

Bilan des différentes approches mises en œuvre lors de l'expertise sur les données bancarisées et propositions d'harmonisation

N. Guigues

Avec la collaboration de D. Soudant, A. Assoumani, J-P.
Ghestem

Mai 2020

Note de synthèse

En partenariat avec



Avec le soutien de :
**AGENCE FRANÇAISE
POUR LA BIODIVERSITÉ**
ÉTABLISSEMENT PUBLIC DE L'ÉTAT





Avec le soutien de :
**AGENCE FRANÇAISE
POUR LA BIODIVERSITÉ**
ÉTABLISSEMENT PUBLIC DE L'ÉTAT



Contexte de programmation et de réalisation

Ce rapport a été réalisé dans le cadre du programme scientifique et technique AQUAREF pour l'année 2019, au titre de l'action « Expertise des données bancarisées » dans le cadre du thème E « Garantir la qualité des données bancarisées ».

Auteurs :

Nathalie Guigues
LNE
Nathalie.guigues@lne.fr

Azziz Assoumati
INERIS
Azziz.assoumani@ineris.fr

Dominique Soudant
IFREMER
Dominique.Soudant@ifremer.fr

Jean-Philippe Ghestem
BRGM
jp.ghestem@brgm.fr

Approbateur :
Sophie Vaslin-Reimann
LNE
sophie.vaslin-reimann@lne.fr

Vérification du document :

Bénédicte Lepot
Ineris
Bénédicte.lepot@ineris.fr

Emilie Gauthier
Ifremer
Emilie.gauthier@ifremer.fr

Les correspondants

AFB : Nicolas Gaury, nicolas.gaury@ofb.gouv.fr

LNE : Sophie Vaslin-Reimann sophie.vaslin-reimann@lne.fr

Référence du document : Guigues N., Soudant D., Assoumani A., Ghestem JP. - Bilan des différentes approches mises en œuvre lors de l'expertise sur les données bancarisées et propositions d'harmonisation- Note de synthèse AQUAREF 2020 - 15 p.

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

SOMMAIRE

1. INTRODUCTION.....	5
2. BILAN DES ÉTUDES AQUAREF RÉALISÉES ENTRE 2011 ET 2018.....	6
2.1 Etat des lieux des bases de données 2007-2009.....	6
2.2 Etudes visant à identifier des effets laboratoires dans les séries de données	7
2.3 Evolution temporelle des performances des laboratoires.....	7
2.4 Estimation des incertitudes de mesure à partir de séries chronologiques. .	8
3. IDENTIFICATION DES POINTS CRITIQUES EN LIEN AVEC LES DONNÉES CENSURÉES.....	8
3.1 stratégies de remplacement des données inférieures à la limite de quantification.....	9
3.2 Cas des séries de données contenant plusieurs LQ.....	10
3.3 Impacts sur les descripteurs statistiques.....	11
4. PROPOSITIONS MÉTHODOLOGIQUES POUR L'ANALYSE DES DONNÉES BANCARISÉES.....	12
5. CONCLUSION.....	13
6. Références bibliographiques.....	13

1. INTRODUCTION

L'amélioration continue de la qualité des données de la surveillance a pour objectif de rendre plus fiables et mieux documentées les données produites dans le cadre de la surveillance des milieux aquatiques. Ces données sont bancarisées dans des bases de données nationales (ADES pour les eaux souterraines, QUADRIGE pour le milieu marin, Naiades pour les eaux de surface) ou propres à chaque Agence ou Office de l'eau. Maîtriser la qualité de la donnée, notamment au travers des métadonnées associées, permet à la fois de faciliter leur exploitation dans le cadre d'expertises mais aussi de mieux les comparer dans l'espace et le temps.

Depuis 2011, différentes études ont été conduites dans le cadre des programmes Aquaref, pour i) vérifier la complétude des données et identifier les erreurs de codification ii) évaluer l'évolution des performances des méthodes analytiques employées et leur conformité au regard des exigences réglementaires (Ghestem, 2012 ; Guigues 2012), iii) identifier d'éventuels effets laboratoires (Bristeau et Ghestem, 2013, 2015 et 2016 ; Ngo et Botta, 2016) ou encore iv) estimer les incertitudes de mesure à partir de séries chronologiques (Soudant et al., 2015 et 2017).

Au travers de ces études, le constat a pu être fait que l'une des questions premières qui se posait lors de l'exploitation des données bancarisées concerne la manière de prendre en compte les données dites « censurées », c'est-à-dire les données inférieures à une limite (détection ou quantification par exemple). De plus, compte-tenu des évolutions des exigences réglementaires et des performances des moyens de mesure, il est très courant d'avoir à exploiter des séries de données caractérisées par des limites de détection ou quantification différentes.

Un autre point qu'il est important d'adresser est le choix des descripteurs statistiques (comme la moyenne, la médiane etc.) à estimer ainsi que le type d'approche en lien avec la distribution des données (par exemple paramétrique pour les données distribuées selon une loi normale, ou non paramétrique dans les autres cas) et les tests statistiques à mettre en œuvre. Ces différents choix sont en général très dépendants du jeu de données (nombre, distribution etc.).

Les objectifs de cette note sont, dans un premier temps, de présenter les différentes études réalisées dans le cadre du programme Aquaref entre 2011 et 2018, dans un deuxième temps, de décrire les approches existantes avec leurs inconvénients et leurs avantages pour répondre notamment à la question des données censurées et des approches statistiques les mieux adaptées à mettre en œuvre et enfin dans un troisième temps de faire des propositions et recommandations pour améliorer l'exploitation des données bancarisées.

2. BILAN DES ÉTUDES AQUAREF RÉALISÉES ENTRE 2011 ET 2018

Dans le cadre des programmes Aquaref de 2011 à 2018, plusieurs études ont été réalisées sur les données bancarisées entre 2007 et 2018 pour répondre principalement à quatre objectifs :

- Evaluer la qualité des données bancarisées en réalisant un état des lieux des champs renseignés dans les bases de données et proposer des pistes d'amélioration ;
- identifier des potentiels effets laboratoires, notamment en comparant des données issues de différentes sources ;
- mettre en évidence les évolutions de performances des laboratoires au cours du temps et leur impact sur les séries temporelles ;
- estimer les incertitudes de mesure à partir de séries chronologiques.

2.1 ETAT DES LIEUX DES BASES DE DONNÉES 2007-2009

Trois études ont été réalisées à partir des données bancarisées entre 2007 et 2009 dans les bases de données ADES sur les eaux souterraines (Ghestem, 2012) et BNDE sur les eaux de surfaces (Guigues, 2012) ainsi qu'à partir d'une compilation de données sur les eaux de surfaces des Agences de l'eau (Strub, 2011 communication personnelle).

Les paramètres concernés étaient les micropolluants et les composés majeurs (ions constitutifs, nutriments).

Au travers de ces études, il s'agissait principalement d'analyser le niveau de renseignement des métadonnées associées aux résultats analytiques permettant non seulement d'apprécier la qualité des données bancarisées mais également de garantir leur exploitabilité.

Les champs suivants ont fait l'objet d'une attention particulière :

- Support et fraction
- LQ
- Incertitude analytique
- Méthode analytique
- Accréditation des laboratoires

Une comparaison entre les LQ renseignées avec, d'une part, les valeurs saisies comme données non quantifiées et, d'autre part, les exigences de l'arrêté du 27 octobre 2011 (transposition de la directive européenne QA/QC) a aussi été réalisée.

Les principales conclusions de ces études sont que les champs LQ, méthode analytique et incertitude analytique sont très peu renseignés, ou s'ils le sont, pas toujours de manière exploitable. Pour certains paramètres comme les nutriments, une coexistence de données analysées sur eau brute et sur fraction dissoute ne permet pas de comparer les données entre elles.

Des recommandations pour améliorer et harmoniser la saisie de ces champs ont été émises au travers de ces études. Par exemple, il a été recommandé de

rendre obligatoire la saisie des champs cités ci-dessus, notamment pour rendre compte de l'application de la directive QA/QC, et de proscrire des codes génériques comme le code 0 pour « inconnu ».

2.2 ETUDES VISANT À IDENTIFIER DES EFFETS LABORATOIRES DANS LES SÉRIES DE DONNÉES

Le BRGM a réalisé trois études depuis 2013 (Bristeau et Ghestem 2013 ; 2015 et 2016) afin de vérifier la qualité des données de surveillance bancarisées, identifier les anomalies, en lien avec un potentiel effet laboratoire, et si possible quantifier leur impact sur la qualité des données.

Afin d'évaluer si les résultats issus de deux laboratoires sont différents, la méthodologie mise en œuvre est basée sur 3 approches distinctes :

- Ecart maximal admissible calculé sur la base des coefficients de variation de reproductibilité (CVR) estimés lors d'essais d'aptitude ;
- Test de justesse utilisant l'incertitude de mesure fournie par les laboratoires ;
- Test de justesse utilisant une incertitude de mesure modélisée.

L'étude d'un jeu de données fourni par l'Agence de l'Eau Seine Normandie correspondant à des prélèvements réalisés en double et analysés par des laboratoires différents dans le cadre de la surveillance DCE et lors de la campagne exceptionnelle de recherche de substances émergentes de 2011, a permis de mettre en évidence qu'environ 30 % à 40 % des écarts observés sont considérés comme significativement différents. (Bristeau et Ghestem, 2013).

La même méthodologie a été mise en œuvre en 2015 (Bristeau et Ghestem, 2015) sur un jeu de données provenant de deux réseaux de mesure : la surveillance DCE et le contrôle sanitaire. Des écarts significatifs ont été observés pour les substances étudiées, allant de 10 % des couples de données (Silice, Bore) à 60 % (Fer). Cependant, ces résultats ont aussi montré qu'il était parfois difficile d'attribuer les écarts observés qu'aux seuls effets analytiques quand les prélèvements n'ont pas été réalisés en même temps ou à un intervalle de temps court au regard de la variabilité des masses d'eau souterraines.

En 2016, afin d'étudier d'éventuels effets laboratoires, une analyse de couples de données issues de deux laboratoires et prélevées dans un délai de 3 jours maximum a été réalisée pour 25 substances (Bristeau et Ghestem, 2016). Pour plusieurs substances comme notamment le boscalide, le DEHP, l'isoproturon et la DEDIA, des effets laboratoires importants ont ainsi pu être identifiés.

2.3 EVOLUTION TEMPORELLE DES PERFORMANCES DES LABORATOIRES

En 2016, deux études ont été réalisées avec pour objectif de s'intéresser aux modifications de séries chronologiques suite à un changement de prestataires d'analyse ou suite à une amélioration des performances analytiques (évolution

des méthodes, changement d'équipement) ou encore suite à une évolution réglementaire (LQ de plus en plus basses exigées) dans le temps (Ngo et Botta, 2018 ; Bristeau et Ghestem, 2016).

Une des études concernait les données acquises par l'Agence de l'Eau Rhône Méditerranée Corse sur les eaux de surface (Ngo et Botta, 2018). Ainsi, il a pu être mis en évidence que pour certaines substances, une diminution de la limite de quantification a entraîné une augmentation des fréquences de quantification (par exemple cadmium, plomb, nickel, naphthalène, nonylphénols, benzo(a)pyrène, benzo(k)fluoranthène, benzo(b)fluoranthène). Pour certaines substances (mercure, benzo(a)pyrène, trichlorobenzènes, trichlorométhane, indeno(1,2,3,cd)pyrène), des modifications autres que le changement de limite de quantification ont eu lieu simultanément.

La deuxième étude adressait les concentrations d'atrazine dans les eaux souterraines de la base ADES (Bristeau et Ghestem, 2016). L'analyse des séries chronologiques n'a cependant pas permis de mettre en évidence sans équivoque d'effet laboratoires car les évolutions de pratiques et de méthodes peuvent avoir lieu au sein d'un même laboratoire.

2.4 ESTIMATION DES INCERTITUDES DE MESURE À PARTIR DE SÉRIES CHRONOLOGIQUES

A partir de l'exploitation des séries temporelles de la base de données Quadrigé, deux études ont été réalisées en 2015 et 2017 sur les eaux littorales afin de quantifier les incertitudes de mesure de paramètres tels que la chlorophylle a (Soudant et al., 2015) et les paramètres hydrologiques comme notamment les nutriments et la turbidité (Soudant et al., 2017). En effet, comme ces séries temporelles présentent la caractéristique d'intégrer toutes les variabilités liées à la fois à la mesure et à l'environnement, leur analyse en termes de signal et bruit permet d'évaluer l'amplitude des incertitudes. Les modèles linaires dynamiques (DLM) ont ainsi été utilisés dans ce but.

Si pour la température et l'oxygène dissous, la part de variabilité liée à l'acquisition de la donnée reste petite (e.g. resp. ± 1.35 °C et ± 0.26 mg/L), elle est souvent élevée et très variable pour les autres paramètres. En ce qui concerne les comptages phytoplanctoniques, leur expression en échelle log est associée à une variabilité qui reste très raisonnable (i.e. $\pm 13.5\%$). Ces variabilités sont également dépendantes des lieux d'échantillonnage et du plan d'échantillonnage.

3. IDENTIFICATION DES POINTS CRITIQUES EN LIEN AVEC LES DONNÉES CENSURÉES

La grande majorité des données environnementales est censurée au niveau des limites de détection (LD) ou de quantification (LQ) : les laboratoires accrédités impliqués dans la surveillance environnementale ne rendent pas les résultats obtenus quand ils sont inférieurs à la LQ. Par conséquent, une stratégie de remplacement de ces données inférieures à la LQ est souvent nécessaire.

Ces stratégies sont présentées dans ce paragraphe avec une estimation de leur impact sur les descripteurs statistiques usuels.

3.1 STRATÉGIES DE REMPLACEMENT DES DONNÉES INFÉRIEURES À LA LIMITE DE QUANTIFICATION

Il existe principalement deux types de manière de remplacer les données inférieures à la limite de quantification :

- Remplacer par une valeur arbitraire

La méthode assez couramment utilisée pour remplacer les données inférieures à la LQ est de les substituer par une fraction de cette LQ, généralement comprise entre 0 et 1.

Cette approche est imposée dans la surveillance DCE par la directive QA/QC pour le calcul des moyennes, en remplaçant les données inférieures à la LQ par $LQ/2$ pour des substances individuelles, et par 0 pour les sommes de substances.

- Reconstruire la série de données

Il est possible de reconstituer la série de données en attribuant des valeurs aux données inférieures à la LQ afin d'obtenir une distribution des données qui ne soit pas tronquée à gauche (Helsel, 2005).

Par exemple un modèle de type MLE (maximum likelihood estimation - maximum de vraisemblance) peut être utilisé à cet effet pour des grandes séries de données ($n > 30-50$). Un exemple de ce type de modèle MLE pour une distribution de type log normal est présenté dans la Figure 1.

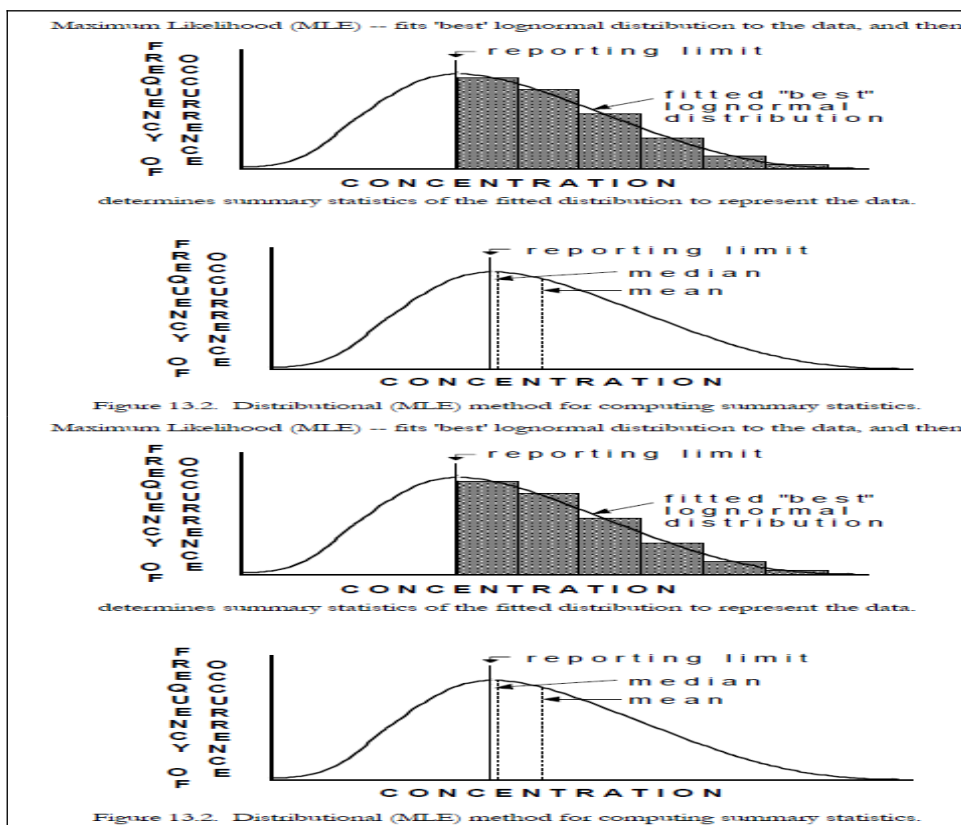


Figure 1 : Exemple d'une approche MLE : reconstruction des données à partir de l'ajustement d'un modèle de loi log normal (en haut) et estimation de la moyenne et de la médiane (en bas), d'après Helsel et Hirsch (2002)

Comme alternative pour des petites séries de données ($n < 30$), il est possible de mettre en œuvre des méthodes non paramétriques comme le test de Wilcoxon ou le coefficient de corrélation τ de Kendall) comme suggéré par Helsel (2006).

3.2 CAS DES SÉRIES DE DONNÉES CONTENANT PLUSIEURS LQ

De par l'évolution des exigences réglementaires, mais aussi l'évolution des performances analytiques au sein d'un même laboratoire, ou parce que plusieurs laboratoires ayant des performances analytiques quelque peu différentes sont impliqués dans l'acquisition des données, il est courant de se retrouver avec une série de données caractérisée par plusieurs limites de quantification.

Quevauvillers (2010) propose d'identifier la LQ la plus élevée, c'est-à-dire la LQmax et de remplacer toutes les données inférieures à cette LQmax, y compris des valeurs quantifiées, par LQmax/2.

Cette approche a le désavantage cependant de conduire à remplacer potentiellement un grand nombre de données quantifiées.

Pour pallier cet inconvénient majeur, une méthodologie a été développée par Assoumani et al. (2018). Elle consiste à définir une LQ seuil, qui est virtuelle, mais qui permet d'optimiser en le limitant, le nombre de données quantifiées inférieure à cette LQ seuil.

Les critères proposés par Assoumani et al. (2018) pour définir cette LQ seuil sont:

- le nombre de données quantifiées au-dessus de LQ seuil doit être le plus important possible (en minimisant ce seuil) ;
- la période d'étude doit être la plus large possible (en veillant cette fois à ce que le seuil ne soit pas trop bas).

Un exemple pour le trichlorométhane est présenté dans la Figure 2.

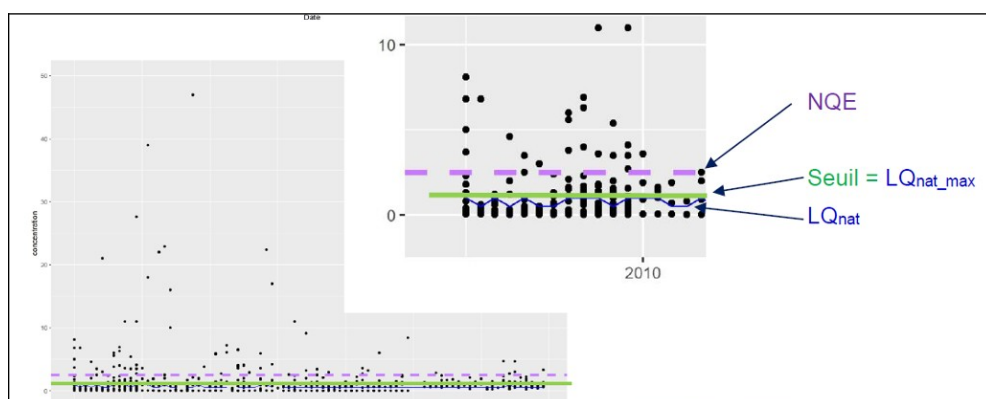


Figure 2 : Chronique des concentrations (données quantifiées) du trichlorométhane pour le réseau RCO, avec application du seuil, sur la période 2009-2015 pour définir la LQ seuil. La ligne verte représente le seuil, la ligne discontinue violette représente la NQE (MA = 2,5 µg/L), et la ligne bleue représente la valeur maximale de la LQ à l'échelle nationale (d'après Assoumani et al., 2018)

3.3 IMPACTS SUR LES DESCRIPTEURS STATISTIQUES

Helsel (2006) a montré que le fait de remplacer des données inférieures à la LQ par une fraction de cette LQ engendrait une distorsion, parfois très importante, de l'estimation de la moyenne et de l'écart type avec des conséquences non négligeables pour les tests paramétriques utilisant ces estimateurs statistiques (Analyse de variance, corrélation linéaire par exemple).

Si le pourcentage de données inférieures à la LQ est inférieur à 50 % du nombre total de données, alors il est possible d'utiliser la médiane comme descripteur statistique, car cette dernière ne sera pas affectée par le nombre de données non quantifiées, contrairement à la moyenne, comme cela est illustré pour le Vanadium (14%) et le Bore (42%) dans le Tableau 1.

Par contre pour des pourcentages de données non quantifiées plus importants, il est préférable de reconstituer la série de données car les descripteurs comme la moyenne ou la médiane sont très impactés, comme c'est le cas du Nickel (65 %) dans le Tableau 1. Ainsi dans ce cas, la moyenne est largement sous-estimée quand les données inférieures à la LQ sont remplacées par 0 et inversement surestimées quand elles sont remplacées par la valeur LQ.

Substance	Nombre de données	LQ (µg/L)	% de données < LQ	Données < LQ remplacées par	Moyenne (µg/L)	Ecart type (µg/L)	Médiane (µg/L)
Vanadium	2477	0.1	14%	0	0.42	0.42	0.32
				LQ/2	0.42	0.41	0.32
				LQ	0.43	0.41	0.32
Bore	2477	10	42%	0	17.1	52.7	12
				LQ/2	19.2	52.1	12
				LQ	21.3	51.6	12
Nickel	6706	0.5	65%	0	0.32	0.70	0
				LQ/2	0.49	0.63	0.25
				LQ	0.65	0.58	0.5

Tableau 1 : Descripteurs statistiques (moyenne, écart type et médiane) pour 3 substances ayant des pourcentages de données inférieures à la LQ compris entre 14 % et 65 %. Les données sont issues des stations du réseau de contrôle de surveillance (RCS) cours d'eau de l'Agence de l'eau Rhône Méditerranée Corse sur la période 2016-2018.

Par ailleurs, il faut garder à l'esprit que les quantiles 5-10 % et 90-95 % sont très sensibles au nombre total de données et seront d'autant plus robustes que le nombre de données est important.

Pour l'étude des séries chronologiques, il existe principalement 2 approches permettant de tenir compte des données inférieures à la LQ (Lopez et Leynet, 2011) :

- travailler en fréquence de dépassement de la LQ : l'information reportée correspond à l'évolution temporelle de la fréquence de dépassement de la LQ
- utiliser les méthodes alternatives proposées par Helsel (2005) qui permettent de reconstruire la série de données, à partir par exemple de la modélisation de la distribution des données et son extrapolation pour les valeurs inférieures à la LQ. Cependant, cette méthode ne permet pas d'attribuer ces valeurs extrapolées à des dates de prélèvement précises. Seule l'estimation des différents descripteurs statistiques de la série chronologique est possible.

4. PROPOSITIONS METHODOLOGIQUES POUR L'ANALYSE DES DONNÉES BANCARISÉES

Avant tout traitement statistique des données, une analyse du jeu de données disponible est primordiale car elle va permettre d'orienter le choix des descripteurs statistiques ainsi que des tests statistiques à mettre en œuvre. Cette analyse peut se dérouler en deux étapes comme décrites ci-après.

- 1) Etape 1 : description du jeu de données à analyser.

En premier, une visualisation graphique des données sous forme de boîte à moustaches par exemple et/ou d'histogramme de distribution permet de mettre

en évidence des distributions de données atypiques et de les comparer entre elles.

Ensuite, il s'agit principalement de décrire le nombre de données disponibles, ainsi que le nombre de données inférieures à une limite de quantification (ou de détection). Ceci permet d'estimer le pourcentage de données inférieures à cette limite de quantification et d'orienter vers la stratégie la mieux adaptée de remplacement de ces données.

Comme dans un jeu de données, plusieurs limites de quantification peuvent coexister, une analyse détaillée de la répartition des données inférieures à chaque limite de quantification permettra de définir une limite virtuelle, ou une LQ seuil, pour ce jeu de données.

2) Etape 2 : remplacement des données inférieures à la limite de quantification.

Si le pourcentage de données inférieures à la limite de quantification est inférieur à 50 %, alors il est possible de remplacer les données inférieures à cette LQ par une valeur égale à la moitié de la limite de quantification. Dans ce cas, seuls les descripteurs statistiques non paramétriques comme la médiane et le quantile 95 % peuvent être utilisés car ils ne sont pas impactés par cette substitution.

Dans le cas opposé, ou si des descripteurs statistiques paramétriques comme la moyenne et l'écart type doivent être estimés, alors il est préférable d'utiliser des modèles de type MLE pour remplacer les données inférieures à la limite de quantification.

Enfin l'identification et la prise en compte des données exceptionnelles doit être réalisée. Ces données exceptionnelles sont souvent le résultat de conditions environnementales et hydrologiques (par exemple un prélèvement réalisé lors d'une importante crue, peut engendrer des concentrations importantes de certains paramètres comme les matières en suspension). A noter que les tests statistiques non paramétriques permettent de ne pas éliminer ces données tout en leur donnant un poids moindre.

5. CONCLUSION

Différentes études ont pu mettre en avant les biais importants engendrés lors de l'estimation des descripteurs statistiques, comme la moyenne, par une stratégie de remplacement des données inférieures à la limite de quantification analytique par une fraction de cette limite. C'est pourtant cette approche qui est recommandée dans le cadre de la DCE alors qu'il faudrait la bannir. Il serait souhaitable d'aborder ce point lors de la révision de la DCE, en proposant des alternatives impactant peu ou pas les descripteurs statistiques habituellement utilisés.

Ainsi, pour estimer des concentrations moyennes à partir de séries de données contenant des valeurs inférieures à la limite de quantification, une approche alternative devrait être envisagée :

- quand le pourcentage de données inférieures à la limite de quantification est inférieur à 50 %, alors la médiane devrait être utilisée car non impactée, contrairement à la moyenne.
- quand le pourcentage de données inférieures à la limite de quantification est supérieure à 50 %, alors une approche de type MLE devrait être mise en œuvre pour permettre l'estimation de la moyenne ou de la médiane. Un module sous R (Nada) existe à cette fin.

6. RÉFÉRENCES BIBLIOGRAPHIQUES

Assoumani A., Salomon M., Jouglet P., Staub P.F., Clavel L., Andrade A. (2018) Bilan du 1er cycle de surveillance de la Directive Cadre sur l'Eau - Evolution des tendances des concentrations, Rapport DRC-18-167427-11774A, 45 p.

Bristeau S. et Ghestem JP. (2013) Etude comparative de données d'analyse de surveillance d'eau souterraine, Rapport Aquaref, 63p.

Bristeau S. et Ghestem JP. (2015) Surveillance de la qualité des eaux souterraines : comparaison de données des réseaux santé et environnement - Rapport Aquaref, 40 p.

Bristeau S. et Ghestem JP. (2016) Etude de données de surveillance d'eau souterraine. Recherche « d'effets laboratoires » à travers l'exploitation de la base ADES. Rapport Aquaref, 41 p.

Directive QA/QC - Directive 2009/90/CE de la Commission du 31 juillet 2009 établissant, conformément à la directive 2000/60/CE du Parlement européen et du Conseil, des spécifications techniques pour l'analyse chimique et la surveillance de l'état des eaux.

Ghestem JP. (2012) Note BRGM AQUAREF sur l'examen des banques de données de surveillance 2007-2009 : banque ADES, 21p.

Guigues N. (2012) Analyse critique des bases de données surveillance pour les paramètres physico-chimiques des eaux de surface (2007 à 2009), Rapport Aquaref, 53 p.

Helsel D.R. et Hirsch R.M. (2002) Statistical Methods , Techniques of Water-Resources Investigations of the United States Geological Survey, Book 4, Hydrologic Analysis and Interpretation, Chapter A3, 510 p.

Helsel D.R. 2005) Nondetects and Data Analysis: Statistics for Censored Environmental Data. John Wiley, New York, 268 p.

Helsel D.R. (2006) Fabricating data: How substituting values for nondetects can ruin results, and what can be done about it, *Chemosphere* 65, 2434-2439

Ngo S. et Botta F. (2016) Impact du changement de laboratoire et des limites de quantification sur le taux d'occurrence des substances - données 1987-2015 de l'Agence de l'Eau Rhône-Méditerranée et Corse, Rapport Aquaref, 42 p.

Quevauviller P. (2010) Protection des eaux souterraines - Législation européenne et avancées scientifiques, Lavoisier, Tec et Doc, 432 p.

Soudant D., Miossec L., Neaud-Masson N., Auby I., Maurer D., Daniel-Scuiller A. (2015) Incertitudes des méthodes d'évaluation « eaux littorales » : utilisation de modèles linéaires dynamiques pour l'évaluation des incertitudes (chlorophylle a, phytoplancton), Rapport Aquaref, 42 p.

Soudant D., Auby I., Daniel A. (2017) Incertitudes des méthodes d'évaluation « eaux littorales » : utilisation de modèles linéaires dynamiques pour l'évaluation des incertitudes des paramètres hydrologiques, Rapport Aquaref, 71 p.