



SURVEILLANCE DE POLLUANTS DANS LES EAUX SOUTERRAINES : IMPACT DE L'INCERTITUDE DE MESURE SUR L'ÉVALUATION DES TENDANCES

Savoirs - Action n° 7

**JP GHESTEM (BRGM) N GUIGUES, C YARDIN (LNE)
Décembre 2010**

Programme scientifique et technique
Année 2010

Document final

En partenariat avec



Avec l'approbation et le soutien de



et le soutien de



Contexte de programmation et de réalisation

Les travaux présentés dans ce rapport ont été réalisés dans le cadre des activités de Service Public du BRGM avec le soutien de l'ONEMA (convention de partenariat n° 1900/09 - Année 2010). Ces travaux entrent dans le cadre du programme de travail d'AQUAREF pour l'année 2010

Jean-Philippe GHESTEM

Chef de Projet - Service Métrologie, Monitoring, Analyse
jp.ghestim@brgm.fr
3 av Claude Guillemin BP36009 45060 ORLEANS CEDEX 2

Nathalie GUIGUES

Chef de Projet Qualité des Eaux-Direction Métrologie Scientifique et Industrielle / Pôle Chimie Biologie
Nathalie.guigues@lne.fr

Et

Catherine YARDIN

Service Mathématiques et Statistiques
Catherine.yardin@lne.fr
Laboratoire national de métrologie et d'essais, 1, rue Gaston Boissier - 75724 Paris cedex 15

Vérification du document

Anne MORIN

Directrice du programme AQUAREF
33 (0)3 44 55 66 32
anne.morin@ineris.fr

Marina COQUERY

Laboratoire d'analyses physico-chimiques des milieux aquatiques UR Milieux Aquatiques, Ecologie et Pollutions 3bis quai Chauveau, CP 220, F-69336 Lyon cedex 09 Tel : 33 (0)4 72 20 86 16 - Cemagref Lyon
Marina.coquery@cemagref.fr

Les correspondants

Onema : Pierre François STAUB, ONEMA, DAST, pierre-françois.staub@onema.fr

Etablissement : Jean Philippe GHESTEM, BRGM, Service MMA, jp.ghestim@brgm.fr

Référence du document : AQUAREF 2010 - Rapport BRGM/RP-59432-FR

GHESTEM JP, GUIGUES N, YARDIN C (2010) - Surveillance de polluants dans les eaux souterraines : impact de l'incertitude de mesure sur l'évaluation des tendances. Rapport BRGM-RP-59342-FR, 108 pages 1 Annexe, 19 Figures, 16 Tableaux.

Droits d'usage :	<i>Accès libre</i>
Couverture géographique :	<i>International</i>
Niveau géographique :	<i>National</i>
Niveau de lecture :	<i>Professionnels, experts</i>
Nature de la ressource :	<i>Document</i>

Surveillance de polluants en eau souterraine : impact des incertitudes de mesure sur l'évaluation des tendances

Rapport final

BRGM/RP-59342-FR
Décembre 2010

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

JPh GHESTEM (BRGM), N GUIGUES, C YARDIN (LNE)

Vérificateur :

Nom : B LOPEZ

Date : 18/01/11

Signature :

Approbateur :

Nom : G. HERVOUET

Date : 24/01/11

Signature :

Mots clés : tendance, eau souterraine, incertitude, analyse, échantillonnage, variance,

En bibliographie, ce rapport sera cité de la façon suivante :

GHESTEM JP, GUIGUES N, YARDIN C (2010) – Surveillance de polluants dans les eaux souterraines : impact de l'incertitude de mesure sur l'évaluation des tendances. Rapport BRGM-RP-59342-FR, 108 pages 1 Annexe, 19 Figures, 16 Tableaux.

Synthèse

La directive cadre européenne sur l'eau fixe aux états membres, des objectifs en termes d'identification des tendances d'évolution des polluants dans les masses d'eau notamment pour les eaux souterraines et les supports biote et sédiments pour les eaux de surface.

Afin d'atteindre ces objectifs et compte tenu des plans d'actions conséquents qui peuvent être mis en œuvre à l'issue des programmes de surveillance relatifs à l'identification des tendances, il semble indispensable de disposer d'éléments techniques permettant :

- de dimensionner un programme de surveillance compatible avec une évaluation de tendance répondant à des objectifs quantitatifs donnés de détection
- de connaître la fiabilité des tendances identifiées sur des chroniques existantes

Ce rapport rédigé dans le cadre du programme de travail d'AQUAREF pour l'année 2010 a pour objectif d'apporter des éléments de réponse à ces questions en s'intéressant à l'impact de la qualité des données de surveillance et notamment leur incertitude sur la capacité à détecter des tendances de façon fiable principalement dans le contexte des eaux souterraines. Il n'a pas pour objectif de discuter de façon détaillée des tests statistiques disponibles et de leurs avantages et limites. Cette discussion est réalisée par ailleurs en ce qui concerne les eaux souterraines (cf Lopez & al. 2010 [12]).

Les différents éléments discutés dans ce rapport conduisent à quelques propositions adaptées aux deux situations suivantes:

1. Programme de surveillance « tendance » en cours de préparation (cas idéal et à favoriser dans la mesure du possible)
 - a. Définir des objectifs quantitatifs pour le programme en termes de durée, de pente à détecter, de fréquence d'acquisition, de l'efficacité du programme, de taux de signification des tests statistiques.
 - b. Disposer d'informations concernant la variabilité des données (bibliographie, études préliminaires, études antérieures, ...)
 - c. Vérifier que le programme mis en œuvre permettra de remplir ces objectifs.
 - d. Si besoin le modifier éventuellement en cherchant à réduire si possible la variabilité des données

2. Programme de surveillance en cours (on dispose des chroniques de données ou bien elles sont en cours d'acquisition) :
 - a. Identification des données devant servir à l'évaluation des tendances
 - b. Expertise des données d'analyse en parallèle de l'exploitation statistique
 - c. Renforcement de la traçabilité documentaire sur ces données
 - d. Mise en place de tests de contrôle des données au fur et à mesure de leur acquisition
 - e. Calcul des écart-types résiduels des données (issus des tests statistiques d'évaluation de tendance) pour estimer leur variabilité et préciser si possible la part de l'incertitude analytique dans cette variabilité
 - f. Calcul du rapport pente/écart-type résiduel comme information supplémentaire empirique sur la fiabilité de la tendance identifiée.
 - g. Estimation de la puissance du test réalisé pour identifier la tendance dans les conditions du programme.

Ces quelques propositions constituent un résumé des différents chapitres du rapport dont les principaux éléments sont restitués ci-dessous.

Les principaux textes réglementaires (directive 2006/11/CE dite directive « eau souterraine » et directive 2008/105/CE dite « directive NQE ») ont tout d'abord été analysés ainsi que leurs guides d'application européens afin de faire ressortir les exigences et recommandations liées à l'estimation de tendances.

Une synthèse bibliographique a ensuite été réalisée avec l'objectif de rechercher des articles, abordant à la fois l'estimation de tendances sur des chroniques temporelles et également l'impact de l'incertitude analytique ou plus largement l'impact de la variabilité des données sur la fiabilité des tendances détectées. Le premier constat est que très peu de références existent sur ce sujet.

La plupart des articles utilisent la notion de puissance de tests qui traduit de façon schématique la capacité du test à détecter une tendance d'intensité donnée en une durée, dans des conditions et pour un degré de signification statistique donnée. Par extension on peut parler de la puissance ou efficacité d'un programme de surveillance comme étant la capacité du programme de surveillance à détecter de façon fiable une tendance d'intensité donnée.

Les principaux facteurs identifiés comme influents pour évaluer l'efficacité d'un programme sont :

- La durée de la chronique
- La variabilité des données (coefficient de variation CV)
- Le nombre de données
- Le degré de signification statistique du test.
- La pente de la tendance.

Que ce soit par simulation, ou bien par exploitation statistique, les références bibliographiques analysées fournissent des informations quantitatives (tableaux, abaques) sur les relations entre ces différents termes et sur la puissance des tests. A travers toutes ces données, il est donc possible par exemple :

- De déterminer la pente que l'on pourra détecter toutes conditions fixées par ailleurs.
- De déterminer le nombre d'années nécessaire à détecter une pente donnée pour un CV, un degré de signification et une puissance de test donnés.
- De déterminer le CV nécessaire pour atteindre un objectif fixé en termes de pente, de nombre d'années, de puissance et de degré de signification donnés.

En ce qui concerne l'influence du coefficient de variation qui est plus précisément l'objet de ce rapport, la grande tendance identifiée est la suivante : la puissance du test est une fonction décroissante du coefficient de variation de la série. En d'autres termes, lorsque la variabilité des données augmente au sein de la série, la puissance du test diminue et il est donc plus difficile de détecter l'existence d'une tendance ou encore, les variations au sein d'une série masquent l'existence d'une tendance.

La notion de variabilité du système (ou des données) évoquée dans le guide européen « sédiment et biote » est la notion principale qui ressort de la plupart des articles étudiés dans ce rapport. C'est la notion qui permet de façon prospective de concevoir un programme de surveillance destiné à identifier des tendances sur la base d'objectifs quantitatifs. C'est également la notion qui permet, de façon rétrospective, sur une chronique existante, de préciser la fiabilité de la tendance identifiée. L'incertitude analytique ou plus globalement l'incertitude sur la mesure incluant les étapes techniques d'échantillonnage et d'analyse est une part plus ou moins importante de la variabilité globale des données en fonction des contextes (paramètre, milieu, concentration, support, ...).

L'estimation de la variabilité des données apparaît donc comme un facteur décisif pour dimensionner un programme de surveillance pour l'évaluation des tendances. Cette

estimation est délicate. Certains auteurs utilisent des études préliminaires spécifiques pour cela ou bien se servent de données bibliographiques. D'autres auteurs déterminent la variance résiduelle issue de l'exploitation de chroniques réelles. Cette variabilité des données dépend des paramètres mesurés, des niveaux de concentration, du milieu, du support, du site étudié.

Outre l'étude bibliographique, des études de cas, issues de données simulées ou bien de l'exploitation de chroniques réelles sont présentées.

Ces données de simulation (chroniques avec tendance et incertitudes variables sur les données) permettent de disposer d'éléments objectifs intégrant la prise en compte de la variabilité des données pour dimensionner ou concevoir des programmes de surveillance ayant des objectifs quantitatifs de détection de tendance.

L'exploitation de chroniques réelles de surveillance (eau souterraine sur la période 1990-2007) semble aussi un outil très intéressant mais qu'il serait nécessaire d'approfondir afin d'estimer la variabilité des données en eau souterraine en fonction du paramètre considéré. Des exploitations de chroniques nitrates et pesticides montrent que la part de l'incertitude analytique dans la variabilité de la donnée est majoritaire (plus de 50%). Dans l'hypothèse où il serait nécessaire d'améliorer la puissance d'un programme de surveillance de ces paramètres en eaux souterraines en termes de détection de tendance, agir, si cela est possible techniquement et financièrement, sur l'incertitude analytique serait donc efficace et le bénéfice de cette action pourrait être quantifié.

Pour d'autres paramètres (Fe, As) la part de l'incertitude analytique dans la variabilité des données des chroniques considérées apparaît faible (moins de 20%). Ceci peut signifier que pour ces paramètres, une variabilité importante est liée aux protocoles d'échantillonnage ou bien que la variabilité des conditions du milieu est prépondérante. La part liée à l'échantillonnage est certainement non négligeable car ces paramètres sont particulièrement sensibles aux différences de pratiques (filtration ou pas, filtration sur le terrain ou au laboratoire, durée avant filtration au laboratoire, acidification avant ou après filtration, ...).

Cette méthodologie d'exploitation de chroniques réelles (basée sur une estimation de la variance résiduelle et une comparaison aux incertitudes analytiques) nécessiterait d'être validée sur le fond y compris au niveau statistique pour mieux prendre en compte les variations cycliques avant une exploitation qui pourrait être très large sur de nombreuses chroniques temporelles en eau souterraine. Ceci permettrait de disposer d'une base de données par paramètre indiquant la part relative de l'incertitude analytique dans la variabilité globale des données.

Les travaux menés au sein d'AQUAREF ont notamment pour objectif la maîtrise et la connaissance des incertitudes liées aux étapes d'analyse et d'échantillonnage. Ils contribuent ainsi à fournir des informations importantes pour l'évaluation de la variabilité des données de surveillance. Par ce biais ils pourront contribuer à évaluer la

Surveillance de polluants dans les eaux souterraines : impact de l'incertitude de mesure sur l'évaluation des tendances

fiabilité des tendances identifiées ou dimensionner des réseaux de surveillance destinés à évaluer les tendances d'évolution des concentrations de polluants.

Sommaire

1. Introduction	15
2. Objectifs.....	19
3. Programmes de surveillance et évaluation de tendances : exigences et recommandations européennes	21
3.1. EAU SOUTERRAINE	21
3.1.1. Directive fille eau souterraine 2006/118/CE.....	21
3.1.2. Guide européen « eau souterraine »	22
3.2. SEDIMENT ET BIOTE.....	22
3.2.1. Directive NQE	22
3.2.2. Guide européen sédiment et biote.....	23
4. Principaux tests statistiques d'identification des tendances	27
4.1. TEST DE MANN KENDALL.....	27
4.2. REGRESSION LINEAIRE	28
5. Etude bibliographique	31
5.1. TENDANCES ET ARTEFACTS ANALYTIQUES	31
5.1.1. "Uncertainty in water quality data and its implications for trend detection: lessons from Swedish environmental data", Wahlin, Grimvall [8].....	31
5.1.2. "Lessons from the Silica "Decline" in Lake Michigan", Shapiro, Swain, 1983 [10]	33
5.2. EXEMPLE D'IMPACT DE L'INCERTITUDE DE MESURE SUR L'EVALUATION DES TENDANCES AU SEIN DE RESEAUX DE SURVEILLANCE	35
5.2.1. "Uncertainty and Trend Analysis—Radium in Ground Water and Drinking Water", Soderberg and Hennet, 2007 [9].....	35
5.2.2. "Modelling the effects of varying data quality on trend detection in environmental monitoring", Sulkava, Luysaert, Rautio & al., 2007 [11] .	39
5.2.3. "Designing Monitoring programmes for detecting temporal trends in contaminants in fish and shellfish", Nicholson & al (1997) [14]	43
5.2.4. « Trend detection in water quality time series of LRTAP – QUEBEC network lakes", Bouchard & al, 1992 [10]	48
5.3. PERFORMANCES DE TESTS STATISTIQUES ET FACTEURS INFLUENTS	51

5.3.1. "Power of the Mann-Kendall and Spearman's rho tests for detecting monotonic trend in hydrological series", Yue, Pilon, Cavadias, 2002 [7]. 51

6. Simulations et exploitation de chroniques réelles	57
6.1. SIMULATIONS	57
6.1.1. Objectif en termes d'évaluation de tendance	57
6.2. DONNEES SIMULEES	58
6.2.1. Mode de simulation	58
6.2.2. Exploitation des données	59
6.2.3. Résultats et interprétations	60
6.3. EXPLOITATION DE CHRONIQUES REELLES	64
6.3.1. Données sur les Nitrates	64
6.3.2. Données sur les pesticides	69
6.3.3. Données sur d'autres paramètres	70
6.3.4. Synthèse	71
7. Pistes pour améliorer et mieux maîtriser les incertitudes de mesure	75
7.1. TRAÇABILITE METROLOGIQUE	75
7.2. INCERTITUDE DE MESURE	76
7.3. PISTES D'AMELIORATION	81
7.3.1. Analyse du processus de mesure	81
7.3.2. Identification des contributions les plus importantes et exemple de mise en place d'action possible pour diminuer ces contributions	83
7.4. CONCLUSION ET RECOMMANDATIONS	92
8. Conclusion et perspectives	95
9. Bibliographie	101

Liste des annexes

Annexe 1 Directives fille « eaux souterraines » et NQE : extraits des chapitres relatifs aux tendances	105
--	-----

Liste des illustrations

Figure 1 : Représentation des concentrations normalisées en Azote total ($\mu\text{g/l}$) de plusieurs rivières suédoises en fonction du temps [8] (azote total estimé soit par analyse directe et azote total estimé par la somme des différentes formes de l'azote).	32
Figure 2 : Représentation schématique simplifiée des concepts développés dans les références [9](approche CV/E) et [10](approche Tr/σ) – cf texte.	37
Figure 3: relation entre la variabilité analytique et la durée nécessaire pour détecter une tendance de pente 0.1 (taux de signification = 0.05 et puissance >50%)	40
Figure 4 : Diagrammes puissance en fonction de CV et N ($\alpha=0.05$ et pente 0.005) puis de CV et de la pente ($\alpha=0.05$ et $N=50$) [7].....	53
Figure 5 : moyenne des pentes estimées pour les simulations en fonction de l'incertitude et du nombre de données (les pentes théoriques sont identifiées par les traits rouges)	61
Figure 6 : Ecarts-types résiduels (en %) estimés sur 38 chroniques temporelles de suivi de nitrates en eau souterraine (cf texte)	68
Figure 7 : Ecarts-types résiduels (en %) estimés sur 12 chroniques temporelles de suivi de pesticides en eau souterraine en fonction de la concentration (cf texte).....	69
Figure 8 : répartition des rapports Tr/σ pour les chroniques nitrates en fonction du caractère significatif ou pas de la tendance identifiée par le test de Mann Kendall	72
Figure 9 : Exemple pour le nickel des résultats obtenus lors de l'essai d'intercomparaison réalisé dans le cadre du projet Euramet 924, Partie III.	79
Figure 10 : Distribution de l'incertitude élargie en quatre classes pour le cadmium, le nickel et le plomb, d'après l'étude Euramet 924, Partie III [25].....	80
Figure 11 : Diagramme des 5 M pour l'analyse du plomb par ICP-MS dans une eau d'alimentation [20]	81
Figure 12 : Budget d'incertitude pour l'analyse du plomb par ICP-MS dans une eau d'alimentation [20]	82
Figure 13 :Schéma de la détermination de la concentration en espèces chimiques dans les eaux	83
Figure 14a et b : Impact des pratiques d'échantillonnage sur les résultats d'analyse du zinc dans une eau souterraine – Essai d'intercomparaison AQUAREF 2009 (La Chapelle Saint Luc) (à gauche : matériel d'échantillonnage fourni aux équipes de prélèvement – à droite : matériel d'échantillonnage des participants) – En vert : moyenne des résultats de contrôles du site effectués par le BRGM avec l'incertitude analytique associée (en orange). Chaque point est le moyenne de 2 résultats obtenus par le participant issus de la réalisation de 2 échantillonnages	84
Figure 15 : Concentrations moyennes (en $\mu\text{g/L}$) et écart type pour le plomb sur la Deule à Don en fonction du nombre de jours entre le prélèvement et la filtration (Etude Aquaref en cours). Une distinction entre des flacons remplis à ras bord et des flacons remplis au $\frac{3}{4}$ a été faite	85
Figure 16 : Evolution des concentrations moyennes (en $\mu\text{g/L}$) en arsenic et en fer dans une eau souterraine en fonction du nombre de jours entre le prélèvement et la filtration	

(Etude Aquaref à paraître). Une distinction entre des flacons échantillonnés remplis à ras bord (RB) et des flacons remplis au $\frac{3}{4}$ (NRB) a été faite	86
Figure 17 : Variabilité des solutions commerciales pour les PFOS (tous composés confondus), d'après [22]	88
Figure 18 : Effet matriciel sur l'analyse instrumentale, l'étape de préparation et sur l'efficacité de la méthode pour 14 pesticides, d'après Krueve et al. [23].....	89
Figure 19 : Comparaison entre les étalonnages externes et internes en solvant ou en matrice pour le α -BHC et le 4,4'-DDT, d'après Sanchez-Brunete et al. [24].....	91

Liste des tableaux

Tableau 1: Concentrations estimées en métaux traces dans l'eau de mer (en $\mu\text{g/l}$)[4]	17
Tableau 2: nombre d'années nécessaires pour détecter une tendance significative de pente donnée en fonction de la variance analytique (d'après[13]).....	43
Tableau 3 : valeurs du rapport b/ψ en fonction de la puissance voulu du test de tendance et du nombre d'années visées pour le programme de surveillance (d'après [14]).....	44
Tableau 4 : valeurs à cibler pour la dispersion ψ des données en fonction de la puissance voulu du test de tendance et du nombre d'années visées pour le programme de surveillance (d'après [14])	45
Tableau 5: Nombre d'années nécessaire pour détecter une tendance de 5% par an en fonction du nombre d'individus analysés annuellement et de différentes hypothèses sur la variance des données. (d'après [14])	46
Tableau 6 : Nombre d'années nécessaires pour détecter une tendance de 5% par an en fonction de différentes hypothèses sur la variance des données. (d'après [14])	47
Tableau 7 : Pour quatre paramètres physico chimiques d'eau de lac, nombre d'années nécessaires pour détecter une tendance dans des conditions définies par le rapport Tr/σ en fonction du nombre données par an (d'après [10])	49
Tableau 8 : Pour quatre paramètres physico chimiques d'eau de lac, rapport Tr/σ détectable en fonction du nombre d'années de surveillance et du nombre de données par an (d'après [10])	50
Tableau 9: Pourcentage de simulations pour lesquelles une tendance significative est identifiée en fonction de l'incertitude sur la donnée et du nombre de données par an (2 ou 4).....	60
Tableau 10: Pourcentage de simulations pour lesquelles une tendance significative est identifiée en fonction de l'incertitude sur la donnée et du nombre de données par an.....	62
Tableau 11 : Ecart-type moyen des données et écart-type de la pente en fonction du nombre de données.....	63
Tableau 12 : Probabilité de détecter une tendance en fonction de l'incertitude sur la donnée et du nombre de données par an	63
Tableau 13 : tableau récapitulatif des caractéristiques des chroniques nitrates utilisées.....	66

Tableau 14 : Ecart-types résiduels en % estimés sur des chroniques temporelles de suivi de différents paramètres en eau souterraine.....	70
Tableau 15 : Synthèse de quelques incertitudes types estimées dans des eaux par 2 laboratoires prestataires (A et B), un laboratoire de référence (LNE) et par différents laboratoires participant au programme EURAMET 924 pour 3 familles de paramètres et pour différentes méthodes d'analyse.	77
Tableau 16 : Synthèse des résultats obtenus lors de la comparaison interlaboratoire Euramet 924 Partie III (\bar{x} : concentration moyenne ; N : nombre de laboratoires ; s / \bar{x} : coefficient de variation ; U : incertitude élargie estimée à partir de l'équation $U = k s / \sqrt{N}$ avec $k = 2$)	78

1. Introduction

La directive cadre européenne sur l'eau (DCE) [1] qui a conduit à la mise en place de programmes de surveillance dans l'ensemble des pays de la communauté européenne exige des états membres des exploitations diverses de ces programmes afin de garantir le bon état chimique et écologique des masses d'eau.

Deux directives filles décrivent les modes d'évaluation des données issues de ces programmes de surveillance

La directive 2008/105/CE du 16 décembre 2008 [2] établit des normes de qualité environnementale (NQE) pour les eaux de surface. Les données issues des programmes de surveillance doivent être comparées aux NQE pour 41 substances prioritaires afin de définir l'état chimique de la masse d'eau. En plus de cette comparaison aux NQE, la directive demande aux états membres une exploitation en tendance des données dans les supports intégrateurs que sont les sédiments et le biote pour un certain nombre de substances hydrophobes. Il s'agit de montrer qu'il n'existe dans les masses d'eau aucune tendance significative à la dégradation du milieu.

En ce qui concerne les eaux souterraines, la directive 2000/60/CE fixe des dispositions générales pour leur protection et leur conservation. Dans son article 17 la directive demande aux états membres d'adopter des mesures de prévention et de contrôle de la pollution des eaux souterraines, notamment des critères pour l'évaluation du bon état chimique (comparaison des concentrations à des normes de qualité) mais aussi pour l'identification des tendances significatives et durables à la hausse des concentrations. Elle demande aussi que les états membres prennent des mesures afin d'établir l'inversion des tendances à la hausse et qu'ils définissent les points de départ de ces inversions. La directive fille 2006/118/CE du 12 décembre 2006 [3] sur la protection des eaux souterraines décrit de façon plus précise ces exigences générales.

La DCE exige donc des états membres des exploitations des données de concentration :

- Instantanées issues de la comparaison entre une concentration et une valeur seuil
- Sur la durée pour identifier d'éventuelles tendances à la hausse de la concentration de polluants.

Jusqu'à présent, les discussions au niveau national se sont portées plus sur les exigences de comparaison des concentrations de polluants avec les normes de qualité. De plus en plus, l'intérêt se porte sur l'évaluation des tendances et sur les méthodologies à mettre en place pour répondre à ces exigences.

Par ailleurs, outre les exigences strictes de la directive sur les évaluations en tendance, toute politique environnementale ayant pour objectif de diminuer la pollution des milieux après la mise en place de mesures de gestion, il paraît indispensable de

disposer d'outils permettant de vérifier la pertinence de ces mesures et donc d'outils permettant de vérifier de façon fiable que les concentrations des polluants concernés diminuent.

De nombreux outils statistiques existent et sont décrits dans la littérature pour identifier des tendances au sein de chroniques temporelles. Cependant, l'utilisation de ces outils et leurs conditions de validité même si elles sont largement répertoriées restent complexes. De très nombreux paramètres influencent la possibilité d'utiliser tel ou tel outil ainsi que la fiabilité des résultats. On peut citer par exemple :

- La loi de distribution des données
- Les fréquences d'acquisition
- La présence de données manquantes, de données aberrantes, de données inférieures aux limites de quantification,
- Le nombre total de données
- ...

De façon générale, l'exercice d'identification des tendances est un exercice difficile qu'il s'agisse du choix des outils ou de la conception des programmes de surveillance qui permettront d'identifier dans l'avenir de façon fiable des tendances significatives au regard des objectifs fixés.

Parmi les facteurs susceptibles d'influencer la capacité à identifier de façon fiable des tendances, la qualité des données analytiques a peu souvent été étudiée. Peu de références existent concernant la prise en compte des incertitudes analytiques ou de façon plus large concernant la variabilité de la donnée.

L'exemple présenté ci après permet de mettre en relief l'impact que peut avoir la qualité des données dans l'interprétation de chroniques temporelles. Cet exemple, avec du recul, peut sembler caricatural compte tenu de l'ampleur des variations observées. Cependant, il semble tout à fait possible que des exemples similaires pour d'autres substances puissent se produire à l'heure actuelle ou dans le futur.

Ainsi le Tableau 1 présente l'évolution des concentrations en quelques métaux traces (en $\mu\text{g/l}$) dans les eaux de la mer du Nord, telles que déterminées par des programmes de surveillance sur une période de vingt ans (d'après [4]). Il est apparu, au début des années 1980, que la diminution brutale des concentrations observées n'était en rien due à une décroissance de la pollution mais plutôt que la majorité des analyses réalisées avant 1980 étaient entachées d'erreurs liées à une contamination des échantillons lors de leur collecte et de leur analyse (ie, source de contamination essentiellement liée aux échantillonneurs et à l'absence de précaution lors de la manipulation des échantillons).

Métal	1965	1975	1983
Pb	0.03	0.03	0.002
Hg	0.03	0.03	0.001
Ni	2	1.7	0.46
Cu	3	0.5	0.25
Zn	10	4.9	0.39

Tableau 1: Concentrations estimées en métaux traces dans l'eau de mer (en µg/l)[4]

N'importe quel outil statistique identifiera sur cet exemple des tendances à la baisse des concentrations de façon très significative. L'intérêt de s'interroger sur la fiabilité de la donnée avant de conclure à la présence d'une tendance de diminution des concentrations apparaît ici évident.

Ce rapport est rédigé en collaboration entre le BRGM et le LNE dans le cadre du programme d'action d'AQUAREF pour l'année 2010. Il est rédigé également dans le cadre des conventions de partenariat ONEMA-BRGM et ONEMA-LNE pour l'année 2010.

Les auteurs du rapport remercient B. LOPEZ et B. BOURGINE (BRGM) pour leur appui et conseils dans la réalisation de ce rapport. B LOPEZ est par ailleurs l'auteur d'un rapport sur les méthodes statistiques d'estimation des tendances et sur des recommandations pour leur utilisation dans le cadre de la DCE. ([12]).

2. Objectifs

Entrant dans le cadre des actions d'AQUAREF, ce rapport a pour objectif de s'intéresser à la qualité des données dans un contexte d'évaluation des tendances. Il s'agit plus particulièrement d'étudier l'impact de la qualité des données de surveillance sur les possibilités d'identification des tendances. L'objectif n'est pas de réaliser une étude approfondie des outils statistiques disponibles, cette synthèse étant réalisée par ailleurs [12]. L'objectif de la présente étude est de contribuer à la mise en place d'une méthodologie nationale pour l'identification des tendances en apportant des informations et des éléments de réflexion sur l'importance de la qualité et la fiabilité des données d'analyse.

Traduit de façon concrète en termes d'incertitude sur les résultats de mesure analytiques, l'objectif de ce rapport est d'essayer de répondre à la question suivante : de quelle qualité de donnée a-t-on besoin pour identifier une tendance de façon fiable en un nombre d'années déterminé: une incertitude de 5% ? 50% ? 100 % ? Peu importe l'incertitude ? Ou bien le niveau d'incertitude peut-il être variable en fonction d'autres paramètres du milieu et des objectifs fixés ?

Le rapport est principalement orienté vers les eaux souterraines. Les exemples seront pris dans ce contexte. Toutefois, les problématiques et les conclusions seront pour la plupart adaptables à d'autres milieux ou supports.

Le premier chapitre du rapport rappelle les exigences de la DCE en termes d'évaluation des tendances ainsi que les éléments techniques qui figurent dans les guides européens rédigés en appui à la mise en place des directives.

Le second chapitre est une synthèse bibliographique sur les études ayant pris en compte sous différents angles la qualité des données d'analyse pour évaluer la fiabilité d'identification des tendances.

Le troisième chapitre présentera quelques exemples de simulations et quelques analyses de chroniques temporelles réelles afin de montrer les implications des niveaux d'incertitudes analytiques sur l'identification des tendances.

Le quatrième chapitre s'intéressera aux solutions techniques ou méthodologiques qui permettraient d'améliorer la prise en compte la qualité des données dans les programmes de surveillance ayant des objectifs en termes d'identification des tendances.

Enfin, les principaux résultats seront synthétisés dans la conclusion.

3. Programmes de surveillance et évaluation de tendances : exigences et recommandations européennes

Ce chapitre rappelle les exigences issues des directives filles eaux souterraines [3] et NQE [2] relatives à l'identification des tendances. Il s'inspire aussi du guide européen qui contient des recommandations sur la mise en place d'une surveillance des eaux souterraines [6]. Le rapport est orienté vers les eaux souterraines cependant les quelques éléments disponibles dans le guide européen sur la surveillance biote et sédiments [5] apportent des informations complémentaires intéressantes.

3.1. EAU SOUTERRAINE

3.1.1. Directive fille eau souterraine 2006/118/CE

Les chapitres de la directive fille eau souterraine en lien avec les exigences d'évaluation de tendances sont rappelés en Annexe 1 dans leur intégralité. A la fin du paragraphe, les principaux éléments ayant un lien direct avec le présent rapport seront mis en relief. En premier lieu, la directive fille sur les eaux souterraines définit le terme suivant :

« Tendance significative et durable à la hausse : toute augmentation significative, sur les plans statistique et environnemental, de la concentration d'un polluant, d'un groupe de polluants ou d'un indicateur de pollution dans les eaux souterraines, pour lequel une inversion de tendance est considérée comme nécessaire conformément à l'article 5. »

La directive insiste sur l'importance de déceler des tendances pas uniquement significatives sur le plan statistique mais également sur le plan environnemental. La seule identification de tendance sur un plan statistique peut dans certains cas ne pas être suffisante et une expertise environnementale ou chimique/analytique paraît indispensable. Des exemples issus de la bibliographie confirmeront ce point (cf §5).

Selon la directive, la détection des tendances doit se faire en premier lieu avec des outils statistiques et en se basant sur des informations permettant de différencier les tendances des variations naturelles avec un niveau de confiance et de précision suffisants. De façon générale, ces niveaux de confiance et précision suffisants doivent être définis pour toute exploitation statistique. C'est la notion d'objectifs quantitatifs préalablement fixés qui paraît importante. Enfin, et en lien avec cette notion d'objectifs, la directive rappelle que les programmes de surveillance doivent être conçus afin de pouvoir répondre aux objectifs qui auront été définis au préalable.

3.1.2. Guide européen « eau souterraine »

Dans le cadre des activités du groupe d'experts européens WG C, un guide a été élaboré sur la détermination de l'état des masses d'eau souterraine et sur l'évaluation des tendances [6].

Ce guide rappelle tout d'abord les grands principes de la directive et insiste sur la nécessité de développer des modèles conceptuels généraux permettant une compréhension claire des conditions requises pour l'atteinte des objectifs environnementaux. Ces modèles se basent notamment sur une bonne connaissance des conditions de débit, de transport et des conditions hydrogéochimiques des milieux.

En ce qui concerne l'évaluation des tendances, les modèles conceptuels doivent permettre de choisir les sites de surveillance ainsi que les fréquences permettant de différencier avec un niveau acceptable de précision les tendances observées des variations naturelles. Ceci est repris dans le paragraphe 6.2.2 du rapport concernant les objectifs à prendre en compte pour la conception des programmes de surveillance (choix des points et fréquence). Un autre objectif cité dans ce paragraphe est la possibilité de détecter une tendance à la hausse en un temps suffisamment court pour que des mesures environnementales puissent être prises.

Le rapport donne également quelques indications sur la durée des chroniques à prendre en compte. Il indique que le choix de cette durée est lié :

- A la façon dont la masse d'eau réagit aux changements de pratiques de surface
- A la puissance du test de tendance utilisé
- Et également à la qualité des données.

En ce qui concerne les méthodes d'évaluation des tendances, le rapport indique que ces évaluations doivent se référer à des méthodes statistiques reconnues comme par exemple la régression linéaire. Le rapport définit également la puissance d'une méthode d'estimation de tendance comme étant la capacité à détecter une augmentation donnée de concentration d'un polluant avec une probabilité donnée.

De façon générale peu d'informations opérationnelles sont données dans ce guide concernant les tendances (méthodes de calcul,..) . Des informations supplémentaires sont données dans le cas de la surveillance biote et sédiments (cf. § 3.2.2).

3.2. SEDIMENT ET BIOTE

3.2.1. Directive NQE

La directive 2008/105/CE [2], dite directive NQE définit des normes de qualité environnementale pour l'état chimique des eaux de surface, et impose aux états

membres dans son article 3 paragraphe 3 quelques éléments relatifs aux évaluations de tendance (Cf Annexe 1).

La directive est peu développée sur la notion de tendance. Elle donne les grands principes qui, sur le plan méthodologique, imposent de définir la notion de tendance significative et/ou fiable. Par ailleurs, les programmes de surveillance doivent être conçus notamment pour la définition de la fréquence des contrôles afin de fournir des données suffisantes pour évaluer ces tendances de façon fiable. Cette notion de conception de programmes adaptés pour l'évaluation de tendances implique la définition d'objectifs précis et quantitatifs.

3.2.2. Guide européen sédiment et biote

Le guide européen sur la surveillance des sédiments et du biote dans le cadre de la DCE, intitulé «Guidance on Chemical Monitoring of sediment and biota under the Water Framework Directive » [5] aborde la question de l'évaluation des tendances dans son chapitre 4. Les notions abordées sont générales mais importantes car elles permettent de structurer la réflexion pour la mise en place de méthodologie sur ce thème. **Elles sont par ailleurs tout à fait applicables à d'autres types de support (matrice eau notamment).**

Dans ce chapitre du guide, la notion d'incertitude est définie de façon générale. Appliquée au contexte de la surveillance des sédiments et du biote, il est rappelé que l'acte de prélever un échantillon dans l'environnement introduit une incertitude dans le résultat de mesure. Les protocoles d'échantillonnage ne sont jamais parfaits car très souvent ils ne peuvent prendre en compte tous les événements susceptibles d'intervenir au moment de l'échantillonnage.

Pour la surveillance du biote et des sédiments, les principales sources d'incertitude sont les variabilités spatiales et temporelles sur le site d'échantillonnage ainsi que la variabilité issue du processus de mesure incluant le prélèvement, les étapes de prétraitement et de stockage ainsi que la mesure des paramètres en elle-même. Le guide indique que, en général, la variabilité naturelle (hétérogénéité spatiale, variabilité biologique, ...) ainsi que l'acte d'échantillonnage sont certainement les facteurs les plus importants d'incertitude et les facteurs les moins contrôlables.

Tandis que l'échantillonnage et la mesure peuvent être contrôlés jusqu'à un certain niveau et leurs variabilités estimées, l'influence de la variabilité naturelle ne peut être évaluée qu'à travers l'acquisition de données sur le site considéré pendant les phases de conception du programme de surveillance. Plus la masse d'eau étudiée sera complexe ou hétérogène, plus grand devra être l'effort d'investigation pour caractériser le site.

Le guide rappelle que la définition claire des objectifs du programme de surveillance est d'une importance capitale car ce sont ces objectifs qui permettront de définir de façon adaptée les sites à surveiller, les fréquences, les durées, les méthodologies d'échantillonnage et d'analyse. Ces objectifs doivent également être connus durant les

phases préliminaires d'investigation sur le site qui sont capitales pour la conception définitive du programme de surveillance.

Le guide aborde ensuite quelques considérations statistiques sur la surveillance.

Bien que sédiments et biote soient moins influencés par de rapides changements comme le support "eau", ils sont sujets à des variations aléatoires, systématiques ou saisonnières. Dans le cas de variations aléatoires, un nombre plus élevé de répétitions permettra de fiabiliser la donnée. Dans le cas de variations systématiques (cycliques) par exemple, le choix de la période d'échantillonnage sera essentiel. Le type de variabilité observée sur un site aura des conséquences directes sur la stratégie d'échantillonnage à mettre en œuvre.

Comme indiqué précédemment, une description précise des objectifs de la surveillance est essentielle pour définir les paramètres du programme de surveillance en termes de fréquence, durée, densité d'échantillonnage, ...

Ces objectifs doivent intégrer entre autres :

- Le choix de la matrice échantillonnée et une définition stricte de l'unité élémentaire échantillonnée et de ce qu'elle représente dans le temps et dans l'espace.
- La définition de la sensibilité choisie du programme de surveillance c'est-à-dire la plus petite variation temporelle à détecter.
- La définition de la puissance statistique ciblée pour détecter cette variation à un niveau de signification donné.

L'efficacité du programme de surveillance diminue quand les sources de variation augmentent (méthodes d'analyse, protocoles d'échantillonnage, facteur environnementaux, ...). En conséquence, afin de calculer par exemple le nombre d'échantillons et la fréquence d'échantillonnage requis pour atteindre les objectifs du programme, une estimation de la variance d'échantillonnage est nécessaire. Cette variance peut être estimée grâce à des programmes similaires ou bien, grâce à un projet pilote sur le site considéré.

L'efficacité nécessaire ou possible d'un programme de surveillance varie avec la substance, la matrice, le site considéré. Il n'est donc pas possible de donner des valeurs fixes valables pour toutes les situations. Il est de la responsabilité du concepteur du programme de surveillance de spécifier la sensibilité et l'efficacité attendue pour le programme et pour ceux qui doivent le mettre en œuvre, ce qui sera possible de faire dans les faits.

A titre d'exemple pour des études temporelles, un objectif peut se décliner de la façon suivante :

- Détecter une diminution de 50% sur une période de 10 ans avec une puissance statistique de 80% et pour un niveau de signification de 5% (50% sur 10 ans représente une diminution annuelle de 7%)

Un niveau de signification de 5% signifie que l'on est prêt à accepter un risque de 5% de conclure des données qu'il y a une tendance alors qu'il n'y en a pas (il s'agit d'un des paramètres du test statistique qui sera utilisé).

Une puissance de 80% signifie que, en construisant ce programme de surveillance on souhaite avoir 80% de chance de conclure sur une diminution de concentration de 50% en 10 ans si celle-ci se produit.

Dans le cas des études temporelles, si aucune tendance n'est identifiée, il est essentiel de savoir si cela reflète une situation stable ou bien si la stratégie d'échantillonnage est trop pauvre pour détecter même des changements majeurs.

L'exploitation statistique de chroniques temporelles ainsi que la détection de tendances par ces outils statistiques doivent toujours être complétées par l'intervention d'experts capables d'une évaluation plus précise des résultats d'analyse chimique.

En résumé, ce guide européen traite de l'exigence d'identification des tendances non pas à travers la présentation d'outils statistiques mais en présentant les grands principes méthodologiques à appliquer pour la conception de programmes de surveillance adaptés à l'identification de tendances. Parmi ces grands principes, on peut retenir :

- **La nécessité de rédiger en termes quantitatifs les objectifs du programme de surveillance pour l'évaluation des tendances en termes notamment de sensibilité et d'efficacité du programme.**
- **La nécessité de disposer d'informations concernant la variabilité du système étudié afin de mettre en place de façon concrète le programme de surveillance sur la base des objectifs fixés.**
- **La variabilité de données issues d'un programme de surveillance environnementale est liée notamment à :**
 - **Une variabilité naturelle décomposable en des variations aléatoires et des variations systématiques (cycliques par exemple).**
 - **Une variabilité liée à la mesure qui inclut les effets liés aux pratiques d'échantillonnage, à la préparation de l'échantillon, son stockage, et à l'analyse.**

- **La variabilité des données est liée à la substance, la matrice, au site étudié.**
- **L'efficacité d'un programme est d'autant plus faible que la variabilité des données est forte.**
- **L'intervention d'experts (environnement, analyse) est indispensable à la confirmation des tendances observées par les outils statistiques.**

L'incertitude analytique ou plus largement l'incertitude sur la mesure incluant la part liée à l'échantillonnage contribue donc à définir l'efficacité du programme de surveillance. A cette incertitude liée à la mesure s'ajoute cependant l'incertitude liée à l'influence des divers processus environnementaux.

4. Principaux tests statistiques d'identification des tendances

Les deux principaux tests statistiques utilisés pour la détection de tendances sont :

- Le test de Mann Kendall
- La régression linéaire

La référence [12] décrit la plupart des tests utilisés pour évaluer des tendances et en donnent les avantages et les limites. Dans les paragraphes qui suivent, les principales notions permettant de mieux comprendre le chapitre 5 sont décrites pour les tests de Mann Kendall et de régression linéaire. Il n'est pas dans l'objectif de ce rapport de décrire les avantages et inconvénients de ces tests. Il s'agit juste ici de donner quelques concepts de base.

De façon schématique le test de Mann Kendall est un test non paramétrique qui ne présuppose pas de modèle de distribution pour les données. Le test de régression linéaire impose au contraire pour son application une hypothèse de distribution normale des données.

4.1. TEST DE MANN KENDALL

Le test de Mann-Kendall (voir par exemple [7]), également appelé test « Kendall's tau », est un test non paramétrique basé sur le rang, qui permet d'évaluer l'existence d'une tendance monotone significative ou non. L'hypothèse nulle H_0 représente le cas d'un échantillon dont les données sont indépendantes, sans corrélation. L'hypothèse alternative H_1 correspond au cas où les données de l'échantillon présentent une tendance monotone.

De façon schématique, le test consiste à étudier 2 à 2 tous les couples de 2 données consécutives de la série temporelle. Le signe de la différence entre ces deux données est utilisé afin de déterminer globalement si une tendance existe. La statistique S du test de Mann-Kendall se définit comme suit:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(X_j - X_i)$$

Avec n le nombre de données de la série et X_j et X_i les valeurs des données séquentielles i et j et

$$\text{Sgn}(t)=1 \text{ si } t>0$$

$$\text{Sgn}(t)=0 \text{ si } t=0$$

$$\text{Et } \text{sgn}(t)=-1 \text{ si } t<0$$

Ce test est particulièrement utile puisqu'il permet d'étudier des séries incomplètes et que les données n'ont pas besoin de se conformer à une distribution statistique, par exemple de type normal (test non paramétrique). Pour $n>10$, la statistique S est reconnue être approximativement normalement distribuée, avec une moyenne et une variance

$$E(S)=0$$

$$\text{Var}(S) = \frac{n(n-1)(2n+5) - \sum_{p=1}^g t_p(t_p-1)(2t_p+5)}{18}$$

où g est le nombre de groupes dont les valeurs sont égales (tied groups) et t_p est le nombre de données dans le p ième groupe (par exemple dans la séquence {23, 24, 6, 6, 2,24, 6}, $g=2$, $t_1=2$ pour la valeur 24 et $t_2=3$ pour la valeur 6).

La variable Z est calculée à partir de S et de $\text{VAR}(S)$:

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

La variable Z est normalement distribuée. Une valeur positive de Z indique une tendance croissante, alors qu'une valeur négative de Z indique inversement une tendance décroissante. L'hypothèse nulle H_0 (il n'existe aucune tendance) est rejetée avec un niveau de signification de 0.05 si $Z > 1.96$ en valeur absolue.

4.2. REGRESSION LINEAIRE

La première hypothèse faite pour appliquer ce test est que les données suivent une loi normale.

La tendance linéaire recherchée dans les données est modélisée par une droite :

$$y = a_0 + b_0 \cdot x$$

où a_0 représente la concentration initiale

x le temps

b_0 la pente de la tendance linéaire, c'est-à-dire l'augmentation de la concentration par unité de temps.

La valeur de la pente dépend de l'unité de temps utilisée. Dans cette étude, la période d'une année est représentée par un pas de 1.

On estime trois paramètres :

- la constante : valeur a et incertitude-type $u(a)$
- la pente : valeur b et incertitude-type $u(b)$
- l'incertitude-type des mesures : $u(y)$

Cette incertitude $u(y)$ intervient dans le calcul des incertitudes des coefficients.

La recherche d'une tendance consiste à tester l'hypothèse que la pente théorique est nulle contre l'hypothèse qu'elle ne l'est pas :

$$H_0 : b_0 = 0$$

$$H_1 : b_0 \neq 0 \text{ (il y a une tendance positive ou négative).}$$

Lorsqu'il n'y a pas de tendance (H_0 vraie) la pente estimée b est proche de 0. Dans le cas contraire, elle est élevée en valeur absolue.

Il s'agit d'un test bilatéral de Student dans lequel le paramètre du test est le rapport de la pente estimée b à son écart-type :

$$t_{\text{observé}} = |b| / u(b)$$

La pente b_0 est déclarée significativement différente de 0 lorsque $t_{\text{observé}}$ excède une valeur seuil lue dans la table de Student (par exemple, $t_{\text{critique}} = 2,201$ à 11 degrés de liberté et pour un niveau de signification de 95%).

5. Etude bibliographique

Une étude bibliographique a été réalisée avec comme objectif de rechercher les références abordant à la fois l'identification de tendance et également l'impact de la qualité des données sur la fiabilité de l'évaluation de ces tendances. Cette recherche a été faite de façon large incluant différents types d'eau mais également des références issues d'autres domaines scientifiques.

De nombreuses références existent sur le traitement statistique de chroniques temporelles mais très peu s'intéressent à la qualité des données et notamment à leur incertitude. En raison du faible nombre de références, il est difficile de réaliser une synthèse identifiant des positions communes aux différents articles. Les articles sont cependant regroupés par grands thèmes :

- Tendances et artéfact analytiques
- Exemple d'impact de l'incertitude de mesure sur l'évaluation des tendances au sein de réseaux de surveillance
- Performances de tests statistiques et facteurs influents

Les références sont successivement décrites dans chaque thème. Les principaux points marquants de chaque article sont résumés.

Quelques articles sont issus d'autres domaines que le domaine de la surveillance des eaux. Les concepts utilisés dans ces articles sont tout à fait adaptables à la surveillance des eaux souterraines et aux autres types de surveillance DCE.

5.1. TENDANCES ET ARTEFACTS ANALYTIQUES

5.1.1. “Uncertainty in water quality data and its implications for trend detection: lessons from Swedish environmental data”, Wahlin, Grimvall [8]

Les tendances statistiques significatives observées (dédites des données de surveillance) sont elles affectées par des erreurs liées à l'échantillonnage, à la manipulation des échantillons ou aux analyses de laboratoire ?

Pour répondre à cette question, Wahlin et Grimvall [8] ont étudié les données de surveillance d'eaux de surface de 34 rivières suédoises réparties sur l'ensemble du pays, pour les composés azotés et phosphorés de 1980 à 2005.

Le même laboratoire est responsable de la surveillance (analyse et échantillonnage) depuis le début du programme. Le laboratoire a une longue pratique d'assurance qualité.

Pour chaque paramètre, les auteurs étudient de façon simultanée les résultats obtenus par régression linéaire sur l'ensemble des sites. Pour ce faire, les données utilisées sont la plupart du temps des concentrations moyennées sur plusieurs ou la totalité des sites pour mieux détecter les tendances.

Ainsi, il est observé pour l'ensemble des sites depuis le milieu des années 1990 une tendance nette à la baisse des concentrations d'azote total obtenues par une méthode de digestion au persulfate. A l'inverse, aucune tendance n'est observée pour l'azote total obtenu par somme de l'azote Kjeldhal, de l'azote NO₂ et NO₃. Les explications possibles sont :

- Un changement de nature de l'azote présent sous forme d'azote organique (plus grande résistance à l'attaque persulfate) – Cette hypothèse est exclue car il est très peu probable que ce changement soit observé de façon aussi générale sur l'ensemble des 34 sites.
- Une dérive liée à la méthode d'analyse

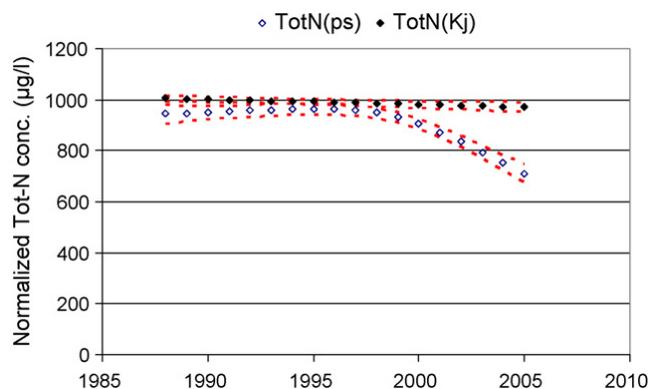


Figure 1 : Représentation des concentrations normalisées en Azote total ($\mu\text{g/l}$) de plusieurs rivières suédoises en fonction du temps [8] (azote total estimé soit par analyse directe et azote total estimé par la somme des différentes formes de l'azote).

Après enquête auprès du laboratoire, il s'est avéré qu'à la même époque (milieu des années 1990) un réactif persulfate de moins bonne qualité a été utilisé. Les auteurs concluent que ce changement de pratique est responsable de la tendance observée sur l'azote total. On peut toutefois remarquer qu'il aurait été intéressant de rechercher plus précisément les causes de l'évolution lente et continue, un changement de réactif engendrant plus probablement une rupture nette dans un suivi analytique.

Les auteurs concluent de leur étude que des outils statistiques puissants peuvent déceler des tendances permettant ainsi d'obtenir de nouvelles interprétations sur l'impact anthropique. Cependant, il peut s'avérer que la plupart des tendances ou effets de long terme concernant les paramètres étudiés (nutriments) soit liée à des artefacts analytiques. Une étude rapide des données avant leur bancarisation aurait permis de détecter la plupart des problèmes. Pour cela, le développement d'outils permettant « en temps réel » d'identifier des valeurs aberrantes dans une chronique serait utile. Par ailleurs, l'analyse simultanée de nombreux sites pour un même paramètre est un outil très efficace pour détecter des tendances réelles ou bien des artefacts analytiques qui n'auraient pas été révélés par l'étude des différentes chroniques prises séparément.

Quelques points importants

La prise en compte de facteurs liés à l'analyse dans la recherche d'explications de tendances observées. Dans les exemples explicités par les auteurs il s'agit de l'explication principale (mais les interprétations sont le plus souvent qualitatives et pas toujours argumentées. Il s'agit souvent d'identification par défaut après élimination des autres causes possibles. Les effets identifiés dans les chroniques étant souvent ponctuels pour une année, l'identification d'un problème analytique ponctuel est cependant possible).

L'utilité d'exploitations groupées de chroniques de surveillance sur plusieurs sites afin de déceler des biais (analytiques par exemple) est démontrée. L'exploitation site par site étant beaucoup moins puissante pour cela. L'exploitation globalisée (dans le temps et dans l'espace) de nombreuses données de surveillance permet de tirer des enseignements importants sur les biais analytiques.

Il ressort également que la traçabilité documentaire des données a permis, de nombreuses années après l'acquisition de la donnée de proposer une explication analytique au problème observé.

5.1.2. “Lessons from the Silica “Decline” in Lake Michigan”, Shapiro, Swain, 1983 [10]

Les auteurs partent du constat effectué en 1967 que la concentration en silice du lac Michigan diminuait. Plusieurs interprétations environnementales et géochimiques ont depuis été données à ce phénomène. Cette observation a été reprise dans de nombreuses publications et ouvrages scientifiques.

En reprenant les données, les auteurs confirment qu'à la première exploitation, la tendance semble indiscutable et que les tests statistiques (régression linéaire)

confirment de façon formelle l'hypothèse (diminution de 4 mg/l en 44 ans soit environ 60%)

Cependant, les auteurs remarquent par exemple qu'en éliminant uniquement les 2 premières années de la série (1926 et 1927), les tendances observées ne sont plus significatives et que par ailleurs, sur la période 1948-1962, aucune tendance n'était identifiée.

L'exploitation de métadonnées comme le nom du laboratoire ayant effectué les analyses révèle que des ruptures semblent liées au changement de laboratoire. D'autres ruptures sont également identifiées au sein d'une même période d'intervention d'un laboratoire.

La conclusion générale de l'article confirme que les précédentes interprétations de tendance étaient erronées car liées à des changements de pratiques analytiques.

Les auteurs insistent sur le fait que les données utilisées n'avaient pas initialement pour objectif de détecter des tendances mais étaient acquises à des fins de vérification de la conformité à des normes de qualité. Il n'y avait notamment à l'époque que peu de soin mis au contrôle de ces données, notamment en terme des calibrations croisées, d'évaluation d'incertitude notamment.

Ils concluent sur l'obligation (souvent non exprimée) faite aux organismes en charge de l'acquisition régulière sur de longues durées de données environnementales, de veiller à la maîtrise de la mesure. Ceci peut par exemple impliquer que les méthodes d'analyse ne devraient pas changer sans raison et également que tout changement devrait être l'objet d'une validation par rapport aux données antérieures.

Quelques points importants

- Exemple de tendance identifiée, publiée, exploitée et apparemment erronée du fait d'un biais analytique.
- Importance de « cibler » les données dont une exploitation en tendance est envisagée (incluant une maîtrise supérieure de la donnée : contrôle de cohérence notamment lors de changement de méthodes ou de laboratoire).
- Importance également des métadonnées et de la traçabilité documentaire des données, permettant la recherche et l'identification de causes.

5.2. EXEMPLE D'IMPACT DE L'INCERTITUDE DE MESURE SUR L'EVALUATION DES TENDANCES AU SEIN DE RESEAUX DE SURVEILLANCE

5.2.1. "Uncertainty and Trend Analysis—Radium in Ground Water and Drinking Water", Soderberg and Hennet, 2007 [9]

Les auteurs insistent largement sur le fait que la capacité à détecter une tendance dépend de la qualité des données et pas seulement du nombre de points et de l'importance de la tendance. Or, les tests classiquement utilisés n'intègrent pas de façon simple la prise en compte des incertitudes de mesure. L'incertitude n'est pas prise en compte dans l'analyse du caractère significatif ou pas de la tendance statistique.

Afin de caractériser l'influence des incertitudes de ces données sur l'évaluation des tendances, les auteurs se sont intéressés aux données de surveillance du radium dans les eaux souterraines et les eaux de consommation. Les données proviennent de 137 sites et ont été obtenues entre 1992 et 2003. Les analyses ont été effectuées dans 8 laboratoires différents.

Pour cela, ils appliquent dans un premier temps le test non paramétrique Mann Kendall sur les différentes chroniques dont ils disposent. Dans un second temps, ils mettent en place un « test de sensibilité ». Ce test est défini de la façon suivante : pour chaque site, 1000 nouvelles chroniques sont simulées. La simulation consiste à remplacer chaque résultat par un résultat simulé de façon aléatoire dans un intervalle de largeur $\pm U$ (incertitude élargie) autour du résultat. La loi de probabilité choisie est une loi rectangle. Ce point nous apparaît comme discutable dans le raisonnement choisi car il suppose que le résultat vrai de l'analyse a autant de chance de se trouver aux bornes de l'intervalle qu'au centre de celui-ci. Généralement, des distributions normales sont présupposées. Cette hypothèse maximise l'influence de l'incertitude. Les résultats inférieurs à la limite de détection sont remplacés par LD/2. L'incertitude quand elle n'est pas donnée par le laboratoire est estimée à partir d'un modèle établi avec l'ensemble des incertitudes connues en fonction de la concentration.

Sur chacune des 1000 chroniques simulées, le test de Mann Kendall est appliqué. Les auteurs définissent un « taux de détection » comme le pourcentage de chroniques simulées pour lesquelles une tendance significative est obtenue avec le test MK. Un « taux de détection » de 90% est considéré comme significatif pour cette exploitation des chroniques simulées.

Sur l'ensemble des chroniques réelles étudiées, le test de Mann Kendall identifie 10 sites avec une tendance significative (niveau de confiance de 90% traduit par la paramètre statistique p value qui doit être inférieur à 1). Sur ces 10 chroniques, seulement 2 donnent un taux de détection supérieur à 90% (100%). Les autres taux de détection sont compris en fonction des sites entre 20 et 77%. Autrement dit, pour 8 des chroniques pour lesquelles le test de Mann Kendall a détecté une tendance significative, moins de 90% des simulations réalisées prenant un compte l'incertitude de mesures, identifient des tendances significatives par application du même test.

Les auteurs concluent à un manque de robustesse de l'identification des tendances avec le test de MK pour les données traitées. Pour certains sites, des tendances significatives sont détectées alors qu'une part faible des simulations réalisées, prenant en compte l'incertitude de mesure, est identifiée comme présentant une tendance significative.

Les auteurs étudient également les corrélations possibles entre les taux de détection pour les données simulées et le résultat p value (statistique du test de Mann Kendall qui caractérise le caractère significatif ou pas du résultat du test ; plus il est faible et plus le résultat du test est significatif ; la valeur seuil est de 0.1 pour un indice de confiance de 90% et de 0.05 pour 95%). De façon qualitative, plus p est élevé plus le taux de détection est faible et réciproquement. Pour un degré de confiance de 90% ($p < 0.1$) les taux de détection sont compris entre 20 et 100%.

D'autres indicateurs de « fiabilité » des tests sont recherchés à travers l'exploitation des données sur les 10 sites montrant une tendance significative avec le test MK. Les auteurs considèrent comme « stables » les situations pour lesquelles les simulations réalisées entraînent de faibles dispersions sur les 1000 « p value » obtenues.

Des diagrammes «écart-type » des p value en fonction du rapport CV/E (CV= coefficient de variation des données et E erreur analytique) sont tracés. De façon schématique, plus le rapport CV/E est grand et plus la dispersion des p value est faible donc selon les auteurs, plus la tendance identifiée est robuste. Un rapport CV/E grand caractérise des chroniques pour lesquelles la dispersion globale des données est forte devant la variabilité analytique. Une valeur de 2 pour ce rapport CV/E est donnée par les auteurs comme critère qualitatif permettant de dire que la tendance identifiée par le test MK est fiable. De façon schématique, ce test rejoint celui décrit plus loin par Bouchard & al. [10]. L'approche décrite ci-dessus semble moins précise que dans ce dernier article. Le paramètre CV est pris en compte. Il s'agit de la dispersion globale des données intégrant l'éventuelle tendance et l'ensemble des facteurs de variations des données : il est peu informatif. L'autre paramètre est le paramètre E, incertitude analytique. Bouchard & al [10] prennent en compte l'étendue de variations des concentrations sur la droite de tendance identifiée entre le premier point de la série et le dernier point. Il divise ce terme T_r par la variance résiduelle σ_r de la série de donnée, c'est-à-dire la variance identifiée après prise en compte de la tendance dans l'analyse des données. Ce terme T_r/σ_r leur sert de critère discriminant de la même façon que le terme CV/E pour les auteurs du présent article. Le fait d'utiliser E dans ce rapport est réducteur. Cela suppose que l'on ait vérifié que l'incertitude analytique est l'unique facteur composant la variabilité des données (pas de variabilité liée à l'échantillonnage à des variabilités du milieu hors tendance, ...). Une représentation graphique de ces différents paramètres est proposée Figure 2.

Un autre diagramme similaire au précédent mais établi à partir de statistiques non paramétriques est tracé (étendue interquartile des p value en fonction du rapport IQR/E avec IQR, étendue interquartile des données divisée par la médiane en %). Ce dernier diagramme montre que les deux sites pour lesquels un taux de détection supérieur à 90% est observé sont les deux sites pour lesquels ce rapport est supérieur à 0.2. Pour les autres sites (test de Mann Kendall positif mais taux de détection inférieur à 90%),

ce rapport est inférieur à 0.2. Les auteurs soulignent que cette limite de 0.2 est une limite très empirique et devrait être validée à partir de grands jeux de données.

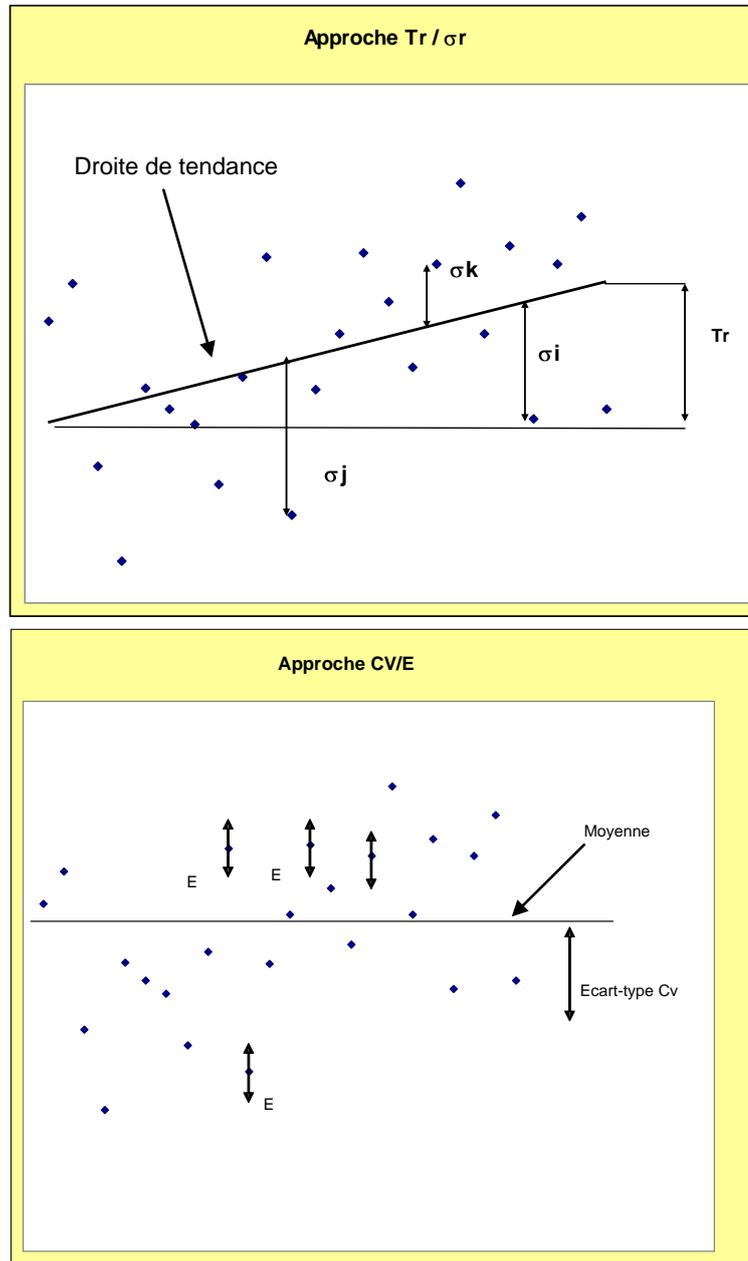


Figure 2 : Représentation schématique simplifiée des concepts développés dans les références [9](approche CV/E) et [10](approche Tr/σ) – cf texte.

Les auteurs évaluent enfin l'influence de l'erreur analytique sur le « taux de détection ». Ils utilisent pour leurs nouvelles simulations des erreurs analytiques variables de 0, 2, 10, 20 et 30%. Pour les 10 sites précités (test de Mann Kendall montrant une tendance significative), ils étudient le nombre de taux de détection supérieurs à 90% en fonction de l'erreur analytique. A titre d'exemple pour une erreur analytique de 20%, 2 sites sur les 10 ont un taux de détection supérieur à 90%. Les résultats sont de 10, 7, 4 et 1 pour des erreurs analytiques de 0, 2, 10 et 30%.

Quelques points importants

- les auteurs proposent une méthode pour tester la fiabilité des résultats du test de tendance de Mann Kendall en prenant en compte l'incertitude analytique sur les données.
- Les résultats sont issus de simulations obtenues à partir de chroniques réelles.
- Les tests effectués sur des chroniques pour lesquelles des tendances sont identifiées, montrent qu'une proportion non négligeable des simulations réalisées donne des résultats différents du test lui-même.
- La règle simplifiée suivante est donnée par les auteurs : les tendances identifiées à travers le test Mann Kendall paraissent significatives dès lors que le rapport entre la dispersion globale des données et l'erreur analytique (CV/E) est supérieure à 2.
- Deux critiques principales peuvent être formulées en plus de la critique ci dessus: le nombre de données par chronique est relativement faible (entre 4 et 20) et la fiabilité des tendances observées est donc certainement limitée. Par ailleurs, les simulations sont effectuées en considérant une distribution rectangle pour la prise en compte de l'incertitude alors qu'une distribution gaussienne semblerait plus adaptée et conduirait à renforcer la fiabilité des tendances observées.

5.2.2. “Modelling the effects of varying data quality on trend detection in environmental monitoring”, Sulkava, Luysaert, Rautio & al., 2007 [11]

Le domaine de travail des auteurs est l'étude de la concentration en certains nutriments dans les feuilles d'arbres. Ils partent du constat que l'interprétation de phénomènes complexes comme des effets de la pollution, du changement climatique nécessite des programmes de surveillance sur de longues durées ainsi que des évaluations de tendance. L'identification de tendance constitue également un outil d'orientation de politiques environnementales.

Les auteurs rappellent que le processus d'acquisition de données intègre un grand nombre d'étapes, chacune apportant sa part d'incertitude sur la donnée finale (voir aussi § 7 sur ce sujet). Par ailleurs, les auteurs indiquent qu'il est largement reconnu que de fortes variabilités sur les données peuvent masquer des tendances et que par conséquent des durées beaucoup plus longues peuvent être nécessaires afin de les identifier. Malgré cela, peu d'études existent permettant de quantifier ces observations.

L'objectif de l'étude proposée est donc de quantifier la durée d'acquisition nécessaire pour détecter des tendances linéaires dans la composition chimique des feuilles d'écosystèmes forestiers dans des conditions fixées d'exactitude et de fidélité des méthodes analytiques utilisées.

Dans de longues chroniques, il est difficile de supposer que les caractéristiques de la méthode analytique sont constantes. Elles évoluent avec l'évolution des techniques, de la maîtrise du personnel, ... Ces évolutions doivent être prises en compte dans l'analyse des données. Ainsi, une régression linéaire pondérée a été utilisée pour estimer le modèle. La pondération est définie comme l'inverse de la variance. Les données présentant une incertitude analytique faible ont donc ainsi un poids supérieur dans le processus d'estimation.

Les changements de la variance analytique au cours du temps ont été simulés de différentes façons. Le premier scénario correspond à une variance constante au cours du temps. Le deuxième scénario correspond à une variance constante pendant une durée de a années puis une évolution linéaire de la variance entre l'année a et l'année c et enfin une variance constante après l'année c .

Les simulations faites font l'hypothèse qu'il existe une tendance linéaire et que les données sont distribuées de façon aléatoire autour de la tendance. Par ailleurs, il est sous entendu dans ces simulations que les variations aléatoires sont uniquement dues à l'erreur analytique. Ceci maximise l'impact que peut avoir l'amélioration de la qualité analytique de la donnée sur l'estimation de tendance.

Les valeurs suivantes sont considérées pour effectuer les simulations :

- Une pente de 0.1

- Une valeur de a comprise entre 1 et 5 ans
- Une valeur de c comprise entre 2 et 10 ans
- Les termes σ_a et σ_c (avec $\sigma_a \geq \sigma_c$) ont été simulées de façon « continue » entre 0 et 1.

A partir de ce modèle et des simulations réalisées, les auteurs calculent le temps minimum nécessaire pour détecter de façon significative (plus de 50% des tendances identifiées comme statistiquement significative) la tendance simulée. Il est à noter que ce niveau de significativité choisi par les auteurs est faible. Par conséquent, le nombre d'année sera plus grand si un niveau de 90 % est considéré.

Le nombre de données temporelles simulées n n'est pas explicitement indiqué. Il semble que l'étude se base sur l'acquisition d'une donnée par an.

Les résultats sont présentés sous forme de diagramme montrant le temps minimum en fonction des paramètres σ_a et σ_c pour 3 couples de valeurs pour a et c (cf Figure 3)

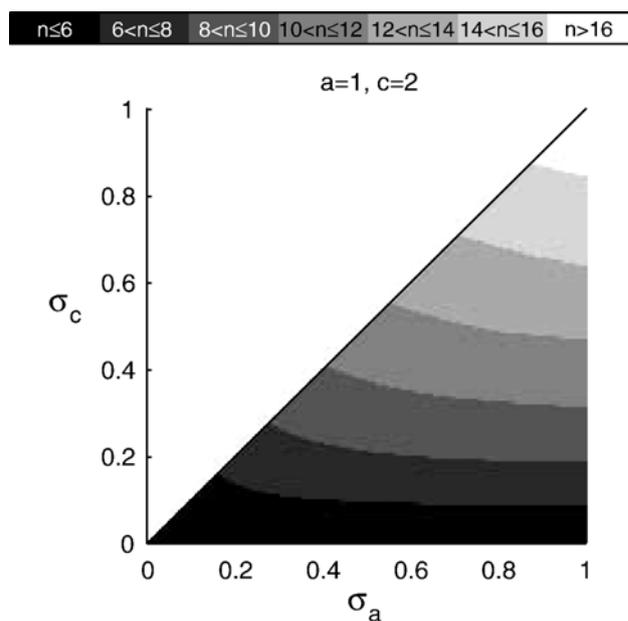


Figure 3: relation entre la variabilité analytique et la durée nécessaire pour détecter une tendance de pente 0.1 (taux de signification = 0.05 et puissance >50%)

Les simulations confirment, tout en apportant des éléments quantitatifs, les idées intuitives suivantes :

- Pour une tendance donnée, des données de meilleure qualité révéleront la tendance de façon plus précoce que des données de moins bonne qualité.

- Pour une tendance donnée, diminuer l'incertitude de la méthode d'analyse, y compris alors que le programme de surveillance est entamé, diminue le temps nécessaire pour pouvoir identifier cette tendance.

A titre d'illustration, les quelques données quantitatives suivantes sont extraites des résultats présentés :

- Pour $\sigma_a = \sigma_c = 0.2$ (incertitude de 40% avec un facteur d'élargissement de 2) il faut 7 années environ pour détecter la tendance suivant le critère choisi par les auteurs
- Pour $\sigma_a = \sigma_c = 0.4$ il faut 10 années
- Pour $\sigma_a = \sigma_c = 0.6$ il faut 13 années. Dans le cas où on conserve $\sigma_a=0.6$ pendant 5 ans et qu'en 2 ans on passe à $\sigma_c=0.2$, le temps nécessaire à détecter la tendance passe de 13 ans à 9 ans.

D'autres modèles de tendances non linéaires ont été testés par les auteurs mais peu de variations sont observées. Le test de Mann Kendall non paramétrique a aussi été utilisé pour traiter les données. Les résultats obtenus sont globalement proches excepté dans les cas pour lesquels l'incertitude sur la donnée change au cours du temps. Dans ces cas, les auteurs identifient une puissance de test inférieure pour le test de Mann Kendall.

Dans les conditions du modèle choisi, les données de simulation présentées précisent l'idée que la sensibilité d'un programme de surveillance est d'autant plus grande que l'incertitude sur les données est faible.

Les auteurs insistent cependant sur un des points délicats du modèle utilisé. Dans ce modèle, la variabilité totale sur la donnée est équivalente à la variabilité analytique. Ceci a déjà été précisé mais il est important de rappeler que dans ces conditions, les considérations théoriques montrent l'effet maximum que peut avoir une amélioration de la qualité de la donnée. Cependant, dans les programmes de surveillance, la variabilité totale sur la donnée provient de la variabilité analytique mais aussi de la variabilité liée à l'échantillonnage et à la variabilité du milieu. Dans le cas où ces deux dernières composantes sont majoritaires dans la variabilité globale de la donnée, les auteurs insistent sur le fait que les résultats présentés ne sont plus applicables.

En conclusion, les auteurs reconnaissent que la capacité à détecter une tendance dépend de la variabilité globale sur la donnée incluant milieu, échantillonnage, analyse. En conséquence cette variabilité globale doit être minimisée dans la mesure du possible pour permettre une détection de tendance rapide et fiable. Cette variabilité globale est de la responsabilité du gestionnaire du programme. Il peut en partie agir sur la variabilité naturelle à travers le choix de site, les périodes d'échantillonnage. Il peut aussi agir sur les composantes « échantillonnage » et « analyse » en veillant au respect de systèmes d'assurance et de contrôle qualité.

Quelques points importants

- Evaluation de la durée d'acquisition nécessaire à détecter une tendance de façon fiable en fonction de la variabilité sur les données.
- La variabilité sur les données est assimilée à l'incertitude analytique (hypothèse forte à vérifier).
- Etude basée sur des simulations de chroniques et sur un test de régression linéaire.
- Le critère de performance choisi est peu exigeant (seulement 50% des simulations présentant une tendance significative).
- Avec ce critère minimaliste, et pour une incertitude analytique de 40% ($k=2$) il faut 7 années pour détecter une tendance de pente 0.1.
- Le paramètre principal ayant un effet sur la sensibilité d'un programme de surveillance à vocation d'identification de tendance est la variabilité globale sur la donnée (incluant milieu, échantillonnage, analyse). Cette variabilité doit être minimisée pour identifier plus précocement une tendance.

La publication ci-dessous est voisine de celle qui vient d'être décrite (même contexte).

« Combining Measurement Quality into Monitoring Trends in foliar Nutrient Concentrations » - Sulkava & al., 2005 [13]

Le degré de signification statistique est estimé à travers un test Fisher et il est fixé à 0.05.

Les résultats sont présentés pour une pente ($\beta=0.05, 0.1, 0.15, 0.2, 0.25$) et une valeur de a données ($a=2, 4, 6, 8, 10$ ans) sous formes de diagramme carré (b en abscisse et c en ordonnée). Dans le diagramme, est reporté le nombre d'années nécessaires pour déceler une tendance significative (au degré de signification de 0.05) sous formes d'aires de couleurs différentes.

A titre d'exemple, en considérant $b=c=0.5$ ou 0.2 c'est-à-dire une variance constante de 50% ou 20% des données sur la durée du programme on obtient le nombre d'années du Tableau 2 avant de détecter une tendance significative:

	Pour une pente de				
Variance	0.05	0.1	0.15	0.2	0.25
50%	>10	>10	8<n<10	Env. 8 ans	6<n<8
20%	Env 10	6<n<8	Env 6	4<n<6	4<n<6

Tableau 2: nombre d'années nécessaires pour détecter une tendance significative de pente donnée en fonction de la variance analytique (d'après[13])

Les auteurs concluent que la précision de la mesure affecte fortement la détection de tendances. Améliorer la qualité des données notamment en termes d'incertitude diminue sensiblement le temps nécessaire pour identifier de façon statistiquement significative une tendance.

5.2.3. “Designing Monitoring programmes for detecting temporal trends in contaminants in fish and shellfish”, Nicholson & al (1997) [14]

Les auteurs présentent des outils statistiques pour concevoir des programmes de surveillance de polluants présents dans des poissons et des crustacés.

Un programme de surveillance doit maîtriser les sources de variations qui ne sont pas d'intérêt pour se focaliser sur celles qui le sont. Dans les programmes de surveillance des poissons et des crustacés, il s'agit par exemple de contrôler au mieux les variations liées à l'échantillonnage, l'analyse et les variations saisonnières afin de se focaliser sur les tendances persistantes.

Différents moyens existent pour maîtriser et diminuer certaines de ces variations. Les variations liées à l'échantillonnage peuvent être diminuées en prélevant plus d'animaux, la variance analytique peut être réduite en effectuant des répétitions et en calculant la moyenne de ces répétitions, ou bien en améliorant les procédures analytiques. Diminuer la variance liée aux saisons peut être réalisée en effectuant les échantillonnages toujours à la même époque de l'année.

Toutes ces notions sont intégrées dans le concept d'efficacité du programme de surveillance. Il s'agit de la probabilité que le programme détecte des variations temporelles données. Si des changements importants ne sont pas détectés, l'efficacité du programme est faible et réciproquement.

Les auteurs présentent donc dans leur article les bases théoriques et des tables permettant d'estimer l'efficacité d'un programme ayant pour objectif de détecter, par régression linéaire, une tendance linéaire donnée en T années. Ils montrent aussi comment ces informations permettent de mesurer l'influence de l'échantillonnage et de la variabilité analytique dans une hypothèse où R animaux sont collectés et analysés séparément à la même période de l'année.

La théorie est basée sur un modèle de tendance linéaire du type : $E[y_t]=a+bt$, avec b la pente de la tendance, Y_t la donnée à la date t et $E[y_t]$ est le résultat attendu à l'année t .

On suppose également que la variance des données est constante et donnée par :

$$\text{Var } [y_t]=\psi^2$$

La tendance est détectée par régression linéaire. Le test de signification de la tendance est un test Fisher au risque 5%. L'efficacité du programme est la probabilité que le test rejette l'hypothèse nulle (hypothèse d'absence de tendance soit $b=0$). Les auteurs indiquent que l'efficacité du programme dépend :

- de T , le nombre total d'années de surveillance
- et du rapport pente sur écart type (b/ψ).

Ils proposent une table (Tableau 3) donnant pour différentes valeurs de T (de 5 à 25 ans) et pour différentes valeurs de puissance P (0.5, 0.6, 0.7, 0.8, 0.9, 0.95 et 0.99) les valeurs du rapport b/ψ . Les tests sont basés sur une acquisition d'une donnée par an.

T	Puissance						
	0,5	0,6	0,7	0,8	0,9	0,95	0,99
5	0,906	1,035	1,176	1,344	1,584	1,786	2,175
6	0,616	0,7	0,791	0,899	1,051	1,178	1,421
7	0,459	0,52	0,586	0,664	0,773	0,864	1,037
8	0,36	0,408	0,459	0,52	0,604	0,674	0,806
9	0,293	0,332	0,373	0,422	0,49	0,546	0,653
10	0,245	0,278	0,312	0,352	0,409	0,455	0,544
11	0,209	0,237	0,266	0,3	0,348	0,388	0,462
12	0,181	0,205	0,23	0,26	0,301	0,335	0,4
13	0,159	0,18	0,202	0,228	0,264	0,294	0,35
14	0,141	0,16	0,179	0,202	0,234	0,261	0,311
15	0,127	0,143	0,161	0,181	0,21	0,233	0,278
16	0,114	0,129	0,145	0,163	0,189	0,211	0,251
17	0,104	0,117	0,132	0,148	0,172	0,191	0,228
18	0,095	0,107	0,12	0,136	0,157	0,175	0,208
19	0,087	0,098	0,11	0,125	0,144	0,16	0,191
20	0,08	0,091	0,102	0,115	0,133	0,148	0,176
21	0,074	0,084	0,094	0,106	0,123	0,137	0,163
22	0,069	0,078	0,088	0,099	0,115	0,127	0,152
23	0,065	0,073	0,082	0,092	0,107	0,119	0,141
24	0,06	0,068	0,077	0,086	0,1	0,111	0,132
25	0,057	0,064	0,072	0,081	0,094	0,104	0,124

Tableau 3 : valeurs du rapport b/ψ en fonction de la puissance voulu du test de tendance et du nombre d'années visées pour le programme de surveillance (d'après [14])

- A titre d'exemple, pour un programme s'étalant sur 10 ans, il y a 90% de chances de détecter une tendance satisfaisant à $b/\psi=0.409$. Si, par hypothèse les données présentent une dispersion de 0.2 (20%), il y a 90% de chances de détecter une tendance de ± 0.08 (8%) par an.
- Pour 10 ans avec une probabilité de 99%, le rapport b/ψ est de 0.544. Pour une dispersion des données de 0.2, la tendance détectable pour une probabilité de 99% est de 0.1 (10%) par an.
- En 5 ans et pour une probabilité de 90%, le rapport b/ψ est de 1.584. Toujours pour la même dispersion de 0.2, la tendance détectable ne serait que de 0.32 (32%).

Les données sont présentées également sous la forme d'un tableau (cf donnant, pour T et P données, la dispersion des données à cibler afin de pouvoir détecter une tendance de ± 0.05 (5% par an).

- Pour un programme de 10 ans et une probabilité de 90%, la dispersion à cibler est de 0.122 (12%).
- Pour un programme de 5 ans, et une probabilité de 90%, la dispersion à cibler est de 0.032 (3.2%).

T	Puissance						
	0,5	0,6	0,7	0,8	0,9	0,95	0,99
5	0,055	0,048	0,043	0,037	0,032	0,028	0,023
6	0,081	0,071	0,063	0,056	0,048	0,042	0,035
7	0,109	0,096	0,085	0,075	0,065	0,058	0,048
8	0,139	0,123	0,109	0,096	0,083	0,074	0,062
9	0,17	0,151	0,134	0,118	0,102	0,092	0,077
10	0,204	0,18	0,16	0,142	0,122	0,11	0,092
11	0,239	0,211	0,188	0,166	0,144	0,129	0,108
12	0,276	0,244	0,217	0,192	0,166	0,149	0,125
13	0,314	0,278	0,247	0,219	0,189	0,17	0,143
14	0,354	0,313	0,279	0,247	0,213	0,192	0,161
15	0,395	0,35	0,311	0,276	0,238	0,214	0,18
16	0,438	0,388	0,345	0,306	0,264	0,237	0,199
17	0,482	0,427	0,38	0,337	0,291	0,262	0,22
18	0,528	0,467	0,416	0,369	0,319	0,286	0,241
19	0,575	0,509	0,453	0,402	0,347	0,312	0,262
20	0,623	0,551	0,491	0,435	0,376	0,338	0,284
21	0,672	0,595	0,53	0,47	0,406	0,365	0,307
22	0,723	0,64	0,57	0,505	0,436	0,392	0,33
23	0,775	0,686	0,611	0,542	0,468	0,421	0,354
24	0,827	0,733	0,652	0,578	0,5	0,449	0,378
25	0,881	0,78	0,695	0,616	0,532	0,479	0,402

Tableau 4 : valeurs à cibler pour la dispersion ψ des données en fonction de la puissance voulu du test de tendance et du nombre d'années visées pour le programme de surveillance (d'après [14])

Suite à la présentation de ces concepts et des données théoriques associées, les auteurs étudient ensuite l'efficacité de programmes à partir de données réelles. Ils considèrent l'objectif suivant : être capable de détecter avec une puissance de 90%, une tendance de 5% pour différents niveaux de variance analytique ou d'échantillonnage. La variable temporelle suivie est la moyenne des log c observés dans R animaux collectés (c étant la concentration d'un contaminant dans un poisson).

- Les variances d'échantillonnage inter-année et intra-année se situent entre 8% et 51% et entre 25 et 44 % respectivement.
- Les variances analyse inter-année et intra-année sont comprises entre 4 et 10% et entre 9 et 23 % respectivement.

Ces variances sont ensuite estimées avec différentes données comme des résultats d'essais interlaboratoires pour les variances analytiques. Les variances « échantillonnage » sont estimées grâce à l'exploitation de 90 séries temporelles et par ajustement des courbes. Les auteurs rappellent que ces estimations de variances peuvent aussi être faites à travers la réalisation d'études pilotes. Chaque variance présente un intervalle de valeurs observées. Cet intervalle est divisé en niveau haut, bas et moyen.

La variance globale des données dépend du nombre R d'individus collectés et analysés. Il est donc possible d'en déduire, en fonction de ce nombre R et pour diverses hypothèses concernant les valeurs des variances élémentaires décrites ci-dessus, le nombre d'années nécessaires pour détecter une tendance de 5% avec une puissance de 90%. Augmenter R diminue le nombre d'années nécessaires à détecter la tendance car cela diminue la dispersion des données. Cependant cet effet est limité dans le cas où les variances « échantillonnage » sont globalement élevées (cf Tableau 5).

	R				
Variances échantillonnage	infini	125	25	5	1
Valeurs basses	12	12	12	14	18
V moyennes	17	17	18	18	22
V hautes	25	25	>25	>25	>25

Tableau 5: Nombre d'années nécessaire pour détecter une tendance de 5% par an en fonction du nombre d'individus analysés annuellement et de différentes hypothèses sur la variance des données. (d'après [14])

Des données sont également fournies pour différentes combinaisons des valeurs de variances analyse et échantillonnage. Pour chaque combinaison, la durée nécessaire pour détecter une pente de 5% (puissance 90% et R=25) est donnée dans le Tableau 6.

		Variances analytiques		
		Basses	Moyennes	Hautes
Variances échantillonnage	Basses	10	12	16
	Moyennes	17	18	20
	Hautes	25	>25	>25

Tableau 6 : Nombre d'années nécessaires pour détecter une tendance de 5% par an en fonction de différentes hypothèses sur la variance des données. (d'après [14])

Le Tableau 6 permet d'observer que, quand les variances liées à l'échantillonnage sont faibles, contrôler les variances analytiques permet d'optimiser le programme en termes de durée. Ce n'est pas le cas pour des variances « échantillonnage » élevées et pour des durées raisonnables.

D'autres simulations sont faites en faisant l'hypothèse que les analyses ne sont pas faites sur chacun des R animaux mais sur un échantillon moyen préparé. La conclusion de ces simulations est que l'analyse d'un échantillon moyen a peu d'effet sur la durée minimum nécessaire pour détecter la tendance mais par contre cette pratique diminue les coûts.

Quelques points importants

- Formalisation d'un objectif précis et quantitatif de la surveillance en termes de tendance : efficacité du programme, pente de la tendance, durée du programme
- Utilisation de la régression linéaire pour l'évaluation des tendances.
- A partir de modèles et d'outils statistiques simples, possibilité de dimensionner le programme de surveillance afin d'atteindre les objectifs fixés
- Le paramètre clé permettant de dimensionner le programme de surveillance et de répondre aux objectifs fixés (efficacité, pente, durée) est la variance globale des données.
- Des tableaux présentent en fonction d'objectif prédéfinis, les variances maximales permettant d'atteindre ces objectifs.
- Estimation des variances analytiques et échantillonnage sur la base d'informations disponibles ou bien d'études pilotes.
- Etude des performances du programme de surveillance en fonction du nombre R de spécimen analysés par an (si R augmente les variances « analyse » et « échantillonnage » diminuent et l'efficacité du programme augmente).

5.2.4. « Trend detection in water quality time series of LRTAP – QUEBEC network lakes”, Bouchard & al, 1992 [10]

Dans le contexte d'études d'impact de pluies acides en Amérique du Nord dans les années 1980, un programme de surveillance de la qualité des lacs du Québec a été mis en place en 1983. Cet article présente l'analyse des 5 premières années de surveillance (25 à 30 données) pour 12 sites. Les paramètres suivis sont pH, alcalinité, sulfates, nitrates, Ca, Mg, Na, K.

L'objectif donné à ce programme était de pouvoir détecter des tendances statistiquement significatives en 7 à 10 ans avec un niveau de confiance de 90%. La conception du programme s'est par ailleurs basée sur une étude pilote au cours de laquelle l'acquisition de données mensuelles pendant 2 années a été réalisée. Ces données de caractérisation initiale de sites ont permis avec l'objectif chiffré ci-dessus de dimensionner le programme en termes de nombre de points pour une période donnée. Une fréquence optimale de 6 échantillons par an a été déterminée.

Les objectifs en termes de tendance à quantifier sont présentés sous la forme d'un rapport T_r / σ (cf. paragraphe 5.2.1 pour une description schématique de ces notions, et 5.2.2) avec :

- T_r : calculé comme étant la valeur finale moins la valeur initiale de la droite de tendance estimée
- σ est l'écart-type résiduel calculé à partir des paramètres de régression. Sans que cela soit explicite dans l'article, ce paramètre définit la variabilité des données hors tendance. Il a très certainement été estimé avec la phase préliminaire d'acquisition de données sur 2 années. Dans les cas où des données antérieures sont traitées, ce paramètre est donné par le logiciel de traitement statistique.

L'objectif initial du programme est de pouvoir détecter une tendance telle que $T_r / \sigma \geq 1$.

Peu de temps après le démarrage du programme une étude a remis en cause la durée au bout de laquelle des résultats significatifs d'identification de tendance pourraient être obtenus. L'explication donnée est principalement liée à des niveaux de significativité différents pris en compte ainsi qu'à l'auto corrélation des données. Le point intéressant de l'étude citée concerne le mode de présentation des résultats ainsi que les ordres de grandeur donnés. Les 2 tableaux ci-dessous (Tableau 7 et Tableau 8) reprennent les résultats de cette étude. Pour 4 paramètres chimiques, ils présentent, pour une puissance de test de 90% et un niveau de signification de 5% :

- Tableau 7 : en fonction d'un objectif exprimé sous la forme T_r / σ , et en fonction d'un nombre d'échantillon annuel, le nombre d'années nécessaires pour identifier la tendance.

- Tableau 8 : en fonction d'un objectif en termes d'années et en fonction du nombre d'échantillons annuels, la tendance (exprimée sous forme Tr/σ) qui sera détectable.

Tr/σ	Nombre d'échantillons annuels	Alcalinité	SO4	Ca+Mg	pH
2	6	8.2	6.9	6.2	5.6
	8	7.3	5.8	5.3	4.5
	13	6.7	4.7	4.3	3.3
	26	6.2	4	3.6	2.5
1.5	6	13.9	11.6	10.4	9.5
	8	12.4	9.9	9	7.6
	13	11.3	8	7.3	5.5
	26	10.5	6.7	6	4.3
1	6	31.4	26.1	23.4	21.3
	8	27.9	22.3	20.2	17.1
	13	25.5	18.1	16.4	12.5
	26	23.7	15.1	13.5	9.6
0.7	6	64	53.3	47.8	43.4
	8	57	45.4	41.2	35
	13	52.1	36.9	33.5	25.5
	26	48.3	30.7	27.6	19.5

Tableau 7 : Pour quatre paramètres physico chimiques d'eau de lac, nombre d'années nécessaires pour détecter une tendance dans des conditions définies par le rapport Tr/σ en fonction du nombre données par an (d'après [10])

Période	Nombre d'échantillons annuels	Alcalinité	SO4	Ca+Mg	pH
5	6	2.61	2.37	2.23	2.13
	8	2.46	2.17	2.06	1.89
	13	2.34	1.94	1.85	1.60
	26	2.24	1.77	1.67	1.39
10	6	1.77	1.62	1.53	1.46
	8	1.67	1.49	1.42	1.31
	13	1.6	1.34	1.28	1.12
	26	1.54	1.23	1.16	0.98
15	6	1.45	1.32	1.25	1.19
	8	1.36	1.22	1.16	1.07
	13	1.3	1.10	1.05	0.91
	26	1.26	1.00	0.95	0.80
20	6	1.25	1.14	1.08	1.03
	8	1.18	1.06	1	0.93
	13	1.13	0.95	0.91	0.79
	26	1.09	0.87	0.82	0.69

Tableau 8 : Pour quatre paramètres physico chimiques d'eau de lac, rapport T_r/σ détectable en fonction du nombre d'années de surveillance et du nombre de données par an (d'après [10])

Ces résultats montrent de façon quantitative que plus une tendance est importante par rapport au terme σ , plus le nombre d'année nécessaire pour l'identifier sera faible.

On peut aussi noter que doubler l'effort d'acquisition de données au sein d'une année n'implique pas une diminution proportionnelle de la durée du programme. De façon générale, un doublement de l'effort de surveillance (fréquence d'échantillonnage) diminue de 10 à 25% seulement la durée du programme.

Les résultats des 5 premières années de surveillance ont été traités avec différents tests statistiques non paramétriques (Mann Kendall, Mann Kendall saisonnier, Spearman, Friedman, ...) en fonction des caractéristiques des données. Les sites et paramètres chimiques pour lesquels des tendances significatives ont été identifiés sont présentés. Dans chaque cas, les rapports T_r/σ sont calculés. Les tendances sont ensuite analysés d'un point de vue environnemental et notamment à travers les effets d'acidification de lacs dus aux pluies acides.

Une autre exploitation concerne la capacité à détecter les tendances. Le paramètre critique utilisé est le rapport T_r/σ . Dans la quasi-totalité des cas pour lesquels des tendances statistiquement significatives ont été détectées, les rapports T_r/σ sont

supérieurs à 1. Seuls deux cas impliquant l'utilisation du test Mann Kendal saisonnier ont des rapports inférieurs à 1.

Dans le cas où aucune tendance significative n'est identifiée, les rapports T_r/σ sont à l'inverse inférieurs à 1. Cependant il existe aussi dans quelques cas (principalement avec les tests Spearman et Hirsch) des rapports T_r/σ supérieurs à 1 alors qu'aucune tendance significative n'est observée.

Malgré les quelques exceptions, la limite de 1 pour le rapport T_r/σ est une valeur de référence pour discriminer les tendances significatives. Les auteurs en concluent que la configuration du réseau est adaptée à l'objectif de départ fixé malgré une exploitation au bout de 5 années seulement alors que 7 années étaient initialement prévues. Une partie de ces bons résultats est liée à l'utilisation de tests non paramétriques alors que la configuration initiale du programme a été réalisée avec des hypothèses d'utilisation de tests paramétriques.

Quelques points importants

- Construction et dimensionnement d'un réseau de surveillance sur la base d'un objectif quantitatif fixé et d'une préétude expérimentale permettant de déterminer la variabilité des données.
- Construction de tableaux d'aide à la décision permettant de déterminer dans des conditions bien définies la durée attendue du programme de surveillance permettant d'atteindre l'objectif fixé en termes de tendances ou bien de déterminer la tendance qui serait détectable en un temps donné. Tout ceci pour des variabilités de données préalablement déterminées (variabilité incluant l'incertitude analytique).
- Utilisation d'un paramètre empirique T_r/σ (variation de concentration sur la période considérée divisée par la variabilité de la donnée) permettant, de façon empirique, de discriminer l'identification de tendances significatives (valeur discriminant de 1 environ)

5.3. PERFORMANCES DE TESTS STATISTIQUES ET FACTEURS INFLUENTS

5.3.1. "Power of the Mann-Kendall and Spearman's rho tests for detecting monotonic trend in hydrological series", Yue, Pilon, Cavadias, 2002 [7]

Deux tests non paramétriques sont très souvent utilisés pour identifier des tendances dans des séries temporelles. Il s'agit des tests de Mann Kendal (cf §4) et de

Spearman. Cette étude s'intéresse à la puissance de ces tests en réalisant des simulations en fonction de différents paramètres comme le niveau de signification choisi, l'importance de la tendance, le nombre de données, et l'importance de la variabilité des données au sein de la série temporelle. Les deux tests sont également comparés.

La raison principale de l'utilisation de tests non paramétriques est que ces tests sont de façon générale plus adaptés à des jeux de données qui ne suivent pas une loi normale (asymétrie, aplatissement, valeurs extrêmes) et à des jeux de données non complets.

Le niveau de signification α d'un test est la probabilité de rejeter l'hypothèse « il n'existe pas de tendance » alors que cette hypothèse serait vérifiée. Par exemple avec un niveau de signification de 5%, il y a 5% de chance d'affirmer qu'il y a une tendance alors qu'en réalité il n'existe pas de tendance.

Des simulations de Monte Carlo sont réalisées par les auteurs afin de tester la puissance des tests. Des jeux de 2000 séries temporelles normalement distribuées sont générés. Les paramètres variant dans ces jeux de données sont les suivants sachant que les données sont choisies par convention comme ayant une moyenne de 1 :

- Nombre de données de la série temporelle (de 10 à 100 par pas de 10)
- Variance des données de 0.1 à 1 par pas de 0.1
- Pente de la tendance : de -0.01 à +0.01

Dans ces jeux de données, il existe donc toujours une tendance. L'hypothèse nulle H_0 « il n'existe pas de tendance » est donc fautive. La puissance du test est déterminée par le nombre de séries pour lesquelles l'hypothèse nulle est rejetée sur le nombre total de simulations.

Les données sont présentées sous forme de diagramme en 3 dimensions représentant la puissance du test en fonction de 2 paramètres, les autres paramètres étant fixés (Figure 4).

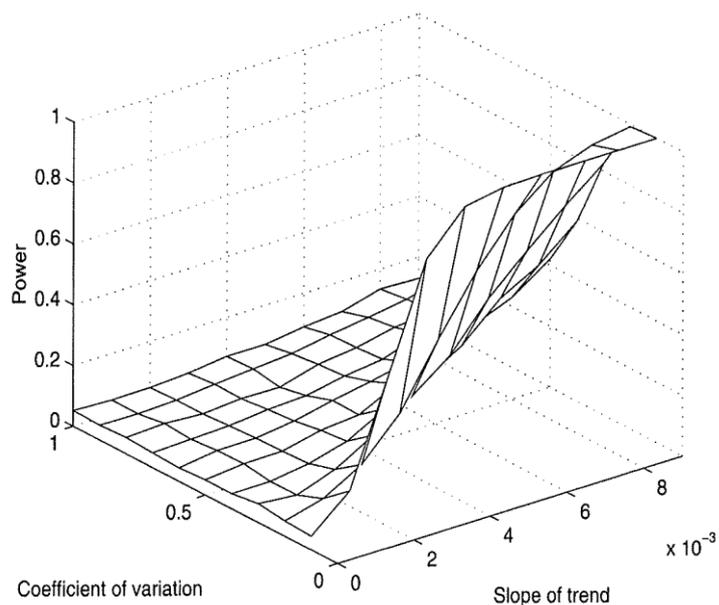
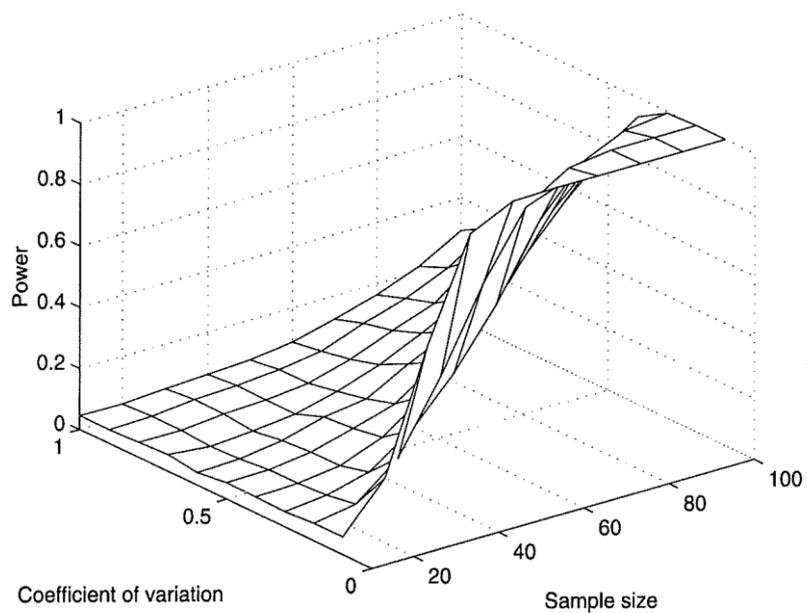


Figure 4 : Diagrammes puissance en fonction de CV et N ($\alpha=0.05$ et pente 0.005) puis de CV et de la pente ($\alpha=0.05$ et N=50) [7]

Quelques observations faites par les auteurs sont présentées ci dessous.

Pour un nombre de données et une variance (ou un coefficient de variation) fixés :

- Pour α fixé, la puissance du test est une fonction croissante de l'intensité de la pente
- Pour une pente fixée, augmenter α (degré de signification) augmente la puissance du test.

Pour un degré de signification et une variance fixés,

- la puissance du test est une fonction croissante de la pente et du nombre de données. Augmenter le nombre de données augmente la puissance du test conduisant à une capacité accrue à identifier l'existence d'une tendance.

Pour un nombre de données N et α fixés

- pour une pente fixée, la puissance du test est une fonction décroissante du coefficient de variation de la série. En d'autres termes, lorsque la variabilité des données augmente au sein de la série, la puissance du test diminue et il est donc plus difficile de détecter l'existence d'une tendance ou encore, les variations au sein d'une série masquent l'existence d'une tendance.

A partir des différents diagrammes il est par exemple possible d'établir le nombre minimum de données nécessaires pour détecter une pente préétablie pour un niveau de signification et une puissance de test donnés.

Différents types de distribution de données sont également étudiées à travers les mêmes simulations décrites ci-dessus. En l'absence de pente, toutes les distributions testées donnent la même puissance de test. Quand la pente augmente, les différences sur la puissance du test en fonction du type de distribution des données augmentent significativement. La distribution normale des données donne la puissance de test la plus faible.

A titre d'exemple, si on considère deux sites pour lesquels il existe une pente de 0.005, un nombre de données de 50, un CV de 0.5, la puissance du test sera double pour le site présentant une distribution non normale par rapport au site présentant une distribution normale : la puissance du test est fortement affectée par les caractéristiques du site quand une tendance existe.

Ces résultats montrent que le test de Mann Kendall n'est pas insensible au type de distribution des données à l'inverse de ce qui est habituellement admis.

Les tests de Mann Kendall et de Spearman sont identifiés comme donnant des résultats très semblables par les simulations réalisées. Ces deux tests sont appliqués

sur une série de chroniques temporelles concernant le débit journalier maximum annuel de 20 rivières.

Les auteurs insistent dans la partie finale de l'article sur l'importance de différencier la significativité statistique et la significativité pratique d'une tendance. Il n'existe pas de lien automatique entre ces deux notions. Par exemple sur un site, une tendance réelle pourra être détectée de façon statistiquement significative si on considère un grand nombre de données alors qu'elle ne sera pas détectée de façon significative avec un jeu de données réduit. Les auteurs illustrent leur propos à travers l'exemple d'un site pour lequel la chronique temporelle est courte (peu d'historique de données) et pour lequel aucune tendance statistique significative n'est identifiée. Pourtant, qualitativement, l'augmentation du débit maximum annuel semble conséquente et dans l'hypothèse où la pente observée serait confirmée, cette augmentation serait d'une importance pratique capitale pour la gestion du bassin.

Les auteurs ajoutent que les estimations statistiques de tendance doivent dans la mesure du possible toujours être confirmées par des avis d'experts s'appuyant sur des considérations environnementales concrètes.

Quelques points importants

Les principaux paramètres considérés comme critiques pour tester la puissance des tests de tendances sont :

- le nombre de données disponibles
- l'intensité de la pente
- le niveau de signification recherché
- la variance des données

Ce dernier paramètre est le paramètre dont l'impact est recherché dans le cadre de ce rapport. Il intègre à la fois la variabilité temporelle « environnementale » des données sur le site considéré mais également la variabilité des données liée à l'échantillonnage et à l'analyse. Dans le cadre de ce rapport, les données de cet article peuvent donc être utilisées pour quantifier l'impact de ce paramètre.

De grandes tendances sont identifiées qualitativement mais aussi quantitativement à la lecture des diagrammes présentés en fonction des paramètres des chroniques. La plupart des observations qualitatives paraissent relativement évidentes de façon intuitive. Les données quantitatives issues des diagrammes sont plus intéressantes car elles pourraient permettre de dimensionner un programme de surveillance destiné à mettre en évidence une tendance en fonction d'objectifs préalablement définis intégrant par exemple un niveau de signification et un niveau de pente. Les données

suivantes sont présentées pour un coefficient α de 0.05, uniquement à titre d'information, et de façon très qualitative car elles sont issues de l'exploitation visuelle des graphiques de l'article.

Pente	CV	N	Puissance
0.005	0.1	45	80
0.005	0.1	50	100
0.005	0.2	50	65
0.005	0.4	100	80
0.005	0.5	50	15
0.01	0.1	50	100
0.01	0.5	20	8
0.01	0.5	40	30
0.01	0.5	50	50
0.01	0.5	60	75
0.01	0.5	65	80
0.01	0.5	100	100
0.01	1	50	10

La lecture de ce tableau fournit par exemple les informations suivantes :

- Pour une pente de 0.005, un CV de 0.1 (10% de variabilité sur les données) et 50 données, la puissance du test est de 100%, c'est-à-dire que la probabilité d'identifier une telle tendance dans ces conditions est de 100%. Dans les mêmes conditions, si le CV est doublé (CV=0.2 ou 20%), la probabilité descend à 65%.
- Pour un pas de temps de 3 mois, une pente de 0.005 correspond à une variation annuelle de 2%. Avec N=45, la durée d'observation de la chronique est de 11 ans environ. Sur cette période, l'évolution de concentration est de 22%. En supposant que la variabilité sur la donnée est uniquement due à l'incertitude de mesure, pour une incertitude élargie de 20% (CV=0.1), la puissance du test sera de 80%.

En prenant comme objectif une puissance de test minimale de 80%, toute donnée dont l'incertitude serait supérieure à 20% (k=2) empêcherait la détection de façon fiable d'une tendance de 22% sur 11 ans (en faisant la même hypothèse que ci-dessus concernant l'incertitude).

6. Simulations et exploitation de chroniques réelles

Les quelques exploitations de données simulées ou de données réelles qui sont décrites dans ce chapitre ont pour objectif d'essayer de quantifier l'influence de la variabilité des données sur l'efficacité d'un programme de surveillance. L'objectif est notamment de déterminer, sur des chroniques temporelles de la qualité des eaux souterraines, de quels types de situation ces chroniques peuvent se rapprocher (situations issues de la synthèse bibliographique, des données simulées, situations notamment en termes de variabilité des données, ...). Pour cela, les facteurs principaux dont il a été montré qu'ils pouvaient être influents concernant les études de tendance seront utilisés dans la mesure du possible (variance résiduelle, variance des données, variance analyse, échantillonnage, ...).

6.1. SIMULATIONS

L'objectif de ces simulations est dans un premier temps de vérifier la bonne mise en œuvre des tests de Mann Kendall et du modèle de régression linéaire sur un jeu de données présentant une tendance à la hausse. Dans un deuxième temps il s'agit de démontrer l'impact que peut avoir l'incertitude sur les données (en faisant varier cette incertitude) sur l'évaluation des tendances.

6.1.1. Objectif en termes d'évaluation de tendance

Une des principales règles de bonne pratique dans la conception d'un réseau à des fins d'évaluation de tendance est de fixer très précisément et quantitativement les objectifs.

Les simulations suivantes ont été réalisées sur la base de l'objectif quantitatif suivant :

Déceler une augmentation de la concentration d'une substance de 20% en 6 ans avec une acquisition de 12 ou 24 données sur la période. Trois niveaux de variabilité, exprimées par le coefficient de variation ($CV = \text{Ecart-type} / \text{Moyenne}$ en %) des données ont été fixées : $CV = 2,5\%$, 10% et 30% . Par hypothèse, ces variabilités correspondent à l'incertitude analytique, provenant de la méthode d'analyse et des opérations techniques d'échantillonnage. L'incertitude élargie est $U = 5\%$, 20% et 60% .

La notion d'incertitude analytique utilisée dans les simulations est une simplification. La variabilité globale sur la donnée, décrite dans la synthèse bibliographique, inclut une part liée à la mesure (analyse et échantillonnage) et une part liée à la variabilité du milieu (cyclicité, ...). Ici, l'hypothèse faite est que la variabilité du milieu est négligeable ou bien que la variabilité globale de la donnée est de 2,5, 10 et 30%.

6.2. DONNEES SIMULEES

6.2.1. Mode de simulation

Dans le cas de $n = 12$ donnés, on suppose un prélèvement tous les 6 mois pendant 6 ans. Le niveau de concentration théoriquement atteint a été calculé pour chacun de ces points en supposant un accroissement linéaire dans le temps et en tenant compte de la fréquence de prélèvement.

Dans l'exemple, on commence avec une concentration d'une substance (pesticides) $C = 0,1 \mu\text{g/l}$ qui atteint $0,12 \mu\text{g/l}$ au bout de 6 ans. Le niveau de concentration au point n est calculé par la formule :

$$C_n = C_{\text{initiale}} * (1 + \mu)^n$$

Avec μ déterminé par : $C_{12} = C_{\text{initiale}} * (1 + \mu)^{12} = C_{\text{initiale}} * (1,20)$

Soit $\mu = 0,01531$ ou 1,53% d'augmentation par semestre relativement au niveau de concentration atteint au semestre considéré.

L'augmentation relative est de 3,1% la première année, puis 6,27% à la fin de la deuxième année,

Puis il a été supposé une distribution normale des valeurs autour de cette concentration théorique avec un écart-type s , dépendant du coefficient de variation considéré. s est déterminé par : $s_n = CV * C_n$

12 variables aléatoires de loi normale, avec une moyenne C_n et un écart-type s_n , représentant les niveaux de concentration possibles au moment des prélèvements ont été configurées dans le fichier de simulation.

La simulation a été lancée dans le logiciel Crystal Ball en indiquant $n = 1000$ simulations du groupe des 12 données. Pour chaque groupe des 12 données, une tendance est estimée avec le test de Mann Kendall et la régression linéaire.

Cette opération de simulation a été réalisée 3 fois pour tenir compte des différents cas de variabilité globale des données : $CV = 5\%, 10\%$ et 20% .

Dans le cas de $n = 24$ données, on suppose un prélèvement tous les 3 mois pendant 6 ans. Le niveau de concentration au point n est calculé par la formule :

$$C_n = C_{\text{initiale}} * (1 + \mu)^n$$

Avec μ déterminé par $C_{24} = C_{\text{initiale}} * (1 + \mu)^{24} = C_{\text{initiale}} * (1,20)$.

Soit $\mu = 0,0076$ ou 0,76% d'augmentation par trimestre relativement au niveau de concentration atteint au trimestre considéré.

6.2.2. Exploitation des données

Pour chacune des 1000 séries simulées, le test de tendance de Mann Kendall et la régression linéaire sont calculés.

- Dans le cas du test de Mann Kendall, la valeur du paramètre Z indique le caractère significatif ou non de la tendance observée. Au seuil de signification 5% choisi, la valeur critique pour Z est 1.96.

Le nombre de tendances statistiquement significatives sur le total des 1000 simulations est enregistré.

- Dans le cas de la régression linéaire, les données sont modélisées par : $y = a + b.x$ où a représente la concentration initiale, x le temps et b la pente de la tendance linéaire.

Dans cette étude, la période d'une année est codée par une valeur de 1 : ainsi, $x = 0$ pour la concentration initiale, $x = 0,5$ le premier semestre de la première année, $x = 1$ le deuxième semestre de la première année, ...

Dans le cas de $n = 12$ données, les valeurs de x sont : 0, 0.5, 1, 1.5, 2, ...,6. En incluant le point 0, la droite est estimée avec 13 points.

Dans le cas de $n = 24$ données, les valeurs de x sont : 0, 0.25, 0.5, 1, 1.25, 1.5, 2, ...,6. La droite est estimée avec 25 points.

Dans notre exemple, la concentration initiale est $C = 0,1 \mu\text{g/l}$, l'augmentation est de 20% en 6 ans et la concentration finale est $0,12 \mu\text{g/l}$. La valeur théorique de la pente est obtenue en divisant l'augmentation de concentration par l'intervalle de temps : $b_0 = 0,02 / 6 = 0,0033$. La droite théorique est : $y = 0,1 + 0,0033*x$

Les données simulées varient autour de cette droite et l'amplitude de variation dépend du niveau d'incertitude fixé lors de la simulation. Un exemple de pente estimée est : $y = 0,0989 + 0,0032 x$

Le test de l'existence d'une tendance est un test bilatéral de Student dans lequel le paramètre du test est le rapport de la pente estimée b à son écart-type :

$$t_{\text{observé}} = | b | / u(b)$$

La pente b_0 est déclarée significativement différente de 0 lorsque $t_{\text{observé}}$ excède une valeur seuil :

- $t_{\text{critique}} = 2,201$ pour $n = 12$ données, 13 points et 11 degrés de liberté
- $t_{\text{critique}} = 2,069$ pour $n = 24$ données, 25 points et 23 degrés de liberté

dans le cas d'un test bilatéral, à 95%.

Le nombre de tendances statistiquement significatives sur le total des 1000 simulations est enregistré.

6.2.3. Résultats et interprétations

Test de Mann Kendall

Les résultats sont donnés dans le Tableau 9 et les graphes de la Figure 5. Ils montrent que :

- Pour un niveau d'incertitude donnée, l'efficacité du programme de surveillance en termes d'évaluation de tendance augmente avec le nombre de données.
- Pour un niveau d'incertitude de 5%, que ce soit avec $n=12$ ou $n=24$ la tendance simulée (20% en 6 ans) est détectée systématiquement (efficacité ou puissance = 100%).
- La puissance du test diminue quand l'incertitude (ou de façon plus générale la variabilité des données) augmente.
- Traduits d'une autre façon et en prenant l'exemple $n=12$ (resp $n=24$) et $U=20\%$, la probabilité de détecter une pente d'environ 20% sur 6 ans est de 31% (resp 65%).
- Si on fixe une probabilité de 90%, l'« incertitude » maximale devrait être de l'ordre de 7-12%.

	Incertitude élargie en %		
	5	20	60
n=12	99%	31%	0,7%
n=24	100%	65%	10%

Tableau 9: Pourcentage de simulations pour lesquelles une tendance significative est identifiée en fonction de l'incertitude sur la donnée et du nombre de données par an (2 ou 4)

Ces simulations permettent donc à la fois de disposer d'outils permettant de réfléchir aux effets des variations des grands facteurs déterminant un programme de surveillance (fréquence d'acquisition des données, variabilité des données, période d'observation, ...). En plus de ces grandes tendances, de telles simulations fournissent des éléments quantitatifs utiles pour la conception de programmes destinés à l'évaluation de tendances.

Chaque fois que des tendances fiables sur le plan statistique sont identifiées (au niveau 5%) par le test de Mann Kendall, les pentes déterminées par le test sont relevées. Puis, la moyenne et l'écart-type de ces pentes sont calculées. Les résultats sont reportés dans la Figure 5.

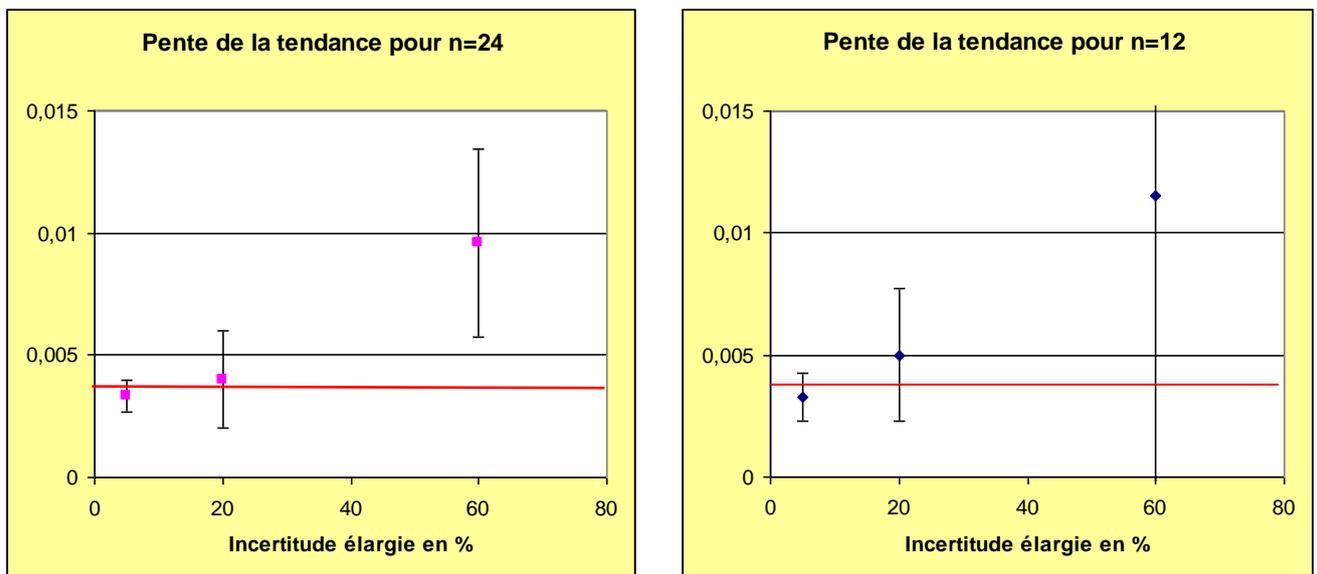


Figure 5 : moyenne des pentes estimées pour les simulations en fonction de l'incertitude et du nombre de données (les pentes théoriques sont identifiées par les traits rouges)

Les résultats montrent que dans le cas où des tendances significatives sont identifiées par le test de Mann Kendall, l'incertitude sur l'estimation des tendances est d'autant plus forte à la fois en variabilité et en justesse que l'incertitude sur la donnée est forte. Pour U=60%, la pente moyenne estimée sur les 1000 simulations est environ 3 fois plus grande que la pente théorique que ce soit pour n=12 ou n=24.

Régression linéaire

L'efficacité (la puissance) du test de tendance réalisé dans le cadre de la régression linéaire a été déterminée selon deux méthodes : calcul empirique sur les simulations, calcul théorique basé sur les quantiles de la loi normale.

Le calcul sur les simulations consiste à déterminer le nombre de pentes déclarées significatives sur les 1000 simulations réalisées. Le calcul est effectué pour les 6 configurations testées. Les résultats sont présentés dans le Tableau 10.

	Incertitude élargie en %		
	5	20	60
n = 12	100%	48%	9%
n = 24	100%	79%	12%

Tableau 10: Pourcentage de simulations pour lesquelles une tendance significative est identifiée en fonction de l'incertitude sur la donnée et du nombre de données par an

Le test de Student sur la pente est réalisé sans préciser de valeur pour la tendance. Néanmoins, plus la tendance est importante, plus elle sera facilement détectée. L'efficacité du test dépend de la valeur de la tendance.

Dans cette étude la tendance simulée est de 20% d'augmentation en 6 ans, ce qui correspond à une pente théorique $b_0 = 0,0033$. Nous calculons l'efficacité du test à cette valeur de pente.

La pente estimée b suit une loi normale dont la moyenne est la pente théorique $b_0 = 0,0033$ et son écart-type est $u(b)$, à déterminer.

Cet écart-type $u(b)$ dépend de l'écart-type des données et du nombre de données utilisées pour estimer la droite.

Dans notre exemple, l'incertitude sur les données est exprimée en valeur relative ; nous calculons un écart-type moyen noté u absolu associé à chaque point.

L'écart-type de la pente estimée est calculé ainsi :

$$u(b) = u \text{ absolu} / (\text{SCE } x)^{0,5}$$

où SCE_x représente la somme des carrés des écarts des valeurs de x à leur moyenne (la variance des x , à un facteur multiplicatif près). Les résultats sont présentés dans le Tableau 11.

Nb données	U relatif	u relatif	u absolu	penne b	u(b)	u(b) en %
n = 12	5%	2.50%	0.00277	0.00333	0.000410	12%
n = 12	20%	10%	0.01107		0.001641	49%
n = 12	60%	30%	0.03321		0.004923	148%
n = 24	5%	2.50%	0.00276	0.00333	0.000306	9%
n = 24	20%	10%	0.01083		0.001201	36%
n = 24	60%	30%	0.03308		0.003670	110%

Tableau 11 : Ecart-type moyen des données et écart-type de la pente en fonction du nombre de données

En termes statistiques, l'efficacité du test est la probabilité de déceler la tendance lorsqu'elle existe :

$$\text{Efficacité} = \text{probabilité} \{ \text{rejetter } b_0 = 0 / b_0 = 0,0033 \}$$

Dans le test, l'hypothèse $b_0 = 0$ est rejetée lorsque $t_{\text{observé}}$ excède la valeur seuil appelée t_{critique} . Intégrer la moyenne $b_0 = 0.0033$ dans le test revient à modifier la valeur seuil. Elle devient :

$$\text{critique-}b_0 = t_{\text{critique}} - 0.0033/u(b)$$

L'efficacité est calculée ainsi :

$$\text{Efficacité} = \text{probabilité} \{ \text{Student} > t_{\text{critique-}b_0} \}$$

Par exemple, dans le cas de $n = 12$ données et une variabilité $U = 10\%$, la valeur seuil est :

$$t_{\text{critique-}b_0} = 0,170$$

$$\text{Efficacité} = \text{probabilité} \{ \text{Student} > 2,201 - 0,0033/u(b) = - 5,924 \} = 43,4\%$$

Le tableau suivant présente les calculs d'efficacité pour les différentes configurations des données.

	Incertitude élargie en %		
	5	20	60
n = 12	100%	43%	7.8%
n = 24	100%	76%	13%

Tableau 12 : Probabilité de détecter une tendance en fonction de l'incertitude sur la donnée et du nombre de données par an

Sur cet exemple, il est à noter que les résultats de l'exploitation par régression linéaire montrent une efficacité supérieure à détecter la tendance. Il n'est toutefois pas possible de généraliser cette observation à partir de ce seul exemple.

6.3. EXPLOITATION DE CHRONIQUES REELLES

A la fois dans le chapitre concernant l'exploitation des références bibliographiques ainsi que dans le paragraphe 6.2 beaucoup d'études sont basées sur des simulations. Il est intéressant d'essayer de déterminer sur un certain nombre de chroniques, comment se situent les données réelles par rapport à ces simulations. (notamment en termes de variabilité des données réelles par rapport aux simulations).

L'objectif principal de ce chapitre est d'estimer la variabilité globale des données pour différents types de substances. Cette variabilité est un des facteurs déterminant l'efficacité d'un programme de surveillance. Cette variabilité sera comparée à la variabilité due à la méthode analytique utilisée pour doser ces substances. Cette comparaison permettra de vérifier si l'incertitude analytique constitue une part prépondérante ou pas dans cette variabilité et donc de connaître l'influence potentielle de l'incertitude analytique sur l'efficacité des programmes de surveillance. Enfin, il s'agit de voir si le rapport pente de la tendance/variabilité (écart type) peut être utilisé comme critère de robustesse des tests de tendance mis en œuvre sur ces chroniques réelles

6.3.1. Données sur les Nitrates

Une quarantaine de sites de surveillance « eau souterraine » ont été choisis parmi des sites de surveillance d'agences de l'eau. Ces sites ont été choisis à l'intérieur d'un ensemble de sites (6000) dont les chroniques ont été exploitées en ce qui concerne les tendances en nitrates dans la cadre d'une autre étude. Les données traitées sont des données acquises entre 1990 et 2007 pour la plupart (quelques unes sont plus anciennes).

A l'intérieur de cette base de données, les critères de choix de ces sites ont été les suivants :

- Répartition égale entre sites identifiés comme montrant des tendances significatives et ceux n'en montrant pas.
- Répartition des pentes de façon régulière sur la gamme des pentes estimées.

L'exploitation des tendances a été réalisée à l'aide du test de Mann Kendall.

A partir de ces 40 chroniques nitrates, les éléments suivants sont fournis dans le Tableau 13.

- Nombre de données de la chronique utilisées pour l'application du test de Mann Kendall
- Pente calculée par le test MK (en mg/l par jour)
- Concentration moyenne en NO₃ (mg/l)

- Tendence identifiée comme significative ou pas par le test de MK
- Nombre d'années couvertes par les données
- Détermination d'un écart-type résiduel σ_r (en % de la concentration moyenne) après application d'une régression linéaire. Cet écart-type permet d'accéder à la variabilité des données après prise en compte de la tendance identifiée pour les données. Cette variabilité intègre la variabilité analytique, l'échantillonnage et les variations naturelles hors tendance de long terme identifiée.
- Le rapport T_r/σ_r qui est utilisé dans certaines publications comme critère de fiabilité de tendance [10] [14]. T_r représente la variation de concentration entre le début et la fin de la chronique en utilisant la droite de tendance (pente annuelle * nombre d'années)
- La présence de variations cycliques ou pas (identifiées de façon qualitative et visuelle)
- A des fins de comparaison avec les objectifs fixés dans le paragraphe 6.1 sur les données simulées, la variation de concentration ramenée à une période de 6 ans.

Le principal paramètre qui sera exploité est le paramètre σ_r . C'est ce paramètre qui caractérise la dispersion des données. Les autres paramètres sont donnés pour information sur le type de chroniques utilisées.

Remarque : l'estimation de la variance résiduelle n'est pas pertinente dans tous les cas. En effet, pour les chroniques montrant des cyclicités marquées, l'analyse des données par régression linéaire, de même que la détermination de la variance résiduelle ne sont pas adaptées. Dans le Tableau 13 : tableau récapitulatif des caractéristiques des chroniques nitrates utilisées toutes les données ont cependant été conservées. Une indication qualitative de la présence plus ou moins importante de variations cycliques dans les chroniques est présentée. Ce sont essentiellement les données avec les plus faibles variances résiduelles qui seront exploitées. En effet l'hypothèse faite est que parmi tous les écarts-types estimés sur les chroniques testées, les valeurs minimales représentent la variabilité de données nitrates dans des contextes relativement stables en ce qui concerne le milieu. La prise en compte des phénomènes périodiques permettraient une meilleure estimation des valeurs de σ_r pour certains sites.

Surveillance de polluants dans les eaux souterraines : impact de l'incertitude de mesure sur l'évaluation des tendances

	Nombre de mesures pour les test	Pente (mg/l/jour)	NO3_Moy (mg/l)	Tendance identifiée comme significative	Nombre année	σ (%)	Cyclique	Tr/sres	Evolution sur 6 ans (%)	Commentaires qualitatifs visuels
Site 1	32	-0,0544	120,1	OUI	10				-99,2	Dispersion énorme - Tendance absente
Site 2	9	-0,0194	47,8	NON	4				-89,1	Chronique impossible à traiter. Données aberrantes ?
Site 3	11	-0,0184	63,1	OUI	5	11,8		-4,1	-63,8	
Site 4	6	-0,0184	105,3	NON	6	40,3		-0,9	-38,3	
Site 5	16	-0,0184	48,1	OUI	3	9,4		-3,7	-83,8	
Site 6	12	-0,0173	27,2	OUI	10	24,2		-9,3	-139,5	
Site 7	9	-0,0162	37,2	NON	5	24,2	+ ?	-3,1	-95,4	Forte dispersion. Présence de cyclicité ? Chronique difficile à traiter.
Site 8	14	-0,0150	37,8	OUI	4	5,8		-9,6	-86,9	
Site 9	7	-0,0146	61,4	NON	3	5,5	+++	-5,0	-52,0	Variation cyclique très marquée - attention pente identifiée n'a aucune signification. Meme chose pour Tr/s
Site 10	12	-0,0100	120,4	NON	3	7,6		-1,4	-18,1	
Site 11	36	-0,0098	34,3	OUI	8	14,2		-5,6	-62,8	
Site 12	9	-0,0098	38,3	NON	8	12,0		-6,0	-56,2	Pente aberrante estimée par le test Mann Kendall
Site 13	13	-0,0027	46,8	NON	9	11,2	++	-1,7	-12,8	
Site 14	10	-0,0027	23,7	OUI	7	22,7	+++	-1,3	-25,1	
Site 15	33	-0,0008	21,7	NON	8	11,5		-1,0	-8,1	
Site 16	10	-0,0003	36,1	NON	5	10,5		-0,2	-2,1	S res = 4,7 % après élimination de 2 valeurs aberrantes
Site 17	11	0,0002	23,7	OUI	8	7,4	++	0,3	1,5	
Site 18	15	0,0002	0,8	NON	4				46,2	Résultats inférieurs à LQ

Tableau 13 : tableau récapitulatif des caractéristiques des chroniques nitrates utilisées

Surveillance de polluants dans les eaux souterraines : impact de l'incertitude de mesure sur l'évaluation des tendances

	Nombre de mesures pour les test	Pente (mg/l/jour)	NO3_Moy (mg/l)	Tendance identifiée comme significative	Nombre année	σ (%)	Cyclique	Tr/sres	Evolution sur 6 ans (%)	Commentaires qualitatifs visuels
Site 19	15	0,0006	25,1	NON	8	14,3	++	0,5	5,5	
Site 20	36	0,0006	17,3	OUI	15	10,6		1,9	8,0	
Site 21	24	0,0009	27,2	NON	9	10,3		1,0	7,5	
Site 22	27	0,0009	27,1	OUI	15	13,9		1,3	7,6	
Site 23	12	0,0010	27,4	OUI	19	12,4		2,1	8,3	S res = 4,1 après élimination d'une valeur aberrante
Site 24	14	0,0011	24,8	NON	6	29,5		0,3	9,3	
Site 25	18	0,0011	27,8	NON	8	6,4	+	1,9	8,7	
Site 26	155	0,0011	33,4	OUI	21	10,0		2,5	7,3	
Site 27	12	0,0015	14,8	OUI	8	9,8		2,8	21,5	
Site 28	11	0,0015	61,8	NON	10	7,8		1,1	5,2	
Site 29	27	0,0015	25,0	OUI	13	9,6	++	3,1	13,3	
Site 30	9	0,0015	43,6	NON	10	22,9		0,6	7,7	En enlevant un ou 2 points qui paraissent non significatifs, visuellement pente forte observée. S=2,6 et tr/s=2,2
Site 31	14	0,0015	34,3	OUI	13	4,1	++	5,1	9,9	
Site 32	12	0,0015	29,6	NON	4	12,8		0,6	11,5	
Site 33	16	0,0025	17,6	NON	9	45,0		1,1	31,2	Forte dispersion
Site 34	20	0,0025	44,3	OUI	9	6,6		2,8	12,5	
Site 35	24	0,0034	25,7	OUI	9	10,7	+++	4,1	28,9	
Site 36	19	0,0035	40,0	NON	5	20,3		0,8	18,9	
Site 37	29	0,0047	22,5	OUI	12		+++		46,0	
Site 38	20	0,0047	49,3	NON	8				21,0	Chronique impossible à traiter. Variations "en cloche" sur la durée

Tableau 13 : tableau récapitulatif des caractéristiques des chroniques nitrates utilisées

Dans la Figure 6, les écarts-types résiduels estimés sur l'ensemble des chroniques nitrates sont reportés par ordre croissant. Cette présentation permet de voir que la plupart des σ_r sont compris entre 5 et 15%.

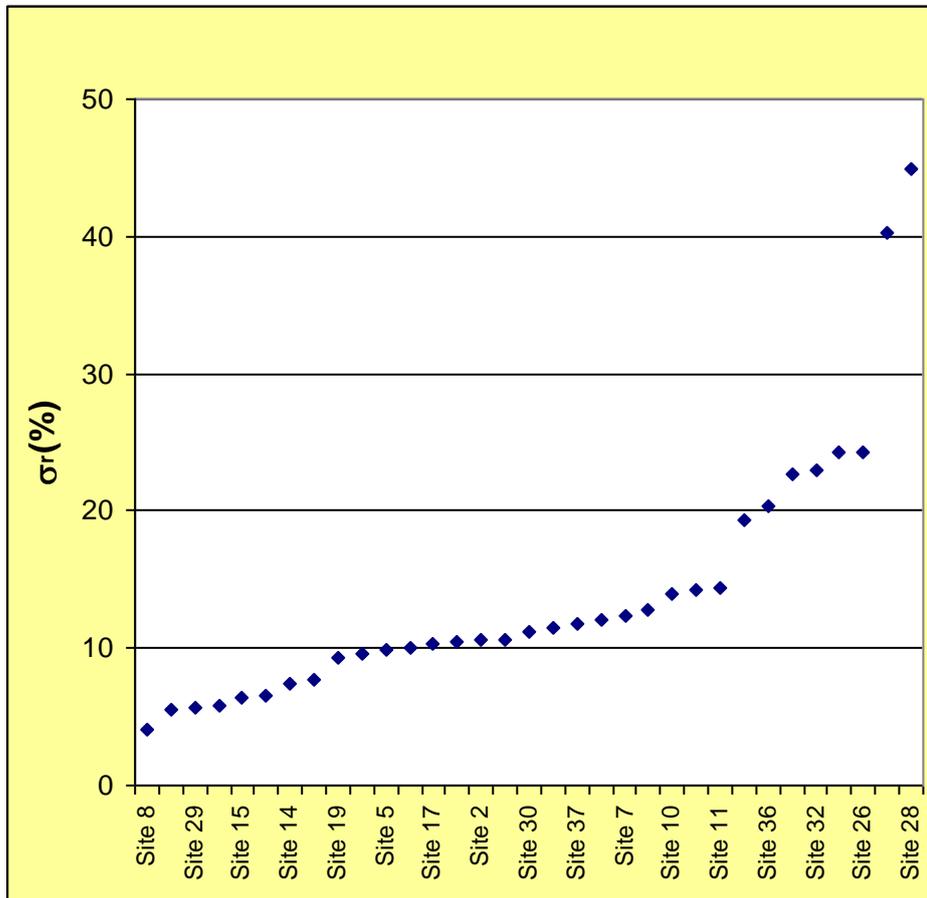


Figure 6 : Ecart-types résiduels (en %) estimés sur 38 chroniques temporelles de suivi de nitrates en eau souterraine (cf texte)

Les valeurs plus élevées ne sont pas significatives (prise en compte des variations cycliques, points aberrants qui entraînent une surestimation de la variance résiduelle ...). Par exemple pour le site 30 après élimination d'un point aberrant, le σ_r de 22.9% tombe à 6 % environ. Certaines estimations sont donc peu fiables c'est pour cela que nous nous intéresserons principalement aux valeurs les plus faibles qui tendent vers un minimum incompressible correspondant à la variabilité analytique.

Les valeurs les plus basses estimées sont d'environ 5-6%. Les données issues d'essais interlaboratoires en termes de dispersion des résultats ou bien encore les incertitudes classiques fournies par les laboratoires pour ce paramètre aux concentrations moyennes observées sont également de l'ordre de 5% ($k=1$). On a

donc de façon qualitative, une bonne adéquation entre la valeur minimale de σ_r observée et la dispersion analytique attendue.

Même si le terme de σ_r est un terme spécifique à un site donné, cette bonne adéquation entre les données issues de chroniques réelles avec les dispersions analytiques permet de considérer une valeur de 5-10% comme référence pour la variabilité des données nitrates. Ceci permet de conclure de façon qualitative que :

- Dans le cas de sites ne présentant pas de cyclicité marquée, la part de variations des données semble majoritairement expliquée par l'incertitude analytique pour ce paramètre.
- Cette fourchette de valeur de 5-10% pourrait être utilisée comme référence pour ce paramètre pour anticiper l'efficacité d'un programme de surveillance à partir des informations issues des simulations ou de la littérature (cf §5).

6.3.2. Données sur les pesticides

Ces exploitations ont été aussi effectuées sur d'autres paramètres mais sur un nombre plus restreint de données. La Figure 7 rassemble les estimations de variances résiduelles de 26 chroniques temporelles de suivi de 4 pesticides : atrazine, DIA, DEA et Simazine. Pour ces composés, les incertitudes variant plus rapidement en fonction de la concentration compte tenu de niveaux mesurés relativement proches des limites de quantification, les résultats sont présentés en fonction de la concentration plutôt que par ordre croissant comme pour les nitrates.

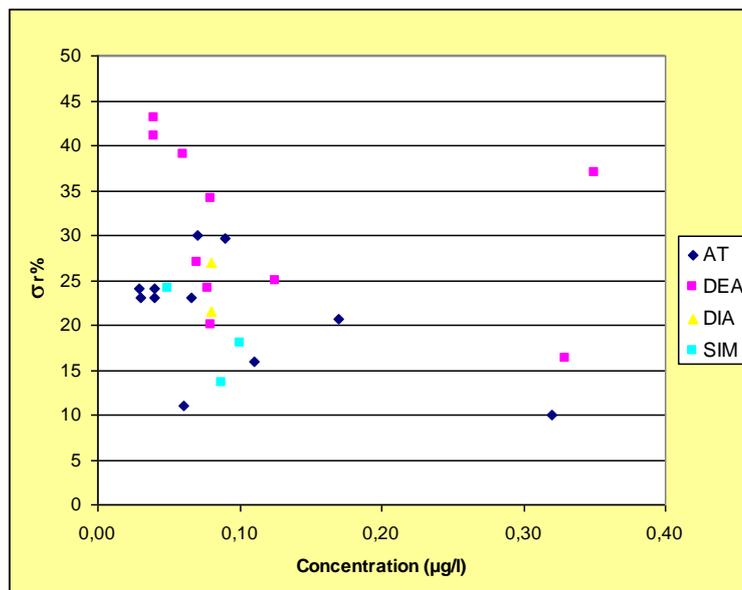


Figure 7 : Ecarts-types résiduels (en %) estimés sur 12 chroniques temporelles de suivi de pesticides en eau souterraine en fonction de la concentration (cf texte)

Pour ces paramètres organiques les écarts-types résiduels sont globalement plus élevés que pour les nitrates. Ils sont de l'ordre de 20-25% ce qui correspond aux dispersions analytiques observés classiquement lors d'essais d'intercomparaison entre laboratoires.

6.3.3. Données sur d'autres paramètres

Le Tableau 14 donne quelques exemples d'estimation d'écarts-types résiduels pour d'autres paramètres. Ces estimations ne sont données que pour trois ou quatre sites. Elles ne sont donc pas nécessairement représentatives. Il serait nécessaire, si la démarche se révèle utile, de l'appliquer à un plus grand nombre de sites.

Paramètre	Ecart-type résiduel (%)	Concentration moyenne
COD site 1	43	0.6
COD site 2	31	0.6
COD site 3	38	1.9
Fe site 1	65	40
Fe site 2	58	45
Fe site 3	40	58
As site 1	53	82
As site 2	44	80
Zn site 1	59	200
Conductivité site 1	8.3	310
Conductivité site 2	2.5	591
Conductivité site 3	6.4	552
Conductivité site 4	8.3	512

Tableau 14 : Ecart-types résiduels en % estimés sur des chroniques temporelles de suivi de différents paramètres en eau souterraine

Ces quelques données très préliminaires montrent:

- Pour le COD (carbone organique dissous) ainsi que pour la conductivité (site 2) des variabilités proches des incertitudes analytiques classiquement observées
- Pour les métaux, les variabilités observées sont nettement supérieures aux incertitudes analytiques classiques (de l'ordre de 10%). Ceci peut être dû soit à des variabilités naturelles pour ces paramètres sur les sites considérés, soit à des variabilités supplémentaires liées à l'échantillonnage. En ce qui concerne le fer, cette deuxième explication est sans doute à prendre en compte car des variations très importantes de concentration peuvent être liées à des différences de pratiques d'échantillonnage (filtration ou pas, filtration sur le terrain ou au laboratoire, durée avant filtration au laboratoire, acidification avant ou après filtration, ...).

6.3.4. Synthèse

Avec l'ensemble de ces données qui doivent être confirmées par l'exploitation de plus nombreuses chroniques, on pourrait ainsi :

- Disposer d'estimation de variances globales des données pour certaines familles de paramètres ce qui permettrait de dimensionner des programmes de surveillance répondant à des objectifs précis. En particulier, le nombre d'années au bout duquel il sera envisageable de détecter une tendance d'une pente donnée peut être calculé.
- Connaissant une estimation même grossière de la part relative de l'incertitude analytique dans la variabilité globale de la donnée, on pourrait alors vérifier dans quelle mesure, une diminution de cette incertitude analytique, si elle est possible techniquement et financièrement, permettrait d'améliorer les performances du programme de surveillance si cela est nécessaire.

La plupart des éléments décrits ci-dessus relatifs à l'exploitation de chroniques réelles ont pour objectif de déterminer la variabilité globale des données et la part de l'incertitude dans cette variabilité afin de déduire les performances de futurs programmes de surveillance ou bien de les concevoir et éventuellement d'adapter les méthodes d'analyse afin de diminuer leur incertitude si cela s'avérait nécessaire : analyse prospective

A l'inverse, on peut se demander si les tendances identifiées par les tests statistiques sont fiables au regard des incertitudes de mesure ou plus largement de la variabilité des données : analyse rétrospective

Dans cet objectif, les éléments pouvant être utilisés sont les suivants :

- Le degré de signification du test statistique
- Le rapport Tr/σ

Dans la publication de Bouchard & al [10], ce facteur est utilisé comme objectif du programme pour l'identification de tendances fiables. L'objectif était de pouvoir détecter des tendances inférieures ou égales à l'écart-type résiduel ($Tr/\sigma < 1$) en 7 à 10 ans et avec une puissance de test de 90%. Après quelques années de surveillance, les auteurs se sont rendu compte que la plupart des tendances identifiées de façon fiable par les tests statistiques respectaient ce critère. L'exploitation des rapports Tr/σ pour les chroniques nitrates de ce chapitre est présentée dans la Figure 8 en fonction du caractère significatif (S) ou pas (NS) sur le plan statistique de la tendance. Mis à part deux chroniques présentant des résultats aberrants (sites 7 et 9, voir commentaires), les rapports Tr/σ sont inférieurs à 2 pour les chroniques ne présentant pas de

tendances significatives et supérieurs à 1.5 majoritairement dans le cas contraire. La chronique présentant une tendance significative et un Tr/s inférieur à 1.5 est particulière. Le traitement préalable de données effectué pour l'exploitation Mann Kendall a réduit l'intervalle utilisé pour l'identification de la tendance. Sur cet intervalle, une légère cyclicité est observée d'où un σ surestimé. Sur une période plus longue le rapport Tr/σ estimé est de 9 (forte tendance à la hausse).

- La puissance du test (ou sa robustesse)

Que ce soit en s'appuyant sur certains tableaux issus de la synthèse bibliographique, sur des calculs statistiques ou des simulations, il est possible pour une chronique donnée de déterminer la puissance du test effectuée, c'est-à-dire la probabilité dans les conditions de la chronique qu'une tendance soit détectée par le test statistique au niveau de signification choisi. Par exemple, pour le site 27, la durée de la chronique considérée est de 8 ans (avec $n=12$ sur cette période), l'écart-type résiduel est de 10% et la pente est de $0.0015 \cdot 365 / 14.8 = 4\%$ par an. Dans ces conditions, en utilisant le tableau du 5.2.3 (p 55), la puissance estimée est d'environ 60% (sans doute un peu plus élevée car le nombre de données $n=12$ est supérieur au nombre de données pris comme hypothèse pour le tableau : $n=8$). Cette probabilité est donc faible par rapport aux probabilités prises comme objectif dans beaucoup de publications (90%).

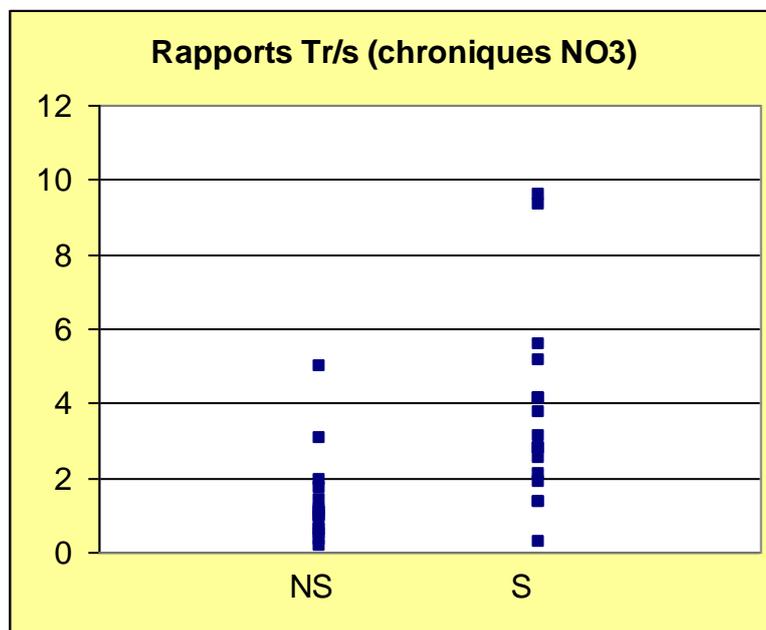


Figure 8 : répartition des rapports Tr/σ pour les chroniques nitrates en fonction du caractère significatif ou pas de la tendance identifiée par le test de Mann Kendall

Les paramètres Tr/σ ou bien la puissance du test, qui prennent en compte la variabilité des données (et donc pour une part plus ou moins importante l'incertitude de mesure)

de façon plus explicite que le test lui-même, ne viennent pas contredire le résultat du test qui est un résultat en lui-même basé sur un niveau de confiance statistique. Ils viennent uniquement le compléter et préciser sa robustesse.

7. Pistes pour améliorer et mieux maîtriser les incertitudes de mesure

L'impact que peut avoir l'incertitude de mesure sur l'évaluation des tendances a été illustré à plusieurs reprises dans ce rapport. (cf 5 et les références [8][11][13][14]). Un dernier exemple parlant qui illustre comment un changement de méthode peut impacter l'évaluation des tendances ou d'une manière plus générale la compréhension des mécanismes mis en jeu dans un écosystème, est la question de l'origine de l'acidification des ruisseaux et des lacs par les dépôts atmosphériques ou par les processus naturels. En effet, une controverse sur l'origine de cette acidification a pris naissance dans les années 80 suite à un changement de méthode analytique. Lorsque l'analyse des sulfates n'a plus été réalisée par colorimétrie mais par chromatographie ionique, un déficit anionique a été observé pour certains échantillons, de l'ordre de 10 à 30 %, fortement corrélé à la teneur en matière organique. Gorham et Detenbeck [15] ont montré que la surestimation de la teneur en sulfate déterminée par colorimétrie était le résultat d'interférences spectrales dues à certains anions organiques. Ils ont aussi montré comment ces observations pouvaient remettre en question les causes d'acidification de zones marécageuses.

En fonction d'objectifs à définir au niveau national pour l'évaluation des tendances, les chapitres précédant donnent des pistes pour estimer la variabilité requise des données ou l'incertitude maximale nécessaire. Dans certains cas, il se pourrait que les incertitudes actuelles des laboratoires ou plus largement de la chaîne de mesure (incluant les opérations d'échantillonnage et d'analyse) soient insuffisantes pour atteindre les objectifs fixés pour l'évaluation de tendance. Les solutions seraient alors, soit de redéfinir ces objectifs, soit de diminuer les incertitudes de mesure si cela est possible. Ce chapitre a pour vocation de dresser un bilan des étapes critiques de la chaîne de mesure et de proposer des pistes permettant de mieux maîtriser chaque étape du processus de mesure, depuis le prélèvement jusqu'au rendu de résultat.

Les notions primordiales de traçabilité métrologique et d'incertitude sont abordées dans ce chapitre.

7.1. TRAÇABILITE METROLOGIQUE

Deux résultats sont comparables non pas parce qu'ils sont du même ordre de grandeur, mais parce qu'ils peuvent être traçables aux unités d'un même système de référence.

Selon la définition du VIM (Vocabulaire International de Métrologie) [26], la traçabilité métrologique est la propriété d'un résultat de mesure selon laquelle ce résultat peut

être relié à une référence par l'intermédiaire d'une chaîne ininterrompue et documentée d'étalonnages dont chacun contribue à l'incertitude de mesure.

Les éléments clés pour établir la traçabilité d'un résultat de mesure sont, d'après EURACHEM [16] :

- La spécification du mesurande, de la portée de la mesure et du niveau d'incertitude demandée
- Le choix d'une méthode appropriée pour l'estimation de la valeur, à savoir une procédure de mesure avec un calcul associé – à savoir une équation – et les conditions de mesure
- La démonstration, à travers la validation, que les calculs et les conditions de mesure incluent toutes les « grandeurs » qui peuvent influencer le résultat ou la valeur assignée à une référence
- L'identification de l'importance relative de toute grandeur d'influence
- L'utilisation des étalons appropriés
- L'estimation de l'incertitude

7.2. INCERTITUDE DE MESURE

L'incertitude est un des points critiques qui entre dans les performances d'une méthode. Le niveau d'incertitude, ainsi que la méthode d'estimation employée sont des informations importantes qui donnent de l'information sur la capacité de la méthode à déterminer l'espèce chimique recherchée et qui doivent par conséquent être maîtrisées, connues et documentées. Différents documents traitent de l'estimation de l'incertitude de mesure : le GUM [17] qui constitue le document de référence, le guide Eurachem/CITAC [18] et la norme XP T90-220 [19] qui sont plus spécifiquement dédiés aux analyses chimiques.

Une synthèse de quelques incertitudes estimées en fonction de la méthode analytique choisie est présentée dans le Tableau 15 pour des paramètres comme les nitrates, les métaux et les pesticides (triazines et phénylurées) dans les eaux:

Paramètre	NQE / seuil réglementaire	Méthode analytique	Incertitude élargie (k=2)	Source
Nitrates	50 mg/L	Electrophorèse Capillaires	5-7%	Labo A
		Chromatographie ionique	10 %	Labo B
Métaux	0.1 – 20 µg/L	ICP-OES, SAA Four	5-15%	Labo A
		ICP-MS	15%	Labo B
(Cd, Ni, Pb)		ICP-MS, SAA Four	0.5-45 %	Euramet 924, Partie III
(Cd, Ni, Pb)		ICP-MS (dilution isotopique)	1-3 %	LNE
Triazines et phénylurées	0.01 – 1 µg/L	CPG MS	10 - 25%	Labo A
		HPLC MSMS	30%	Labo B
		SPE/DI/LC-MS ²	1-5 %	Fiches méthodes Aquaref MA17 / MA 18 (LNE)

Tableau 15 : Synthèse de quelques incertitudes types estimées dans des eaux par 2 laboratoires prestataires (A et B), un laboratoire de référence (LNE) et par différents laboratoires participant au programme EURAMET 924 pour 3 familles de paramètres et pour différentes méthodes d'analyse.

Les données utilisées dans ce tableau proviennent de 2 laboratoires prestataires (A et B), de la comparaison interlaboratoires réalisée dans le cadre du programme Euramet 924 (Part III) [25] pour les métaux (incluant des laboratoires prestataires et laboratoires de référence) et de données du LNE (laboratoire de référence).

Il apparaît que pour les nitrates, les incertitudes de 2 laboratoires pris comme exemple sont relativement faibles, de l'ordre de 5-10%. Ces ordres de grandeur sont typiques d'autres informations obtenues à partir de résultats issus d'inter comparaison française notamment les circuits d'inter comparaison AGLAE et BIPEA.

En comparaison avec les méthodes analytiques utilisées en routine, les méthodes développées par les laboratoires de métrologie, basées sur la dilution isotopique, permettent d'obtenir des incertitudes très faibles, de l'ordre de 1 à 5 % pour les micropolluants organiques et métalliques.

En ce qui concerne les métaux, l'étude réalisée dans le cadre d'Euramet 924, Partie III, montre que, pour un même élément, les estimations d'incertitude des laboratoires couvrent une grande plage, de 0.5 à 45 %. Cette dispersion peut provenir à la fois de

différences au niveau du protocole analytique utilisé ou bien de la méthode d'estimation des incertitudes.

Exemple pour la détermination du cadmium, du nickel et du plomb dans l'eau : Projet EURAMET 924 (Partie III) [25]

Le projet EURAMET 924 visant à démontrer que la traçabilité aux unités du système international permet d'assurer la comparabilité des données a été mis en œuvre au niveau européen. La partie III de ce projet concernait l'analyse dans une eau souterraine de quatre métaux prioritaires de la Directive Cadre Européenne sur l'Eau (DCE) à des concentrations proches des normes de qualité environnementales (NQE).

Trois niveaux de laboratoires ont été considérés : les laboratoires de métrologie qui assurent la traçabilité au système international, les laboratoires experts qui agissent comme intermédiaire dans la chaîne de traçabilité et enfin les laboratoires de routine qui réalisent les analyses dans le cadre des programmes de surveillance de la DCE.

Les résultats obtenus pour le cadmium, le nickel et le plomb sont regroupés dans le Tableau 16. Il est important de noter que la valeur de référence a été estimée par gravimétrie.

	Cadmium				Nickel				Plomb			
	\bar{x} ng/L	N 1	s/\bar{x} %	U ng/L	\bar{x} µg/L	N 1	s/\bar{x} %	U ng/L	\bar{x} µg/L	N 1	s/\bar{x} %	U ng/L
Valeur de référence	126.6			3.0	27.15			0.94	12.313			0.069
Tous les participants	127.3	99	36	9.2	26.10	116	11.5	0.56	12.28	115	21.6	0.49
Instituts de métrologie	125.6	3	1.1	1.6	27.37	3	1.4	0.44	12.23	3	1.0	0.14
Laboratoires experts	125.3	12	15	10.5	26.28	13	3.3	0.48	12.24	13	2.9	0.20
Laboratoires de routine	127.6	84	39	11.0	26.04	100	12.3	0.64	12.28	99	23.3	0.57

Tableau 16 : Synthèse des résultats obtenus lors de la comparaison interlaboratoire Euramet 924 Partie III (\bar{x} : concentration moyenne ; N : nombre de laboratoires ; s/\bar{x} : coefficient de variation ; U : incertitude élargie estimée à partir de l'équation $U = k s / \sqrt{N}$ avec $k = 2$)

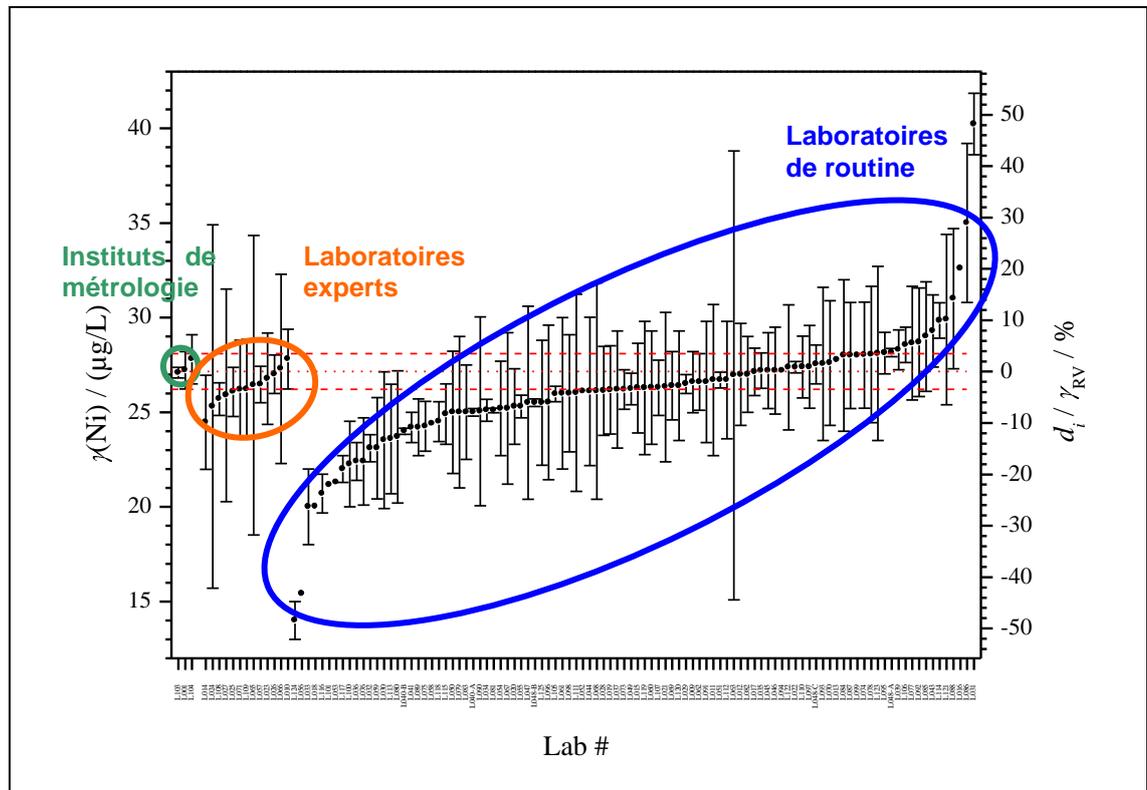


Figure 9 : Exemple pour le nickel des résultats obtenus lors de l'essai d'intercomparaison réalisé dans le cadre du projet Euramet 924, Partie III.

La comparaison interlaboratoire réalisée dans le cadre de ce projet Euramet 924, Partie III, a permis de voir que la dispersion des valeurs est croissante en fonction du type de laboratoire considéré. En effet, pour le nickel par exemple, la dispersion (s/\bar{x}) des valeurs des laboratoires de métrologie est de 1.4%, alors qu'elle est de 3.3 % pour les laboratoires experts et enfin de 12.3% pour les laboratoires de routine (voir Figure 9). Pour le cadmium, l'écart entre les instituts de métrologie, les laboratoires experts et les laboratoires de routine est encore plus marqué, avec des dispersions respectives de 1.1 %, 15 % et 39 %.

En ce qui concerne la justesse, 64-66% des laboratoires ont des résultats satisfaisants au regard des critères définis pour l'essai (basés sur l'écart normalisé EN [27]). Cependant, des différences apparaissent selon le type de laboratoire considéré. Ainsi, les laboratoires experts ont de meilleures performances que les laboratoires de routine et les performances des instituts de métrologie sont largement supérieures aux laboratoires experts.

Enfin, pour cet essai, il était demandé à l'ensemble des laboratoires de fournir leur résultat avec une incertitude associée. Entre 8 et 10 % des laboratoires ont rendu leurs résultats sans incertitude associée pour le cadmium, le nickel et le plomb.

La compilation des incertitudes fournies par les laboratoires permet d'établir une distribution selon 4 grandes classes, montrant que la majorité des incertitudes se situent entre 5 et 20 %, avec cependant 20 à 30 % des laboratoires qui ont une incertitude inférieure à 5 % (Figure 10)

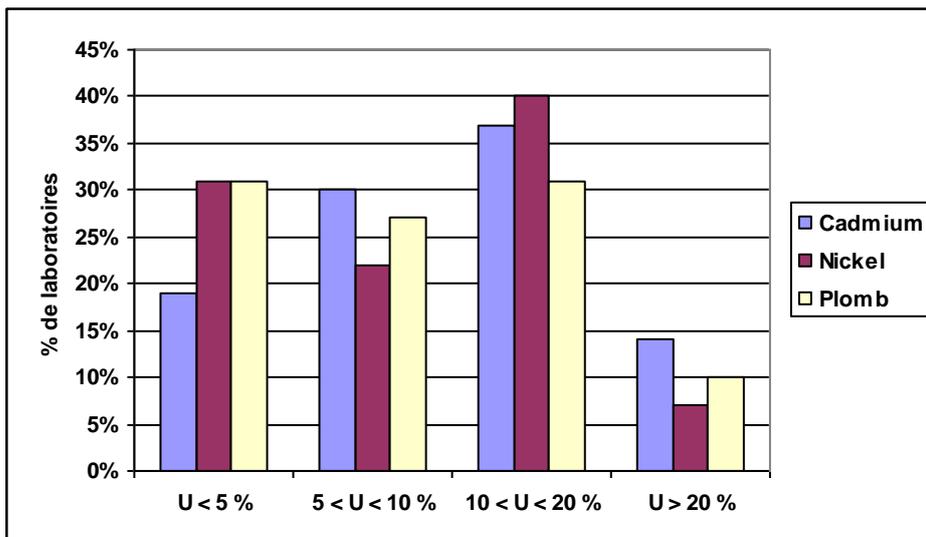


Figure 10 : Distribution de l'incertitude élargie en quatre classes pour le cadmium, le nickel et le plomb, d'après l'étude Euramet 924, Partie III [25]

Cette étude illustre le fait qu'il y a un réel besoin de mieux maîtriser les protocoles analytiques et d'harmoniser la méthode d'estimation des incertitudes afin de réduire cette plage d'incertitude. Ceci est particulièrement important si un des critères de sélection de la méthode analytique est l'incertitude de mesure (par exemple pour des objectifs ayant trait à l'évaluation de tendances). En effet, en se basant sur les résultats de cette étude, il semble difficile pour le moment de sélectionner des laboratoires prestataires sur la base de leur estimation d'incertitude comme indicateur de performance d'une méthode.

7.3. PISTES D'AMELIORATION

7.3.1. Analyse du processus de mesure

La première étape essentielle consiste à établir clairement et sans ambiguïté ce qui doit être mesuré et de donner une expression quantitative reliant la valeur du mesurande aux paramètres dont il dépend.

L'analyse du processus de mesure permet, dans un deuxième temps, de recenser les différentes sources élémentaires d'incertitude contribuant à l'incertitude globale. Une première analyse des facteurs d'influence permet d'identifier les facteurs susceptibles d'avoir une contribution significative sur le résultat de mesure.

La technique des 5 M permet, à partir d'une réflexion approfondie et d'une très bonne connaissance du processus de mesure, de recenser toutes ces causes. Cette technique 5M consiste à se poser 5 questions : quels sont les facteurs liés aux moyens, à la méthode, au milieu, à la matière et à la main d'œuvre qui influencent le résultat de mesure ?

Le but étant ensuite de réduire l'influence des facteurs qui auront été identifiés. Les méthodes pour diminuer l'influence de ces facteurs sont les suivantes :

- Les facteurs de nature aléatoire (erreurs aléatoires) sont généralement diminués en répétant les mesurages et en calculant la moyenne arithmétique des lectures ou des mesures.
- Les facteurs de nature systématique (erreurs systématiques) sont diminués en appliquant des corrections.

Un exemple de diagramme des 5 M est présenté Figure 11.

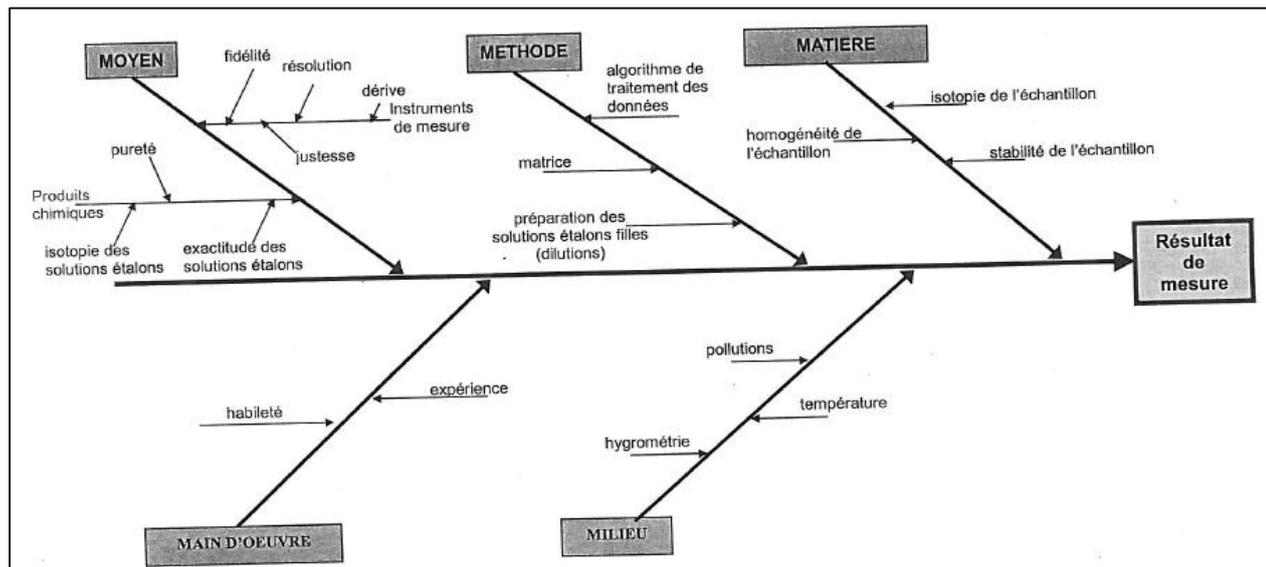


Figure 11 : Diagramme des 5 M pour l'analyse du plomb par ICP-MS dans une eau d'alimentation [20]

Par exemple, les différentes sources identifiées pour l'analyse du plomb par ICP-MS dans une eau d'alimentation sont (Figure 12):

- La fidélité des instruments de mesure
- La dérive des instruments de mesure
- La justesse des instruments de mesure
- L'exactitude des solutions étalons mères
- L'exactitude des solutions étalons filles
- La matrice de l'échantillon
- L'algorithme de traitement des données

Le modèle du processus de mesure établi en tenant compte de l'ensemble des facteurs ayant une influence significative sur le résultat final est le suivant :

$$C_{Pb} = \text{moyenne}_{3 \text{ mesures}} \times f_{\text{dérive}} \times f_{\text{exactitude des solutions}} \times f_{\text{justesse de la méthode}}$$

La quantification des différentes contributions à l'incertitude permet de voir que pour diminuer l'incertitude globale, ce sont la dérive de l'appareil et l'exactitude des solutions sur lesquelles il faut porter les efforts en premier pour diminuer leur contribution.

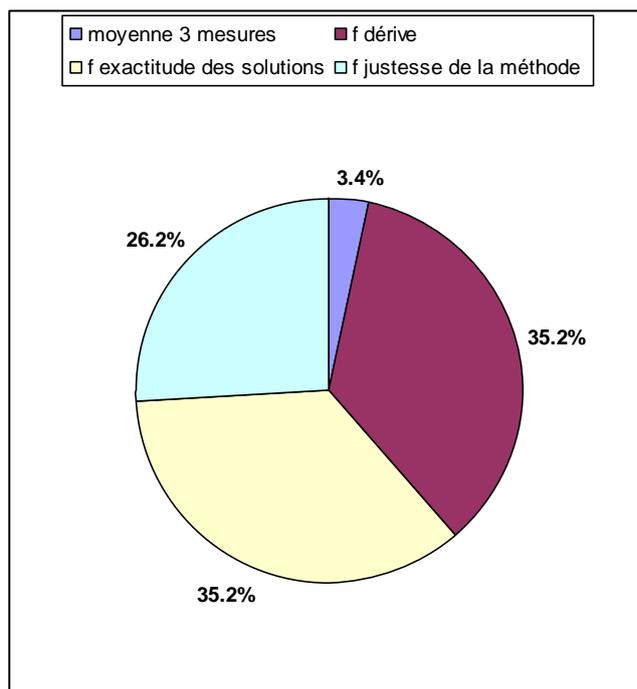


Figure 12 : Budget d'incertitude pour l'analyse du plomb par ICP-MS dans une eau d'alimentation [20]

7.3.2. Identification des contributions les plus importantes et exemple de mise en place d'action possible pour diminuer ces contributions

La détermination des concentrations en espèces chimiques dans les eaux suit un protocole qui comprend plusieurs étapes, certaines sur le terrain et d'autres en laboratoires (voir Figure 13). Chaque étape doit être maîtrisée afin de garantir la fiabilité des données acquises.

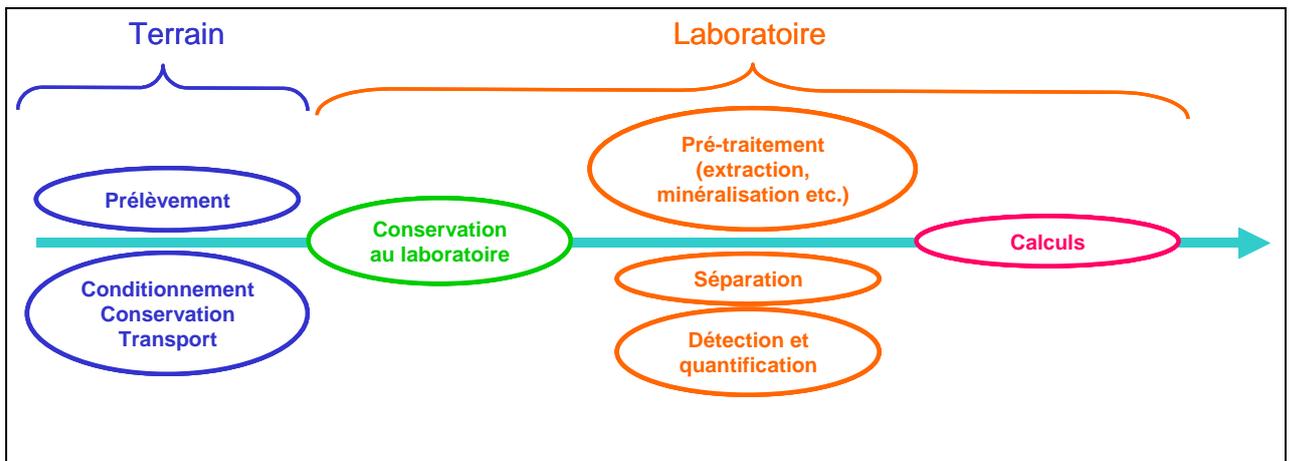


Figure 13 :Schéma de la détermination de la concentration en espèces chimiques dans les eaux

Des exemples illustrant l'importance de bien maîtriser chaque étape du protocole analytique ainsi que quelques pistes pour réduire les erreurs aléatoires ou systématiques sont présentés dans la suite de ce paragraphe.

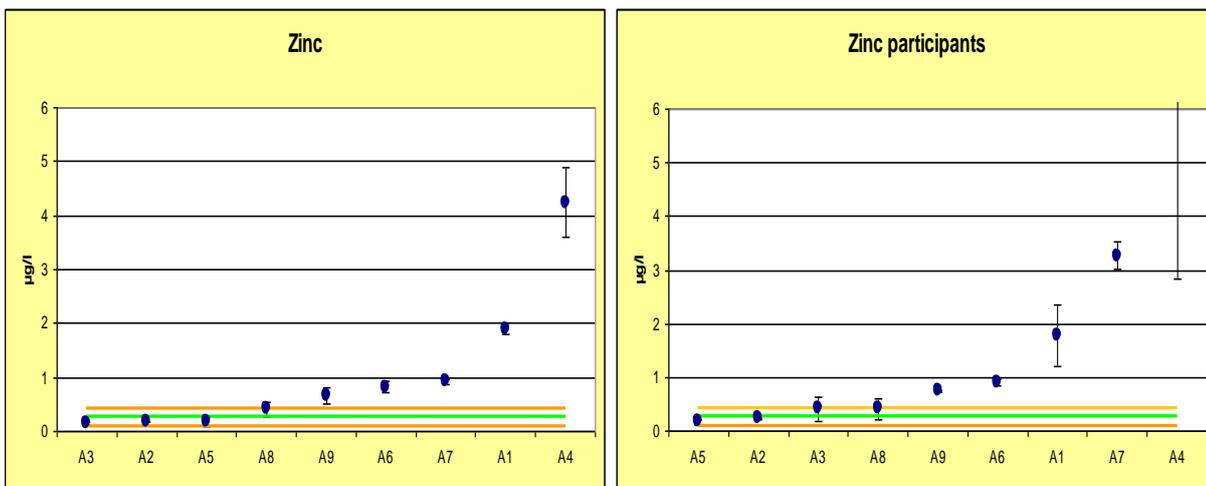
❖ Prélèvement et conditionnement

Il est primordial de maîtriser les protocoles de prélèvement et de conservation des échantillons. En effet de cette étape va découler le résultat du protocole analytique mis en œuvre au laboratoire.

A titre d'exemple, un essai collaboratif sur l'échantillonnage en eau souterraine a été réalisé en 2009 par le BRGM et le LNE dans le cadre d'AQUAREF [21]. Un des objectifs était d'évaluer l'impact des opérations de prélèvement sur la variabilité des résultats. 9 équipes de préleveurs ont réalisé selon leur procédure habituelle des prélèvements en double en vue d'analyse physico-chimiques. Les échantillons ont ensuite été pris en charge par le BRGM qui a réalisé les analyses.

Un effet « préleveur » significatif a pu être mis en évidence pour le zinc et le cuivre. Les figures 14a et 14b montrent que pour les préleveurs A9, A6, A7 et surtout A1 et A4, les concentrations mesurées dans les échantillons prélevés et conditionnés dans des flacons fournis par le BRGM sont significativement supérieures aux analyses de contrôle effectuées par le BRGM ainsi qu'aux résultats obtenus par les laboratoires A3, A2 et A5.

Par ailleurs, pour les métaux, un double système a été mis en place pour le flaconnage : flacons, matériel de filtration et acide fournis par le BRGM ou par chaque participant. Un effet flaconnage/matériel/acide a ainsi été mis en évidence pour le zinc et le cuivre : une dispersion plus importante pour certains participants (A3, A8, A1, A4), ainsi qu'une concentration moyenne mesurée plus forte pour trois d'entre eux (A3, A4, A7) ont été observées. Ceci montre bien que l'impact du flaconnage peut être non négligeable et qu'il est nécessaire de bien maîtriser cette étape de conditionnement de l'échantillon prélevé. Il est évident que de telles dispersions de résultat dans le cadre de programmes de surveillance destinés à l'évaluation de tendance auraient un effet désastreux sur la capacité à détecter ces tendances. Bien évidemment les effets peuvent également être importants dans le cadre d'une comparaison ponctuelle à des valeurs seuils (NQE par exemple).



❖ Filtration

Quand le programme de surveillance impose une analyse sur la fraction dissoute, les échantillons prélevés doivent être filtrés (cas des métaux pour les eaux de surface dans la DCE). Cette filtration peut être réalisée, d'un point de vue technique, sur le terrain immédiatement après l'échantillonnage ou bien à réception au laboratoire.

Dans le cadre du programme AQUAREF 2010, des essais ont été réalisés sur 3 types d'eau (cours d'eau, plan d'eau et eau souterraine) afin de tester l'influence du délai entre échantillonnage et filtration pour l'analyse des métaux dissous.

Pour chaque site sélectionné, une comparaison a été effectuée entre les résultats obtenus sur l'échantillon filtré et acidifié au moment du prélèvement et ceux obtenus sur l'échantillon filtré et acidifié au laboratoire avant l'analyse. Afin de mettre en évidence une évolution possible de la concentration des analytes en fonction du temps, la filtration au laboratoire a été effectuée après 24h, après 3 jours et après une semaine.

Les Figure 15 et Figure 16 illustrent des résultats issus de cette étude AQUAREF pour le plomb en eau de surface et pour le fer en eau souterraine.

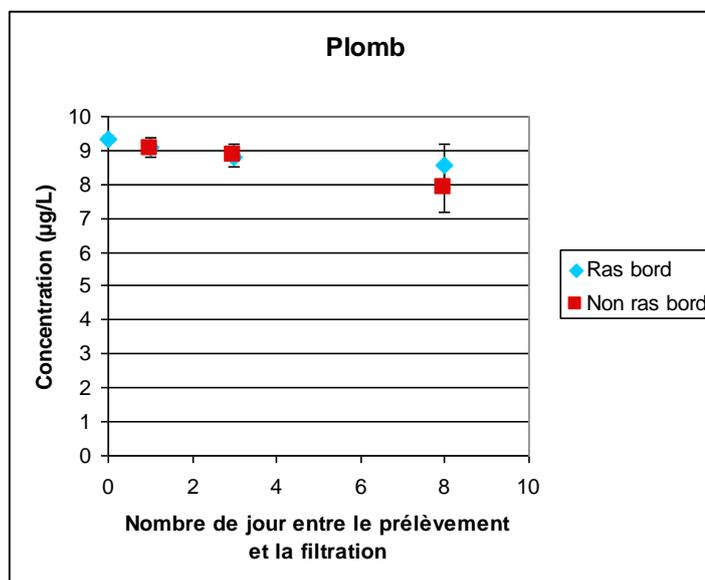


Figure 15 : Concentrations moyennes (en µg/L) et écart type pour le plomb sur la Deule à Don en fonction du nombre de jours entre le prélèvement et la filtration (Etude Aquaref en cours). Une distinction entre des flacons remplis à ras bord et des flacons remplis au ¾ a été faite

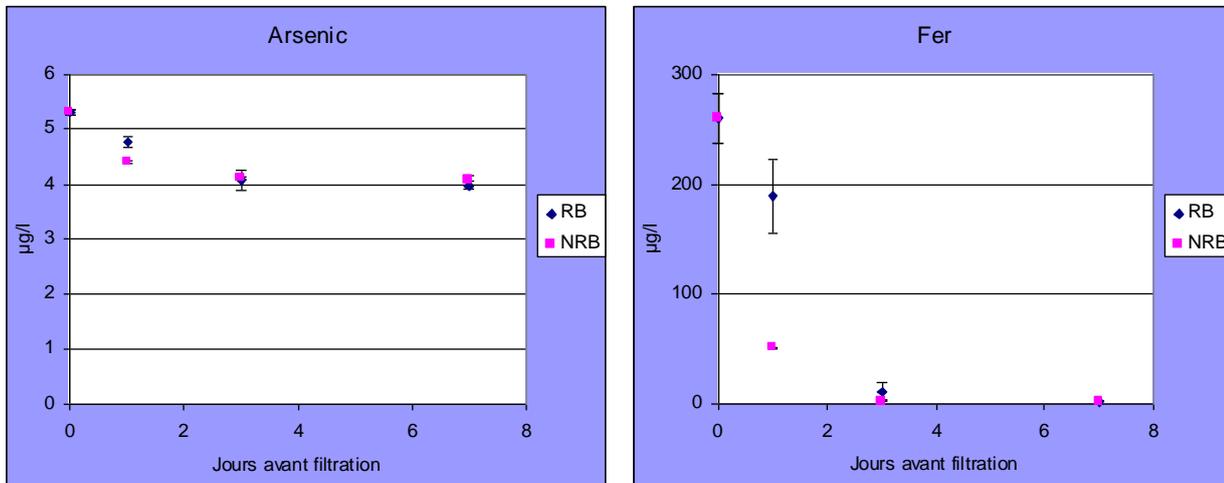


Figure 16 : Evolution des concentrations moyennes (en µg/L) en arsenic et en fer dans une eau souterraine en fonction du nombre de jours entre le prélèvement et la filtration (Etude Aquaref à paraître). Une distinction entre des flacons échantillonnés remplis à ras bord (RB) et des flacons remplis au ¼ (NRB) a été faite

Une diminution significative de la teneur en plomb, de 12% après une semaine, a pu être mise en évidence sur la Deule à Don (59). En ce qui concerne les eaux souterraines, la teneur en fer II dissous chute dramatiquement en 24h sur un des sites choisis. Ainsi, la cinétique d'évolution d'un échantillon non stabilisé est très variable et dépend fortement des conditions physico-chimiques de chaque site (conditions d'oxydoréduction pour les eaux souterraines, teneurs en matière en suspension et carbone organique pour les eaux de surfaces etc.). La recommandation dans le cas d'une analyse sur dissous est donc de filtrer au plus vite, de préférence sur site pour minimiser l'évolution de l'échantillon prélevé.

Il apparaît également clairement que dans un objectif de minimiser la variabilité globale sur les données acquises, l'harmonisation des pratiques concernant la filtration (harmonisation vers une filtration maîtrisée sur site) serait nécessaire.

❖ Conservation

Le délai de conservation des échantillons (délai entre échantillonnage et analyse) est également, pour certaines substances, un facteur prépondérant de la qualité de l'analyse. Il s'agit donc là également d'un paramètre critique à maîtriser afin de limiter la dispersion des résultats.

Une étude de stabilité peut être réalisée pour le type d'eau considéré (eaux de consommation, eaux de surface, eaux souterraines, eaux résiduaires etc.) afin de

s'assurer qu'aucune évolution significative des paramètres analysés n'a eu lieu entre le prélèvement et l'analyse en laboratoire.

❖ Pureté des produits

La pureté des produits utilisés, et notamment des solutions mères pour la préparation des gammes d'étalons peut engendrer des erreurs systématiques sur l'analyse et donc potentiellement des écarts entre résultats de différents laboratoires. Un exemple pour les PFOS permet d'illustrer l'importance du choix des produits utilisés (Figure 17).

Une comparaison interlaboratoire a été organisée par dans le but d'évaluer les performances de 21 laboratoires américains et européens pour l'analyse des PFOS dans l'eau et les poissons. L'intérêt de cette comparaison est qu'elle a permis notamment d'étudier les étalons commerciaux utilisés par les différents laboratoires. Pour cela les laboratoires ont utilisé une solution étalon fournie par l'organisateur (Wellington Laboratories), appelée solution de référence, pour la quantification et l'ont comparée à celle qu'ils utilisent en routine.

Les résultats montrent que la moyenne du rapport entre la solution étalon de référence et les solutions étalons utilisées par les laboratoires en routine est globalement entre 95 et 105 %, tous composés confondus. Cependant, des différences importantes ont été observées pour quelques laboratoires : certaines solutions étalons comme Fluorochem et Acros sous estiment de 72-85% alors que d'autres surestiment de 116% (Alta Aesar) par rapport à la solution de référence.

Par ailleurs cette étude a permis aussi de montrer d'une part qu'il pouvait exister un effet significatif entre lot pour un même fabricant, et d'autre part que selon les composés analysés, la sous estimation ou la surestimation n'était pas identique pour une même solution étalon.

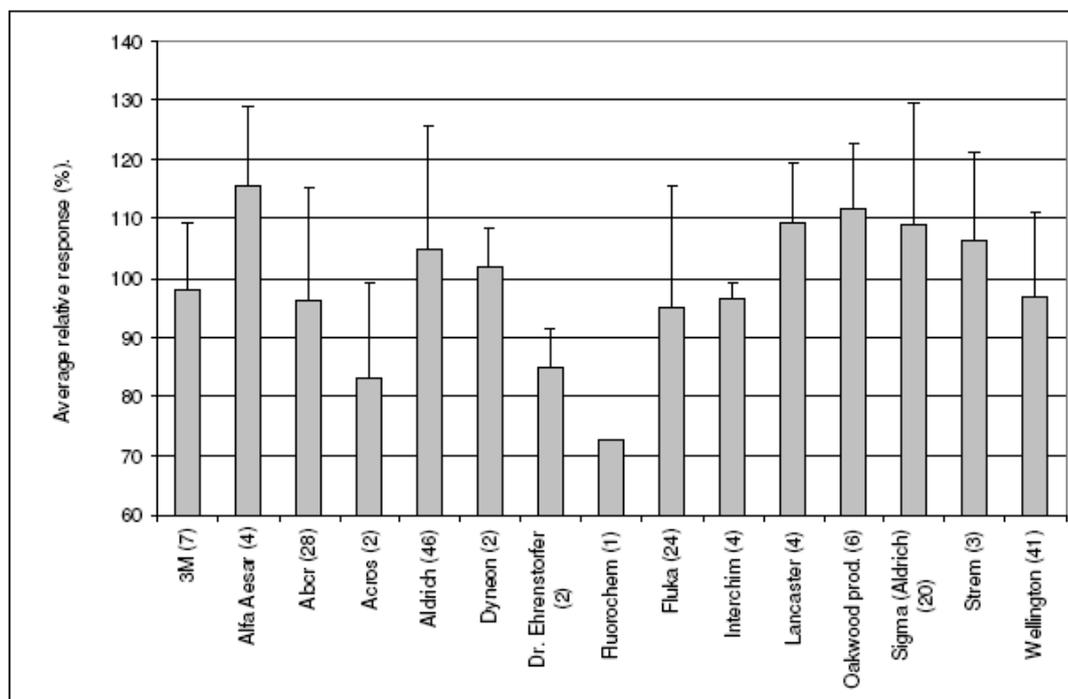
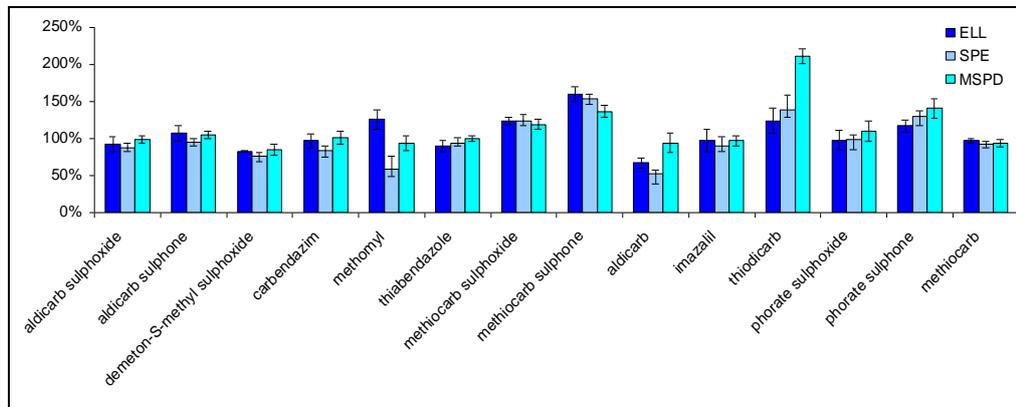


Figure 17 : Variabilité des solutions commerciales pour les PFOS (tous composés confondus), d'après [22]

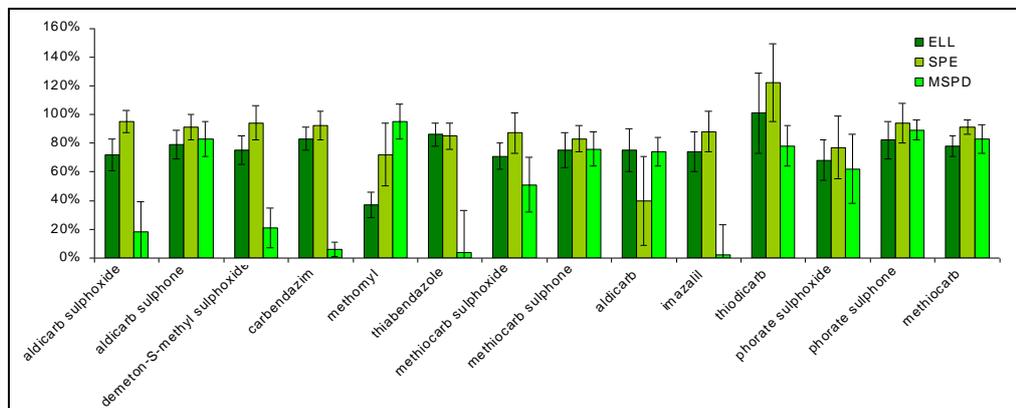
❖ Extraction et effets de matrice

Kruve et al. [23] ont étudié l'influence des effets de matrices sur l'analyse (partie instrumentale) et sur l'extraction (préparation de l'échantillon) pour la détermination de 14 pesticides dans des fruits. En ce qui concerne la préparation de l'échantillon, ils ont évalué trois techniques d'extraction : l'extraction liquide-liquide (ELL), l'extraction sur phase solide (SPE) et l'extraction sur phase solide dispersée (MSPD).

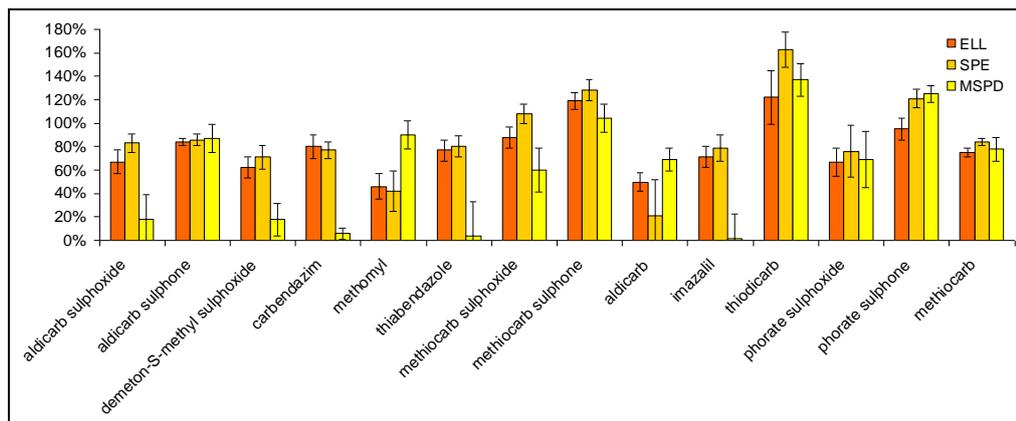
Le premier des 3 graphes de la Figure 18 montre que les effets matriciels sont dépendants de la technique de préparation choisie pour certains composés uniquement comme le methonyl, l'aldicarb ou le thiodicarb. Ceci peut être dû à une extraction plus ou moins sélective de composés constituant la matrice des échantillons, et pouvant interférer ou non lors de l'analyse.



% Effet matriciel sur l'analyse – extrait de pomme supplémenté à 0.01 mg/kg comparé à une référence



% Rendement extraction d'un échantillon de pomme supplémenté à 0.01 mg/kg



$$\% \text{ Efficacité de la méthode} = (\text{Effet matriciel} * \text{rendement}) / 100\%$$

Figure 18 : Effet matriciel sur l'analyse instrumentale, l'étape de préparation et sur l'efficacité de la méthode pour 14 pesticides, d'après Krueve et al. [23]

Par ailleurs, pour un même composé, les 3 techniques d'extraction testées ne donnent pas toujours des rendements d'extraction équivalents comme cela peut se voir sur la deuxième figure. En particulier, la MSPD donne des rendements particulièrement faibles (< 20%) pour 5 des 14 pesticides.

Enfin, en considérant l'efficacité de la méthode, qui résulte de la combinaison de l'étape de préparation et de l'analyse instrumentale et qui traduit l'ensemble du processus analytique, il apparaît que la méthode d'extraction par SPE est la mieux adaptée, ainsi que l'extraction ELL, avec 104 % et 83% d'efficacité en moyenne respectivement.

Ces considérations sont capitales dans le cas de matrices complexes comme peuvent l'être des matrices de type sédiment, eaux résiduaires ou même des eaux naturelles fortement chargées en matières en suspension.

❖ Mode d'étalonnage

Le choix du mode d'étalonnage est une étape critique car il peut engendrer un biais analytique. En effet les effets de matrice peuvent avoir une influence non négligeable sur l'étalonnage. Ceci est illustré par un exemple dans lequel différents modes d'étalonnage ont été comparés par Sanchez-Brunete et al. [24] (Figure 19): étalonnage externe dans le solvant ou dans la matrice et étalonnage interne dans le solvant ou dans la matrice pour différents insecticides organochlorés analysés dans les boues de station d'épuration urbaine.

L'étalonnage externe a été réalisé comme suit :

- préparation de solutions étalons dans un solvant
- dopage à plusieurs niveaux de concentration d'extraits pour prendre en compte la matrice (technique des ajouts dosés)

L'étalonnage interne a été réalisé de la même manière mais en incluant comme des étalons internes marqués (une molécule marquée par composé analysé), c'est-à-dire en utilisant la dilution isotopique.

Les résultats pour 2 insecticides (α -BHC et le 4,4'-DDT) montrent bien que pour l'étalonnage externe la matrice a un effet non négligeable sur la pente de la droite d'étalonnage, en particulier pour le 4,4'-DDT. La mise en œuvre de la dilution isotopique permet de compenser l'effet de la matrice, les droites d'étalonnage solvant et matrice étant confondues.

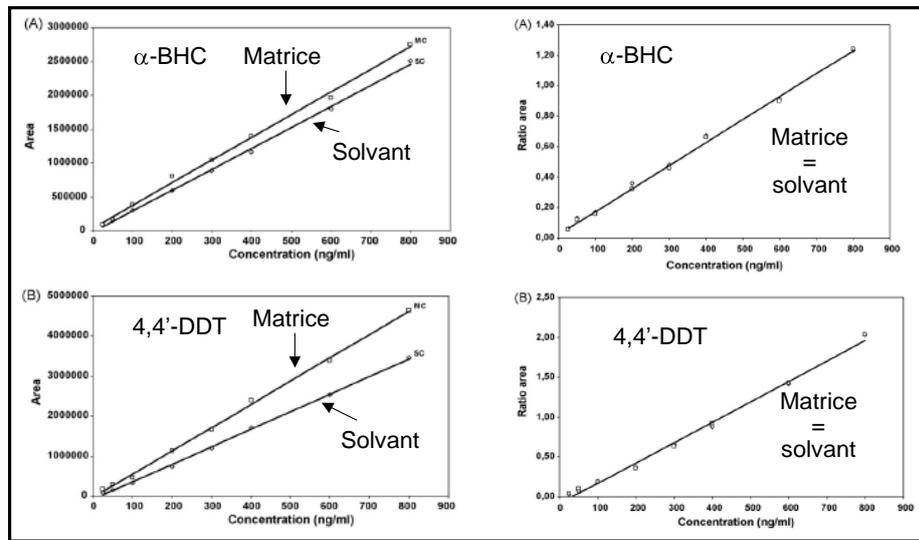


Figure 19 : Comparaison entre les étalonnages externes et internes en solvant ou en matrice pour le α -BHC et le 4,4'-DDT, d'après Sanchez-Brunete et al. [24]

Ainsi les phénomènes matriciels peuvent être compensés, soit par l'ajout d'étalons internes, soit par l'étalonnage dans la matrice si celle-ci reste constante pour l'ensemble des échantillons analysés ou encore si le volume d'échantillon le permet par la technique des ajouts dosés.

❖ Blancs

La problématique du contrôle des «blancs» dont on sait qu'ils peuvent représenter un élément de poids dans la validation et la qualité des données de mesure, notamment dans le contexte de la Directive Cadre sur l'Eau DCE (niveaux de concentration faibles recherchés) est cruciale. Ces procédures de contrôle sont globalement mal encadrées d'un point de vue normatif. En conséquence, il existe une grande hétérogénéité des pratiques au sein des laboratoires pour ce qui concerne les blancs: leur nature, leur fréquence de détermination, leurs critères d'exclusion ainsi que l'usage qu'il en est fait : correction ou non de la valeur mesurée par la valeur de blanc. Le poids de cette hétérogénéité sera d'autant plus important que le niveau de concentration de la molécule dans le milieu est faible et que l'écart entre les blancs de laboratoire et les concentrations mesurées dans l'environnement sont proches.

❖ Fidélité

L'augmentation du nombre de répétitions permet d'obtenir une meilleure estimation de la moyenne et de l'écart type, et par conséquent de réduire l'erreur aléatoire. Cette stratégie analytique, utilisés par les auteurs de l'article présenté au paragraphe 5.2.3, pourrait être utilisée pour diminuer la variabilité des données. La diminution d'incertitude est proportionnelle à la racine carrée du nombre de répétitions.

❖ Justesse

L'utilisation de Matériaux de Référence Certifiés (MRC) permet d'assurer la traçabilité métrologique des mesures effectuées, mais aussi de minimiser la contribution de la justesse à l'incertitude de mesure. En effet, si l'écart normalisé EN (différence entre la valeur mesurée et la valeur de référence normalisée par la somme de carrés de incertitudes de deux valeurs) calculé à partir des incertitudes élargies est inférieur à 1, alors l'incertitude type de justesse est égale à l'incertitude sur le Matériaux de Référence (MR). Selon le type de MR (certifié ou non certifié), les incertitudes associées peuvent être très différentes selon la méthode utilisée pour les quantifier.

Par exemple, pour les HAP, le NIST calcule une incertitude élargie pour son MRC 1647e qui comprend une correction pour la pureté des produits et qui intègre la différence entre la préparation par gravimétrie et les mesures faites par chromatographie. L'incertitude élargie sur les composés de la solution mélange certifiée (JBR-0005) n'est par contre estimée que pour la préparation par gravimétrie.

7.4. CONCLUSION ET RECOMMANDATIONS

Plusieurs articles scientifiques traitent de la qualité de la donnée et de son impact sur la détermination des tendances (voir chapitre 5).

Différentes pistes permettant de mieux maîtriser chaque étape du processus de mesure, depuis le prélèvement jusqu'au rendu de résultat, ont été présentées dans ce chapitre. Ainsi, une meilleure maîtrise du processus de mesure a pour conséquences de diminuer l'incertitude et d'améliorer la qualité de la donnée.

Une des premières conditions à l'évaluation d'une tendance est de bien définir l'analyte considéré. En effet, il peut coexister sous un même numéro CAS plusieurs isomères (par exemples des isomères chiraux) et il est très important de s'assurer que c'est toujours le même mesurande qui est analysé, par exemple la somme des isomères ou bien un seul des isomères.

Une traçabilité documentaire des méthodes utilisées apparaît comme nécessaire afin de suivre l'évolution des méthodes analytiques dans le temps et pouvoir tracer les changements qui ont eu lieu pendant la chronique et pouvoir s'assurer que les tendances détectées ne sont pas imputables à un changement dans les méthodes analytiques utilisées (Wahlin et Grimwall, 2008).

Ainsi, les métadonnées qu'il paraît indispensable de connaître incluent :

- Les composés analysés, en incluant tous les isomères
- les limites de quantification (LQ)
- les incertitudes de mesures U ainsi que la méthode employée pour les estimer

- les protocoles de prélèvement
- les conditions de conservations
- les protocoles analytiques, en détaillant les étapes critiques comme l'extraction et la minéralisation, en spécifiant la pureté des réactifs utilisés pour préparer les solutions étalons, l'appareil, le modèle d'étalonnage et le type d'étalonnage (externe, interne, dans la matrice), les blancs etc.

Par ailleurs, visualiser graphiquement dans les chroniques temporelles, les changements apparus dans les méthodes employées, mais aussi les améliorations analytiques, permettrait d'écarter ou non plus rapidement l'hypothèse d'une origine analytique sur la détection des tendances.

Wahlin et Grimwall (2008) proposent de réaliser régulièrement une analyse rétrospective des protocoles analytiques afin de ne pas découvrir 20 ans après qu'il y a eu une dérive analytique et que les tendances détectées sont en fait principalement dues aux méthodes utilisées et non à une évolution du milieu comme ils l'ont démontré sur les rivières suédoises. Cette analyse rétrospective peut notamment s'appuyer sur une vérification régulière des performances annoncées des méthodes, par exemple via des audits des laboratoires impliqués dans les programmes de surveillance en vue d'évaluer les tendances.

8. Conclusion et perspectives

La directive cadre européenne sur l'eau fixe aux états membres, des objectifs d'identification des tendances d'évolution des polluants dans les masses d'eau notamment pour les eaux souterraines et les supports biote et sédiments pour les eaux de surface.

Ce rapport a pour objectif de s'intéresser à l'impact de la qualité des données de surveillance et notamment leur incertitude sur la capacité à détecter des tendances de façon fiable principalement dans le contexte des eaux souterraines.

Les principaux textes réglementaires (directive 2006/11/CE dite directive « eau souterraine » et directive 2008/105/CE dite directive NQE) ont été dans un premier temps analysés ainsi que leurs guides d'application européens.

Pour les documents relatifs aux « eaux souterraines », cette analyse a permis de rappeler que la directive demande l'utilisation d'outils statistiques permettant de différencier les tendances des variations naturelles du milieu avec un niveau de confiance et de précision suffisants. Il est également demandé de concevoir le programme de surveillance de manière à ce que les tendances à la hausse significatives et durables des concentrations de polluants puissent être décelées.

La directive NQE donne des grands principes sur les objectifs en termes de tendance et rappelle également que, les états membres doivent fixer la fréquence de contrôle à effectuer (donc concevoir le programme de surveillance) de manière à fournir des données suffisantes pour effectuer des analyses de tendance fiable.

Le guide d'application de la directive NQE pour les sédiments et le biote est le texte le plus complet sur le sujet de l'évaluation des tendances. Il ne présente pas les outils statistiques mais décrit les grands principes à respecter pour mettre en place des programmes de surveillance adaptés à l'identification de tendance. Il insiste tout d'abord sur l'importance de fixer des objectifs quantitatifs pour le programme en termes notamment d'efficacité. Par exemple on peut vouloir détecter une diminution de 50% de concentrations sur une période de 10 ans avec une puissance statistique de 80% et pour un niveau de signification de 5%. Le guide insiste également sur la nécessité de disposer d'informations sur la variabilité du système étudié afin de mettre en place de façon concrète le programme de surveillance sur la base des objectifs fixés. Cette variabilité du système dépend :

- De la variabilité naturelle du milieu
- De la variabilité liée à la mesure qui inclut les effets liés aux pratiques techniques d'échantillonnage, à la préparation de l'échantillon, au stockage, et à l'analyse.

La variabilité du système pourra notamment être évaluée à travers des phases d'investigation préliminaires qui seront d'autant plus complexes que la masse d'eau sera elle aussi complexe ou hétérogène. Enfin, le guide insiste sur l'importance de l'intervention d'experts capables de compléter les exploitations purement statistiques par une vision plus techniques sur les résultats d'analyse.

La notion de variabilité du système évoquée dans le guide européen « sédiment et biote » est la notion principale qui ressort de tous les chapitres de ce rapport. C'est la notion qui permet de façon prospective de concevoir un programme de surveillance destiné à identifier des tendances sur la base d'objectifs quantitatifs. C'est également la notion qui permet, de façon rétrospective, sur une chronique existante, de préciser la fiabilité de la tendance identifiée. L'incertitude analytique ou plus globalement l'incertitude sur la mesure incluant les étapes techniques d'échantillonnage est une part, plus ou moins importante en fonction des contextes (paramètre, milieu, concentration, support, ...) de cette variabilité globale des données.

Une synthèse bibliographique a été réalisée avec l'objectif de rechercher des articles, abordant à la fois l'estimation de tendances sur des chroniques temporelles et également l'impact de l'incertitude analytique ou plus largement l'impact de la variabilité des données sur la fiabilité des tendances détectées.

De façon générale, peu de références existent sur ce sujet.

Certains articles abordent la question sur un plan qualitatif. Ils mettent en avant des exemples de tendances identifiées sur le plan statistique mais biaisées par des artefacts analytiques. Les enseignements tirés de ces publications sont les suivants :

- L'intérêt d'intégrer une expertise analytique associée à l'exploitation statistique des données. Il semble en effet important de porter un regard critique sur la qualité des données en termes d'incertitude, d'homogénéité des méthodes d'analyse, de traçabilité des mesures, ...
- L'importance d'identifier, si possible, dans les programmes de surveillance les données qui devront être exploitées en tendance afin de vérifier que leur qualité est suffisante pour l'objectif fixé (par exemple mise en place de contrôles adaptés, renforcés,...)
- L'importance de la traçabilité documentaire des données qui devra permettre la recherche et l'identification d'artefacts analytiques gênant l'identification de tendances. On peut par exemple envisager que les changements les plus significatifs (laboratoires, méthodes) soient identifiés de façon claire dans les bases de données et que ces changements apparaissent graphiquement sur les chroniques temporelles.

- La proposition de réfléchir à des tests (contrôles de cohérence des données, contrôles de calibration croisée lors de changement de méthodes ou de laboratoire, variations aberrantes, ...) avant la bancarisation de données destinées à l'évaluation de tendances.
- L'utilité d'exploitations statistiques groupées multi sites pour identifier plus facilement des biais analytiques ou autres biais.

Les articles abordant le sujet des tendances sur un plan plus quantitatif utilisent principalement la régression linéaire et le test non paramétrique de Mann Kendall.

La plupart des articles utilisent la notion de puissance de tests qui traduit de façon schématique la capacité du test à détecter une tendance d'intensité donnée en une durée, dans des conditions et pour un niveau de signification statistique donné. Par extension on peut parler de la puissance (ou efficacité) d'un programme de surveillance qui intègre des conditions bien définies et auquel on peut associer un test d'identification de tendances.

Les principaux facteurs identifiés comme influents pour ces études sont :

- La durée de la chronique
- La variabilité des données (coefficient de variation CV)
- Le nombre de données
- Le degré de signification statistique du test.
- La pente de la tendance.

Que ce soit par simulation, ou bien par exploitation statistique, les références bibliographiques analysées fournissent des informations quantitatives (tableaux, abaques) sur les relations entre ces différents termes et sur la puissance des tests. A travers toutes ces données, il est donc possible par exemple :

- De déterminer la pente que l'on pourra détecter toutes conditions fixées par ailleurs.
- De déterminer le nombre d'années nécessaire à détecter une pente donnée pour un CV, un degré de signification et une puissance de test donnés.
- De déterminer le CV nécessaire pour atteindre un objectif fixé en termes de pente, de nombre d'années, de puissance et de degré de signification donnés.

En ce qui concerne l'influence du CV qui est plus précisément l'objet de ce rapport, la grande tendance identifiée est la suivante : la puissance du test est une fonction décroissante du coefficient de variation de la série. En d'autres termes, lorsque la variabilité des données augmente au sein de la série, la puissance du test diminue et il

est donc plus difficile de détecter l'existence d'une tendance ou encore, les variations au sein d'une série masquent l'existence d'une tendance.

Pour plusieurs auteurs, un des critères discriminant est le rapport entre l'étendue des variations de concentrations sur la période considérée (en prenant comme base la droite de tendance détectée) et l'écart-type résiduel des données (équivalent au CV). Ce critère n'apporte pas d'argument décisif pour l'exploitation des tendances ou pour l'analyse de leur fiabilité mais il permet de fixer une limite très empirique pour une évaluation fiable de tendance : de façon schématique, on peut détecter une tendance de façon fiable dès lors que la variation de concentrations sur la période considérée est supérieure ou égale à la variabilité CV des données (ou mieux au double de cette variabilité).

L'estimation de la variabilité des données apparaît donc comme un facteur décisif pour dimensionner un programme de surveillance pour l'évaluation des tendances. Cette estimation n'est pas facile. Certains auteurs utilisent des études préliminaires spécifiques pour cela ou bien encore se servent de données bibliographiques. D'autres auteurs déterminent la variance résiduelle issue de l'exploitation de chroniques réelles. Cette variabilité des données dépend du type de paramètres, du milieu, du support, du site étudié.

L'incertitude sur une mesure incluant la variabilité analytique et la variabilité liée aux opérations techniques d'échantillonnage est une part plus ou moins importante de la variabilité globale sur la donnée. Certains auteurs font l'hypothèse pour leur étude que ces deux termes sont confondus mais ils attirent l'attention sur le caractère simplificateur de cette hypothèse.

La suite du rapport a permis d'illustrer concrètement ces concepts par la réalisation de simulations ou par l'exploitation de quelques chroniques réelles.

La simulation de chroniques à partir d'une configuration donnée en termes de pente et de durée (20% sur 6 ans) et de nombre de données par année (2 ou 4) a permis d'estimer la puissance de tests pour une hypothèse d'incertitude (ou de variabilité sur la donnée) de 2,5, 10 et 30%. A titre d'exemple, pour une incertitude de 20% ($k=2$), la puissance du test est de 30% environ pour $n=12$ données sur 6 ans. Dans 70% des simulations réalisées, le test de Mann Kendall appliqué ne détecte pas la tendance simulée.

Ces données de simulation permettent de disposer d'éléments objectifs intégrant la prise en compte de la variabilité des données pour dimensionner ou concevoir des programmes de surveillance ayant des objectifs quantitatifs de détection de tendance. Elles ont aussi un caractère démonstratif permettant d'appréhender plus facilement l'effet de tel ou tel paramètre d'un programme de surveillance.

L'exploitation de chroniques réelles semble un outil très intéressant mais qu'il serait nécessaire d'affiner afin d'estimer la variabilité de données en eau souterraine en fonction du paramètre considéré. Des exploitations de chroniques nitrates et pesticides tendent à montrer que la part de l'incertitude analytique dans la variabilité de la donnée est majoritaire. Dans l'hypothèse où il serait nécessaire d'améliorer l'efficacité d'un programme de surveillance en termes de détection de tendance, agir, si cela est possible techniquement et financièrement, sur l'incertitude analytique serait donc efficace et cette action pourrait être quantifiée.

Pour d'autres paramètres (Fe, As) la part de l'incertitude analytique dans la variabilité des données des chroniques considérées apparaît faible. Ceci peut signifier que pour ces paramètres, une variabilité importante est liée aux protocoles d'échantillonnage ou bien que la variabilité des conditions du milieu est prépondérante. La part liée à l'échantillonnage est certainement non négligeable car ces paramètres sont particulièrement sensibles aux différences de pratiques (filtration ou pas, filtration sur le terrain ou au laboratoire, durée avant filtration au laboratoire, acidification avant ou après filtration, ...).

Cette méthodologie d'estimation de la variance résiduelle et de comparaison aux incertitudes analytiques nécessiterait d'être validée sur le fond y compris au niveau statistique pour mieux prendre en compte les variations cycliques avant une exploitation qui pourrait être très large sur de nombreuses chroniques temporelles en eau souterraine. Ceci permettrait de disposer d'une base de données par paramètre.

Pour résumer les liens qui peuvent exister entre tendance et qualité des données, il est possible de distinguer deux cas :

On dispose des chroniques ou bien elles sont en cours d'acquisition. Les actions possibles sont :

- Expertise des données d'analyse en parallèle de l'exploitation statistique
- Identification des données devant servir à l'évaluation des tendances
- Renforcement de la traçabilité documentaire sur ces données
- Mise en place de test de contrôles des données au fur et à mesure de l'acquisition des données
- Calcul des écart-types résiduels des données pour estimer leur variabilité et préciser si possible la part de l'incertitude analytique dans cette variabilité
- Calcul du facteur Tr/σ comme information supplémentaire empirique sur la fiabilité de la tendance identifiée.

- Estimation de la puissance du test réalisé pour identifier la tendance dans les conditions du programme.

On ne dispose pas de chroniques et le programme de surveillance est en cours de préparation

- Définir des objectifs quantitatifs pour le programme en termes de durée, de pente à détecter, de fréquence d'acquisition, de l'efficacité du programme, de taux de signification des tests statistiques.
- Disposer d'informations concernant la variabilité des données (bibliographie, études préliminaires, études antérieures, ...)
- Vérifier que le programme mis en œuvre permettra de remplir ces objectifs.
- Si besoin le modifier éventuellement en cherchant à réduire par exemple, et si possible, la variabilité des données

Il apparaît donc essentiel que pour l'évaluation des tendances à long terme, des objectifs quantitatifs précis soient fixés et qu'ensuite, le programme de surveillance soit dimensionné notamment en termes de performances de la chaîne de mesure. Dans l'hypothèse où les performances seraient insuffisantes pour atteindre les objectifs fixés, il pourrait être nécessaire soit de réviser les objectifs, soit d'améliorer les performances notamment en termes d'incertitude de mesure. Le dernier chapitre de ce rapport expose les principaux points critiques dont l'amélioration permettrait une diminution des incertitudes de mesure.

Enfin une piste de réflexion concerne les échantillons qui peuvent être conservés dans des conditions de bonne stabilité sur plusieurs années. La bancarisation de ces échantillons pourrait être utilisée pour identifier plus facilement des tendances dès lors que la variabilité analytique s'avérerait prépondérante dans la variabilité globale des données. En effet, l'analyse dans une même série analytique, profitant des dernières avancées technologiques, d'un ensemble d'échantillons archivés depuis de longues années permettrait de réduire au minimum la variabilité analytique et donc de favoriser la détection de tendances.

9. Bibliographie

- [1] 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, JO L327 du 22/12/2000, 1-73
- [2] Directive 2008/105/CE du parlement européen et du conseil du 16 Décembre 2008 établissant des normes de qualité environnementale dans le domaine de l'eau, modifiant et abrogeant les directives du Conseil 82/176/CEE, 83/513/CEE, 84/156/CEE, 84/491/CEE, 86/280/CEE et modifiant la directive 2000/60/CE, JO L348 du 24/12/2008, 84-97.
- [3] 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, JO L372/19 du 27/12/2006
- [4] Quality assurance in analytical chemistry – Prichard & Barwick – Analytical Techniques in the sciences, Wiley Ed, 2007
- [5] Guidance on chemical monitoring of sediment and biota under the water framework directive – Chemical Monitoring Activity – 16/06/2010
- [6] Guidance on groundwater status and trend assessment - J Grath, R Ward and Ph Quevauviller avec l'aide des participants au groupe WG C – 15/10/2008
- [7] Yue S, Pilon P, Cavadias G, Power of the Mann-Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series, Journal of Hydrology, 259 (2002) 254-271.
- [8] Wahlin K., Grimvall A., Uncertainty in water quality data and its implications for trend detection: lessons from Swedish environmental data, Environmental Science & Policy 11 (2008) 115 – 124
- [9] Soderberg K and. Hennet R, Uncertainty and Trend Analysis—Radium in Ground Water and Drinking Water”, Ground Water Monitoring & Remediation 27, no. 4/ 2007/pages 122–129
- [10] Bouchard A, Haemmerli J., Trend detection in water quality time series of LRTAP – QUEBEC network lakes”, Water, Air and Soil Pollution (1992), 62:89-110.
- [11] Sulkava M., Luysaert S., Rautio P. & al, Modelling the effects of varying data quality on trend detection in environmental monitoring”, Ecological Informatics 2 (2007) 167-176
- [12] Lopez B, Leynet A., Blum A., Baran N. - Evaluation des tendances d'évolution des polluants dans les eaux souterraines. Revue des méthodes statistiques existantes

et recommandations pour la mise en œuvre de la DCE. – Rapport BRGM (à paraître Février 2011)

- [13] Sulkava M., Rautio P., and Hollmén H. - Combining Measurement Quality into Monitoring Trends in foliar Nutrient Concentrations – Lecture Note in Computer Science - 2005
- [14] Nicholson M.D., Fryer R.J., Ross C.A - Designing Monitoring programmes for detecting temporal trends in contaminants in fish and shellfish - Marine Pollution Bulletin, 34 (10), p 821-826 (1997)
- [15] Gorham E. et Detenbeck N.E. (1986) Sulphate in bog waters : a comparison of ion chromatography with Mackereth's cation-exchange technique and a revision of earlier views on cause of bog acidity. J. Ecology,74/3 , 899-903
- [16] Eurachem/CITAC (2003) Traceability in Chemical Measurement. A Guide to achieving comparable results in chemical measurements, 297p.
- [17] NF ENV 13005 (1999) Guide pour l'expression de l'incertitude de mesure (GUM)
- [18] Eurachem (2000) Quantifying uncertainty in analytical measurements, 2nd Edition
- [19] Afnor XP T90-220 (2003) Protocole d'estimation de l'incertitude de mesure associée à un résultat d'analyse pour les méthodes d'analyses physico-chimiques
- [20] Rivier C. Lalere B. (2003) Guide méthodologique pour l'estimation des incertitudes en analyse chimique. Rapport C370 X18 (LNE)
- [21] Ghestem J.P., Fiscaro P., Champion R. (2009) – Essai collaboratif sur l'échantillonnage en eau souterraine. BRGM/RP-57687-FR (http://www.aquaref.fr/publications_prelevements)
- [22] Van Leeuwen S.P.J., Swart K., van der Veen I., de Boer J. (2008) Significant improvements in the analysis of perfluorinated compounds in water and fish. Results form an interlaboratory method evaluation study. Institute for Environmental Studies , Vrije University, The Netherlands
- [23] Krueve A ., Künnapas A., Herodes K., Leito I. (2008) Matrix effects in pesticide multi-residue analysis by liquid chromatography – mass spectrometry. Journal of Chromatography A, 1187, 58-66
- [24] Sanchez-Brunete C., Miguel E., Tadeo J.L. (2008) Determination of organochlorine pesticides in sewage sludge by matrix solid-phase dispersion and gas chromatography-mass spectrometry, Talanta, 74, 1211-1217
- [25] Euramet 924, Partie III (2010), Determination of Hg, Cd, Ni and Pb in natural water at concentration levels required by the European Water Framework Directive., Rapport final

[26] Guide ISO/CEI 99 (2007), Vocabulaire international de métrologie – Concepts fondamentaux et généraux et termes associés (VIM).

[27] ISO 13528 (2005), Méthodes statistiques utilisées dans les essais d'aptitude par comparaisons interlaboratoires

Annexe 1

Directives fille « eaux souterraines » et NQE : extraits des chapitres relatifs aux tendances

Cette annexe retranscrit les passages de la directive fille sur les eaux souterraines [3] et de la directive NQE[2] relatifs à l'évaluation des tendances.

Directive fille Eau souterraine

Article 5

Identification des tendances à la hausse significatives et durables et définition des points de départ des inversions de tendance

1. 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.

2. Conformément à l'annexe IV, partie B, 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.

3. Les États membres définissent le point de départ des inversions de tendance sous la forme d'un pourcentage du niveau établi par les normes de qualité des eaux souterraines fixées à l'annexe I et les valeurs seuils fixées conformément à l'article 3, sur la base de la tendance identifiée et des risques environnementaux associés à cette tendance, conformément à l'annexe IV, partie B, point 1.

4. Les États membres résument, dans les plans de gestion de district hydrographique à présenter conformément à l'article 13 de la directive 2000/60/CE:

a) la manière dont l'évaluation de tendance effectuée à partir :

de certains points de surveillance au sein d'une masse ou d'un groupe de masses d'eau souterraine a contribué à établir, conformément à l'annexe V, point 2.5, de ladite directive, que ces masses subissent d'une manière significative et durable une tendance à la hausse des concentrations d'un polluant quelconque ou le renversement d'une telle tendance; et

b) les raisons sous-tendant les points de départ définis conformément au paragraphe 3.5. Lorsque cela est nécessaire pour évaluer l'impact des panaches de pollution constatés dans les masses d'eau souterraine et susceptibles de menacer la réalisation des objectifs énoncés à l'article 4 de la directive 2000/60/CE, et en particulier des panaches résultant de sources ponctuelles de pollution et de terres contaminées, les États membres effectuent des évaluations de tendance supplémentaires pour les polluants identifiés, afin de vérifier que les panaches provenant de sites contaminés ne s'étendent pas, ne dégradent pas l'état chimique de la masse ou du groupe de masses d'eau souterraine et ne présentent pas de risque pour la santé humaine ni pour l'environnement. Les résultats de ces évaluations sont résumés dans les plans

de gestion de district hydrographique à présenter conformément à l'article 13 de la directive 2000/60/CE.

L'annexe IV de la directive détaille les exigences sur l'évaluation des tendances.

ANNEXE IV

IDENTIFICATION ET INVERSION DES TENDANCES À LA HAUSSE SIGNIFICATIVES ET DURABLES

Partie A

Identification des tendances à la hausse significatives et durables

Les États membres identifient les tendances à la hausse significatives et durables dans toutes les masses d'eau souterraine ou tous les groupes de masses d'eau souterraine caractérisés comme étant à risque, conformément à l'annexe II de la directive 2000/60/CE, en tenant compte des exigences ci-après:

1) conformément à l'annexe V, point 2.4, de la directive 2000/60/CE, **le programme de surveillance est conçu de manière à ce que les tendances à la hausse significatives et durables des concentrations de polluants identifiées en vertu de l'article 3 de la présente directive puissent être décelées.**

2) la procédure d'identification des tendances à la hausse significatives et durables est fondée sur les éléments suivants:

a) les fréquences et les lieux de surveillance sont choisis de façon à être suffisants pour:

i) **fournir les informations nécessaires pour garantir la possibilité de distinguer ces tendances à la hausse des variations naturelles, avec des degrés de confiance et de précision suffisants;**

ii) permettre d'identifier en temps utile ces tendances à la hausse afin que des mesures puissent être mises en oeuvre en vue de prévenir, ou au moins d'atténuer autant que possible, les dégradations de la qualité des eaux souterraines ayant une incidence sur l'environnement. Un premier exercice d'identification aura lieu au plus tard en 2009, si possible, en tenant compte des données existantes, dans le contexte du rapport sur l'identification de tendances dans le cadre du premier plan de gestion de district hydrographique visé à l'article 13 de la directive 2000/60/CE, et au moins tous les six ans par la suite;

iii) tenir compte des caractéristiques physiques et chimiques temporelles de la masse d'eau souterraine, y compris les conditions d'écoulement des eaux souterraines et les vitesses d'infiltration, ainsi que le délai de percolation à travers le sol ou le sous-sol;

b) les méthodes de surveillance et d'analyse utilisées sont conformes aux principes internationaux de contrôle de la qualité, y compris éventuellement aux méthodes du CEN ou aux méthodes nationales normalisées, pour garantir la fourniture de données d'une qualité scientifique et d'une comparabilité équivalentes;

c) **l'évaluation est 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;

d) afin d'éviter de fausser l'identification des tendances, la moitié de la valeur de la limite de quantification 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.

3) l'identification des tendances significatives et durables à la hausse des concentrations de substances à la fois naturellement présentes et résultant de l'activité humaine prendra en compte les points de départ de l'identification et, lorsqu'elles sont disponibles, les données recueillies avant le démarrage du programme de surveillance aux fins de l'identification de tendances dans le cadre du premier plan de gestion de district hydrographique prescrit à l'article 13 de la directive 2000/60/CE.

Partie B

Points de départ des inversions de tendance

Les États membres inversent les tendances à la hausse significatives et durables, une fois identifiées, conformément à l'article 5, en respectant les exigences ci-après:

1) Le point de départ de la mise en œuvre de mesures visant à inverser des tendances à la hausse significatives et durables correspond à une concentration du polluant qui équivaut à 75 % des valeurs des paramètres relatifs aux normes de qualité des eaux souterraines établies à l'annexe I et des valeurs seuils fixées conformément à l'article 3, sauf si:

a) un point de départ plus précoce est nécessaire pour que les mesures d'inversion de tendance puissent prévenir de la façon la plus économique qui soit, ou au moins atténuer autant que possible, toute dégradation de la qualité des eaux souterraines ayant une incidence sur l'environnement;

b) un point de départ différent se justifie lorsque la limite de détection ne permet pas, à 75 % des valeurs des paramètres, de démontrer l'existence d'une tendance; ou

c) le taux d'accroissement et la réversibilité de la tendance sont tels que le choix d'un point de départ plus tardif pour les mesures d'inversion de tendance permettrait encore de prévenir de la façon la plus économique qui soit, ou au moins d'atténuer autant que possible, toute dégradation de la qualité des eaux souterraines ayant une incidence sur l'environnement. Le cas échéant, le choix d'un point de départ plus tardif n'empêche pas de respecter les échéances fixées pour atteindre les objectifs environnementaux.

Pour les activités relevant de la directive 91/676/CEE, le point de départ de la mise en œuvre de mesures destinées à inverser les tendances à la hausse significatives et durables est établi conformément à ladite directive et à la directive 2000/60/CE et, notamment, en adhérant aux objectifs environnementaux de protection des eaux visés à l'article 4 de la directive 2000/60/CE.

2) Une fois un point de départ établi pour une masse d'eau souterraine caractérisée comme étant à risque conformément à l'annexe V, point 2.4.4, de la directive 2000/60/CE et au point 1 ci-dessus, il ne sera plus modifié au cours du cycle de six ans du plan de gestion de district hydrographique prescrit à l'article 13 de la directive 2000/60/CE.

3) *Les inversions de tendance doivent être démontrées, compte tenu des dispositions pertinentes en matière de surveillance figurant partie A, point 2.*

Directive fille NQE

Article 3 §3

Les États membres procèdent à l'analyse tendancielle à long terme des concentrations des substances prioritaires énumérées à l'annexe I, partie A, qui ont tendance à s'accumuler dans les sédiments et/ou le biote, en prêtant tout particulièrement attention aux substances no 2, 5, 6, 7, 12, 15, 16, 17, 18, 20, 21, 26, 28 et 30, et en se fondant sur la surveillance de l'état des eaux assurée conformément à l'article 8 de la directive 2000/60/CE. Sous réserve de l'article 4 de la directive 2000/60/CE, ils prennent les mesures nécessaires pour veiller à ce que de telles concentrations n'augmentent pas de manière significative dans les sédiments et/ou le biote pertinent.

Les États membres déterminent la fréquence des contrôles à effectuer dans les sédiments et/ou le biote de manière à fournir des données suffisantes pour effectuer une analyse tendancielle à long terme fiable. À titre indicatif, les contrôles devraient avoir lieu tous les trois ans, à moins qu'un autre intervalle ne se justifie sur la base des connaissances techniques et des avis des experts.



Centre scientifique et technique
Service Métrologie, Monitoring, Analyse
3, avenue Claude-Guillemin
BP 36009 – 45060 Orléans Cedex 2 – France – Tél. : 02 38 64 34 34