



Développement d'outils d'aide à l'évaluation des tendances dans les eaux souterraines au titre de la DCE

Rapport final

Auteur(s)

**B. LOPEZ, N. CROISSET, N. SURDYK, A. BRUGERON
(BRGM)**

Janvier 2013

Document élaboré dans le cadre de :
La Directive Cadre sur l'Eau

- **AUTEURS**

Benjamin LOPEZ, chef de projet (BRGM), b.lopez@brgm.fr

Nolwenn CROISSET, ingénieur hydrogéologue (BRGM), n.croiset@brgm.fr

Nicolas SURDYK, ingénieur agronome (BRGM), n.surdyk@brgm.fr

Alexandre BRUGERON, ingénieur hydrogéologue (BRGM), a.brugeron@brgm.fr

- **CORRESPONDANTS**

Onema : **Cendrine DARGNAT**, cendrine.dargnat@onema.fr

DEB/MEDDE : **Sarah Bonneville**, bureau des eaux souterraines et de la ressource en eau (GR1), sarah.bonneville@developpement-durable.gouv.fr

Partenaire : **Laurence GOURCY**, correspondante Onema (BRGM), l.gourcy@brgm.fr

Droits d'usage : Accès libre

Niveau géographique : national

Couverture géographique : France métropolitaine et DOM

Citations locales :

Niveau de lecture : Professionnels, experts

- **RESUME**

Le calcul des tendances d'évolution des polluants dans les eaux souterraines constitue une étape essentielle pour répondre aux exigences de la Directive Cadre sur l'Eau (DCE). Les travaux présentés dans ce rapport font suite à une étude menée en 2010 (Lopez et Leynet, 2011) qui a permis de faire le point sur les méthodes statistiques d'identification des tendances des séries temporelles existantes et d'examiner leur applicabilité au domaine des eaux souterraines.

Dans une première partie, ce rapport présente les « fiches tendance » réalisées pour les agences de l'eau par masse d'eau afin de rendre compte des tendances d'évolution des nitrates dans les eaux souterraines.

Dans une seconde partie, ce rapport décrit le fonctionnement d'un outil qui permet la caractérisation des séries temporelles et l'identification des tendances d'évolution de n'importe quel contaminant dans les eaux souterraines. Cet outil, développé sous environnement R (logiciel libre et gratuit), est facilement utilisable par les gestionnaires (agences de l'eau en particulier).

Afin de tester l'outil sur des chroniques variées, des chroniques caractéristiques des données utilisées pour le suivi DCE ont été extraites d'ADES. L'outil a été appliqué sur ces chroniques. Les résultats sont présentés et montrent que l'outil peut être utilisé sur une grande variété de chroniques.

- **MOTS-CLES : Outil d'aide à la décision, polluants, eaux souterraines, tendances**

Développement d'outils d'aide à l'évaluation des tendances dans les eaux souterraines au titre de la DCE

Rapport final

BRGM/RP-61855-FR
Janvier 2013

B. Lopez, N. Croiset, N. Surdyk, A. Brugeron

Vérificateur :

Nom : L. Gourcy

Date : 19/12/2013

Signature :



Approbateur :

Nom : N. Dorfliger

Date : 10/01/2014

Signature :



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

Le système de management de la qualité du BRGM est certifié AFAQ ISO 9001:2008.

Mots-clés : Outil d'aide à la décision, polluants, eaux souterraines, tendances

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

Lopez B., Croiset N., Surdyk N., Brugeron A. (2013) – Développement d'outils d'aide à l'évaluation des tendances dans les eaux souterraines au titre de la DCE. Rapport final. BRGM/RP-61855-FR, 98 p., 45 ill., 1 ann.

Synthèse

Le calcul des tendances d'évolution des polluants dans les eaux souterraines constitue une étape essentielle pour répondre aux exigences de la Directive Cadre sur l'Eau (DCE). Les travaux présentés dans ce rapport et réalisés dans le cadre de la convention ONEMA-BRGM 2012 font suite à une étude menée en 2010 également sous convention avec l'ONEMA (Lopez et Leynet, 2011) qui a permis de faire le point sur les méthodes statistiques d'identification des tendances des séries temporelles existantes et d'examiner leur applicabilité au domaine des eaux souterraines.

Dans une première partie, ce rapport présente les « fiches tendance » réalisées pour les agences de l'eau par masse d'eau afin de rendre compte des tendances d'évolution des nitrates dans les eaux souterraines.

Dans une seconde partie, ce rapport décrit le fonctionnement un outil qui permet la caractérisation des séries temporelles et l'identification des tendances d'évolution de n'importe quel contaminant dans les eaux souterraines. Cet outil, développé sous environnement R (logiciel libre et gratuit), est facilement utilisable par les gestionnaires (agences de l'eau en particulier). Un premier module permet de visualiser les données et extrait les principales caractéristiques des chroniques. Le module principal propose et exécute une série de tests statistiques permettant de caractériser les tendances d'évolution des concentrations, la présence de ruptures dans les chroniques et la variabilité entre saisons. Trois modules complémentaires d'utilisation avancée sont également proposés : l'un permet d'évaluer la présence de tendances en tenant compte du caractère saisonnier des chroniques, un autre permet d'étudier la présence de tendance à l'échelle régionale en utilisant des données acquises sur plusieurs points d'eau et un dernier module permet de rééchantillonner des données en ne gardant que des données ayant une fréquence d'échantillonnage régulière.

Afin de tester l'outil sur des chroniques variées, des chroniques caractéristiques des données utilisées pour le suivi DCE ont été extraites d'ADES. L'outil a été appliqué sur ces chroniques de concentrations de nitrates et atrazine. Les résultats sont présentés et montrent que l'outil peut être utilisé sur une grande variété de chroniques.

Sommaire

1. Introduction.....	9
1.1. CONTEXTE REGLEMENTAIRE	9
1.2. ETAT DE L'ART	10
1.3. OBJECTIFS	11
1.3.1. Aide à l'évaluation des tendances d'évolution des nitrates à l'échelle des masses d'eau.....	11
1.3.2. Développement d'un outil permettant la description des chroniques temporelles et l'évaluation des tendances d'évolution	11
1.3.3. Validation de l'outil par application sur des données réelles.....	12
2. Edition de fiches « tendance NO₃ » par masse d'eau.....	13
2.1. METHODES MISES EN ŒUVRE POUR LE RAPPORTAGE DES TENDANCES D'EVOLUTION DES CONCENTRATIONS EN NITRATE.....	13
2.1.1. Méthodes sélectionnées	13
2.1.2. Spécificités liées à l'approche par bassin.....	14
2.1.3. Organisation physique des données livrées.....	14
2.2. PRESENTATION DES FICHES « TENDANCE NO ₃ » PAR MASSE D'EAU	15
2.2.1. Organisation générale des fiches « tendances » par masse d'eau.....	16
2.2.2. Zone d'identification	16
2.2.3. Zone de localisation	17
2.2.4. Zone de résultats des calculs de tendances.....	18
2.3. SYNTHESE DES RESULTATS A L'ECHELLE NATIONALE	22
3. Outil de caractérisation des chroniques et d'identification des tendances d'évolution de la qualité des eaux souterraines.....	23
3.1. PRESENTATION DE L'OUTIL.....	23
3.2. QUELQUES RAPPELS STATISTIQUES	23
3.2.1. Stationnarité d'une chronique et tendance	23
3.2.2. Le test statistique	24
3.2.3. L'hypothèse nulle	24
3.2.4. Niveau de confiance	24
3.3. MISE EN FORME DES DONNEES.....	25
3.4. MODULE « CARACTERISATION »	27
3.4.1. Calcul des statistiques de bases de la chronique.....	27
3.4.2. Représentation graphique.....	30

3.5. MODULE « TENDANCES ET RUPTURES »	32
3.5.1. Tests de tendance	33
3.5.2. Modification des tests de tendance dans le cas d'autocorrélation.....	36
3.5.3. Tests de rupture	39
3.5.4. Sorties générées par l'outil	42
3.6. MODULES COMPLEMENTAIRES (UTILISATION AVANCEE)	45
3.6.1. Module « Saisonnier »	45
3.6.2. Module « Régional ».....	46
3.6.3. Module « Rééchantillonnage »	46
4. Utilisation de l'outil sur des chroniques représentatives	49
4.1. SELECTION DE CHRONIQUES REPRESENTATIVES	49
4.1.1. Contexte et méthode	49
4.1.2. Extraction et données utilisées	49
4.1.3. Résultats des classifications.....	51
4.2. APPLICATION DE L'OUTIL.....	63
4.2.1. Contexte et méthode	63
4.2.2. Application.....	63
5. Conclusion.....	83
6. Bibliographie	85

Liste des illustrations

Illustration 1 : Arbre décisionnel pour le choix de la (ou des) méthode(s) statistique(s) à appliquer sur les chroniques d'évolution de la qualité des eaux souterraines	10
Illustration 2 : Exemple de fiche « tendance NO ₃ » (Masse d'eau FRAG001).....	15
Illustration 3 : Localisation des principales composantes de la fiche « tendance NO ₃ ».....	16
Illustration 4 : Zone d'identification de la fiche tendance NO ₃	17
Illustration 5 : Identification de la zone de localisation sur la fiche tendance NO ₃	17
Illustration 6 : Description détaillée de l'encart concernant le test Mann-Kendall par point.....	18
Illustration 7 : Exemple de tableau de synthèse des résultats Mann-Kendall	19
Illustration 8 : Exemple de visualisation des données prises en compte par le test Kendall régional	20
Illustration 9 : Exemple de visualisation par courbe de fréquence d'apparition des pentes agrégées pour le test Kendall régional	21
Illustration 10 : Calcul des déciles	28
Illustration 11: Test de normalité de la distribution des données	29

Illustration 12 : Test d'autocorrélation.....	30
Illustration 13 : Exemple de sortie graphique pour une chronique de nitrates	31
Illustration 14 : Aide à la lecture d'un diagramme en « boîte à moustache »	32
Illustration 15 : Schéma récapitulatif des différents tests effectués dans le module « Tendances et ruptures ».....	33
Illustration 16 : Test de tendance de Mann-Kendall (Kendall, 1938)	34
Illustration 17 : Calcul de la pente de Sen et de l'ordonnée à l'origine pour le test de Mann-Kendall	35
Illustration 18 : Régression linéaire	36
Illustration 19 : Test de Mann-Kendall modifié pour la prise en compte de l'autocorrélation.	37
Illustration 20 : Analyse de la variance (ANOVA) à un facteur.....	38
Illustration 21 : Test de Kruskal-Wallis.	39
Illustration 22 : Test d'homogénéité de Pettitt	40
Illustration 23: Valeurs critiques de Z du test de Buishand	41
Illustration 24 : Test paramétrique d'homogénéité (Buishand, 1982).....	41
Illustration 25 : Test d'inversion de tendance (Darken,1999).....	42
Illustration 26 : Description d'une sortie graphique obtenue avec le module « Tendances et ruptures » pour une chronique de nitrates au point BSS 05943X00008.	44
Illustration 27 : Test de Kendall saisonnier.....	45
Illustration 28 : Labels des classes et nombre de chroniques par classe (en rouge) pour les concentrations en atrazine dans la classification manuelle.....	53
Illustration 29 : Nombre de prélèvements par site pour le paramètre atrazine	56
Illustration 30 : Nombre de sites par taux de détection pour le paramètre atrazine.....	56
Illustration 31 : Labels des classes et nombre de chroniques par classe (en rouge) pour les chroniques de concentration en nitrate dans la classification manuelle.....	59
Illustration 32 : Nombre de sites en fonction du nombre d'années de suivi pour le paramètre atrazine	62
Illustration 33 : Nombre de sites par taux de détection pour le paramètre atrazine.....	62
Illustration 34 : Copie d'écran de l'outil pour l'analyse de la chronique d'atrazine pour le point 01821X0012/HY (Classe C1)	64
Illustration 35 : Copie d'écran de l'outil pour l'analyse de la chronique d'atrazine pour le point 05005X0042/S-FOND (Classe C3).....	66
Illustration 36 : Copie d'écran de l'outil pour l'analyse de la chronique d'atrazine pour le point 02393X0030/HY (Classe C4)	68
Illustration 37 : Copie d'écran de l'outil pour l'analyse de la chronique d'atrazine pour le point 02522X0003/S1 (Classe C7)	69
Illustration 38 : Copie d'écran de l'outil pour l'analyse de la chronique d'atrazine pour le point 00675X0141/F.AEP (Classe C9)	71
Illustration 39 : Copie d'écran de l'outil pour l'analyse de la chronique d'atrazine pour le point 09043X0001/F (Classe C10)	72
Illustration 40 : Copie d'écran de l'outil pour l'analyse de la chronique d'atrazine pour le point 10142X0022/S (Classe C18)	74

Illustration 41 : Copie d'écran de l'outil pour l'analyse de la chronique de nitrates pour le point 02097X0010/C1 (Classe C1)	75
Illustration 42 : Copie d'écran de l'outil pour l'analyse de la chronique de nitrates pour le point 03538X0005/P (Classe C3)	77
Illustration 43 : Copie d'écran de l'outil pour l'analyse de la chronique de nitrates pour le point 01546X0065/F (Classe C4)	78
Illustration 44 : Copie d'écran de l'outil pour l'analyse de la chronique de nitrates pour le point 02598X0009/P1 (Classe C7)	80
Illustration 45 : Copie d'écran de l'outil pour l'analyse de la chronique de nitrates pour le point 09574X0026/F (Classe C9)	81

Liste des annexes

Annexe 1 Contenu du fichier ZIP «Tendances_NO3_ONEMA_XXXX»	87
--	----

1. Introduction

1.1. CONTEXTE REGLEMENTAIRE

La surveillance accrue de la qualité des eaux souterraines depuis plusieurs années a mis en évidence une contamination plus ou moins importante des aquifères par les nitrates et les produits phytosanitaires notamment. De plus, grâce à la pérennité des programmes réglementaires de mesures, des chroniques de concentrations en contaminants dans les eaux souterraines de plus en plus longues sont disponibles, permettant une analyse approfondie de l'évolution de ces signaux.

La connaissance des niveaux de contamination des eaux souterraines et de leur évolution est un objectif aujourd'hui transcrit dans les textes européens et repris au niveau national. Ainsi, le principal objectif fixé par la Directive Cadre sur l'Eau (2000/60/CE) est que les masses d'eau présentent d'ici 2015 un bon état quantitatif et qualitatif. La Directive impose aux Etats Membres non seulement de caractériser le niveau de la contamination des eaux souterraines mais aussi d'étudier les tendances d'évolution des concentrations des polluants. Parallèlement à la Directive Cadre sur l'Eau, la Directive Nitrate (91/676/EEC) demande aux Etats Membres d'identifier les eaux souterraines contenant plus de 50 mg/L ou susceptibles de dépasser cette limite si des mesures préventives ne sont pas engagées. Un rapport de la commission européenne datant de 2003 indique que 20% des stations européennes ont des concentrations supérieures à cette limite et que 40 % dépassent la valeur guide de 25 mg/L. Dans ce rapport, la France est indiquée comme étant un Etat Membre dans lequel les tendances sont à la hausse. Les Etats Membres doivent ainsi mettre en place un programme d'actions afin d'atteindre le bon état. En cas de non atteinte du bon état, les Etats Membres devront être en mesure d'expliquer les raisons de ce constat et des demandes de dérogation de délai étayées devront être réalisées.

L'évaluation des tendances des concentrations de polluants dans les eaux souterraines est nécessaire dans différents exercices distincts de mise en œuvre de la Directive Cadre sur l'Eau au niveau français :

- 1- Pour les tests « AEP » et « intrusion saline » de l'évaluation chimique de l'état des MESO – guide bon état chimique et tendances¹
- 2- Pour déterminer si une MESO est à risque ou pas (RNAOE chimique) – cf. guide national sur la mise à jour des états des lieux 2013²
- 3- Pour identifier les MESO ayant des tendances à la hausse significatives et durables de concentration de polluants qu'il faut inverser conformément à l'article R.212-21-1 du code de l'environnement. – cf. guide bon état chimique et tendances¹
4. Pour vérifier que les panaches de pollution au sein des MESO, en particulier ceux provenant de sources ponctuelles, ne s'étendent pas (tendances pour les paramètres pertinents) et ne dégradent pas l'état chimique de la masse d'eau.

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

² Direction de l'Eau et de la Biodiversité, MEDDE, mars 2012. Guide pour la mise à jour de l'état des lieux. 124p.

Le calcul des tendances d'évolution des polluants dans les eaux souterraines constitue ainsi un des préalables essentiel pour répondre aux exigences de la Directive Cadre sur l'Eau (DCE). Les travaux réalisés durant cette étude s'inscrivent donc pleinement dans le cadre de la DCE en proposant dans un premier temps d'aider les agences de l'eau à rendre compte des tendances d'évolution des nitrates dans les eaux souterraines et, dans un second temps, un outil complet de caractérisation des séries temporelles et d'identification des tendances d'évolution des contaminants dans les eaux souterraines.

1.2. ETAT DE L'ART

Les travaux présentés dans ce rapport et réalisés dans le cadre de la convention ONEMA-BRGM 2012 font suite à une étude menée en 2010 également sous convention avec l'ONEMA (Lopez et Leynet, 2011) qui a permis de faire le point sur les méthodes statistiques d'identification des tendances des séries temporelles existantes et d'examiner leur applicabilité au domaine des eaux souterraines.

L'analyse de la littérature scientifique sur le sujet a permis de retracer l'historique du développement et de l'amélioration de ces méthodes et révèle certaines limites pour leur application sur des données de qualité des eaux. La non normalité des données et leur autocorrélation apparaissent par exemple comme des phénomènes très perturbateurs des résultats des tests statistiques classiques. Une revue complémentaire des travaux réalisés en France et à l'international pour évaluer les tendances d'évolution de la qualité des eaux superficielles et souterraines a aussi révélé les biais qui existent entre l'évolution de la recherche scientifique et les applications faites par les gestionnaires.

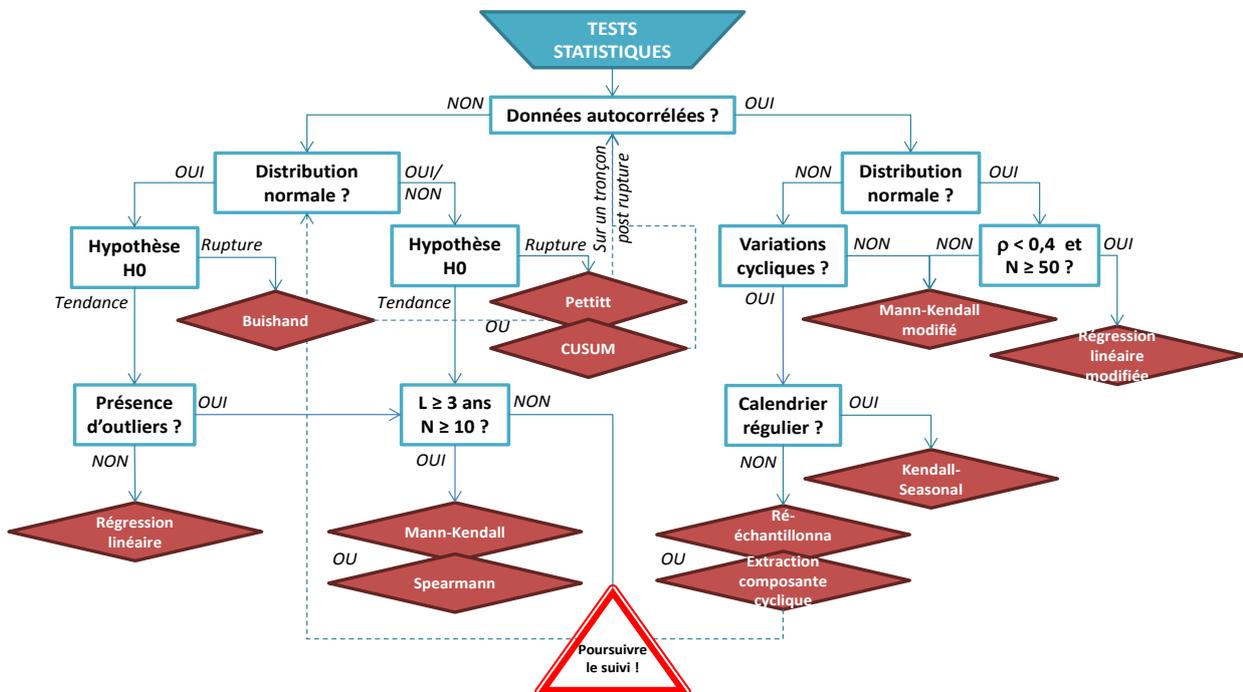


Illustration 1 : Arbre décisionnel pour le choix de la (ou des) méthode(s) statistique(s) à appliquer sur les chroniques d'évolution de la qualité des eaux souterraines

Aussi, afin d'accompagner les gestionnaires (les agences de l'eau en particulier), une procédure de traitement des séries temporelles a été proposée à l'issue de cette étude (Illustration 1). Le choix des tests statistiques à utiliser est argumenté par une analyse préalable de leur robustesse et de leur puissance face aux caractéristiques spécifiques des données de qualité des eaux souterraines. Ces caractéristiques sont issues de l'analyse des données bancarisées dans la base de données nationale ADES sur la qualité des eaux souterraines. L'outil global d'aide à la décision qui conclut l'étude est ainsi nourri des dernières avancées de la Recherche sur le domaine, propose l'utilisation des méthodes statistiques les plus robustes en fonction des caractéristiques des données à traiter et permet aux opérateurs d'être accompagnés pas à pas pour mettre en œuvre une analyse statistique des chroniques d'évolution des concentrations en polluants dans les eaux souterraines.

Néanmoins, à l'issue de ce premier travail, les gestionnaires ont émis le souhait de disposer d'un outil informatique pour la mise en œuvre de la procédure, proposée sous la forme d'un arbre décisionnel dans un premier temps. Les utilisateurs cibles n'étaient en effet pas assez sensibilisés aux tests statistiques recommandés pour pouvoir prendre en charge rapidement leur mise en œuvre dans la perspective de la révision de l'état des lieux en 2013.

1.3. OBJECTIFS

1.3.1. Aide à l'évaluation des tendances d'évolution des nitrates à l'échelle des masses d'eau

Sur proposition du BRGM et au vu des contraintes de temps et d'interopérabilité des logiciels utilisés par les agences, il a été acté que le BRGM fournirait aux bassins pour mi-juin 2012 une fiche par masse d'eau d'identification des tendances nitrates, en format pdf et accompagnée des données brutes correspondantes.

Ces fiches compilent les résultats de l'application de tests robustes de tendances aux échelles d'analyse ponctuelles et à la masse d'eau. Les outils d'identification des tendances appliqués sont :

- Le test de Mann-Kendall par point (code couleur en fonction de la tendance) et comparaison avec la méthode dite de « généralisation » à la masse d'eau telle que mentionnée dans la DCE (si + de 20% des points sont à la hausse alors la masse d'eau est notée à la hausse) ;
- Le test Kendall régional appliqué à l'échelle de la masse d'eau.

1.3.2. Développement d'un outil permettant la description des chroniques temporelles et l'évaluation des tendances d'évolution

Dans la perspective de la révision de l'état des lieux en 2013, les agences et les offices de l'eau ont exprimé le besoin d'avoir un outil leur permettant d'appliquer facilement ces méthodes à partir des données d'ADES. L'automatisation de celles-ci requiert cependant un travail préalable visant à trier les données d'entrée et à tester la bonne applicabilité des tests sur différents paramètres. A l'issue de ce travail, les méthodes retenues pourront être intégrées au SEEE (système d'évaluation de l'état des eaux). Pour faciliter cette intégration, l'outil proposé dans cette étude a été développé sous R (R Development Core Team, 2008) Ce travail de conception et de validation des spécifications de l'outil est essentiel pour permettre et faciliter son intégration dans SEEE.

Les méthodes statistiques d'identification et de quantification des tendances d'évolution des concentrations en contaminants présélectionnées lors de l'étude réalisée en 2010 ont été compilées dans un outil intégré et appliquées sur les chroniques réelles extraites d'ADES représentatives de la diversité des caractéristiques des séries temporelles rencontrées. Les méthodes concernent à la fois l'identification des dates de ruptures de pente des chroniques ainsi que l'estimation des tendances d'évolutions des contaminants dans les eaux souterraines, à l'échelle du point de prélèvement et à une échelle régionalisée.

1.3.3. Validation de l'outil par application sur des données réelles

La base de données nationale sur les eaux souterraines ADES constitue le jeu de données initial à partir duquel les données sont extraites pour l'évaluation des tendances d'évolution des contaminants dans les eaux souterraines. Les travaux réalisés en 2010 (Lopez et Leynet, 2011) ont montré la sensibilité des méthodes statistiques envisageables dans le contexte « eaux souterraines » aux caractéristiques des données d'entrée (longueur, fréquence de prélèvement, distribution, autocorrélation, cyclicité).

Des chroniques réelles d'évolution des concentrations en contaminants dans les eaux souterraines et représentatives de la diversité des typicités rencontrées dans ADES ont été sélectionnées. Deux polluants (nitrates et atrazine) présentant des caractéristiques très contrastées sont considérés. Pour cet exercice, des groupes de chroniques ont été constitués par analyses multicritères et de correspondance sur les caractéristiques des données constitutives des séries temporelles, potentiellement préjudiciables à l'application des tests statistiques. Les règles de tri définies permettent d'extraire d'ADES les chroniques les plus représentatives de chacun des groupes de typicité. L'outil a ensuite été testé sur ces chroniques représentatives.

2. Edition de fiches « tendance NO₃ » par masse d'eau

2.1. METHODES MISES EN ŒUVRE POUR LE RAPPORTAGE DES TENDANCES D'EVOLUTION DES CONCENTRATIONS EN NITRATE

2.1.1. Méthodes sélectionnées

Pour rappel, une tendance d'évolution peut être identifiée de deux manières, visuellement ou par l'application d'une méthode statistique.

La première méthode, généralement associée à l'élaboration de graphiques et/ou cartes, présente l'avantage d'être un moyen de communication simple sur l'existence de tendances, surtout pour les non-initiés. Néanmoins, cette approche ne permet pas d'estimer la « véracité » de la tendance ou de quantifier la pente de celle-ci. A l'opposé et malgré les nombreuses contraintes sur les conditions initiales d'application des tests et les erreurs possibles, l'approche par application d'une méthode statistique permet d'identifier des tendances difficiles à voir, propose un résultat répétable et permet d'estimer les probabilités de se tromper.

Conformément aux recommandations énoncées dans les rapports Lopez et Leynet (2011) et Lopez (2011), les méthodes qui ont été menées dans l'optique du rapportage des tendances d'évolution des concentrations en nitrates sont les suivantes :

- test de Mann-Kendall par point : ce test non-paramétrique constitue le meilleur compromis en terme de robustesse, surtout dans le cas où un nombre important de chroniques doit être testé, et que le temps vient à manquer pour décrire de façon individuelle chacune des séries temporelles. Les résultats sont présentés sous la forme d'une carte de la masse d'eau avec les points d'eau et un code couleur en fonction de la tendance calculée par Mann-Kendall,
- test de Kendall Régional : variante du test précédent permettant d'appréhender la tendance globale d'un ensemble de chroniques et permettre la caractérisation de la zone définie (région, masse d'eau,...) par un groupe de points.

Ces deux tests présentent l'avantage d'être appropriés aux données environnementales dont les distributions s'écartent souvent de la normalité, permettent de s'affranchir du « nettoyage » des chroniques étudiées, ne sont pas trop sensibles aux outliers et leurs résultats restent acceptables même dans le cas d'autocorrélation des données (Lopez et Leynet, 2011). Pour le Kendall Régional, un indice de confiance de l'évaluation (probabilité d'erreur dans la définition de la pente) est également ajouté.

Outre les résultats de ces deux tests statistiques, la fiche « tendance NO₃ » comporte également un tableau de synthèse des résultats (cf. 2.2.4) comprenant un certain nombre d'informations permettant l'application de la méthode dite de « généralisation ». Cette dernière est citée dans le guide européen dans le cadre de la caractérisation d'une masse d'eau à partir d'informations ponctuelles, sans pour autant concerner le point spécifique des tendances. Elle peut se résumer à l'affirmation suivante : « Si plus de 20% des points associés à une masse d'eau sont à la hausse pour un paramètre donné, alors cette masse d'eau est considérée à la hausse vis-à-vis de ce paramètre ».

Lors de la réunion du sous-groupe national « QUALITE » du GT DCE Eaux Souterraines du 20 septembre 2011, il a été acté :

- L'édition de l'atlas des fiches tendances NO₃ qui ne traitent que du paramètre nitrates, sur les données de 1996 à aujourd'hui (plus précisément, du 1/10/1996 au 1/10/2011). Cette période retenue pour le calcul des tendances est, pour cette identification des tendances, spécifique à la révision de l'état des lieux, 1996 à aujourd'hui. La date de début des calculs correspond en effet au premier SDAGE, au progrès des laboratoires et permet une bonne répartition entre années sèches et années humides
- Le développement d'un outil tendance simplifié dont le rendu ressemblerait aux résultats par masse d'eau présenté dans ces fiches avec la méthode de « généralisation » selon la DCE, du Kendall Régional et du Vecteur Régional (cf. chapitre 3).

2.1.2. Spécificités liées à l'approche par bassin

Pour le cas spécifique des masses d'eau trans-bassins, les données prises en compte pour l'analyse des tendances correspondent à la concaténation des données fournies par chaque agence concernée. Les fiches des masses d'eau souterraine trans-bassins sont rendues à toutes les agences qui possèdent une partie de la masse d'eau sur leur bassin.

De plus, certaines agences de bassin se sont engagées dans des travaux de re-délimitations de leurs masses d'eau souterraine. L'agence de l'eau Rhône-Méditerranée & Corse a souhaité l'identification des tendances sur la base de la nouvelle version de leur référentiel Masse d'eau. Dès lors, un travail conséquent d'association des points d'eau aux nouvelles masses d'eau a été mené au préalable par l'agence puis l'ensemble de ces points ainsi qu'une version sub-finalisée des masses d'eau pour ce bassin ont été fournis au BRGM pour le calcul des tendances ainsi que l'édition des fiches.

Pour la Réunion, les fiches ont été éditées en se basant sur les entités de niveaux 2 de la BDLISA qui seront amenées à devenir les masses d'eau souterraine version 2. Tous les points fournis par le SGR Réunion ont été préalablement rattachés à une entité BDLISA.

Malgré les travaux menés de modification des limites de ses masses d'eau souterraine, l'agence de l'eau Loire-Bretagne n'a pas souhaité travailler sur un autre support que celui du Référentiel version 1.

Les autres agences ne sont pas concernées par un changement de référentiel Masse d'eau.

2.1.3. Organisation physique des données livrées

Les fiches ont été regroupées par bassins et livrées à chaque agence concernée :

- Le 20 juillet 2012 pour les masses d'eau des bassins Artois-Picardie, Rhin-Meuse, Seine-Normandie, Loire-Bretagne et Adour-Garonne ;
- Le 25 juillet 2012 pour les masses d'eau des bassins Martinique, Guadeloupe, Réunion et Guyane (pas de calcul à Mayotte car le nombre d'analyses est insuffisant pour permettre une application des tests de tendance) ;

- Le 24 octobre 2012 pour les masses d'eau du bassin Rhône-Méditerranée & Corse (délai dû à la restitution tardive des nouvelles masses d'eau souterraine du bassin et de leurs points d'eau associés).

Pour chaque bassin, l'envoi a été réalisé sous la forme d'un fichier zip comportant :

- Un dossier « Fiches » compilant les fiches « tendances NO₃ » au format PDF et déclinées selon les codes des masses d'eau souterraines (MESO) ;
- Un fichier MXD (format ArcGIS 10.0 ou inférieur selon les demandes spécifiques de l'agence concernée) compilant la majorité des données des bases avec une symbologie rappelant celle des fiches ;
- Une géodatabase Access contenant les données brutes utilisées ainsi que les fichiers résultats exploitées par la fiche pdf.
- Un document « A_lire_avant_utilisation » décrivant le contenu de la géodatabase (cf Annexe 1).

2.2. PRESENTATION DES FICHES « TENDANCE NO₃ » PAR MASSE D'EAU

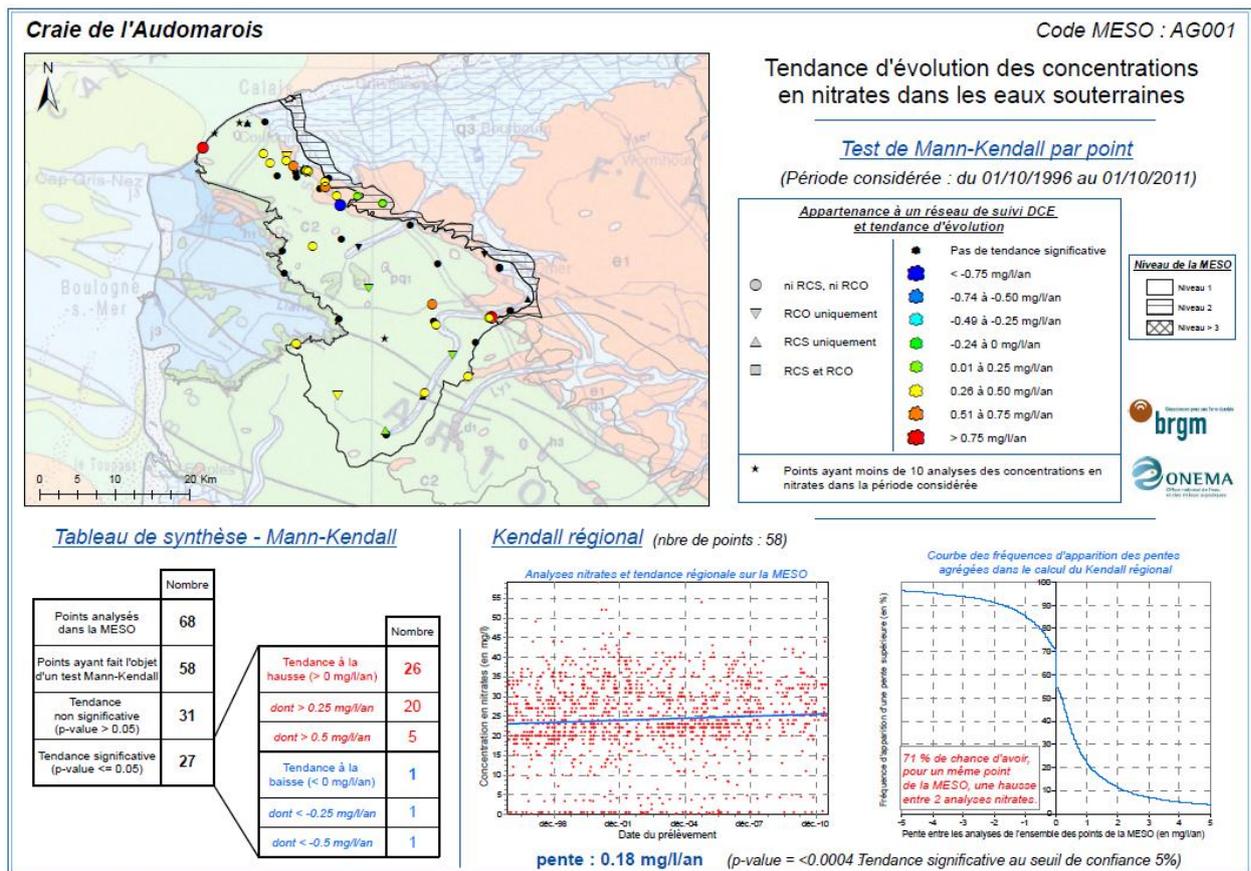


Illustration 2 : Exemple de fiche « tendance NO₃ » (Masse d'eau FRAG001)

2.2.1. Organisation générale des fiches « tendances » par masse d'eau

Le fiche peut se décomposer en 3 zones principales :

- Une partie « Identification de la masse d'eau » ;
- Une partie « Localisation de la masse d'eau » ;
- Une partie « Résultats des calculs des tendances ».

Pour cette dernière, 3 encarts sont définis :

- Résultats du test Mann-Kendall à l'échelle des points de prélèvements ;
- Tableau de synthèse des tendances aux points ;
- Résultats du test Kendall Régional appliqué à l'échelle de la masse d'eau.

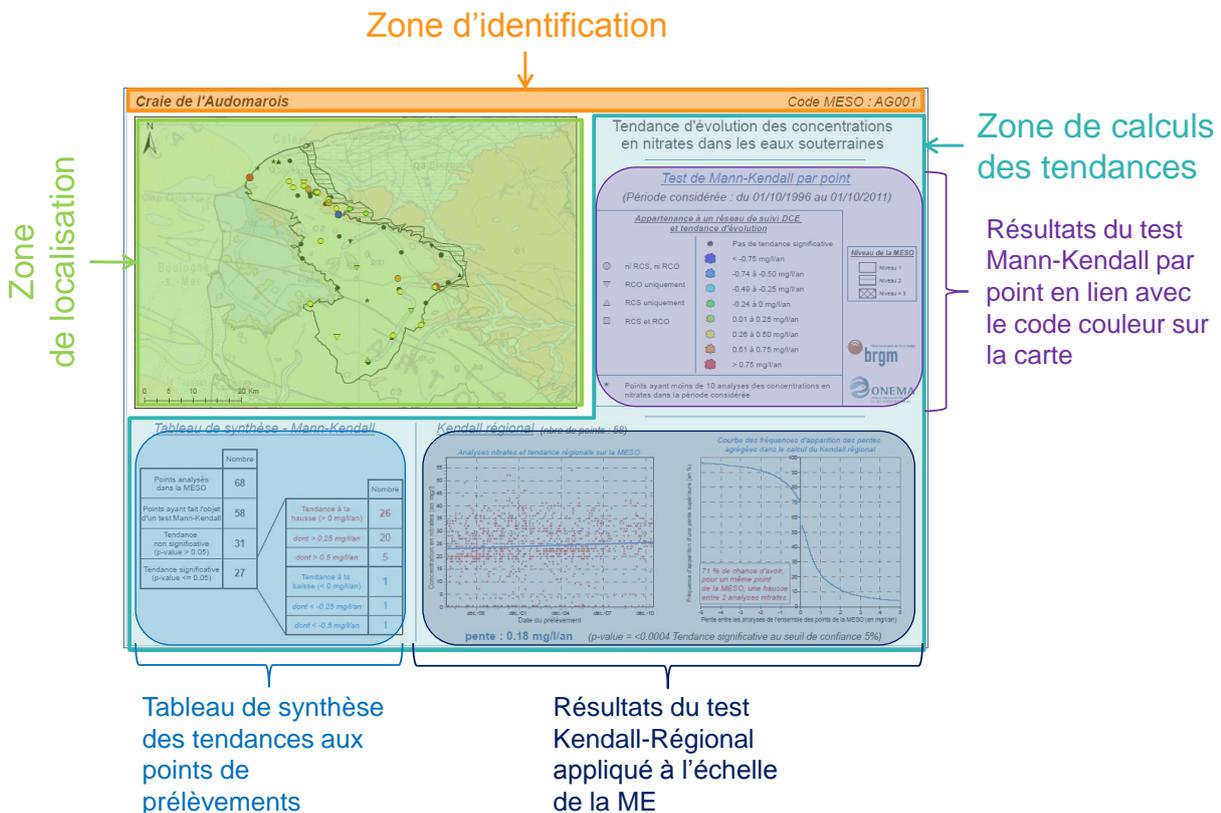


Illustration 3 : Localisation des principales composantes de la fiche « tendance NO₃ »

2.2.2. Zone d'identification

Cette zone comporte le libellé ainsi que le code français de la masse d'eau considérée, telle qu'elle est définie dans le référentiel Masse d'eau Version 1 (version diffusée sur le site du SANDRE).

Craie de l'Audomarois

Code MESO : AG001

Illustration 4 : Zone d'identification de la fiche tendance NO_3

Dans les cas spécifiques des bassins RM&C et Réunion, les codes et libellés utilisés sont ceux fournis par les agences de bassin, à savoir :

- des codes et libellés provisoires des masses d'eau souterraine version 2 pour RM&C,
- les codes et libellés des entités BDLISA de niveaux 2 pour la Réunion.

2.2.3. Zone de localisation

La zone de localisation correspond à la carte de localisation de la masse d'eau souterraine considérée, avec les points d'eau qui lui sont associés et qui ont fait l'objet des tests de tendance.

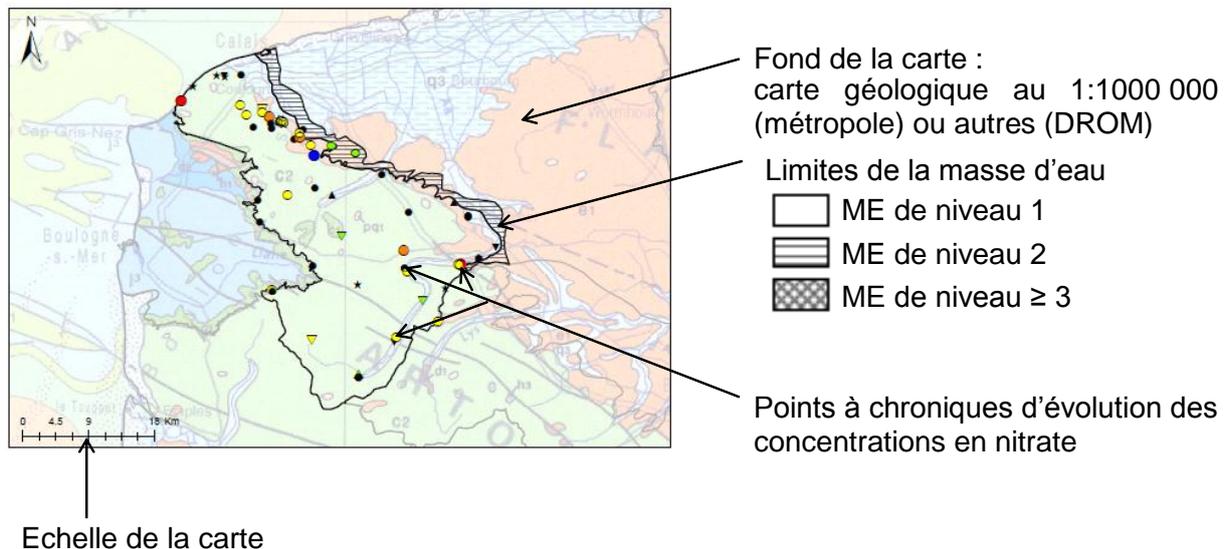


Illustration 5 : Identification de la zone de localisation sur la fiche tendance NO_3

Chaque masse d'eau souterraine considérée est représentée selon ses niveaux de profondeur mais par souci de lisibilité, toutes les parties de niveau supérieur ou égal à trois sont regroupées en une seule classe.

Chaque carte est centrée sur l'extension totale de la masse d'eau souterraine (tout niveau confondu) et comporte une échelle et une flèche d'orientation.

En France métropolitaine, le fond de carte utilisé correspond à la carte géologique au 1/1000 000 du BRGM. Pour les DROM, les fonds de cartes sont les suivants :

- Martinique : carte géologique au 1/50 000 ;
- Guadeloupe : carte géologique au 1/50 000 ;

- Réunion : carte géologique au 1/100 000 ;
- Guyane : carte géologique au 1/500 000 ;

Les points de la masse d'eau pour lesquels les tests statistiques de tendance nitrates ont été calculés sont représentés selon un code couleur et des figurés explicites, détaillés dans le chapitre suivant.

2.2.4. Zone de résultats des calculs de tendances

Résultats du test Mann-Kendall par point

La forme, la taille et la couleur du point sur la carte indiquent le type de point, le sens de la tendance et sa pente. L'illustration 6 ci-dessous détaille cette zone qui est en réalité la légende de la carte de la « Zone de localisation ».

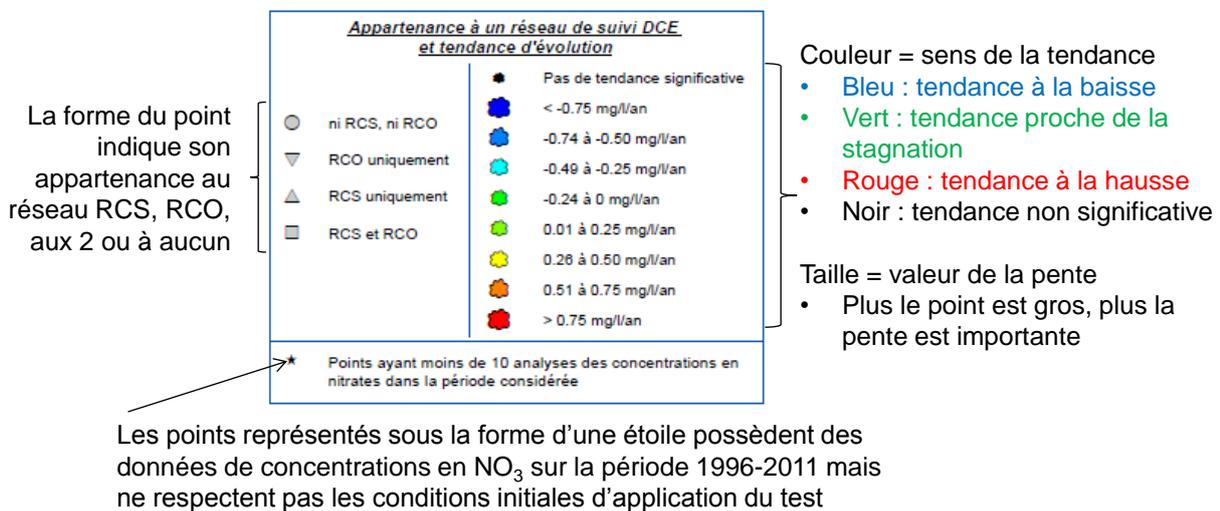


Illustration 6 : Description détaillée de l'encart concernant le test Mann-Kendall par point

Le code couleur correspond au sens de la tendance et la taille du figuré correspond à la valeur prise par la pente d'évolution.

L'ensemble des chroniques ayant au moins 10 données dans la période considérée est pris en compte, quelles que soient la concentration moyenne et la concentration maximale au point. Les points dont les figurés sont des étoiles ne respectent pas cette unique condition.

La possibilité de prendre en compte les chroniques qui possèdent moins de 10 données a été discuté mais l'exercice a été rapidement abandonné, pour l'édition de ces fiches, car il nécessitait des développements complémentaires qu'il était difficilement envisageable de réaliser vu leur délai de livraison (début 2nd semestre 2012).

Tableau récapitulatif du calcul des tendances à l'échelle des points de prélèvements

Le tableau reporté sur les fiches correspond à une synthèse des résultats de l'identification des tendances aux points par la méthode de Mann-Kendall par point (cf. Illustration 7).

Tableau de synthèse - Mann-Kendall

	Nombre
Points analysés dans la MESO	68
Points ayant fait l'objet d'un test Mann-Kendall	58
Tendance non significative (p-value > 0.05)	31
Tendance significative (p-value ≤ 0.05)	27

	Nombre
Tendance à la hausse (> 0 mg/l/an)	26
dont > 0.25 mg/l/an	20
dont > 0.5 mg/l/an	5
Tendance à la baisse (< 0 mg/l/an)	1
dont < -0.25 mg/l/an	1
dont < -0.5 mg/l/an	1

Illustration 7 : Exemple de tableau de synthèse des résultats Mann-Kendall

Ce tableau est composé des informations suivantes :

- Le nombre de point de la masse d'eau qui comporte au moins une analyse nitrates dans la période considérée ;
- Le nombre de point ayant fait l'objet d'un test Mann-Kendall (en d'autre termes, nombre de points avec plus de 10 analyses dans la période considérée) ;
- Le nombre de points pour lesquels la tendance déterminée via Mann-Kendall n'est pas significative (p-value > 0,05) ;
- Le nombre de points pour lesquels la tendance déterminée via Mann-Kendall est significative (p-value ≤ 0,05). Parmi ces derniers, un second tableau détaille le sens et l'importance de la tendance :
 - Nombre de tendance à la hausse (dont >0,25 mg/l/an et >0,5 mg/l/an)
 - Nombre de tendance à la baisse (dont <-0,25 mg/l/an et <-0,5 mg/l/an)

Résultats du test Kendall-Régional à l'échelle de la masse d'eau

Jusqu'à présent, les résultats du test Kendall-Régional étaient souvent présentés sous la forme d'un code couleur appliqué à la zone considérée et dépendante de la valeur de la pente globale déterminée.

Néanmoins, et dans un souci de lisibilité, il a été choisi de privilégier ce type de représentation cartographique pour le test de Mann-Kendall par points, ce qui apporte ainsi une information complémentaire sur la distribution géographique des points de la masse d'eau. Des parties de masses d'eau peu renseignées peuvent ainsi être mises en évidences

Pour le Kendall régional, deux types de représentations complémentaires des résultats de l'application du test ont donc été adoptées.

Pour rappel, le nombre de points utilisés par le test Kendall régional correspond à celui des points ayant fait l'objet d'un test Mann-Kendall (2^{nde} case à gauche du tableau de synthèse).

Le test de Kendall Régional est un dérivé du test de Mann-Kendall par points. Les premières étapes de calcul sont similaires jusqu'à l'obtention, pour chacun des points, de l'ensemble des pentes entre les données d'un même point. L'ensemble de ces pentes est regroupé (comme s'il s'agissait d'un seul point fictif) et la valeur médiane est recherchée. Cette dernière constitue la valeur résultat du test Kendall Régional.

Comme pour l'identification par point, l'ensemble des chroniques ayant au moins 10 données dans la série est pris en compte pour le calcul de la tendance régionale et ce, quelles que soient les concentrations en nitrate

- **Visualisation des données prises en compte et représentation de la tendance régionale**

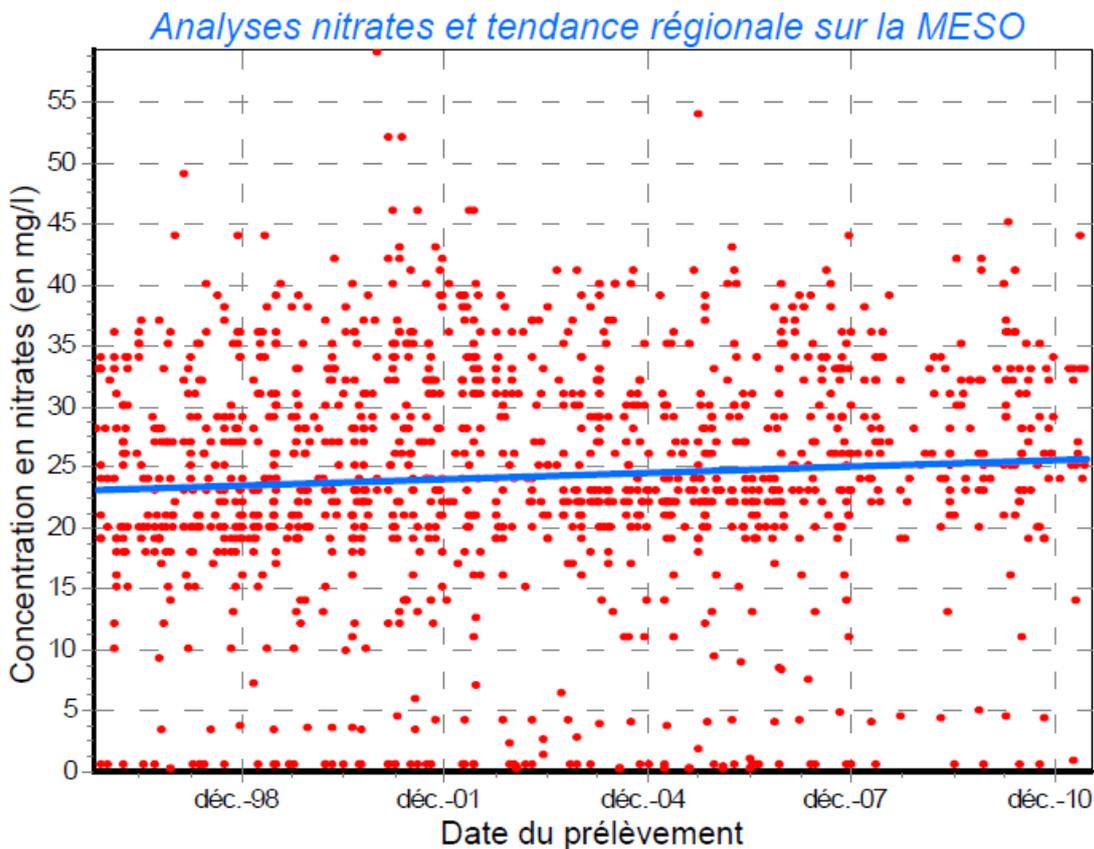


Illustration 8 : Exemple de visualisation des données prises en compte par le test Kendall régional

Cette représentation graphique comporte :

- L'ensemble des couples « valeurs / date d'analyse » pour les points utilisés ;

- La droite ayant pour pente la valeur résultat de la tendance Kendall Régional pour la masse d'eau considérée. **Attention, cette droite ne correspond pas à la régression linéaire du nuage de points !**

- **Représentation probabiliste des résultats du test Kendall-Régional**

Cet autre type de représentation correspond à la courbe des fréquences cumulées d'apparition des pentes entre les données d'un même point, pour l'ensemble des points de la masse d'eau souterraine (cf. Illustration 9).

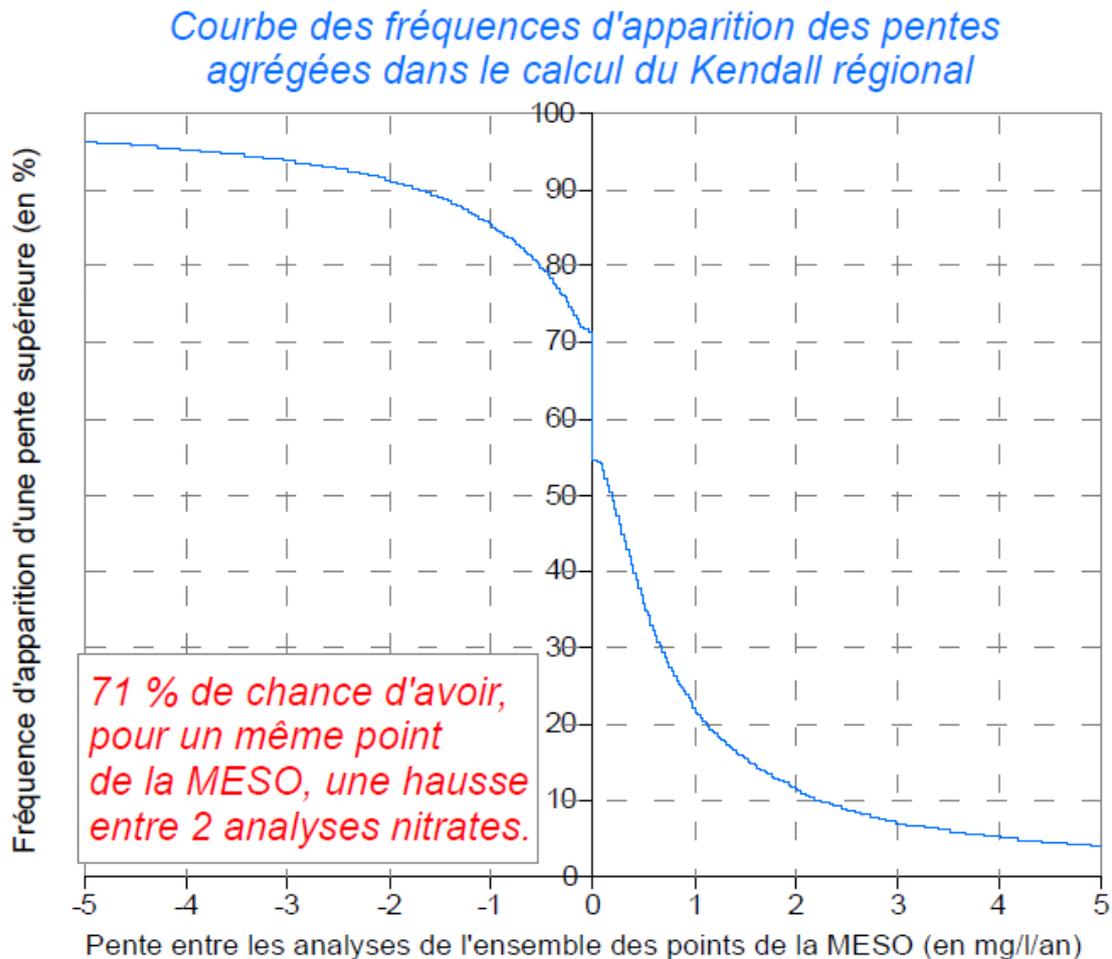


Illustration 9 : Exemple de visualisation par courbe de fréquence d'apparition des pentes agrégées pour le test Kendall régional

Sur l'exemple proposé, cette représentation permet de répondre aux questions suivantes :

- « Pourcentage de probabilité d'avoir, sur un même point de prélèvement de la ME, une hausse entre 2 analyses nitrates » → 71 % ;
- « Pourcentage de probabilité d'avoir, sur un même point de prélèvement de la ME, une hausse entre 2 analyses supérieure à 1 mg/L/an ? » → 20%.

- **Valeur de la pente et significativité statistique**

pente : 0.18 mg/l/an (*p-value = <0.0004 Tendance significative au seuil de confiance 5%*)

Une correspondance entre le paramètre Z des tests de Mann-Kendall et Kendall Régional et la p-value permet d'estimer la significativité du test Kendall Régional.

Si la p-value > 0,05 (5%), alors la tendance n'est pas significative au seuil de confiance 5%.

2.3. SYNTHÈSE DES RESULTATS A L'ECHELLE NATIONALE

	Points						Masses d'eau					
	Nombre de tendances significatives		Hausse		Baisse		Nombre de tendances significatives		Hausse		Baisse	
	N	% ¹	N	% ²	N	% ²	N	% ³	N	% ⁴	N	% ⁴
Adour-Garonne	356	12%	252	71%	91	26%	54	51%	43	80%	11	20%
Artois-Picardie	346	21%	289	84%	56	16%	14	78%	13	93%	1	7%
Guadeloupe	4	13%	1	25%	3	75%	2	33%	1	50%	1	50%
Guyane	1	5%	1	100%	0	0%	1	8%	1	100%	0	0%
Loire-Bretagne	696	15%	397	57%	287	41%	87	61%	51	59%	36	41%
Martinique	8	13%	3	38%	5	63%	2	33%	1	50%	1	50%
Réunion	30	14%	28	93%	2	7%	8	57%	8	100%	0	0%
Rhin-Meuse	385	8%	238	62%	141	37%	18	69%	15	83%	3	17%
Seine-Normandie	462	12%	310	67%	146	32%	32	60%	26	81%	6	19%
National	2288	13%	1519	66%	731	32%	218	56%	159	73%	59	27%
Rhône-Méditerranée-Corse	514	7%	223	44%	289	56%	94	37%	47	50%	47	50%

¹ Pourcentage par rapport au nombre total de point avec au moins une analyse nitrates, pour le bassin considéré.

² Pourcentage par rapport au nombre de points ayant une tendance significative, pour le bassin considéré.

³ Pourcentage par rapport au nombre de masses d'eau souterraine dans le bassin considéré.

⁴ Pourcentage par rapport au nombre de masses d'eau souterraine ayant une tendance significative, pour le bassin considéré.

3. Outil de caractérisation des chroniques et d'identification des tendances d'évolution de la qualité des eaux souterraines

3.1. PRESENTATION DE L'OUTIL

L'outil développé dans cette étude intègre les méthodes statistiques d'identification et de quantification des tendances d'évolution des concentrations en contaminants présélectionnées lors de l'étude réalisée en 2010 (Lopez et Leynet, 2011). Les méthodes concernent à la fois l'identification des dates de ruptures de pente des chroniques ainsi que l'estimation des tendances d'évolutions des contaminants dans les eaux souterraines, à l'échelle du point de prélèvement et à une échelle régionalisée.

Par rapport à l'étude de 2010, plusieurs ajouts ont été effectués : la recherche d'une inversion de tendance, l'analyse de la variabilité entre saison et le rééchantillonnage des chroniques pour extraire des sous-ensembles d'analyses.

L'outil permet de traiter une ou plusieurs chroniques de qualité pour n'importe quel élément ou paramètre (conductivité, température par exemple) d'intérêt. Plusieurs modules sont proposés qui peuvent être effectués indépendamment les uns des autres.

Un premier module permet d'extraire les principales caractéristiques des chroniques et propose une sortie sous forme de graphiques et une sortie sous forme de tableau.

Le module principal permet d'effectuer des tests de tendance et de rupture sur des chroniques.

Les autres modules permettent des analyses plus fines de tout ou partie des chroniques :

- Test de tendance saisonnier qui permet de rechercher des tendances dans des chroniques présentant des cycles saisonniers.
- Test régional qui permet de traiter des ensembles de chroniques.
- Ré-échantillonnage des chroniques pour extraire des sous-ensembles d'analyse à partir d'une chronique. Une fois les chroniques ré-échantillonnées, l'utilisateur peut lancer les autres modules sur ces chroniques partielles.

Avant de présenter l'outil en lui-même, quelques concepts théoriques et termes spécifiques au domaine sont rappelés.

3.2. QUELQUES RAPPELS STATISTIQUES

Ces définitions sont extraites du rapport BRGM de Lopez et Leynet (2011).

3.2.1. Stationnarité d'une chronique et tendance

Une des grandes questions dans l'étude de séries temporelles (ou chronologiques) est de savoir si celles-ci suivent un processus stationnaire. On entend par là le fait que la structure du processus sous-jacent supposé évolue ou non avec le temps. Si la structure reste la même, le

processus est alors stationnaire. Autrement dit, une chronique stationnaire aura toujours le même comportement que l'on regarde à la date t ou à la date $t+k$. Par exemple, une chronique dont la moyenne n'évolue pas avec le temps est stationnaire. Il en est de même si la variance et/ou l'autocorrélation n'évoluent pas avec le temps.

Les causes de non stationnarité peuvent être :

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

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

3.2.2. Le test statistique

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

3.2.3. L'hypothèse nulle

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

3.2.4. Niveau de confiance

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

Le niveau de confiance correspond à la probabilité qu'a la valeur du test statistique d'égaliser ou de dépasser la valeur observée lorsque l'hypothèse H_0 (pas de tendance) est vraie. Par

exemple, pour $\alpha = 0,05$ la valeur critique du test statistique est celle qui serait dépassée par 5 % des valeurs des tests statistiques réalisés sur des données aléatoires. Si la valeur du test statistique est plus importante que la valeur critique, alors l'hypothèse H_0 est rejetée.

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

Ainsi le niveau de confiance peut être interprété comme suit :

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

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

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

3.3. MISE EN FORME DES DONNEES

Les données à traiter doivent être sous format texte. Le caractère séparateur de colonne peut être :

- soit la barre verticale « | », (appelé tube ou *pipe* en anglais), comme c'est le cas dans les fichiers provenant d'exports ADES,
- soit une tabulation, ce que l'on obtient lorsque l'on enregistre un fichier au format Excel en un fichier au format texte.

Le fichier peut comporter un nombre de colonnes illimité mais certaines sont obligatoires. Les titres des colonnes doivent se situer sur la première ligne et doivent correspondre **exactement** aux titres décrits ci-dessous. Ils correspondent aux titres des colonnes que l'on obtient dans un export ADES, ce qui permet de traiter directement un fichier exporté d'ADES sans avoir à la manipuler au préalable.

Les colonnes obligatoires sont :

- le code BSS, appelé « CODE_BSS » ;
- la date d'analyse, appelée « DATE_DEBUT_PRELEVEMENT » ;
- un code correspondant à la qualification du résultat comme renseigné dans ADES, appelé « CODE_SIGNE ». Ce paramètre permet notamment le calcul des taux de quantification. Le tableau ci-dessous décrit les valeurs prises par ce paramètre selon le résultat de l'analyse.

CODE_SIGNE	
1	Résultat supérieur au seuil de quantification
2	Résultat inférieur au seuil de détection
10	Résultat inférieur au seuil de quantification
7	Traces (<seuil de quantification et > seuil de détection)

- la valeur du résultat, appelé « RESULTAT »
- une colonne précisant l'unité du résultat. Pour cette colonne, il y a deux possibilités :
 - o renseigner directement l'unité par son abréviation qui apparaîtra dans les légendes des graphiques. Le titre de la colonne doit être alors UNITE_GRAPH
 - o renseigner l'unité par son libellé complet – ce qui est le cas dans un export ADES – dans une colonne appelée « UNITE ». Le programme fera alors appel au fichier texte « Unite_SANDRE.txt », livré avec l'outil pour convertir cette unité en son abréviation.

Pour effectuer un test de Mann-Kendall régional, une colonne précisant l'unité spatiale, appelée « UNITE_SPATIALE » est également obligatoire, sinon, le test sera effectué sur la totalité des points.

Afin de renseigner la légende des axes sur les sorties graphiques, une colonne précisant le paramètre étudié est nécessaire. Ainsi, si le paramètre étudié est un élément chimique, on trouvera en légende le terme « Concentration », si c'est la conductivité, on trouvera « Conductivité », etc. Si le code paramètre n'est pas renseigné, par défaut, le terme « Concentration » est utilisé. S'il est donné, le paramètre doit l'être par son code –suivant les dénominations SANDRE³ - dans une colonne appelée « CODE_PARAMETRE ».

Les analyses renseignées comme non faites, ne disposant pas de résultats, sont écartées de l'analyse.

³ On pourra se reporter au site Internet du SANDRE pour connaître la nomenclature utilisée <http://www.sandre.eaufrance.fr/Parametres>

3.4. MODULE « CARACTERISATION »

Après avoir sélectionné le jeu de données pour la détermination de tendances, la première étape consiste à étudier leurs caractéristiques. Ce module permet d'appréhender la structure des données en proposant d'une part la représentation graphique des données et d'autre part le calcul des statistiques de base.

3.4.1. Calcul des statistiques de bases de la chronique

Dans cette première étape, l'outil permet de calculer les caractéristiques principales d'une chronique de qualité : **nombre d'analyses, longueur de la chronique, première et dernière date de prélèvement, moyenne et écart-type des valeurs, médiane, 1er et dernier décile, fréquence de quantification, limite de quantification minimum et maximum, normalité de la distribution des données, autocorrélation des données, moyenne des écarts entre deux dates d'analyses (en jours) et écart-type des écarts entre deux analyses.**

Les paragraphes suivants décrivent plus précisément ces différents paramètres.

a) *Moyenne*

Lorsque le résultat de l'analyse est inférieur à la limite de détection ou de quantification, la valeur de concentration prise en compte est la moitié de la limite de quantification ou de détection indiquée en accord avec les recommandations DCE.

b) *Ecart-type*

L'écart-type calculé est **l'écart-type d'un échantillon, non biaisé**, défini par la formule suivante :

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$
 où n est le nombre d'observations, x_i le vecteur des données et \bar{x} leur moyenne.

c) *Médiane/Déciles*

Lorsque le nombre de données est impair, la valeur de la médiane est la valeur centrale. Lorsque le nombre de données est pair, la valeur de la médiane est la moyenne arithmétique des deux valeurs centrales.

Les déciles sont les valeurs qui divisent le jeu de données en 10 parts égales. Le premier décile est la valeur qui sépare le jeu de données entre les 10% les plus bas et le reste des analyses. Le dernier décile sépare les 10% des analyses les plus élevées du reste. La procédure de calcul est détaillée sur l'illustration 10.

Calcul des déciles

De nombreuses définitions existent pour le calcul des déciles. Hyndman and Fan (1996) listent 9 méthodes différentes utilisées dans les logiciels de calcul statistique. Nous avons choisi d'utiliser la définition notée « définition 7 » dans l'article de Hyndman et Fan qui est communément utilisée. La définition est donnée ci-dessous

Soit n le nombre de valeurs dans la chronique et x_1, x_2, \dots, x_n les valeurs ordonnées de la plus petite à la plus grande.

Le i ème décile est défini par interpolation entre x_j et x_{j+1} . où j est la partie entière de $i/10 * (n - 1) + 1$ et g est la partie fractionnaire de ce nombre. On a donc $g = i/10 * (n - 1) + 1 - j$

Le i ème décile est ainsi défini par :

$$Q_i = x_j + g(x_{j+1} - x_j)$$

Illustration 10 : Calcul des déciles

Dans les colonnes « remarque sur la médiane/déciles » un texte est ajouté si la médiane ou l'un des déciles est supérieur à la limite de quantification maximale.

d) Fréquence de quantification

Le résultat d'une analyse est considéré comme une quantification dès lors que le code remarque est égal à 1. Aucune vérification n'est effectuée sur la valeur du résultat de l'analyse.

La fréquence de quantification est égale au rapport du nombre d'analyses supérieures à la limite de quantification (c'est-à-dire ayant un code remarque égal à 1) par le nombre total d'analyses.

e) Limites de quantification

Deux couples de limites de quantification minimum et maximum sont indiquées. Le premier correspond aux analyses indiquant une concentration inférieure à la limite de quantification (code signe égal à 10). Le second correspond à toutes analyses qui ne sont pas des quantifications (code signe différent de 1). Cette dernière limite peut alors correspondre à une limite de détection.

f) Fréquence d'échantillonnage

Pour caractériser la fréquence d'échantillonnage, deux valeurs caractéristiques sont calculées : la moyenne et l'écart-type des écarts entre deux analyses.

L'écart-type calculé est l'écart-type non biaisé (pour la définition, voir page 27).

g) Normalité de la distribution des données

La normalité de la distribution des données est testée, notamment afin de savoir s'il faut appliquer des tests paramétriques ou non paramétriques à ces données.

Elle est testée pour les chroniques disposant d'au moins 3 analyses en appliquant le test de Shapiro-Wilk (voir Illustration 11).

Si la p-value du test de Shapiro est inférieur à 0,05, on considèrera qu'on peut rejeter l'hypothèse nulle selon laquelle les données sont normalement distribuées.

Le test de normalité de Shapiro-Wilk (Shapiro & Wilk, 1965)

Le test de Shapiro-Wilk teste l'hypothèse nulle selon laquelle un échantillon provient d'une population normalement distribuée.

La statistique calculée est :

$$W = \frac{(\sum_{i=1}^n a_i x'_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

où \bar{x} est la moyenne de l'échantillon, x' est le vecteur contenant les données triées (x'_i est donc la i ème valeur la plus petite), n est le nombre d'observations, et :

$(a_1, \dots, a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{1/2}}$, où $m = (m_1, \dots, m_n)^T$ et les m_i sont les valeurs attendues d'une distribution normale pour un échantillon de taille n et V est la matrice de variance-covariance correspondante.

La p-value du test est une valeur exacte pour $n = 3$, sinon des approximations différentes sont utilisées pour $3 < n < 12$ d'une part et $n \geq 12$ d'autre part. L'algorithme utilisé pour le calcul de la p-value est celui proposé par Royston (1995).

Illustration 11: Test de normalité de la distribution des données

Pour les chroniques qui ne suivent pas une distribution normale, il faudra utiliser de préférence des tests non paramétriques pour la recherche de tendances et de ruptures.

h) Autocorrélation des données

Estimer l'autocorrélation des données revient à se demander si une valeur observée à un temps t dépend de ce qui a été observé dans le passé.

Il est important d'estimer cette autocorrélation des données car elle peut biaiser le niveau de significativité des tests statistiques (voir par exemple Renard, 2006). Dans des conditions environnementales, l'autocorrélation est toujours positive, ce qui tend à diminuer la significativité des tests statistiques.

L'intervalle de confiance à 95% est calculé pour être comparé à l'autocorrélation au rang 1.

Le test d'autocorrélation

L'autocorrélation au rang k se calcule par la formule suivante :

$$r_k = C_k/C_0, \text{ où } C_k = \frac{1}{n} \sum_{i=1}^{n-k} (x_i - \bar{x})(x_{i+k} - \bar{x}).$$

L'autocorrélation au rang 0 est égale à 1. Plus les données sont autocorrélées, plus l'autocorrélation est proche de 1 pour les rangs suivants.

La valeur des autocorrélations est ensuite comparée à la valeur limite définie ci-dessous à un seuil de significativité donné :

$$r_{lim} = \frac{1}{\sqrt{n}} qnorm\left(\frac{1+\alpha}{2}\right), \text{ où } qnorm \text{ est la fonction quantile d'une loi normale centrée réduite. } \alpha \text{ est le seuil de significativité. Nous l'avons choisi égal à } 0,95.$$

Si $r_k > r_{lim}$ l'autocorrélation au rang k est considérée comme significative.

Illustration 12 : Test d'autocorrélation

3.4.2. Représentation graphique

Au choix de l'utilisateur, il est possible de créer une représentation graphique des chroniques ainsi que leur représentation en boîte à moustaches (voir Illustration 13). Ces graphiques sont créés dans un document PDF regroupant l'ensemble des chroniques.

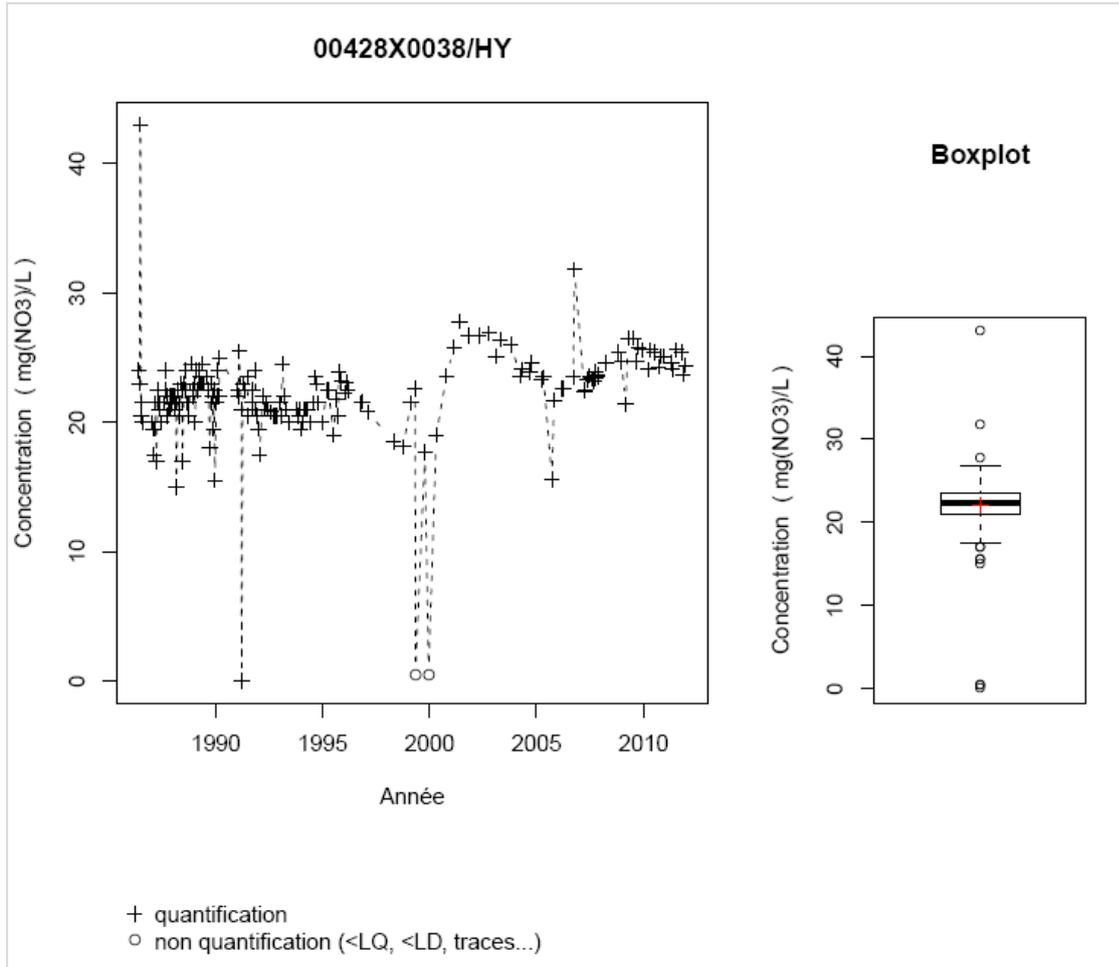


Illustration 13 : Exemple de sortie graphique pour une chronique de nitrates

a) Chronique

L'ensemble des analyses réalisées à un point BSS est représenté sur un graphe. Les données correspondant à des quantifications sont distinguées des données non quantifiées (<LQ, <LD, traces...).

« Boîte à moustaches »

La représentation en boîte à moustaches est un moyen rapide de représenter plusieurs caractéristiques d'une série de données. Elle a été inventée en 1977 par John Tukey. Ce diagramme (Illustration 14) résume les caractéristiques suivantes : médiane, quartiles, minimum, maximum et valeurs extrêmes (outliers en anglais).

Le rectangle représente l'écart entre le premier quartile et le troisième quartile. Il est coupé par la médiane. Sont ensuite ajoutées des segments (les moustaches) aux extrémités de la boîte menant jusqu'à la valeur minimum de la série vers le bas et maximum vers le haut. Les valeurs minimum et maximum sont déterminées sans prise en compte des valeurs extrêmes. Ces valeurs extrêmes correspondent à des valeurs inférieures à $QL - 1,5 \cdot IQR$ ou supérieure à $QU + 1,5 \cdot IQR$ avec :

IQR : différence entre les quartiles, QL : valeur du premier quartile (bas de la boîte) et QU : valeur du troisième quartile (haut de la boîte)

Ces valeurs extrêmes sont représentées en points individuels sur le diagramme.

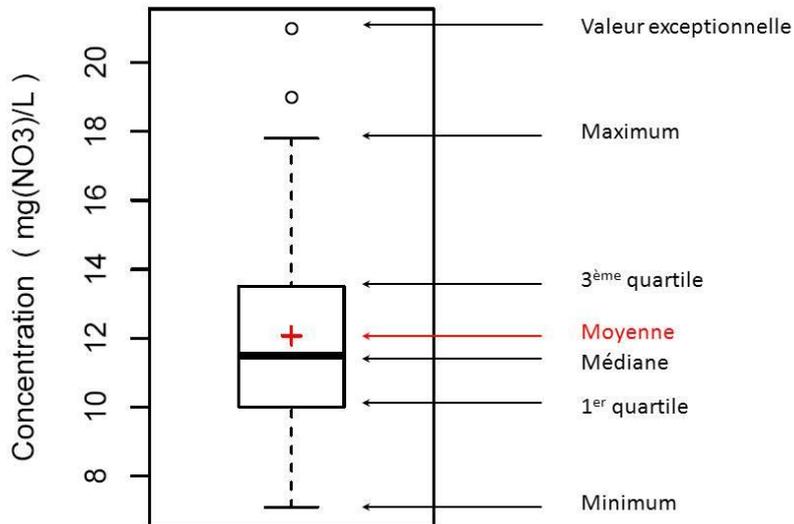


Illustration 14 : Aide à la lecture d'un diagramme en « boîte à moustache »

3.5. MODULE « TENDANCES ET RUPTURES »

Ce module permet d'appliquer des tests statistiques de détection de tendance et de rupture sur les chroniques. Il permet également de tester la variabilité des données suivant les saisons. Dans le cas où les saisons sont significativement différentes les unes des autres, il peut être intéressant d'appliquer un test de Kendall saisonnier pour la recherche de tendance ou de rééchantillonner la chroniques. Ces procédures font partie des modules complémentaires décrits dans le paragraphe 3.6.

Le tableau suivant récapitule les différents tests effectués dans ce module. Les tests appliqués dépendent des caractéristiques des chroniques. Les différents tests sont détaillés dans les paragraphes suivants.

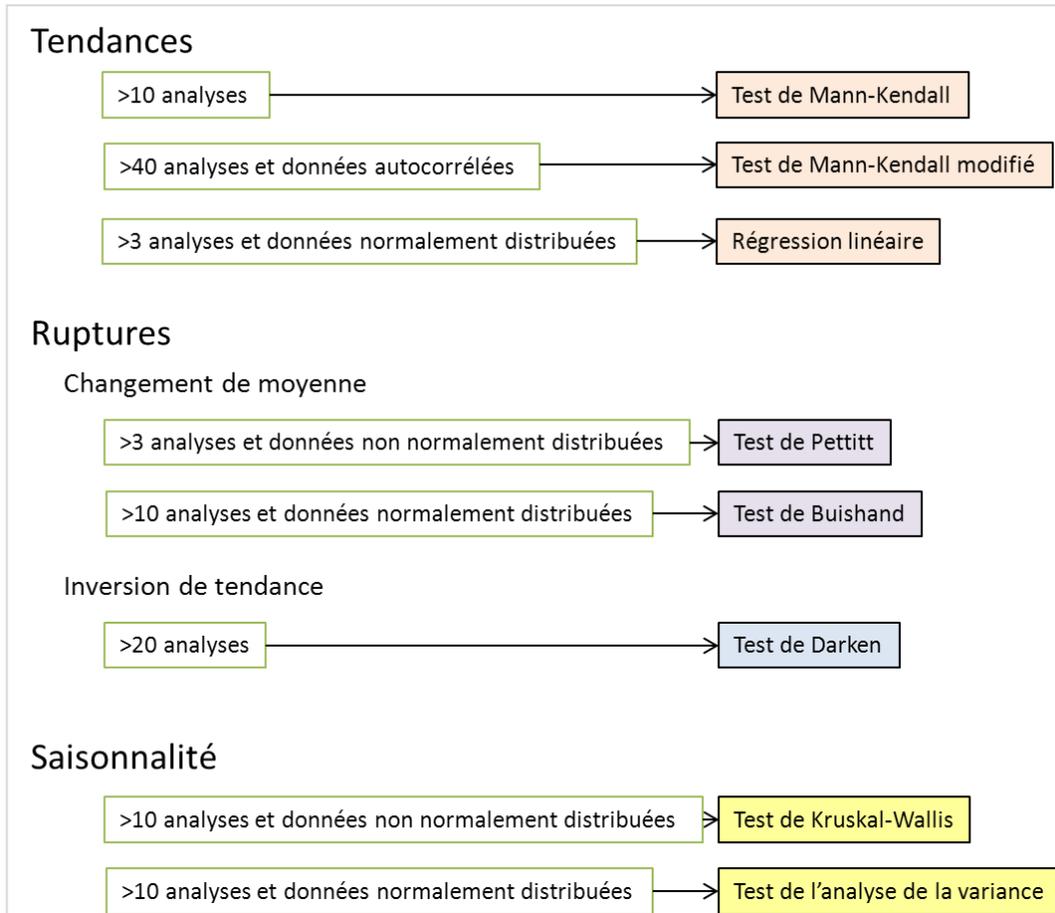


Illustration 15 : Schéma récapitulatif des différents tests effectués dans le module « Tendances et ruptures ».

Pour ces calculs, tous les résultats sont pris en compte, qu'il s'agisse de réelles valeurs de quantification ou de limite de quantification/détection.

3.5.1. Tests de tendance

La recherche de tendance s'effectue par la mise en œuvre du test de Mann-Kendall sur l'ensemble des chroniques quelles que soient leurs distributions.

La régression linéaire est appliquée en complément sur les données normalement distribuées, ce test s'étant révélé plus puissant dans ces conditions spécifiques (Lopez et Leynet, 2011).

a) Test de tendance non paramétrique (Mann-Kendall)

Le test de Mann-Kendall (décrit dans l'encart ci-dessous) est associé au calcul de la pente de Sen (décrit sur l'illustration 17). Il est appliqué sur les toutes les chroniques non stationnaires disposant d'au moins 10 analyses.

La tendance est dite significative d'un point de vue statistique lorsque la p-value du test est inférieure à 5%.

Le test de Mann-Kendall (Kendall, 1938, repris par Renard, 2006)

L'hypothèse H_0 testée est l'absence de tendance.

La statistique calculée est définie comme suit :

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \operatorname{sgn} [(y_j - y_i)(x_j - x_i)]$$

où la fonction sgn est la définie par : $\operatorname{sgn}(X) = 1$ pour $X > 0$; $\operatorname{sgn}(X) = 0$ pour $X = 0$ et $\operatorname{sgn}(X) = -1$ pour $X < 0$.

Mann (1945) et Kendall (1975) ont démontré que

$$E(S) = 0$$

$$\operatorname{Var}(S) = n(n-1)(2n+5)/18$$

Dès que l'échantillon contient une dizaine de données, la loi de la statistique de test Z ci-dessous peut-être approché par une gaussienne centrée-réduite.

$$Z = \frac{S-1}{(\operatorname{Var}(S))^{1/2}} \text{ si } S > 0$$

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

$$Z = \frac{S+1}{(\operatorname{Var}(S))^{1/2}} \text{ si } S < 0$$

S'il y a des ex-aequo dans la série, la variance de S est corrigée de la façon suivante :

$$\operatorname{Var}(S) = 1/18 \left[n(n-1)(2n+5) - \sum_{p=1}^g t_p(p-1)(2p+5) \right]$$

où t_p est le nombre d'égalités impliquant p valeurs.

Illustration 16 : Test de tendance de Mann-Kendall (Kendall, 1938)

Calcul de la pente de Sen et de l'ordonnée à l'origine

La pente de la droite de régression (appelée pente de Kendall-Theil ou pente de Sen) est effectué par la méthode de Sen (Sen, 1968).

La pente est la médiane de toutes les pentes calculées entre chaque paire de point. L'estimation de l'ordonnée à l'origine peut être calculée de différentes manières. La méthode retenue est celle recommandée par Helsel et Hirsch (2002) utilisant la pente médiane et la médiane des variables (Conover, 1980).

$$pente_{Sen} = \text{mediane}_{i < j} \left\{ \frac{y_j - y_i}{x_j - x_i} \right\}$$

$$\text{ordonnée}_{orig} = \text{mediane}(y) - pente_{Sen} * \text{mediane}(x)$$

Illustration 17 : Calcul de la pente de Sen et de l'ordonnée à l'origine pour le test de Mann-Kendall

b) Test de tendance paramétrique : la régression linéaire

La régression linéaire (détaillée sur l'illustration 18) est calculée pour les chroniques disposant d'au moins 5 analyses et ayant des données normalement distribuées. Les hypothèses sous-jacentes à l'application d'une régression linéaire sont : la normalité de la distribution des données, l'homogénéité de la variance et une relation linéaire entre la variable expliquée et la variable explicative

Le r^2 est calculé ainsi que la pente de régression.

La p-value de la régression est donnée. La tendance est dite significative d'un point de vue statistique lorsque la p-value du test < 0.05 (5%).

La régression linéaire (selon Renard, 2006)

L'hypothèse H_0 testée est que les données ne sont pas linéairement dépendantes du temps.

Ce test est basé sur le modèle paramétrique suivant :

$$X = \alpha + \beta t + \varepsilon$$

où les erreurs ε suivent une loi normale centrée.

La valeur de la pente et l'ordonnée à l'origine sont définies de façon à minimiser la somme des carrés des écarts entre les valeurs observées et les valeurs de la droite de régression.

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

$$\hat{\alpha} = \bar{y} - \hat{\beta} * \bar{x}$$

La variance de l'estimateur de tendance peut être estimée par :

$$Var(\hat{\beta}) = \frac{12 \sum_{i=1}^n (y_i - \bar{y} - \hat{\alpha} - \hat{\beta} x_i)^2}{(n-2)n(n^2+1)}$$

Le test de la régression linéaire consiste alors à vérifier que l'estimateur du coefficient β est proche de 0. Pour cela on compare la statistique Z définie ci-dessous aux quantiles d'une loi de Student à $n-2$ degré de liberté :

$$Z = \frac{\hat{\beta}}{\sqrt{Var(\hat{\beta})}}$$

Illustration 18 : Régression linéaire

3.5.2. Modification des tests de tendance dans le cas d'autocorrélation

Si les données présentent une autocorrélation significative, le niveau de significativité des tests statistiques peut être biaisé. La présence d'autocorrélation tend à augmenter la significativité des tests statistiques, il convient donc d'apporter quelques modifications aux tests précédemment décrits.

Plusieurs approches sont possibles pour travailler avec des chroniques présentant de l'autocorrélation :

- Appliquer les tests de tendance modifiés. Ces tests sont décrits ci-dessous.
- Appliquer un test prenant en compte les variations saisonnières : le test de Mann-Kendall saisonnier. Pour cela, une première étape est de tester la variabilité entre saisons.
- Travailler sur un sous-échantillon de la série.

Les deux dernières méthodes sont présentées dans un troisième module (§3.6, page 45).

a) Test de Mann-Kendall modifié (Hamed, 1998)

Ce test ne peut être appliqué que si la chronique dispose de 40 analyses ou plus.

Le principe repose sur une modification du test S de Mann-Kendal plutôt que de modifier les données elles-mêmes :

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

où γ est un facteur correctif appliqué à la variance.

Ainsi, seule la p-value du test de Mann-Kendall est modifiée ; la pente est la même que celle calculée pour le test de Mann-Kendall non modifié.

Test de Mann-Kendall modifié (Hamed, 1998)

La modification du test correspond au fait qu'un échantillon autocorrélé positivement de taille n se comporte comme un échantillon indépendant de taille $n^* < n$ (et inversement pour un échantillon autocorrélé négativement).

Plusieurs méthodes de calcul de γ sont relevées dans la littérature. Nous retiendrons la méthode proposée par Hamed et Rao (1998), légèrement plus puissante que la formule proposée par Yue et Wang (2004), d'après les tests effectués par Renard (2006). La formule proposée par Hamed et Rao se base sur une formule empirique spécifiquement calculée pour corriger la statistique de Mann-Kendall. Elle prend en compte les autocorrélations des résidus de régression calculées aux différents rangs si celles-ci sont significatives :

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

où n est le nombre de données et ρ_k est l'autocorrélation à l'ordre k , si elle est significative, $\rho_k = 0$ sinon

Le seuil de significativité choisi pour l'autocorrélation est 5%.

Illustration 19 : Test de Mann-Kendall modifié pour la prise en compte de l'autocorrélation.

b) Régression linéaire modifiée

En théorie, il est possible d'appliquer des corrections à la régression linéaire. En pratique, il existe de nombreuses méthodes très différentes pour les corrections à apporter (Green, 2000, Woodward et Gray, 2003, Renard, 2006). Un travail important reste à faire pour déterminer quel test il est pertinent d'appliquer sur les données de qualité des eaux souterraines.

Aucune méthode n'est implémentée dans l'outil.

c) Analyse de la variabilité entre période

Deux tests sont utilisés par l'outil pour étudier la variabilité entre périodes : un test paramétrique, l'analyse de la variance (ANOVA), si les données sont normalement distribuées et un test non paramétrique, le test de Kruskal-Wallis si les données ne sont pas normalement distribuées.

Ce test permet d'évaluer si une variable explicative a une influence sur les données. Deux variables explicatives qui correspondent à deux périodes sont testées dans l'outil : le trimestre (saison) et le mois.

- **ANOVA**

Ce test est appliqué pour les chroniques présentant au moins 10 valeurs et ayant une distribution normale.

Si la p-value du test est inférieure à 0.05, on peut considérer qu'au moins deux périodes sont significativement différentes l'une de l'autre.

Analyse de la variance (ANOVA) à un facteur

L'hypothèse nulle de ce test est que les moyennes des différentes périodes sont identiques.

La statistique calculée est la statistique de Fisher, définie comme suit :

$F = \frac{s_{\alpha}^2}{s_r^2}$, où s_{α}^2 et s_r^2 sont respectivement la variance due au facteur (ou variance inter-classe) et la variance résiduelle (ou variance intra-classe).

$$s_{\alpha}^2 = \frac{\sum_{i=1}^g n_i (\bar{x}_i - \bar{x})^2}{k - 1}$$

$$s_r^2 = \frac{\sum_{i=1}^g (n_i - 1) \sigma_i^2}{n - k}$$

g est le nombre de périodes disposant de n_i échantillons, \bar{x}_i est la moyenne et σ_i^2 la variance des valeurs dans le groupe i . \bar{x} est la moyenne globale de l'échantillon.

La statistique est ensuite comparée aux quantiles d'une loi de Fisher à $(g - 1)$ degrés de liberté au numérateur, et $(n - g)$ degrés de liberté au dénominateur.

Illustration 20 : Analyse de la variance (ANOVA) à un facteur.

- **Test de Kruskal-Wallis**

De même que le test d'analyse de la variance, le test de Kruskal-Wallis permet de déterminer si les données d'une période sont significativement différentes d'une autre. Il est appliqué pour les chroniques dont les données ont une distribution non normale et disposant d'au moins 10 valeurs.

Si la p-value du test est inférieure à 0.05, il y a une différence entre les données d'au moins deux périodes.

Test de Kruskal-Wallis

L'hypothèse nulle est que les périodes ne sont pas différentes les unes des autres.

La statistique calculée est définie comme suit :

$$K = (n - 1) \frac{\sum_{i=1}^g n_i (\bar{r}_i - \bar{r})^2}{\sum_{i=1}^g \sum_{j=1}^{n_i} (r_{ij} - \bar{r})^2}$$

où g est le nombre de période, n_i est le nombre d'observation dans la période i , r_{ij} est le rang de l'observation j dans le groupe i . n est le nombre total de données, $\bar{r}_i = \frac{\sum_{j=1}^{n_i} r_{ij}}{n_i}$ et $\bar{r} = \frac{1}{2(n+1)}$

La statistique est ensuite comparée aux quantiles d'une loi du chi 2 à $(g-1)$ degrés de liberté.

Illustration 21 : Test de Kruskal-Wallis.

3.5.3. Tests de rupture

Deux types de rupture sont recherchés : une rupture dans la moyenne et une inversion dans la tendance.

a) Changement de moyenne.

Pour rechercher un changement de moyenne dans la chronique, deux tests d'homogénéité peuvent être appliqués. Si les données sont normalement distribuées, le test appliqué est le test paramétrique de Buishand. Dans le cas contraire, on applique le test non paramétrique de Pettitt.

Si une rupture significative est détectée, les moyennes sur les tronçons pré- et post-rupture sont calculées.

- **Test non paramétrique de Pettitt**

Le test de Pettitt est un test non paramétrique qui dérive du test de Mann-Whitney. Ce test est appliqué sur les chroniques non stationnaires disposant d'au moins 3 données et dont la distribution est non normale.

La rupture est dite significative d'un point de vue statistique lorsque la p-value du test est inférieure à 5%.

Le test de Pettitt (Pettitt, 1979)

Le test de Pettitt est non paramétrique. Il dérive du test de Mann-Whitney. L'hypothèse nulle est l'absence de rupture dans la chronique. Elle est testée par la statistique $U_{t,n}$ considérée pour l'ensemble des valeurs de t telles que $1 \leq t \leq n$:

$U_{t,n} = \sum_{i=1}^t \sum_{j=t+1}^n D_{ij}$ où : $D_{ij} = \text{sgn}(X_i - X_j)$ où X_i est le vecteur des données trié par date et la fonction sgn est définie par :

$\text{sgn}(X) = 1$ pour $X > 0$; $\text{sgn}(X) = 0$ pour $X = 0$ et $\text{sgn}(X) = -1$ pour $X < 0$

On utilise alors la variable K_n pour tester H_0 telle que $K_N = \max |U_{t,n}|$.

Si k correspond à la valeur de K_n , la probabilité de dépassement de la valeur k est donnée par :

$$\Pr(K_n > k) \sim 2 \exp \left[-\frac{6k^2}{(n^3 + n^2)} \right]$$

Si α est supérieur à cette probabilité, H_0 est rejetée. La série présente alors une rupture au temps t définissant K_n .

Illustration 22 : Test d'homogénéité de Pettitt

- **Test paramétrique de Buishand**

Le test de Buishand est un test paramétrique. Il est appliqué pour les chroniques disposant d'au moins 10 valeurs et dont la distribution est normale. Ce test suppose un non changement de la variance de la série.

Le test de Buishand (Buishand, 1982, 1984)

L'hypothèse H_0 est l'absence de rupture dans la chronique.

Ce test est construit à partir des écarts cumulés à la moyenne jusqu'à un rang k :

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

La statistique de test est obtenue par la division des valeurs S_k par la déviation standard :

$$Z = \max \left(\frac{|S_k|}{\sqrt{n} \sqrt{\sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n}}} \right)$$

Une valeur de Q élevée est un signe d'une rupture dans la chronique. La significativité du test est calculé en comparant la valeur de Z à des valeurs critiques.

Les valeurs critiques prises en compte sont celles évaluées par Buishand (1982) par la génération de séquences aléatoires. Ces valeurs sont données dans l'illustration 23.

N	Q/ \sqrt{n} au niveau de confiance		
	$\alpha = 0,10$	$\alpha = 0,05$	$\alpha = 0,01$
10	1,05	1,14	1,29
20	1,10	1,22	1,42
30	1,12	1,24	1,46
40	1,13	1,26	1,50
50	1,14	1,27	1,52
100	1,17	1,29	1,55
∞	1,22	1,36	1,63

Illustration 23: Valeurs critiques de Z du test de Buishand

Illustration 24 : Test paramétrique d'homogénéité (Buishand, 1982)

b) Inversion de la tendance

Darken (1999) propose dans sa thèse méthode pour détecter un changement de signe de la pente (et non un changement dans la magnitude de la tendance sans changement de signe) qui est valable pour des données non normalement distribuées. Le test est décrit ci-dessous.

Si la p-value du test est inférieure à 5%, la chronique présente une inversion de tendance significative. La date de rupture est déterminée ainsi que les tendances avant et après cette date de rupture.

Changement de pente de Darken (1999)

Darken (1999) propose dans sa thèse deux méthodes basées sur le tau de Kendall.

Pour un changement de tendance (= changement de signe), il propose la statistique suivante :

$$Z = \frac{\tau_1 - \tau_2}{\sqrt{\text{Var}(\tau_1) + \text{Var}(\tau_2)}}$$

Les variances sont calculées comme décrit par Kendall (1976).

La date de rupture la plus probable est la date pour laquelle Z est maximum.

La p-value du test est calculée en comparant la statistique Z pour la date de rupture identifiée aux quantiles d'une loi normale centrée réduite (Darken, 1999).

Illustration 25 : Test d'inversion de tendance (Darken, 1999)

- **Remarque sur la recherche d'un changement de pente (sans inversion de signe)**

Plusieurs méthodes existent pour déterminer un changement de pente de tendance. Sen (1980) a développé une méthode semi-paramétrique basée sur l'analyse des résidus de la régression. Elle ne peut pas traiter des séries autocorrélées et suppose qu'il n'y a pas d'ex-aequo.

Darken (1999) propose une méthode, applicable avec des hypothèses moins contraignantes, basée sur le calcul des intervalles de confiance sur les pentes de régression calculées avant et après rupture. Salmi (2002) propose une méthode similaire mais pour laquelle les intervalles de confiance sont calculés différemment.

Etant donné qu'il n'y a pas de consensus sur la méthode de calcul des intervalles de confiance, un travail important serait nécessaire pour déterminer la méthode la plus adaptée aux chroniques de qualité des eaux souterraines. Aucune de ces méthodes n'est donc implémentée dans l'outil.

3.5.4. Sorties générées par l'outil

a) *Tableau récapitulatif*

Un tableau récapitulatif est créé par l'outil. Il contient, pour chaque point BSS, les principales caractéristiques des chroniques, l'ensemble des tests effectués ainsi que leurs résultats en termes de significativité. Si les tests sont significatifs, les pentes et dates de rupture déterminées sont également données.

Le tableau contient également le résultat des tests d'analyse de la variabilité entre périodes

Sortie graphique

La sortie graphique est un fichier PDF, composé d'une page par point BSS.

L'illustration 26 présente un exemple de sortie graphique obtenue pour une chronique de nitrates. Le graphique présente la chronique étudiée, les différents tests de rupture et de

tendance effectués, leurs résultats et dans le cas où ils sont significatifs, les tendances et/ou ruptures détectées sont tracées sur le graphique.

S'il y a un changement de moyenne significatif détecté, les moyennes avant et après rupture sont représentées. S'il y a une inversion significative de la tendance, les tendances avant et après rupture sont représentées si elles sont significatives.

Dans le cas où le taux de quantification sur l'ensemble d'une chronique est inférieur à 50%, un avertissement apparaît en rouge sous le graphe. En effet, les résultats des tests sont à considérer dans ce cas avec précaution.

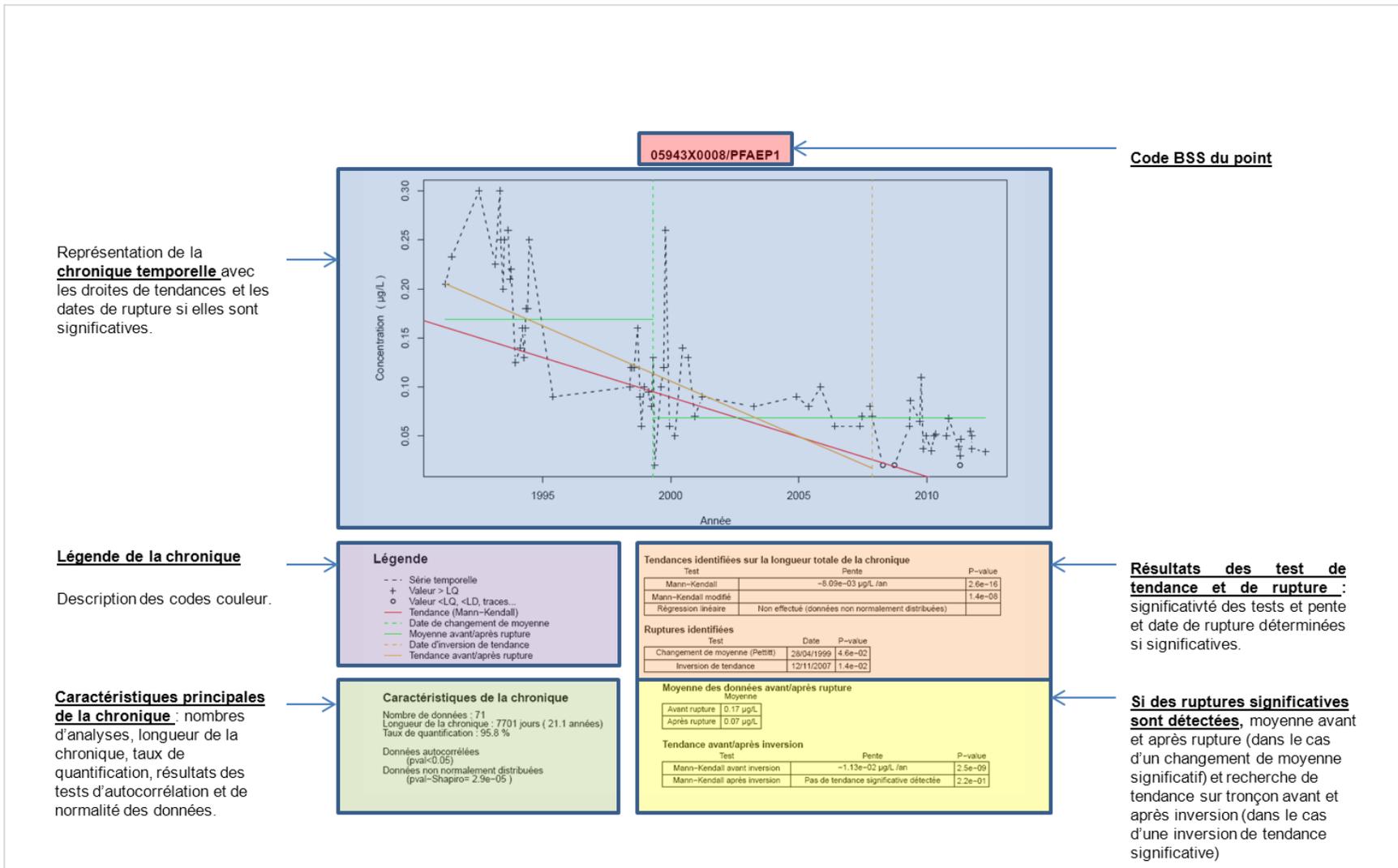


Illustration 26 : Description d'une sortie graphique obtenue avec le module « Tendances et ruptures » pour une chronique de nitrates au point BSS 05943X00008.

3.6. MODULES COMPLEMENTAIRES (UTILISATION AVANCEE)

Plusieurs modules supplémentaires permettent d'effectuer des traitements plus avancés sur tout ou partie d'une chronique.

Deux modules : « Saisonnier » et « Régional », permettent d'effectuer des tests avancés de tendance: un test saisonnier et un test régional.

Un dernier module « Rééchantillonnage » permet d'extraire des sous-ensembles d'une chronique.

3.6.1. Module « Saisonnier »

Ce module permet d'effectuer un test de Kendall saisonnier. Ce test, proposé par Hirsch et al. (1982) permet d'estimer des tendances de séries cycliques saisonnières. Le principe est identique à celui du test de Mann-Kendall mais le caractère saisonnier de la série est pris en compte. Autrement dit pour des données mensuelles ayant une saisonnalité de 12 mois, on ne va pas chercher à savoir s'il y a une croissance au global sur la série, mais simplement si, d'un mois de janvier à l'autre, d'un mois de février à l'autre, et ainsi de suite, il y a une tendance. L'outil peut être utilisé sur deux périodes différentes : le trimestre (de janvier à mars, avril à juin, juillet à septembre et octobre à décembre) et le mois calendaire.

La statistique S_k de Kendall se calcule à partir de la somme des statistiques pour chaque saison (Hirsch, 1982).

$S_k = \sum_{i=1}^s S_i$ où s est le nombre de saisons.

Et $\tau = \frac{\sum_{i=1}^s S_i}{\sum_{i=1}^s D_i}$

La statistique calculée est :

$$Z_{S_k} = \frac{S_k}{\sigma_{S_k}}$$

Si le produit du nombre de saisons par le nombre d'années est supérieur à 25, la distribution des S_k peut être approximé par une distribution normale avec une moyenne nulle et la variance égale à la somme des variances. S_k est standardisé par soustraction de sa moyenne (égale à 0) et division par son écart-type. Le résultat est évalué en comparant à une table de distribution normale standard. L'hypothèse nulle est à rejeter à un niveau de significativité α si $|Z_{S_k}| > Z_{crit}$ où Z_{crit} est la valeur de la distribution normale avec une probabilité de dépassement de $\alpha/2$.

Illustration 27 : Test de Kendall saisonnier

Le calcul de la pente de Sen et l'ordonnée à l'origine est légèrement modifié par rapport au test de Mann-Kendall (cf Illustration 17) : la pente est la médiane des pentes calculées entre deux

analyses **au sein d'une même période**. Les pentes calculées entre des analyses contenues dans des périodes différentes ne sont pas prises en compte.

3.6.2. Module « Régional »

Ce module permet d'effectuer un test de Kendall régional.

Ce test permet d'étudier la présence de tendance à l'échelle d'une région d'étude, appelée ici unité spatiale, comprenant plusieurs sites.

Pour utiliser ce module, il faut que le tableau des données contienne une colonne «UNITE_SPATIALE » précisant à quelle unité spatiale est rattaché le point.

Le principe du test est de déterminer si une tendance cohérente peut être mise en évidence à partir des différentes chroniques. Il suit exactement le même principe que le test de Kendall saisonnier. On pourra donc se reporter au paragraphe précédent pour la mise en application de ce test ; la « saison » est à remplacer par l'« unité spatiale ».

3.6.3. Module « Rééchantillonnage »

Ce module permet d'extraire d'une chronique les analyses effectuées avec une périodicité particulière. Les données ainsi extraites sont écrites dans un fichier résultat sous une forme lisible par les modules précédemment cités. L'ensemble des tests de caractérisation de la chronique et de recherche de tendances et de ruptures peuvent donc être effectués sur ces chroniques partielles.

a) Périodicité annuelle

Il peut être intéressant de travailler avec un sous-ensemble de la chronique comprenant les analyses ayant lieu à la même période de l'année.

L'outil identifie les sous-ensembles d'analyses ayant eu lieu au même moment plus ou moins 35 jours, quel que soit le nombre d'année d'écart entre les analyses.

La période peut être choisie par l'utilisateur ou alors définie automatiquement par le logiciel comme la période disposant du plus d'analyses.

Si pour une année, la chronique dispose de plusieurs analyses, la valeur prise en compte est la moyenne des résultats des analyses. Si, parmi ces analyses, il y a au moins une quantification, le résultat pour la période est considéré comme une quantification.

Le module crée les fichiers suivants :

- Dans le cas où la recherche est faite de manière automatique, un fichier xls contenant un tableau récapitulatif des sous-ensembles d'analyses effectuée à la même période de l'année qui comprend :
 - o Le point BSS
 - o La date qui définit la période de recherche
 - o Le nombre d'années disposant d'une analyse

- La longueur totale (en année) du sous-ensemble identifié
- Dans tous les cas, un fichier texte reprenant les analyses effectuées avec une périodicité annuelle pour les points BSS disposant d'une chronique annuelle de plus de 10 années. Seuls les champs obligatoires (code BSS, date, code signe, résultat, unité et code paramètre si présent initialement) sont repris dans ce fichier. Dans le cas où le traitement est automatique, seule la chronique la plus longue est prise en compte pour chaque point BSS.

b) Périodicité mensuelle

De même, l'outil peut identifier des sous-ensembles de chroniques ayant une périodicité mensuelle. L'outil identifie des sous-ensembles d'analyses ayant eu lieu le même jour du mois +/- 5 jours, pour des mois consécutifs ou presque : l'outil ne retient que les chroniques pour lesquelles 80% des mois disposent d'une analyse – par exemple, pour une chronique de deux ans, il faut qu'il y ait au moins 19 analyses.

4. Utilisation de l'outil sur des chroniques représentatives

4.1. SELECTION DE CHRONIQUES REPRESENTATIVES

4.1.1. Contexte et méthode

Le contexte de cette partie de l'étude est de sélectionner des chroniques réelles d'évolution des concentrations en contaminants dans les eaux souterraines représentatives de la diversité des typicités rencontrées dans la base de données nationale ADES sur la qualité des eaux souterraines.

Différents paramètres de la base de données ADES ont été sélectionnés. Les extractions ont été réalisées en ne prenant en compte que les points du réseau de Contrôle de surveillance de l'état chimique des eaux (FR_SOS), utilisé pour le suivi lié à la DCE. Ces chroniques contiennent l'ensemble des points de prélèvement affiliés au réseau. Les chroniques obtenues ont été décrites statistiquement grâce à l'outil présenté ci-dessus. Un ensemble de caractéristiques a ainsi été obtenu pour chaque point de prélèvement. Cinq caractéristiques ont été jugées intéressantes pour décrire les séries temporelles et les analyser. Ces caractéristiques sont : le nombre d'analyses dans la chronique, la moyenne des écarts entre deux analyses, le résultat du test de normalité, le résultat du test d'autocorrélation et la fréquence de détection. Grâce à cet ensemble de caractéristiques décrivant les points de prélèvement, il a été possible de réaliser des groupes puis d'en extraire les individus les plus représentatifs.

4.1.2. Extraction et données utilisées

a) Extraction

Une extraction de données sur le réseau de contrôle de surveillance de l'état chimique des eaux (FR_SOS) a été réalisée pour obtenir l'ensemble des chroniques disponibles dans la base de données ADES. Ce réseau a été sélectionné car c'est celui qui fait référence dans le cadre de rapportage à l'Europe pour l'état chimique des eaux souterraines. Ainsi, la représentativité et la qualité des points de prélèvement de ce réseau sont censées être bonnes sur l'ensemble du territoire français.

L'extraction a été réalisée le 12 décembre 2012. 1824 points de prélèvement, différenciés par des codes BSS différents, étaient contenus dans le réseau.

Paramètres

Sur le grand nombre de paramètres présents dans la base de données ADES, deux paramètres présentant des caractéristiques différentes ont été sélectionnés. Il s'agit des nitrates (code sandre 1340) et de l'atrazine (code sandre 1107). Il s'agit de deux contaminants d'origine essentiellement agricole.

Le paramètre "nitrates" a été sélectionné car il est souvent recherché et souvent détecté dans les nappes (Blum et al., 2011). Plus de 20% des masses d'eau ont été déclassées, durant le

dernier rapportage, à cause d'une concentration trop élevée en nitrates (SOeS, 2012). Le paramètre "atrazine" a été sélectionné car ce produit phytosanitaire est souvent recherché et détecté dans les eaux souterraines (SOeS, 2011). C'est un herbicide sélectif surtout utilisé sur maïs, mais il est interdit en France depuis 2003.

Réalisation des groupes

- **Approche automatique**

L'approche automatique a consisté à réaliser les groupes de points de prélèvement à partir des caractéristiques des chroniques temporelles, à l'aide d'une Classification Ascendante Hiérarchique (CAH) réalisée sous Excel avec le module Excel Stat.

La classification ascendante hiérarchique (CAH) est une méthode de classification itérative en phases qui se répètent. D'abord, il s'agit de calculer de la dissimilarité entre les N objets. Ensuite, les deux objets, dont le regroupement minimise un critère d'agrégation donné, sont regroupés, créant ainsi une classe comprenant ces deux objets. Le calcul de la dissimilarité entre cette classe et les N-2 autres objets en utilisant le critère d'agrégation est ensuite réalisé. Enfin, les deux objets ou classes d'objets dont le regroupement minimise le critère d'agrégation sont regroupés. Les calculs de dissimilarité puis de regroupement se poursuivent ainsi de suite jusqu'à ce que tous les objets soient regroupés. Ces regroupements successifs produisent un arbre binaire de classification (dendrogramme), dont la racine correspond à la classe regroupant l'ensemble des individus. Ce dendrogramme représente une hiérarchie de partitions.

Dans cette étude, la Distance Euclidienne a été utilisée pour le calcul en tant qu'indice de dissimilarité. La Distance Euclidienne est probablement le type de distance le plus couramment utilisé en statistique. Il s'agit simplement d'une distance géométrique dans un espace multidimensionnel. La méthode du lien simple et la méthode de Ward ont été testées en tant qu'indice d'agrégation.

Aucune des méthodes automatiques qui ont été essayées n'ont permis d'aboutir à de bons résultats. Les tentatives de regroupement automatique ont juste permis de différencier les sites de suivi DCE qui ont plusieurs années d'existence, de ceux qui sont en place depuis très peu de temps. Ainsi, la première classe contient plus de 1000 individus et est caractérisée par un nombre d'analyses important. A l'inverse, les autres classes contiennent très peu d'individus et sont caractérisées par un nombre d'analyses faible. Cette approche de classification n'avait aucun intérêt pour cette étude, l'approche manuelle pour la classification a été privilégiée.

- **Approche manuelle**

L'approche manuelle a consisté à réaliser des tris successifs sur la liste de code BSS. L'ordre des tris est celui proposé dans l'arbre décisionnel pour le choix de la méthode statistique à appliquer sur les chroniques d'évolution de la qualité des eaux souterraines proposé dans le rapport BRGM de Lopez et Leynet (2011).

Dans cette approche manuelle, la première classification est réalisée sur les individus (c'est-à-dire les points de prélèvement) à partir du résultat du test d'autocorrélation. Ensuite, sur les classes créées, un nouveau tri est réalisé. Celui-ci est basé sur le test de normalité.

Dans les deux cas, trois classes sont créées : résultat positif au test (ex: la chronique est normale), résultat négatif au test (ex: la chronique n'est pas normale) et aucun résultat (ex : pas assez de données pour réaliser le test de normalité).

Un dernier tri a été réalisé sur les classes déjà créées. Ce tri est basé sur le taux de quantification pour créer deux classes. Les chroniques, qui ont un taux de quantification supérieur à 50%, sont séparées de celles qui ont un taux inférieur à 50%.

La sélection de l'individu le plus représentatif est réalisée à partir du nombre d'analyses et de la moyenne des écarts entre deux analyses. L'individu se rapprochant le plus des caractéristiques moyennes de sa classe sur ces deux caractéristiques a été sélectionné.

4.1.3. Résultats des classifications

a) Atrazine

- **Classification automatique**

Sur les 1824 points de prélèvement affiliés au réseau de Contrôle de surveillance de l'état chimique des eaux, seuls 1809 ont au moins une donnée de concentration pour l'atrazine.

La Classification Ascendante Hiérarchique a été réalisée sans limitation durant la phase de paramétrage. Cela a abouti à la création de 17 classes. La variabilité inter-classes est de 68% et la variabilité intra-classe de seulement 32%. Cette classification est donc relativement peu performante.

La première classe contient 1450 individus, elle est caractérisée par des individus ayant des chroniques ayant en moyenne 31 analyses. La deuxième classe rassemble 255 individus. Les autres classes contiennent moins de 50 individus. Sept classes ne contiennent qu'un individu.

Comme expliqué précédemment, cette approche de classification automatique n'est pas utile pour cette étude. La classe 1 contient beaucoup trop d'individus dont des caractéristiques importantes pour les choix de la méthode statistique (normalité, autocorrélation) sont mélangées. La plupart des autres classes ne sont pas représentatives de l'ensemble des chroniques de RCS, ainsi, sept classes ne représentent qu'une seule chronique. La classification manuelle à partir des nœuds de l'arbre de décision est un moyen plus sûr de déterminer des chroniques représentatives au regard des méthodes statistiques à employer.

- **Classification manuelle**

Différents groupes sont créés à partir des caractéristiques présentées ci-dessus (paragraphe 4.1.20). L'arbre de décision présentant les différents groupes ainsi que le nombre de chroniques par groupe obtenu est présenté sur l'*Illustration 28*.

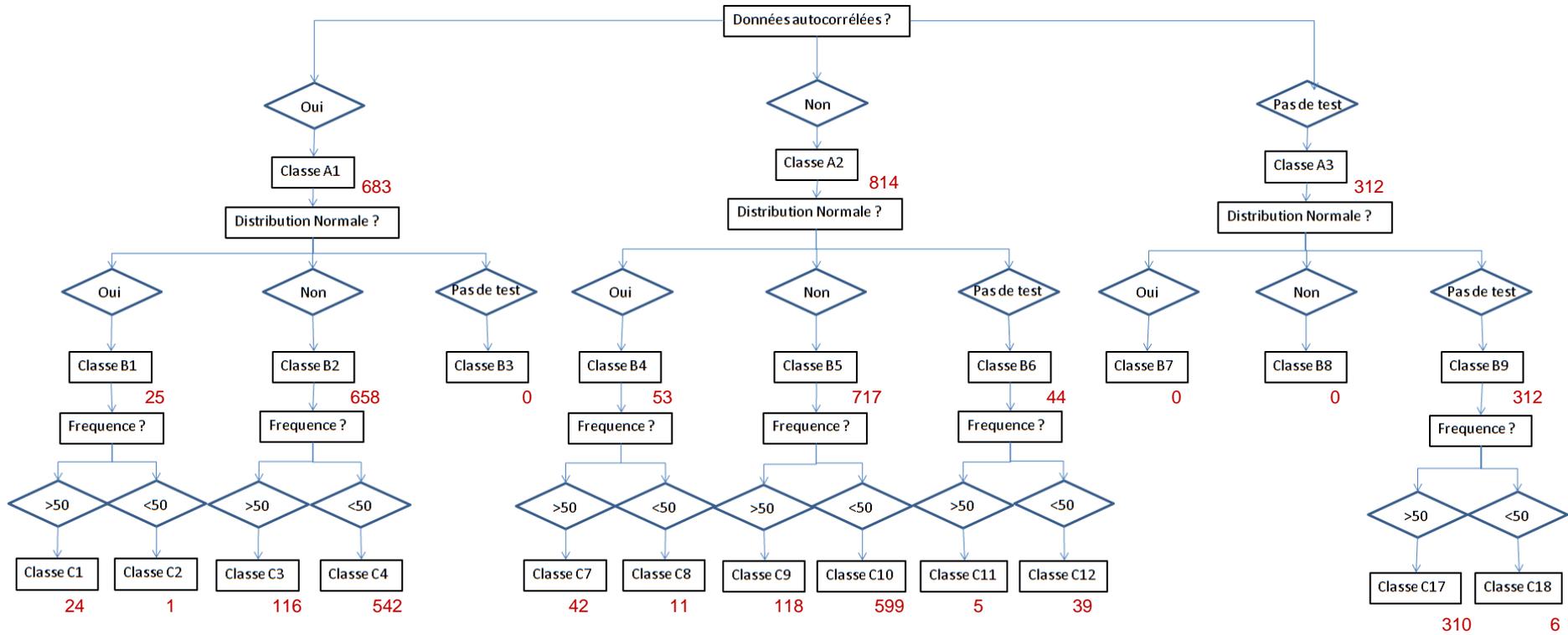


Illustration 28 : Labels des classes et nombre de chroniques par classe (en rouge) pour les concentrations en atrazine dans la classification manuelle

L'illustration 28 montre que le nombre de chroniques par groupe est très variable. Certains groupes sont peu représentés. Par exemple, par construction, la classe C2 ne pourra jamais contenir beaucoup d'individus. En effet, pour que les calculs de normalité et d'autocorrélation donnent des résultats positifs tous les deux, il faut que les séries temporelles présentent certaines caractéristiques rarement réunies lorsque le taux de quantification est inférieur à 50%.

Le choix des individus représentatifs de chaque groupe, sur lequel est testé l'outil s'est fait à partir du nombre d'analyses et du nombre de jours d'écart moyen entre deux analyses, qui devaient être proches des valeurs moyennes du groupe. Pour certains groupes avec peu d'individus (C17, C11, C12, C8 et C2) il n'a pas été choisi d'individu représentatif.

L'individu qui représente le mieux la classe C18 est le point de prélèvement 10142X0022/S. Son nombre moyen d'analyse est de 12 et sa moyenne des écarts entre deux analyses est de 303 jours

L'individu qui représente le mieux la classe C10 est le point de prélèvement 09043X0001/F. Son nombre moyen d'analyses est de 25 et sa moyenne des écarts entre deux analyses est de 184 jours.

L'individu qui représente le mieux la classe C11 est le point 00675X0141/F.AEP. Sa moyenne des écarts entre deux analyses est de 185 jours et son nombre d'analyse est de 30 – pour la classe, la moyenne des écarts entre deux analyses est de 187 jours et le nombre moyen d'analyse est 30,4.

L'individu qui représente le mieux la classe C7 est le point 02522X0003/S1. Il dispose de 20 analyses et sa moyenne des écarts entre deux analyses est de 97 jours - pour la classe, la moyenne des écarts entre deux analyses est de 206 jours et le nombre moyen d'analyse est 19,8.

L'individu qui représente le mieux la classe C4 est le point 02393X0039/HY. Il dispose de 33 analyses et sa moyenne des écarts entre deux analyses est de 123 jours - pour la classe, la moyenne des écarts entre deux analyses est de 157 jours et le nombre moyen d'analyse est 33,6.

L'individu qui représente le mieux la classe C3 est le point 05005X0042/S-FOND. Il dispose de 59 analyses et sa moyenne des écarts entre deux analyses est de 131 jours - pour la classe, la moyenne des écarts entre deux analyses est de 124 jours et le nombre moyen d'analyse est 58,4.

L'individu qui représente le mieux la classe C1 est le point 01823X0012/HY. Il dispose de 33 analyses et sa moyenne des écarts entre deux analyses est de 133 jours - pour la classe, la moyenne des écarts entre deux analyses est de 136 jours et le nombre moyen d'analyse est 32.

L'amplitude du nombre d'analyses pour les individus représentatifs est assez peu étendue ; entre 12 et 59. Cela provient du fait que tous les sites du réseau RCS ont environ le même nombre d'analyses (*Illustration 29*). De manière générale, il y a peu de disparité dans le nombre de mesures dans les sites du réseau RCS. Ce réseau a été mis en place pour assurer un suivi régulier sur une longue période. Cette particularité du réseau peut expliquer les résultats obtenus.

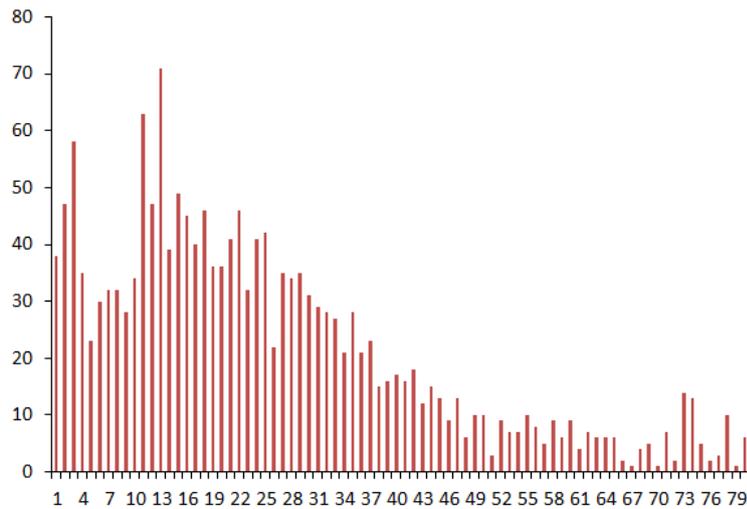


Illustration 29 : Nombre de prélèvements par site pour le paramètre atrazine

Si les sites sont classés par catégorie de taux de détection, de grande disparité apparaissent (Illustration 30). L'analyse du nombre d'individus par classe montre qu'il y a peu d'individus dans les catégories concernant les taux de quantification supérieur à 50%. Comme tous les produits phytosanitaires, la fréquence de détection de l'atrazine est assez faible, c'est-à-dire que sur le nombre d'analyses réalisées, le nombre d'analyses pour lesquelles la molécule est réellement détectées (c'est-à-dire ayant un code remarque égal à 1 dans ADES) est relativement faible. L'illustration 30 montre que dans le réseau FR_SOP plus de 1400 points de prélèvement ont une fréquence de détection comprise entre 0 et 10%. A l'inverse, seuls 8 points de prélèvement détectent l'atrazine à chaque recherche.

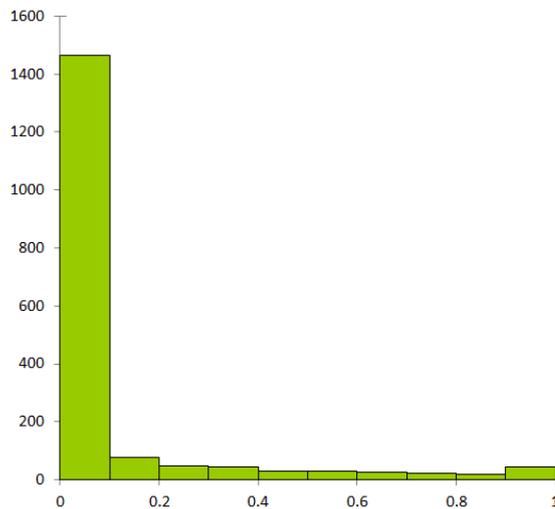


Illustration 30 : Nombre de sites par taux de détection pour le paramètre atrazine

Nitrates

- Groupe automatique

Après l'atrazine, des chroniques représentatives pour les nitrates ont été recherchées. Sur les 1824 points de prélèvement affiliés au réseau de Contrôle de surveillance de l'état chimique des eaux, tous ont au moins une donnée de concentration pour les nitrates

Une première CAH a été réalisée sans limitation lors du paramétrage (pas de troncature sur le nombre de catégories, par exemple). Cela a mené à un très grand nombre de classes (222) difficilement interprétable. De plus, la variabilité inter-classes était de 95%.

Une seconde CAH a été réalisée en limitant le nombre de classe à 5. Le pourcentage de variabilité intra-classe était de 54% et la variabilité inter-classes de 46%. Néanmoins, l'interprétation et la recherche de chroniques représentatives n'étaient pas possibles. La première classe rassemblait 1817 individus et les quatre autres classes rassemblaient moins de 3 individus. L'approche manuelle a donc été privilégiée.

- **Groupe manuel**

Différents groupes sont créés à partir des caractéristiques présentées ci-dessus (paragraphe 0 4.1.20). L'arbre de décision présentant les différents groupes ainsi que le nombre de chroniques par groupe obtenu est présenté sur l'illustration 31.

Le choix des individus représentatifs de chaque groupe, sur lequel est testé l'outil s'est fait à partir du nombre d'analyses et du nombre de jours d'écart moyen entre deux analyses, qui devaient être proche des valeurs moyennes du groupe. Pour certains groupes avec peu d'individus (C17, C18, C11, C12) il n'a pas été choisi d'individu représentatif.

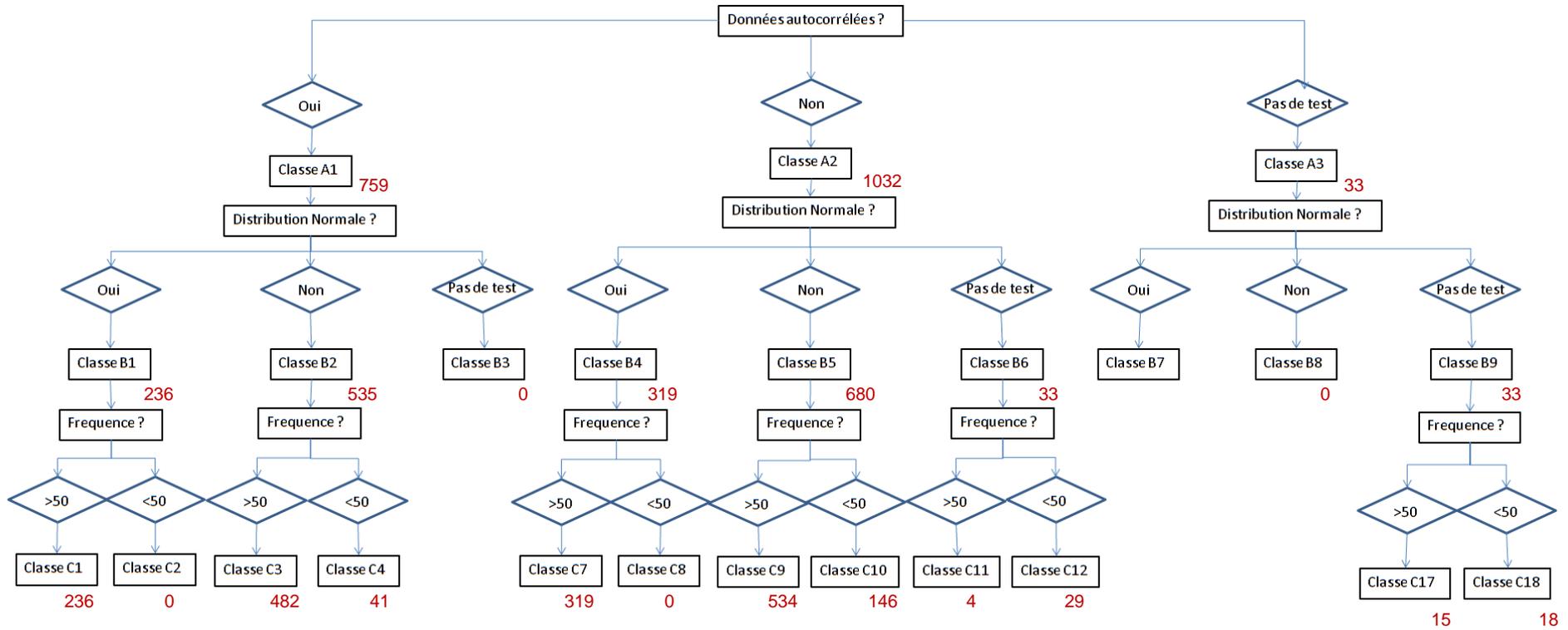


Illustration 31 : Labels des classes et nombre de chroniques par classe (en rouge) pour les chroniques de concentration en nitrate dans la classification manuelle

L'individu qui représente le mieux la classe C10 est le point 08762X0025/F2. Il dispose de 23 analyses et sa moyenne des écarts entre deux analyses est de 255 jours - pour la classe, la moyenne des écarts entre deux analyses est de 256 jours et le nombre moyen d'analyse est 23,7.

L'individu qui représente le mieux la classe C9 est le point 09574X0026/F. Il dispose de 34 analyses et sa moyenne des écarts entre deux analyses est de 168 jours - pour la classe, la moyenne des écarts entre deux analyses est de 168 jours et le nombre moyen d'analyse est 36,4.

L'individu qui représente le mieux la classe C7 est le point 02598X0009/P1. Il dispose de 22 analyses et sa moyenne des écarts entre deux analyses est de 236 jours - pour la classe, la moyenne des écarts entre deux analyses est de 211 jours et le nombre moyen d'analyse est 21,8.

L'individu qui représente le mieux la classe C4 est le point 01546X0065/F. Il dispose de 23 analyses et sa moyenne des écarts entre deux analyses est de 242 jours - pour la classe, la moyenne des écarts entre deux analyses est de 258 jours et le nombre moyen d'analyse est 25,1.

L'individu qui représente le mieux la classe C3 est le point 03538X0005/P. Il dispose de 69 analyses et sa moyenne des écarts entre deux analyses est de 131 jours - pour la classe, la moyenne des écarts entre deux analyses est de 135 jours et le nombre moyen d'analyse est 68,2.

L'individu qui représente le mieux la classe C1 est le point 02097X0010/C1. Il dispose de 42 analyses et sa moyenne des écarts entre deux analyses est de 141 jours - pour la classe, la moyenne des écarts entre deux analyses est de 167 jours et le nombre moyen d'analyse est 43,2.

L'amplitude du nombre d'analyses pour les individus représentatifs est assez peu étendue (23 à 69 analyses). Cela provient du fait que, comme pour l'atrazine, tous les sites du réseau RCS ont environ le même nombre d'analyses. Le fait que tous les sites aient approximativement le même nombre d'analyses est lié aux dates de début de suivi des nitrates qui sont similaires dans la plupart des sites du réseau RCS (*Illustration 32*). Il existe un certain nombre de points ayant des périodes entre 30 et 50 ans, néanmoins la majorité des durées de suivi est entre 10 et 20 ans.

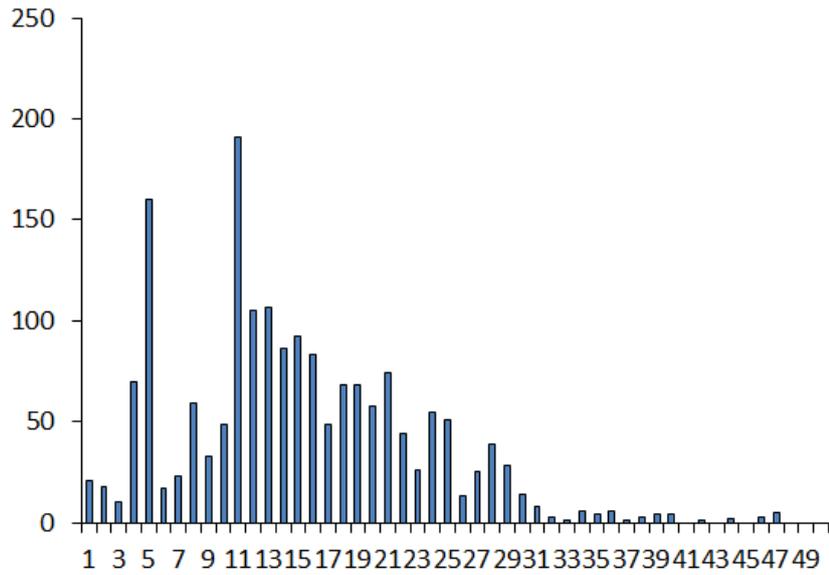


Illustration 32 : Nombre de sites en fonction du nombre d'années de suivi pour le paramètre atrazine

L'analyse du nombre d'individus par classe de taux de quantification montre qu'il y a peu d'individus dans les catégories concernant les taux de quantification inférieurs à 50%. L'analyse du taux de détection des nitrates montre un profil complètement différent de celui de l'atrazine. En effet, pour les nitrates, la fréquence de détection est pratiquement toujours égale à 1. Sur les 1824 points du réseau de surveillance, 1327 détectent les nitrates à chaque analyse.

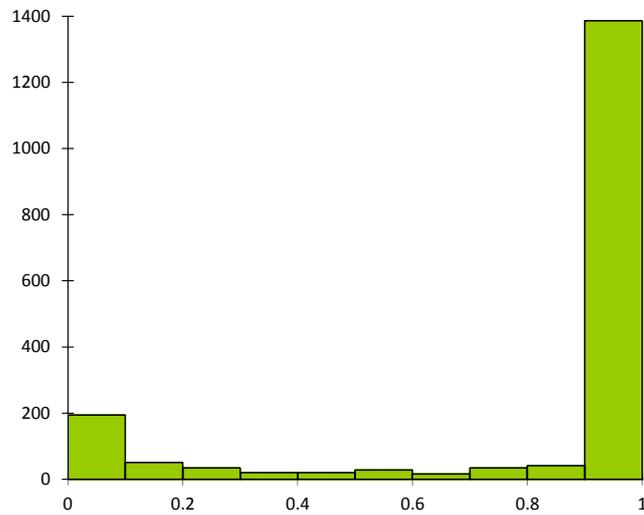


Illustration 33 : Nombre de sites par taux de détection pour le paramètre atrazine

4.2. APPLICATION DE L'OUTIL

4.2.1. Contexte et méthode

Le module principal de l'outil – recherche de tendances et de rupture - a été appliqué sur les individus centraux des classes déterminées manuellement par l'approche présentée précédemment.

L'outil a été appliqué aux individus centraux de chaque molécule. La description suivante présente le test de la méthode, pour chaque paramètre ADES identifié et pour chaque individu central déterminé.

4.2.2. Application

a) *Atrazine*

- **01821X0012/HY**

La première chronique analysée est la chronique de classe C1. Elle représente la classe des chroniques aux données autocorrélées, avec une distribution normale et dont le taux de quantification est supérieur à 50%. Les indications sur la normalité et sur l'autocorrélation sont indiquées dans la rubrique "Caractéristique de la chronique" (*Illustration 34*). Cette chronique présente 33 analyses et sa moyenne des écarts entre deux analyses est de 133 jours.

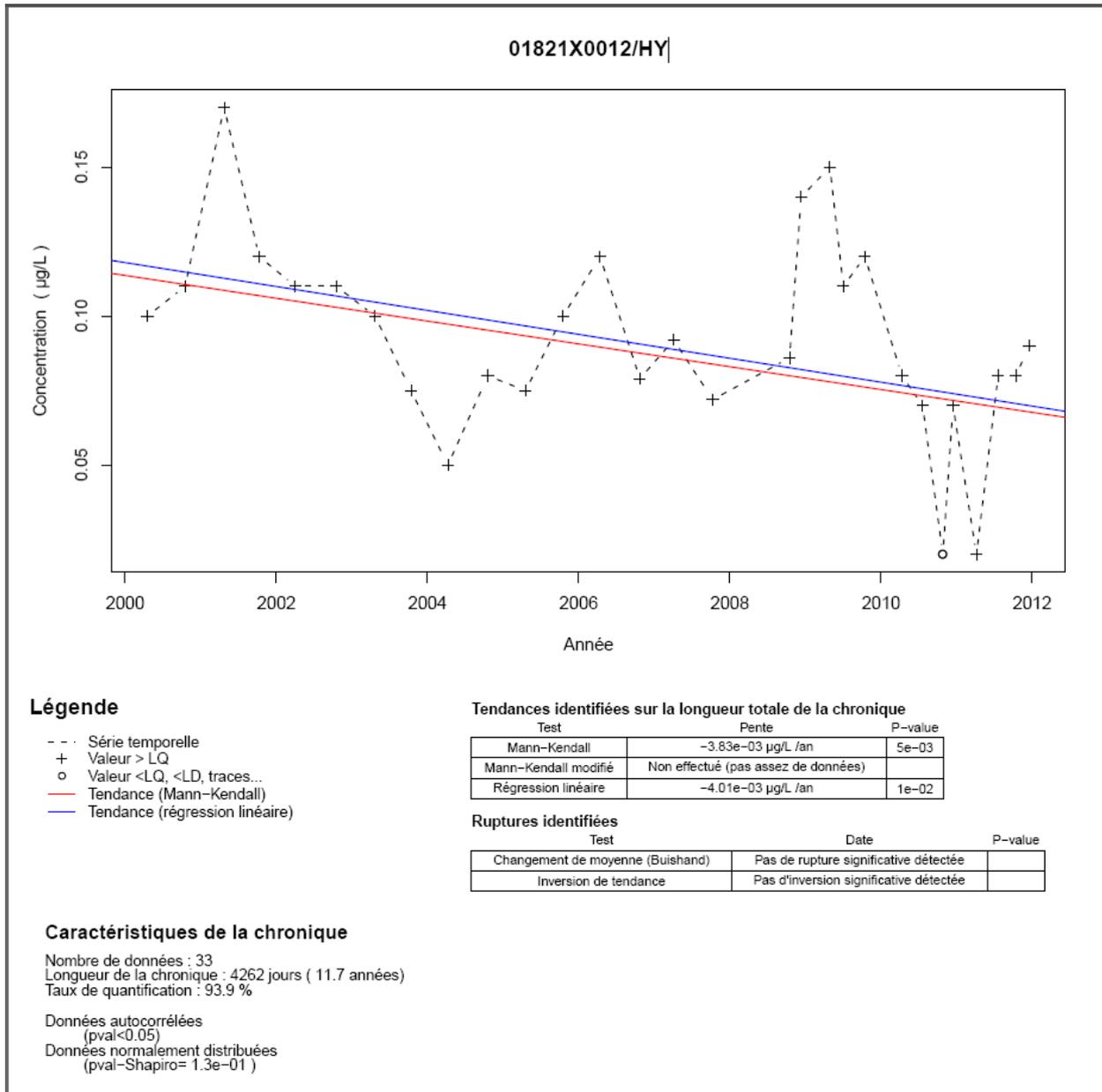


Illustration 34 : Copie d'écran de l'outil pour l'analyse de la chronique d'atrazine pour le point 01821X0012/HY (Classe C1)

Cette chronique est autocorrélée, a une distribution normale et son nombre d'analyses est inférieur à 50. Les tests appliqués sont la régression linéaire ou le test de Mann-Kendall pour la recherche de tendance et le test de Buishand et celui de Darken pour la recherche de rupture ainsi que proposé dans le rapport.

En ce qui concerne la rupture, l'outil n'en détecte pas selon le test de Darken. Le texte proposé est "Pas d'inversion significative détectée" (Illustration 34). De plus, il ne détecte pas de rupture avec le test de Buishand. Ainsi, il propose le texte suivant " *Changement de moyenne (Buishand) | Pas de rupture significative détectée*". Puisqu'aucune rupture n'est identifiée, elles ne peuvent pas être tracées sur le graphique (Illustration 34).

Pour la recherche de tendances, l'outil applique une régression linéaire et un test de Mann-Kendall modifié. Les deux méthodes fournissent la même tendance globale d'évolution, à savoir une diminution sur toute la période. L'outil fournit le texte suivant pour la méthode Mann-Kendall "Mann-Kendall | $-3.83e-03 \mu\text{g/L/an}$." et pour la régression linéaire le texte "*Régression linéaire* | $-4.01e-03 \mu\text{g/L/an}$ ". Le signe de la pente permet de connaître si la tendance est orientée à la hausse ou à la baisse. Un signe négatif de la pente indique une baisse de la tendance. Les tendances étant significatives, elles sont tracées sur le graphique. La régression est affichée en bleu, la tendance Mann-Kendall, en rouge (Illustration 34).

Ainsi, selon l'outil, les pentes des deux tendances proposées sont significatives. Une diminution de la concentration en produits phytosanitaires est observable, selon les deux méthodes, malgré les cycles qui peuvent perturber la lecture de la chronique. En effet, la première valeur mesurée en 2000 ($0,1 \mu\text{g/l}$) est peu différente de la dernière en 2012 ($0,09 \mu\text{g/l}$). C'est l'ensemble de toutes les valeurs mesurées qui permettent de tracer la tendance.

- **05005X0042/S-FOND**

Cette chronique représente la classe C3. C'est la classe des chroniques dont les données sont autocorrélées, ont une distribution non normale et dont les taux de quantification sont supérieurs à 50%. Les données sur la normalité et l'autocorrélation sont rappelées dans l'encart "Caractéristiques de la chroniques" en bas à gauche du graphique (Illustration 35). Le nombre d'analyses de cette chronique est de 59 et sa moyenne des écarts entre deux analyses est de 131 jours. Cette chronique est un bon reflet de certaines chroniques du réseau RCS du point de vue de la fréquence d'analyse. En effet, bien que commençant tôt dans les années 1990, l'effort d'analyses n'est réel qu'à partir des années 2000.

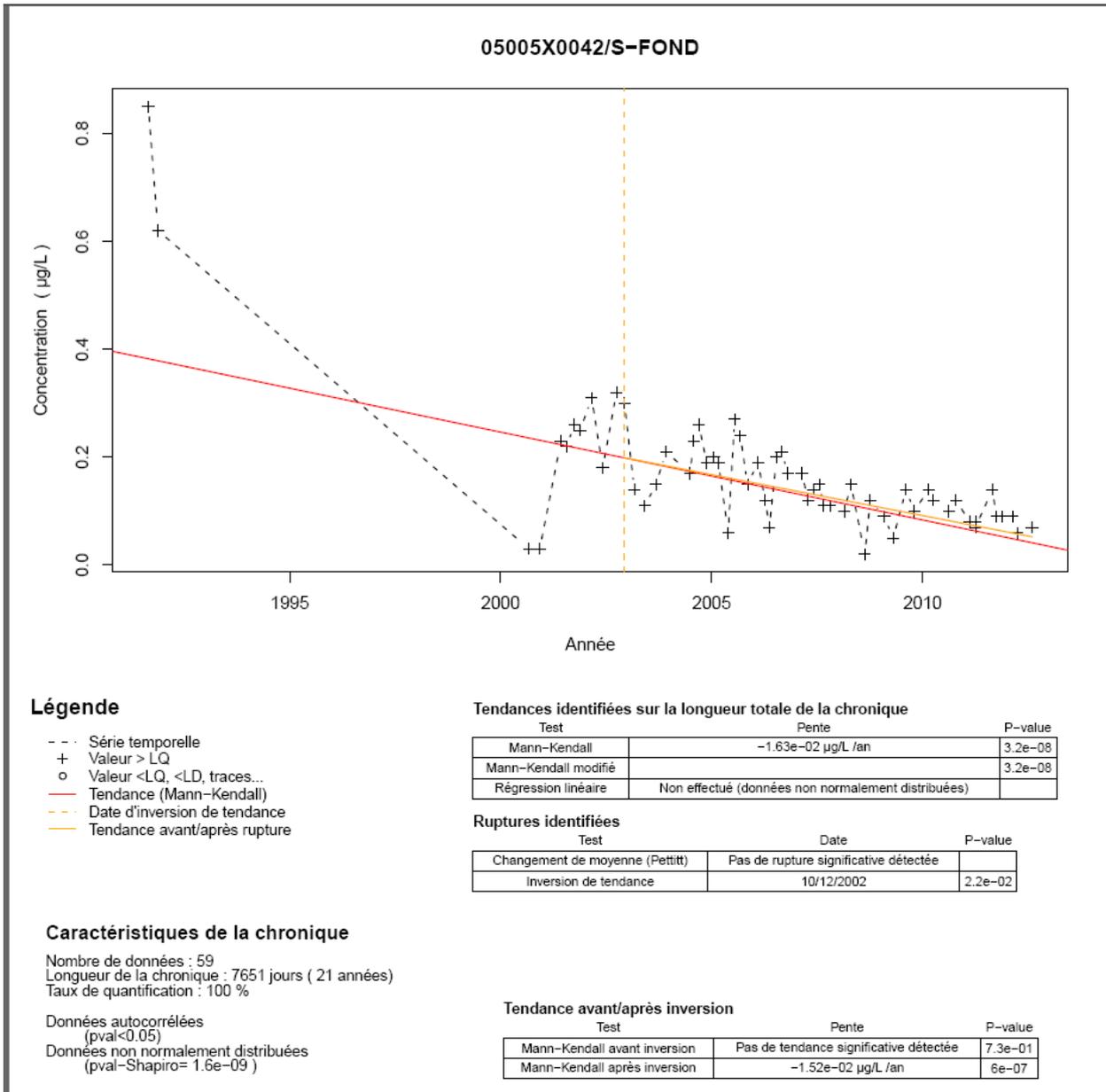


Illustration 35 : Copie d'écran de l'outil pour l'analyse de la chronique d'atrazine pour le point 05005X0042/S-FOND (Classe C3)

Cette chronique est autocorrélée et a une distribution qui n'est pas normale. Le test de Mann-Kendall modifié doit être privilégié pour la recherche de la tendance. Pour la recherche de rupture les tests de Pettitt et de Darken sont utilisés.

Selon le test de Pettitt, aucune rupture n'est détectée. Ainsi, l'outil affiche le texte suivant "Changement de moyenne (Pettitt) | Pas de rupture significative détectée". De plus, aucune courbe ne s'affiche pour la rupture sur le graphique (Illustration 35). Selon le test de Darken, une inversion est détectée et le texte suivante est proposé "Inversion de tendance | Date = 10/12/200 | 2.2e-02". Un seul segment est détecté, il s'affiche, en jaune, à la baisse en 2002 et 2011 sur le graphique (Illustration 35). La date de rupture (10/12/2002) est représentée sur le graphique par un trait horizontal jaune pointillé (Illustration 35).

Pour la détection de tendance, une seule méthode est proposée par l'outil, la méthode Mann-Kendall. La distribution n'étant pas normale, il n'est pas recommandé de réaliser de régression linéaire. La courbe ne s'affiche donc pas pour la régression linéaire. Selon l'outil, une pente existe et une tendance à la diminution est détectée. Le texte suivant est fourni " *Mann-Kendall | -1.63e-02 µg/L/an*" (Illustration 35). La courbe correspondante s'affiche, en rouge, sur le graphique (Illustration 35).

Bien qu'il y ait une grande différence entre l'effort d'analyses avant et après 2000, l'outil est capable de prendre en compte la différence d'individus. Cette différence entre le nombre d'analyses par an, n'influe pas non plus sur le calcul de rupture, ainsi l'outil ne détecte pas de rupture juste parce que l'effort d'analyses change. Il est à noter que le calcul de la tendance sur toute la période s'accorde sur la tendance à la baisse dans la dernière partie de la chronique (2002-2012).

- **02393X0030/HY**

La chronique suivante est celle représentant la classe C4. Cette classe représente les chroniques dont des données sont autocorrélées, ont une distribution non normale et dont les taux de quantification sont inférieurs à 50% (*Illustration 36*). Cette chronique a un nombre d'analyses égal à 33 et sa moyenne des écarts entre deux analyses est de 123 jours. Cette chronique représente bien les points de prélèvement dont le taux est inférieur à 50%. En effet, cela signifie souvent qu'ils contiennent de nombreuses données inférieures ou égales à la LQ.

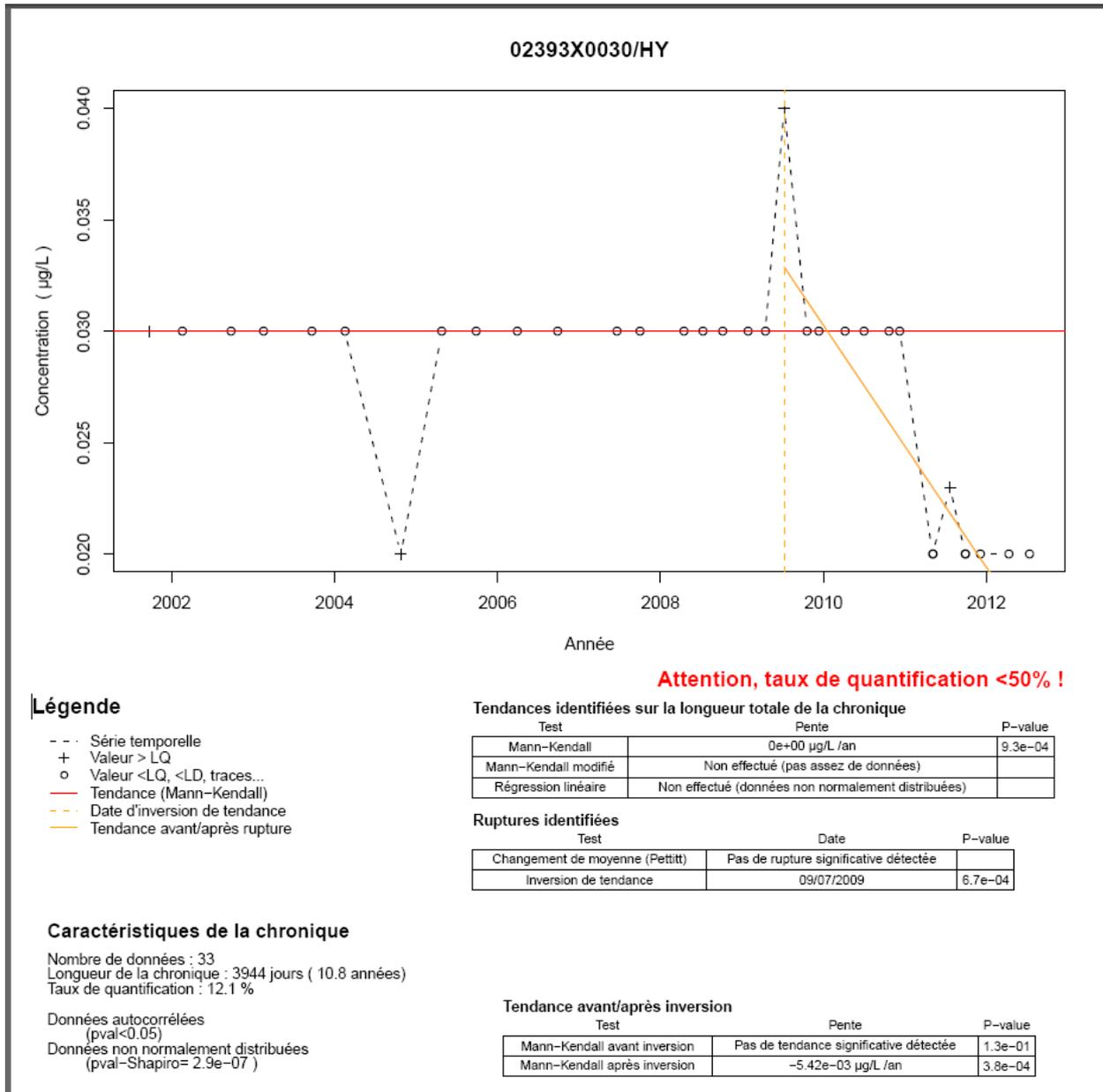


Illustration 36 : Copie d'écran de l'outil pour l'analyse de la chronique d'atrazine pour le point 02393X0030/HY (Classe C4)

Les résultats donnés par l'outil ont ici peu d'importance. En effet, cette chronique montre les limites de la méthode proposée. Une tendance à la baisse a été tracée par l'outil sur la dernière partie la chronique (2009-2012) uniquement à cause d'un changement de méthode analytique qui a fait passer la LQ pour l'atrazine de 0,03 à 0,02. Au final, cette chronique ne présente que 4 données dont le code remarque est égal à 1. Leur valeur est 0,03 (en 2001), 0,04 (en 2010), 0,023 (en 2009) et 0,02 (en 2004). L'outil indique par la phrase "Attention, le taux de quantification <50%" qu'un nombre important de données inférieures à la LQ ont été utilisées. Si l'utilisateur se sert de la chronique, il doit avoir conscience du type de données qui ont servi au calcul. Ici, il y a 33 données et un taux de quantification de 12%. Cela signifie que 4 valeurs seulement sont supérieures au taux de quantification. Utiliser 4 valeurs n'est pas suffisant pour donner une tendance. Une suite de valeurs proches ou égales à la LQ est une source de

difficulté dans le calcul de la tendance. Les variations de valeurs de la LQ en fonction des laboratoires et du temps rendent encore le calcul de tendance et de rupture souvent complexe.

• **02522X0003/S1**

La chronique suivante est celle représentant la classe C7. Cette classe représente les chroniques qui ont des données non autocorrélées, normales et qui ont des taux de quantification supérieurs à 50% (*Illustration 37*). Cette classe est assez peu représentative de l'ensemble du réseau RCS, elle a un taux de quantification de 90%, ce qui est rare pour une chronique de produits phytosanitaires. Dans cette chronique, seuls deux points n'ont pas un code remarque égal à 1. Cette chronique intègre un nombre d'analyses égal à 20 et sa moyenne des écarts entre deux analyses est de 97 jours.

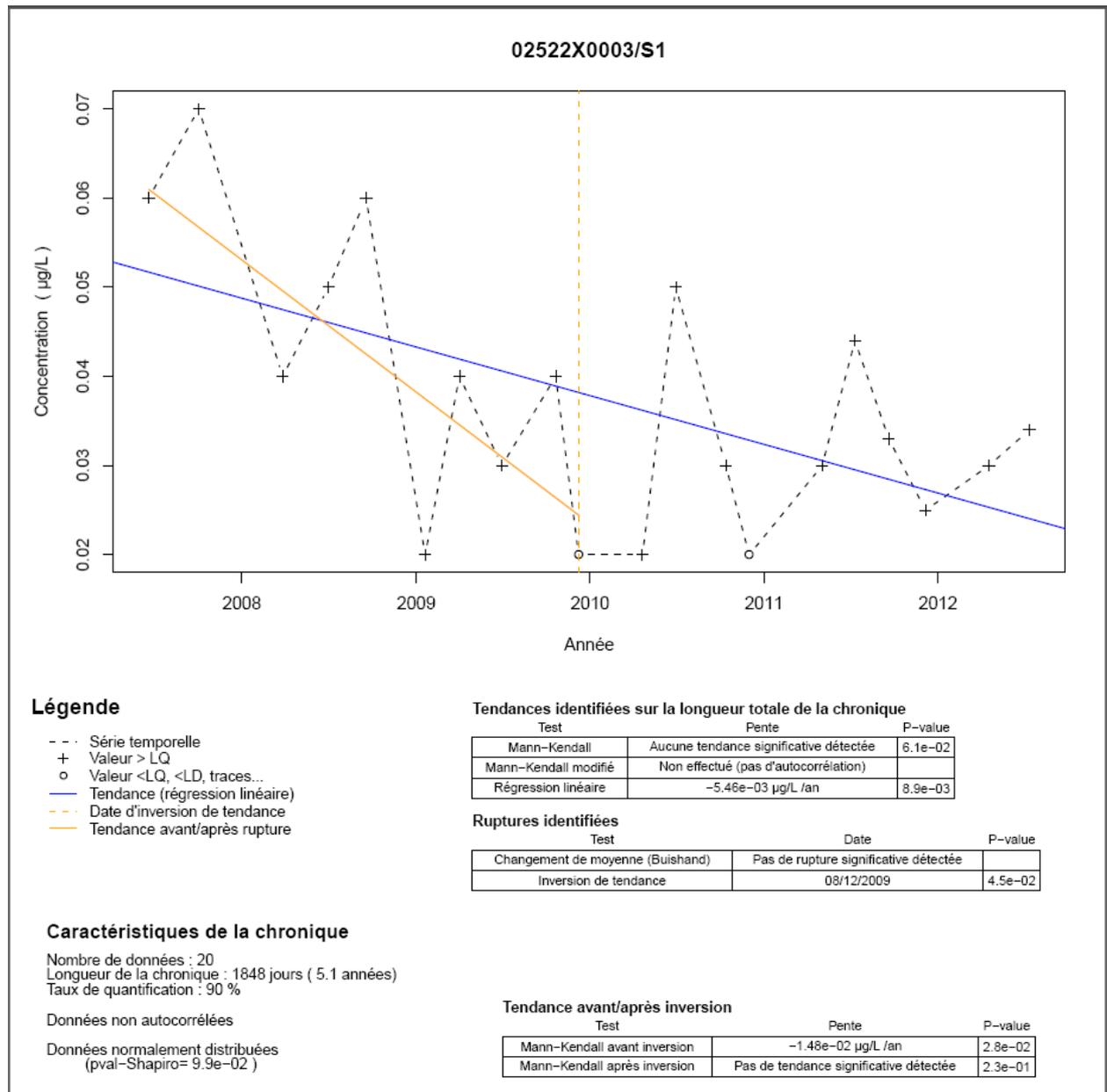


Illustration 37 : Copie d'écran de l'outil pour l'analyse de la chronique d'atrazine pour le point 02522X0003/S1 (Classe C7)

Cette chronique possède des données non autocorrélées qui ont une distribution normale. En considérant ces caractéristiques, des tests de rupture de pente de Darken et de Buishand doivent être envisagés. De plus, une recherche de tendance par régression linéaire et par un test de Mann-Kendall doit être tentée.

Il n'y a pas de rupture selon le test de Buishand. Ainsi l'outil propose le texte suivant "*Changement de moyenne (Buishand) | Pas de rupture significative détectée*") dans la rubrique Rupture à droite (*Illustration 37*). Selon le test de Darken, un seul segment est tracé. Une tendance à la baisse est observée entre 2007 et 2009. Il n'y a pas de segment tracé après 2009. Pour la recherche de tendance, l'outil réalise une régression linéaire en plus du test de Mann-Kendall. Le test selon la méthode de Mann-Kendall n'est pas jugé significatif et donc n'est pas illustré sur le graphique (*Illustration 37*). Le texte proposé pour la méthode de Mann-Kendall est "*Aucune tendance significative n'est observée*". En revanche, l'outil détecte une tendance à la baisse lors de l'utilisation de la régression linéaire. Le texte proposé pour cette méthode est "*Régression linéaire | -5.48e-03 µg/L/an | 8.9e-03*" (*Illustration 37*).

Les différences entre les méthodes ne sont donc pas anodines. Ici, entre Mann-Kendall et la régression linéaire, une tendance peut ou ne peut pas être jugée significative.

- **00675X0141/F.AEP**

Ce point est représentatif des chroniques non autocorrélées, non normales et dont le taux de quantification est supérieur à 50% (*Illustration 38*). Ces chroniques sont contenues dans la classe C9. La chronique du point 00675X0141/F.AEP a une moyenne des écarts entre deux analyses valant 185 jours et un nombre d'analyses de 30. Son taux de quantification est de 90% ce qui reste élevé. Comme expliqué précédemment, seuls 20% des points du réseau RCS ont un taux de quantification supérieur à 10%.

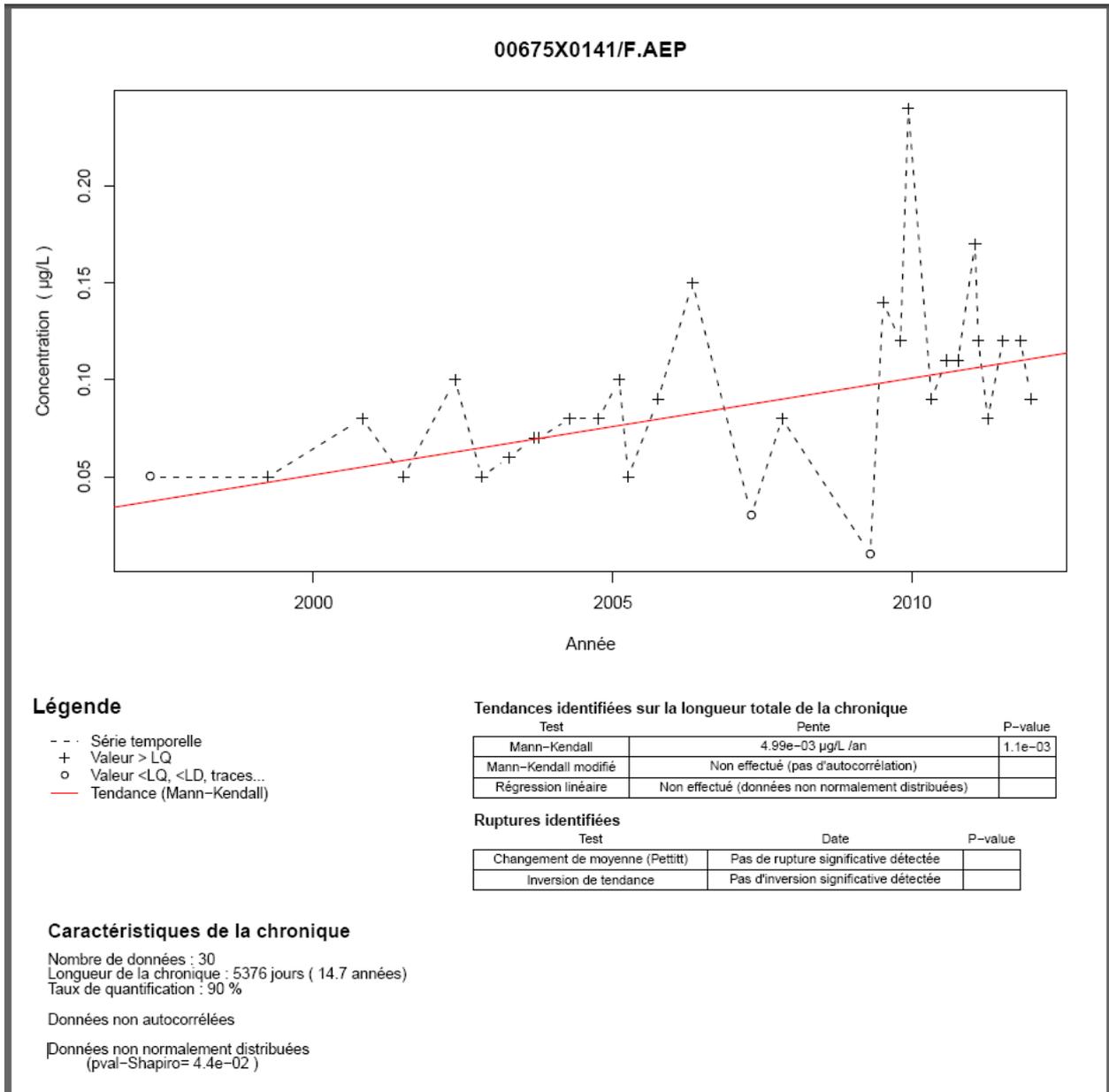


Illustration 38 : Copie d'écran de l'outil pour l'analyse de la chronique d'atrazine pour le point 00675X0141/F.AEP (Classe C9)

Cette chronique possède des données non autocorrélées et qui n'ont pas de distribution normale. Un test de rupture de pente de Darken ou de Pettitt doit être considéré. De plus, une recherche de tendance par un test de Mann-Kendall doit être envisagée.

Selon les données de cette chronique, il n'y a pas de rupture. Ainsi l'outil propose le texte suivant "Pas de rupture significative détectée". De plus il n'y a pas d'inversion. Le texte suivant s'affiche "Pas d'inversion significative détectée" (Illustration 38). Le test de Mann-Kendall a été utilisé pour tester l'existence d'une tendance. Selon l'outil, une tendance à la hausse est déterminée. Le texte suivant est fourni par l'outil "Mann-Kendall | 4.99e-03 µg/L/an | 1.1e-03" (Illustration 38).

- **09043X0001/F**

Ce point est représentatif de la classe C10, c'est-à-dire de la classe des chroniques non autocorrélées, non normales et dont le taux de quantification est inférieur à 50% (Illustration 39). Cette classe est celle qui contient le plus d'individus (599 membres). Cette chronique est représentative des chroniques rencontrées dans le domaine des analyses de produits phytosanitaires. Les concentrations mesurées oscillent entre des valeurs supérieures à la LQ (code remarque 1), des valeurs en dessous de celle-ci (code remarque 2 ou 10). Son nombre moyen d'analyses est de 24, sa moyenne des écarts entre deux analyses est de 184 jours, son taux de quantification est de 32%.

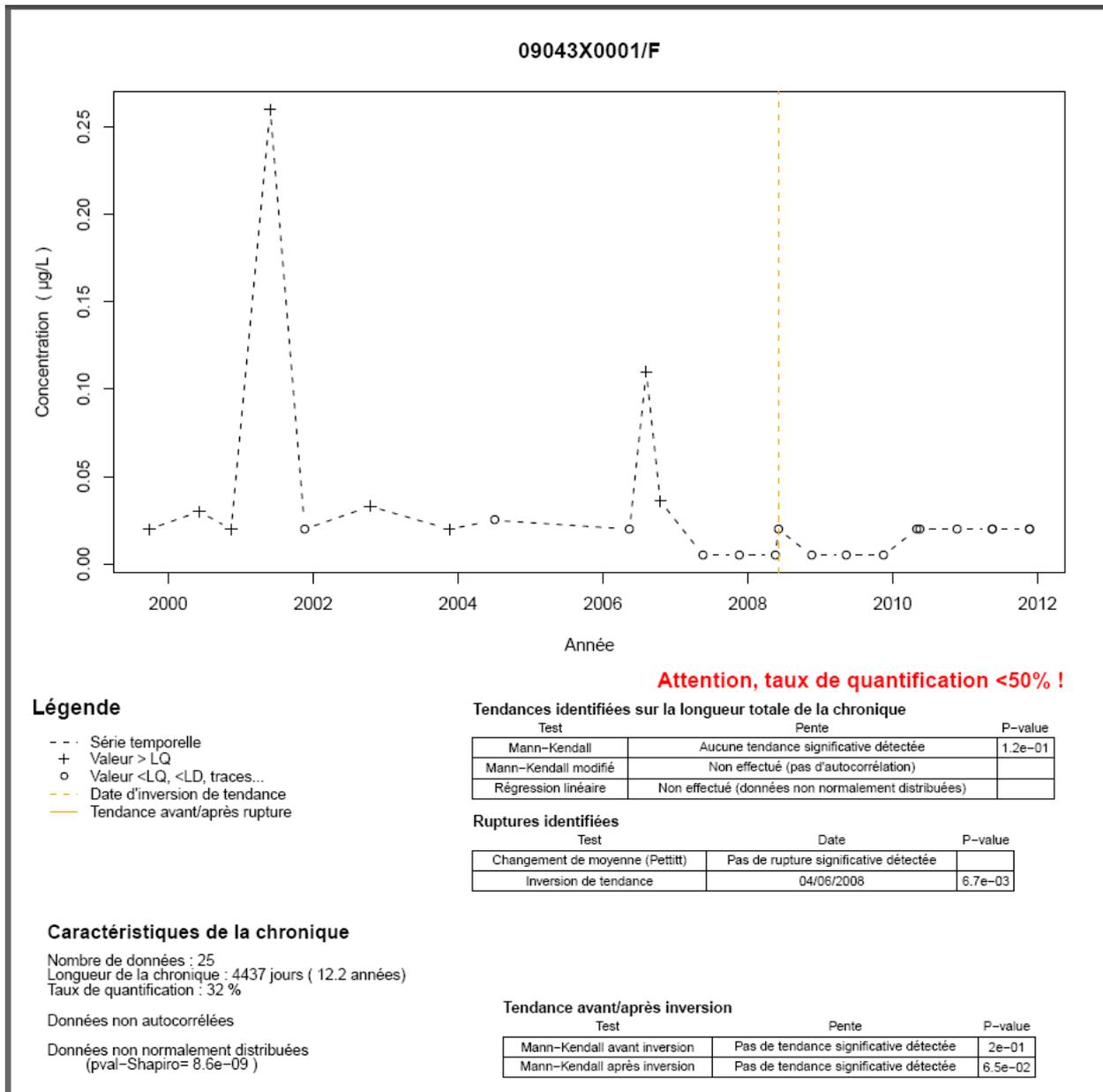


Illustration 39 : Copie d'écran de l'outil pour l'analyse de la chronique d'atrazine pour le point 09043X0001/F (Classe C10)

Cette chronique possède des données non autocorrélées et qui ont une distribution non normale. De ce fait, un test de rupture de pente de Darken ou de Pettitt doit être considéré. De plus, une recherche de tendance par un test de Mann-Kendall doit être envisagée.

Il semble que, selon le test de Pettitt données, il n'y ait pas de rupture. Ainsi l'outil propose le texte suivant "*Pas de rupture significative détectée*". Selon le test le Darken, une rupture est détectée en 2008. Le texte suivant est proposé "*Inversion de tendance | 04/06/2008 | pval 6.7e-03*". Mais ni le segment avant ni le segment après cette date ne sont significatif c'est pour cela qu'ils ne sont pas tracés. (*Illustration 39*). Le test de Mann-Kendall est réalisé par l'outil mais aucune tendance ne peut être détectée. L'outil propose la phrase suivante "*Aucune tendance significative n'est observée*". La distribution n'étant pas normale, la régression linéaire n'a pas été appliquée. L'outil rappelle cela en affichant le texte "*Régression linéaire | Non effectué (données non normalement distribuées)*" (*Illustration 39*).

Cet exemple indique qu'il est parfois impossible de déterminer une tendance. De plus, le fait que de nombreuses valeurs soient proches ou égales à la LQ perturbe les tests car cela revient à ajouter des séries de valeurs identiques. Le faible taux de quantification est rappeler par l'outil par la phase "Attention, taux de quantification <50%"

- **10142X0022/S**

Ce point est représentatif de la classe des chroniques pour lesquelles ni le test de normalité, ni le test d'autocorrélation n'ont pu être appliqués (*Illustration 40*). Ces chroniques sont rassemblées dans la classe C18. Comme expliqué ci-dessus, cela est surtout dû au fait que ces chroniques ont un nombre de données insuffisant. Cela concerne près de 310 individus. Le nombre d'analyses de cette chronique (10142X0022/S) est de 12 et sa moyenne des écarts entre deux analyses est de 303 jours. Elle est typique des sites pour lesquels le produit phytosanitaire est mesuré à la valeur de la LQ depuis la mise en place de la station de mesure.

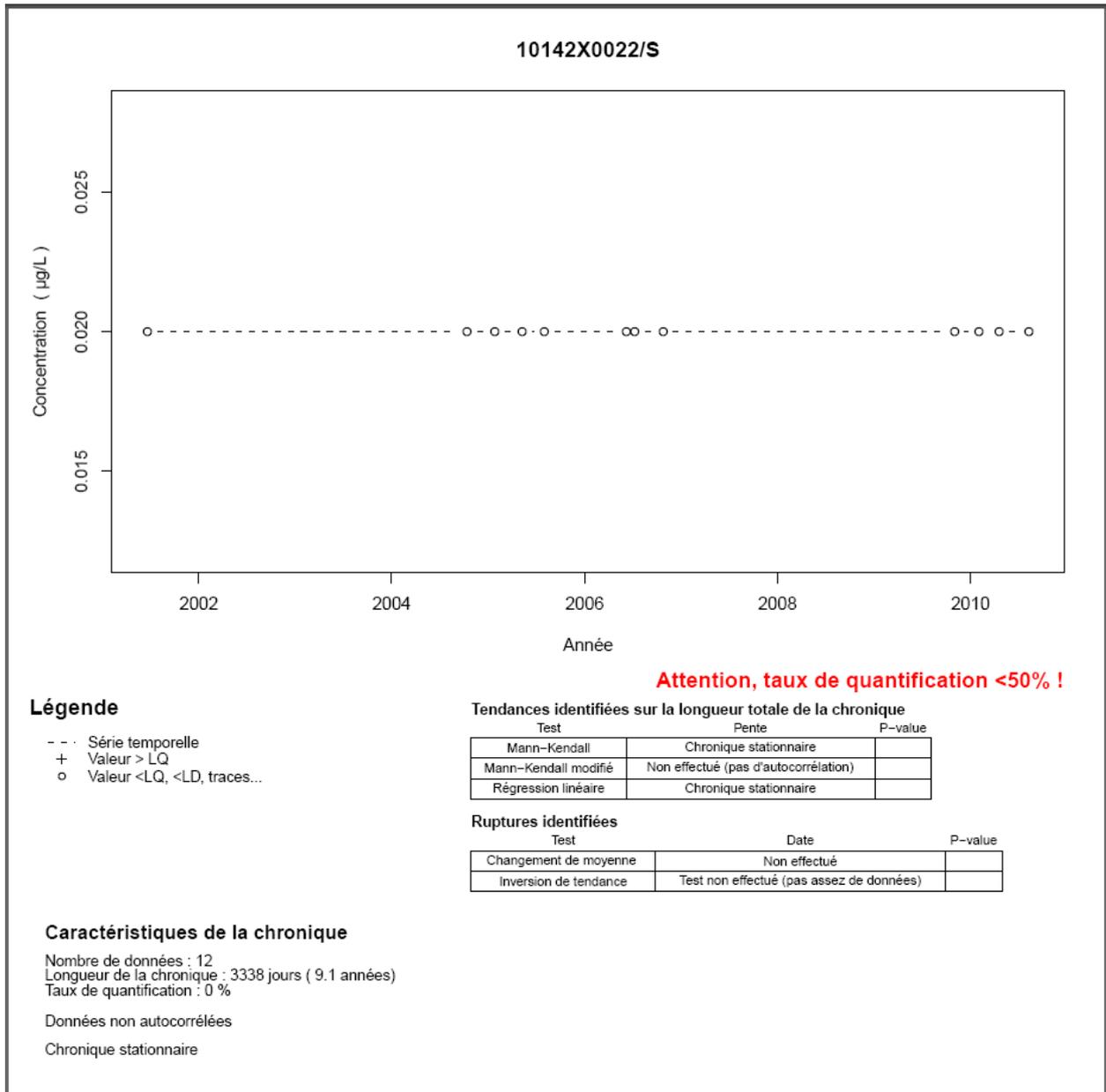


Illustration 40 : Copie d'écran de l'outil pour l'analyse de la chronique d'atrazine pour le point 10142X0022/S (Classe C18)

Aucun test de tendance ni de rupture n'est pertinent dans le cas d'une telle chronique. L'ensemble des données est inférieur à la LQ et toutes les valeurs sont identiques.

Nitrates

- **02097X0010/C1**

La première chronique analysée pour les nitrates est la chronique de classe C1 (Illustration 41). Elle représente la classe des chroniques autocorrélées, normales et dont le taux de quantification est supérieur à 50%. Le nombre d'analyses de la chronique du site

02097X0010/C1 est de 42 et sa moyenne des écarts entre deux analyses est de 141 jours. Son taux de quantification est de 100%, ce qui est classique pour un élément comme le nitrate.

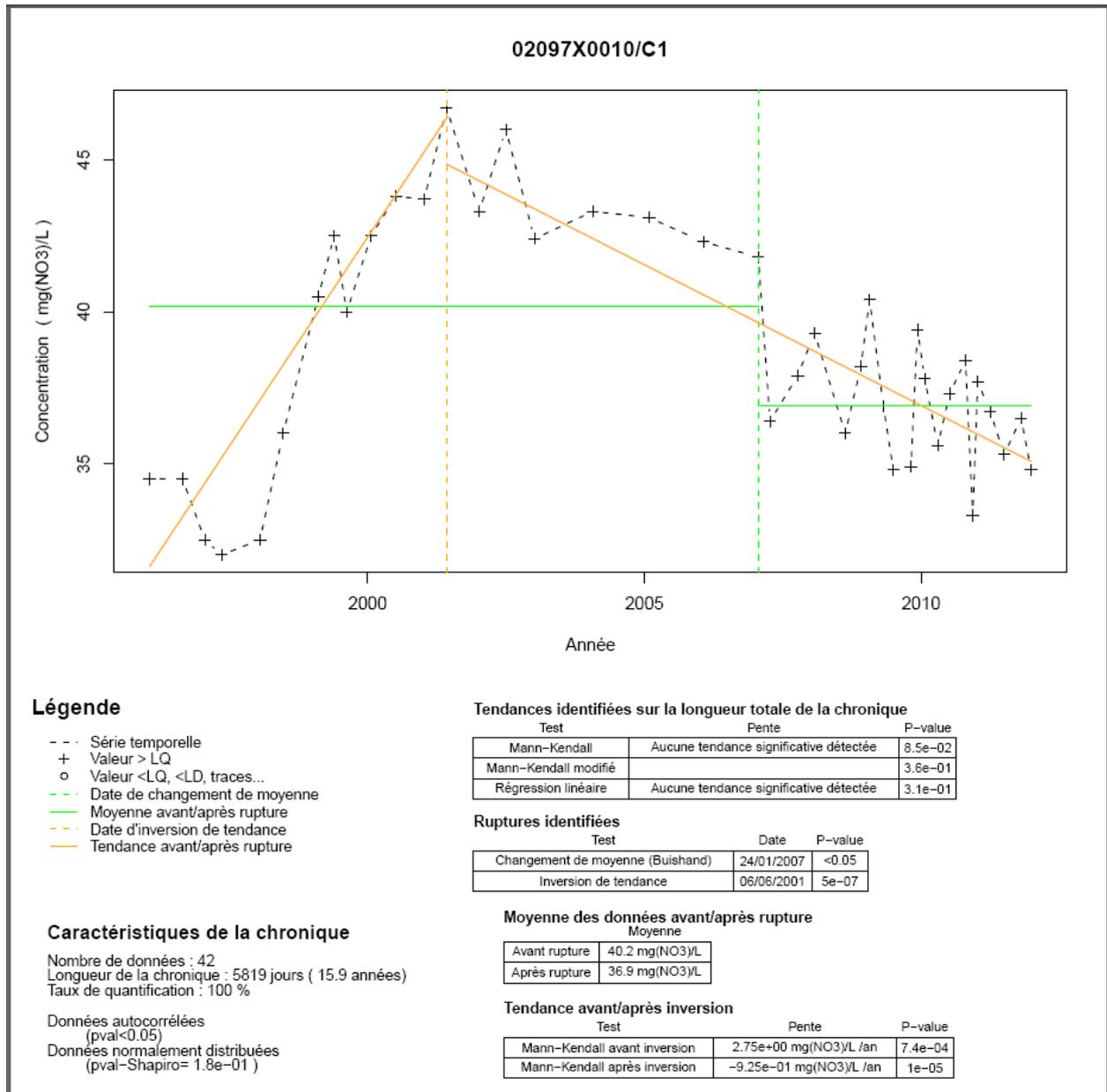


Illustration 41 : Copie d'écran de l'outil pour l'analyse de la chronique de nitrates pour le point 02097X0010/C1 (Classe C1)

Cette chronique est autocorrélée et a une distribution normale. En considérant cela, une recherche de rupture par le test de Buishand et de Darken doit être considérée. De plus, une recherche de tendance par régression linéaire et selon le test de Mann-Kendall modifié doit être envisagée.

La méthode de Buishand est appliquée pour rechercher un changement de moyenne dans la chronique et la méthode de Darken est appliquée pour rechercher une inversion de tendance. Selon la méthode de Buishand, une rupture est constatée. L'outil affiche le texte suivant "Changement de moyenne (Buishand) | 24/01/2007 | <0.05" Ici, une moyenne (40,2 mg/l) est

calculée de 1996 à 2007 puis une autre plus faible (36,9 mg/l) est calculée de 2007 à 2011. Ces informations sont disponibles dans l'encart "Moyenne des données avant/après rupture". Une diminution est constatée. Selon la méthode Darken, une inversion est constatée, 06-06-2001. Le texte suivant est proposé "*Inversion de tendance | 24/01/2007 | 5e-07*" Les concentrations augmentent de 1996 à 2001 puis diminuent de 2001 à 2011 (*Illustration 41*). Les données sur les pentes avant et après rupture sont disponible dans l'encart "Tendance avant/après rupture". Pour la recherche de tendances, l'outil essaie une régression linéaire en plus du test de Mann-Kendall modifié. Pour ces deux méthodes, aucune tendance significative n'est détectée comme l'indique les informations données par l'outil (*Illustration 41*).

- **03538X0005/P**

Cette chronique représente la classe C3, c'est la classe des chroniques dont les données sont autocorrélées, non normales et dont les taux de quantification sont supérieurs à 50% (*Illustration 42*). Cette chronique a un nombre d'analyses égal à 69 analyses et une moyenne des écarts entre deux analyses égale à 131 jours. Son taux de quantification est de 100%.

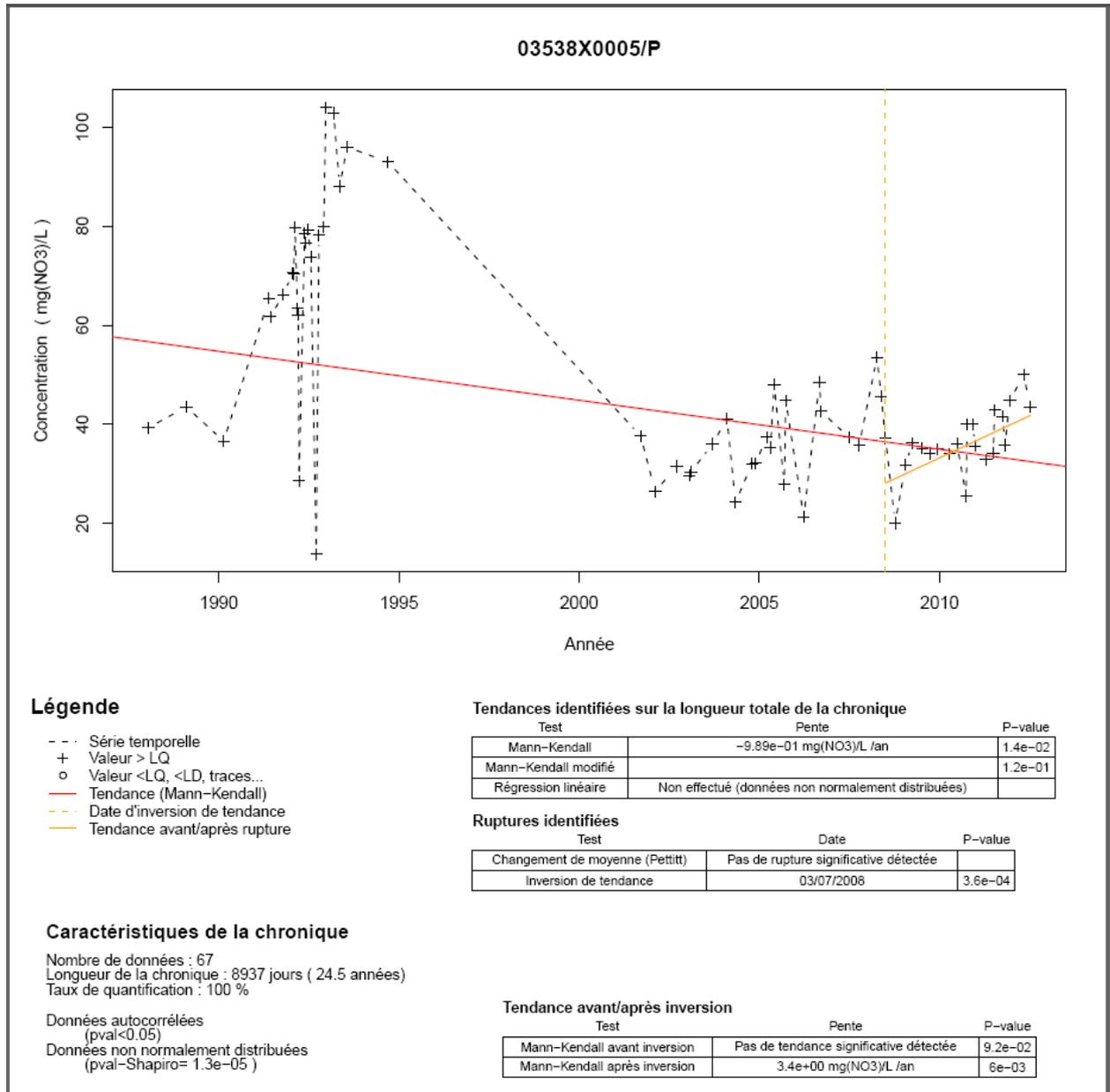


Illustration 42 : Copie d'écran de l'outil pour l'analyse de la chronique de nitrates pour le point 03538X0005/P (Classe C3)

Cette chronique est autocorrélée et a une distribution qui n'est pas normale. Selon ces caractéristiques, le test de Mann-Kendall modifié doit être privilégié pour la recherche de la tendance. Pour la recherche de rupture de pente, les tests de Darken et de Pettitt sont à utiliser.

La méthode de calcul de Pettitt est réalisée, par l'outil, pour la recherche de rupture de pente. A partir des données, l'outil conclut à une absence de rupture selon ce test. Il affiche le texte suivant "Pas de rupture significative détectée". Selon la méthode de Darken, une date de rupture est détectée le 03-07-2008. Un seul segment, ayant une tendance significative à la hausse, est détecté entre 2008 et 2011. L'outil n'arrive pas à détecter deux segments significatifs, (Illustration 42). Pour cette chronique, la méthode de calcul de tendance de Mann-Kendall, est réalisée. L'outil affiche le texte suivant "Mann-Kendall |-9.89e-01 mg(NO₃)/L/an |

1.4e-02". Le signe de la pente étant négatif, cela indique qu'une tendance significative à la baisse est détectée (*Illustration 42*).

Cet exemple est intéressant puisqu'une tendance globale à la baisse a été mise en avant alors que, sur la dernière période (2008-2011), il apparaît que les concentrations augmentent.

• **01546X0065/F**

La chronique suivante est celle représentant la classe C4. Cette classe représente les chroniques qui ont des données autocorrélées, non normales et qui ont des taux de quantification inférieurs à 50% (*Illustration 43*). Cette chronique est définie par un nombre d'analyses de 23 et une moyenne des écarts entre deux analyses de 242 jours. Cette chronique n'est pas très représentative des concentrations en nitrates du réseau RCS dans le sens où son taux de quantification est seulement de 13%, ce qui est très bas. De plus, les concentrations observées (entre 0 et 5 mg/l) sont éloignées de la moyenne des points du RCS. Ceci correspond à un point d'eau sous conditions réductrice avec possibilité de dénitrification expliquant les faibles concentrations en nitrates.

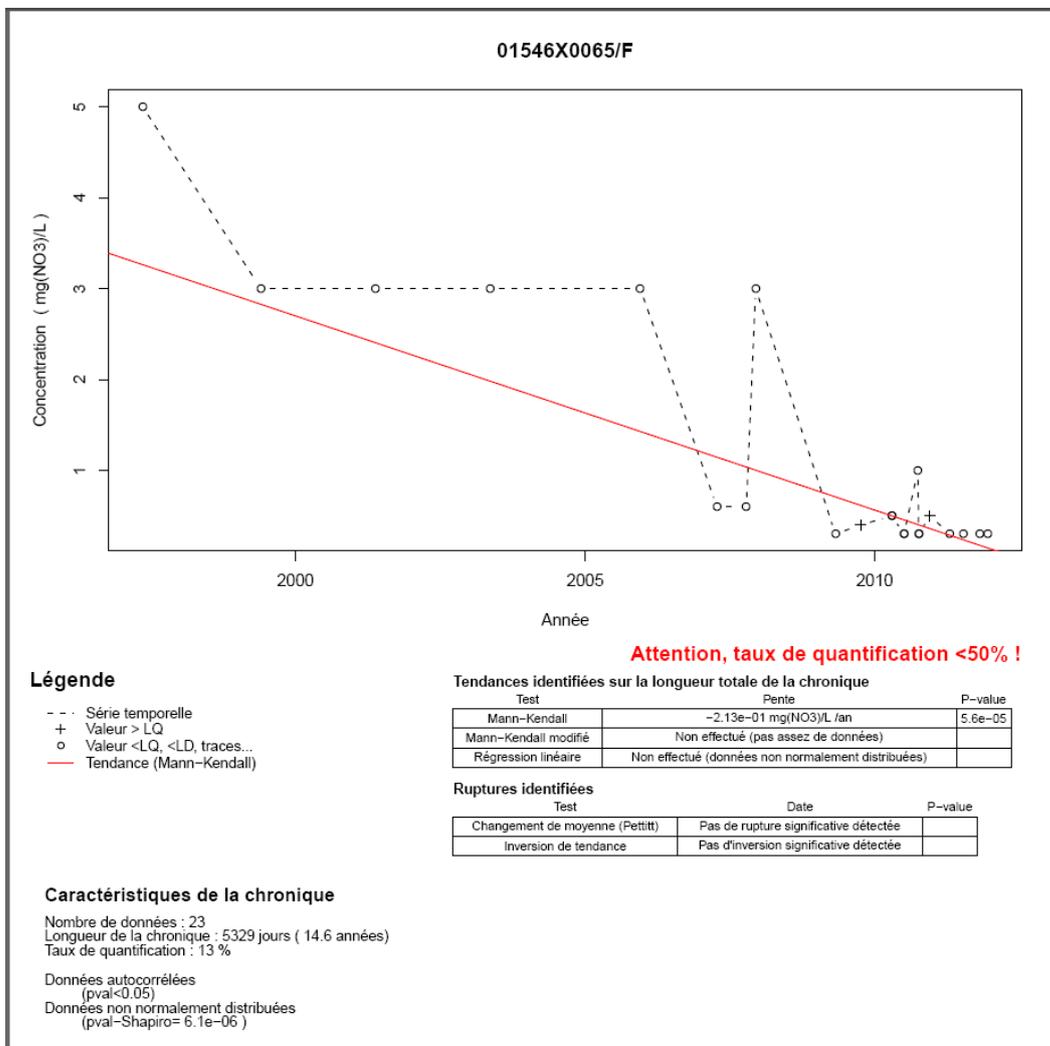


Illustration 43 : Copie d'écran de l'outil pour l'analyse de la chronique de nitrates pour le point 01546X0065/F (Classe C4)

Cette chronique est autocorrélée et a une distribution qui n'est pas normale. Le test de Darken est à envisager pour recherche de rupture de pente. De plus, le test de Mann-Kendall doit être privilégié pour la recherche de la tendance.

Selon les résultats d'analyses de cette chronique, il n'y a pas de rupture. Ainsi l'outil propose le texte suivant "*Pas de rupture significative détectée*". De plus, il n'y a pas d'inversion, le texte suivant est affiché "*Pas d'inversion significative détectée*". Ces deux phrases sont rappelées dans l'encart Ruptures (*Illustration 43*). Selon l'outil une pente existe et une tendance à la diminution est détectée. Le texte suivant est fourni dans l'encart Tendances "*Mann-Kendall | - 2.13e-01 mg(NO3)/L/an | 5.6e-05*" (*Illustration 43*).

Ces tendances ne peuvent pas réellement être considérées comme valides au vu du nombre de point inférieur à la LQ. Aucune conclusion ne peut être tirée à partir d'une chronique comme celle-ci. Les tendances à la baisse ont été tracées par l'outil uniquement en raison d'un changement de traitement analytique des échantillons qui a fait passer la LQ de 5 mg/l à 0,5 mg/l en passant par 3 mg/l et 1 mg/l. Une phrase renseigne explicitement lorsque le taux de quantification est inférieur à 50%.

- **02598X0009/P1**

La chronique suivante est celle représentant la classe C7. Cette classe représente les chroniques qui sont non autocorrélées, normales et qui ont des taux de quantification supérieurs à 50% (*Illustration 44*). Son nombre d'analyses est de 22 et sa moyenne des écarts entre deux analyses est de 236 jours. Son taux de quantification est de 100%.

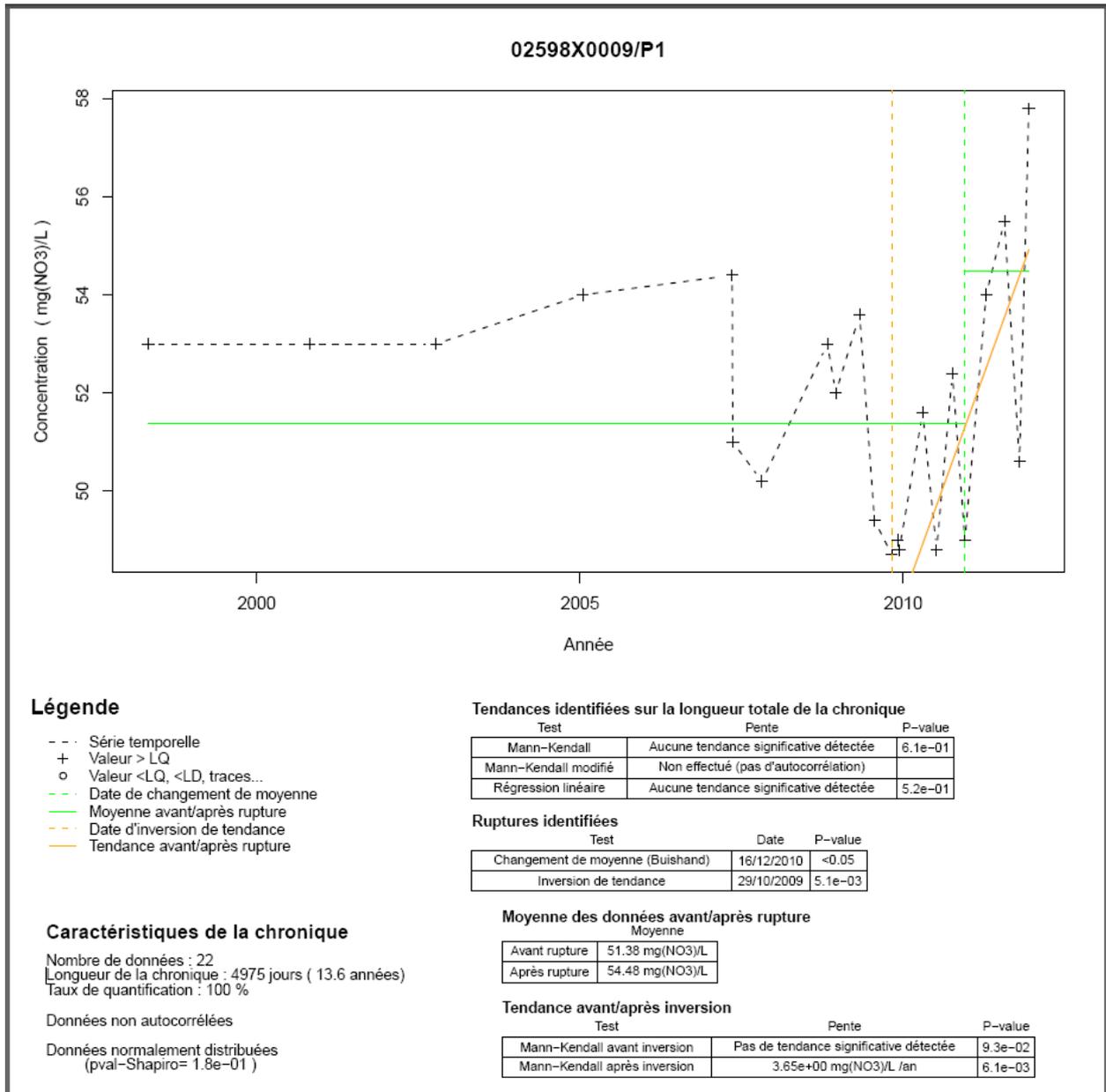


Illustration 44 : Copie d'écran de l'outil pour l'analyse de la chronique de nitrates pour le point 02598X0009/P1 (Classe C7)

Cette chronique possède des données non autocorrélées et qui ont une distribution normale. Ainsi, des tests de rupture de pente de Darken et de Buishand doivent être envisagés. De plus, une recherche de tendance par régression linéaire et par le test de Mann-Kendall doit être réalisée.

Selon la méthode de Buishand, une rupture est constatée. L'outil affiche le texte suivant "Changement de moyenne | 16/12/2010 | <0.05". Ici, une moyenne de concentrations en nitrates de 51,38 mg/l est calculée de 1998 à 2010 puis une autre plus faible (54,48 mg/l) est calculée de 2010 à 2011. De même, pour le test de Darken, une inversion est détectée le 29-10-2009. Selon ce test, il y a un segment à la hausse à partir de 2009. L'outil propose ensuite deux possibilités de calcul pour la tendance. (Illustration 44). Aucune des deux méthodes

utilisées n'identifie de tendance comme l'indique l'outil dans la rubrique Tendances (*Illustration 44*).

Les deux tests de ruptures (Buishand et Pettitt) tendent à montrer une hausse des concentrations sur une période récente.

• **09574X0026/F**

Ce point est représentatif des chroniques non autocorrélées, non normales et dont le taux de quantification est supérieur à 50% (*Illustration 45*). Ces chroniques sont contenues dans la classe C9. La moyenne des écarts entre deux analyses du point 09574X0026/F est de 168 jours et son nombre d'analyses est de 34 analyses.

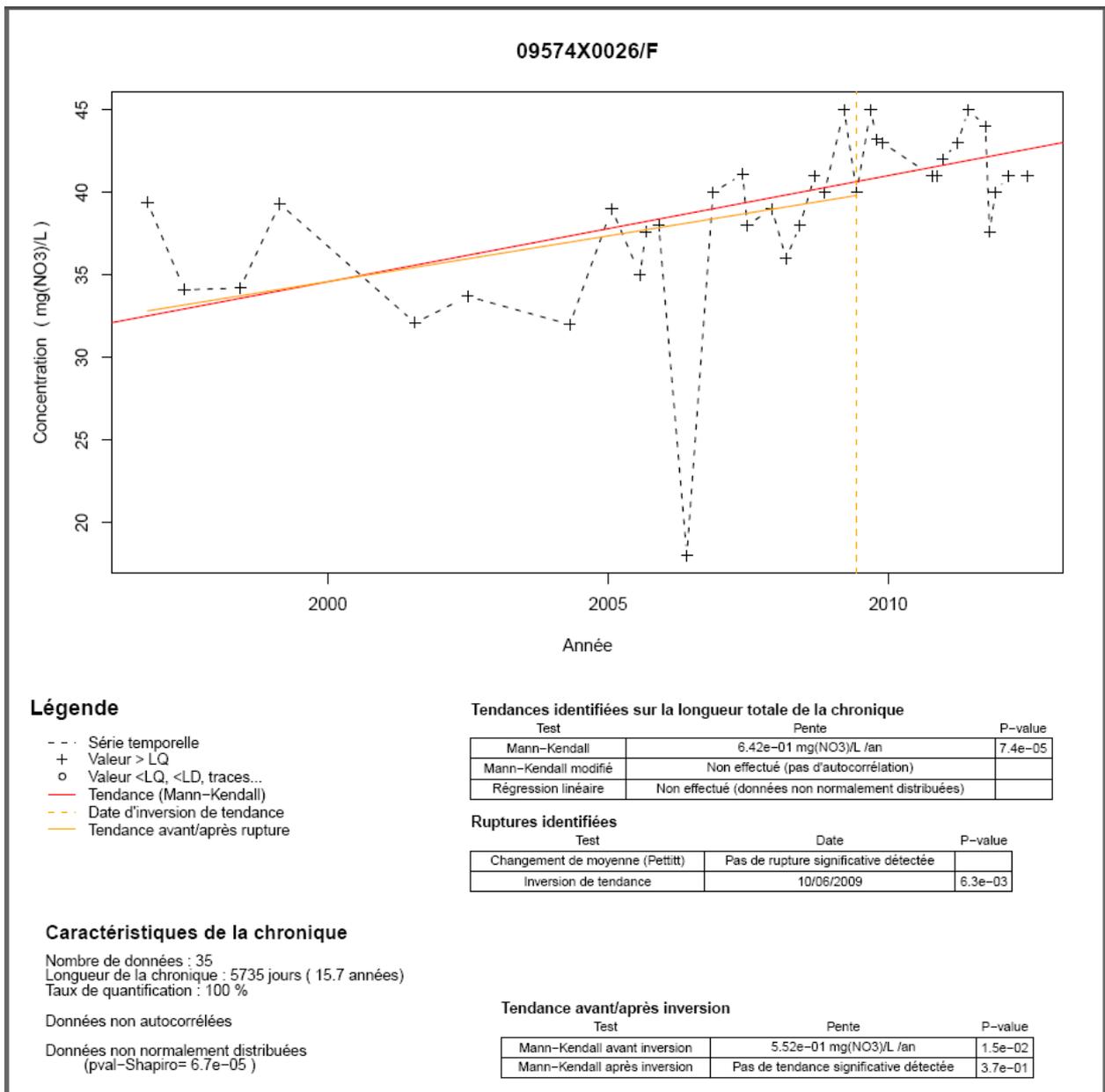


Illustration 45 : Copie d'écran de l'outil pour l'analyse de la chronique de nitrates pour le point 09574X0026/F (Classe C9)

Cette chronique possède des données non autocorrélées et qui n'ont pas de distribution normale. Cela implique que des tests de rupture de pente de Mann-Kendall et de Pettitt doivent être considérés. De plus, une recherche de tendance par un test de Darken peut être envisagée.

Selon les résultats d'analyse de cette chronique, il n'y a pas de rupture. Ainsi l'outil propose le texte suivant "*Pas de rupture significative détectée*". Selon la méthode de Darken, une rupture est détectée le 10-06-2009. Une tendance à la hausse est détectée de 1996 à 2009. Le segment, indiquant la tendance, est tracé sur le graphique (*Illustration 45*). L'outil indique le résultat du test de Mann-Kendall pour la tendance. Selon ce test, une tendance à la hausse significative existe. Le texte suivant est affiché par l'outil dans la rubrique Tendances "*Mann-Kendall |6,42e-01 mg(NO3)/L/an | 7.4e-05*" (*Illustration 45*).

5. Conclusion

Cette étude a permis de préparer des fiches d'évaluation des tendances en nitrates sur chaque masse d'eau souterraines et de développer un outil à destination des gestionnaires de la qualité des eaux souterraines.

Dans sa première phase et dans la perspective de la révision de l'état des lieux en 2013, un atlas de fiches d'identification des tendances d'évolution des concentrations en nitrate à l'échelle des points de prélèvements et de la masse d'eau, a été fourni pour chaque bassin (à l'exception de Mayotte par manque de données). Ces fiches présentent, entre autres, les résultats des tests statistiques Mann-Kendall par points et Kendall Régional par masse d'eau, sous forme cartographique et graphique, considérés comme étant les tests les plus robustes pour ce type de données environnementales. Les données brutes et valorisées ont également été fournies dans une géodatabase sous format Access.

Dans la seconde phase de l'étude, un outil a été développé sous R pour l'analyse des évolutions des concentrations en contaminants dans les eaux souterraines.

Les différents modules proposés dans cet outil permettent d'effectuer les tests de recherche de tendances et de rupture les plus adaptés aux caractéristiques des données étudiées. Ainsi, suivant le nombre de données, la distribution des données et la présence ou non d'autocorrélation dans les données, des tests différents sont proposés. Le module principal permet la recherche de tendance monotone d'évolution, de changement de moyenne, d'inversion de la pente de tendance, et permet d'étudier la variabilité des données entre saisons. Des modules complémentaires permettent de rechercher des tendances d'évolution en tenant compte du caractère saisonnier des données, de rechercher des tendances d'évolution à une échelle régionale et un dernier module propose de rééchantillonner les données pour ne garder que des analyses pour lesquelles la fréquence d'échantillonnage est régulière.

Cet outil permettra donc aux gestionnaires d'effectuer simplement et de manière automatisée des tests statistiques pour la détection de ruptures et de tendances dans les chroniques d'évolution des concentrations quelques soient les contaminants et les caractéristiques des données. D'autre part, il pourra aider à coordonner les actions menées dans le domaine de l'évolution des concentrations en contaminants dans les eaux souterraines, en permettant à l'ensemble des acteurs nationaux d'appliquer une méthodologie unique. Il est alors prévu de former les gestionnaires à l'utilisation de l'outil dans le courant des années 2013 et 2014. Cette formation prévoit aussi d'aborder le problème de l'interprétation des résultats issus de l'outil. En effet, compte tenu de la complexité des tests statistiques retenus d'une part et de la grande variabilité des concentrations dans les eaux souterraines d'autre part (mise en jeu de nombreux facteurs comme la pluie, les propriétés de la zone non saturée, l'évolution des pressions anthropiques....), l'expertise du thématicien est indispensable à la bonne utilisation de l'outil.

6. Bibliographie

Blum A., Chassiot L. (2011). Rapportage des premiers plans de gestion des districts hydrographiques - Analyse des éléments rapportés sur les eaux souterraines le 22 mars 2010. Rapport BRGM, RP-59860-FR. 151p.

Conover W.L. (1980). Practical nonparametric statistics, 2d ed.: New York, John Wiley and Sons, 493p.

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

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

Helsel D.R., Frans L.M. (2006). Regional Kendall Test for Trend. *Environmental Science and Technology*, 40, 13. 4066-4073

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

Hyndman R., Fan Y. (1996). Sample quantiles in Statistical Packages, *The American Statistician*, 50 (4), 361-365.

IFEN (2002). Les pesticides dans les eaux, bilan annuel 2002. IFEN. Etudes et Travaux, n°36. 25p.

Lopez, B. Leynet, A., Blum, A., Baran, N. (2011). Evaluation des tendances d'évolutions des concentrations en polluants dans les eaux souterraines - Revue des méthodes statistiques existantes et recommandations pour la mise en œuvre de la DCE. Rapport BRGM, BRGM/RP-59515-FR. 166p.

Lopez, B. (2011). Evaluation des tendances d'évolutions des concentrations en polluants dans les eaux souterraines - Guide méthodologique. Rapport BRGM, RP-59931-FR. 29p.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

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

Royston P. (1982). An extension of Shapiro and Wilk's *W* test for normality to large samples. *Applied Statistics*, 31, 115–124.

Sen P.K., (1968). Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*, 63, 1379-1389.

Sen P.K. (1980). Asymptotic theory of some tests for a possible change in the regression slope occurring at an unknown time point. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 52, 203-218.

Shapiro S.S. et Wilk M.B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52,2 and 3, 591-597.

SOeS (2011). Bilan de présence des micropolluants dans les milieux aquatiques continentaux - Période 2007-2009. SOeS (Service de l'Observation et des Statistiques). Études & documents. 117p.

SOeS (2012). Mise en oeuvre de la directive cadre sur l'eau : position de la France en Europe en 2009. SOeS (Service de l'Observation et des Statistiques). Observation et Statistiques. 10p.

Yue S., Wang C. (2004). The Mann-Kendall Test Modified by Effective Sample Size to Detect Trend in Serially Correlated Hydrological Series. *Water Resource Management* 18, 201-218.

Annexe 1

Contenu du fichier ZIP «Tendances_NO3_ONEMA_XXXX»

Extrait du document « A_Lire_Avant_Utilisation » fourni à chaque agence de bassin.

- Dossier « **Fiches** » : fiches Tendances NO₃ en format PDF et déclinées selon les codes des masses d'eau souterraines (MESO) ;
- Fichier MXD (format ArcGIS) « **MESO_XXXX** » compilant la majorité des données des bases avec une symbologie rappelant celle des fiches ;
- Geodatabase Access « **BD_MESO_XXXX** » contenant les informations suivantes:
 - Une couche de points « **Kendall_Pts_XXXX** » qui correspond à l'ensemble des points pour lesquels un test Mann-Kendall par point a été réalisé (leur nombre d'analyses nitrates est supérieur ou égal à 10 entre le 01/10/1996 et le 01/10/2011). Sa table attributaire contient les informations suivantes :
 - CODE_BSS
 - XOUVL2E : Coordonnées X en Lambert II Etendu
 - YUOVL2E : Coordonnées Y en Lambert II Etendu
 - CODE_MESO : Code numérique de la MESO rattachée au point
 - NOM_MESO : Nom de la MESO rattachée au point (*optionnel*)
 - RCS : Appartenance au Réseau national de Contrôle de Surveillance Qualité
 - RCO : Appartenance au Réseau national de Contrôle Opérationnel Qualité
 - val_Mediane : Valeur de la tendance d'évolution de la concentration en Nitrates (en mg/l/an)
 - ZS : paramètre déterminant la significativité (ou non) de la tendance calculée – si $|ZS| > 1.96$ alors tendance significative au seuil de confiance de 5%
 - nb_analyses : nombre d'analyses dans la période considérée
 - min_analyse : Concentration minimale en Nitrates, mesurée dans la période considérée
 - max_analyse : Concentration maximale en Nitrates, mesurée dans la période considérée
 - date_min : Date de la 1^{ère} mesure réalisée dans la période considérée
 - date_max : Date de la dernière mesure réalisée dans la période considérée
 - Une couche de points « **Kendall_exclusion_XXXX** » qui correspond à l'ensemble des points pour lesquels le test Mann-Kendall par point n'a pas pu être réalisé (leur nombre d'analyses nitrates est inférieur strict à 10 entre le 01/10/1996 et le 01/10/2011). Sa table attributaire contient les informations suivantes :
 - CODE_BSS
 - XOUVL2E : Coordonnées X en Lambert II Etendu
 - YUOVL2E : Coordonnées Y en Lambert II Etendu
 - CODE_MESO : Code numérique de la MESO rattachée au point

- NOM_MESO : Nom de la MESO rattachée au point (*optionnel*)
 - RCS : Appartenance au Réseau national de Contrôle de Surveillance Qualité
 - RCO : Appartenance au Réseau national de Contrôle Opérationnel Qualité
 - Moy_ : Moyenne des concentrations en Nitrates mesurées durant la période considérée
 - Max_ : Concentration maximale mesurée durant la période considérée
 - nb_analyses : nombre d'analyses dans la période considérée
- Une couche de polygones « **MESO_AIINIV_XXXX** » qui correspond à l'ensemble des masses d'eau souterraine considérées, avec subdivision selon les ordres ;
 - Une couche de polygones « **MESO_AIINIV_Dissolve_XXXX** » qui correspond à l'ensemble des extensions maximales des masses d'eau souterraine considérées (pas de notion d'ordre);
 - Une table « **PTSBSS_INFOS_XXXX** » qui correspond aux données brutes descriptives des points. Sa table attributaire comporte des champs très similaires à ceux décrits précédemment mais également un champ « SOURCE_POINT » décrivant la provenance du point (extraction ADES, Données Agence...etc...) ;
 - Une table « **PTSBSS_VALEURS_XXXX** » qui correspond aux données brutes des analyses nitrates pour les points de la table « PTSBSS_INFOS_XXX ». Sa table attributaire comporte :
 - CODE_BSS
 - DATE_PRELEVEMENT
 - CODE_PARAM : code sandre du paramètre – toujours 1340 pour cette étude
 - VALEUR
 - REMARQUE_ANALYSE
 - LIMITE_DETECTION
 - UNITE
 - SOURCE_ANALYSE : décrit la provenance de l'analyse (Extraction ADES, Données agence...)
 - Une table « **Stats_Kendall_regional_1XXX** » qui comporte les informations suivantes :
 - CODE_MESO
 - Date_Min : Date de la 1^{ère} mesure réalisée dans la période considérée et pour les points pris en compte dans le test de Kendall régional
 - Date_Max : Date de la dernière mesure réalisée dans la période considérée et pour les points pris en compte dans le test de Kendall régional
 - Tendence_REG : Valeur de la tendance régionale d'évolution de la concentration en Nitrates pour la MESO (en mg/l/an)

- Abs_ZS : Valeur absolue du paramètre Zs, intermédiaire à la détermination de la pValue
 - p_Value : pValue du test de Kendall régional pour la MESO considérée
 - p_Value_COM : Commentaires sur la significativité de la tendance selon la valeur de la pValue et le seuil de confiance 5%
- Une table « **Stats_MK_par_points_1XXX** » qui comporte les informations suivantes :
- CODE_MESO
 - NB_Points : Nombre de points avec au moins une analyse nitrates dans la période considérée
 - NB_MK : Nombre de points pour lesquels un test Mann-Kendall par points a pu être entrepris (nb analyses NO3 > =10)
 - NB_nonTendance : Nombre de points pour lesquels un test Mann-Kendall par points a pu être entrepris mais qui n'ont pas montré de tendance significative ($|ZS| < 1.96$ soit $pValue > 0.05$)
 - NB_Tendance : Nombre de points pour lesquels un test Mann-Kendall par points a pu être entrepris et qui ont montré une tendance significative ($|ZS| > 1.96$ soit $pValue < 0.05$)
 - NB_Hausse : Nombre de points à tendance significative à la hausse (valeur du champ « Val_Mediane » > 0 mg/l/an)
 - NB_Sup025mgl : Nombre de points à tendance significative à la hausse avec une valeur du champ « Val_Mediane » > 0.25 mg/l/an
 - NB_Sup05mgl : Nombre de points à tendance significative à la hausse avec une valeur du champ « Val_Mediane » > 0.5 mg/l/an
 - NB_Baisse : Nombre de points à tendance significative à la baisse (valeur du champ « Val_Mediane » < 0 mg/l/an)
 - NB_Infm025mgl : Nombre de points à tendance significative à la hausse avec une valeur du champ « Val_Mediane » < -0.25 mg/l/an
 - NB_Infm05mgl : Nombre de points à tendance significative à la hausse avec une valeur du champ « Val_Mediane » < -0.5 mg/l/an



Centre scientifique et technique
Direction Eau, Environnement et Ecotechnologie
3, avenue Claude-Guillemain
BP 36009 – 45060 Orléans Cedex 2 – France – Tél. : 02 38 64 34 34
www.brgm.fr