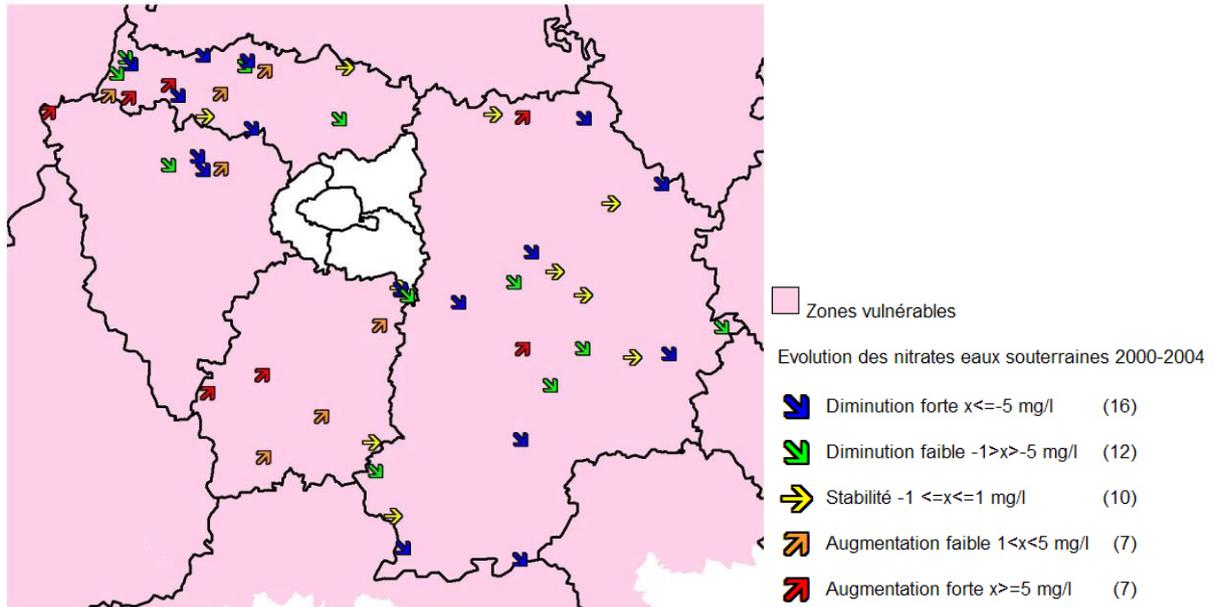


Illustration 15 : Evolution des teneurs moyennes en nitrates des eaux souterraines entre les campagnes 2000-2001 et 2004-2005

Afin d'amorcer une réflexion sur les moyens d'identifier et d'évaluer les tendances d'évolution de la qualité des eaux souterraines et de se mettre en conformité avec les exigences de la DCE, les Agences de l'eau Loire-Bretagne et Seine-Normandie réalisent actuellement, sous convention avec le BRGM, des études sur les tendances d'évolution des pollutions diffuses d'origine agricole à l'échelle des grands bassins hydrogéologiques.

Sur le bassin Loire-Bretagne (Baran et al., 2009), les approches sont les suivantes. Dans un premier temps, la datation des eaux souterraines effectuée de manière synchrone à l'échelle du bassin Loire-Bretagne permet d'apprécier la variabilité du temps de transfert de l'eau d'un secteur du bassin à un autre. Dans un deuxième temps, la sectorisation du bassin Loire-Bretagne est entreprise. Cette sectorisation s'appuie sur des critères autres que la lithologie et intègre le type de réponse piézométrique (existence ou non de cycles intra et interannuels étudiés grâce au calcul des variogrammes des chroniques piézométriques) ainsi que la pression anthropique (qui peut avoir un impact sur les concentrations en nitrate mais pas les types de fluctuations). Cette sectorisation est nécessaire pour travailler sur des zones où les types de réponse (contamination en nitrate et évolution) sont pressentis comme homogènes au sein de chaque zone. Dans un troisième temps, sur la base de cette sectorisation, différents tests statistiques sont mis en œuvre : test de Mann-Kendall pour l'identification des tendances par points, test de Kendall-Régional pour une étude spatialisée des tendances, la méthode de Theil (cf § 2.2.2) pour la quantification des pentes des tendances et la

représentation graphique Box Plot pour une comparaison des états chimiques des secteurs entre deux périodes.

Le calcul des tendances par points est réalisé pour toute la période considérée ainsi que sur des tronçons chevauchants d'une durée de 10 ans. Ainsi, comme le montre l'illustration 16, il est possible de suivre l'évolution des tendances décennales des concentrations en nitrate dans les eaux souterraines et donc de déterminer les périodes d'inversion.

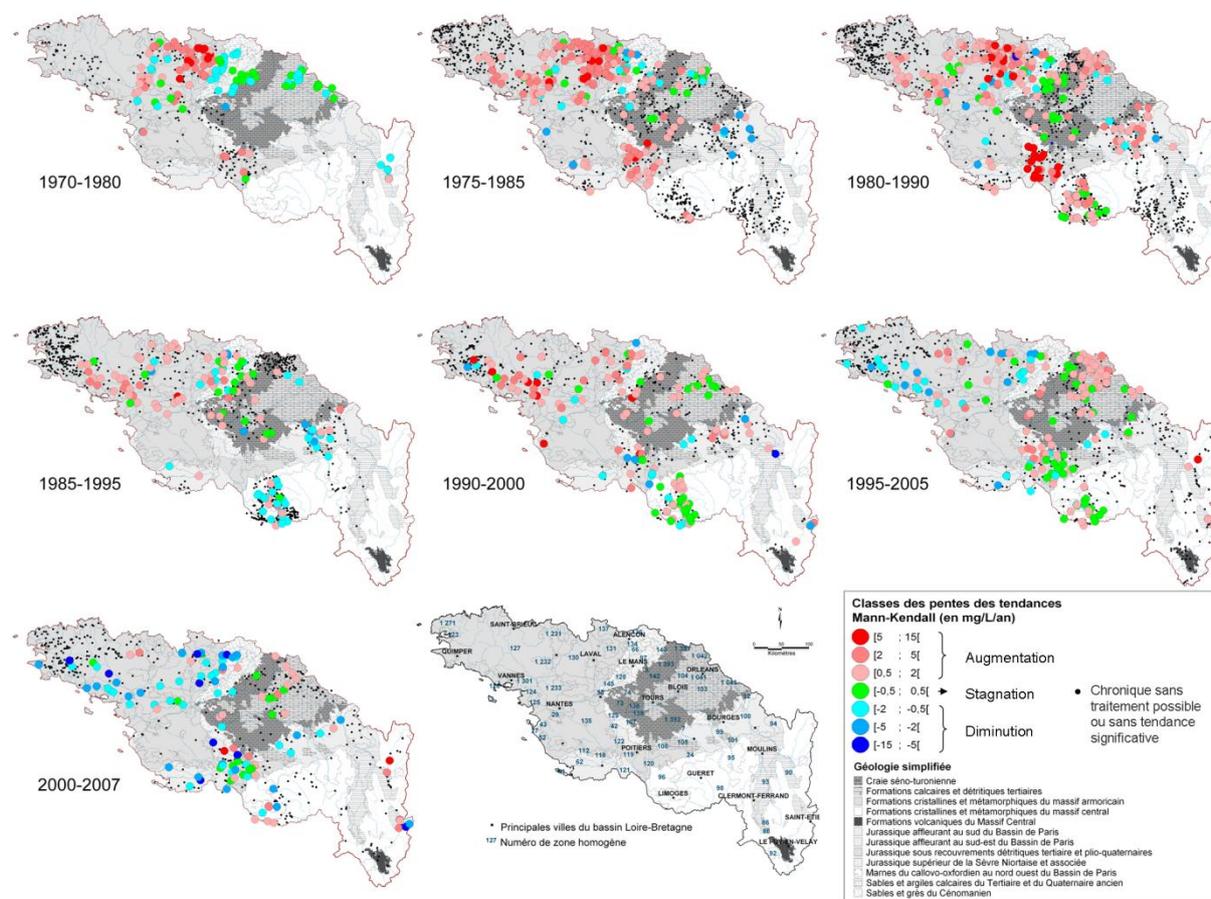


Illustration 16 : Tendances d'évolution des teneurs en nitrate sur le bassin Loire-Bretagne et pentes des tendances (en mg/L/an) étudiées par décennie sur la période 1970-2007 suivant le test statistique de Mann-Kendall régional (extrait de Baran et al., 2009).

Par la suite, le test de Kendall-Regional est mis en œuvre sur les secteurs identifiés comme ayant des réponses homogènes en termes de comportement piézométrique et de pressions. Deux périodes sont comparées (1975-1993 et 1994-2007) afin toujours de rendre compte de possibles changements de tendances. L'illustration 17 et l'illustration 18 montrent les résultats du test statistique Kendall-Regional sur les données de concentrations en nitrate dans les eaux souterraines du bassin Loire-Bretagne. Au-delà des tendances calculées, il

apparaît de nombreuses zones où le calcul est impossible en raison, soit de données insuffisantes, soit de données à pas de temps trop irrégulier pour que le test soit significatif d'un point de vu statistique. Les conclusions sur la nécessité d'améliorer les réseaux et notamment les programmes de suivis sont donc bien illustrées par ce type de résultats.

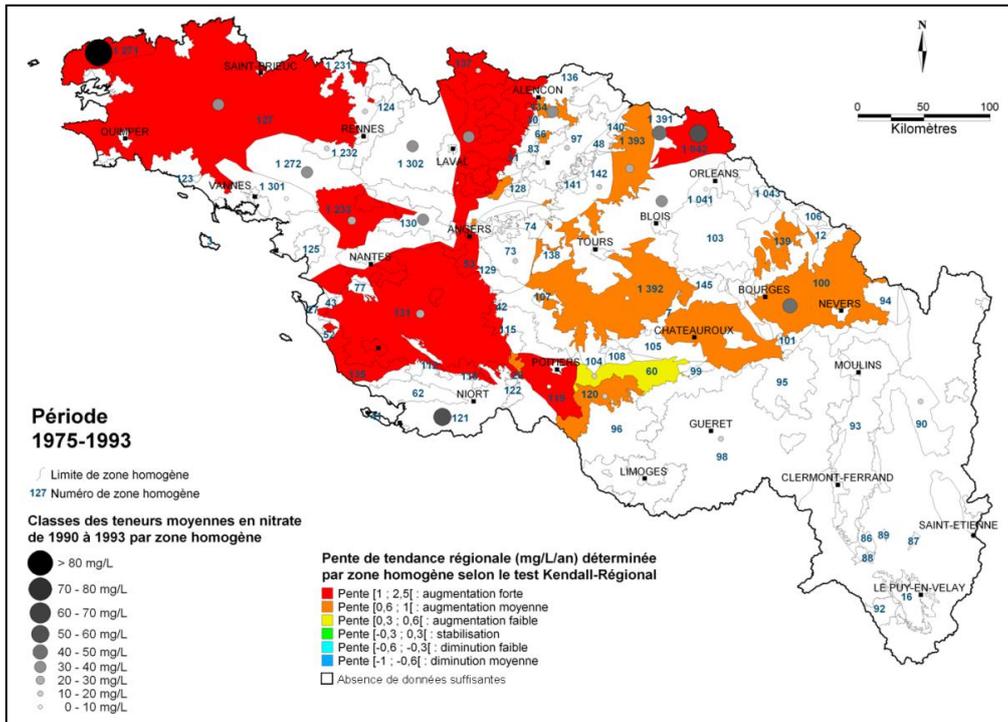


Illustration 17 : Estimation sur le bassin Loire-Bretagne des tendances d'évolution des concentrations en nitrate par zone sur la période 1975-1993 suivant le test Kendall-régional (extrait de Baran et al., 2009).

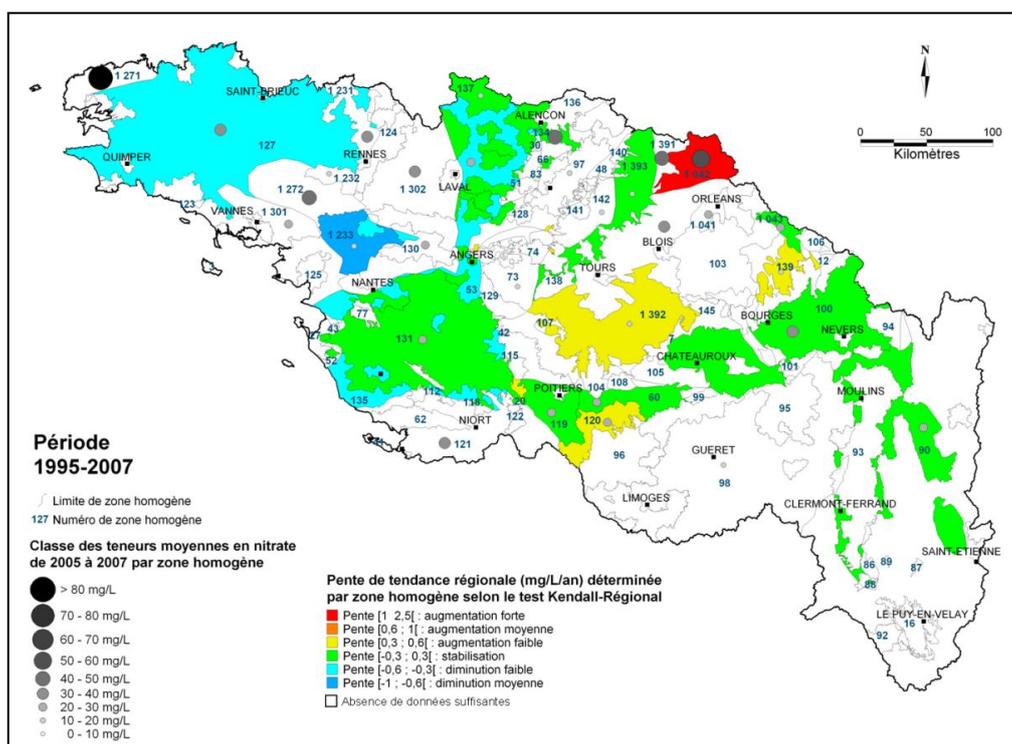


Illustration 18 : Estimation sur le bassin Loire-Bretagne des tendances d'évolution des concentrations en nitrate par zone sur la période 1995-2007 suivant le test Kendall-régional (extrait de Baran et al., 2009).

De plus, un travail d'analyse sur quelques points choisis de la relation entre l'évolution de la piézométrie et l'évolution de la concentration en nitrate montre la nécessité de travailler sur les facteurs explicatifs des tendances observées. En effet, la piézométrie apparaît clairement comme un des facteurs prépondérants dans l'explication des variations en nitrate observées dans les aquifères à nappe libre.

Enfin, l'interprétation des données CFC, non pas dans leur ensemble mais en prenant en compte la sectorisation proposée, permet de mettre en évidence l'existence de relation entre l'âge de l'eau et les concentrations en nitrate dans certaines zones alors que ce type de relation peut ne pas être perçu dans d'autres zones.

Une étude à peu près similaire est actuellement en cours sur le bassin Seine-Normandie (partenariat BRGM-AESN). Des améliorations ont toutefois été apportées à l'étude car après avoir travaillé par décennie sur le bassin Loire-Bretagne et comparé les tendances ainsi obtenues, les travaux réalisés sur le bassin Seine-Normandie permettent d'aller un peu plus loin dans l'évaluation des changements de pente. En effet, le test de CUSUM est appliqué sur l'ensemble des chroniques nitrates afin d'identifier la date de changement majeur de comportement. Le test de Mann-Kendall est alors appliqué sur les derniers tronçons des

chroniques à comportement homogène. Les tendances ainsi déterminées correspondent aux évolutions les récentes et ne sont pas biaisées par de possibles variations brusques dans les données.

Les travaux sur les facteurs explicatifs sont aussi améliorés, notamment par l'automatisation du calcul de la covariance croisée entre les évolutions des concentrations en nitrate dans les eaux souterraines et les fluctuations piézométriques des nappes.

L'étude étant toujours en cours, les résultats ne sont pas encore disponibles.

Cette revue bibliographique montre tout l'intérêt à créer un lien entre la recherche scientifique qui développe et améliore les outils statistiques d'évaluation et de quantification des tendances et les organismes gestionnaires qui n'utilisent que très peu les outils adaptés à leurs problématiques. Plus globalement, cette bibliographie montre l'importance d'adopter une démarche scientifique pour évaluer les tendances d'évolution des concentrations en polluants dans les eaux souterraines.

4 Caractéristiques des données de qualité des eaux souterraines : implications pour l'identification des tendances d'évolution

Le champ d'application de la présente étude est celui de la qualité des eaux souterraines. Or, l'appréciation de l'état qualitatif des aquifères est encore aujourd'hui le plus souvent réalisée par l'intermédiaire d'analyses de la chimie des eaux issues de prélèvements ponctuels tant dans l'espace que dans le temps. La relative difficulté de mise en œuvre des campagnes de prélèvements et le coût engendré par les analyses de la chimie des eaux impliquent la collecte de jeux de données discontinues à partir desquelles sont basées les interprétations. Contrairement à d'autres domaines intéressés par l'analyse de séries chronologiques, l'économie et la météorologie par exemple, où les données sont facilement accessibles et les chroniques sub-continues (flux financier, données de température de l'air par exemple), les données de qualité des eaux souterraines présentent des caractéristiques spécifiques qui dépendent de l'accessibilité aux eaux souterraines, de la variété des paramètres analysés et du potentiel analytique des laboratoires d'analyse. Ces spécificités impactent fortement le choix des méthodes à appliquer pour l'analyse des chroniques et la détermination des tendances d'évolution des concentrations en éléments dissous.

D'après l'article 5.1 de la DCE Eaux souterraines, « les États membres identifient les tendances à la hausse significatives et durables **des concentrations de polluants, groupes de polluants ou d'indicateurs de pollution observées** dans les masses ou groupes de masses d'eau souterraines identifiées comme étant à risque ... ».

La phrase n'est pas très explicite mais dans la mesure où l'évaluation de la tendance ne concerne que les masses d'eau à risque, il apparaît logique que celle-ci ne soit définie **que pour les paramètres à risque** (d'après l'état des lieux de 2004 mis à jour depuis en fonction des nouvelles informations disponibles). Ce choix est d'autant plus réaliste que le point d'inversion de tendance est basé sur un pourcentage de la valeur seuil, elle-même calculée pour les paramètres à risque uniquement. Ces arguments sont également ceux du guide européen qui recommande d'appliquer l'évaluation des tendances et des inversions de tendances aux paramètres à risque (European Commission, 2009).

Il est donc choisi de présenter certaines caractéristiques des données de qualité des eaux souterraines qui peuvent impacter les méthodes de détermination des tendances en sélectionnant des paramètres à risque de non atteinte du bon état qualitatif des masses d'eau souterraines. Les chroniques sont extraites de la base de données ADES sur les eaux souterraines en France (date d'extraction : septembre 2010). L'objectif est de rendre compte pertinences des données disponibles en France pour l'évaluation des tendances, sachant que les caractéristiques des chroniques détaillées orienteront le choix de la (ou des) méthode(s) statistique(s) à appliquer pour l'étude de l'évolution des paramètres.

4.1 Variété et variabilité des données à traiter

Les substances « à risque de non atteinte du bon état qualitatif des masses d'eau souterraines » sont listées par chaque Agence de l'eau lors des états des lieux des masses d'eau. Ces substances appartiennent à des familles très diverses (ions majeurs, pesticides, métabolites pertinents de pesticides, métaux et métalloïdes, polluants émergents...) qui présentent des caractéristiques intrinsèques différentes à la fois au sein d'une même famille mais aussi entre familles de substances. Ces caractéristiques intrinsèques impactent fortement le comportement de chacune des substances dans les différents milieux naturels (vitesse de transfert, stockage, dégradation) ainsi que les concentrations au-delà desquelles ces substances sont considérées comme dangereuses pour l'environnement et/ou la santé humaine. De plus, l'utilisation de ces substances est elle aussi très variable, tant en terme de doses utilisées que de surfaces impactées. Par exemple, les substances d'origine agricole engendrent généralement des pollutions dites diffuses dont les entrées dans les milieux naturels occupent de grandes surfaces. Les substances industrielles engendrent quant à elles le plus souvent des pollutions dites ponctuelles dont les points d'entrée dans les milieux naturels sont plus restreints.

Cette variété de substances, et donc de comportements, ainsi que d'usages se retrouve dans les images partielles de leur présence et des évolutions de leur présence dans les eaux souterraines qu'il est possible d'obtenir grâce aux suivis ponctuels de leurs concentrations (tant dans l'espace que dans le temps). Ainsi, certaines substances facilement analysables et reconnues depuis longtemps comme problématiques montrent des chroniques de longues durées alors que d'autres, plus récemment utilisées ou bien reconnues comme problématiques sont caractérisées par des chroniques courtes.

De plus, certains paramètres montrent des chroniques peu variables dans le temps alors que d'autres sont caractérisés par des fluctuations de grandes amplitudes sur des cycles plus ou moins longs. Ces variations sont souvent dépendantes des milieux dans lesquels les suivis sont effectués et des comportements des substances vis-à-vis de la phase aqueuse. Les aquifères karstiques libres et bien développés sont par exemple plus réactifs face aux événements pluvieux que les aquifères profonds faiblement poreux. Une substance très hydrophile montrera ainsi des variations plus importantes et corrélées avec les phénomènes climatiques dans un milieu réactif que dans un milieu à réponses hydrodynamiques inertielles.

Aussi, les substances très toxiques nécessitent d'être mesurées à de très faibles concentrations de sorte qu'elles peuvent être détectées dans les eaux souterraines sans pour autant dépasser les limites de quantification des laboratoires. Les chroniques sont alors affectées par des données qualitatives du type « inférieur à la LQ » et non plus quantitatives. Enfin, la variabilité spatiale caractérise souvent les données de qualité des eaux souterraines. Cette caractéristique peut avoir une importance capitale lors de la détermination de tendances spatialisées. Elle peut être parfois confondue avec la variabilité temporelle de sorte que les facteurs contributifs à la variabilité spatiale doivent être bien connus s'ils ne sont pas contrôlés.

Toutes ces caractéristiques inhérentes aux données de qualité des eaux souterraines sont à prendre en compte lors des exercices d'analyses statistiques des chroniques temporelles. Aussi, afin d'aller plus loin dans la démonstration, les données de 6 paramètres à risque de non atteinte du bon état qualitatif + 1 paramètre nouvellement recherché* sont extraites d'ADES en septembre 2010 :

- 2 ions majeurs (**Nitrate** NO_3^- et **Sulfate** SO_4^{2-})
- 2 pesticides, (**Atrazine** de la famille des triazines et **Isoproturon** de la famille des urées substituées)
- 1 substance industrielle émergente (**Trichloroéthylène** de la famille des solvants organiques chlorés)
- 1 élément trace (**Arsenic**)
- *1 médicament (**Carbamazépine** de la famille des antiépileptiques)

Les données sont triées et analysées afin d'extraire leurs caractéristiques critiques face à l'application des méthodes statistiques d'étude des séries temporelles.

4.2 Nombre de données, de points et longueur des chroniques

La condition première pour l'étude de l'évolution de la qualité des eaux souterraines est de disposer de suffisamment de données pour que l'analyse soit significative soit d'un point de vue temporel (recul historique suffisant), soit d'un point de vue spatial (nombre et répartition des chroniques suffisant sur le territoire investigué). Or le nombre de données, le nombre de chroniques, le nombre moyen de données par chroniques ainsi que les longueurs moyennes de suivi sont très variables en fonction du type de paramètre comme le montre le Tableau 5.

Tableau 5 : Nombre de données, nombre de chroniques, nombre moyen de données par chronique et longueur moyenne des chroniques pour 7 paramètres extraits d'ADES en septembre 2010.

Paramètre	Nombre de données	Nombre de chroniques	Nombre moyen de données par chronique	Longueur moyenne des chroniques (an)
Nitrate	397 763	30 512	13	9,7
Sulfate	295 402	29 494	10	8,5
Atrazine	116 723	20 608	5,7	4,9
Isoproturon	95 016	19 101	5	3,6
Trichloroéthylène	122 040	25 520	4,7	3,8
Arsenic	113 365	26 928	4,2	4,3
Carbamazépine	46	9	5,1	2,6

Les ions majeurs sont les paramètres pour lesquels l'opérateur dispose à la fois du plus grand nombre de données, du plus grand nombre de points de prélèvements ainsi que des chroniques les plus renseignées. Les pesticides bien connus que sont l'atrazine et l'isoproturon sont suivis sur environ 1/3 de points en moins que les ions majeurs et les chroniques sont, environ et en moyenne, deux fois moins renseignées. Les valeurs sont à peu près identiques pour l'Arsenic et le Trichloroéthylène. La carbamazépine, non encore intégrée dans les réseaux de suivi nationaux de la qualité des eaux souterraines ne dispose que de très peu de données disponibles dans ADES bien que les 9 points sur lesquels des mesures ont été réalisées présentent un nombre moyen de données identique aux pesticides.

La longueur des chroniques étudiées est une caractéristique qui peut influencer grandement sur la qualité des résultats de l'application des méthodes statistiques. Par exemple, seules des tendances de fortes magnitudes peuvent être détectées sur des périodes très courtes et des

chroniques plus longues sont nécessaires à l'examen détaillé de tendances plus ténues. De la même façon, les tests de changement de moyenne ou de variance demandent des chroniques longues, par ailleurs nécessaires à la description de cycles saisonniers ou pluriannuels.

Comme le montre le Tableau 5, les ions majeurs sont de façon attendue les paramètres pour lesquels les suivis sont les plus importants et donc pour lesquels le recul est le plus grand. La première donnée bancarisée dans ADES qui concerne la concentration en nitrate dans les eaux souterraines est ainsi datée du 26 octobre 1945. Pour les sulfates, la première valeur disponible est encore plus ancienne : 26 janvier 1932.

Les autres paramètres à risque extraits d'ADES montrent des longueurs moyennes de chroniques relativement similaires variant de 4,9 ans pour l'atrazine à 3,6 ans pour l'isoproturon. Le suivi des concentrations en pesticides dans les eaux souterraines a débuté au milieu des années 80, la première analyse en atrazine bancarisée dans ADES étant datée du 12 mars 1986 et celle de l'isoproturon du 23 août 1988. Quelques chroniques longues sont ainsi disponibles pour les pesticides, notamment pour l'atrazine avec plus de 23 ans de données au point 03966X0001/FAEP à Champigny en Beauce par exemple.

L'arsenic, bien qu'étudié depuis de nombreuses années (première analyse le 9 décembre 1972) ne montre pas une longueur moyenne des chroniques importante. Ce résultat est en réalité biaisé par le fait qu'un nombre important de points n'est suivi que depuis peu de temps, masquant ainsi la présence dans ADES de chroniques longues. Il en est de même pour le trichloroéthylène dont la première mesure est elle aussi (idem atrazine) datée du 12 mars 1986.

La carbamazépine est un paramètre suivi sur trop peu de points et depuis trop peu de temps (première analyse datée du 14 mai 2002) pour que les chroniques soient aujourd'hui exploitables par une méthode statistique. Ce constat est le même pour de nombreuses substances émergentes dont le suivi n'a débuté que très récemment. Toutefois, afin d'anticiper les futurs traitements statistiques sur ces chroniques temporelles, les suivis de ces substances pourront être adaptés en fonctions des conclusions des tests de robustesse et de puissance des méthodes statistiques réalisés au chapitre 5. Cette remarque est valable pour l'ensemble des paramètres de la qualité des eaux souterraines, mais prend toute son importance lorsque les suivis sont en cours de construction, ce qui est généralement le cas pour les substances émergentes ou, plus particulièrement, les substances qui seront nouvellement intégrées dans les programmes de mesure après leur révision en 2013.

4.3 Pas de temps des prélèvements

La fréquence et la régularité des prélèvements sont des caractéristiques de chroniques qui peuvent influencer le choix de la méthode statistique d'analyse des séries temporelles en vue d'évaluer les tendances d'évolution. L'analyse des chroniques disponibles dans ADES laisse entrevoir de possibles variations importantes dans les fréquences de prélèvements et les longueurs des pas de temps, autant entre les chroniques que lors du suivi d'un paramètre en un point unique. Ces variations peuvent être expliquées par l'évolution des programmes de surveillance en fonction de la révision des réseaux ainsi que par l'appartenance d'un point à plusieurs réseaux. Dans ce dernier cas, la bancarisation pour un même point d'analyses réalisées à des fréquences de prélèvements différentes engendre de grandes variations dans les pas de temps. Certains points ont aussi été suivis historiquement, abandonnés, puis réintégrés dans un réseau récent de mesure de la qualité des eaux souterraines. Ces chroniques sont alors caractérisées par des périodes plus ou moins longues durant lesquelles aucune donnée n'est disponible.

L'illustration 19 montre en exemple l'évolution des concentrations en nitrate au point 02507X0505/P en Mayenne.

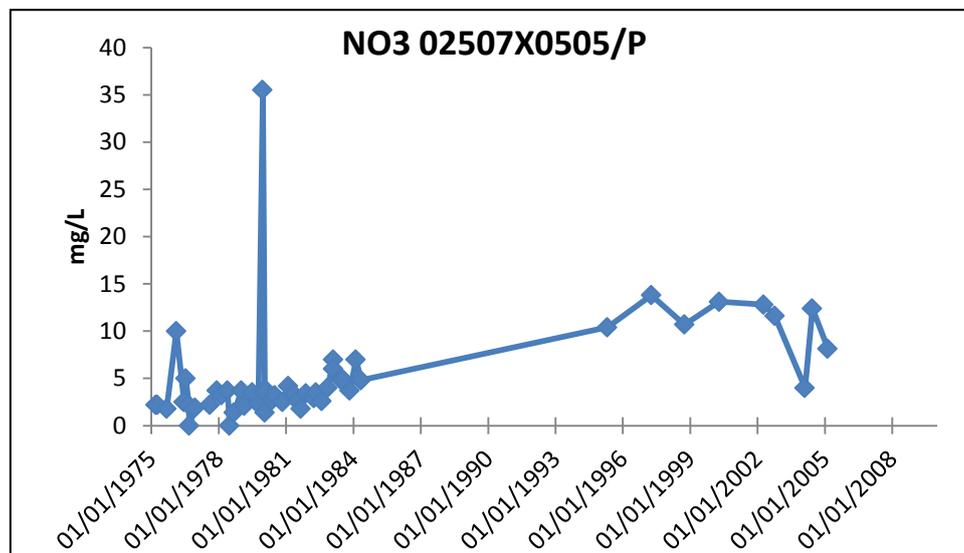


Illustration 19 : Evolution des concentrations en nitrate au point 02507X0505/P (Pre-En-Pail, Mayenne -53-) de 1975 à 2005.

Sur cette chronique particulièrement critique pour la mise en œuvre des tests statistiques, les prélèvements ont débuté en 1975 lorsque ce point fut intégré dans la base de données sur les eaux souterraines ONQUES. Les prélèvements étaient effectués au rythme de 2 par

an de sorte que la chronique apparaît bien renseignée sur la période de suivi [1975 ; 1984]. Les prélèvements en ce point furent abandonnés en 1984 et ce, jusqu'en 1995, période durant laquelle aucune donnée de concentration en nitrate n'est disponible. A partir de 1995, les données sont de nouveau disponibles à la fréquence de 1 mesure tous les deux ans sauf certaines années où les prélèvements sont multipliés (2002 par exemple). Sur une telle chronique, il faudra bien se poser la question de la sensibilité aux variations des pas de temps de prélèvements de la méthode statistique pressentie pour l'étude de la stationnarité. Il est par exemple reconnu que les méthodes basées sur un classement par rang (généralement le cas des méthodes non paramétriques) sont moins sensibles aux variations de longueur des pas de temps que les méthodes paramétriques. Dans ce derniers cas, il sera peut-être nécessaire, soit de régulariser les pas de temps de mesure en évitant toutefois une perte d'informations trop importante, soit de ne réaliser un test de tendance que sur un tronçon de chronique.

Des études poussées menées sur la définition de réseaux de mesure de la qualité des eaux souterraines et superficielles (Lettenmainer, 1978 ; Ryberg and Vecchia, 2006) ont conduit à préconiser des fréquences d'échantillonnages relativement hautes. Ainsi, un pas de temps mensuel semble le compromis idéal pour favoriser la puissance des tests de tendance. Une fréquence élevée des prélèvements permet en effet de comprendre le comportement des paramètres aux points de mesures et est nécessaire pour mettre en évidence l'autocorrélation possible des données. En effet, dans le cas où le paramètre étudié évolue de façon cyclique périodique par exemple, une fréquence élevée des prélèvements permet de bien décrire les cycles d'évolution des concentrations (période et amplitude) et facilite l'analyse de l'autocorrélation du signal. La multiplication des prélèvements au point augmente donc l'information sur le comportement du paramètre étudié et fournit les clefs pour le choix et l'adaptation de la méthode statistique à utiliser.

4.4 Données hors domaine de validité

L'augmentation du nombre de paramètres analysés pour qualifier la qualité des eaux fait émerger un nouveau problème lors des travaux de détermination des tendances. Des substances présentes en des concentrations très faibles dans les eaux naturelles et dans les eaux souterraines en particulier sont aujourd'hui recherchées de sorte que les valeurs reportées peuvent être inférieures aux limites de quantifications tout en étant supérieures aux limites de détection des laboratoires. Les capacités d'analyse des laboratoires

constituent alors un facteur limitant de l'évaluation et de l'interprétation des évolutions de la qualité des eaux souterraines. D'autant que l'évolution des capacités d'analyse des laboratoires vient complexifier le problème avec des limites de quantification et/ou de détection variables dans le temps pour un même laboratoire et souvent différentes entre les organismes d'analyse. Les chroniques affectées par ces mesures sont pourtant sources d'informations et des tests statistiques spécifiques sont développés afin d'exploiter ce type de données (Cf. paragraphe 2.3).

Ce problème ne concerne pas (ou très rarement) les ions majeurs dont l'analyse est relativement simple et les concentrations dans les eaux naturelles pratiquement toujours supérieures aux limites de quantification des laboratoires. Par contre, les substances recherchées à de très faibles concentrations comme les pesticides, certaines substances issues de l'industrie et les métaux et métalloïdes sont souvent bancarisées dans ADES sous la dénomination « concentration inférieure à la LQ ». Le Tableau 6 montre ainsi, pour les 5 substances extraites d'ADES recherchées à de faibles concentrations, le nombre et le pourcentage d'analyses dans le domaine de validité (valeurs vraies), le nombre et le pourcentage de données inférieures à la limite de détection (absence) et le nombre et le pourcentage de données en-dessous des seuils de quantification (présence avérée mais impossibilité de données une valeur vraie). Ces résultats doivent toutefois être considérés avec prudence en raison de la confusion souvent faite entre <LQ et <LD lors de la bancarisation des données dans ADES.

Tableau 6 : Nombre et pourcentage de données de concentration de l'atrazine, de l'isoproturon, du trichloroéthylène, de l'arsenic et de la carbamazépine bancarisées dans ADES (septembre 2010) dans le domaine de validité, inférieures à la limite de détection et en-dessous des seuils de quantification

Paramètre	Données dans le domaine de validité	Données <LD	Données <LQ	Autres
Atrazine	34 045 (29,2%)	59 534 (51%)	22 452 (19,2%)	692 (0,6%)
Isoproturon	3 069 (3,2%)	62 605 (65,9%)	29 267 (30,8%)	75 (0,1%)
Trichloroéthylène	35 846 (29,4%)	63 442 (52%)	22 727 (18,6%)	25 (0%)
Arsenic	26 997 (23,8%)	61 713 (54,5%)	24 618 (21,7)	37 (0%)
Carbamazépine	13 (28,3%)	33 (71,7%)	0 (0%)	0 (0%)

L'exploitation des données d'ADES montre qu'il n'est pas rare de rencontrer des données inférieures aux seuils de quantification et de détection. Pour toutes les substances présentées, il s'agit même du cas le plus courant : l'atrazine, le trichloroéthylène et la

carbamazépine ont des données en dehors du domaine de validité dans plus de 70% des cas, pourcentage qui approche 80% pour l'arsenic et même près de 97% pour l'isoproturon. Ces données, même si elles ne sont pas quantitatives, sont porteuses d'informations et demandent à être valorisées.

Une analyse plus poussées des données d'atrazine montre que seules 1984 chroniques ont 100% de données supérieures à la limite de quantification, ce qui représente 9,4% du nombre total de chroniques. De la même façon, 5,5% des chroniques d'atrazine bancarisées dans ADES ont 100% de données inférieures à la limite de quantification et 59% ont toutes les données inférieures à la limites de détection. Par différence, on montre que plus de 23% des chroniques d'évolution des concentrations en atrazine dans les eaux souterraines sont formées par un composite de données dans le domaine de validité, de données inférieures aux seuils de détection et/ou aux seuils de quantification. En l'état actuel de bancarisation dans ADES, une chronique type de concentration en atrazine est formée de 18% en moyenne de données dans le domaine de validité, de 71% en moyenne de données inférieures aux limites de détection et de 11% en moyenne de données inférieures aux seuils de quantification.

La même analyse réalisée sur les chroniques d'arsenic montre des résultats à peu près similaires avec 12% de chroniques composées à 100% de données dans le domaine de validité, 58,6% de chroniques formées de 100% de données inférieures aux limites de détection et 8,2% de chroniques à 100% de données inférieures aux seuils de quantification. Ainsi, 21,2% des chroniques de concentration en arsenic sont constituées d'un composite de données de différents types (quantitatif et qualitatif). La chronique type de concentrations en arsenic extraite d'ADES et constituées de 19% en moyenne de données dans le domaine de validité, de 68% en moyenne de données inférieures aux seuils de détection et de 13% en moyenne de données inférieures aux seuils de quantification.

L'illustration 20 montre en exemple une chronique d'évolution des concentrations en Arsenic au point 05008X0132/AEP à Poncey-Les-Athee en Cote-D-Or (21).

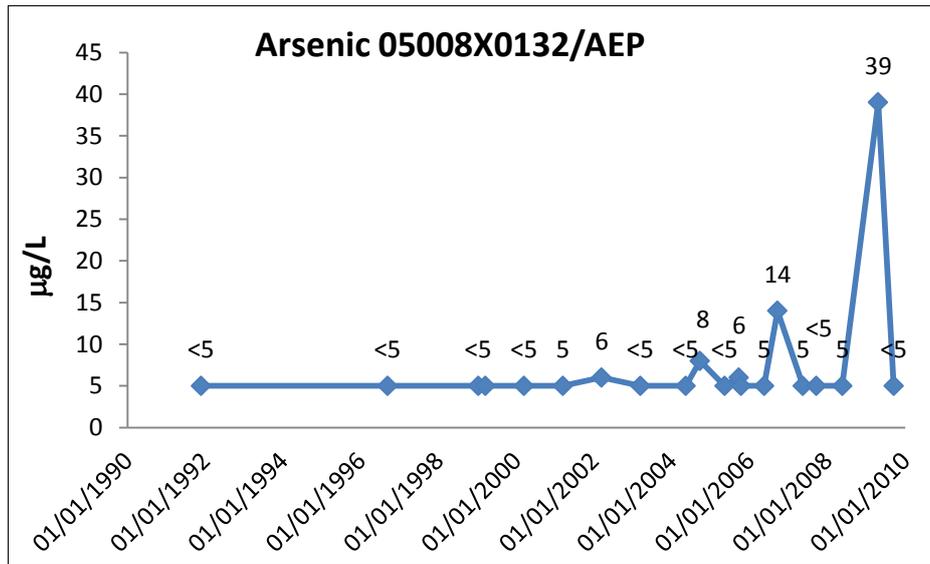


Illustration 20 : Evolution de la concentration en Arsenic ($\mu\text{g/L}$) au point 05008X0132/AEP (à Poncey-Les-Athee, 21).

L'arsenic est un paramètre déclassant de la masse d'eau n°6320 bien que ses concentrations soient souvent reportées inférieures à la limite de quantification (LQ = 5 $\mu\text{g/L}$ dans l'exemple). La chronique est du type : <5 ; <5 ; <5 ; <5 ; <5 ; 5 ; 6 ; <5 ; <5 ; 8 ; <5 ; 6 ; <5 ; 5 ; 14 ; 5 ; <5 ; 5 ; 39 ; <5 ; c'est-à-dire composée à 55% de données inférieures à la limite de quantification. La question de l'exploitation possible de ce type de données pour caractériser l'évolution des concentrations en arsenic au sein de cette masse d'eau est alors largement posée. Des pistes de réflexions sont développées au paragraphe 2.3 afin de voir dans quelles mesures il est possible de valoriser ces chroniques.

4.5 Distribution des données

Le type de distribution des données au sein d'une chronique oriente le choix de la (ou des) méthode(s) statistique(s) qu'il est possible d'utiliser pour étudier sa stationnarité. En effet, l'une des conditions initiales fortes mentionnée pour l'utilisation des tests paramétriques est une distribution normale des données. Or, il n'est pas évident que toutes les chroniques d'évolution des concentrations en polluants dans les eaux souterraines ont des données qui suivent une distribution normale.

Afin de rendre compte de la distribution des données constitutives des chroniques d'évolutions des concentrations en polluants dans les eaux souterraines bancarisées dans ADES, des tests de normalité sont effectués sur les paramètres nitrate et sulfate. Il est

décidé de ne travailler que sur ces deux paramètres car, parmi les 7 paramètres extraits d'ADES en septembre 2010, ce sont les seuls dont pratiquement la totalité des données est reportée dans le domaine de validité analytique des laboratoires.

Les tests de normalités sont réalisés sur deux échantillons de 100 chroniques chacun (1 échantillon pour les nitrates et un échantillon pour les sulfates). Les chroniques sont prises au hasard parmi tous les points qui disposent d'au moins une valeur de concentration en nitrate ou sulfate. Les méthodes appliquées pour évaluer la normalité des données sont les tests de Shapiro-Wilk, d'Anderson-Darling, de Lilliefors et de Jarque-Bera cités au paragraphe 2.2. Toutes ces méthodes testent la même hypothèse H0 (les données suivent une distribution normale). Lorsque l'hypothèse H0 est rejetée, les données suivent une autre distribution que la loi normale limitant ainsi les potentialités d'applications des méthodes statistiques paramétriques.

Le Tableau 7 résume les résultats des tests de normalité sur les chroniques nitrate et sulfate.

Tableau 7 : Tests de normalité des données issues de 100 chroniques de concentration en nitrate et de 100 chroniques de concentration en sulfate prises aléatoirement dans ADES en septembre 2010.

Pourcentage de rejet de H0				
Paramètre	Test de Shapiro-Wilk	Test d'Anderson-Darling	Test de Lilliefors	Test de Jarque-Bera
NO₃⁻	44%	41%	38%	20%
SO₄²⁻	24%	28%	24%	19%

En admettant que les 2 x 100 chroniques prises aléatoirement constituent des échantillons représentatifs de l'ensemble des populations « chroniques nitrates » et « chroniques sulfates », l'analyse montre que, suivant les tests appliqués, entre 20% et 44% des chroniques d'évolution des concentrations en nitrate ont des données qui ne suivent pas une distribution normale. Pour les sulfates, entre 19% et 28% des chroniques ont des données non normales. Ces chiffres sont tout à fait significatifs et indiquent que dans environ ¼ des cas, les conditions initiales requises pour l'utilisation des tests paramétriques d'évaluation de la stationnarité des chroniques ne sont pas respectées.

Il y a donc tout intérêt à prévoir des méthodes qui ne soient pas biaisées par une distribution autre que normale des données afin de s'affranchir de cette vérification avant de débiter l'analyse statistique.

4.6 Autocorrélation des données

Les concentrations en éléments dissous dans les eaux naturelles, qui définissent la « qualité des eaux », montrent généralement de grandes variations. Ces variations peuvent être expliquées soit par des facteurs naturels, soit par des facteurs anthropiques comme les sources de pollution ponctuelles ou diffuses, soit encore par une combinaison de ces différents facteurs. Les fluctuations de concentrations en éléments dissous peuvent être corrélées avec les variations hydrodynamiques des aquifères, les usages des sols dans les zones de recharge, les contextes géologiques ainsi qu'avec les pratiques de gestions des aires d'alimentation des nappes. De plus, les facteurs qui guident l'évolution de la qualité des eaux souterraines et les relations qui lient ces différents paramètres peuvent varier d'un élément à l'autre pour un même contexte.

De ce fait, la détection des tendances d'évolution de la qualité des eaux naturelles est souvent compliquée par la présence de variations plus ou moins cycliques. Ces évolutions, généralement liées aux comportements hydrodynamiques des aquifères, peuvent être importantes au regard des changements de la qualité des eaux liés à l'activité humaine.

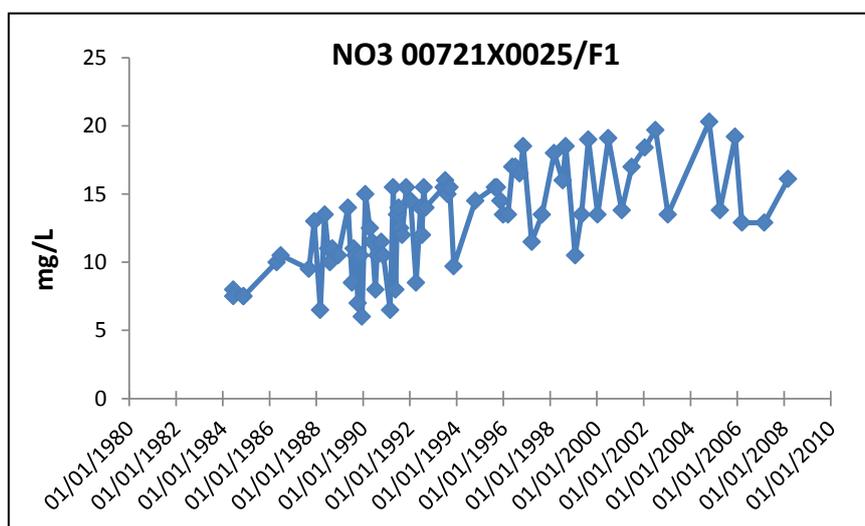


Illustration 21 : Evolution des concentrations en nitrate au point 00721X0025/F1 dans la masse d'eau 3507 à l'Est de Cherbourg.

Dans l'exemple de l'illustration 21, les concentrations en nitrate au point 00721X0025/F1, dans la masse d'eau n°3507 (« socle du bassin versant des cours d'eau côtiers »), montrent une cyclicité saisonnière importante au regard de l'évolution pluriannuelle. Les concentrations de basses eaux sont en effet plus importantes que les données de hautes

eaux de 5 à 6 mg/L en moyenne, alors que l'augmentation générale depuis 1984 peut être estimée visuellement à environ 8 mg/L. Cette périodicité dans le signal engendre de l'autocorrélation dans les données qui perturbe généralement les tests statistique classiques de stationnarité. Estimer l'autocorrélation des données revient à se demander si une valeur observée à un temps t dépend de ce qui a été observé dans le passé ou non. L'autocorrélation des données est estimée en calculant l'autocorrélogramme de la série temporelle. Lorsque la valeur de l'autocorrélation dépasse les bornes des intervalles de confiance (95% la plupart du temps) alors la série est affectée par ce phénomène. L'autocorrélogramme de l'évolution des concentrations en nitrate au point 00721X0025/F dont la chroniques est présentée en Illustration 21, est visible dans l'illustration 22.

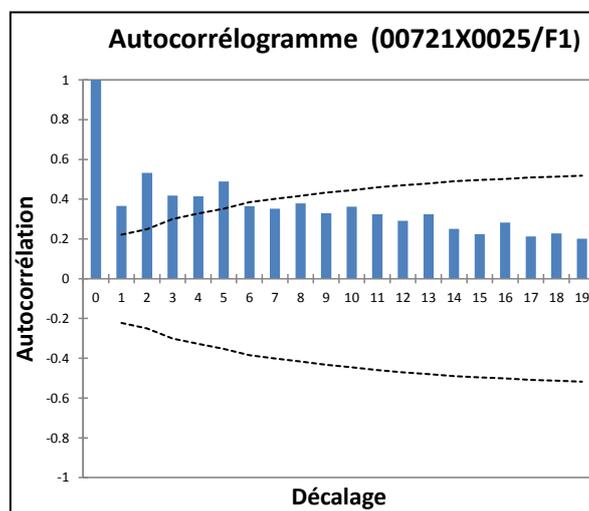


Illustration 22 : Autocorrélogramme (fonction d'autocorrélation FAC) de la chronique d'évolution des concentrations en nitrate au point 00721X0025/F1 et intervalles de confiance.

Les résultats montrent une autocorrélation importante des données jusqu'au 5^{ème} ordre, c'est-à-dire qu'une donnée observée à l'instant t dépend en moyenne dans la série des valeurs des 5 données qui la précèdent. Sur cette chronique, une cyclicité saisonnière forte est mise en évidence. Toutefois, les variations temporelles cycliques de la qualité des eaux souterraines, généralement liées à l'hydrodynamisme des nappes, présentent des périodes pouvant être très différentes d'une chronique à l'autre. Les échelles de temps rencontrées sont la journée, la saison, l'année, la décennie voire des périodes de plus longues durées encore. Or, ces cycles peuvent largement influencer la détermination des tendances d'évolution de la qualité des eaux souterraines, notamment lorsqu'ils sont établis sur des longues périodes. Le danger provient de la généralisation d'une tendance d'évolution de la qualité des eaux qui ne serait déterminée que sur une demi-période de cycle, c'est-à-dire sur la partie croissante ou décroissante uniquement du signal sinusoïdal.

Dans l'exemple présenté en Illustration 23, les fluctuations des concentrations en nitrate au point 02552X0023/AEP évoluent selon un cycle pluriannuel dont la période peut être estimée grossièrement à 6 années.

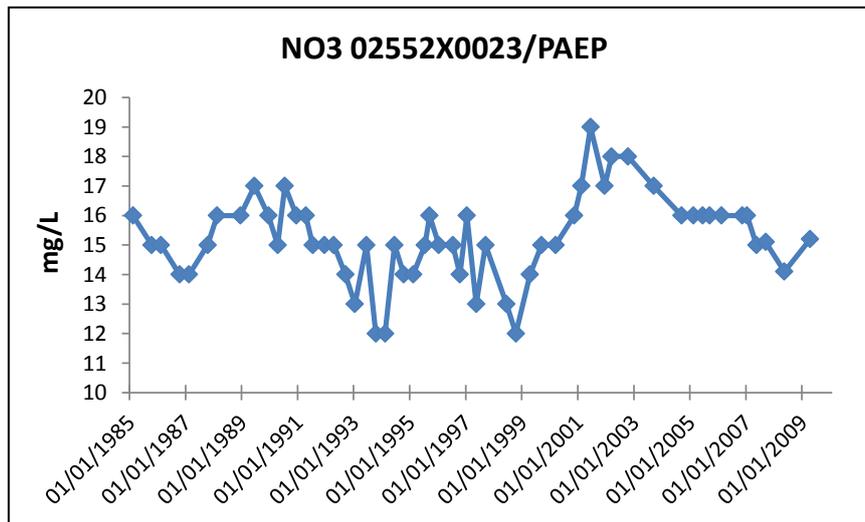


Illustration 23 : Evolution des concentrations en nitrate au point 02552X0023/AEP à Jouy, en Eure-et-Loir.

L'autocorrélogramme de cette chronique, Illustration 24, révèle une autocorrélation des données jusqu'au troisième ordre engendrée par une cyclicité différente de celle du premier exemple présenté.

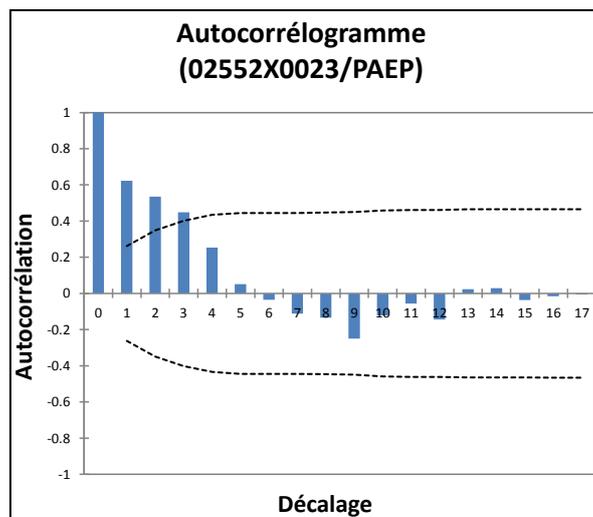


Illustration 24 : Autocorrelogramme (fonction d'autocorrélation FAC) de la chronique d'évolution des concentrations en nitrate au point 02552X0023/AEP et intervalles de confiance.

Comme pour les autres caractéristiques critiques des données et afin d'estimer la qualité des données sur les eaux souterraines dans ADES, les autocorrélogrammes ont été calculés pour les 100 chroniques nitrates et les 100 chroniques sulfates extraites aléatoirement de la base de données ADES (extraction septembre 2010). Le Tableau 8 montre le pourcentage de chroniques autocorrélées (1^{er} et 2nd ordre) par rapport à l'ensemble des chroniques extraites.

Tableau 8 : Autocorrélation du 1^{er} et 2nd ordre des données issues de 100 chroniques de concentration en nitrate et de 100 chroniques de concentration en sulfate extraites d'ADES

Autocorrélation des données		
Paramètre	1 ^{er} ordre	2 ^{ème} ordre
NO₃⁻	31%	19%
SO₄²⁻	25%	4%

Il apparaît que les données de nitrate ont une plus grande propension à être autocorrélées que les chroniques d'évolution des concentrations en sulfate dans les eaux souterraines. Près d'un tiers des chroniques nitrates présentent de l'autocorrélation du premier ordre alors que ce phénomène ne concerne qu'un quart des séries temporelles de sulfate. L'autocorrélation de second ordre affecte très peu les données de sulfate alors qu'1/5^{ème} des chroniques nitrates montre ce type d'autocorrélation.

L'autocorrélation n'est donc pas un phénomène marginal dans les chroniques d'évolution de la qualité des eaux souterraines, tout du moins en ce qui concerne l'évolution des polluants majeurs.

La mise en évidence de l'autocorrélation est favorisée par un nombre important de données dans les chroniques. Ainsi, comme le montre l'illustration 25, les différences entre la valeur d'autocorrélation observée (jusqu'au 1^{er} ordre (en bleu) et jusqu'au 2nd ordre (en rouge)) et la borne supérieure de l'intervalle de confiance 95% diminuent suivant une loi logarithmique en fonction de l'augmentation du nombre de données dans les séries. Cette différence traduit l'importance de l'autocorrélation, considérée comme nulle pour toutes les valeurs > 0 et présente dans les séries temporelles pour toutes les valeurs < 0. Il est alors possible de déterminer des valeurs seuils de nombre de données dans les séries à partir desquelles la probabilité de mettre en évidence de l'autocorrélation devient très importante. Ces valeurs correspondent aux coordonnées des intersections entre les lois logarithmiques décroissantes et les abscisses. Pour les nitrates, la probabilité de mettre en évidence de

l'autocorrélation des données jusqu'au premier ordre augmente fortement lorsque le nombre de points dans la chronique dépasse 27 (41 pour une autocorrélation jusqu'au 2nd ordre). Pour les sulfates, le seuil de points est de 35 pour l'autocorrélation jusqu'au 1^{er} ordre et de 60 pour le 2nd ordre.

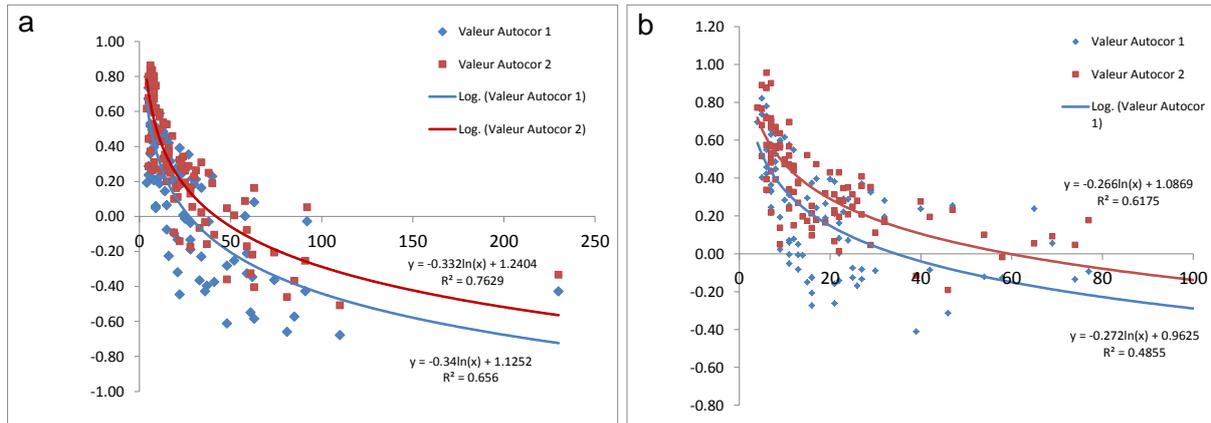


Illustration 25 : Evolution de la différence entre la valeur observée d'autocorrélation et la borne supérieure de l'intervalle de confiance 95% en fonction du nombre de points dans les chroniques pour les autocorrélations du 1^{er} ordre (en bleu) et du 2nd ordre (en rouge) pour a) les 100 chroniques nitrates extraites d'ADES et b) les 100 chroniques sulfates extraites d'ADES.

Cette démonstration confirme l'intérêt de choisir (ou bien d'initier l'acquisition) des chroniques aux fréquences de prélèvements les plus fortes, c'est-à-dire qui portent le maximum d'informations, pour comprendre le comportement des polluants dans les eaux souterraines. Pour la mise en œuvre des tests statistiques lorsque les données sont autocorrélées, il est possible d'utiliser des méthodes qui prennent en compte ce phénomène et dont le principe repose sur un ré-échantillonnage (calcul d'un nombre équivalent de données) des données brutes acquises initialement.

D'une manière générale, lors des exercices de détection des tendances d'évolution de la qualité des eaux, il est important de prendre le temps de bien analyser les chroniques mises à disposition. Toutes les caractéristiques des données détaillées dans ce chapitre peuvent en effet influencer le choix des tests à appliquer. Leur connaissance peut aussi aider à identifier les possibles facteurs qui expliquent les tendances observées. En effet, si des phénomènes d'autocorrélation sont mis en évidence dans une chronique, il est possible que le signal étudié soit corrélé avec d'autres phénomènes externes plus ou moins cycliques responsables d'une part importante des évolutions observées. La compréhension du comportement global d'un polluant peut ainsi, dès l'analyse des

caractéristiques des données s'y rapportant, être déjà grandement améliorée et ce, même avant la mise en œuvre des tests statistiques.

5 Tests de robustesse et de puissance des outils statistiques les plus couramment utilisés dans le domaine de la qualité des eaux

La revue bibliographique des travaux menés pour l'évaluation des tendances d'évolution de la qualité des eaux montre que les tests choisis par la majorité des opérateurs sont peu nombreux : régression linéaire, Mann-Kendall et ses déclinaisons (Kendall Saisonnier, Kendall modifié), test de Spearman et Boîtes à moustaches le plus souvent. Or, le choix de la méthode à appliquer est très dépendant de la qualité des données à tester. Seront ainsi préférées les méthodes robustes compte tenu des caractéristiques des données qui ont tendance à biaiser le résultat des tests (augmentation du risque de première espèce). De la même manière, seront recherchés les tests les plus puissants, c'est-à-dire les méthodes qui ont la plus grande probabilité de rejeter à raison l'hypothèse H_0 .

Les biais qui peuvent perturber le bon fonctionnement des tests sont engendrés par certaines caractéristiques (dites critiques relativement aux tests) des séries chronologiques telles que :

- le nombre de données disponibles dans la chronique,
- la distribution non normale des données,
- les cycles marqués dans l'évolution des données,
- les changements abrupts,
- les données aberrantes dans la série,
- les données manquantes,
- l'autocorrélation des données.

Afin de rendre compte de la robustesse (risque de première espèce) des tests les plus couramment utilisés pour l'évaluation des tendances d'évolution de la qualité des eaux face à ces caractéristiques, il est proposé d'identifier les méthodes qui ont la propriété de rester non biaisées lorsque les observations ne respectent plus les conditions initiales d'applicabilité du test.

Parallèlement, les méthodes couramment employées seront testées afin de rendre compte de leur puissance à détecter à raison des tendances dans les séries à données critiques.

Parmi les tests rencontrés dans la littérature, la régression linéaire par moindre carré, les tests de Mann-Kendall et Mann-Kendall modifié, le test de Spearman, le test de Cox-Stuart et, dans un second temps le test Kendall-Saisonnier, sont sélectionnés pour l'évaluation de

leur robustesse et de leur puissance face aux caractéristiques critiques des données (listées ci-dessus) des chroniques d'évolution de la qualité des eaux souterraines.

5.1 Processus des tests de robustesse et de puissance

Devant l'impossibilité de sélectionner, parmi l'ensemble des données de qualité des eaux souterraines disponibles dans ADES, des jeux de chroniques (appelés « collections » par la suite) suffisamment importants pour être significatifs d'un point de vue statistique et spécifiques chacun d'une caractéristique critique pouvant biaiser les tests statistiques, il a été décidé de simuler les séries temporelles sur lesquelles les tests sont réalisés.

Cette approche présente plusieurs avantages :

- il n'est pas nécessaire de balayer l'ensemble des bases de données existantes pour extraire les chroniques d'intérêts,
- il est possible de simuler un nombre très important de chroniques ayant toutes les mêmes caractéristiques (longueur, nombre de mesures, écart type de la série, distribution, amplitude et durée de cycles ...)
- toutes les caractéristiques qui définissent les chroniques sont contrôlées et modifiables à volonté de sorte qu'il est possible de faire varier un seul paramètre en fixant les autres à une valeur donnée.

Les collections de chroniques sont simulées en reprenant des caractéristiques réelles mises en évidence après analyse des données d'ADES (Cf. chapitre 4). D'autres travaux menés par le BRGM et les Agences de l'eau Loire-Bretagne et Seine-Normandie sur la caractérisation du comportement des chroniques d'évolution des concentrations en nitrates ont aussi servi de base à la simulation des séries à variations cycliques marquées (Baran et al., 2009). De la sorte, même si les chroniques sur lesquelles sont réalisés les tests de robustesse et de puissance sont purement fictives, leurs caractéristiques sont toujours proches de ce qu'il est possible de rencontrer dans la réalité et les résultats des tests sont directement transférables au domaine de la qualité des eaux souterraines.

Dans ce chapitre, les termes suivants sont employés :

- **Calendrier** : suite croissante de dates
- **Chronique** : suite de valeurs selon un calendrier donné
- **Collection** : ensemble de couples de calendriers et chroniques.

Les calendriers et chroniques d'une même collection, lorsque cette dernière est créée par simulation, suivent généralement les mêmes lois de simulation.

L'algorithme se divise en deux parties :

- Les collections sont d'abord créées. Elles sont stockées dans les différents fichiers d'un même répertoire. Chaque fichier, en plus de la collection, contient les paramètres utilisés pour sa création.
- Les collections sont ensuite testées. Itérativement, chaque fichier est ouvert, la collection est soumise aux tests, et le résultat est écrit dans un fichier texte.

Ce processus de test permet de créer divers cas de figures théoriques simulant des chroniques rencontrées dans les bases de données existantes sur la qualité des eaux. Les tests de tendance sont appliqués sur les collections qui regroupent chacune 10 000 couples calendrier/chronique. Au sein d'une même collection, les calendriers et chroniques suivent tous la même loi de création et possèdent ainsi les mêmes caractéristiques critiques pour l'application des tests statistiques.

5.2 Nomenclature des collections

5.2.1 Modèles de calendriers et de chroniques

Les modèles de calendriers de prélèvements et de chroniques sont créés afin de couvrir toutes les situations possibles et ainsi mieux évaluer les performances des méthodes statistiques face à la réalité de la qualité des données des eaux souterraines. Le Tableau 9 et le Tableau 10 listent les modèles de calendriers (caractérisés par leur longueur L et leur nombre de prélèvements N) et les modèles de chroniques (moyenne m , écart type e , pente de tendance a et composante pépite p) utilisés pour la création des collections. Les modèles utilisés pour les calendriers permettent de reproduire la variété des fréquences de prélèvements observées (Cf. paragraphe 4.3). Les modèles de chroniques, basés à la fois sur l'observation des distributions des données dans ADES (Cf. paragraphe 4.5) et sur les études menées par le BRGM et les Agences de l'eau Loire-Bretagne et Seine-Normandie sur les tendances d'évolution des nitrates dans les eaux souterraines (Baran et al., 2009 ; Baran et al., à paraître), permettent de générer une variété de types de séries temporelles qui couvre l'ensemble des cas réels observés.

Le premier cas de figure (calendrier régulier **R** / chronique à loi gaussienne pure **G0**) correspond au cas théorique le plus « idéal » mais peu réaliste. Les autres modèles de calendriers sont générés afin de simuler des pas de temps de prélèvement plus ou moins réguliers et/ou aléatoire (calendriers uniforme **U** et bruité **B**) ainsi que des fréquences de mesures plus ou moins grandes (calendrier troué **T**).

Tableau 9 : Modèles de calendrier

Code	Modèle	Paramètres	Description
R	régulier	L, N	espacement régulier des points de mesure (cas particulier de B) de 1 à L
U	uniforme	L, N	tirage des points de mesure selon une loi uniforme entre 1 et L
T	troué	L, N0, N	sélection, dans un calendrier régulier de N0 points de 1 à L, de N points au hasard
B	bruité	L, N, s	espacement régulier, décalé par un bruit gaussien, avec réordonnancement des points ; le premier point est toujours 1, le dernier L, les autres entre 1 et L

Tableau 10 : Modèles de chroniques

Code	Modèle	Paramètres	Description	Stationnarité	Autocorrelation
G0	loi gaussienne pure	m, e	$\mathcal{L}(t) = \mathcal{G}(m, e)$	Oui	Non
Ga	Loi gaussienne + tendance linéaire	m, e, a	$\mathcal{L}(t) = \mathcal{G}(m, e) + a \cdot t$	Non	Oui
Ln0	loi log normale	m, e	$\mathcal{L}(t) = \mathcal{LN}(m, e)$	Oui	Non
Lna	loi log normale + tendance linéaire	m, e, a	$\mathcal{L}(t) = \mathcal{LN}(m, e) + a \cdot t$	Non	Oui
V0	variogramme	m, e, p	$\mathcal{L}(t) = \mathcal{N}(\gamma_1) + m$	Oui	Oui
Va	variogramme + tendance linéaire	m, e, p, a	$\mathcal{L}(t) = \mathcal{N}(m, \gamma_1) + m + a \cdot t$	Non	Oui
W0	variogramme à cycle long	m, e, p	$\mathcal{L}(t) = \mathcal{N}(\gamma_2) + m$	Oui	Oui

Les différents modèles de chroniques permettent de simuler des distributions de données normale (**G0**), log normale (**Ln0**) et cyclique à plus ou moins longues périodes par

l'intermédiaire de variogrammes (**V0** et **W0**). Les modèles de chroniques stationnaires sont doublés en ajoutant une tendance linéaire de pente a (**Ga**, **Lna** et **Va**). Cette tendance supprime la stationnarité des chroniques et augmente l'autocorrélation des données temporelles. Le phénomène d'autocorrélation est présent par défaut dans les chroniques à distributions simulées par variogrammes.

Par construction, les modèles de chroniques **G0**, **Ln0** et **V0** sont des cas particuliers des modèles de chroniques **G**, **Ln** et **V** pour des pentes $a = 0$.

Les modèles de variogrammes sont inspirés des études menées par le BRGM et les Agences de l'eau Loire-Bretagne et Seine-Normandie sur les tendances d'évolution des nitrates dans les eaux souterraines (Baran et al., 2009 ; Baran et al., à paraître). Les auteurs ont en effet montré la possibilité de décrire la distribution des données d'évolution des concentrations en nitrates dans les eaux souterraines par des variogrammes théoriques à cycles plus ou moins longs de la forme :

- $\gamma_1 = e^2 \times (\text{Pépite}(p) + (1-p) \times (\text{Cosinus}(0,9 ; 365 j) + \text{Exponentielle}(0,1 ; 3650 j)))$ pour des cycles annuels et
- $\gamma_2 = e^2 \times (\text{Pépite}(p) + (1-p) \times (\text{Cosinus}(0,9 ; 3650 j) + \text{Exponentielle}(0,1 ; 3650 j)))$ pour des cycles à période de 10 ans.

Les composantes des variogrammes sont les suivantes :

- Pépite (p) = $h \mapsto p$ si $h \neq 0$, 0 si $h = 0$
- Cosinus (palier_relatif, période) = $h \mapsto \frac{1}{2} \times \text{palier} \times (1 - \cos(h \div \text{période}))$
- Exponentielle (palier_relatif, portée) = $h \mapsto \text{palier_relatif} \times (1 - e^{-h \div \text{portée}})$

5.2.2 Paramètres utilisés

Les paramètres des calendriers et chroniques sont expliqués dans le Tableau 11. Pour chaque paramètre, il a été choisi un cas de base, dont la valeur est indiquée en gras dans le tableau.

La moyenne a a été prise fixe à zéro. En effet, la moyenne n'a pas d'influence sur la robustesse et la puissance des tests de tendance.

Le modèle de calendrier **R** est le cas de base de **B**. Les tests sont donc réalisés uniquement selon **R**, excepté lors d'études de l'influence de l'écart-type **e**.

Les valeurs des paramètres choisis engendrent 58 320 combinaisons possibles (sans tenir compte des lois de création des calendriers et chroniques). Il est décidé de faire varier les paramètres séparément avec pour les autres paramètres, la valeur de base indiquée en gras dans le Tableau 11. Cette approche représente 39 combinaisons possibles.

Tableau 11 : Paramètres des modèles de calendriers et de chroniques

Code	Nom	Description	Valeurs possibles
L	longueur	durée du calendrier	en années (à multiplier par 365) 20 ; 10 ; 5 ; 3 ; 1 ; 0,5
N	effectif	nombre de points du calendrier	240 ; 100 ; 50 ; 20 ; 10 ; 9 ; 8 ; 7 ; 6
N0	effectif complet	effectif de la liste à partir de laquelle est tiré le calendrier	3000
s	bruit de calendrier	écart-type du bruit appliqué sur le calendrier	0 ; 3 ; 7 ; 10
m	moyenne	moyenne de la chronique sans tendance	0
e	écart-type	écart-type de la chronique sans tendance	0,4 ; 2 ; 4 ; 10 ; 20 ; 40
a	pente	pente de la tendance appliquée à la chronique stationnaire	sur 20 ans (à diviser par 7 300) 0 ; 0,1 ; 0,5 ; 1 ; 2
p	pépite	pourcentage de pépite dans le variogramme choisi	(%) 5 ; 10 ; 20 ; 30 ; 50 ; 70 ; 80 ; 90 ; 95
t	extension (nbcol)	nombre de simulations d'une collection	10 000

Remarques :

- i) La valeur de pépite $p = 100\%$ aurait pu constituer le cas de base. Dans ce cas de figure, G_0 , resp. G , deviennent les cas de base de V_0 , resp. V . Cependant, cette configuration accroît de façon trop importante le temps de calcul.
- ii) Il aurait été possible de prendre comme paramètre, non pas **N** mais directement la fréquence de mesure $N \div L$. Toutefois, les paramètres choisis permettent de tester indépendamment l'effet du caractère plus ou moins aléatoire des calendriers et celui de fréquences de mesures plus ou moins grandes.

5.2.3 Compatibilité entre calendriers et chroniques

Comme le montre le Tableau 12, la méthode de génération des calendriers dépend du choix de la loi de distribution des données pour la création des chroniques temporelles.

Tableau 12 : Méthodes de génération des calendriers selon les lois de distribution des chroniques

	G0	Ga	Ln	Lna	V0, W0	Va
régulier	gaussien	gaussien tendance	gaussien	gaussien tendance	cholesky	cholesky tendance
uniforme	gaussien	gaussien tendance	gaussien	gaussien tendance		
troué					cholesky_ sélection	cholesky_séle ction tendance
bruité	gaussien	gaussien tendance	gaussien	gaussien tendance		

La méthode «gaussien» correspond au tir de valeurs selon une loi gaussienne. La méthode «cholesky» utilise la décomposition de Cholesky de la matrice des covariances, elle-même définie par le variogramme, l'écart-type, et le calendrier. Il s'agit donc de simulations conditionnelles des chroniques temporelles. Les méthodes «tendance» ajoutent une tendance linéaire.

Le modèle **U** ne peut pas être utilisé pour des simulations de chronique utilisant des variogrammes. En effet le calcul par la méthode «covariance» demande un temps de calcul trop important. C'est pourquoi il a été proposé le modèle **T**, qui permet d'utiliser la méthode «cholesky_sélection».

De même, le modèle de calendrier **B** est utilisé uniquement avec les modèles de chronique **G0** et **Ga**. En effet, l'appliquer à des simulations variographiques demanderait de calculer, pour chaque chronique, la décomposition de Cholesky d'une matrice $N \times N$, ce qui, là aussi, demande un temps de calcul important.

5.2.4 Génération des collections

Il est choisi de créer des collections qui s'écartent du cas de base selon un seul paramètre (autre que les choix de modèles de calendriers et chroniques). L'ensemble des paramètres d'une même collection sont ainsi fixés à leur valeur de base sauf celui dont l'effet sur la

robustesse et la puissance des tests de tendance est analysé. De la sorte, 395 collections sont créées et testées.

5.2.5 Méthodes statistiques d'évaluation des tendances testées

Les tests de robustesse et de puissance concernent les six méthodes statistiques d'évaluation des tendances présentées dans le Tableau 13. Ces méthodes sont choisies en raison de leur utilisation courante dans le domaine de l'évaluation des tendances d'évolution de la qualité des eaux, mais aussi en raison de la disponibilité des codes de calculs dans MATLAB®, logiciel de calcul numérique utilisé pour cet exercice. Tous les tests sont réalisés avec un risque acceptable de première espèce $\alpha = 5 \%$.

Tableau 13 : Méthodes statistiques d'évaluation de tendance testées

Code	Test	Paramétrique	Cible
RL	Régression linéaire par moindres carrés ordinaires	oui	tendance
MK	Mann-Kendall	non	tendance
MKM	Mann-Kendall modifié	non	tendance
KS	Kendall-Seasonal	non	tendance
Spearman	Spearman	non	tendance
CS	Cox-Stuart	non	tendance

Remarques :

- i) *les tests sur des collections à chroniques gaussiennes longues montrent un biais de la méthode Mann-Kendall modifié. Pour compenser cela, il est choisi un risque de première espèce pour MKM de $\alpha' = 2,5 \%$. Les mêmes tests pratiqués sur la méthode Cox-Stuart semblent montrer également un taux de rejet inférieur à 5 %, mais le taux de rejet tend vers 5 % quand l'effectif et l'extension augmentent très fortement : il ne semble donc pas s'agir d'un biais systématique.*
- ii) *La comparaison des tests de Mann-Kendall et Kendall-Seasonal est présentée dans un second temps en raison du problème spécifique pour lequel le dernier test est développé.*

5.3 Résultats

Dans le cas de l'évaluation de la présence de tendance dans la série de données temporelles, le terme de « non rejet de H0 » est utilisé lorsque la méthode statistique estime que la chronique est stationnaire. Dans le cas contraire où la méthode détecte une tendance, le terme de « rejet de H0 » est employé.

Les graphiques résultats des tests de robustesse et de puissance montrent la proportion de rejet de H0 $\hat{\alpha}$ en fonction de différents paramètres. **Chaque résultat** (chaque point sur tous les graphiques résultats suivants) **correspond à l'application de la méthode testée sur 10 000 chroniques dont les caractéristiques fixées sont identiques.**

Les résultats doivent être interprétés de deux manières complémentaires :

- Les méthodes statistiques doivent reconnaître les chroniques effectivement stationnaires par construction (lois de génération sans tendance). La marge d'erreur α , dite risque de première espèce, mesure la probabilité de refuser une chronique construite selon des lois stationnaires. La valeur de α est fixée à 5 % ; cependant, des biais dans les hypothèses (lois non-gaussiennes, nombre faible de points, autocorrélations, faible nombre de chroniques testées) font que la proportion de rejet peut diverger de cette valeur. La méthode statistique est dite **robuste** si elle conserve un taux de rejet de H0 $\hat{\alpha}$ proche de 5 % lorsque les conditions de son application s'écartent des conditions idéales requises.
- Les méthodes statistiques doivent rejeter H0 lorsque les chroniques sont non-stationnaires par construction (lois de génération avec tendance linéaire). Le taux de rejet $\hat{\alpha}$ doit théoriquement croître vers 100 % à mesure que les chroniques s'éloignent de la stationnarité. On dit qu'un test est **puissant** s'il rejette à raison l'hypothèse H0 dans le cas de chroniques non-stationnaires.

5.3.1 Collections de chroniques gaussiennes (distribution normale des données)

Dans un premier temps, l'analyse de la robustesse des méthodes statistiques est réalisée pour des chroniques à loi de distribution gaussienne sans tendance **G0** et des calendriers réguliers **R**. Les paramètres qui ne sont pas indiqués sont pris à leurs valeurs de base.

Cette première analyse permet de tester la robustesse des méthodes statistiques sur des chroniques qui respectent les conditions initiales d'application des tests paramétriques.

- **Effet de la longueur des chroniques**

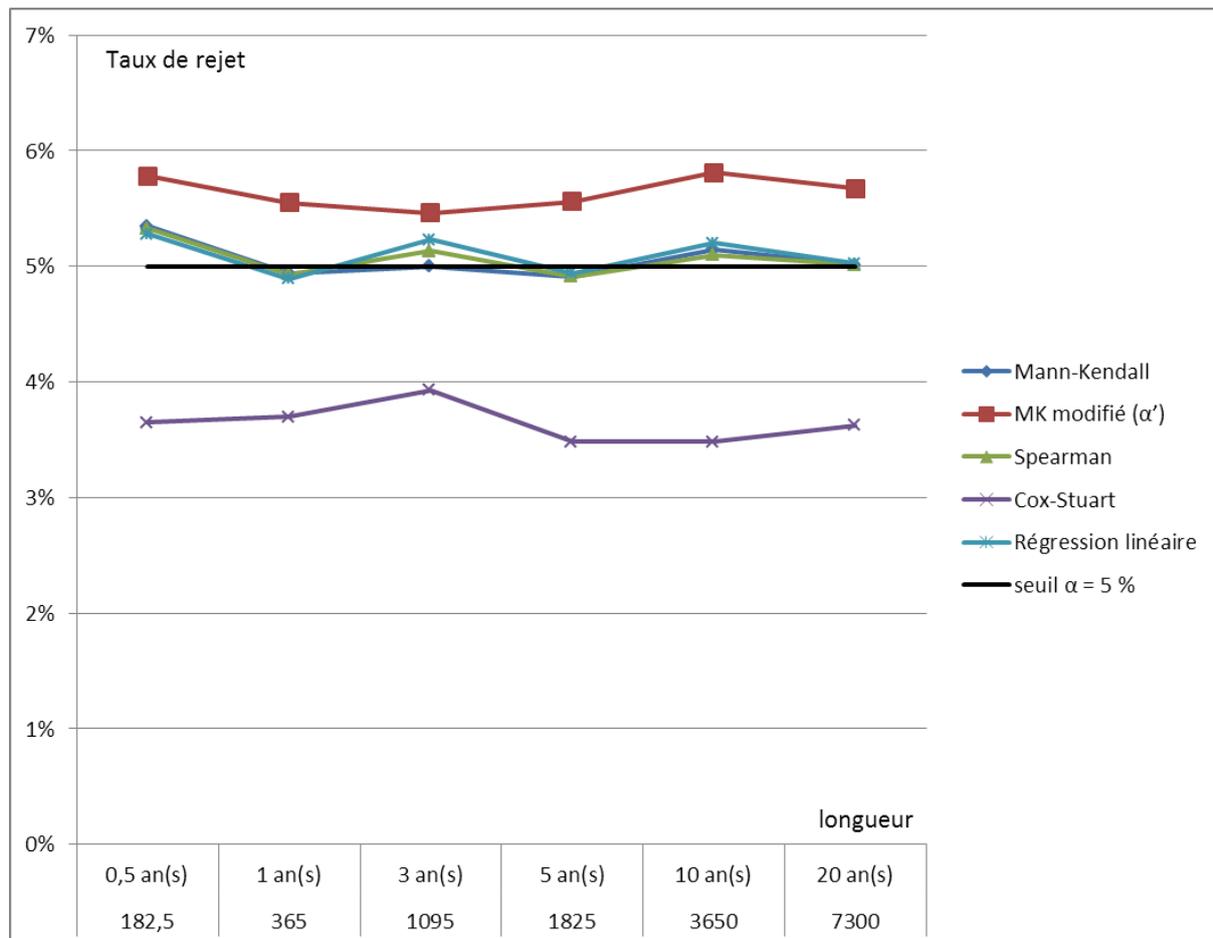


Illustration 26 : Taux de rejet de H0 selon les longueurs de chroniques à distribution de données gaussiennes sans tendance (taux de rejet attendu = 5%).

L'illustration 26 montre que les rejets de H0 des différentes méthodes sont indifférents à la variation de longueur des chroniques lorsque la distribution des données est normale. Cette constatation est sans surprise pour les tests non paramétriques, dont aucun ne prend en compte le calendrier (classement par rang). La constance de la régression linéaire montre que cette méthode est robuste sous des conditions idéales d'application.

Il est à noter que le Mann-Kendall modifié, même avec un $\alpha' = 2,5\%$, ainsi que le Cox-Stuart, sont constamment biaisés. La méthode Mann-Kendall modifié a tendance à rejeter trop souvent H0 alors qu'elle est vraie (erreur de type 1). La méthode Cox-Stuart a tendance à ne pas rejeter assez souvent H0 (erreur de type 2, faible puissance). Les collections sont

en effet créées de sorte que 5% des 10 000 chroniques ne respectent pas la stationnarité, résultat attendu en terme de rejet de H_0 . Le test de Cox Stuart est dit trop « conservateur ».

- **Effet des écarts-types**

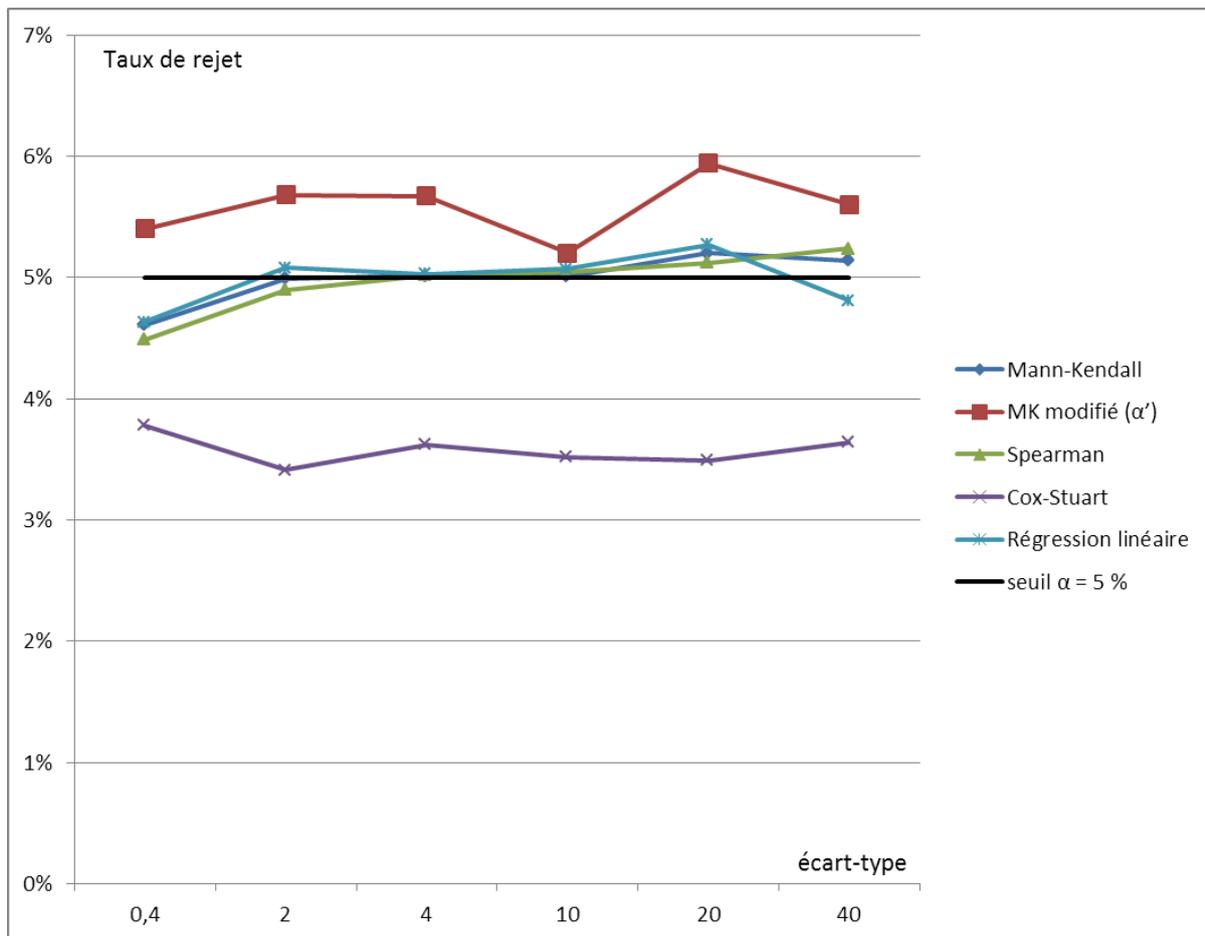


Illustration 27 : Taux de rejet de H_0 en fonction de l'augmentation de l'écart type des données de chroniques à distribution gaussienne sans tendance (taux de rejet attendu = 5%).

Comme précédemment, l'illustration 27 montre une indépendance du taux de rejet de H_0 en fonction de l'augmentation de l'écart-type des données des chroniques construites selon une distribution gaussienne sans tendance.

Les tests sont, au moins jusqu'à ce point, robustes et puissants, avec un bémol pour les méthodes Mann-Kendall modifié et Cox-Stuart.

- **Effet du nombre de données dans la chronique (effectif)**

L'illustration 28 montre que toutes les méthodes ne restent pas robustes à mesure que les effectifs des chroniques décroissent. Le test de Cox-Stuart est le premier à montrer des

signes de faiblesse, puisque son taux de rejet s'éloigne de 5% dès que l'on descend en-dessous de la vingtaine de points de mesure. Le test de Mann-Kendall reste robuste jusqu'à un minimum de 9 points alors que le Mann-Kendall modifié s'éloigne du taux de rejet de 5% en-dessous d'un effectif de 10 données. La régression linéaire ne semble pas affectée, même à des effectifs de 6 points.

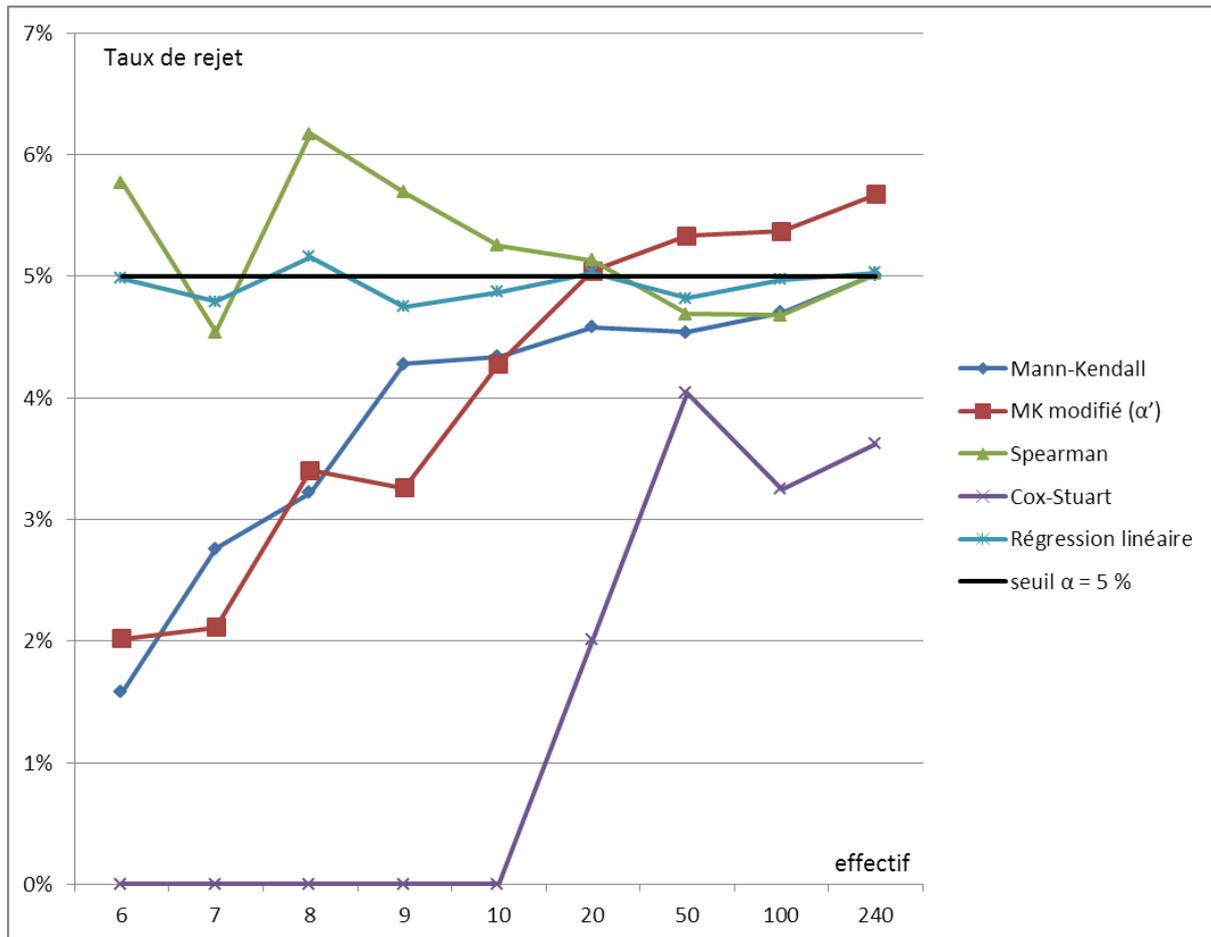


Illustration 28 : Taux de rejet de H_0 en fonction du nombre de données (effectif) des chroniques à distribution gaussienne sans tendance (taux de rejet attendu = 5%).

Il est à noter également que les comportements divergent pour les effectifs pairs et impairs, notamment dans le cas du Mann-Kendall modifié (sans prise en compte des ex aequo dans les séries toutefois).

- **Effet du type de calendrier (pas de temps et fréquence des prélèvements)**

L'illustration 29 montre que si le fait d'utiliser un calendrier régulier ou réparti uniformément ne semble pas avoir d'incidence sur la robustesse et la puissance des méthodes statistiques

de détermination de tendance, l'augmentation du bruit du calendrier fait monter légèrement le taux de rejet de H0 de certaines méthodes, comme le Mann-Kendall et le Mann-Kendall modifié.

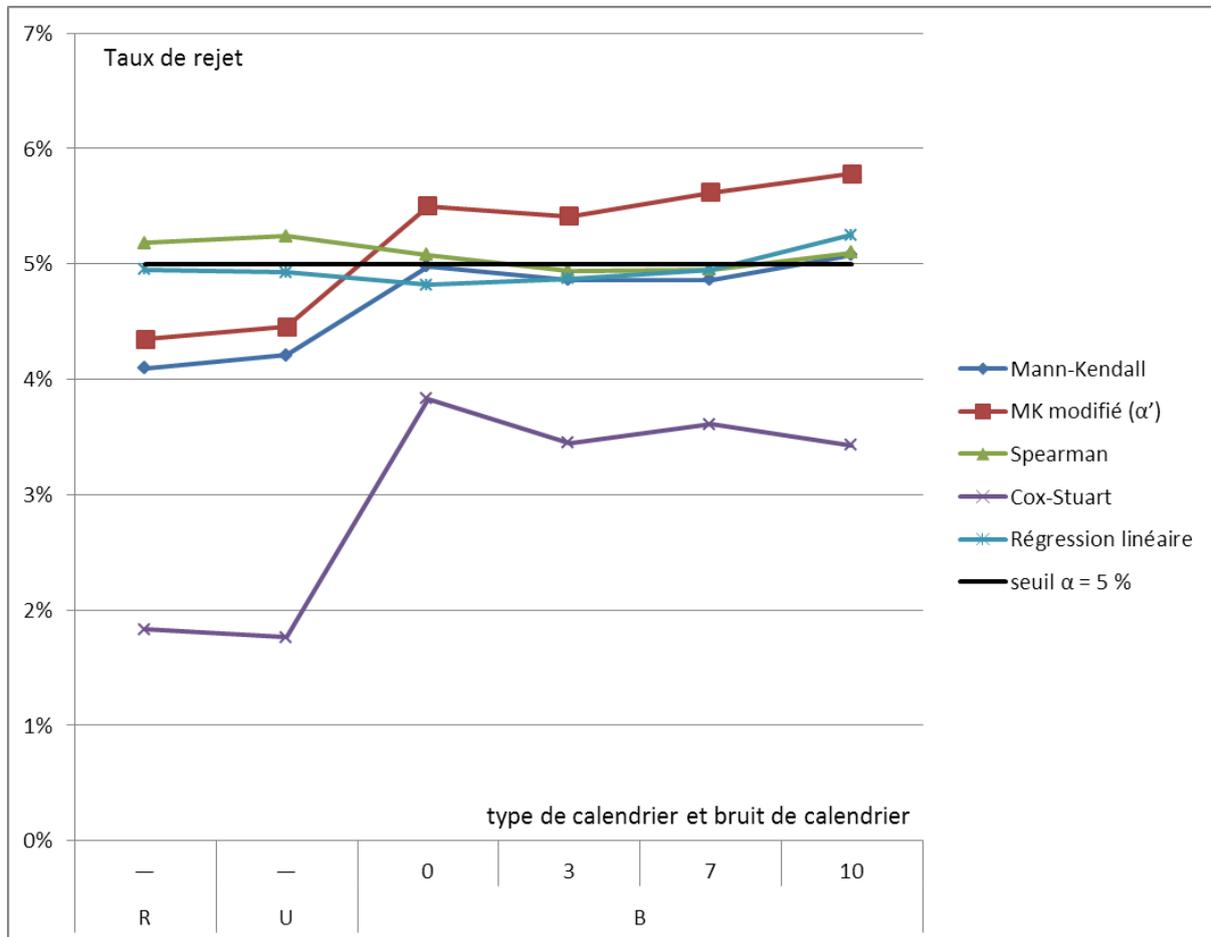


Illustration 29 : Taux de rejet de H0 en fonction du type et du bruit des calendriers pour des chroniques à distribution gaussienne sans tendance (taux de rejet attendu = 5%).

Dans une étude plus détaillée postérieure au travail présenté ici, il serait intéressant d'étudier l'influence sur ces variations de la fréquence de mesure (relation entre l'effectif et la longueur des calendriers).

- **Effet de la valeur de la pente de la tendance a**

Afin de tester l'effet de la pente de la tendance sur la puissance et la robustesse des méthodes statistiques pour la détection de tendance dans le cas d'une distribution de données gaussienne, le modèle de génération des chroniques utilisé est le modèle **Ga**. Il est

ainsi attendu un taux de rejet de H0 proche de 100% en raison de la non-stationnarité avérée des chroniques testées.

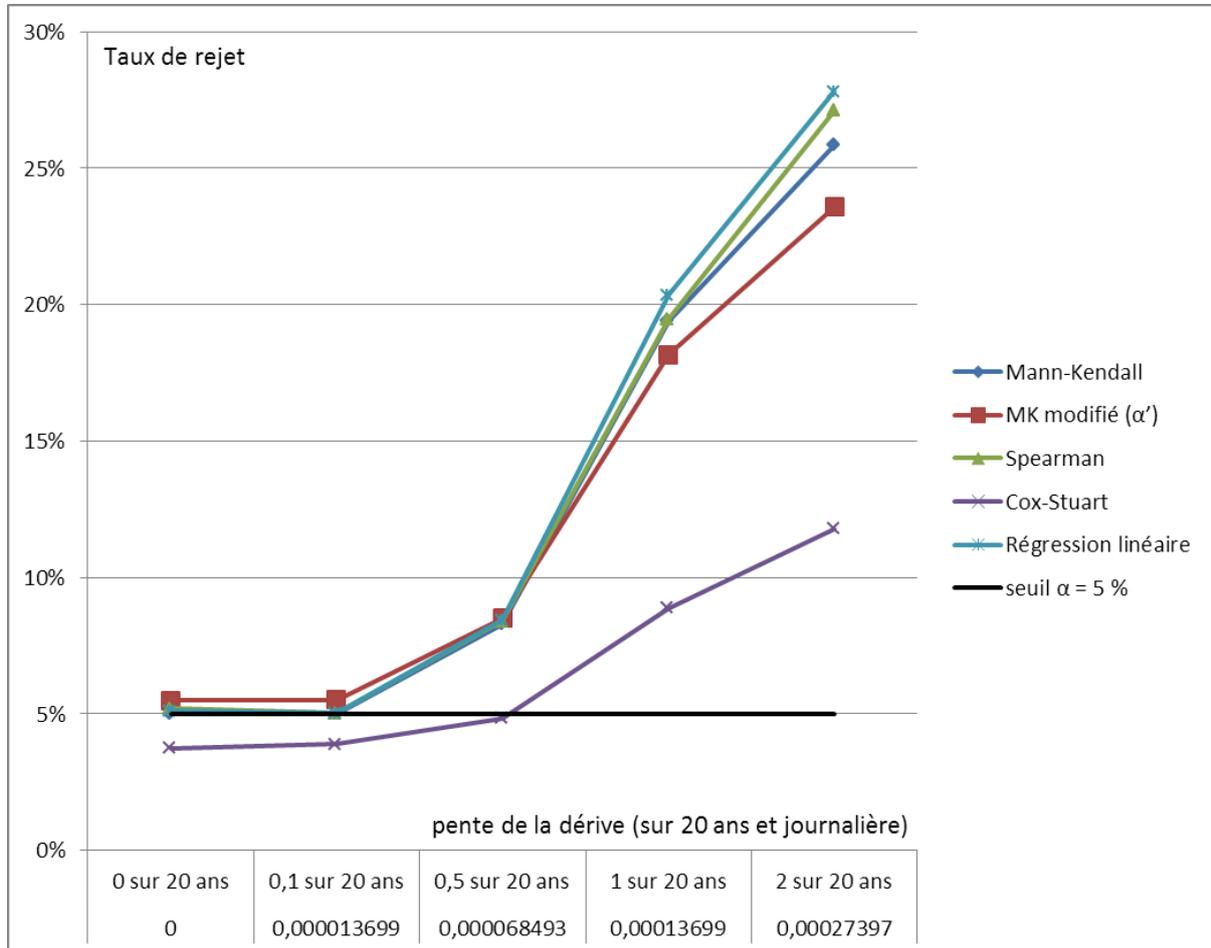


Illustration 30 : Taux de rejet de H0 en fonction de la valeur de la pente a de la tendance pour des chroniques à distribution gaussienne + tendance (taux de rejet attendu = 100%).

L'illustration 30 montre qu'il faut attendre une pente d'au moins 7.10^{-5} par jour (soit 0,5 sur 20 ans) pour noter un éloignement significatif du seuil de 5 %, c'est dire un début de détection de tendance. Une pente de $2,7.10^{-4}$ par jour (soit 2 sur 20 ans) engendre une augmentation du taux de rejet de H0 qui reste tout de même très en dessous des 100% attendu pour des pentes importantes.

Ainsi, malgré des pentes de tendances faibles (0,5 sur 20 ans au minimum), les méthodes statistiques testées les détectent. De plus, l'ensemble des méthodes statistiques se comportent de la même manière pour la détection de tendance sous l'effet de l'augmentation des pentes dans le cas d'une distribution gaussienne des données, excepté le test de Cox-

Stuart qui se distingue par une faible puissance à détecter des tendances dans le cas de pentes faibles.

- **Conclusion sur la puissance et la robustesse des méthodes statistiques dans le cas d'une distribution normale des données**

D'une manière globale, les méthodes statistiques testées montrent des comportements comparables pour l'estimation de la stationnarité des chroniques à distribution normale des données temporelles. Seul le test de Cox-Stuart s'est révélé systématiquement moins puissant face aux différentes caractéristiques critiques des séries temporelles.

La longueur de la chronique et l'écart-type des données ne semblent pas affecter le taux de rejet de H0 des différentes méthodes.

La régression linéaire est peu sensible au nombre de données qui constituent la chronique alors que les tests de Mann-Kendall, Mann-Kendall modifié et Spearman dérivent en-dessous d'un effectif de 10 valeurs (voir 9 dans le cas du test de Mann-Kendall). **Ce seuil de 10 valeurs constitue le nombre de données minimales dans une série temporelle pour l'application non biaisée des méthodes statistiques non paramétriques dans le cas d'une distribution normale des données.**

Le type et le bruit du calendrier n'a que peu d'effet sur la robustesse et la puissance des méthodes statistiques, tout juste note-t-on une légère augmentation du taux de rejet de H0 (à tort) pour les méthodes de Mann-Kendall et Mann-Kendall modifié sous l'effet de l'augmentation du bruit du calendrier.

Enfin, les méthodes statistiques testées commencent toutes (excepté la méthode Cox-Stuart) à détecter une tendance dans les chroniques à distribution normale des données à partir d'une pente de 7.10^{-5} par jour (soit 0,5 sur 20 ans). Les taux de rejet deviennent réellement significatifs à partir d'une valeur de pente de $2,7.10^{-4}$ par jour (soit 2 sur 20 ans) et ce, quel que soit la méthode testée.

Dans le cas d'une distribution normale des données (loi de distribution gaussienne), le test le plus efficace (le plus robuste et puissant) est la régression linéaire qui présente l'avantage d'accepter des chroniques à faible nombre de données. Les méthodes de Mann-Kendall et de Spearman sont tout aussi robustes et puissantes mais demandent respectivement un minimum de 9 et 10 valeurs dans la chronique pour être appliquées sans risques. Le test de Cox-Stuart et, dans une moindre mesure, le test de Mann-Kendall modifié, se sont révélés moins robustes et puissants pour l'estimation de la stationnarité des chroniques sous l'effet de la variation de l'ensemble des caractéristiques critiques des séries temporelles. Ces deux

dernières méthodes seront alors écartées lors du choix du test à appliquer pour l'évaluation de la stationnarité de chroniques à distribution normale au profit de la régression linéaire, du test de Mann-Kendall ou du test de Spearman.

5.3.2 Collections de chroniques à distribution log-normale des données

La robustesse et la puissance des méthodes statistiques d'évaluation de la stationnarité des chroniques sont testées pour des séries temporelles à distribution de données Log-normale. Cette distribution particulière des données est souvent rencontrée pour des chroniques de paramètres environnementaux. Il s'agit de paramètres dont la probabilité d'apparition des valeurs élevées est moins importante que celle des valeurs faibles, ce qui est généralement le cas des données qui qualifient la qualité des eaux. La distribution log-normale des données est un cas particulier de distribution non normale.

La loi de génération des données est la loi Log-normale affectée par une tendance linéaire de pente a (**Lna**). L'objectif étant de rendre compte de la robustesse et de la puissance des méthodes à détecter des tendances dans le cas de distributions non normale des données, seuls les résultats issus des modèles de chroniques Lna sont présentés. Comme précédemment, les paramètres non explicités sont pris à leurs valeurs de base.

Les conditions initiales idéales d'application des tests paramétriques ne sont alors plus respectées (distribution des données non normale). Il est ainsi intéressant de comparer l'efficacité des méthodes statistiques paramétriques et non-paramétriques en conditions initiales critiques pour les premières citées. Les tests seront dits efficaces si le rejet de H_0 s'écarte significativement du seuil de 5%, l'idéal étant 100% de rejet de H_0 sur les 10 000 couples calendrier/chronique d'une collection.

- **Effet de la longueur des chroniques**

L'évolution du taux de rejet de H_0 des méthodes statistiques de détection des tendances en fonction de la longueur des chroniques à distribution Log-normale des données affectées par une tendance linéaire est présentée dans l'illustration 31.

D'une manière globale, les résultats montrent une puissance bien supérieure des tests non paramétriques pour la détection de la non stationnarité des chroniques que les tests paramétriques (représentés par la régression linéaire) et ce, quelle que soit la

longueur des chroniques. Dans le détail, et pour n'importe quelle longueur de chronique, le test de Mann-Kendall est le plus efficace pour la détection de tendance en condition de distribution non normale des données. Le test de Spearman présente pratiquement les mêmes taux de rejet de H0 que le test de Mann-Kendall. Vient ensuite le test de Cox-Stuart qui présente un comportement similaire aux deux tests précédemment cités face à l'évolution de la longueur des chroniques avec toutefois une efficacité moindre pour les valeurs faibles ou fortes de longueurs de chroniques. Le test de Mann-Kendall modifié montre une efficacité moindre alors que **la régression linéaire n'est pas adaptée à l'exercice.**

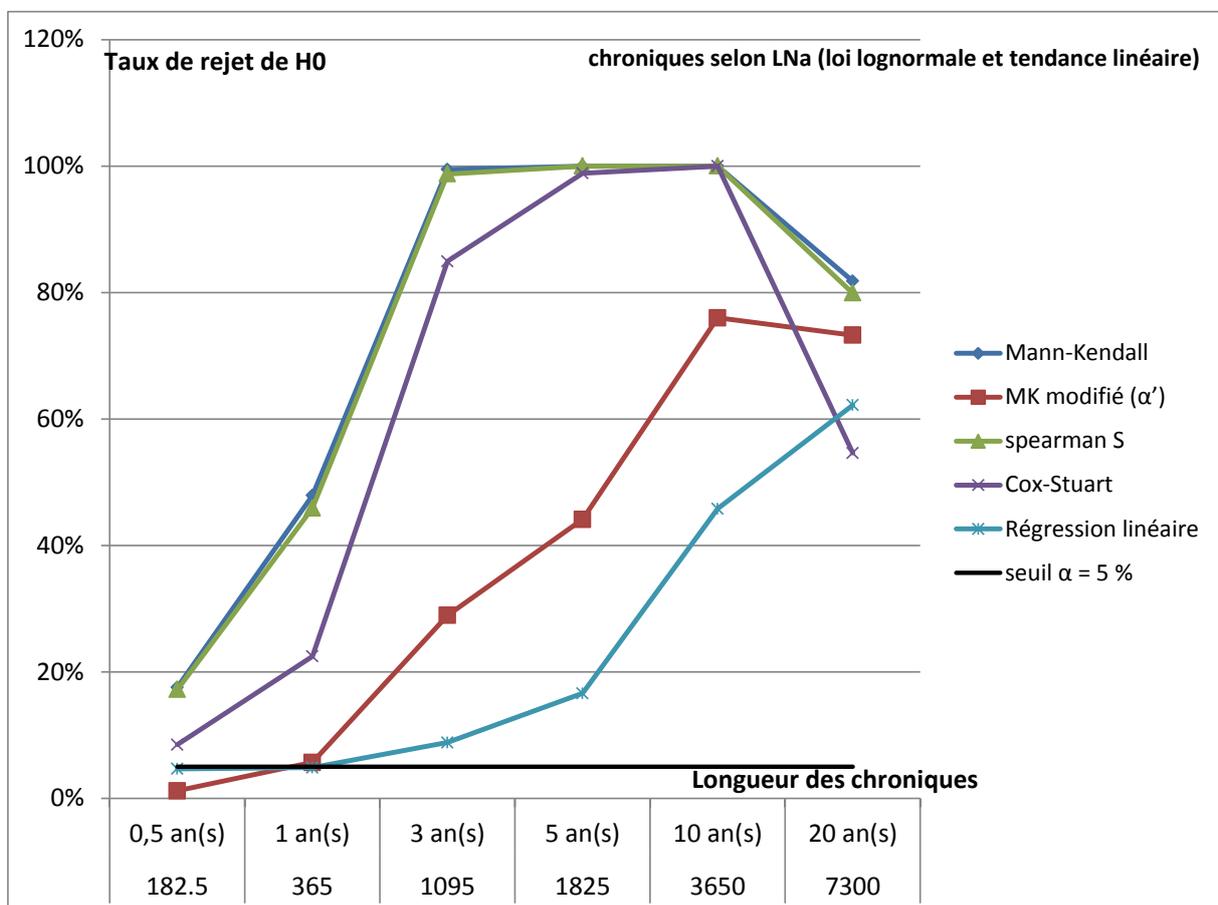


Illustration 31 : Taux de rejet de H0 selon les longueurs de chroniques à distribution de données Log-normale + tendance (taux de rejet attendu = 100%).

Contrairement aux données à distribution normale, le paramètre « longueur de chronique » affecte de manière très significative la détection de la tendance par les différentes méthodes. Pour les méthodes les plus puissantes (Mann-Kendall, Spearman et Cox-Stuart) il semble se dessiner une gamme de longueurs de chroniques optimales comprises entre 3 et 10 ans et

pour laquelle ces méthodes détectent 100% des tendances (longueurs de chroniques optimales entre 5 et 10 ans pour le test de Cox-Stuart).

Pour les deux autres tests, le taux de rejet de H0 augmente avec l'augmentation de la longueur des chroniques avec un palier atteint à 10 ans pour le test de Mann-Kendall modifié et un taux maximum de rejet de H0 à la longueur maximum de 20 ans pour la régression linéaire. Toutefois, les taux de rejet maximum restent relativement faibles pour ces deux tests, notamment pour la régression linéaire qui ne dépasse pas 65% de détection de tendance pour des longueurs de chroniques de 20 ans.

Remarque : Il est possible que la tendance crée de l'autocorrélation des données. De ce fait, si de l'autocorrélation est fictivement créée, alors le test de Mann-Kendall modifié peut effectivement rejeter l'existence de la tendance. En effet, comme le mentionne Renard (2006), le problème de distinguer ce qui relève de l'autocorrélation ou d'une tendance déterministe dans la détection de la non-stationnarité est un problème quasiment insoluble. Dans l'idéal, il faudrait faire le choix d'un test qui permet ou ne permet pas de prendre en compte l'autocorrélation, en fonction des connaissances du comportement du polluant au point notamment.

- **Effet des écarts-types**

L'illustration 32 montre l'évolution du taux de rejet de H0 des méthodes statistiques sous l'augmentation de l'écart-type des données à distribution Log-normale.

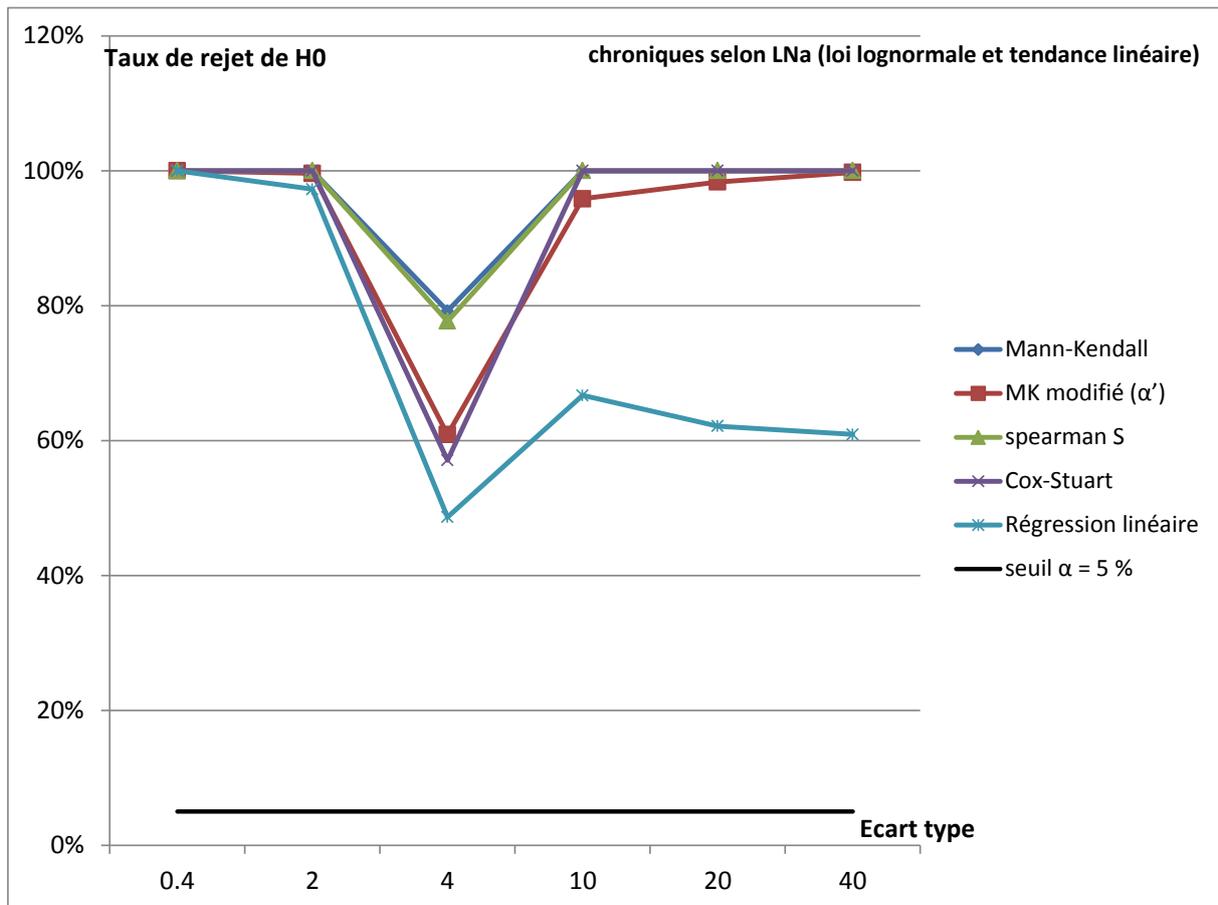


Illustration 32 : Taux de rejet de H0 en fonction de l'augmentation de l'écart type des données de chroniques à distribution Log-normale + tendance (taux de rejet attendu = 100%).

De façon non encore expliquée, une valeur d'écart-type de 4 à tendance à diminuer l'efficacité de l'ensemble des tests pour détecter une tendance dans les chroniques à distribution Log-normale des données. Toutefois, cet effet est moindre sur les méthodes de Mann-Kendall et de Spearman que sur les tests de Mann-Kendall modifié et de Cox-Stuart, eux-aussi moins affectés par cette diminution que la régression linéaire.

Il est à noter que pour des valeurs d'écart-type supérieures à 4, l'ensemble des tests retrouve un taux de détection des tendances proche ou égal à 100%, excepté la régression linéaire qui montre un risque de première espèce élevé et relativement constant.

- **Effet du nombre de données dans la chronique (effectif)**

La capacité des méthodes statistiques à détecter des tendances de chroniques à distribution Log-normale est dépendante du nombre de données qui constituent la série. L'illustration 33 rend bien compte de ce phénomène déjà évoqué dans le cas de chronique à distribution normale des données.

Comme précédemment, les tests les plus robustes face aux variations d'effectifs sont les tests de Mann-Kendall et de Spearman. Pour ces deux méthodes, la détection de tendance dépasse 60% lorsque l'effectif est supérieur ou égal à 9, le taux de rejet de H0 devenant très bon (proche de 100%) à partir de 20 données dans la série.

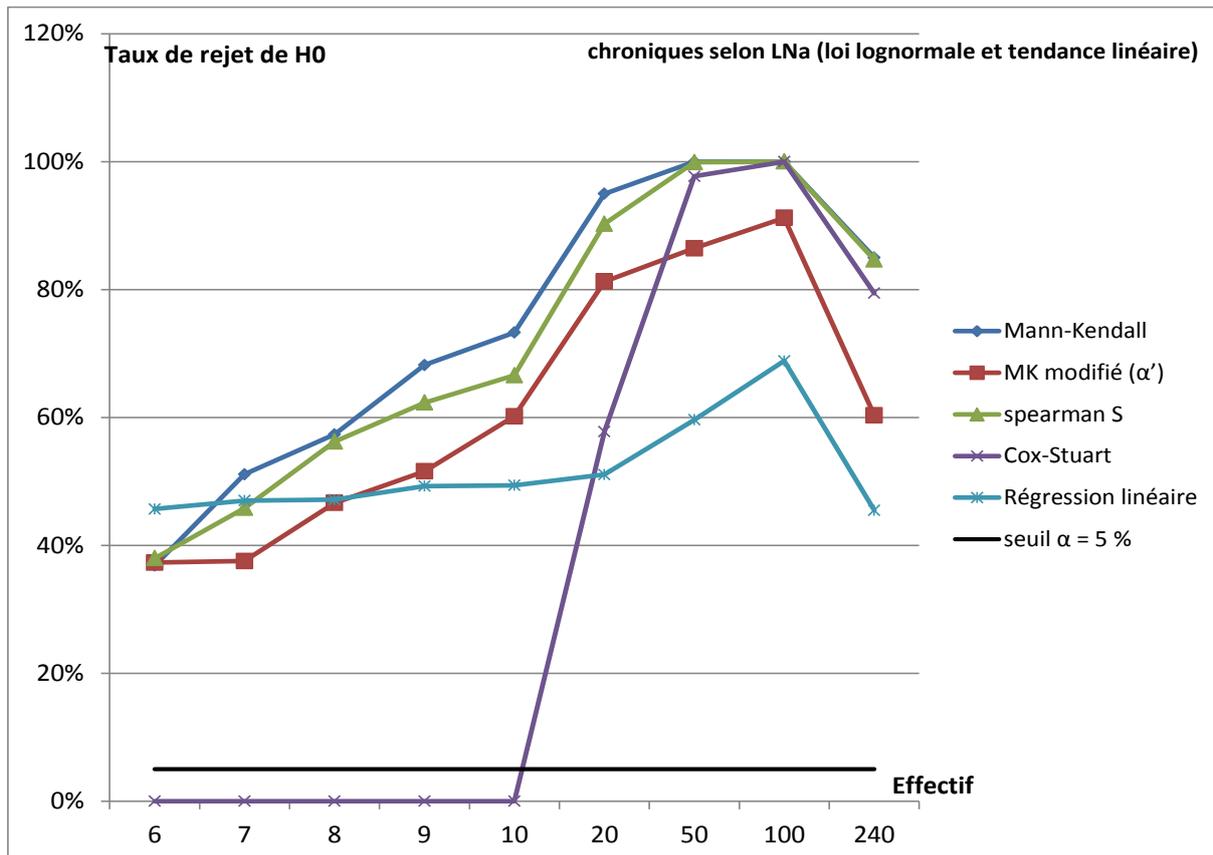


Illustration 33 : Taux de rejet de H0 en fonction du nombre de données (effectif) des chroniques à distribution Log-normale + tendance (taux de rejet attendu = 100%).

Le test de Mann-Kendall modifié montre le même comportement, avec cependant des taux de rejet de H0 toujours inférieurs à ceux des méthodes Mann-Kendall et Spearman.

La régression linéaire est le test le moins affecté par les variations d'effectif mais les taux de rejet de H0 sont trop faibles pour que la méthode soit dite robuste et puissante.

Le test de Cox-Stuart présente quant à lui un comportement trop conservateur pour un effectif inférieur ou égal à 10, c'est-à-dire qu'il ne rejette pas H0 alors que cette hypothèse est fautive par définition des chroniques. Le taux de rejet de H0 augmente de façon spectaculaire pour un effectif de 20 données (taux de rejet de H0 $\approx 55\%$) et le test atteint 100% de détections environ pour 50 données dans les séries temporelles.

- Effet de la valeur de la pente de la tendance

Les tests de Mann-Kendall et de Spearman se révèlent être encore une fois les méthodes les plus robustes et puissantes pour détecter des tendances à faibles pentes dans le cas de chroniques à distribution Log-normale des données (Illustration 34).

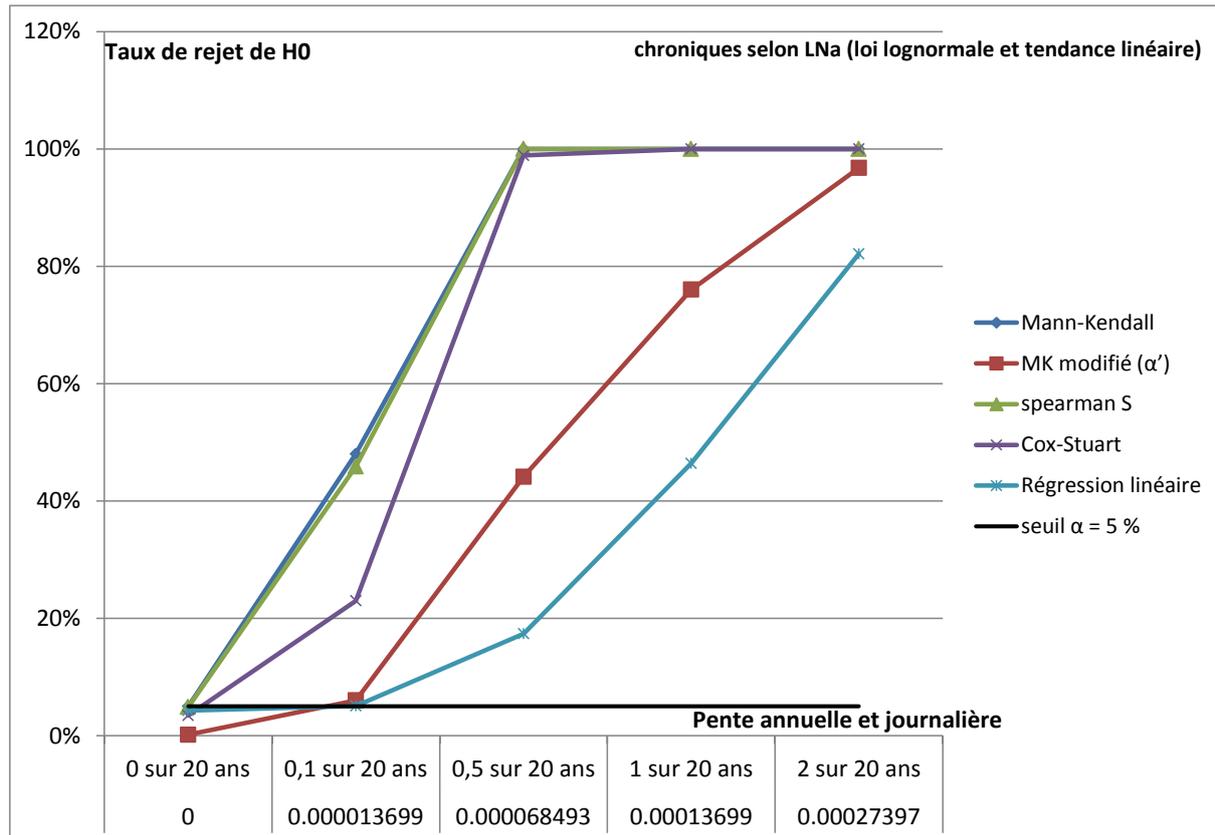


Illustration 34 : Taux de rejet de H0 en fonction de la valeur de la pente a de la tendance pour des chroniques à distribution Log-normale + tendance (taux de rejet attendu = 100%).

Vient ensuite le test de Cox-Stuart, qui, avec les deux méthodes précitées sont les seules à rejeter significativement H0 pour des pentes faibles de 0,1 sur 20 ans. Ces trois tests détectent 100% des tendances lorsque la pente est égale ou supérieure à 0,5 sur 20 ans.

Le test de Mann-Kendall modifié n'apparaît pas très efficace pour identifier des tendances de faibles pentes dans le cas de distribution de données non-normale. Le taux de rejet de H0 ne devient en effet proche de 100% que pour des valeurs de pente de 2 sur 20 ans, soit la valeur maximale testée.

Enfin, les résultats confirment le caractère inadapté de la régression linéaire pour l'évaluation des tendances lorsque l'on s'écarte des conditions initiales d'application de ce test, dont la faible robustesse est ici démontrée.

- **Conclusion sur la robustesse et la puissance des méthodes statistiques dans le cas d'une distribution Log-normale des données**

L'ensemble des résultats des tests sur des chroniques à distribution Log-normale des données montre une forte disparité de l'efficacité des méthodes statistiques suivant si elles sont paramétriques ou non paramétriques. La régression linéaire apparaît en effet peu robuste face à la variation des caractéristiques critiques des séries temporelles et sa puissance est toujours plus faible que les autres méthodes pour la détection des tendances. **La régression linéaire n'est donc pas adaptée à l'étude de la stationnarité de chroniques à distribution non gaussienne des données.**

Parmi les méthodes statistiques non paramétriques, **les tests de Mann-kendall et de Spearman sont les plus efficaces pour la détection des tendances de chroniques à distribution Log-normale des données.** En considérant une fréquence de prélèvement mensuelle, la gamme de longueurs de chroniques optimales pour l'application de ces deux tests est [3 ; 10 ans]. Ces tests sont particulièrement robustes à partir d'un effectif de 20 données mais peuvent être utilisés sans trop de biais pour des chroniques constituées de 10 données minimum. Ils détectent tous deux les plus faibles pentes testées (0,1 sur 20 ans) avec 100% de détection pour des pentes de 0,5 sur 20 ans.

La méthode de Cox-Stuart montre un comportement conservateur avec une efficacité plus faible que les tests de Mann-Kendall et Spearman bien que son utilisation dans le cas de chroniques à distribution Log-normale des données n'est pas à exclure. La gamme de longueurs de chroniques optimum est cependant plus restreinte [5 ; 10 ans] et le nombre de données minimum dans les séries est plus important (50 données minimum). Enfin, ce test détecte mal les pentes inférieures à 0,5 sur 20 ans. La méthode de Cox-Stuart peut donc être préconisée dans le cas de chroniques à distribution non gaussienne des données seulement si les séries sont longues et si elles disposent d'un nombre important de données (> 50).

5.3.3 Collections de chroniques autocorrélées et variogrammes

Dans ce paragraphe, la robustesse et la puissance des méthodes statistiques sont testées dans le cas de chroniques à variations cycliques, annuelles dans un premier temps puis à cycles longs (10 ans) dans un second temps. Ces variations cycliques entraînent une

autocorrélation des données, phénomène réputé très perturbateur de l'efficacité des tests de stationnarité.

Les collections sont constituées de chroniques simulées à l'aide des modèles **V0** pour reproduire des cycles annuels sans tendance, **Va** pour des cycles annuels avec tendance linéaire et **W0** pour les chroniques à variations cycliques de période 10 ans. Comme précédemment, sont attendus des taux de rejet de H0 proches de 5% dans les cas V0 et W0, et proches de 100% dans le cas Va. Devant l'impossibilité de faire varier directement l'autocorrélation des données, le parti est pris d'étudier l'effet d'une augmentation de ce phénomène en faisant varier la composante pépité des variogrammes. Celle-ci peut être assimilée au complémentaire de l'autocorrélation : lorsque la composante pépité augmente, l'autocorrélation des données diminue de sorte qu'une chronique construite selon un variogramme γ_1 avec 100% de pépité ne montre plus d'autocorrélation. Les modèles de calendriers réguliers **R** et troués **T** sont comparés et présentés côte-à-côte le cas échéant.

- **Effet de l'autocorrélation (complémentaire de la pépité)**

L'illustration 35 permet de comparer le taux de rejet de H0 des méthodes statistiques pour la détection de la stationnarité de chroniques cycliques annuelles sans tendance (modèle V0) avec un calendrier régulier ou troué, sous l'effet de l'augmentation de la pépité, c'est-à-dire une diminution de l'autocorrélation.

Les résultats montrent que, d'une manière globale, il suffit d'une très faible autocorrélation pour que les tests s'avèrent inefficaces. Tous les tests se comportent de manière globalement similaire excepté le test de Cox-Stuart.

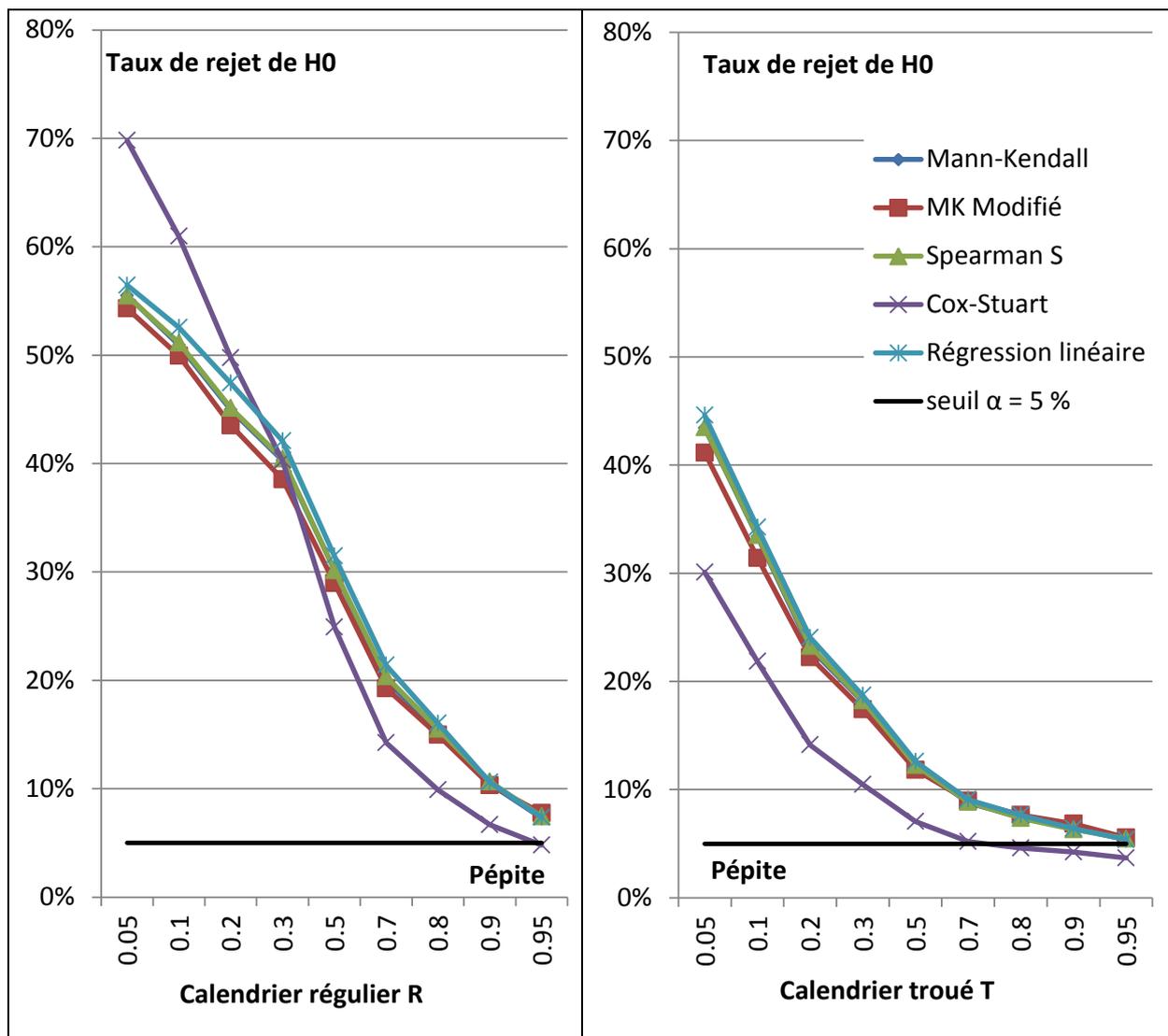


Illustration 35 : Comparaison du taux de rejet de H0 sous l'effet de l'augmentation de la composante pépité (=diminution de l'autocorrélation) pour un calendrier régulier R et troué T dans le cas de chroniques cycliques annuelles sans tendance V0 (taux de rejet attendu = 5%).

Dans le cas d'un calendrier régulier, l'ensemble des tests rejette à tort H0 dès que la pépité est inférieure à 90%. Le test de Cox Stuart montre un risque de première espèce inférieur que les autres tests jusqu'à une pépité de 50%, mais ce risque devient supérieur aux autres tests pour les fortes autocorrélations (pépité \leq 20%). Dans le détail, la méthode Mann-Kendall modifiée se révèle un peu meilleure que les autres tests, la régression linéaire étant la moins robuste dans le cas de chroniques cycliques annuelles et calendrier régulier.

Dans le cas d'un calendrier troué, les méthodes statistiques sont moins affectées par l'autocorrélation et la pseudo-cyclicité des chroniques que lorsque le calendrier est régulier. En effet, la sélection « au hasard » de N points parmi les N0 points initiaux du calendrier

régulier tend à rompre l'autocorrélation des données. Ainsi, tous les tests présentent un risque de première espèce encore acceptable (< 10%) pour une pépité de 70% alors que le taux de rejet atteint plus de 10% (sauf Cox Stuart) dès que la pépité est inférieure ou égale à 90% dans le cas d'un calendrier régulier. Pour un calendrier troué, le test de Cox Stuart apparaît le plus puissant, et ce, quelle que soit la valeur de la pépité. Vient ensuite la méthode de Mann-Kendall modifiée un peu plus puissante que les tests de Mann-Kendall, Spearman et régression linéaire qui présentent tous les trois un comportement identique.

- **Effet de la longueur des chroniques**

L'illustration 36 confirme que les méthodes statistiques allouées à l'étude de la stationnarité supportent mal les chroniques cycliques annuelles simulées par variogramme γ_1 (modèle V_0 , cycles annuels sans tendance). Ceci est particulièrement vrai lorsque le calendrier de prélèvements est régulier. Les taux de rejet à tort de H_0 sont alors supérieurs à 30% pour tous les tests, et ce quelle que soit la longueur des chroniques (excepté le test de Cox-Stuart pour une longueur de 5 ans). Ce phénomène est vraisemblablement engendré par la périodicité des chroniques générée par la composante cosinus du variogramme.

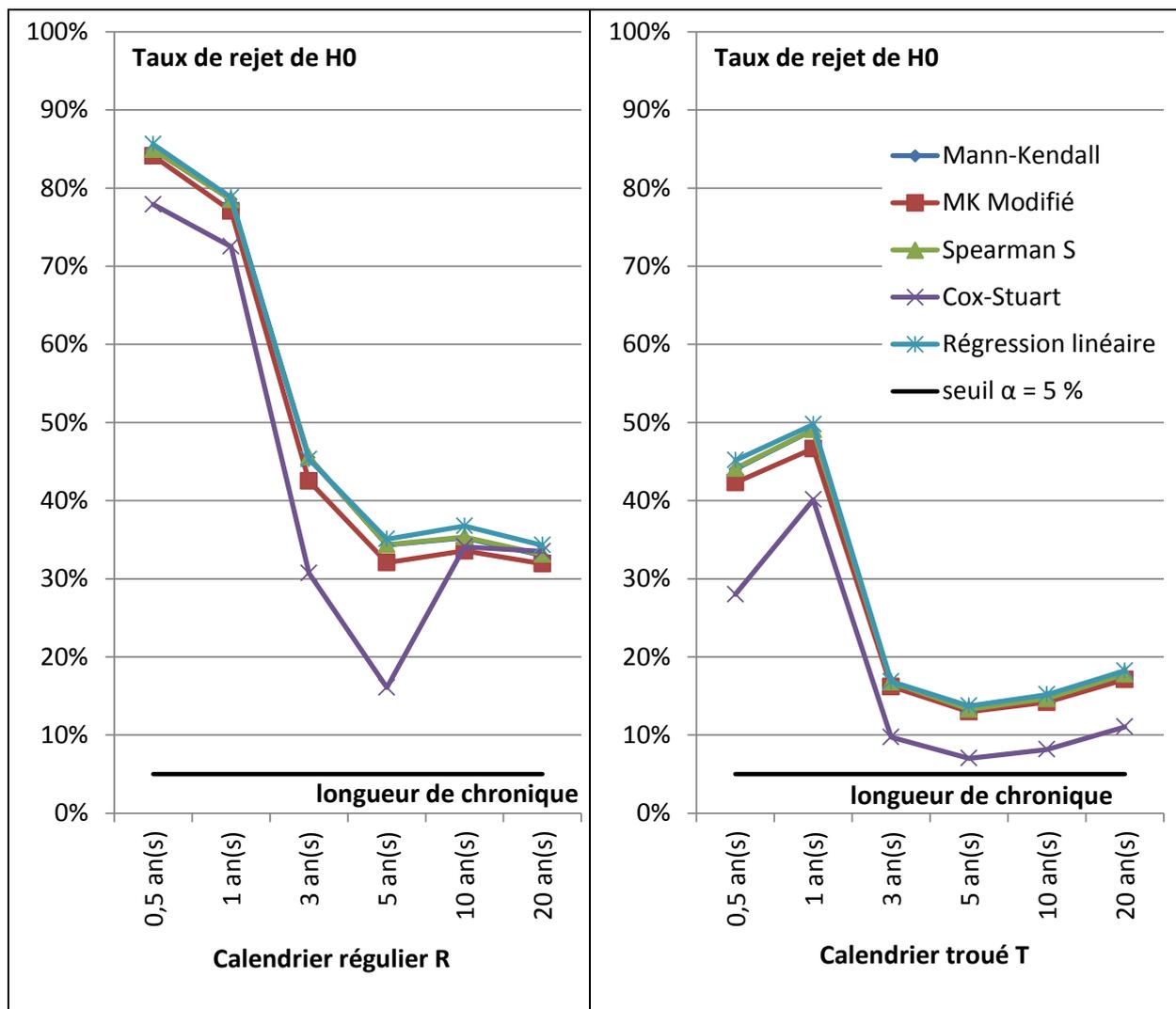


Illustration 36 : Comparaison du taux de rejet de H0 en fonction de la longueur des chroniques cycliques annuelles sans tendance (modèle V0) pour un calendrier régulier R et troué T (taux de rejet attendu = 5%).

Comme précédemment et pour les mêmes raisons, les méthodes sont moins affectées par l'autocorrélation lorsque le calendrier est troué.

Pour tous les tests, en calendrier régulier ou troué, les résultats montrent des taux de rejet de H0 élevés pour des longueurs de chroniques de 0,5 et 1 an. En effet, l'extraction d'une chronique simulée par variogramme γ_1 sur un demi cycle n'est pas stationnaire (partie croissante ou décroissante d'un cycle). Il en est de même lorsque l'extraction est réalisée sur un an et que le cycle ne débute pas à la première date de mesure du calendrier. Ces résultats sont donc cohérents avec le processus mis en œuvre pour tester les méthodes statistiques. Lorsque les chroniques ont une longueur supérieure à 1 an, les meilleurs

résultats sont obtenus pour des longueurs de 5 ans. Cette valeur pourrait correspondre à la longueur minimum préconisée pour l'application des tests de stationnarité sur des chroniques cycliques annuelles.

Dans le détail, la méthode de Cox Stuart montre les meilleurs résultats pour l'étude de la stationnarité de chroniques cycliques annuelles, que le calendrier soit régulier ou troué et quelle que soit la longueur des chroniques. Toutefois, le comportement de ce test est surprenant avec une évolution presque aléatoire du taux de rejet de H0 en fonction de l'augmentation de la longueur des chroniques. Ce phénomène reste encore inexpliqué. La comparaison des autres tests montre que la méthode de Mann-Kendall modifiée est un peu plus puissante que les tests de Mann-Kendall, Spearman et régression linéaire, surtout lorsque le calendrier est régulier. Les méthodes Mann-Kendall et Spearman possèdent une efficacité identique. La régression linéaire est la méthode la moins robuste face aux variations de longueur des chroniques à cycles annuels.

- **Effet de l'écart-type des données**

L'augmentation de l'écart type des données n'a pas réellement d'effet sur la capacité des méthodes statistiques à évaluer la stationnarité des chroniques cycliques annuelles comme le montre l'illustration 37. En calendrier régulier toutefois, et comme observé dans le cas de distributions des données gaussiennes, une valeur d'écart-type de 4 change le résultat de l'ensemble des tests sans qu'il n'y ait aujourd'hui d'explications définitives à ce phénomène. Les résultats confirment la diminution des biais dus à l'autocorrélation lorsque le calendrier est irrégulier. Dans ce dernier cas, le test de Cox Stuart apparaît comme le plus robuste alors que c'est la méthode de Mann-Kendall modifiée qui montre le risque de première espèce le plus faible dans le cas d'un calendrier régulier.

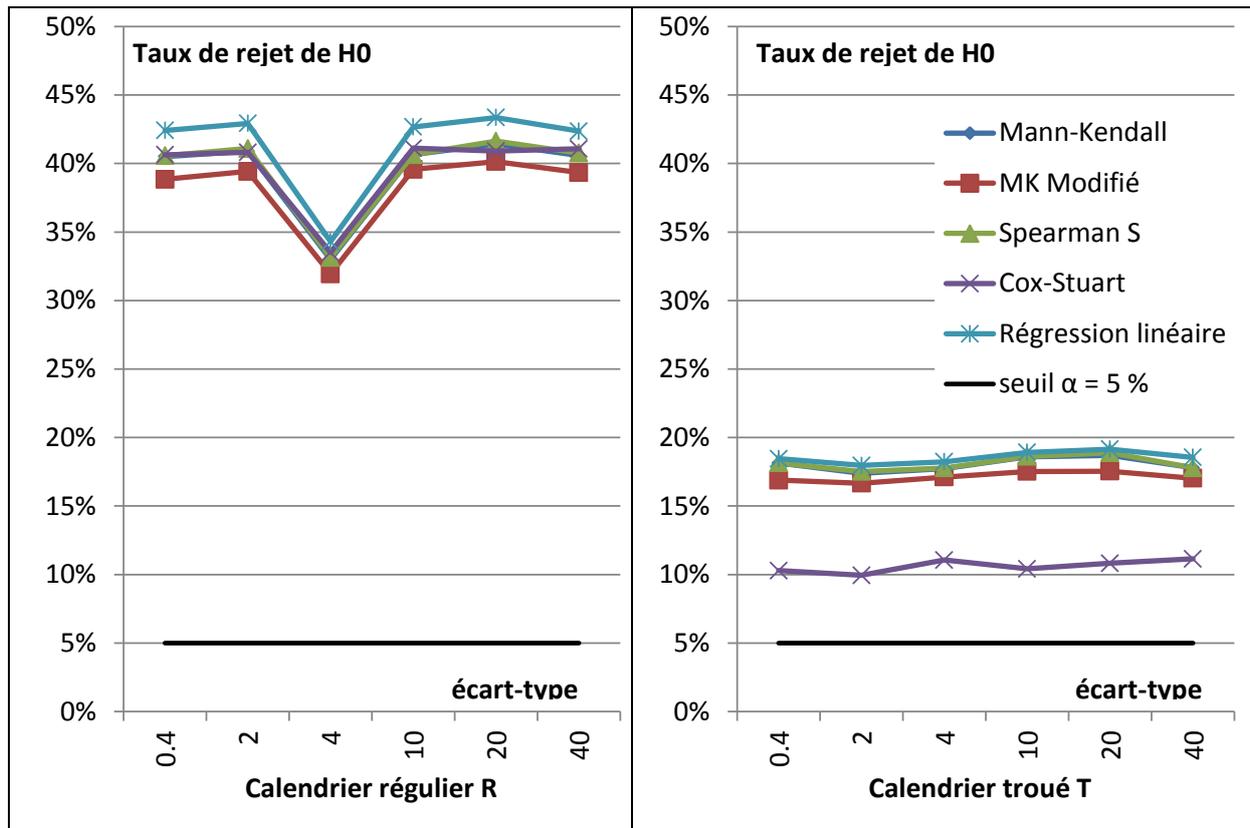


Illustration 37 : Comparaison du taux de rejet de H0 sous l'effet de l'augmentation de l'écart type des données de chroniques cycliques annuelles sans tendance (modèle V0) pour un calendrier régulier R et troué T (taux de rejet attendu = 5%).

- **Effet du nombre de données dans la chronique (effectif)**

Le comportement des méthodes statistiques appliquées sur des chroniques cycliques annuelles est extrêmement dépendant des effectifs (Illustration 38). Toutefois, et de manière paradoxale au premier abord, les méthodes statistiques testées sont plus robustes pour évaluer la stationnarité des chroniques (taux de rejet de H0 proche de 5%) lorsque les densités de points sont faibles. Il s'avère en effet que le caractère pseudo-périodique des données s'exprime d'autant plus faiblement que le nombre de données dans les séries est faible. L'autocorrélation est donc moindre ce qui augmente l'efficacité des méthodes statistiques.

Notons également le comportement étonnant des tests pour un effectif de 50 points (soit 2,5 points par an), uniquement pour les calendriers réguliers. Il y a en effet dans ce cas deux contre-cycles de données, déphasées entre elles de ½ an. Ce phénomène tend alors à

réduire l'autocorrélation des données et, en corollaire, à réduire le risque de première espèce des méthodes statistiques.

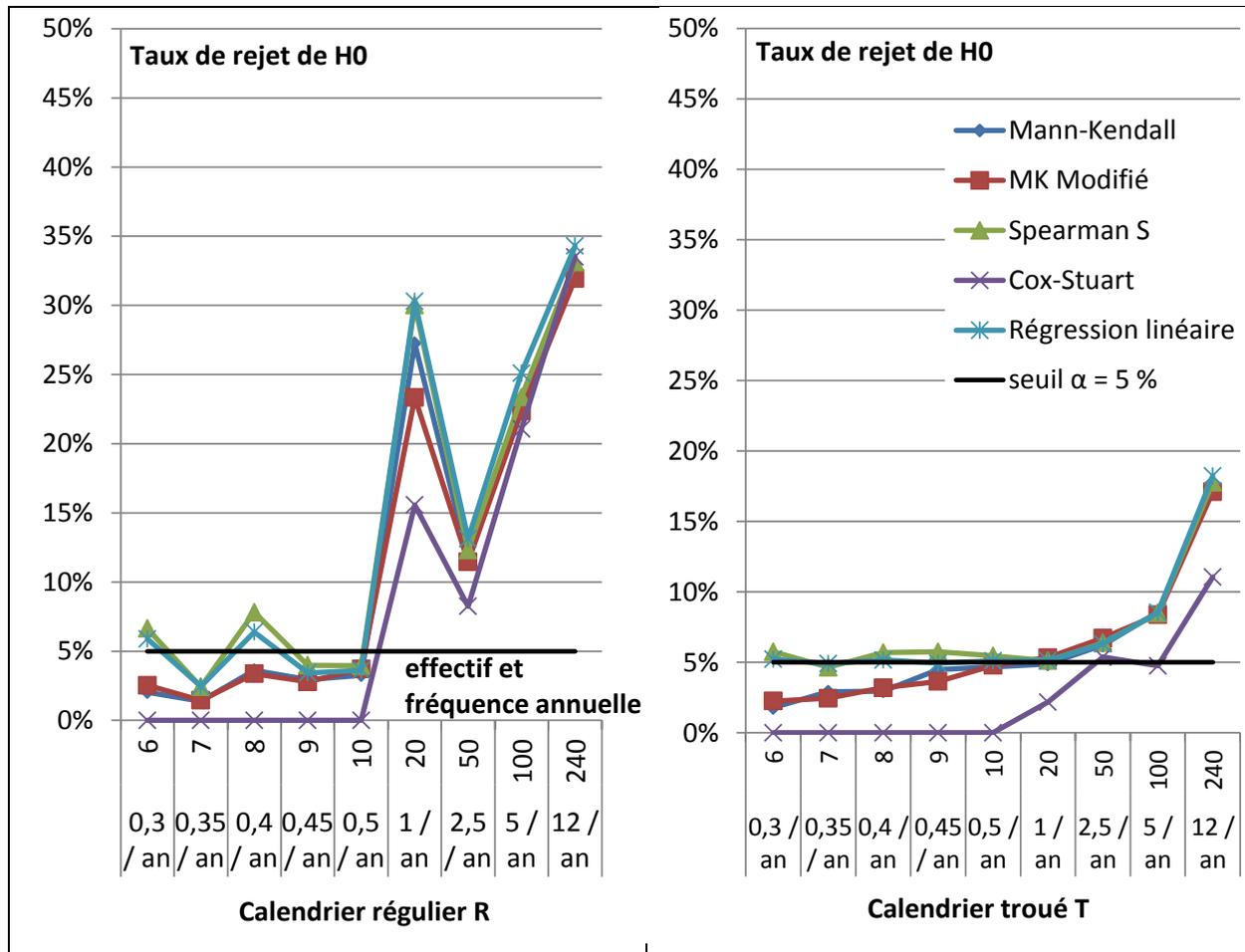


Illustration 38 : Comparaison du taux de rejet de H0 sous l'effet de l'augmentation du nombre de données de chroniques cycliques annuelles sans tendance (modèle V0) pour un calendrier régulier R et troué T (taux de rejet attendu = 5%).

Dans le détail et comme précédemment, la méthode Mann-Kendall modifiée sera préférée aux autres tests pour l'évaluation de la stationnarité des chroniques cycliques annuelles. Le test de Cox Stuart est en effet trop conservateur pour les faibles effectifs et les tests de Mann-Kendall, de Spearman et la régression linéaire sont moins robustes, pour un calendrier régulier particulièrement.

Remarque : Comme pour l'étude de l'effet des longueurs de chroniques cycliques annuelles sur l'efficacité des tests de stationnarité, une étude plus poussée serait nécessaire pour bien comprendre et cadrer les réponses des méthodes statistiques lors de leurs applications sur des chroniques simulées. Il faudrait en effet prendre garde, lors de l'utilisation des tests, à ne

pas faire émerger de fausses réponses à cause de phénomènes de pseudo-cyclicité générés artificiellement.

- **Effet de la valeur de la pente de la tendance**

Dans cette partie, les collections sont constituées de chroniques simulées selon le modèle **Va** (variogramme γ_1 + tendance linéaire) et des calendriers réguliers **R** et troués **T**. Un taux de rejet de H_0 proche de 100% est attendu.

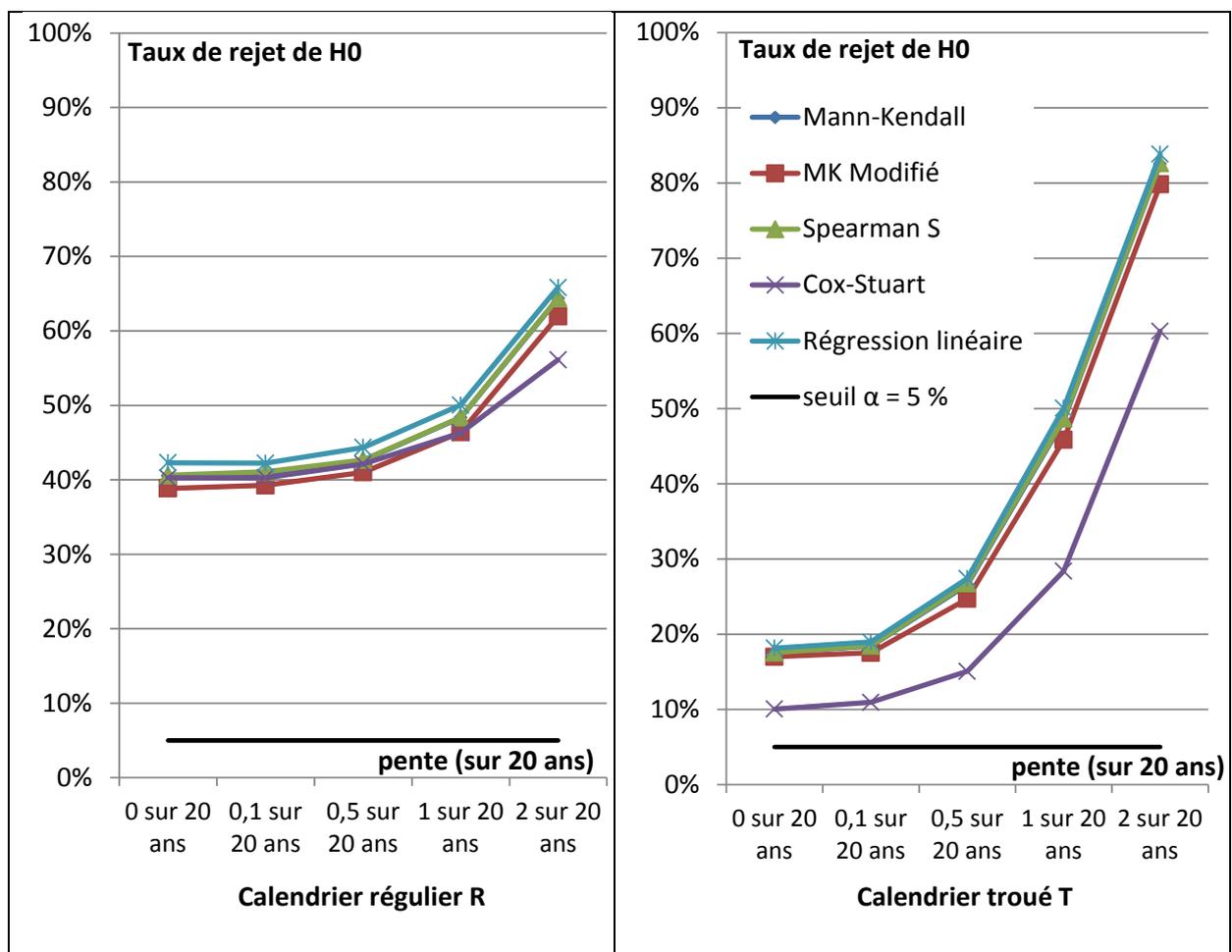


Illustration 39 : Comparaison du taux de rejet de H_0 sous l'effet de l'augmentation de la pente des tendances de chroniques cycliques annuelles avec tendance (modèle Va) pour un calendrier régulier R et troué T (taux de rejet attendu = 100%).

Même s'il apparaît peu pertinent d'étudier le risque de seconde espèce alors que les méthodes statistiques sont globalement peu robustes pour l'évaluation de la stationnarité de chroniques cycliques annuelles, il est toutefois à noter que, sur l'illustration 39, pour un

calendrier régulier, les taux de rejet de H0 augmentent sensiblement à partir d'une valeur de pente de 0,5 sur 20 ans. Ce résultat est cohérent avec les observations déjà réalisées dans le cas de chroniques à distribution des données gaussienne et Log-normale. Un calendrier irrégulier donne de meilleurs résultats, et ce, pour l'ensemble des méthodes statistiques autant pour détecter la stationnarité lorsque la pente a est nulle que pour détecter la non-stationnarité lorsque la pente de la tendance est supérieure à 1 sur 20 ans.

Toutes les méthodes présentent des taux de rejet de H0 pratiquement identiques en fonction de l'augmentation de la pente en calendrier régulier. Lorsque le calendrier est troué, le test de Cox Stuart se démarque négativement en montrant une puissance moindre que les autres tests. Ces derniers ont des valeurs de taux de rejet de H0 identiques lorsque la pente de la tendance augmente de sorte qu'il est impossible, à la vue de ces seuls résultats, d'en préconiser un plus qu'un autre.

- **Effet de la longueur des cycles**

Les résultats des tests de stationnarité sur des collections de chroniques à variogrammes γ_1 et γ_2 (modèles de chroniques V0 et W0 respectivement) sont comparés. Par défaut, le calendrier régulier est utilisé pour la simulation des collections. La robustesse des méthodes statistiques est évaluée dans les deux cas de figures, en fonction de l'évolution de l'autocorrélation dans un premier temps, puis en fonction de la longueur des chroniques et enfin suivant la fréquence de mesure des données. Les méthodes seront dites robustes si leur taux de rejet de H0 se rapproche de 5%.

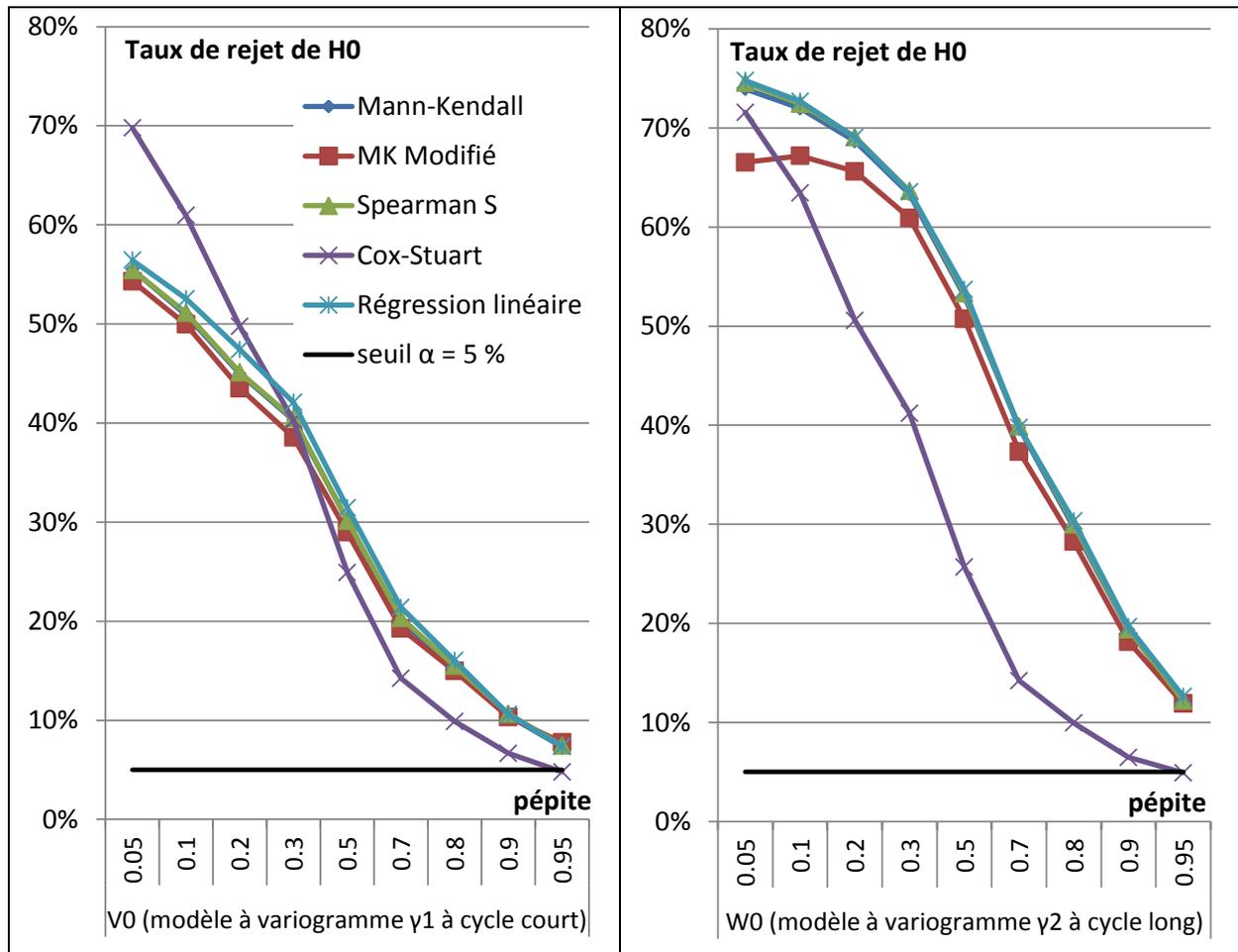


Illustration 40 : Comparaison de l'évolution du taux de rejet de H0 pour des chroniques à cycles annuels sans tendance (modèle V0) et à cycles de 10 ans sans tendance (modèle W0) en fonction de l'augmentation de la composante pépité des variogrammes (taux de rejet attendu = 5%).

L'illustration 40 montre que la qualité des réponses des tests est plus soumise à une diminution de la pépité pour le modèle W0 (cycles longs) que pour le modèle V0 (cycles courts). En effet, la composante Cosinus de W0 a une périodicité de 10 ans, alors que les chroniques ont une durée de 20 ans. Il y a donc moins d'effet de cyclicité que dans le modèle V0 à cycles annuels.

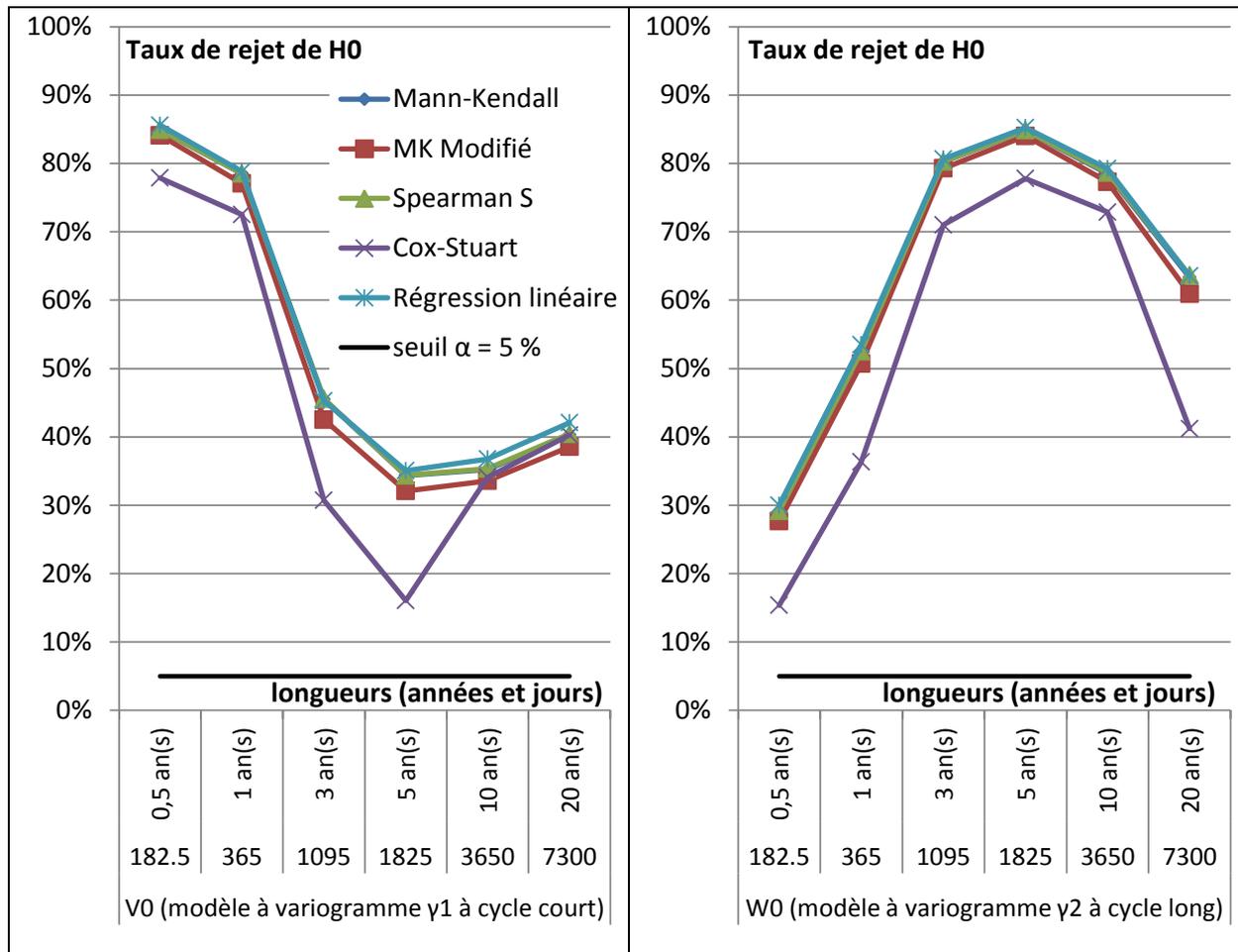


Illustration 41 : Comparaison de l'évolution du taux de rejet de H0 pour des chroniques à cycles annuels sans tendance (modèle V0) et à cycles de 10 ans sans tendance (modèle W0) en fonction de la longueur des chroniques (taux de rejet attendu = 5%).

Les comportements des méthodes statistiques appliquées sur des collections de types V0 et W0 sont totalement différents entre eux selon la longueur des calendriers. L'illustration 41 montre que si les taux de rejet de H0 diminuent en fonction de l'augmentation de la longueur des chroniques lorsque ces dernières sont cycliques annuelles, il n'en est pas de même pour les chroniques à cycles longs. Les méthodes statistiques ne sont alors robustes que pour des longueurs soit faibles (0,5 an) soit élevées (20 ans). Entre ces deux bornes, l'ensemble des méthodes a tendance à rejeter H0 alors que cette hypothèse est vraie. La tendance détectée est en réalité générée artificiellement par la troncature d'un cycle complet, phénomène déjà observé pour des chroniques à cycles annuels et de longueurs inférieures à 1 an. Ces observations montrent toute l'importance d'évaluer la longueur des cycles qui peuvent affecter les chroniques avant de mettre en œuvre un ou des test(s) de stationnarité. Il sera alors préconisé des longueurs de chroniques au minimum deux fois supérieures aux longueurs des cycles.

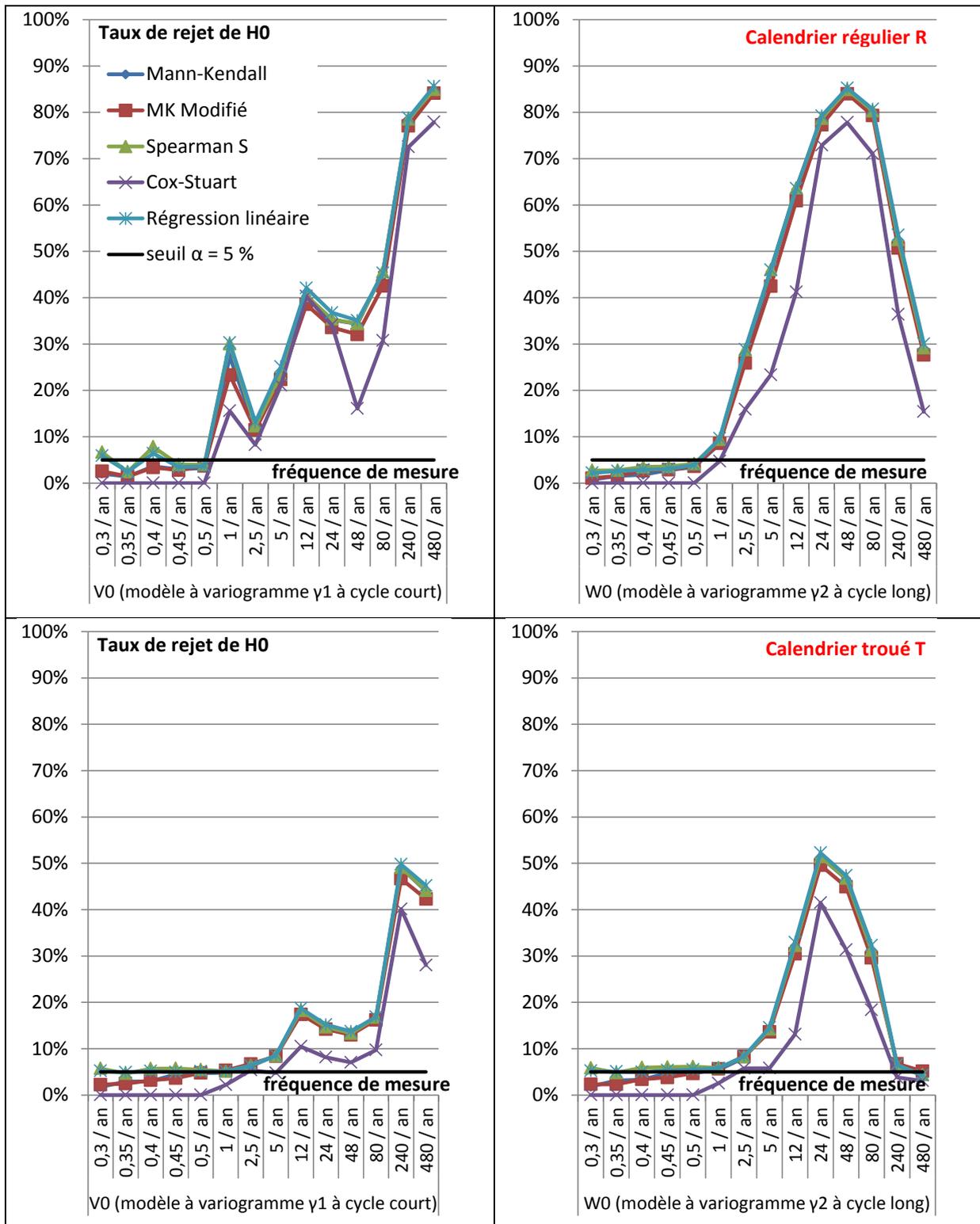


Illustration 42 : Comparaison de l'évolution du taux de rejet de H0 pour des chroniques à cycles annuels sans tendance (modèle V0) et à cycles de 10 ans sans tendance (modèle W0) en fonction de la fréquence de mesure des données en calendrier régulier (en haut) et en calendrier troué (en bas, taux de rejet attendu = 5%) .

L'illustration 42 rend compte de l'intérêt d'acquérir les données de qualité des eaux de façon irrégulière lorsque le comportement du signal chimique au point de mesure est de type cyclique. Cette méthode de prélèvement permet de « casser » l'autocorrélation des données engendrée par la périodicité du signal et donc augmente sensiblement l'efficacité des méthodes statistiques à évaluer la stationnarité des chroniques. Il est en effet montré que, pour des chroniques longues (20 ans) à cycles courts ou longs, les méthodes statistiques rejettent à tort H_0 pour une fréquence de mesure de 2,5/an ou plus. Lorsque le calendrier est troué, les méthodes statistiques montrent encore des bons résultats pour des fréquences de mesures de 5/an, voire beaucoup plus dans le cas de chroniques cycliques annuelles.

Au-delà de la seule détection de la stationnarité, il sera toujours préféré de fortes fréquences de mesures qui permettent d'augmenter l'information et donc la connaissance de l'évolution de la qualité de l'eau au point de mesure. Le choix d'un test qui prenne en compte ou non l'autocorrélation doit se faire en fonction des connaissances du comportement du paramètre étudié. Or, cette connaissance est améliorée par la multiplication des prélèvements dans le temps et part un recul important sur l'historique d'évolution des concentrations dans la nappe.

- **Conclusion sur la robustesse et la puissance des méthodes statistiques dans le cas de chroniques pseudo-périodiques**

Les essais réalisés sur des chroniques à cycles annuels et pluriannuels montrent que l'autocorrélation engendrée par la simulation *via* un variogramme γ_1 et un variogramme γ_2 respectivement fausse globalement l'ensemble des résultats des méthodes statistiques. L'autocorrélation se manifeste sous deux aspects :

- Une cyclicité, ou pseudo-périodicité, due à la composante Cosinus des variogrammes.
- Une autocorrélation de proximité, due aux deux composantes Cosinus (pour des longueurs d'onde inférieures à 1 an, ou 10 ans pour le modèle W_0) et Exponentielle (pour des longueurs d'onde de 10 ans).

La plus grande prudence doit donc être de mise pour l'utilisation de ces tests sur des chroniques autocorrélées. Tout au plus peut-on remarquer que la méthode Mann-Kendall modifiée donne des résultats légèrement meilleurs sur ce type de chroniques que les autres tests. Les résultats montrent aussi que l'efficacité des méthodes statistiques est nettement améliorée lorsque le calendrier de mesure est irrégulier. Ce mode de prélèvement permet en

effet de rompre artificiellement l'autocorrélation des données qui tend à fausser les résultats de détermination de la stationnarité des chroniques.

L'autocorrélation peut en effet être traitée de différentes manières :

- La chronique peut être nettoyée de son autocorrélation (utilisation de filtres) ; cependant le faible nombre de points généralement observés dans les chroniques rend périlleuse cette approche. Il existe cependant des tests qui prennent en compte le caractère cyclique de l'évolution des données temporelles. Il s'agit notamment du test Kendall-Saisonnier dont la robustesse face à l'autocorrélation est comparée à celle du test de Mann-Kendall classique dans le paragraphe 5.4.
- Un modèle de chronique peut être supposé ; il peut alors être proposé un test de conformité de la chronique à ce test. L'inconvénient est que le rejet de l'hypothèse ne permet pas de conclure à la non-stationnarité, mais seulement à l'inadéquation du modèle proposé.
- Un modèle peut également être estimé à partir de données agrégées régionalement, par exemple à l'échelle d'un aquifère. Cela suppose l'homogénéité du comportement observé à l'échelle de l'aquifère, ainsi qu'une méthode d'agrégation pertinente, ce qui n'est pas acquis.

Au-delà de ces méthodes relativement poussées de traitement de l'autocorrélation, la première préconisation simple qu'il est possible de faire dans le cas de chroniques cycliques est d'utiliser la méthode de Mann-Kendall modifiée en privilégiant les plus grandes longueurs possibles (au minimum deux fois supérieures aux longueurs des cycles) et un échantillonnage irrégulier.

5.4 Comparaison des tests Mann-Kendall et Kendall Saisonnier pour la détection de tendances de séries à cycles marqués

Les résultats des tests de robustesse et de puissance sur des chroniques cycliques ont révélé l'inefficacité globale des méthodes statistiques pour l'évaluation de la stationnarité de ces chroniques construites avec des calendriers de prélèvement réguliers. Ce cas de figure n'est pourtant pas rare dans le domaine de la qualité des eaux souterraines. Il est donc nécessaire de trouver une méthode statistique qui puisse prendre en compte les cycles des séries temporelles, c'est-à-dire traiter le problème de l'autocorrélation des données. Cette réflexion a déjà été menée par les statisticiens qui ont développé, à partir du test de Mann-Kendall, une méthode simple d'évaluation de la stationnarité qui permet de désaisonnaliser

les chroniques cycliques : la méthode Kendall Saisonnier. Son principe est détaillé dans le paragraphe 2.2.2.

Avant de recommander ou non l'utilisation de ce test dans les conditions particulières de chroniques cycliques à calendrier régulier, il est intéressant de comparer son efficacité par rapport au test Mann-Kendall classique sur des chroniques théoriques particulièrement critiques pour l'évaluation de leur stationnarité.

En premier lieu, les puissances des tests Kendall Saisonnier et Mann-Kendall sont comparées pour la détection de la tendance d'une chronique théorique cyclique annuelle d'une longueur de 20 ans prélevée tous les jours. La chronique à tester est construite selon un variogramme de type : $L \times (1 - (\cos(2\pi/p) \times t)) + \exp(0,1) + a.t + b$

Avec L = longueur totale de la chronique (20 ans soit 7300 jours dans le cas étudié)

p = période du cycle (ici 365 jours)

a = pente de la tendance

b = bruit du signal, pris au hasard pour chaque valeur de t entre $[-0,3\sigma ; +0,3\sigma]$

L'illustration 43 permet de comparer la puissance des tests Kendall Saisonnier et Mann-Kendall pour la détection des tendances des chroniques cycliques annuelles à calendriers réguliers journaliers. Un taux de rejet de H_0 proche de 100% indique que la méthode statistique détecte bien la tendance.

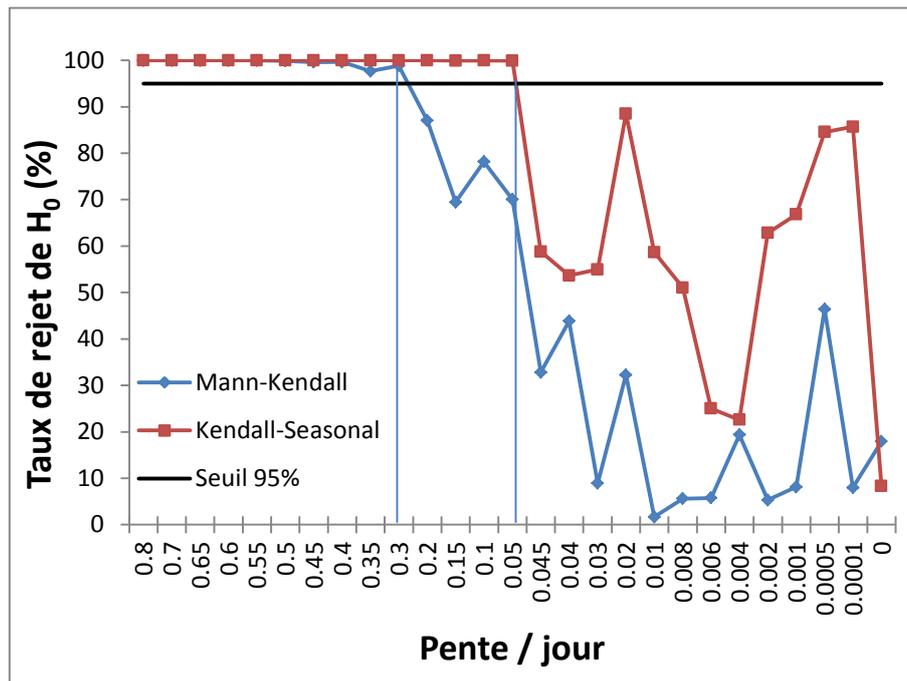


Illustration 43 : Comparaison du taux de rejet de H0 des méthodes Mann-Kendall et Kendall Saisonnier dans le cas de chroniques cycliques annuelles à calendriers réguliers journaliers sous l'effet de la diminution de la pente des tendances (taux de rejet attendu = 100%).

Ces premiers résultats permettent de déterminer la pente minimale à partir de laquelle la méthode ne détecte plus de tendance alors qu'elle existe. Pour le test de Mann-Kendall, cette valeur est établie à 0,3 / an alors que le test Kendall Saisonnier détecte des pentes faibles jusqu'à 0,05 / an soit un gain d'un facteur 6. Ce graphique montre déjà tout l'intérêt de l'utilisation du test Kendall Saisonnier en lieu et place du test de Mann-Kendall pour l'évaluation de la stationnarité de chroniques cycliques à courts termes et à calendrier de prélèvement régulier.

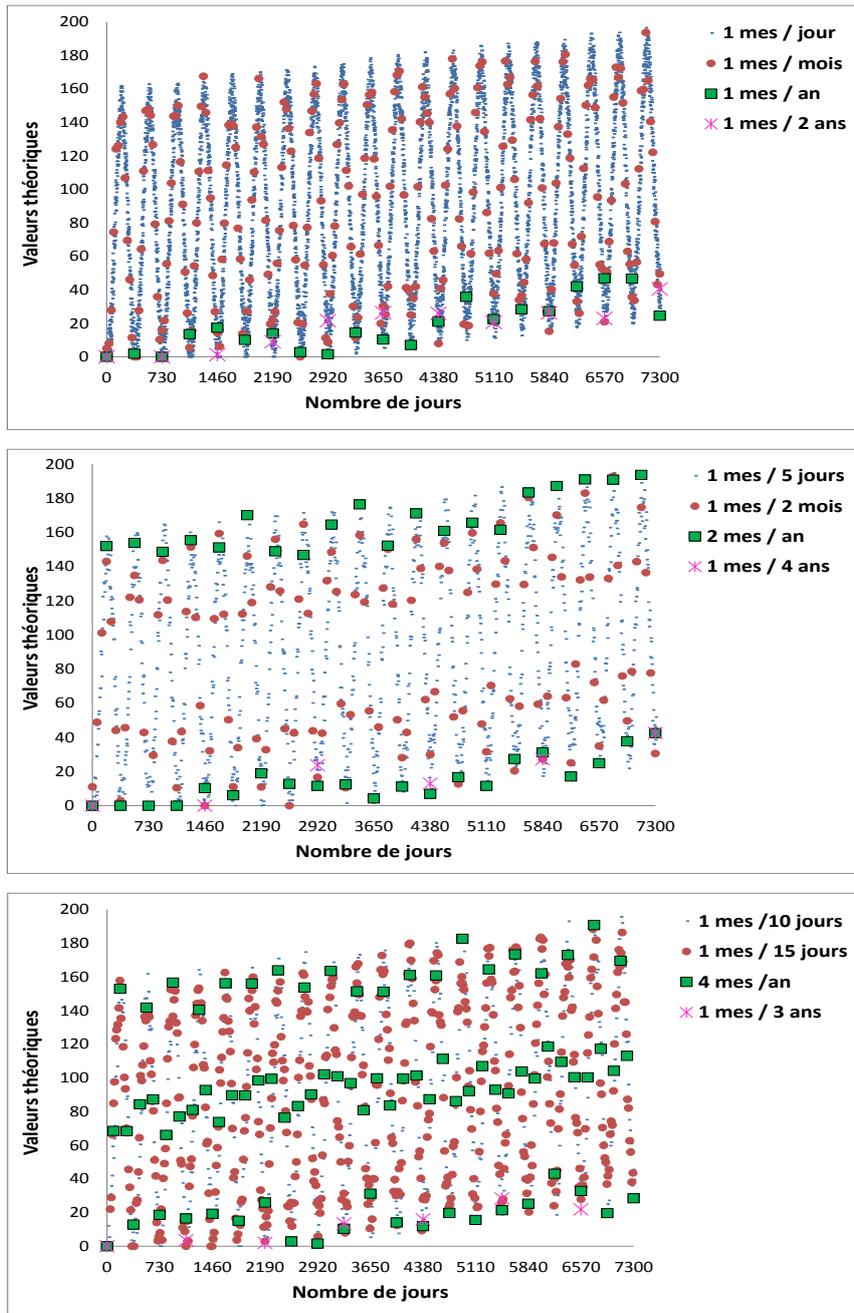


Illustration 44 : Chroniques cycliques annuelles à tendance linéaire de 0,5/an simulées selon 12 fréquences de prélèvements différentes.

Par la suite, les comportements des tests face à la fréquence de prélèvements sont comparés pour trois pentes « a » de tendance données (0,5/an ; 0,2/an et 0,1/an). En partant du même modèle de chroniques par variogramme que précédemment, 12 fréquences de prélèvements sont simulées (1 mesure/jour ; 1 mes/5 jours ; 1 mes/10 jours ; 1 mes/15 jours ; 1 mes/mois ; 1 mes/2 mois ; 4 mes/an ; 2 mes/an ; 1 mes/an ; 1 mes/2 ans ; 1 mes/3 ans ; 1 mes/4 ans). Les trois graphiques présentés en Illustration 44 illustrent l'allure des

chroniques pour une pente de tendance identique de 0,5/an. Chaque graphique représente 4 chroniques à fréquences de prélèvement différentes.

Les méthodes Mann-Kendall et Kendall Saisonnier sont appliquées sur chacune de ces chroniques affectées par des tendances linéaires de 0,5 ; 0,2 et 0,1/an. L'illustration 45 montre le résultat de ces tests.

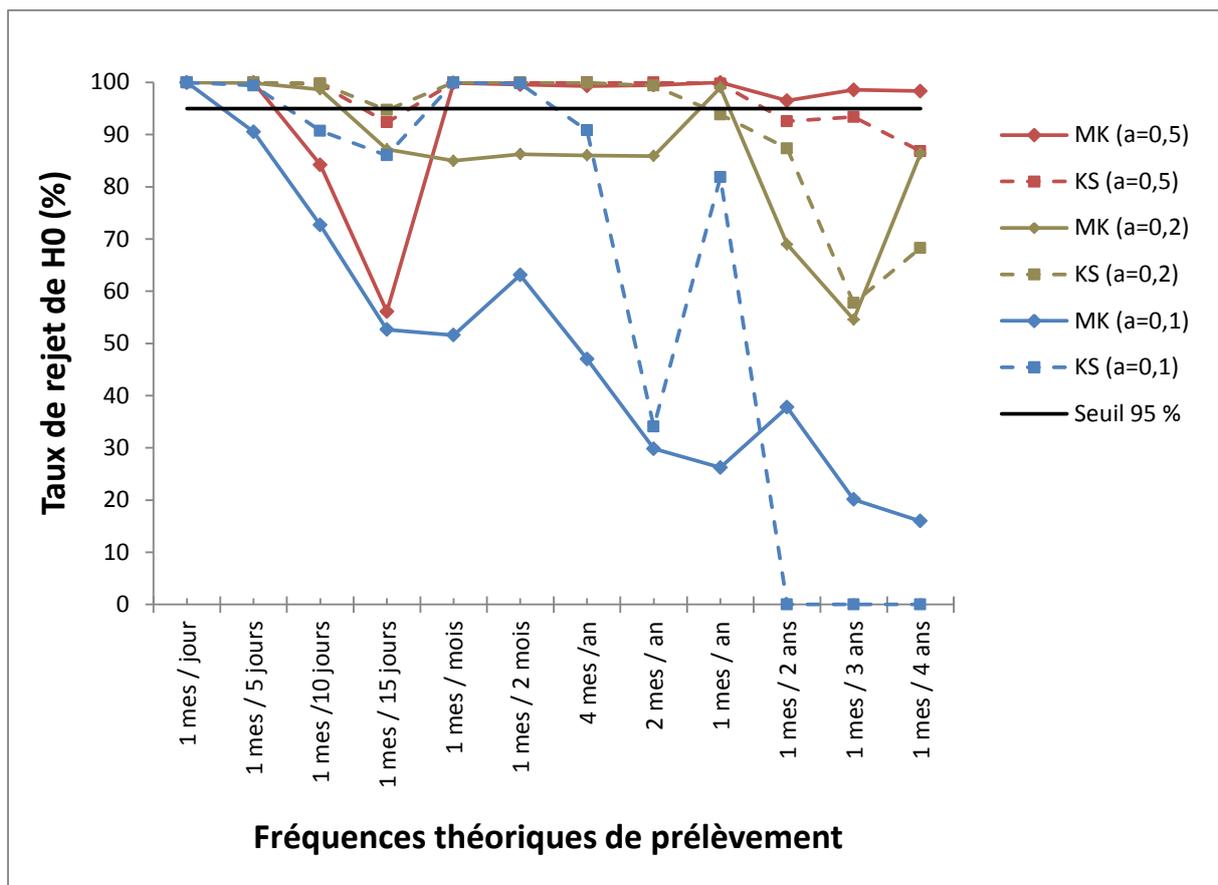


Illustration 45 : Comparaison du taux de rejet de H0 des tests de Mann-Kendall (MK) et Kendall Saisonnier (KS) sur des chroniques cycliques annuelles à tendances linéaires de pentes a en fonction de la fréquence de prélèvement (taux de rejet attendu = 100%).

Pour chaque pente de tendance testée, des fréquences de prélèvements optimales se dégagent :

- Pour les pentes de 0,5/an, les méthodes de Mann-Kendall et Kendall Saisonnier détectent bien les tendances pour les fréquences de prélèvements testées comprises entre 1mes/mois et 1mes/an voire au-delà pour le Mann-Kendall. Pour les fréquences inférieures, les deux méthodes montrent une faible efficacité lorsque les prélèvements sont réalisés au pas de temps de 15 jours, notamment le Mann-Kendall qui ne montre pas non plus de bons résultats pour des prélèvements tous les 10

jours. Les tendances sont toutefois à nouveau détectées par les deux méthodes statistiques pour des prélèvements journaliers ou tous les 5 jours. Ce phénomène n'est aujourd'hui pas encore expliqué.

- Pour les pentes de 0,2/an, la méthode Mann-Kendall ne détecte les tendances que pour les fréquences de prélèvements testées inférieures ou égales à 1 mesure tous les 10 jours ou bien égales à 1 mesure par an. La méthode Kendall Saisonnier montre quant à elle de bons résultats pour toutes les fréquences de prélèvements testées inférieures ou égales à 2 mesures par an. La plage optimum de fréquence de prélèvement est donc bien supérieure pour la méthode Kendall Saisonnier que la méthode Mann-Kendall.
- Pour des pentes de 0,1/an, seule une fréquence de prélèvement journalière permet au test de Mann-Kendall de détecter les tendances alors que la méthode Kendall Saisonnier fonctionne pour des fréquences de 1mes/jour, 1mes/5jours, 1mes/mois et 1mes/2mois. Dans le cas de pentes de tendances faibles, la méthode Kendall Saisonnier accepte une plage de fréquences de prélèvements bien plus importante que le Mann-Kendall pour étudier la stationnarité des chroniques cycliques annuelles.

Ces résultats révèlent une puissance de la méthode Kendall Saisonnier plus importante que la méthode Mann-Kendall pour détecter des tendances de faibles pentes lorsque les chroniques sont de type cyclique annuelle et les calendriers de prélèvements de type régulier et à fréquences de prélèvements comprises entre 1 mesure tous les 5 jours et 2 mesures par an. Si les pentes sont très faibles (0,1/an testée) la plage optimale de fréquences de prélèvements pour l'application du test Kendall Saisonnier est réduite à 1 mesure par mois ou 1 mesure tous les 2 mois.

S'il est ainsi démontré la plus grande efficacité du test Kendall Saisonnier par rapport au test de Mann-Kendall dans le cas de chroniques cycliques annuelles, l'inconvénient majeur de l'utilisation de cette méthode est qu'elle nécessite un calendrier de prélèvement régulier, l'idéal étant vraisemblablement des prélèvements mensuels. Or, dans la réalité, l'opérateur est souvent confronté à des calendriers de prélèvements irréguliers qui imposent :

- soit de régulariser les pas de temps de prélèvements lorsque ceux-ci se rapprochent de la régularité (quelques trous éparses seulement dans le calendrier)
- soit d'utiliser une autre méthode d'évaluation de la stationnarité de la chronique.

Toutefois, il a été démontré au paragraphe 5.3.3 qu'un calendrier irrégulier tend à rompre l'autocorrélation des données et donc favorise l'efficacité des tests classiques non

paramétriques d'évaluation de la stationnarité. Il sera donc recommandé l'utilisation de la méthode Kendall Saisonnier pour la détection des tendances de chroniques cycliques à pas de temps de prélèvements réguliers, et la Mann-Kendall voire Mann-Kendall modifiée pour la détection des tendances de chroniques cycliques à pas de temps de prélèvements très irréguliers.

6 Recommandations pour la détermination des tendances d'évolution de la qualité des eaux souterraines

La revue bibliographique sur les méthodes statistiques existantes pour l'évaluation et la quantification des tendances d'évolution de la qualité des eaux, l'analyse des études déjà réalisées dans ce sens, la revue bibliographique des tests de robustesse et de puissance de ces méthodes statistiques ainsi que les tests complémentaires réalisés spécifiquement après analyse de la qualité des données rencontrées dans la base de données nationale sur les eaux souterraines en France, permettent de faire des préconisations à la fois sur une démarche à adopter et sur des tests statistiques à choisir afin d'optimiser l'analyse des tendances d'évolution de la qualité des eaux souterraines. Néanmoins, la liste des outils proposés n'a pas la prétention d'être exhaustive, les recommandations les plus fortes portant plus sur la démarche à adopter pour l'analyse des chroniques d'évolution de la qualité des eaux souterraines.

6.1 Etapes principales pour l'analyse des tendances d'évolution de la qualité des eaux souterraines

Comme démontré par les tests de robustesse et de puissance réalisés au chapitre 5, la quantité et la qualité des données disponibles jouent un rôle prépondérant dans l'analyse des tendances et des ruptures de pentes. Un effort important doit donc être consenti afin d'analyser et de valider le jeu de données disponibles avant la mise en œuvre des tests statistiques. Tennakoon et al. (2009) proposent ainsi une approche systématique d'analyse et de traitement des données de qualité des eaux afin 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 :

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

6.1.1 Validation 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 et 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 9, 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 toutes analyses 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 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 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 le test de Mann-Kendall ou le test de Spearman.

6.1.2 Analyse visuelle 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 techniques permettent d'ajuster les données brutes comme la moyenne mobile, la régression linéaire, la régression quadratique ou encore le lissage de LOWESS (Cleveland, 1979). Il est aussi possible de tracer le variogramme 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 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 monotone 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-Seasonal.

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 (paragraphe 2.2.8) 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.

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.

6.1.3 Analyse statistique

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 au chapitre 6.3 permet de l'aider à orienter son choix en fonction de toutes ces caractéristiques.

6.1.4 Détermination des facteurs explicatifs

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 de ces variations. L'étape qui consiste à expliquer les tendances d'évolutions ou les changements de la qualité des eaux est donc primordiale dans le processus global 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. La gestion intégrée de la qualité des eaux souterraines demande à la fois une bonne connaissance des contextes géologiques et

environnementaux ainsi que de l'historique d'évolution des pratiques de gestion et d'usages du bassin versant hydrogéologique considéré.

Ainsi, dans le cadre de la recherche des facteurs qui expliquent les évolutions des concentrations en polluants dans les eaux souterraines, il est intéressant de tester dans quelles mesures ces évolutions peuvent être rapprochées de celles d'autres phénomènes potentiellement associés. Il est en effet reconnu que dans certains contextes climatiques et hydrogéologiques, les fluctuations des niveaux des nappes peuvent impacter les variations de concentrations en polluants stockés dans la zone non saturée et mobilisables lors des remontées de nappes. De plus, dans des contextes à circulations rapides d'eaux récentes, des fortes variations de la chimie des eaux souterraines peuvent apparaître lors de forts événements pluvieux. Ces fluctuations peuvent être la conséquence d'une dilution des eaux souterraines par les eaux de pluies peu chargées 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.

Plusieurs approches permettent 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 dans les eaux en fonctions 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. 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.

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.

6.2 Questions préliminaires pour le choix et la mise en œuvre d'un test de tendance et de rupture

Imaginons qu'après avoir collecté une série de données chronologiques sur l'évolution d'un paramètre de la qualité des eaux souterraines, après avoir validé ces données et réalisé leur analyse visuelle et exploratoire, une tendance semble se dessiner après une rupture évidente de la pente. Visuellement, il semble possible d'estimer à la fois la date du changement de comportement, le sens et la valeur de la pente de la tendance. Toutefois, ces estimations sont très subjectives et ne reposent sur aucun calcul précis. Afin de tester ces hypothèses, d'estimer le niveau de significativité de celles-ci et de disposer de données chiffrées sur la date de changement et la pente de la tendance, il est nécessaire de faire appel aux outils statistiques d'analyse des séries chronologiques. Vient alors la question du choix du test statistique à appliquer sur la série de données.

- **Hypothèse sur la série de donnée – tests paramétriques ou non paramétriques ?**

Les tests paramétriques nécessitent, pour être appliqués dans de bonnes conditions, que le modèle d'ajustement des chroniques soit connu, que les « outliers » soient absents ou supprimés et que la distribution des données soit normale (ou rendue normale). Or, les tests non paramétriques sont efficaces et puissants dans le cas d'une distribution normale des données, montrent des avantages potentiels assez faibles lorsque les données s'écartent de façon imperceptible de la normalité mais montrent de grands avantages lorsque les données sont à distribution non normale. De plus, ils sont très peu sensibles à la présence de données extrêmes et ne nécessitent pas d'un modèle a priori d'ajustement aux chroniques.

Dans le cas d'une analyse réalisée sur un grand jeu de chroniques, les tests non paramétriques seront préférés car il sera ainsi possible d'appliquer le même test statistique sur toutes les séries, qu'elles soient à distribution normale ou non, que les données soient affectées par des valeurs extrêmes ou non. Ces vérifications n'auront en effet pas besoin d'être effectuées.

Dans le cas d'une analyse sur une seule chronique, l'utilisation d'un test paramétrique sera recommandée. Dans le cas d'une distribution non normale des données il pourra être tenté une transformation des données pour les rendre normalement distribuées. Les tests paramétriques sont en effet généralement plus informatifs que les tests non-paramétriques dans les cas de figures très complexes.

Dans le cas du traitement d'un nombre important de chroniques, il n'est pas recommandé de transformer les données car la transformation est subjective et serait différente pour l'ensemble des séries analysées. Il deviendrait alors impossible de comparer des résultats issus de processus de traitements différents.

- **Type de tendance – tendance par paliers ou tendance monotone ?**

Deux grands types de tendances peuvent être considérés pour les tests d'hypothèses et l'estimation des variations des séries temporelles : les tendances par paliers et les tendances monotones.

La première hypothèse est l'existence de tendances par paliers. Cette hypothèse suppose que les données collectées avant une date spécifique appartiennent à une population différente de celles collectées après cette date. La distinction entre les deux populations est

caractérisée par des moyennes (ou des médianes) significativement différentes. Les méthodes statistiques adaptées au test de cette hypothèse sont ainsi basés sur l'analyse des différences entre les moyennes (ou les médianes) de plusieurs échantillons d'une même population (constituant une série temporelle).

La deuxième hypothèse est que la population varie de façon monotone (pas d'inversion de direction) sur la durée totale investiguée. Cette hypothèse ne spécifie pas en revanche si la variation est continue, linéaire ou si elle relève d'autres structures spécifiques.

L'hypothèse d'une tendance par palier est plus spécifique que celle d'une tendance monotone car elle requiert la connaissance de la (ou des) date(s) du (ou des) palier(s). Dans ce cas particulier, cette donnée doit être estimée lors de la phase exploratoire des données. Les méthodes de détection des tendances par palier doivent être utilisées dans deux cas de figures particuliers.

Le premier correspond aux chroniques naturellement scindées en deux périodes distinctes avec un intervalle relativement important entre elles. Il n'existe pas de règle pour déterminer la longueur minimum d'intervalle impliquant le choix d'une telle méthode. Mais si sa longueur dépasse 1/3 de la longueur totale de la chronique les méthodes de tendance par palier semblent être les plus appropriées, même si la tendance apparaît linéaire. D'une manière générale, si les périodes renseignées semblent courtes en comparaison avec la longueur de l'intervalle entre les deux périodes, alors le test d'une tendance par palier doit être envisagé.

Le second cas de figure correspond aux chroniques dont il est possible de définir un « avant » et un « après » événement connu. Cet événement peut être l'introduction d'une nouvelle source de contamination, la mise en place d'une action visant à réduire l'utilisation d'un polluant ou bien tout autres événements impactant fortement le comportement du paramètre étudié dans les eaux souterraines.

- **Type de méthode – suppression des sources de variations différentes de la tendance ?**

Il a été montré que, dans les eaux souterraines, l'évolution de certains paramètres peut être corrélée avec des phénomènes extérieurs cycliques, comme la variation du niveau des nappes ou bien l'intensité des pluies efficaces en zone de recharge. Cette corrélation peut parfois masquer la tendance d'évolution (non naturelle) du paramètre étudié en abaissant significativement la puissance et la robustesse des tests statistiques appliqués sur de telles séries de données. Il est alors possible de retirer aux données leur composante cyclique et de procéder au test sur une série dite « désaisonnalisée ». Un moyen de faire cette

désaisonnalisation est de faire une décomposition du signal pour en extraire la part expliquant les variations cyclique. Dans le cas d'une corrélation avec les variations saisonnières du niveau de nappe, les résidus obtenus par la régression entre les concentrations et les fluctuations piézométriques seraient ajoutés à la moyenne des concentrations observées sur la période de mesure.

Toutefois, cet exercice ne peut être réalisé que si le phénomène qui explique les variations cycliques est connu et enregistré sur la même période et à la même fréquence (ou plus forte) que la chronique étudiée.

6.3 Arbres décisionnels pour le choix d'un test statistique de stationnarité

La revue bibliographique sur les méthodes statistiques employées dans le domaine de la qualité des eaux, l'analyse des caractéristiques des données spécifiques aux eaux souterraines en France et les tests de robustesse et de puissance des méthodes statistiques sur des séries reprenant ces caractéristiques ont permis de définir une méthodologie robuste d'analyse des chroniques en vue d'évaluer les tendances d'évolution. Il est ainsi préconisé de travailler en quatre phases :

- 1) validation des données
- 2) analyse exploratoire des séries chronologiques
- 3) mise en œuvre des tests statistiques
- 4) estimation des facteurs explicatifs (peut être réalisée avant afin d'orienter le choix du test)

Les trois premières phases, détaillées dans le rapport, sont présentées sous la forme de schémas simplifiés (Illustration 46, Illustration 47 et Illustration 48). Ils constituent la synthèse des travaux réalisés pour cette étude et permettent d'accompagner l'opérateur dans son exercice d'analyse des séries chronologiques. Faire l'étude de chroniques de la qualité des eaux revient alors à se poser un certain nombre de questions dont chacune des réponses renseigne soit sur une orientation particulière à prendre pour la suite de l'analyse, soit sur une mauvaise utilisation potentielle d'une méthode liée à une caractéristique de la chronique, soit, pour la troisième phase, sur le (ou les) test(s) statistique(s) à appliquer sur la (ou les) chronique(s).

L'Illustration 46 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 validation des données, l'opérateur maîtrise 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.

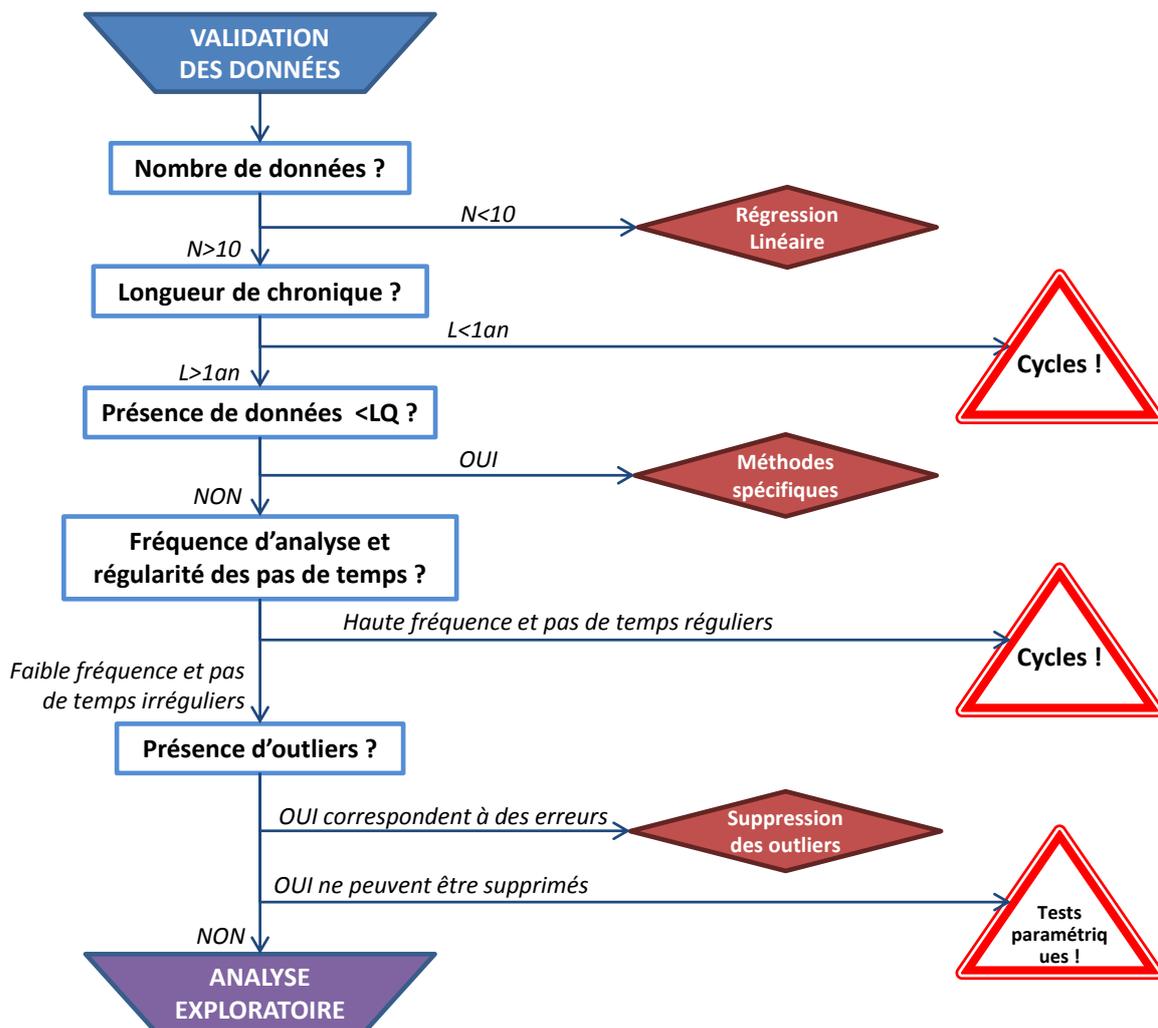


Illustration 46 : Schéma de validation des données brutes. A gauche, questions successives pour la validation ; au centre, orientations préliminaires ; à droite, alertes sur des processus ou l'utilisation de certaines méthodes (attention aux [...]).

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 47. 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.

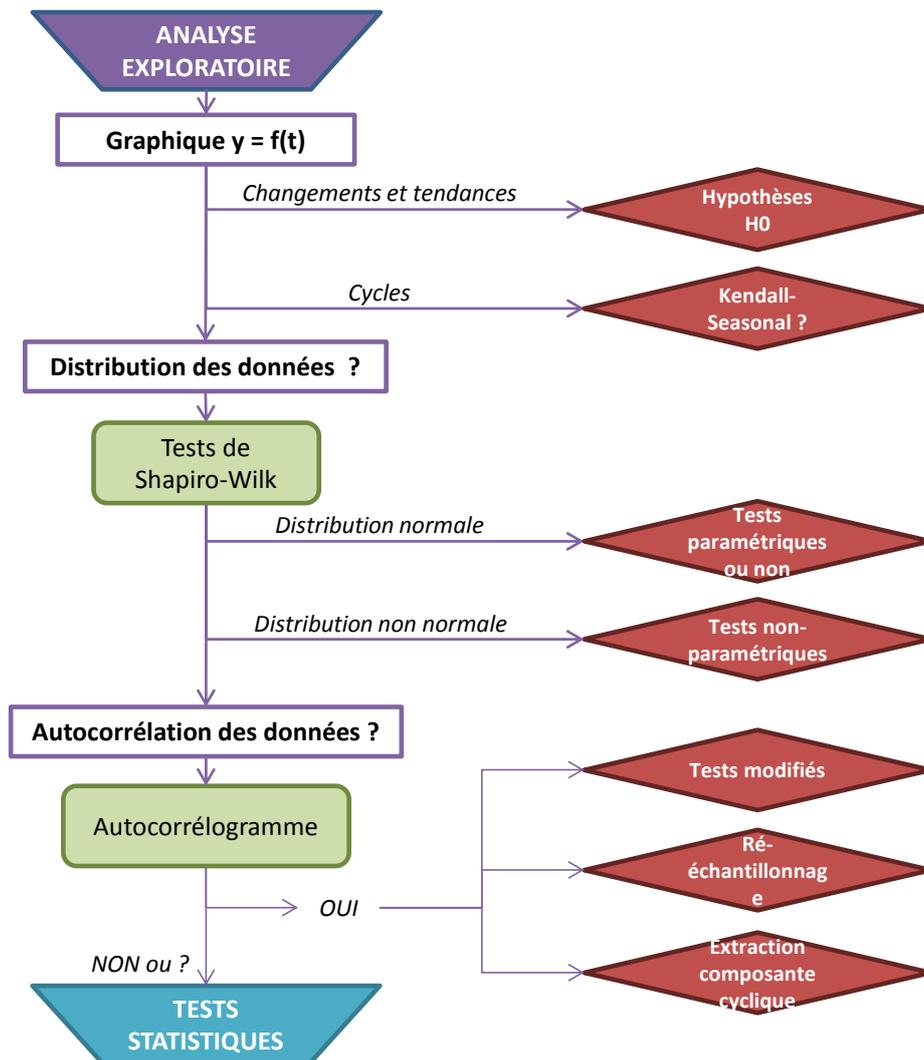


Illustration 47 : Schéma d'analyse exploratoire des données. A gauche, questions successives avec indication de la méthode à appliquer ; à droite, orientations préliminaires sur les actions à réaliser ou les méthodes pressenties pour l'analyse statistique.

La série chronologique à analyser étant bien connue à ce stade, il est possible de choisir la méthode statistique pertinente en fonction des caractéristiques des données et de l'hypothèse à tester. L'illustration 48 montre le schéma de sélection du test statistique à appliquer en fonction des connaissances sur la série acquises lors des deux précédentes phases. 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.

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 s'est montré moins efficace que la version modifiée dans le cas de données autocorrélées, les résultats sont restés 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 possible de mettre en œuvre le 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 son efficacité dans le cas de données fortement autocorrélées n'est pas trop mauvaise. Il sera juste nécessaire de préciser que les processus d'autocorrélation n'ont pas été vérifiés impliquant une possible surestimation des tendances identifiées à la hausse

7 Conclusions et perspectives

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 et 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 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 polluant au point de mesure est suffisante.

En France, les organismes publics de gestion de la qualité des eaux sont actuellement très attachés à l'évaluation de l'état des masses d'eau (superficielles et souterraines) en rapport avec les exigences de la DCE. Ainsi, l'estimation de la qualité des eaux est essentiellement basée sur l'outil SEQ-Eau développé dès 1971 pour les cours d'eau. Les agences de l'eau et le Ministère en charge de l'Environnement ont souhaité mettre en place un outil similaire pour les eaux souterraines. Ce travail a abouti à la publication en 2002 du SEQ - eaux souterraines version 0. Cette version du SEQ adaptée aux eaux souterraines rend compte des spécificités de ces dernières, tout en restant cohérent avec le système mis au point pour les eaux superficielles. Pour faire suite au SEQ-Eaux (reprit et adapté par d'autres pays européens comme la Belgique et la Pologne) un nouvel outil baptisé SEEE pour Système d'Évaluation de l'État des Eaux est en cours de développement. Le pilotage du projet est assuré par l'ONEMA. A cette occasion, avec le soutien du groupe national « DCE eaux souterraines » qui réunit notamment la Direction de l'Eau et de la Biodiversité du MEDDTL et les Agences de l'eau, un recensement des besoins pour le traitement de données et la production d'indicateurs statistiques a été mené. Conformément au périmètre prévu pour l'outil SEEE, ce recensement exclue les indicateurs cartographiques et graphiques.

L'outil SEEE est encore en cours de développement mais conformément aux résultats du recensement des besoins l'outil permettra la réalisation de statistiques assez simples : concentrations maximales, minimales, médiane, percentiles, fréquences de dépassement d'un seuil, taux de quantification, évolution des dépassements, etc... L'accent étant mis sur la production d'un indicateur conforme aux exigences de la DCE, il pourrait être intéressant de réfléchir à l'élaboration d'un outil permettant le traitement statistique des chroniques d'évolution de la qualité des eaux souterraines et notamment l'estimation des tendances d'évolution des concentrations en polluants. Ce rapport peut constituer une base de réflexion solide sur les méthodes qui pourraient être intégrées ou bien qui pourraient constituer un module d'accompagnement de l'outil SEEE. Les préconisations sur le choix des tests statistiques à appliquer en fonction des caractéristiques des données à traiter pourront aussi être reprises pour l'élaboration du guide méthodologique pour l'évaluation des tendances d'évolution des polluants dans les masses d'eau souterraines dont la rédaction devrait débiter en 2011.

Enfin, au-delà de la présentation et du test d'un panel important de méthodes statistiques, l'étude montre la quantité importante de données sur les eaux souterraines bancarisées dans ADES qui pourraient être valorisées selon une approche statistique. Les caractéristiques de ces données sont en effet tout à fait compatibles avec les potentialités d'application de méthodes adaptées. Les longueurs moyennes ainsi que les fréquences moyennes de prélèvements des chroniques d'évolution des concentrations en éléments majeurs polluants (NO_3^- , SO_4^{2-}) permettent de travailler sur des tendances récentes après mise en évidence de ruptures de comportements. Des méthodes existent pour prendre en compte des données dont la distribution n'est pas gaussienne ainsi que pour s'adapter au phénomène d'autocorrélation. Le nombre important de points renseignés sur ces paramètres permet en outre de réaliser des études spatialisées des tendances d'évolution des concentrations.

En ce qui concerne les micropolluants, des chroniques de plus en plus longues sont disponibles de sorte que le recul nécessaire à la compréhension de leur comportement dans le milieu naturel est atteint pour certaines substances. La mise en œuvre des tests statistiques est toutefois compliquée, dans la majeure partie des cas, par la présence de données inférieures à une limite (quantification ou détection) qu'il est nécessaire de prendre en compte. Des pistes de réflexions sont explorées dans le rapport, tant pour l'exploitation de ces données en termes de fréquence d'occurrence que pour la reconstitution des chroniques par l'application de modèles de distribution des données. Il serait ainsi

intéressant de tester les différentes approches proposées sur des jeux de données réels de la qualité des eaux souterraines, dont on sait qu'ils peuvent se prêter à l'analyse statistique.

ONEMA
Hall C – Le Nadar
5 square Félix Nadar
94300 Vincennes
01 45 14 36 00
www.onema.fr

BRGM service EAU
3, avenue Claude-Guillemin
B.P. 36009
45060 Orléans cedex 2
02 38 64 34 34
<http://www.brgm.fr>

Evaluation des tendances d'évolution des polluants dans les eaux souterraines, B. Lopez, A. Leynet

Bibliographie

ADES Banque Nationale d'Accès aux Données sur les eaux souterraines.

<http://www.ades.eaufrance.fr>

Arrêté du 17/12/08 établissant les critères d'évaluation et les modalités de détermination de l'état des eaux souterraines et des tendances significatives et durables de dégradation de l'état chimique des eaux souterraines. JO n° 5 du 7 janvier 2009.

Battle-Aguilar J., Orban Ph., Dassargues A., Brouyère S. (2007) - Identification of groundwater quality trends in a chalk aquifer threatened by intensive agriculture in Belgium, *Hydrogeology Journal*, 15: 1615-1627.

Baran N., Gourcy L., Lopez B., Bourguin B., Mardhel V., (2009) - Transfert des nitrates à l'échelle du bassin Loire-Bretagne. Phase 1 : temps de transfert et typologie des aquifères. Rapport BRGM RP-54830-FR, 105 p.

Bernard-Michel C., (2006) - Indicateurs géostatistiques de la pollution dans les cours d'eau. Thèse de l'Ecole des Mines de Paris spécialité « Géostatistiques ».

Bexfield L. M., (2008) - Decadal-Scale Changes of Pesticides in Ground Water of the United States, 1993–2003. *Journal of Environment Quality* 37: 226-239.

Bouchard A. et Haemmerli J., (1991) - Trend detection in water quality time series of LRTAP-Quebec network lakes. *Water, Air, and Soil Pollution* 62: 89-110.

Boyer J. N., Sterling P., Jones R. D., (2000) - Maximizing Information from a Water Quality Monitoring Network through Visualization Techniques. *Estuarine, Coastal and Shelf Science* 50: 39–48.

Darken P. F., Zipper C. E., Holtzman G. I., Smith E. P., (2002) - Serial correlation in water quality variables: Estimation and implications for trend analysis. *Water Resources Research* 38(7): 1117.

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

Cleveland, W.S. (1979) - Robust locally weighted regression and smoothing scatter plots. Journal of American Statistical Association Vol. 74 (368), 829-836.

Frans L., (2008) - Trends of pesticides and nitrate in ground water of the Central Columbia Plateau, Washington, 1993-2003. Journal of Environmental Quality, 37: 273-280.

Frapporti G., Vriend S. P., Van Gaans P. F. M., (1994) - Qualitative time trend analysis of ground water monitoring networks. An example from the Netherlands. Environmental Monitoring and Assessment 30: 81-102.

Ghestem J-P., Guigues N., Yardin C., (2010) - Surveillance de polluants dans les eaux souterraines : impacts de l'incertitude de mesure sur l'évaluation des tendances. Rapport BRGM-RP-59342-FR, 100p.

Grath J., Scheidleder A., Uhlig S., Weber K., Kralik M., Keimel T., Gruber D., (2001) - The EU Water Framework Directive: Statistical aspects of the identification of groundwater pollution trends, and aggregation of monitoring results, Final Report, 63p. www.wfdgw.net

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

Hamed K. H., (2008) - Trend detection in hydrologic data: The Mann–Kendall trend test under the scaling hypothesis. Journal of Hydrology 349: 350-363.

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.

Hiez G., (1977) - L'homogénéisation des données pluviométriques. Cahiers Orstom, série hydrologie, 14(2): 129-172.

Hirsch, R.M. et Slack J. R., (1984) - A nonparametric trend test for seasonal data with serial dependence. *Water Resources Research* 20: 727-732.

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.

Larpin O., Chery L., Elsass P., Kloppann W., (2003) - Contribution à la caractérisation des états de référence géochimique des eaux souterraines. Application de la méthodologie à la nappe du Rhin. BRGM/RP-52163-FR.

Libiseller C. et Grimvall A., (2002) - Performance of partial Mann-Kendall tests for trend detection in the presence of covariates. *Environmetrics* 13: 71-84.

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.

Loftis J. C., Taylor C. H., Chapman P. L., (1994) - Multivariate Tests for Trend in Water Quality. *Water Resources Research* 27: 419-429.

Margane A., (2004) - Management, Protection and Sustainable Use of Groundwater and Soil Resources in the Arab Region. Volume 7. Guideline for groundwater monitoring. ACSAD-BGR Technical Cooperation Project. 328p.

McLeod-Hipel & Assoc. Ltd., (1991) - Trend analysis methodology for water quality time series. R.A.C. Project no. 278C prepared for Environment Ontario.

Olmedo J. G., Luque Espinar J. A., (2010) - Guidelines for statistical analysis of hydrological data according to the requirements of the Groundwater Directive. AQUAmundi - Am01003: 013 – 021.

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).

Ryberg, K.R., and Vecchia, A.V., (2006) - Water-quality trend analysis and sampling design for the Devils Lake Basin, North Dakota, January 1965 through September 2003: U.S. Geological Survey Scientific Investigations Report 2006-5238, 64 p.

Smitz J., Everbecq E., Deliege J. F., Descy J. P., Wollast R., Vanderborght J. P., (1997) - Pégase, une méthodologie et un outil de simulation prévisionnelle pour la gestion de la qualité des eaux de surface = Pégase, a methodology and a previsional simulation tool for management of surface water quality. La Tribune de l'eau 50(588) : 3-72 83-87 73-82.

Steele T. D., (1974) - Converting water quality information goals into statistical design criteria. Monitoring to Detect Changes in Water Quality Series (Proceedings of the Budapest Symposium, July 1986). IAHS Publ. no. 157.

Stuart M.E., Chilton P.J., Kinniiburgh D.G., Cooper D.M., (2007) - Screening for long-term trends in groundwater nitrate monitoring data. Quaterly Journal of Engineering Geology and Hydrogeology, 40: 361-376.

Tang S. M., Macneil I. B. (1991) - Monitoring statistics which have increased power over reduced time range. Environmental Monitoring and Assessment 23: 189-203

Tennakoon S., David Robinson D., Shen S. (2009) - Decision Support System for Temporal Trend Assessment of Water Quality Data. 18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009.

Thas O., Vooren L. V., Ottoy J. P. (1998) - Nonparametric Test Performance for Trends in Water Quality with Sampling Design Applications. Water resources bulletin 34: 347-357.

Tukey, J. W., (1977) - Exploratory Data Analysis. Addison-Wesley Pub., Reading MA, 506 p.

Van Belle G. et Hughes J.P., (1984) - Nonparametric tests for trend in water quality. Resources Research, 20(1): 127-136.

Visser A., Dubus I., Broers H. P., Brouy S., Korcz M., Orban P., Goderniaux P., Battle-Aguilar J., Surdyk N., Amraoui N., Job H., Pinault J-L., Bierkensab M., (2009) - Comparison of methods for the detection and extrapolation of trends in groundwater quality. Journal of Environmental Monitoring 11: 2030–2043.

Yu Y-S., Zou S., Whittemore D., (1993) - Non-parametric trend analysis of water quality data of rivers in Kansas. Journal of Hydrology, 150: 61-80.

Yue S., Pilon P., Phinney B., Cavadias G. (2002) - The influence of autocorrelation on the ability to detect trend in hydrological series. Hydrological Processes 16: 1807-1829.

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.

