



HAL
open science

Définition des sources d'incertitude du protocole diatomées en cours d'eau : Evaluation de la variabilité inter-opérateurs et de la variabilité interannuelle des notes

M. Wach, J Guéguen, N. Dagens, Juliette Tison-Rosebery, François Delmas

► To cite this version:

M. Wach, J Guéguen, N. Dagens, Juliette Tison-Rosebery, François Delmas. Définition des sources d'incertitude du protocole diatomées en cours d'eau : Evaluation de la variabilité inter-opérateurs et de la variabilité interannuelle des notes. [Rapport de recherche] irstea. 2018, pp.110. hal-02608301

HAL Id: hal-02608301

<https://hal.inrae.fr/hal-02608301v1>

Submitted on 16 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Programme scientifique et technique
Fiche-Action AQUAREF B / Années 2013-2016

Action B3 - Transférer les méthodes de bio-indication :

**Définition des sources d'incertitude du protocole
diatomées en cours d'eau**

**Evaluation de la variabilité inter-opérateurs et de la variabilité
interannuelle des notes**

M. Wach & J. Guéguen, N. Dagens, J. Rosebery & F. Delmas
Avec la participation des DREAL et Bureaux d'Étude impliqués

29-06-2018

Rapport final – V1

**AGENCE FRANÇAISE
POUR LA BIODIVERSITÉ**

Établissement public du ministère de l'Environnement

Avec le soutien de :



et de :

Contexte de programmation et de réalisation

Dans le cadre de la programmation pluriannuelle Aquaref, Irstea est impliqué depuis 2012 dans une action inter-organismes (Fiche-Action B) qui vise à évaluer et à quantifier les incertitudes liées aux méthodes hydrobiologiques utilisées dans le cadre de la surveillance et de l'évaluation des hydrosystèmes de France.

Durant les deux premiers cycles de gestion des Masses d'Eau, les résultats des méthodes abiotiques et biologiques de surveillance des hydrosystèmes étaient utilisés directement, sans estimation d'intervalle de confiance spécifique aux méthodes élémentaires. Faute de processus plus raffiné, l'incertitude attachée à l'évaluation des Masses d'Eau était abordée par un système simple et global d'appréciation du niveau de confiance attribuable à chaque étape du processus de classification, qui s'appuyait sur un arbre de décision faisant partie intégrante des Circulaires et Arrêtés d'Evaluation successivement en vigueur (2005, puis 2010). Cependant, cette pratique ne répondait que partiellement aux attentes de la DCE, qui demande notamment de mieux cerner l'incertitude associée à chaque méthode biologique indiciaire, afin d'assortir le résultat obtenu d'un intervalle de confiance permettant de consolider l'évaluation d'état écologique.

Dans ce contexte général, sous l'égide d'Aquaref, un travail d'état de l'art, de bilan d'étape, de réflexion et de proposition méthodologique pour une meilleure prise en compte de l'incertitude en hydrobiologie a été réalisé en 2012, qui s'est appuyé sur les réflexions menées par le réseau partenarial inter-organismes de ce consortium dans le cadre du groupe de travail national « Incertitude ». Ce travail s'est conclu par l'édition d'un rapport à dominante méthodologique intitulé « L'incertitude en hydrobiologie : concepts et applications ». A la suite de cette réflexion structurante menée par le collectif Aquaref, il a notamment été décidé de lancer, à partir de l'année 2013, un programme national de collecte de données inter-opérateurs permettant de travailler sur les incertitudes liées à l'application du protocole d'acquisition de la donnée hydrobiologique pour deux méthodes végétales : la méthode « diatomées en cours d'eau » et la méthode « macrophytes en cours d'eau ».

C'est dans ce cadre que la présente étude, qui avait pour but de quantifier l'incertitude liée à l'**effet inter-opérateurs** dans le cadre de l'application de l'**IBD 2007** sur **cours d'eau**, a démarré par la réalisation de 2 campagnes annuelles de prélèvements 2013 et 2014, réalisées conjointement par 3 types de partenaires différents (diatomistes des DREAL, de Bureaux d'Étude et d'Irstea). Un rapport d'étape, rédigé par Irstea fin 2015, établissait un bilan d'avancement et tirait les premiers enseignements opérationnels de ce premier volet d'étude.

Le présent rapport final **reprend l'exploitation de cette première démarche centrée sur l'effet inter-opérateurs** et aborde ensuite les deux **aspects complémentaires** suivants :

Volet 2 : Etude de la **variabilité interannuelle** des notes d'indices attribuable à la **variabilité naturelle du milieu**,

Volet 3 : Evaluation d'impact de la variabilité des notes liées à l'effet inter-opérateur sur **l'incertitude d'attribution d'une classe d'état écologique sur un site** à partir du compartiment diatomique, à 2 niveaux différents de résolution temporelle (à l'année, puis par demi-Plan de Gestion des Masses d'Eau).

Au final, les résultats de cette étude considérée dans sa globalité permettent de mieux cerner l'incertitude occasionnée par deux sources principales de variabilités sur les résultats d'une méthode hydrobiologique d'évaluation (l'IBD₂₀₀₇ sur cours d'eau), et de cadrer statistiquement la robustesse globale de l'évaluation d'état écologique ainsi produite dans le cadre d'application de la DCE.

Auteurs :

Wach Marie
Irstea – Centre de Bordeaux
marie.wach@irstea.fr

Guéguen Julie
Irstea – Centre de Bordeaux
julie.quequen@irstea.fr

Dagens Nina
Irstea – Centre de Bordeaux
nina.dagens@irstea.fr

Delmas François
Irstea – Centre de Bordeaux
francois.delmas@irstea.fr

Rosebery Juliette
Irstea – Centre de Bordeaux
juliette.rosebery@irstea.fr

Vérification du document :

Sophie Lardy-Fontan
LNE
sophie.lardy-fontan@lne.fr

Les correspondants :

Onema : Yorick Reyjol, yorick.reyjol@onema.fr

Etablissement Irstea : Christian Chauvin, christian.chauvin@irstea.fr

Référence du document :

Marie Wach & Julie Guéguen, Nina Dagens, Juliette Rosebery & François Delmas, – Rapport final AQUAREF 2017 – Action B3 « Définition des sources d’incertitude du protocole diatomées en cours d’eau » – 109 p

Droits d’usage :	Document en accès libre
Couverture géographique :	National
Niveau géographique :	National
Niveau de lecture :	Professionnels, experts
Nature de la ressource :	Document

Table des matières

Introduction.....	5
1. Eléments de contexte - Genèse de cette étude.....	7
2. Variabilité liée à l'opérateur.....	13
2.1. Etat des connaissances, objectif	13
2.2. Données disponibles	15
2.3. Sélection des données	18
2.4. Traitements statistiques	21
2.5. Résultats.....	21
2.5.1. Aspects qualitatifs : occurrence des taxons.....	22
2.5.2. Aspects quantitatifs : abondances moyennes des taxons	27
2.5.3. Effet du type d'opérateur sur la richesse spécifique	32
2.5.4. Discussion d'étape sur les résultats des 3 types d'analyses relatives à la diversité taxonomique des assemblages	33
2.6. Notes d'IBD ₂₀₀₇	38
2.7. Etat écologique	43
2.8. Discussion - conclusion du volet inter-opérateur	49
2.8.1. Aspects qualitatifs :	50
2.8.2. Réalisation et rapportage des opérations de terrain :	54
2.8.3. Aspects quantitatifs (inventaires, évaluation) :	55
3. Variabilité interannuelle	Erreur ! Signet non défini.
3.1. Problématique.....	56
3.2. Assise de données, méthodologie	57
3.3. Etude des résultats de variabilité interannuelle, discussion.....	68
3.3.1. Classes de qualités des relevés du jeu de données sélectionné.....	68
3.3.2. Analyse des écarts 2 à 2	69
3.3.3. Analyse des écarts de classification d'état au relevé.....	74
3.3.4. Analyse des écarts à la moyenne au site	78
3.4. Conclusions de l'étude de variabilité temporelle de l'IBD et mise en comparaison avec les résultats de variabilité inter-opérateur	80
4 Calcul de la précision et de la confiance de l'évaluation faite à partir de l'IBD₂₀₀₇	85
4.1. Eléments de contexte sur l'évaluation.....	85
4.2. Etat de l'Art et approches méthodologiques.....	86
4.2.1. Références bibliographiques.....	86
4.2.2. Méthodologie publiés et terminologie associée.....	87
4.2.3. Relation entre écart-type et moyenne.....	89
4.2.4. Loi de probabilité d'Etat en fonction de la valeur d'EQR	91
4.2.5. Confiance d'appartenance aux classes de qualité	92
4.3. Résultats de l'approche menée dans le cas de la France.....	93
4.3.1. Réflexions préalables sur la manière d'aborder la confiance de classement d'état	93

4.3.2. Relation entre écart-type et moyenne.....	96
4.3.3. Loi de probabilité et calcul de la confiance des classes	98
4.3.4. Exemples illustratifs	99
4.4. Propagation de la méthodologie à l'évaluation trisannuelle.....	100
5. CONCLUSIONS.....	101

Table des illustrations

Figure 1 : Diagramme schématique des sources de variabilité dans les protocoles de suivis hydrobiologiques (cf. REF. 5).....	9
Figure 2 : Carte des sites diatomées 2013-2014 : nombre de relevés disponibles pour les analyses statistiques sur les 61 sites d'étude	16
Figure 3 : Classement d'état écologique antérieur des sites sélectionnés pour les prélèvements diatomiques inter-opérateurs	17
Figure 4 : Récapitulatif des écarts temporels de prélèvements pour les 2 campagnes 2013 et 2014	18
Figure 5 : Processus de sélection des données pour l'analyse.....	19
Figure 6 : Taxons communs entre les opérateurs sur l'ensemble des relevés (nombres de taxons en maigre, pourcentages en gras).....	23
Figure 7 : Distribution des pourcentages de taxons totaux (à gauche) ou de taxons indicateurs de l'IBD 2007 (à droite) en commun entre les 3 opérateurs, exprimés en occurrence.	25
Figure 8 : Taxons les plus abondants communs entre les opérateurs (zone1) ou trouvés exclusivement par chaque opérateur (zones 2, 3, 4)	26
Figure 9 : Abondances relatives des taxons indicateurs communs entre les opérateurs (%).....	28
Figure 10 : Nombre d'observations de chaque taxon par tous les opérateurs simultanément	29
Figure 11 : Variabilité de la richesse spécifique selon les opérateurs (barre centrale des box-plots = valeur médiane).....	32
Figure 12 : Distribution des assemblages diatomiques du jeu de données d'étude : a) en occurrence des taxons ;b) en abondance relative moyenne de chaque taxon	36
Figure 13 : Variabilité des notes d'IBD2007 selon les opérateurs.....	39
Figure 14 : Distance de Bray-Curtis sur la matrice spécifique en fonction des écarts de notes d'IBD2007	40
Figure 15 : Valeurs absolues des écarts 2 à 2 obtenus entre opérateurs sur les notes d'IBD2007 pour chaque site d'étude.....	41
Figure 16 : Valeur absolue des écarts à la moyenne des notes d'IBD par relevé au même site.....	43
Figure 17 : Comparaison des classes d'état écologique obtenues sur les 61 sites d'étude (soit 63 inter-comparaisons d'état).....	44
Figure 18 : Cartes d'état écologique par opérateur	46
Figure 19 : Valeur absolue des écarts à la moyenne des notes IBD ₂₀₀₇ par classe d'état.....	48
Figure 20 : Exemple de signalétique de station à réactualiser (code de site anonymé) Erreur ! Signet non défini.	
Figure 21 : Nombre de relevés diatomiques disponibles par année dans toute la base Pandore.....	59
Figure 22 : Analyse en Composante Principale (ACP) de la chimie associée aux relevés diatomiques	61
Figure 23 : Boxplot des coefficients de variations de chaque site, sur les 3 plans pour chaque site	64
Figure 24 : Influence de la valeur limite du coefficient de variation sur le nombre de sites sélectionnés.....	65
Figure 25 : Nombre de suivis par site à hydrochimie stable	66
Figure 26 : Projection cartographique des 103 sites "stables" retenus pour l'étude de la variabilité interannuelle des notes d'IBD ₂₀₀₇	67
Figure 27 : Effectif et pourcentage de relevés par classe d'état écologique (TBE : Très bon état, BE : Bon état, EM : Etat moyen, em : état médiocre, ME : Mauvais état)	68
Figure 28 : Distribution des écarts au même site, années prises 2 à 2, des notes d'IBD2007	70
Figure 29 : Nombre de sites suivant le nombre d'écart de classe d'état écologique	74
Figure 30 : Variabilité naturelle interannuelle de l'IBD ₂₀₀₇ : Statistique des cas rencontrés par écart de classes d'état	76
Figure 31 : Variabilité interannuelle naturelle de l'IBD2007 : Valeur absolue des écarts à la note moyenne obtenue au site (sur 103 sites)	78
Figure 32 : Courbe de densité des écarts 2 à 2 des EQR, pour la variabilité inter-opérateurs et la variabilité interannuelle.	81
Figure 33 : Comparaison des box-plots de distribution des écarts 2 à 2 en EQR pour la variabilité	82

Figure 34 : Probabilité d'appartenance à une classe de qualité	85
Figure 35 : Diagramme schématique illustratif de la base de calcul de la confiance du classement	87
Figure 36 : Variabilité intra-site : mise en relation de l'écart-type de la variation au site par rapport à la moyenne des EQR au site. (Kelly et al., 2009).	89
Figure 37 : Confiance dans la prédiction de classe correcte de qualité pour les cours d'eau du Royaume-Uni en fonction de l'EQR observé (d'après Kelly et al, 2009, Figure 3)	92
Figure 38 : Confiance dans la prédiction de classe correcte de qualité : Risque correspondant de mauvaise classification pour des rivières du Royaume-Uni en fonction du nombre de répliqués	92
Figure 39 : Relation entre l'écart-type et les moyennes des EQR des sites et la fonction polynomiale associée (courbe en bleue) pour la variabilité inter-opérateurs.....	96
Figure 40 : Loi de probabilité des EQR dans le cadre de la variabilité inter-opérateurs.....	98
Figure 41 : Probabilité d'appartenance à une classe écologique pour 3 sites différents	99

Introduction

L'objectif général de la fiche-action Aquaref B est l'évaluation de la qualité des données obtenues à l'aide des protocoles hydrobiologiques mis en œuvre sur les réseaux de surveillance des cours d'eau français, dans le cadre d'application de la DCE

Dans ce cadre, un programme spécifique (Action B1-1) a été élaboré en 2013 pour étudier les notions d'incertitudes liées à l'application de la méthode « Diatomées en cours d'eau », et plus précisément les sources de variabilité qui impactent la note IBD₂₀₀₇ (Indice Biologique Diatomées) et l'évaluation de l'état écologique d'un site qui en résulte, à partir du maillon diatomique.

Pour rappel, ce programme a permis dans un premier temps d'étudier la **variabilité liée à l'opérateur**, à partir de données de campagnes de terrain organisées en 2013 et 2014, au cours desquelles une soixantaine de stations ont été prospectées par 3 opérateurs différents (Irstea, Dreal, Bureaux d'études) dans le créneau temporel le plus proche possible compte-tenu des contraintes d'organisation (souvent la même semaine). Ce premier volet d'étude consistait en une étude de la variabilité inter-opérateurs de la donnée issue de relevés diatomiques, intégrant de façon globale les différentes étapes du processus d'obtention de cette donnée : le relevé de terrain, le travail en laboratoire de préparation et de détermination de l'échantillon, l'obtention de la liste taxinomique et le calcul de l'indicateur. Un rapport d'avancement publié fin 2015 par Irstea faisait un point d'étape de cette étude inter-opérateur et formulait les premières conclusions qui ressortaient de ce travail.

Le présent rapport final récapitule et prolonge l'exploitation de cette première démarche, et aborde ensuite deux aspects complémentaires qu'il était prévu d'étudier pour mieux cerner l'impact de l'incertitude de cette méthode biologique sur l'évaluation d'Etat Ecologique :

- **Volet 2** : Etude de la variabilité interannuelle des notes d'indices attribuable à la **variabilité naturelle du milieu et des flores**. Le présent rapport développe donc la méthode utilisée pour la sélection et l'analyse des données à disposition dans le but de satisfaire cet objectif, ainsi que les principaux résultats obtenus. Ces résultats sont ensuite comparés et mis en perspective par rapport aux résultats antérieurs issus de l'analyse de la variabilité inter-opérateurs de la note d'IBD₂₀₀₇ (cf rapport d'avancement 2015).

- **Volet 3** : Exploitation de la variabilité des notes liées à l'effet inter-opérateur pour cerner et quantifier **l'incertitude d'attribution d'une classe d'état écologique à partir du compartiment diatomique** sur un site de réseau de surveillance classique. Deux cas de figure ont été envisagés successivement : l'état évalué à partir d'un seul relevé diatomique collecté chaque année ou, selon la méthodologie d'évaluation actuellement en vigueur pour les sites du Réseau de Contrôle et de Surveillance faisant l'objet d'un suivi pluri-annuel (Arrêté Evaluation du 27-07-2015), sur l'agrégation temporelle du résultat de 3 relevés annuels collectés par demi-Plan de Gestion des Masses d'Eau.

Dans le contexte général d'application de la DCE à notre niveau national, ce programme apporte les éléments permettant au final de mieux cerner la robustesse de l'évaluation de l'état écologique des cours d'eau réalisée à partir du maillon diatomique, dans le cadre du fonctionnement classique des réseaux de surveillance français.

1. Eléments de contexte - Genèse de cette étude

La DCE (JOCE, 2001) (**REF. 1**) demande aux Etats membres d'intégrer dès que possible le concept de confiance et de précision dans l'évaluation de l'état écologique de leurs masses d'eau : « Estimates of the level of confidence and precision of the results provided by the monitoring programmes shall be given in the plan » (DCE, Annexe V, chap. 1.3).

L'application des méthodes d'évaluation comprend toute une chaîne méthodologique couvrant de l'acquisition des données brutes sur le terrain jusqu'au choix des règles d'agrégation des indicateurs permettant l'attribution d'une classe d'état écologique à chaque masse d'eau, avec de nombreuses étapes intermédiaires (préparation, reconnaissance taxinomique des espèces, comptage de la composition des communautés etc...). Or chaque étape intervenant dans le processus d'évaluation mis en œuvre par un pays-membre comporte sa part d'incertitude, qui peut être inhérente à la mise en œuvre des protocoles, intrinsèque à l'outil biologique lui-même ou liée à la démarche d'évaluation de l'état écologique à partir de résultats unitaires de la méthode indicielle.

Compte-tenu de ce qui précède, l'UE demande qu'une estimation des incertitudes en hydrobiologie soit intégrée dans les règles nationales d'évaluation utilisées pour l'état des lieux, le rapportage et le suivi de l'état écologique, le but étant de procurer une évaluation assortie de son intervalle de confiance.

Au cours des deux premiers cycles de gestion des Masses d'Eau, soit avant la sortie du dernier Arrêté national Evaluation du 27-07-2015, la palette d'outils utilisables pour la surveillance et l'évaluation des différents types d'hydrosystèmes de notre pays était encore très incomplète. De ce fait, le travail de mise au point de nouveaux bioindicateurs venant compléter les lacunes du dispositif existant s'imposait comme le chantier le plus prioritaire du moment.

De la même façon, l'étude scientifique de l'incertitude des méthodes hydrobiologiques était à l'époque jugé moins d'actualité que d'autres chantiers inter-Etats-membres nécessitant une organisation collective sous la pression de l'Union Européenne, comme la mise en place de méthodes hydrobiologiques manquantes ou les exercices d'intercalibration des méthodes nationales d'évaluation biologique existantes.

Du fait qu'1/6^{ème} seulement des masses d'eau « cours d'eau » héberge un site de réseau de surveillance, beaucoup de cours d'eau de France ne font l'objet d'aucun suivi direct. De ce fait, il a tout d'abord été nécessaire d'envisager de façon globale l'incertitude attachée à l'évaluation d'état des Masses d'Eau, à l'aide d'un système simple d'appréciation du niveau de confiance attribuable à chaque étape du processus de classification. Ce système s'appuyait sur un arbre de décision régulièrement actualisé, publié en annexe des versions successives de Circulaires et Arrêtés d'Evaluation (Circulaire Evaluation 2005, puis Arrêté Evaluation 2010 réactualisé en Arrêté Evaluation du 27 Juillet 2015, **REF. 2**) ayant force d'application pour la mise en œuvre nationale de la DCE.

L'objectif de cet arbre de décision, basé en grande partie sur des règles de décision s'appuyant sur le peu d'information disponible (souvent extrapolée d'autres hydrosystèmes faisant l'objet d'un suivi et estimés comparables), sur de l'expertise et sur quelques règles de bon sens, est d'évaluer la probabilité d'état écologique et les risques de non-atteinte du Bon Etat Ecologique (RNABE) de chaque masse d'eau de ce type au titre du Plan de Gestion en cours.

L'Arrêté Evaluation actuellement en vigueur va à nouveau faire l'objet d'actualisations en vue d'application réglementaire pour la 2^{ème} moitié du PGME 2016-2021 en cours (publication au J.O. prévue courant 2018).

Ces dernières années, de nouveaux outils biologiques proposés pour l'évaluation des hydrosystèmes sont venus combler les principales lacunes antérieures du dispositif national, certains intégrant d'ailleurs d'emblée une évaluation d'incertitude ou un degré de confiance dans les résultats produits. C'est notamment le cas des indices biologiques multimétriques nouvellement mis au point pour l'évaluation de l'état écologique des cours d'eau, comme l'I2M2 (eg **REF. 3** : *Mondy et al, 2012*) sur le maillon des invertébrés benthiques des cours d'eau, ou comme l'IPR+ (**REF. 4** : *Marzin et al, 2015*) sur les cortèges de poissons des cours d'eau.

Dans ce contexte, et compte-tenu du fait que d'autres urgences importantes pour la mise en application de la DCE étaient progressivement résolues, **l'étude des notions d'incertitude** appliquée aux **méthodes d'évaluation biologique pré-existantes** et largement employées sur les réseaux de surveillance français, ainsi que la **répercussion de leur l'impact sur l'évaluation d'état écologique** requise par la DCE, sont devenues un chantier national tout-à-fait prioritaire.

Un travail d'état de l'art, de bilan d'étape, de réflexion et de proposition méthodologique permettant de cadrer et d'orienter ce chantier de travail a donc été entrepris en 2012 sous l'égide d'Aquaref, afin de faire progresser la pratique nationale vers une plus grande prise en compte des incertitudes et du degré de confiance attribué aux évaluations hydrobiologiques.

Ce travail a pu être mené en s'appuyant sur le réseau partenarial inter-organismes de ce consortium. Il a tiré bénéfice d'une certaine antériorité de mise en place de ce genre de démarches dans le monde de la surveillance chimique des milieux et, dans le cadre collaboratif du groupe de travail national « Incertitude », il s'est enrichi des réflexions menées par les scientifiques des différents organismes partenaires.

Le rapport qui en est issu, intitulé : « L'incertitude en hydrobiologie : concepts et applications », est actuellement en fin de rédaction (*Chauvin et al, Comm. Pers.*). Il rappelle la demande faite aux Etats-Membres, dans le cadre d'application de la DCE, d'intégrer le concept de confiance et précision dans l'évaluation de l'état écologique des masses d'eau. Il établit ensuite un état de l'art des actions antérieures réalisées dans ce sens sur les différents maillons hydrobiologiques des différents types d'hydrosystèmes à surveiller. En se basant sur les échanges des experts inter-organismes appartenant au consortium Aquaref, il propose un schéma théorique des différentes sources d'incertitude pouvant intervenir au long des protocoles de suivis hydrobiologiques et venir impacter les évaluations d'état écologique qui en découlent.

Le schéma conceptuel qu'il contient (cf Figure 1 en page suivante), déjà présenté en communication orale à l'occasion de plusieurs colloques (e.g. **REF. 5** : *Reyjol & Chauvin, 2012*) présente l'analyse théorique des différentes composantes de l'incertitude pouvant affecter *in fine* les résultats d'une méthode d'évaluation hydrobiologique.

Quatre parties y sont distinguées : la variabilité du milieu évalué et des biocénoses, le protocole d'acquisition des données hydrobiologiques, le calcul des indicateurs biologiques, et la classification de l'état écologique. Les sources d'incertitudes sont détaillées au sein de chaque partie.

Ces 4 grandes sources de variabilités ont été listées et décrites dans le détail dans le rapport précité, il est possible de s'y référer pour plus d'information.

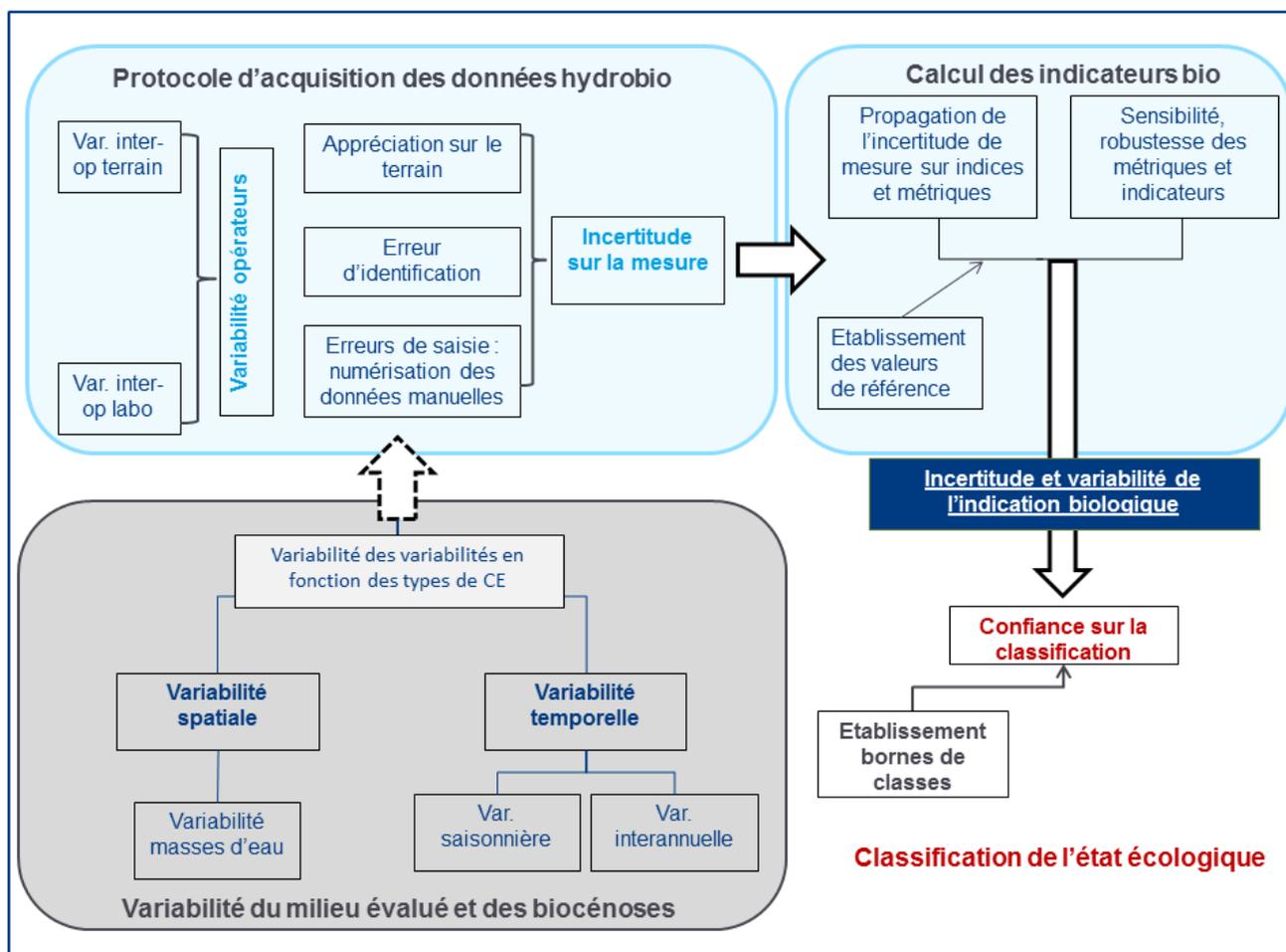


Figure 1 : Diagramme schématique des sources de variabilité dans les protocoles de suivis hydrobiologiques (cf. REF. 5)

Enfin, afin de mieux se conformer aux prescriptions de la DCE, qui demande d'intégrer et de prendre en compte les notions de confiance et de précision dans l'évaluation de l'état écologique des masses d'eau, ce rapport recommandait la mise en place dès 2013 de certaines études prioritaires visant à rajouter une évaluation d'incertitude pour venir conforter l'évaluation d'état écologique actuellement réalisée sur les hydrosystèmes français à partir de méthodes hydrobiologiques pré-existantes et largement diffusées dans le cadre des réseaux de surveillance.

C'est dans ce cadre qu'il a été décidé de lancer, à partir de l'année 2013, un programme de travail sur les incertitudes liées à l'application du protocole d'acquisition de la donnée hydrobiologique pour deux méthodes hydrobiologiques végétales servant à évaluer l'état écologique des cours d'eau de France dans le cadre des réseaux de surveillance : la méthode « diatomées benthiques » basée sur l'IBD₂₀₀₇, et la méthode « macrophytes cours d'eau », s'appuyant sur l'indice macrophytique IBMR.

La présente étude, qui entre dans le cadre programmatique de la **Fiche-Action Aquaref B**, correspond au **programme d'étude pluriannuel B1-1 « Incertitudes Diatomées - Intégration dans les REEE » (2013-2016)**. Ce programme, qui vise à évaluer les incertitudes de la méthode d'évaluation de la qualité des cours d'eau à partir du maillon des diatomées benthiques, a débuté concrètement à partir de 2013.

Il s'est organisé selon les 3 volets suivants :

- 1) Un volet d'étude de la **variabilité inter-opérateurs globale**, construit à partir de l'organisation de 2 campagnes de terrain 2013 et 2014 au cours desquelles une soixantaine de stations ont été prospectées, autant que possible dans le même créneau temporel, par 3 types d'opérateurs différents (Dreal, Bureaux d'études, Irstea). Les opérations de détermination-comptage des échantillons biologiques ainsi collectés ont ensuite été réalisées, le plus souvent durant les périodes hivernales. L'année 2015 a permis de rapatrier la totalité des données de ces 2 campagnes d'étude de la variabilité inter-opérateurs et de réaliser un premier traitement statistique de celles-ci, qui a fait l'objet de la fourniture d'un rapport d'étape fin 2015.
- 2) L'étude de la **variabilité interannuelle de la méthode IBD₂₀₀₇** (réalisation 2016), pour laquelle il n'a pas été prévu de collecte spécifique de données. Ce volet d'étude vise à déterminer la variabilité des résultats de cette méthode diatomique sous l'influence de la variabilité naturelle interannuelle de sites aux conditions hydrochimiques jugées suffisamment stables. Cette variabilité de la réponse biologique et sa traduction en variation de note de l'outil indiciel ou de l'EQR qui en découle influe, au même titre que l'effet inter-opérateur, sur la précision de l'évaluation de qualité biologique et sur l'intervalle de confiance qui doit lui être attribué. Malgré plusieurs difficultés inhérentes à ce type d'exercice, ce volet d'étude a été réalisé en s'appuyant sur des données historiques d'inventaires diatomiques stockées dans la base de données Pandore, dont la plupart a été collectée dans le cadre des réseaux de surveillance nationaux.
- 3) L'évaluation de l'impact de la variabilité des notes due à l'effet **inter-opérateur** sur **l'incertitude d'attribution d'une classe d'état écologique sur un site à partir du compartiment diatomique** (réalisation entre 2016 et 2017). Dans ce cadre, 2 aspects de l'évaluation ont été envisagés successivement : a) l'état évalué à partir d'**un seul relevé diatomique** collecté annuellement à un site donné de réseau ; b) l'état diatomique moyen calculé à partir de **l'agrégation temporelle du résultat de 3 relevés annuels** collectés sur un même site du RCS, selon la méthodologie officielle d'évaluation actuellement en vigueur (*cf* Arrêté Evaluation du 27-07-2015), qui permet d'évaluer l'état de ce site sur un demi-Plan de Gestion des Masses d'Eau.

Des études antérieures (*e.g.* **REF. 6 : Roussel et al, 2011**) avaient clairement montré, par rapport aux nombreuses origines possibles d'incertitude, l'aspect dominant de la variabilité inter-opérateur sur la variabilité du résultat de la méthode diatomique.

Indirectement, les trois volets pré-listés de la présente étude sont légitimés par (et représentatifs de) la pratique des réseaux de surveillance, qui peut être confiée à des opérateurs économiques différents selon la portée temporelle des marchés de surveillance et le hasard des lots remportés une année donnée par un intervenant donné.

Il existe donc un assez fort effet aléatoire, au gré de ces attributions de Marchés Publics, sur quel intervenant économique aura à les exécuter et même, au sein d'une entreprise donnée, sur quel intervenant taxonomiste aura effectivement en charge les déterminations.

Le Chapitre 2 du **présent rapport** ré-intègre et approfondit l'exploitation de la première démarche focalisée sur **l'effet inter-opérateurs**, réalisée à partir des données collectées pendant les 2 campagnes de terrain 2013 et 2014 et qui a déjà fait l'objet d'un livrable d'étape Aquaref fin 2015.

Ce rapport final aborde ensuite successivement deux autres aspects venant compléter l'évaluation de l'effet des incertitudes de la méthode diatomique IBD 2007 sur l'évaluation d'état écologique, à savoir :

- l'étude de la **variabilité interannuelle des notes d'indices** attribuable à la **variabilité naturelle du milieu**,
- l'évaluation de l'impact de la variabilité des notes liées à **l'effet inter-opérateur** sur **l'incertitude d'attribution d'une classe d'état écologique** pour un site faisant l'objet d'une **évaluation à partir du compartiment diatomique**. Deux cas de figure ont été considérés : **1)** l'état évalué à partir d'un seul relevé diatomique collecté annuellement, ou **2)** après application de la méthodologie d'évaluation actuellement en vigueur pour les sites du Réseau de Contrôle et de Surveillance suivis en pluri-annuel (Arrêté Evaluation du 27-07-2015), l'état évalué sur l'agrégation temporelle du résultat de 3 relevés annuels collectés par demi-Plan de Gestion des Masses d'Eau.

Au bilan, les résultats de cette étude considérée dans sa globalité permettent de mieux cerner l'incertitude occasionnée par deux grandes catégories de variabilités sur les résultats d'une méthode hydrobiologique d'évaluation (l'IBD₂₀₀₇ sur cours d'eau) et de cadrer statistiquement la robustesse globale de l'évaluation de l'état écologique déterminée au sein des réseaux de surveillance nationaux, dans le cadre d'application de la DCE.

2. Variabilité liée à l'opérateur

2.1. Etat des connaissances, objectif

Une étude précédente basée sur des données d'exercices d'intercomparaison de DREAL, acquises à l'occasion de sessions annuelles de formation continue organisées par un taxonomiste référent de niveau international, avait permis à *Roussel et al., 2011 (REF. 6)* de réaliser une première approche de la variabilité inter- et intra-opérateur de la note d'IBD2007.

Cependant, dans l'optique de l'appui à la mise en application de la DCE, ces données étaient trop restreintes sur leur assise spatiale et temporelle pour être complètement conclusives et transférables à la large échelle du territoire Français.

D'autre part, elles étaient systématiquement issues d'intervenants spécialistes-maillon de Services Extérieurs de l'Etat (DREAL), donc pas forcément représentatives de la variabilité de données acquises sur les réseaux de surveillance de routine du type RCS. En effet, dans le cadre de ce réseau, la collecte et la détermination sont prises en charge de façon très dominante par des intervenants privés de type Bureaux d'Etude.

Enfin, un autre frein à la généralisation de ces résultats était qu'en fonction de l'objectif principal de formation continue axé sur la reconnaissance taxonomique, les sites d'étude retenus étaient le plus souvent des sites de référence ou de très bonne qualité intéressants par leur **flore originale et très spécifique**, mais dont les inventaires produisaient presque exclusivement de hautes notes d'indice. Or, dans l'optique d'attribuer un intervalle de confiance à un résultat obtenu sur des réseaux classiques de surveillance dans le cadre d'application de la DCE, il est certainement plus représentatif de traiter des incertitudes attachées à l'usage d'un outil hydrobiologique comme l'IBD₂₀₀₇ sur un domaine d'application diversifié couvrant tout le gradient de qualité possible, depuis des sites de référence jusqu'à des situations parmi les plus altérées qu'il est possible de trouver dans ces réseaux de surveillance.

Quoi qu'il en soit, cette première étude avait produit des résultats quantifiés qu'il est important de garder en tête et de mettre en perspective par rapport à ceux qui ont été obtenus dans le cadre de la présente étude.

Une **variabilité inter-opérateurs globale de l'IBD 2007 avec un coefficient de variation de ± 1 point avait été estimée** (tests statistiques), cette variabilité finale résultant de la combinaison de différentes origines (échantillonnage, méthode de traitement, montage de la lame, qualité du matériel d'observation, méthode de comptage).

D'autres résultats obtenus sur un petit nombre de situations, mais faisant intervenir une stratification adéquate des protocoles, montraient une **dominance très nette de l'effet inter-opérateur** (plus de 80% de la variabilité de la note) contre une **variabilité quatre fois moindre** (un peu moins de 20 %) **pour l'effet intra-opérateur**. Concernant l'effet intra-opérateur, le **comptage** était l'étape du protocole responsable de la plus grande variabilité, suivi du **prélèvement** (application du protocole d'échantillonnage) et enfin, du protocole de réalisation de la lame permanente à partir d'un échantillon plus important contenu dans un pilulier.

Une autre conclusion forte de cette étude préalable était que la **variabilité liée au protocole d'échantillonnage est apparue importante** et ce, qu'il s'agisse de la dimension inter-opérateur ou même intra-opérateur (intervenant comme la 2^{ème} source d'incertitude dans ce cas, comme indiqué juste au-dessus).

Pour les exercices précédents inter-DREAL, un élément de minimalisation des erreurs et de l'incertitude par rapport à la présente étude tenait au fait que tous les participants à l'exercice étaient présents sur le même site au même moment, ce qui **annulait l'erreur de localisation de la station, ainsi que l'influence de la variabilité temporelle**. Par ailleurs, dans toutes les études de ce type, les participants aux exercices essaient d'appliquer le plus correctement possible la même norme AFNOR NF 90 T-354, ce qui conduit en principe à limiter les marges d'interprétation et à rendre plus homogène la pratique des différents opérateurs.

Malgré ces facteurs de convergence, le dernier constat sur le degré assez sensible de variabilité du résultat d'échantillonnage vient conforter le fait que, malgré la disparition de deux des sources d'erreur et de variabilité qui existent dans d'autres situations (risque d'écart sur la station et le faciès, intervention d'une variation temporelle) et malgré le fléchage normatif des échantillonnages vers un faciès de prédilection (faciès le plus lotique de la station, si possible sur un radier, sur substrats durs de type blocs, galets ou graviers), une importante variabilité des échantillons composites collectés subsiste, y compris au cours d'échantillonnages pratiquement concomittants. Cela peut résulter d'une part de **raisons naturelles ayant trait à la distribution des communautés**. En effet, les diatomées se sélectionnent et s'assemblent en petits patchs selon les conditions micro-locales de forçage (faciès, substrat, conditions hydrodynamiques etc...), ce qui peut conduire à une part de variation aléatoire de la composition spécifique d'un échantillon composite, au hasard des substrats effectivement prélevés lors de chaque opération unitaire de prélèvement. D'autre part, il est probable que la **façon de procéder de chaque individu pour la réalisation pratique de l'opération d'échantillonnage** puisse jouer aussi un rôle non négligeable sur le contenu de l'échantillon composite produit.

En résumé opérationnel de ces analyses de données préalablement disponibles, terrain et laboratoire confondus, la **variabilité inter-opérateurs** est sensiblement plus importante que la variabilité intra-opérateur. Et, concernant les différentes étapes qui interviennent dans la mise en oeuvre de la méthode, le choix du site et des substrats pour l'échantillonnage, la façon de procéder pour l'échantillonnage mais aussi l'identification taxonomique engendrerait une variabilité assez conséquente. Enfin, la localisation du site d'échantillonnage au sein d'un bassin (ce qui renvoie à la représentativité du choix initial de l'emplacement de la station de réseau par rapport à une masse d'eau donnée), ainsi que l'année d'échantillonnage (ce qui renvoie à la variabilité interannuelle), semblent avoir un impact non négligeable sur l'évaluation du cours d'eau.

Compte-tenu de ces acquis, la première décision de programmation d'action nouvelle sur ce maillon diatomique a concerné la mise en place, dès 2013, d'un volet d'étude visant à déterminer la **variabilité inter-opérateurs** relative à la mise en oeuvre de la méthode IBD 2007 sur diatomées benthiques des cours d'eau.

Les moyens financiers et le temps des intervenants étant limités, **ce volet d'étude n'a pas pu faire l'objet d'un protocole stratifié** visant à isoler de façon détaillée la contribution de chaque étape intervenant dans la mise en oeuvre de cette méthode sur la composition de l'incertitude inter-opérateur.

Cela aurait été beaucoup trop coûteux au regard du budget indiqué et aurait mobilisé trop de temps d'intervenants taxonomistes des 3 types de structures impliquées, alors que cette ressource très sollicitée limite de fait la quantité de travail possible dans ce domaine.

Compte-tenu de ce contexte, il s'est surtout attaché à déterminer l'**incertitude globale inter-opérateur** résultant de l'intervention de 3 catégories différentes d'opérateurs sur l'évaluation d'état écologique à partir de la méthode IBD₂₀₀₇, chaque type d'intervenant prenant complètement en charge toute la chaîne de réalisation couvrant depuis la localisation de la station et la réalisation du prélèvement composite jusqu'à la fourniture du résultat détaillé d'inventaire servant au calcul d'état diatomique, selon la pratique classique d'un réseau de surveillance de type RCS.

Sa réalisation s'est donc appuyée sur la collecte spécifique d'échantillons diatomiques indépendants, collectés sur une soixantaine de stations au cours de 2 campagnes de terrain 2013 et 2014, dans un créneau temporel le plus réduit possible, par 3 types d'opérateurs différents (DREAL, Bureaux d'Etudes, Irstea). Suite à ces réalisations de terrain, l'année 2015 a permis de récupérer la totalité des données de ces campagnes et de réaliser un premier traitement statistique de celles-ci. Ces premières exploitations de données ont permis d'alimenter un rapport d'étape relatif à la variabilité inter-opérateurs, qui a été rendu fin 2015.

Le présent rapport final récapitule les réalisations de ce volet d'étude et en prolonge l'exploitation jusqu'à la fourniture d'éléments opérationnels permettant de cadrer la confiance et la précision adjointe à la classification écologique d'un site, quel que soit son niveau de classement.

2.2. Données disponibles

Au cours des 2 années d'étude, un dispositif de 61 stations différentes a fait l'objet de campagnes de prélèvement de terrain à réaliser par les 3 types d'intervenants dans un délai temporel maîtrisé (cf Figure 2 en page suivante).

Même s'il a pu arriver que 2 intervenants se retrouvent sur un site donné, de façon convenue à l'avance ou fortuite, cela a été très occasionnel. En règle générale, les participants à l'exercice sont intervenus de façon indépendante sur les sites d'étude et chacun d'entre eux a agi à sa main, c'est-à-dire qu'il a lui-même localisé le site et le faciès de prélèvement à l'aide de la signalétique de station fournie, dans le cadre de la programmation temporelle de sa propre mission hebdomadaire. Il est très probablement intervenu un petit nombre d'erreurs ou d'écarts de localisation, notamment en cas de signalétique de station défectueuse ou perfectible, pas forcément traçables avec certitude à chaque fois (lorsque l'erreur de site n'était pas absolument certaine, les résultats n'ont pas été retirés des analyses...), mais qui ont probablement impacté un peu à la hausse la statistique d'écarts de résultats inter-opérateurs.

Des consignes préalables invitaient chaque intervenant, autant que possible, à échantillonner une série de sites convenus au cours d'une semaine donnée. Les relevés de terrain étaient donc faits à l'idéal au cours de la même semaine par le triplet d'intervenants. Cependant, en fonction d'impératifs organisationnels propre à l'organisme lui-même (problèmes d'intervenant terrain disponible ou non, de congés estivaux à positionner dans le cadre d'un collectif...), ou de survenue d'un évènement climatique marquant venant s'interposer (de type orage isolé ou série d'épisodes orageux), il est arrivé que ce délai idéal soit dépassé plus ou moins largement.

Lorsque l'écart était trop important avec au moins un des intervenants, le doublet exploitable a été analysé et, lorsque possible, le site a parfois été re-programmé et ré-échantillonné par les 3 types d'intervenants à une date ultérieure plus convergente.

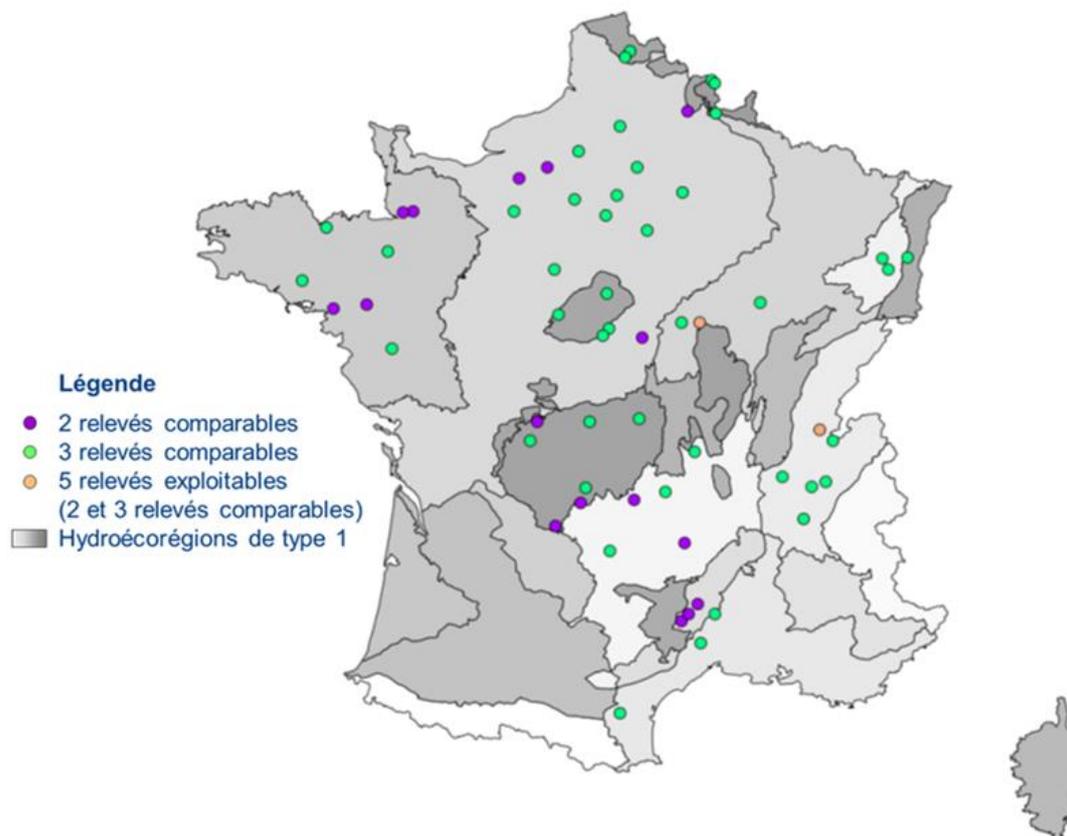


Figure 2 : Carte des sites diatomés 2013-2014 : nombre de relevés disponibles pour les analyses statistiques sur les 61 sites d'étude

L'organisation annuelle des campagnes de prélèvements a été la suivante :

- 27 stations, réparties dans 10 HER1 différentes, échantillonnées en 2013,
- 34 stations, réparties dans 12 HER1 différentes, échantillonnées en 2014.

La répartition spatiale des sites échantillonnés (cf. Figure 2) a pris en compte certaines contraintes d'organisation et notamment des possibilités ou non d'intervention de DREAL, qui ne disposent pas toutes de spécialistes diatomistes au sein de leur laboratoire régional. La couverture du territoire national n'est donc pas complète. Cependant, elle permet la représentation d'un large assortiment de flores régionales associant hydrochimie peu minéralisée et eaux fortement minéralisées, cours d'eau de zones à reliefs (Alpes, Massif Central) et de zones plus planes, appartenant à d'assez nombreuses HER1.

De même, en se basant sur des données pré-existantes d'évaluation écologique (cf. Figure 3 en page suivante), et même si le classement d'évaluation diatomique obtenu sur ces sites à l'issue de la présente étude est apparu plutôt moins sévère au final que l'état historique présenté sur cette figure, il a été apporté un soin particulier faire porter l'étude inter-opérateurs sur un assortiment de sites présentant des **classements écologiques très variés** selon le référentiel antérieurement disponible.

Cette pratique, décidée dans le cadre du collectif des intervenants concernés par l'étude, a été mise en œuvre afin de ne pas donner des résultats trop conditionnés par une catégorie particulière de qualité des milieux mais au contraire, pour distribuer l'incertitude d'évaluation de façon la plus représentative possible sur toute la gamme d'états écologiques qu'il est possible de rencontrer en France.

On peut ainsi voir que le réseau d'étude choisi représentait autant que possible toutes les classes de qualité hydrobiologique, évaluées à partir du compartiment diatomique, qu'il est possible de rencontrer sur le territoire Français.

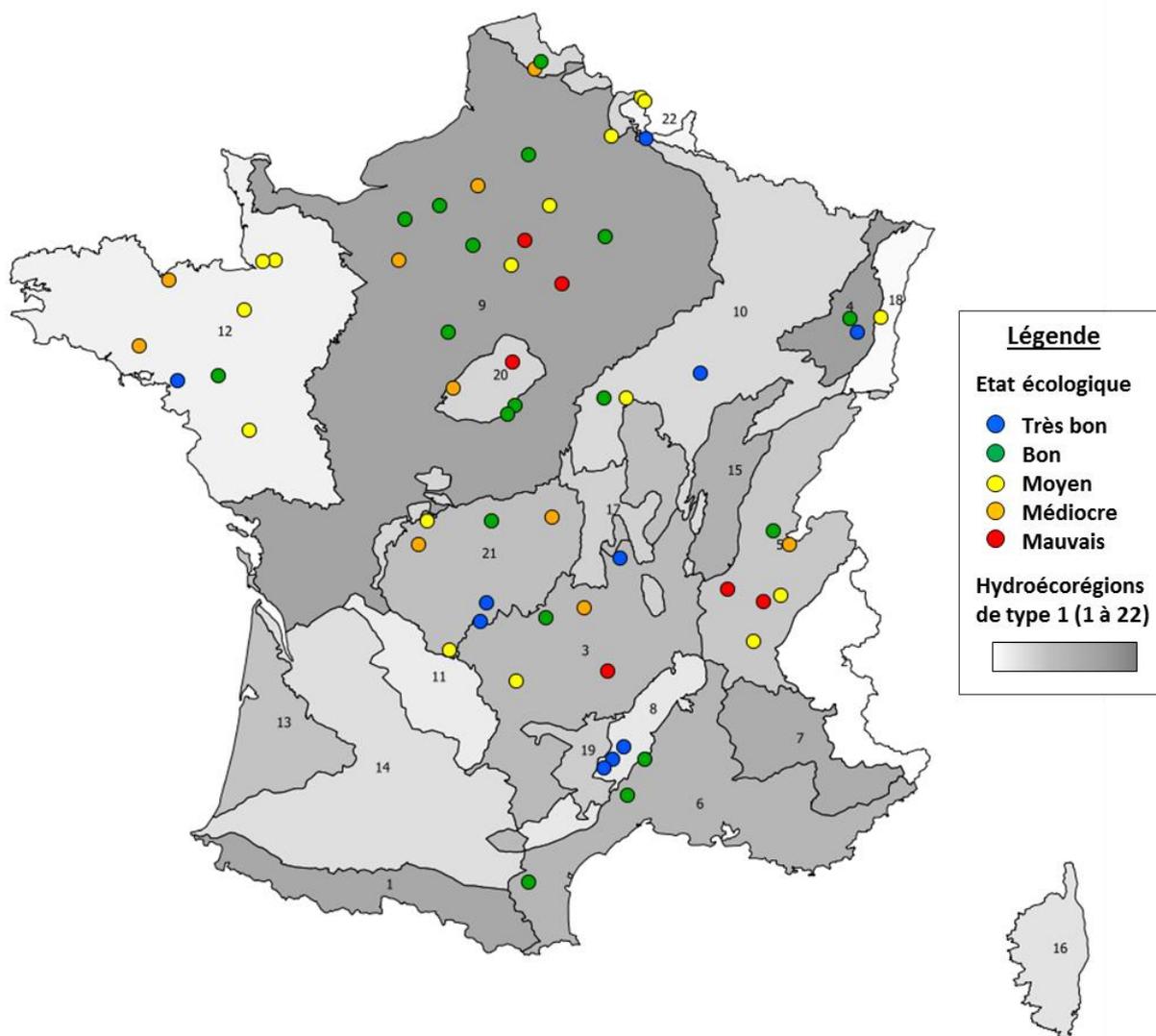


Figure 3 : Classement d'état écologique antérieur des sites sélectionnés pour les prélèvements diatomiques inter-opérateurs

La Figure 4 en page suivante fait l'état de l'écart temporel existant entre les prélèvements diatomiques inter-opérateurs pour la campagne 2013 (en bleu) et pour la campagne 2014 (en rouge).

L'écart temporel-cible visé dans cette étude était de moins de la semaine afin de pouvoir considérer cet effet comme négligeable et de tenter d'isoler le mieux possible l'effet inter-opérateur. On peut constater que, pour différentes raisons, ce créneau temporel-cible a pu être dépassé, parfois largement, risquant de générer une variabilité additionnelle indésirable.

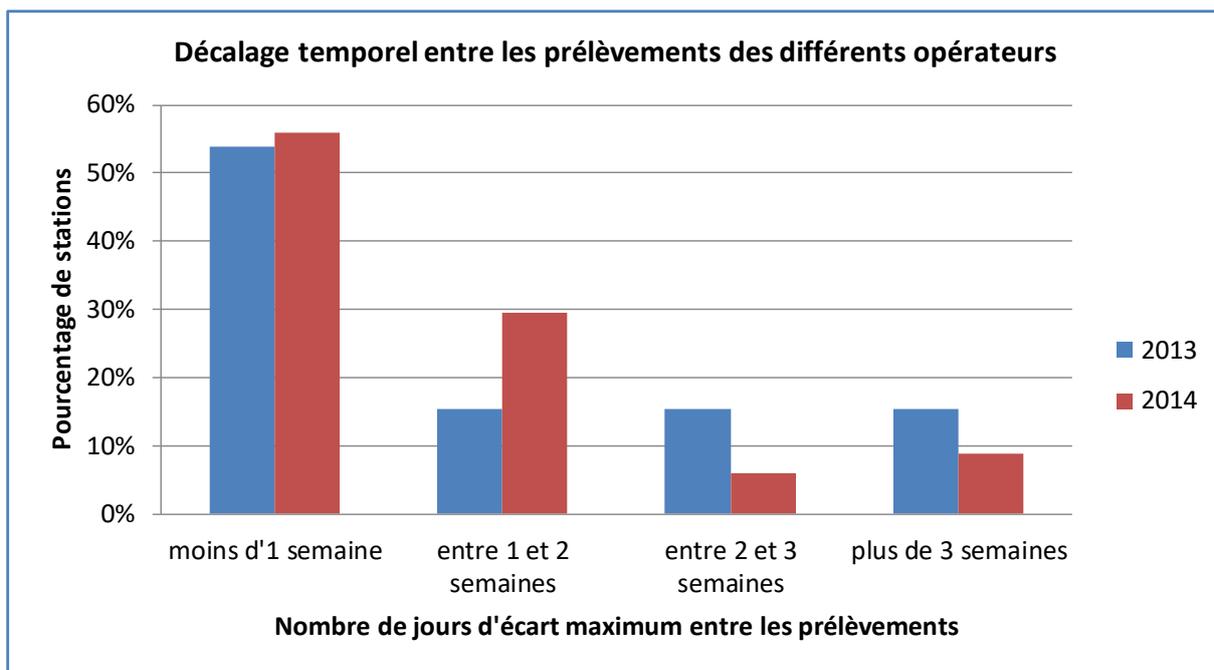


Figure 4 : Récapitulatif des écarts temporels de prélèvements pour les 2 campagnes 2013 et 2014

Les raisons principales de ces écarts ont pu tenir à un agenda de missions terrain planifié longtemps à l'avance et peu adaptable, en particulier pour les intervenants de type « Bureau d'Etude ».

Il a aussi pu se poser, dans une structure donnée, un problème d'indisponibilité temporaire d'un type d'intervenant pour raison de congés estivaux, sans possibilité de substitution par un collègue la même semaine.

D'autres fois, des causes parfaitement naturelles ont pu perturber le déroulement normal des actions programmées. Par exemple, l'intervention d'un épisode orageux provoquant une crue alors que les 3 opérateurs n'étaient pas encore tous passés, a pu générer un délai de carence normatif d'au moins 15 jours (cas d'un bouleversement physique modéré), voire plus long (sur constat de remaniement de substrats durs).

Au vu des écarts importants assez souvent constatés en première année (30% des situations dépassaient les 2 semaines d'écart), un effort plus particulier d'adaptation de programmation visant à une meilleure convergence de dates a été effectué en 2014, réduisant de moitié la statistique des écarts temporels dépassant 2 semaines. Cependant, un petit nombre de cas particuliers s'est encore posé, en liaison avec des impondérables (intervention d'un épisode de pluie d'orage avant que les 3 intervenants ne soient passés, un marché public passé très tardivement, décalant obligatoirement l'intervention du bureau d'études en automne ...).

2.3. Sélection des données

Après réception des données « variabilité inter-opérateurs » complètes dans le courant de l'année 2015 et après expertise des données disponibles, une sélection a été effectuée afin de limiter les biais indésirables.

La Figure 5 ci-dessous récapitule le processus de sélection réalisé et les effectifs de stations représentant les différents cas de figure rencontrés pour la comparaison de données et l'évaluation des variabilités.

Dans ce contexte de référentiel limité (61 stations), mais qui a été raisonné pour être représentatif de la variabilité des intervenants, des conditions naturelles et de l'étendue du gradient d'altération présent dans les réseaux de surveillance de France, il était important de tirer le meilleur parti possible des données collectées.

Ainsi, deux stations incluses dans l'exercice collectif et ayant aussi servi de support à un exercice d'inter-comparaison des DREAL constituaient un cas un peu particulier, du fait qu'elles ont conduit à la disponibilité de 5 relevés diatomiques (dont 3 relevés d'opérateurs DREAL). Les relevés ont été ventilés de la façon indiquée en bas de la Figure 5 ci-dessous pour pouvoir en faire une utilisation optimale dans les inter-comparaisons.

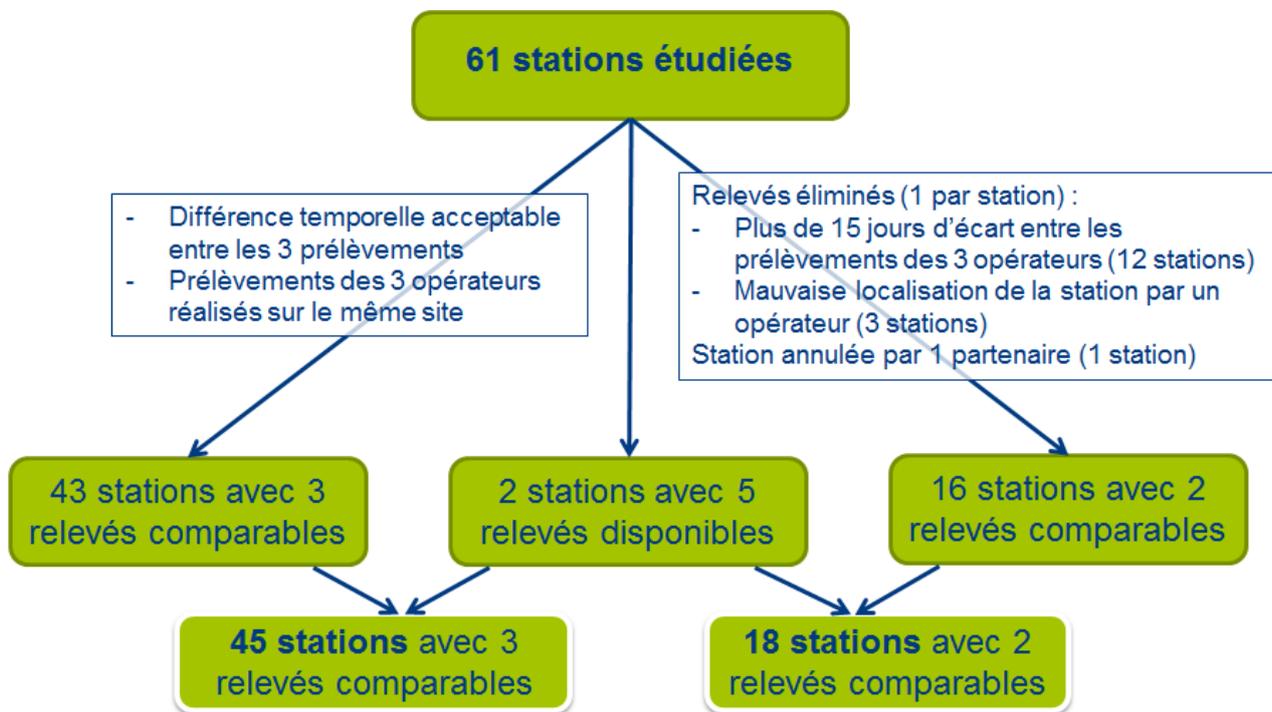


Figure 5 : Processus de sélection des données pour l'analyse

L'élimination de relevés pour raisons de non-comparabilité est le plus souvent intervenue sur la base de critères temporels et/ou climatiques objectifs (trop de temps entre 2 relevés, parfois lié à un problème de passage de marché public, un épisode orageux signalé survenu entre le passage de 2 intervenants et du 3^{ème} etc...).

A noter qu'il a parfois subsisté un doute assez fort sur la prospection de faciès comparables, voire même du même cours d'eau au même site (2 relevés à espèces dominantes et notes d'IBD plutôt comparables, le 3^{ème} intervenant obtenant des résultats très différents à peu près aux mêmes dates).

L'éclaircissement de ce type de situation douteuse aurait nécessité la bonne implémentation systématique, par les 3 intervenants de terrain, des quelques renseignements supplémentaires qu'il était demandé d'associer au relevé, notamment :

- 1) saisie et report des coordonnées géoréférencées effectivement prises sur le terrain,
- 2) utilité des données de physico-chimie associées, non-obligatoires dans un protocole normatif banal de réseau mais très importantes, dans ce type d'étude à enjeu national, comme élément informatif complémentaire susceptible d'appuyer *a posteriori* des recroisements, validations ou éliminations de relevés...).

Sur les triplets de données possibles, dans à peu près 1 cas sur 2, il manquait une partie de ces données d'accompagnement chez un voire 2 intervenants, souvent de façon croisée (un intervenant capturant et reportant les coordonnées géoréférencées de terrain mais pas la physico-chimie, et vice-versa).

De plus, malheureusement, les manques ont factuellement plus concerné les cas où un relevé génèrait un doute que les cas où les 3 relevés présentaient une convergence normale, de sorte que des recoupements fiables 3 à 3 ont rarement été possibles pour confirmer ou infirmer une erreur potentielle suggérée par la composition des prélèvements. En l'absence de critère annexe objectif permettant de spécifier **à coup sûr** une erreur sur un site prospecté, il a été pris le parti de **conserver ces relevés un peu douteux pour l'analyse des écarts**. Certains écarts, assez rares mais d'ampleur exceptionnelle, s'élèvent donc très nettement au-delà de la variabilité habituelle de cette méthode (jusqu'à 6,6 points d'IBD et 2 classes de qualité d'écart entre les 3 évaluations au même site). Ils ont participé à l'établissement de la variabilité inter-opérateurs obtenue dans cette étude, qui s'en trouve statistiquement un peu renforcée, ainsi que l'incertitude associée.

Du fait du manque initial de quelques relevés comparables et suite à la sélection réalisée en fonction des principes évoqués plus haut, **les données comparables entre trois opérateurs au même site**, qui étaient l'objectif-cible du protocole, **sont exploitables pour seulement 45 sites, soit 74 % des inter-comparaisons possibles**, ce qui matérialise un **pourcentage de perte assez conséquent (26 %)** compte-tenu de l'enjeu de ce type d'étude nationale et de la force de travail consacrée en inter-organismes.

Il a donc été pris le parti d'utiliser et d'analyser aussi les variabilités des **doublets de relevés comparables**, collectés sur 18 sites différents (voir partie droite de la Figure 5), dans un souci de valorisation maximale des données collectées dans le cadre de cet exercice collectif.

La base de données finale reste bien équilibrée du point de vue de la répartition spatiale des sites échantillonnés. Toutes les données collectées ont été bancarisées dans une base spécifique liée à la base de données Pandore (liste des sites prospectés en annexe).

Nota :

Même s'il n'y a pas eu de remontée organisée de remarques de type retour-terrain, cette étude a permis de tester l'utilisabilité facile ou non des **signalétiques de stations** mises à disposition des 3 opérateurs de l'étude en vue de trouver un site de réseau non connu à l'avance par certains intervenants. La plupart des fiches pouvaient être qualifiées de très satisfaisantes à satisfaisantes, mais 4 à 6 d'entre elles, soit une faible proportion (moins de 10 %) justifieraient pour le moins une actualisation, voire seraient nettement perfectibles pour éviter les pertes de temps inutiles de la part d'intervenants ne connaissant pas les sites à l'avance, voire jouer un rôle non-négligeable dans une erreur de localisation.

A ce titre, les coordonnées géographiques sont bien sûr nécessaires, mais ne constituent pas un élément de fiche-station suffisant en soi. Outre qu'un petit nombre d'entre elles (2 ou 3) étaient très certainement erronées (impossibilité de tomber sur cette rivière avec les coordonnées indiquées sur la fiche), se pose le problème concret du système référentiel de coordonnées utilisé sur la fiche station d'une part (pas forcément homogène sur toutes les fiches-stations utilisées dans l'étude) et par l'appareil de localisation utilisé par l'opérateur de terrain d'autre part (plusieurs systèmes cohabitent selon les types d'appareils, les fabricants et leur nation d'origine).

En plus de coordonnées géoréférencées, il est hautement appréciable pour l'opérateur de terrain d'avoir des informations sur l'accès le plus pratique à la station, ainsi que sur quelques éléments concrets facilement identifiables dans l'environnement de la station et lui permettant de conforter sa pratique. Les renseignements suivants, notamment, ainsi que des clichés bien choisis et suffisamment actuels, peuvent éviter les pertes de temps inutiles et une erreur sur la localisation précise du faciès prospecté (par quelle berge accéder, par quel nom de voie, l'unité d'observation est-elle située à l'amont ou à l'aval du pont, à quelle distance, etc...), voire sur le site lui-même.

Cet aspect concernant les informations préalables utiles et le besoin d'optimisation de certaines fiches-stations sera repris au chapitre « Discussion » de cette étude inter-opérateur. En effet ce volet, un peu indépendant de la variabilité intrinsèque de la méthode IBD et facilement perfectible dans un petit nombre de cas, peut contribuer sensiblement à limiter ou au contraire, à renforcer bien inutilement la variabilité inter-opérateur et l'hétérogénéité des données collectées dans le cadre des réseaux.

2.4. Traitements statistiques

Plusieurs types d'analyses ont été effectués sur les données sélectionnées afin de décrire la variabilité inter-opérateurs des résultats de la méthodologie indicielle à différents niveaux. Cette variabilité a donc été étudiée sur plusieurs aspects de la mesure, à savoir sur la comparaison du contenu en espèces des inventaires diatomique (liste taxonomique basée sur l'occurrence des taxons, analyse prenant en compte l'abondance relative de chaque taxon, richesse spécifique des assemblages), puis sur les notes et EQRs d'IBD₂₀₀₇, et enfin sur l'évaluation d'état écologique et la probabilité d'appartenance à une classe de qualité.

Dans ces différentes analyses, toutes les données jugées comparables et exploitables ont été prises en compte, qu'elles comprennent deux ou trois relevés comparables par site.

Pour étudier la significativité des différences inter-opérateurs au niveau de la richesse spécifique, des tests paramétriques de Student pour la comparaison de moyennes deux à deux ont été employés ici, au vu de la nature et des caractéristiques des données (normalité et homoscedasticité).

En revanche, les notes IBD₂₀₀₇ ne respectant pas la distribution d'une loi Normale, c'est le test non paramétrique de comparaison de médianes deux à deux de Wilcoxon-Mann-Whitney qui a été utilisé pour vérifier si des différences significatives sont révélées ou non dans l'évaluation inter-opérateurs.

2.5. Résultats

Ce volet d'étude vise en premier lieu à **situer l'incertitude inter-opérateur**, en vue de mieux qualifier les résultats d'évaluation obtenus sur un site faisant partie d'un réseau de surveillance et d'affecter à celui-ci la probabilité d'appartenance à telle ou telle classe de qualité. Cet aspect, traité en fin de rapport, répond assez directement à une requête européenne sollicitant une meilleure spécification de l'évaluation d'état écologique réalisée par les Etats-Membres dans le cadre de la DCE.

Dans une démarche d'évolution qualitative et de progrès pour l'avenir, qui sous-tend la plupart des réalisations entreprises sous la bannière d'Aquaref, ce type d'étude nationale vise aussi à **repérer quelles sont les hypothèses sous-tendant les plus fortes variabilités**, afin d'en identifier les origines et de suggérer les principales voies d'amélioration possibles.

L'interprétation des résultats de cette étude attache donc un poids particulier au **diagnostic des écarts les plus importants rencontrés**, l'objectif étant d'en repérer et d'en identifier les causes les plus probables. Le présent rapport débouche donc aussi sur des recommandations permettant à l'avenir de limiter le plus possible les écarts exagérés par rapport à la gamme de variabilité habituelle attachée à l'application normative de la méthode.

Comme convenu dans le projet initial, pour l'exploitation des résultats de cet exercice inter-opérateur, l'anonymat individuel de chaque intervenant est garanti. Cependant, même si cet aspect est forcément traité de façon globalisée en fonction de ce qui précède, l'analyse des résultats inclut nécessairement, entre autres origines possibles, la source de variabilité liée à la catégorie socio-professionnelle des intervenants, afin de diagnostiquer si certaines des variabilités analysées ont une liaison de cause à effet avec le type d'intervenant impliqué dans l'étude.

Au terme de l'étude, ce rapport doit aussi déboucher sur un train de recommandations opérationnelles à mettre en application pour améliorer la valeur et la fiabilité des référentiels collectés, dans l'optique de resserrer la fourchette de l'état écologique diagnostiqué et remonté au niveau de l'Europe d'une part, et d'améliorer l'efficacité de ces suivis réalisés et des référentiels collectés sur base de crédits publics d'autre part.

Pour les représentations de résultats qui suivront, les regroupements des intervenants par type d'acteurs ont été faits de la façon suivante :

- **Intervenants A** : Opérateurs économiques de type Bureaux d'Etude, compétents sur le maillon diatomées et amenés à intervenir en routine pour la réalisation de marchés publics de surveillance.
- **Intervenants B** : Agents diatomistes spécialistes des Services Extérieurs de l'Etat (DREAL) possédant une bonne expertise taxinomique.
- **Intervenant C** : Agent diatomiste spécialiste appartenant au milieu de la recherche (Irstea), faisant partie du laboratoire ayant généré la méthode indicielle.

Si des exploitations comparatives globales inter-type d'intervenant ont été établies sur le contenu taxinomique quantitatif des relevés, il convient aussi de préciser ici qu'aucun intervenant de cette étude n'a été considéré comme référent plus valide qu'un autre sur le plan taxinomique.

2.5.1. Aspects qualitatifs : occurrence des taxons

Un premier diagramme de Venn (Figure 6 en page suivante), basé sur l'occurrence des espèces, permet de présenter le nombre et le pourcentage de taxons communs entre les trois grands types d'opérateurs sur l'assise complète des triplets de relevés disponible à l'issue de l'étude.

Ce type de figuration indique le nombre et le pourcentage de taxons identifiés en commun entre les 3 types d'opérateurs à l'échelle des 45 triplets de relevés disponibles (zone de recoupement centrale), puis ceux identifiés 2 à 2 entre 2 types d'opérateurs, et enfin le nombre de taxons déterminés par le seul type d'opérateur concerné (partie qui lui est propre, sans intersection avec un autre type d'intervenant). Le nombre total de taxons observés par type d'opérateur sur tout ce jeu de triplets de relevés est porté entre parenthèses en-dessous de l'identifiant opérateur.

Les 18 doublets de relevés comparables n'ont pas été pris en compte dans cette exploitation car ils n'ont pas la même assise ni la même signification statistique que les triplets. En effet, même en cas de forte convergence des déterminations d'espèces entre les 2 opérateurs qui ont examiné les relevés, si le 3^{ème} relevé est inexistant ou était non-comparable pour une raison autre que l'opération de détermination-comptage (voir les raisons de l'absence ou de l'invalidation d'un relevé sur les 3 en partie droite de la Figure 5), l'intersection à 3 opérateurs pour ce doublet serait forcément vide, venant à la fois fausser à la baisse cette statistique inter-opérateurs et faire perdre son sens à son résultat.

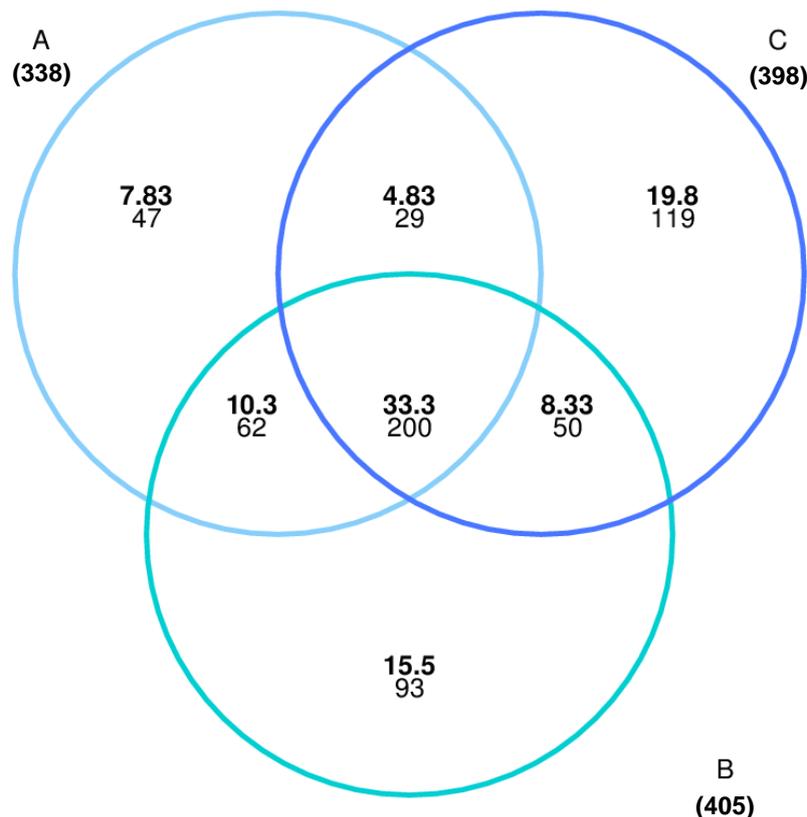


Figure 6 : Taxons communs entre les opérateurs sur l'ensemble des relevés (nombres de taxons en maigre, pourcentages en gras)

Attention : Le diagramme-type utilisé ci-dessus a une **géométrie fixe**. Il permet de définir les espaces communs de recouvrements entre opérateurs individuels et des espaces concernant spécifiquement chaque type d'opérateur, mais il n'est pas géré pour présenter des surfaces d'interface proportionnelles aux valeurs chiffrées. Il ne faut donc pas se baser sur sa physionomie générale et sur les surfaces graphiques qu'il présente, mais bien s'attacher à regarder les valeurs chiffrées figurant dans les domaines. Le diagramme mentionne à la fois les valeurs en pourcentage par rapport au nombre total de taxons (en gras) et le nombre de taxons différents que cela représente par domaine (en police maigre juste en-dessous).

Au niveau de l'ensemble des relevés, la catégorie d'intervenant ayant dénombré le plus de taxons différents est l'opérateur de type DREAL (B), suivi de près par l'intervenant de type scientifique (C).

Le nombre total de taxons identifiés par les intervenants de type DREAL pourrait être lié à la marge à une meilleure connaissance de flores spécifiques prospectées et documentées à l'occasion de l'élaboration des atlas floristiques régionaux.

Sur la totalité des stations du jeu de données, on peut constater que seulement un tiers des taxons observés sont communs entre les trois opérateurs au même site, et seulement un peu plus de la moitié des taxons (soit 56,8%) sont communs à au moins 2 opérateurs (somme des zones communes).

Ces résultats qui, à première vue, mettent en évidence des bilans inter-opérateurs sensiblement différents au niveau de l'occurrence des espèces, feront l'objet d'un développement spécifique au chapitre de discussion des résultats car ils méritent d'être mis en perspective avec la structure classique des communautés de diatomées, puis relativisés. En effet, ce type d'exploitation basé sur l'occurrence donne autant de poids, dans le bilan inter-opérateur, à un taxon très dominant au niveau de l'assemblage qu'à un taxon à effectif modeste, voire à des taxons de « queue d'assemblage » à très faible effectif, dont l'observation devient forcément de plus en plus aléatoire sur la base d'une pression d'observation limitée à 400 formes diatomiques par inventaire.

Un constat ressort nettement à l'analyse de ces résultats d'occurrence : pour la partie qui lui est propre, chaque catégorie d'utilisateur identifie un nombre ou un pourcentage de taxons identifiés par lui seul d'un niveau sensiblement différent. L'opérateur de type A (Bureau d'Etude) en identifie sensiblement moins (7,8%) que l'opérateur de type B (DREAL), qui en identifie un peu moins (15,5 %) que l'opérateur de type C (19,8 %).

La dispersion du pourcentage de taxons communs aux trois opérateurs sur l'ensemble des stations prospectées est présentée Figure 7 en page suivante.

La partie gauche de la figure présente le pourcentage de taxons en commun par rapport à la totalité des taxons trouvés, et la partie droite le pourcentage en commun de taxons indicateurs utilisés dans l'IBD₂₀₀₇.

De même que la Figure 6, cette figure, construite à partir de la contribution des taxons aux assemblages exprimée en occurrence (à savoir le pourcentage d'espèces diatomiques présentes communes aux 3 opérateurs), fait primer l'information sur la biodiversité des relevés dans l'absolu, donnant le même poids aux taxons rencontrés à faible effectif par rapport aux taxons dominant l'assemblage.

Cependant la construction du calcul diffère dans le sens où la part commune aux 3 opérateurs en occurrence n'est pas calculée globalement sur tout le jeu de triplets de relevés, mais sur la part de taxons communs aux 3 opérateurs pour chaque triplet individuel de relevés, ce qui conduit à l'obtention de pourcentages plus bas. En effet, même à supposer que la détermination taxonomique soit complètement comparable en inter-opérateurs, les taxons à faible ou très faible effectif sur une station vont être détectés de façon aléatoire en inter-opérateur comme en intra-opérateur (*i.e.* le même opérateur qui reproduirait exactement la même procédure d'échantillonnage - sous-échantillonnage en rapport avec le montage d'une nouvelle lame - détermination-comptage), ce qui contribue à faire baisser la proportion de déterminations communes à la même station.

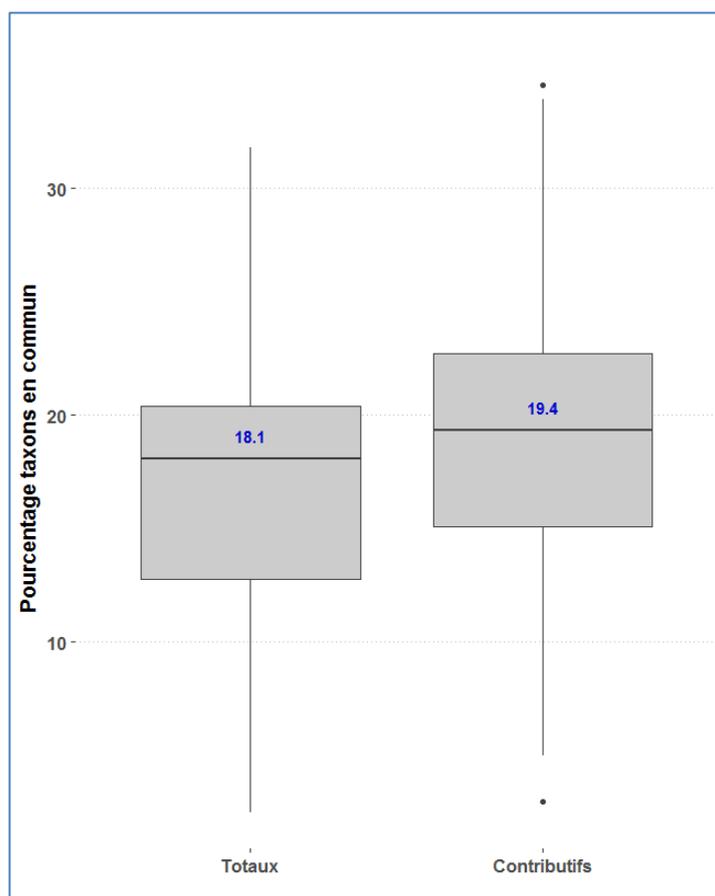


Figure 7 : Distribution des pourcentages de taxons totaux (à gauche) ou de taxons indicateurs de l'IBD 2007 (à droite) en commun entre les 3 opérateurs, exprimés en occurrence.

Sur cette base de comparaison, le pourcentage moyen de taxons indicateurs communs entre les opérateurs est significativement un peu plus élevé (au risque $\alpha = 0,05$) que le pourcentage moyen de taxons totaux communs entre les opérateurs, la médiane des taxons indicateurs communs s'établissant à 19,4% alors que la médiane des taxons totaux des assemblages est de 18,1%.

A l'autre extrémité de cette comparaison, les zones 2, 3 et 4 spécifiées sur la Figure 8 contiennent les taxons trouvés par un seul opérateur. Ce sont souvent des taxons à faible effectif, dont le dénombrement devient aléatoire dans un échantillon sur lequel s'applique une pression de comptage limitée à 400 valves. Mais il est aussi envisageable pour certains d'entre eux qu'une divergence taxonomique intervenant entre différents opérateurs puisse contribuer à amplifier ce score de taxons identifiés par un seul opérateur (noms différents correspondant à une espèce plus difficile à reconnaître avec certitude, notamment en fonction de la qualité de l'équipement de microscopie optique utilisé, ou résolution taxonomique différente selon l'opérateur allant parfois jusqu'à une sous-espèce ou à une forme spéciale).

Entre ces 4 zones décrites, qui constituent les extrêmes en matière de consensualité (zone 1) ou de différence inter-opérateurs (zones 2, 3 et 4), on retrouve des zones d'intersection 2 à 2 renseignées en Figure 6, correspondant à des espèces identifiées de concert entre 2 opérateurs, mais pas par le 3^{ème} (leurs listes n'ont pas été détaillées ici) :

- Il peut s'agir d'espèces en effectifs modestes dont le repérage à tout coup commence à devenir aléatoire (sa distribution à qualité de répartition moyenne, ou en effectifs faibles à moyens, préserve une bonne probabilité pour que 2 intervenants la dénombrent au même site alors qu'un troisième risque ne ne pas l'observer dans son comptage).
- Il est aussi possible, mais moins probable, que 2 opérateurs donnent le même nom à un même taxon observé et qu'un troisième lui affecte un nom différent ou une détermination taxinomique à un niveau de résolution différent, n'arrivant pas sur le même code-métier Omnidia et pas forcément sur le même profil de qualité dans la méthode. C'est seulement ce dernier cas de figure (arrivée sur un profil de qualité différent dans l'IBD) qui peut conduire à une variation non-anecdotique de la note d'IBD obtenue.

2.5.2. Aspects quantitatifs : abondances moyennes des taxons

La Figure 9 (cf. page suivante) ré-utilise le principe du diagramme de Venn sur les 45 triplets de relevés disponibles, en intégrant cette fois les abondances moyennes relevées par les opérateurs pour chaque taxon, et non plus seulement les listes taxonomiques en présence-absence. Par rapport aux analyses présentées au Chapitre 2.5.1, le diagramme construit sur cette base conduit à une analyse plus représentative de la reproductibilité de la pratique inter-opérateur. En effet, elle donne un poids de l'espèce proportionnel à l'effectif de frustules de cette espèce effectivement compté par chaque opérateur. A ce titre, les résultats obtenus ici se rapprochent plus du mode de calcul de l'IBD, qui mobilise les effectifs cumulés des taxons en abondances relatives pour attribuer sa note. Indirectement, ce type de comparaison donne aussi plus de poids, pour le recoupement des domaines inter-opérateurs, aux taxons à effectifs bien représentés qu'aux taxons à présence faible dont la détection conjointe restera forcément plus aléatoire. Ce principe de présentation s'appuyant sur les abondances relatives des taxons communs entre les relevés est très probablement un témoin plus fidèle du degré de comparabilité des pratiques de bio-indication diatomique entre les différents types d'opérateurs que les représentations précédentes basées sur les occurrences d'espèces (Figures 6 et 7), qui présentent l'inconvénient de donner le même poids à chaque taxon dans cette inter-comparaison, qu'il soit rare ou abondant.

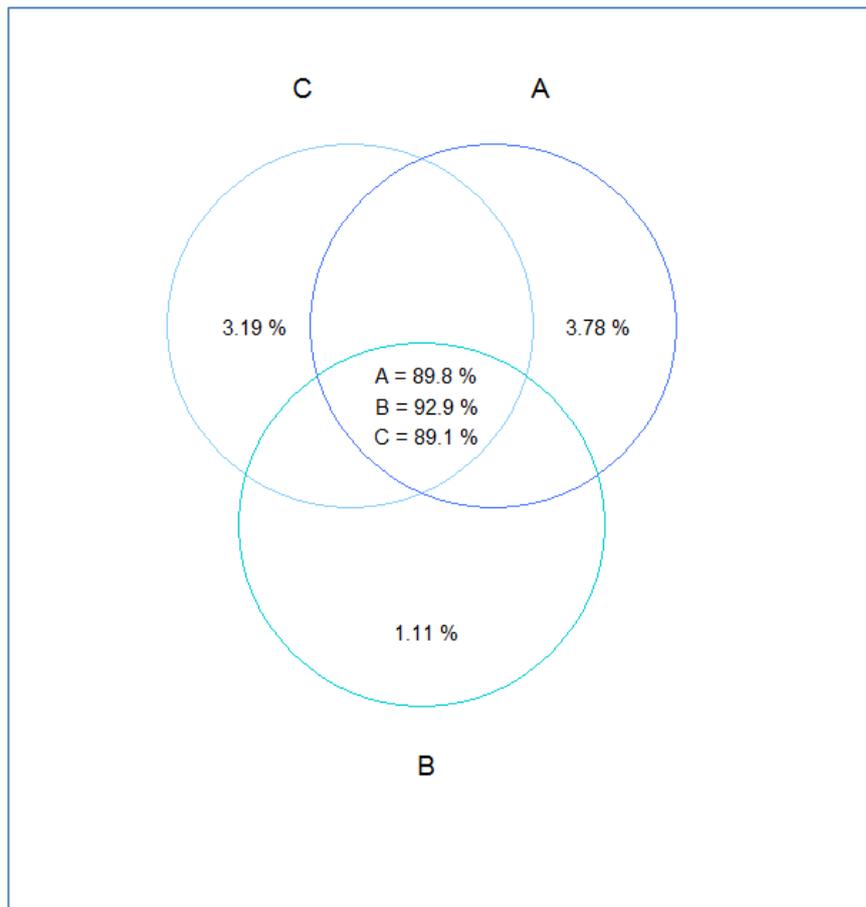


Figure 9 : Abondances relatives des taxons indicateurs communs entre les opérateurs

Si l'on examine plus en détail les résultats, la part pour chaque opérateur de l'abondance des taxons indicateurs retrouvés en commun par les 3 opérateurs est cette fois plus élevée (variation entre 89.1 et 92.9 %). Il est normal que les valeurs de chaque opérateur diffèrent un peu car, même si on leur faisait compter la même lame déjà préparée, ce qui n'est pas le cas ici (ils ont chacun procédé à leur propre échantillonnage diatomique et à leur propre montage de lame, dont le contenu va donc un peu différer), il serait pratiquement impossible que les 3 opérateurs obtiennent exactement les mêmes abondances relatives cumulées de chacun des taxons de la lame, cette considération valant pour tous les taxons comptés sur la lame et en particulier pour les taxons observés en commun.

Dans le cas des intercomparaisons réalisées ici, chacun des opérateurs ayant compté sa propre lame, les résultats inter-opérateurs sur un relevé prélevé au même site vont avoir une part de divergence liée au hasard d'échantillonnage de terrain, du sous-échantillonnage associé à la préparation d'une lame d'observation à partir du pilulier qui contient l'échantillon complet, du comptage et de l'aléa de répartition spatiale des champs d'observation sur la lame.

L'information dominante qui ressort de cette Figure 9 est qu'avec ce mode de présentation des résultats, la partie commune entre les 3 opérateurs s'est sensiblement accrue par rapport à ce qui était obtenu en Figure 6 (ou cette intersection comportait seulement 33% de taxons en commun).

De même, la prise en compte de l'abondance relative des taxons plutôt que de leur seule occurrence diminue aussi le poids relatif des domaines couverts par chaque intervenant en

propre (taxons comptés par un seul opérateur). Par exemple, 89.8 % de l'abondance relative des taxons observés par l'opérateur A sont associés à des taxons communs entre les 3 opérateurs, et seulement 3.78 % des effectifs de taxons qu'il a identifiés et comptés sont retrouvés uniquement par lui-même au niveau de la détermination spécifique, ce qui est nettement moins divergent que les résultats de la Figure 6.

Ce type de figuration donne donc un résultat sensiblement plus conforme au niveau usuel de reproductibilité de la bio-indication diatomique tel que connu des praticiens expérimentés (en effet, les calculs indiciels effectués en fin de procédure avec l'IBD₂₀₀₇ mobilisent eux aussi les résultats de comptages exprimés en abondances relatives).

Les taxons ont ensuite été classés selon le nombre de sites dans lequel ils ont été observés par tous les opérateurs simultanément (cf. Figure 10 ci-dessous).

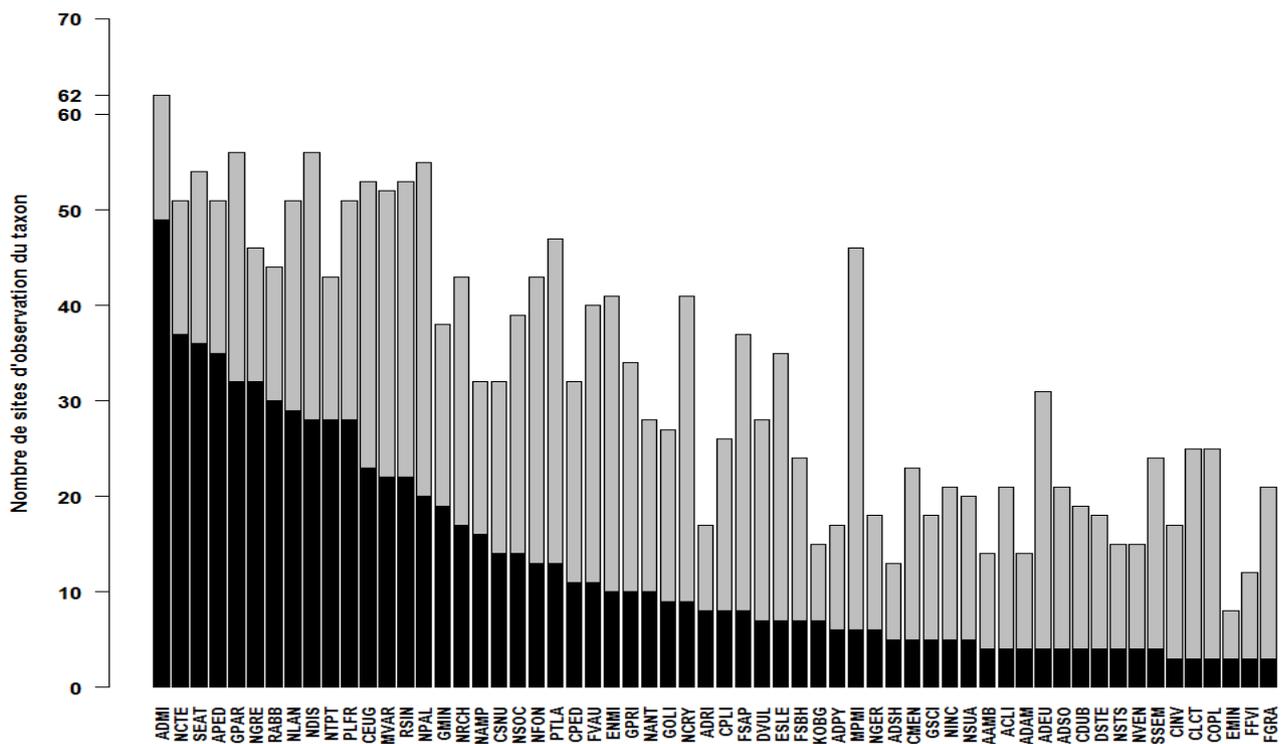


Figure 10 : Nombre d'observations de chaque taxon par tous les opérateurs simultanément

Sur cet histogramme, les barres noires représentent les nombres de sites dans lesquels les taxons les plus souvent rencontrés sont retrouvés simultanément par les 3 opérateurs et les barres grisées font état du nombre de sites dans lesquels l'espèce est retrouvée par au moins un opérateur (classement par ordre décroissant de la gauche vers la droite).

Comme c'est classique en matière de structure des communautés diatomiques, quelques espèces en partie gauche de la figure, à large répartition et qui peuvent dominer assez nettement les effectifs de certains assemblages, ont beaucoup plus de chances d'être observées sur un site et aussi, d'y être observées simultanément par les 3 types d'intervenants.

C'est tout particulièrement le cas de l'espèce (ou du complexe) ADM1, qui est observé au moins une fois dans 62 des 63 triplets d'observateurs, et par les 3 opérateurs simultanément pour 49 des 63 triplets.

Ce constat établi sur notre jeu de données d'étude est fidèle au constat habituel fait sur les réseaux de surveillance nationaux. En effet, ADMI est une entité taxonomique très répandue sur les cours d'eau de l'hydrosystème Français, à la fois sur le plan des occurrences et de son abondance relative moyenne.

C'est aussi le cas, dans une moindre mesure, d'une bonne dizaine d'espèces très abondantes et courantes dont la taxonomie *a priori* est souvent assez consensuelle (*Rhoicosphenia abbreviata* -RABB- par exemple).

Sur cette figure, on peut aussi observer que les 11 premières espèces en partant de la gauche de cette figure sont comptées par les 3 opérateurs dans plus de la moitié des sites où elles ont été repérées et comptées (jusqu'à l'espèce *Planothidium frequentissimum*, PLFR, déterminée de concert par 3 intervenants sur 28 sites sur 51 où elle a été trouvée).

On peut aussi voir qu'ensuite, la situation devient rapidement de plus en plus aléatoire côté reproductibilité de l'observation par les 3 intervenants. En effet, juste après cette espèce, le nombre de sites sur lesquels les 3 intervenants trouvent un taxon devient nettement inférieur à la moitié des sites où l'espèce a été observée.

Donc, mis à part cette courte liste de 11 espèces «dominantes» qui, quand elles sont présentes sur un site donné, ont plus d'une chance sur 2 d'être trouvées simultanément par les 3 opérateurs (statistique qui induit déjà un niveau assez conséquent de variabilité inter-opérateur), très vite, compte-tenu de la pression d'observation normative relativement modeste de 400 formes séparées identifiables / site, la détection simultanée d'un taxon donné par 3 opérateurs devient encore beaucoup plus aléatoire. La Figure 10 est largement tronquée, elle ne présente en effet que la distribution des 60 premiers taxons les plus retrouvés, sur un total de **729 taxons** constitutifs des 173 relevés diatomiques mobilisés. Au-delà de sa partie droite, il existe une longue « queue de comète » de taxons dont la distribution de moins en moins abondante conduit à ce qu'ils soient observés et comptés de façon de plus en plus aléatoire par les uns ou les autres des 3 intervenants. Cet état de fait conduit inévitablement, au final, à une variabilité assez importante de la composition spécifique des assemblages déterminés par chacun.

La variabilité de détection et de comptage dans un relevé donné peut tout d'abord être liée à des espèces non rares, mais qui ont une forte disparité spatiale de leur répartition au niveau de la station de prélèvement (voire temporelle, dans les cas de relevés faits à 2 ou 3 semaines d'écart...). Il est en effet notoire que certaines espèces, en raison d'une nature et d'une qualité de surface de substrat, de caractéristiques hydrodynamiques micro-locales etc..., colonisent des supports par petits patches à répartition aléatoire et irrégulière.

Une autre raison susceptible de conduire à ce type de résultat est la rareté relative des effectifs du taxon dans un assemblage donné, qui conduit à sa détection de plus en plus aléatoire dans cet assemblage, tous intervenants confondus. Selon les effectifs présents et l'aléa d'échantillonnage et de préparation d'une lame donnée, de tels taxons peuvent, soit garder des chances d'être rencontrés et comptés par 2 intervenants sur les 3, soit être peu abondants et risquer de n'être vus, de façon de plus en plus aléatoire et à très faible effectif, que par un intervenant sur les 3.

A l'extrême, des taxons faiblement présents et non régulièrement répartis sur la station pourront avoir été collectés dans un pilulier d'échantillonnage, mais pas dans les 2 autres, et même dans ce cas, n'avoir que peu de chance de figurer dans le sous-échantillonnage servant de base à la préparation de la lame d'observation (aléa croissant d'observation du taxon en liaison avec la baisse de ses effectifs).

Or malheureusement, il y a beaucoup de taxons de ce type dans un assemblage diatomique classique, et c'est une raison importante conduisant à une divergence d'assemblages, qui va encore se renforcer considérablement si l'on ne tient compte que de l'occurrence des taxons (cf. Figures 6 et 7).

Une dernière série de raisons pouvant conduire à des relevés différents peut être d'ordre taxonomique, en liaison avec des espèces dont la taxonomie est insuffisamment stabilisée, ou qui présentent une large plasticité phénotypique (cas de certains *Gomphonema*, de certains *Cocconeis...*), ou dont la qualité de la détermination, du fait de leur très petite taille, va être impactée par la qualité de l'équipement microscopique utilisé.

Par exemple, des espèces comme *Fistulifera saprophila* (FSAP) ou *Mayamaea permitis* (MPMI) semblent peu retrouvées simultanément par les trois opérateurs. Or ces espèces saprobes influencent grandement la note d'indice. En effet, dans le cadre du travail réalisé en 2011 par *Roussel et al.* concernant l'exploitation des exercices d'intercomparaison des DREAL, l'espèce FSAP avait été repérée comme responsable d'une part importante de la variabilité inter-opérateur des notes d'IBD₂₀₀₇ obtenues.

Il est possible que la reproductibilité défectueuse de leur détection selon les intervenants repose en partie sur un problème d'observation et de bonne reconnaissance pour les raisons techniques pré-évoquées (très petites espèces, microscope de qualité insuffisante ou mal réglé...).

Mais ce problème souvent rencontré par un seul intervenant sur les 3 pourrait aussi suggérer l'arrivée de cet opérateur sur un lieu de station ou sur un choix de faciès prospecté un peu différent des autres (cf. importance du contenu de la fiche signalétique de station). Il pourrait aussi renvoyer sur la technicité de la personne qui réalise l'opération d'échantillonnage sur le terrain. Plusieurs raisons de nature différente peuvent donc contribuer à faire diverger les résultats de relevés de surveillance, sans qu'il soit forcément possible de caractériser à coup sûr la raison principale.

Remarque : Dans le cadre de ce **volet d'étude de la variabilité inter-opérateur**, chaque relevé ayant été effectué une seule fois par type d'observateur, la variabilité totale d'assemblage au même site **sera interprétée à tout coup comme une mesure de l'effet inter-opérateur**. Mais il est certain qu'une part assez conséquente de cette variabilité, liée à l'aléa de prélèvement et d'observation d'une espèce donnée lorsqu'elle est modestement présente et abondante, **serait aussi rencontrée par un même opérateur** qui répliquerait toute la séquence allant du prélèvement à la fabrication des lames et à l'obtention du résultat de comptage.

En toute rigueur, la variabilité inter-opérateur obtenue dans cette étude est donc **un peu surestimée** car elle intègre aussi une part de variabilité aléatoire qui va inévitablement s'exercer pour tout opérateur réalisant plusieurs fois d'un bout à l'autre la procédure complète d'application de la méthode IBD₂₀₀₇ et ce, que l'on soit en inter-opérateur ou en intra-opérateur. Cependant, cette nuance n'est pas gênante par rapport à l'objectif finalisé de cette étude, puisque ce sont les résultats globaux de l'exercice qui ont été utilisés pour établir la quantification de l'incertitude de la méthode, que son origine soit intra ou inter-opérateur, et pour en décliner l'effet au final sur la confiance d'attribution d'une classe d'état écologique.

2.5.3. Effet du type d'opérateur sur la richesse spécifique

La variabilité globale de la richesse spécifique pour chaque type d'opérateur a ensuite été examinée (cf. Figure 11 ci-dessous)

Afin de procéder de façon rigoureuse, l'analyse portant sur les sites comprenant trois relevés comparables et celles portant sur les sites comportant deux relevés comparables ont dû être séparées afin de tenir compte de l'hétérogénéité des combinaisons d'opérateurs pour les stations à deux relevés comparables.

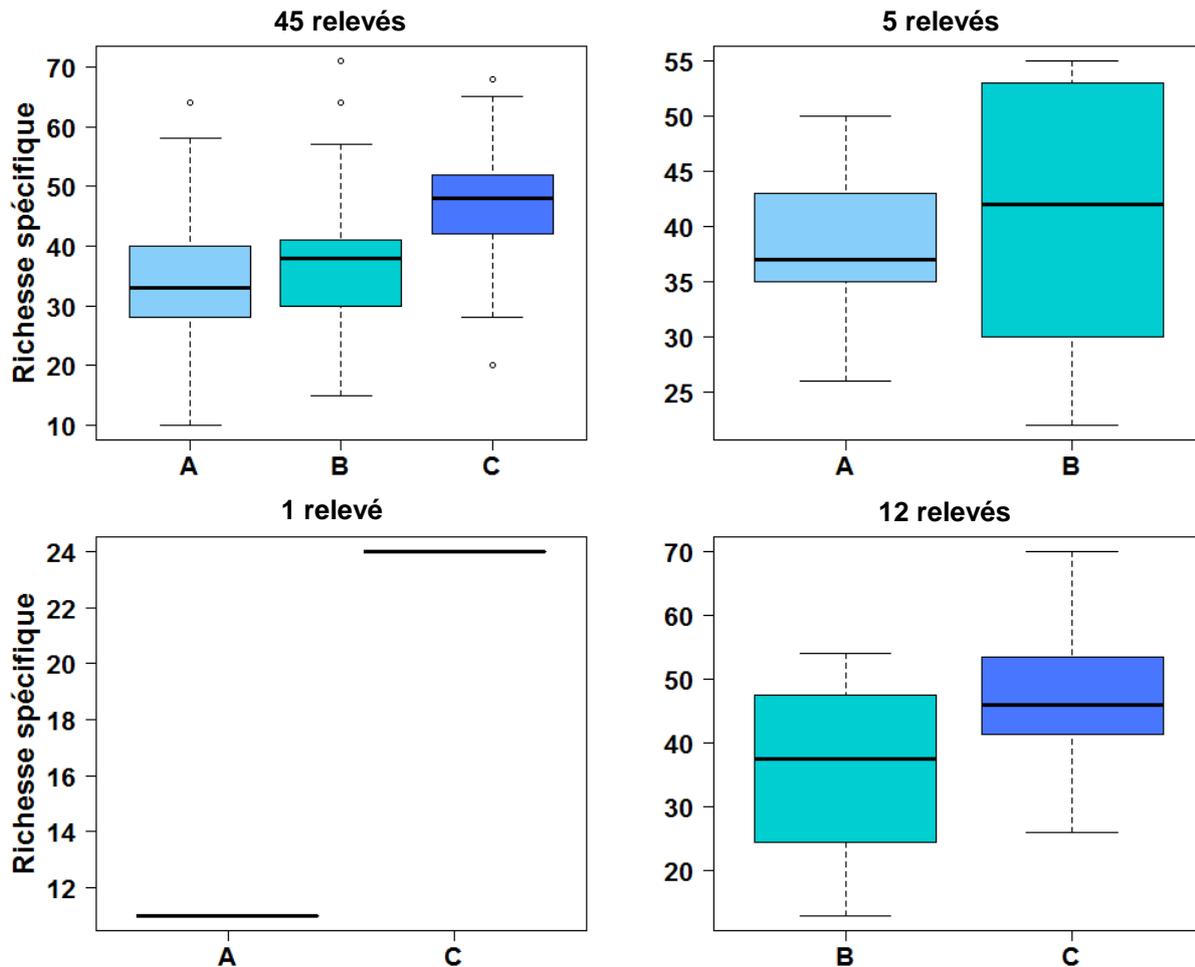


Figure 11 : Variabilité de la richesse spécifique selon les opérateurs
(barre centrale des box-plots = valeur médiane)

De ce fait, cette figure est structurée en 4 parties. La fenêtre 1 (en haut à gauche) représente la dispersion des valeurs de richesse spécifique des relevés réalisés de façon comparable par les 3 opérateurs. Les fenêtres de 2 à 4 représentent les combinaisons d'opérateurs dans les cas où les relevés ne sont issus que de 2 opérateurs. Les effectifs de relevés comparés 2 à 2 sont reportés en haut de chaque fenêtre.

La fenêtre 1 en haut à gauche, qui porte sur la comparaison de 45 triplets de relevés, présente l'analyse statistique la plus assise. Ce comparatif révèle une différence significative (p .value < 0.05) des moyennes de richesse spécifique. Celle-ci est significativement plus élevée pour

l'opérateur C que pour les deux autres types d'opérateurs. Les autres comparatifs suivent une tendance générale similaire à celle dégagée par la comparaison de triplets mais, assis sur un effectif plus réduit de doublets de relevés, les analyses des différences inter-opérateurs ne sont pas significatives. Pour le comparatif Opérateur A / Opérateur C (en bas à gauche de la figure), qui révèle un fort écart de richesse spécifique pouvant passer du simple à plus du double, le comparatif ne porte que sur un seul doublet de relevés sans répliqués et n'a donc pas pu faire l'objet d'une analyse de variance.

Les résultats de comptage des relevés réintégrés en richesse spécifique de chaque assemblage (Figure 11) donnent une vision plutôt convergente par rapport à ceux présentés en Figure 6 et dans une moindre mesure, en Figure 9, à savoir que l'opérateur C trouve souvent un peu plus de taxons, et le plus de taxons identifiés par lui seul, dans un relevé donné.

Sa pratique est plus proche de celle de l'opérateur B que de celle de l'opérateur A, qui trouve à la fois les assemblages les moins variés et moins de taxons identifiés par lui seul.

2.5.4. Discussion d'étape sur les résultats des 3 types d'analyses relatives à la diversité taxonomique des assemblages

Dans le cadre de ce volet d'étude de la variabilité inter-opérateurs, en fonction de l'ampleur de cette étude nationale (61 sites représentant déjà, pour chacun d'entre eux, 3 relevés à déterminer-compter dans l'intention initiale), du cadre financier contraint, ainsi que du temps forcément limité des personnes-ressources compétentes mobilisées pour les déterminations-comptages diatomiques, le protocole n'a pas pu être complètement stratifié pour localiser plus précisément les sources d'écarts et leurs origines, ce qui aurait fait rapidement exploser le nombre d'inventaires à identifier à l'unité taxonomique. Il n'est donc pas possible d'affecter dans le détail la contribution de chacune de ces sources aux différences significatives obtenues sur les résultats, et sur l'incertitude qui leur est associée.

Rappelons qu'une telle stratification avait été mise en place dans des études antérieures à portée plus limitée, dans le cadre d'exercices d'inter-comparaison réalisés entre les spécialistes-maillon des DREAL. Les tendances fortes des résultats obtenus, ainsi que les éléments relatifs à la hiérarchisation des sources (travaux de *Roussel et al, 2011 - REF. 6 -*) ont fait l'objet d'un récapitulatif en pages 9 et 10 du présent rapport.

L'objectif opérationnel principal visé à la fin de la présente étude étant d'attribuer une marge d'incertitude à l'évaluation d'état écologique réalisée à partir du compartiment diatomique des cours d'eau sur les sites du réseau de surveillance (le RCS) dédié au rapportage européen, le niveau de variabilité inter-opérateur a été abordé ici de façon globale et sera répercuté de façon globale sur l'évaluation (au relevé, en moyenne triennale) procurée sur ce site.

Cependant, un autre objectif associé à une telle étude nationale, menée sous l'égide de l'AFB et dans le cadre d'AQUAREF, est de repérer les points perfectibles dans la pratique de la surveillance diatomique telle que réalisée dans le cadre des réseaux, afin de la faire progresser vers une meilleure reproductibilité de ses résultats et vers une meilleure homogénéité à terme du contenu des référentiels collectés dans les bases de données nationales.

Il est donc utile de formuler et de débattre les hypothèses les plus plausibles susceptibles de sous-tendre les différences de composition taxinomique des assemblages diatomiques collectés, préparés et dénombrés par les 3 catégories d'intervenants selon les 3 niveaux d'exploitation qui

précèdent, dans le but de faire progresser à l'avenir la pratique sur les points apparaissant perfectibles.

Ces hypothèses se basent sur la **physionomie générale de ces résultats**, et pas sur tel ou tel cas particulier ayant pu affecter un doublet ou un triplet spécifique (par exemple délai trop important entre les passages de terrain, intervention d'un épisode hydrologique intercalé, possibilité d'erreur de localisation de station ou d'écart plus modéré ayant conduit au choix d'un faciès de prélèvement différent, etc...), sur lesquels diverses marges de progression existent, mais n'étaient pas forcément maîtrisables dans l'état par l'opérateur (cf aspect facilement perfectible de certaines signalétiques de station, séquence hydroclimatique particulière de type orage ayant interféré avec le positionnement comparable de l'opération de terrain, etc...).

a) Comparabilité des relevés :

Si l'on met de côté les 45 opérations de terrain aux conditions jugées suffisamment homogènes entre les 3 opérateurs, qui représentent 74 % de l'objectif initial de l'étude et conduisent à une comparaison de triplets sans problème particulier dans la Figure 11 (fenêtre 1 en haut à gauche), l'examen des nombres de doublets inter-intervenants des 3 autres fenêtres est porteur d'**informations quantitatives intéressantes**. En effet, la fenêtre 4 comporte 12 doublets, pour 5 dans la fenêtre 2 et un seul pour la fenêtre 1.

Pour établir une comparaison quantitative sur une base rigoureuse, il convient logiquement de diminuer l'effectif de la fenêtre 4 de 2 unités (il reste au final à considérer 10 doublets entre les opérateurs B et C), du fait que 2 stations ont subi 2 échantillonnages DREAL surnuméraires, dont un a été traité en doublet (cf bas de la Figure 5) et un a été sorti des comparaisons.

b) Identification et comptage de taxons :

Si l'on cible cette fois le propos sur les **comparaisons de richesse spécifique globale** présentée par les triplets ou doublets de relevés en fonction du type d'opérateur, l'examen de la Figure 11 montre des différences sensibles de résultats. Si l'origine ou les origines ne peuvent pas en être décryptées avec certitude faute de stratification suffisante du protocole, il est cependant possible de faire appel à diverses hypothèses explicatives.

Un aspect qui se révèle au niveau de l'analyse globale est que les richesses spécifiques observées par relevé et au niveau de tous les relevés comparables diffèrent assez sensiblement selon le type d'intervenant concerné. Bien sûr, compte-tenu de la pression d'observation de 400 valves, de l'effet de la réalisation de chaque prélèvement individuel et du faible effectif de bon nombre de taxons dans les assemblages, indépendamment de tout problème mettant en jeu un aspect de compétence taxonomique, une proportion assez sensible de taxons peu représentés peuvent, de façon aléatoire, faire partie du comptage ou au contraire, ne pas avoir été trouvés par un ou 2 types d'intervenant sur les 3.

Toutefois, à pratique équivalente, l'aléa de trouver ou de ne pas trouver telle espèce ou une autre dans tel ou tel relevé devrait logiquement à peu près se compenser entre intervenants sur le plan de la richesse spécifique. En effet, chaque taxon identifié, quel qu'il soit, intervient de façon équipondérée dans ce mode d'expression, qui s'établit uniquement en fonction des occurrences d'espèces différentes dans un relevé donné. Donc le résultat observé par un type d'intervenant donné ne devrait que peu varier au niveau de la biodiversité globale observée et ce, même dans le cas où un intervenant donné appellerait un taxon d'un nom éventuellement un

peu différent d'un ou de deux autres intervenants. Cela en changerait seulement la zone d'affectation dans les diagrammes de Venn.

Un fait évident qui ressort globalement de l'examen des figures 6, 9 et 11 est que l'opérateur de type A trouve moins de taxons par assemblage et en moyenne que l'opérateur de type B, qui en trouve un peu moins lui-même que l'opérateur C.

Si la tendance lourde qui se dégage sur le plan général est bien celle-là, l'importance des écarts entre types d'opérateurs peut cependant s'inverser selon qu'on considère la richesse spécifique des assemblages à l'échelle du regroupement de triplets (Figure 11, fenêtre 1, résultats plus proches entre intervenants A et B) ou l'occurrence de taxons identifiés à l'espèce qui, attribuant une valeur particulière à chaque taxon précis identifié, rend compte de façon plus robuste des similitudes ou divergences de résultats sur le plan taxonomique (Figure 6, résultats plus proches entre intervenants B et C).

La Figure 9, qui considère les résultats d'abondance relative des taxons réellement identifiés par chaque intervenant, donne l'assise taxonomique la plus solide aux résultats puisqu'elle se rapproche le plus de la pratique de bio-indication cible de l'étude (application de l'IBD₂₀₀₇, dont le calcul d'appuie lui aussi sur les abondances relatives des taxons constitutifs de l'assemblage. Selon le principe de présentation adopté, cette figure révèle une pratique sensiblement plus proche entre les 3 types d'intervenants que les résultats exprimés sur la base de la biodiversité des relevés (partie centrale, commune aux 3 intervenants intercomparés, approchant ou dépassant les 90% de recoupement pour seulement 33% selon le mode de présentation de la Figure 6.

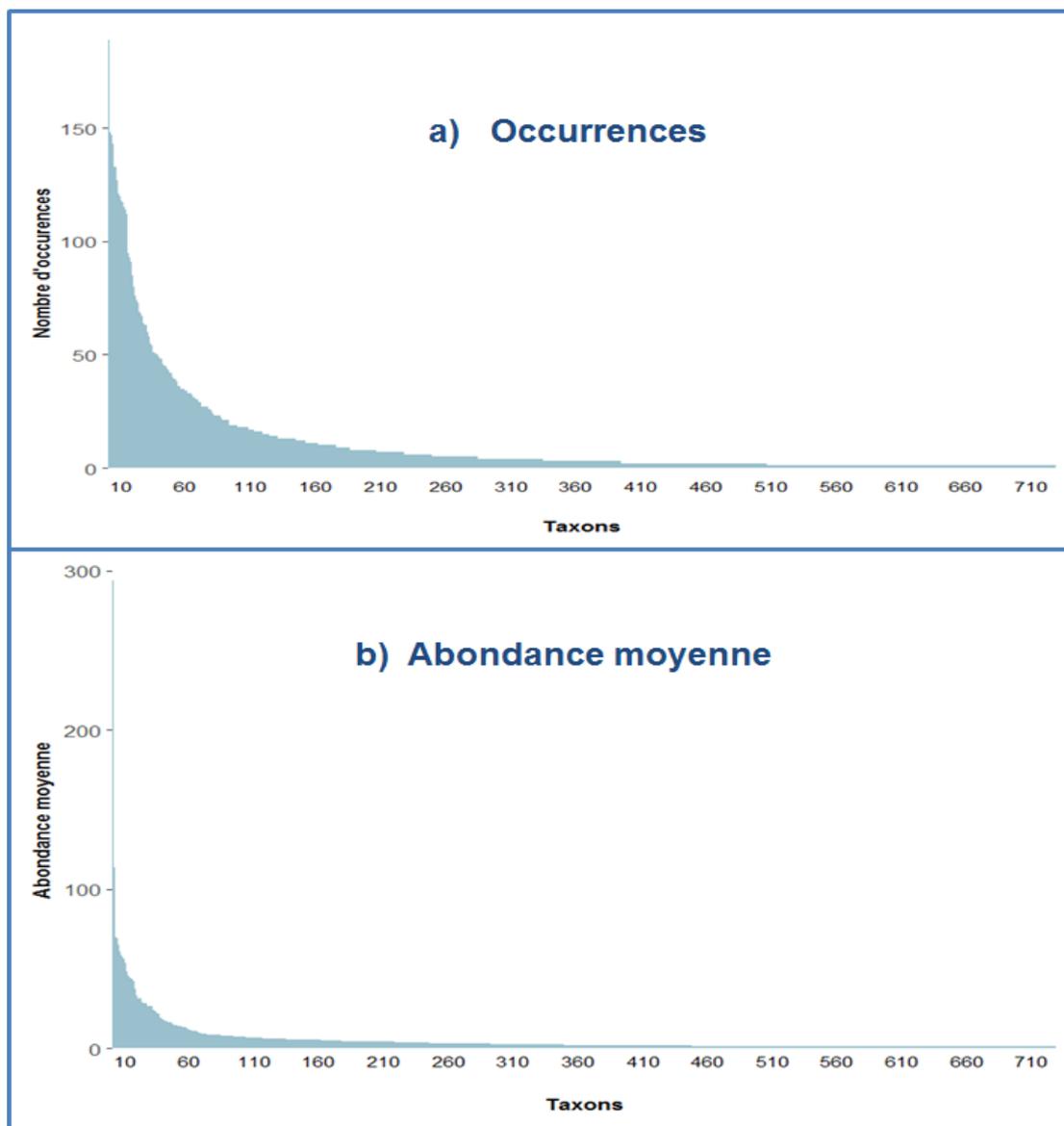


Figure 12 : Distribution des assemblages diatomiques du jeu de données d'étude : a) en occurrence des taxons ;b) en abondance relative moyenne de chaque taxon

La Figure 12, obtenue à partir du jeu de relevés estimés valides pour les intercomparaisons de la présente étude inter-opérateurs, illustre de façon représentative la distribution-type des espèces de diatomées dans le jeu de données diatomiques d'étude, d'abord selon un critère d'occurrence (en haut), puis selon un critère d'abondance relative moyenne (en bas). Son assise porte sur le total des 729 taxons qui contribuent à la composition des 173 relevés diatomiques individuels utilisés.

Un premier examen global portant sur ces 2 parties de figure confirme que les espèces fortement occurrentes, ou respectivement fortement abondantes, représentent une faible proportion des assemblages rencontrés dans les relevés échantillonnés.

En effet, ces courbes de distribution présentent un nombre plutôt modeste de taxons atteignant d'assez fortes valeurs en partie gauche des histogrammes, puis les valeurs s'éboulent rapidement pour les taxons suivants jusqu'à représenter une longue queue de comète de taxons de peu à très peu représentés.

Si l'on s'intéresse d'abord à la distribution **en occurrences** (partie haute de la figure), on peut observer que, mis à part le taxon dominant ADMI, présent dans une nette majorité des relevés (148 sur 173), seulement une dizaine de taxons sont occurrence, parfois à forts effectifs, mais parfois aussi à faibles effectifs, dans plus de 50% des relevés (soit 86 à 87 relevés).

Ensuite, les occurrences concernent très vite une proportion beaucoup plus faible des relevés : 73 taxons (soit 10 % du total), ne sont représentés que dans 25 relevés (soit moins de 15% du total des relevés). 90% des taxons, distribués en allant plus vers la gauche de l'histogramme, ne sont donc plus présents que dans moins à nettement moins de 25 relevés sur les 173 (plus on va vers la droite, plus ils concernent un effectif faible de relevés, jusqu'à ne concerner que 2 ou un seul relevé). En fonction de cette structure d'assemblages diatomiques, qui est tout-à-fait habituelle même quand un seul intervenant taxonomiste a réalisé toutes les déterminations (cas des chantiers DOM-TOM récents, qui n'introduisaient pas de variation inter-opérateur sur le plan de la détermination des espèces), il est donc tout-à-fait normal de trouver une assez grande divergence d'assemblages taxonomiques exprimés sous forme d'occurrences des espèces trouvées. En effet, selon ce mode d'expression qui sous-tend en particulier l'élaboration des figures 6, 7 et 10, tous les taxons situés vers la droite de l'histogramme des occurrences contribuent autant, lorsque un opérateur les a trouvés, que des taxons nettement plus importants dans la structure d'un assemblage donné, et même s'ils ont été trouvés avec un effectif de 1 dans un seul relevé.

Il est évident que ces nombreux taxons à faible ou très faible occurrence dans le jeu de données sont trouvés avec un important effet aléatoire entre les 3 opérateurs, même sur le site ou l'un d'entre eux a pu en observer. C'est donc cette intervention aléatoire de taxons peu représentés dans les échantillons qui est la principale explication du contenu des domaines individualisés de chaque opérateur dans les figures 6 et 9 (taxons trouvés par un seul intervenant).

En **abondances relatives moyennes** (cf. partie b) de la figure) ou absolues (phénomène encore plus marqué, non représenté ici), cette courbe de distribution s'éboule rapidement, de façon comparable ou encore plus vite qu'en occurrences par relevé. Cependant, dans ce cas, les taxons contributifs au calcul de l'indice vont être pris en compte en fonction de leur abondance relative dans l'assemblage concerné, ce qui se rapproche plus du mode de calcul de l'IBD, et les taxons à plus fort effectif dans le relevé vont contribuer à déterminer la part commune entre intervenants à hauteur de leur abondance relative moyenne dans le triplet de relevés inter-comparés (i.e. les 3 relevés échantillonnés par les 3 types d'intervenants au même site).

Mis à part le complexe taxonomique ADMI, qui présente une abondance relative moyenne proche de 300 pour mille dans les relevés où il est présent (taxon très occurrence et abondant, présent dans 148 sur 173 relevés), seule une vingtaine de taxons dépasse 50 pour mille en abondance moyenne (soit plus d' $1/20^{\text{ème}}$ de l'assemblage dans lequel ils ont été identifiés). En conséquence, au hasard des échantillonnages et sous-échantillonnages (reprise d'une petite partie de l'échantillon composite pour la fabrication d'une lame permanente), les relevés diatomiques présentent une forte biodiversité spécifique, qui avoisine 40 taxons en moyenne (variation entre 10 taxons seulement, pour le relevé d'un intervenant A ou a été trouvée la plus faible biodiversité spécifique, et 71 taxons pour un relevé du type d'intervenant B ou a été trouvée la plus grande diversité spécifique).

Malgré cette assise importante de biodiversité spécifique par relevé, dont une partie va forcément être captée de façon aléatoire entre les 3 observateurs (en particulier les espèces peu occurrence et peu abondantes, avec l'aléa de leur distribution sur la station, du protocole

d'échantillonnage, du sous-échantillonnage réalisé pour monter les lames d'observation, puis du hasard de leur répartition sur une lame donnée), l'exploitation de l'inventaire en abondances relatives va minimiser l'écart des relevés en inter-opérateurs, donnant plus de poids aux espèces les plus fréquemment trouvées (en fonction de leurs effectifs dénombrés) par rapport aux taxons peu fréquents et peu abondants, qui interviendront de façon nettement plus aléatoire dans les relevés et les inter-comparaisons de relevés.

C'est ce qui conduit au résultat présenté en Figure 9 (diagramme de Venn construit en fonction des abondances des taxons dans les relevés), où la part comparable des échantillonnages entre les 3 intervenants approche ou dépasse en moyenne les 90% de la composition totale des triplets de relevés intercomparés, contre 33,3% en occurrences de taxons selon le même mode de représentation (cf. Figure 6).

Outre d'éventuels écarts de pratique taxonomique, qui peuvent exister mais qui se réduisent lorsque les déterminateurs capitalisent de l'expérience et se réfèrent aux mêmes documents-guides (notamment guide taxonomique de l'IBD et atlas régionaux...), c'est cette intervention aléatoire de taxons peu représentés dans l'échantillon qui est la **principale explication du contenu des domaines individualisés de chaque opérateur** dans les figures 6 et 9 (taxons trouvés par lui seul), des taxons à distribution un peu plus abondante ou régulière contribuant beaucoup aux recouvrements 2 à 2. Seuls, les taxons distribués de façon suffisamment régulière sur le site et suffisamment abondants contribueront à alimenter significativement la partie centrale (zone de détermination commune entre les 3 types d'intervenants).

2.6. Notes d'IBD₂₀₀₇

La variabilité des notes d'IBD₂₀₀₇ obtenues à partir des relevés élémentaires est présentée en Figure 13 (cf. ci-dessous) selon le même principe de présentation que celui adopté en Figure 11 (traitement séparé des sites comprenant trois relevés comparables et de ceux comprenant deux relevés comparables), et avec les mêmes effectifs de relevés par combinaison.

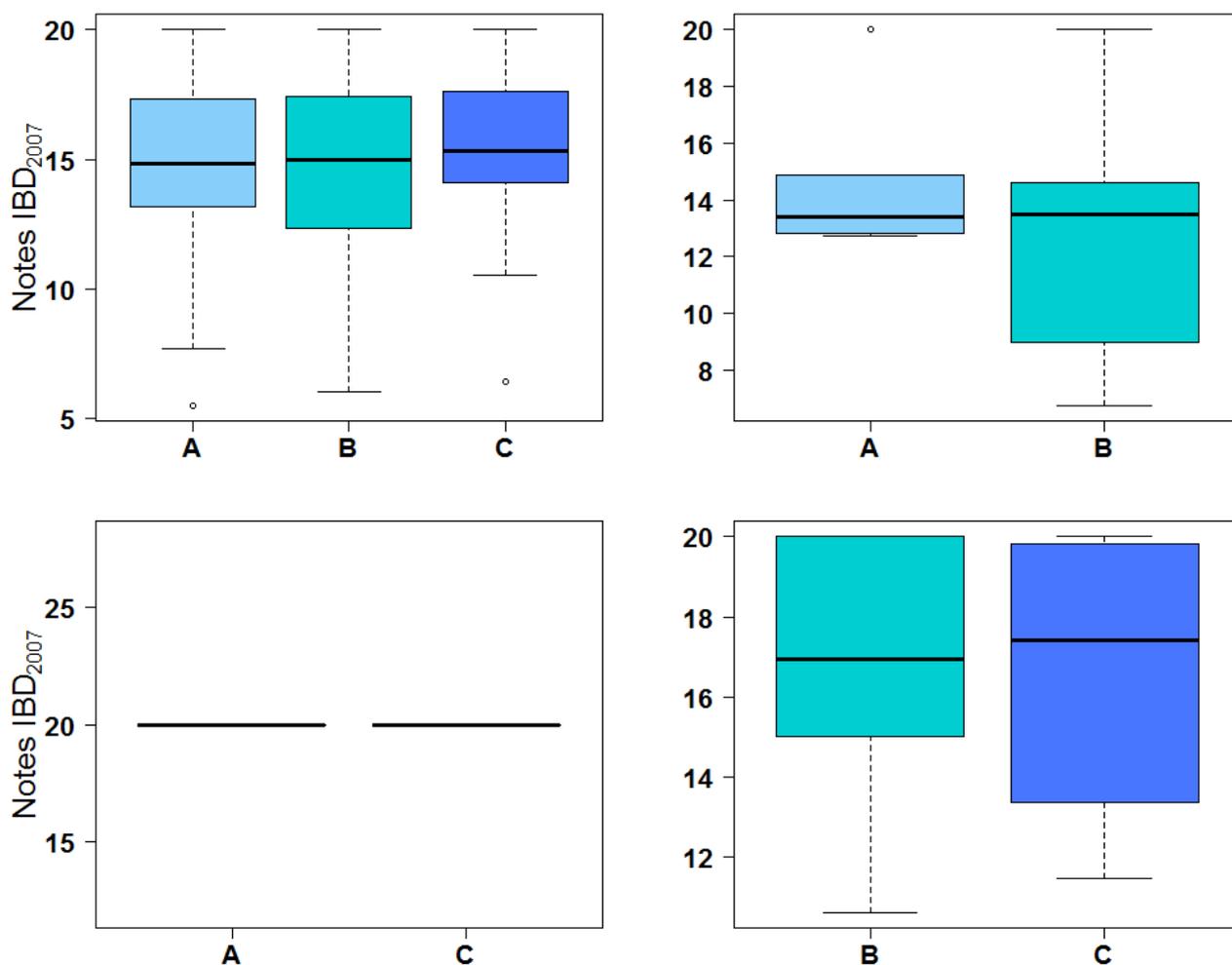


Figure 13 : Variabilité des notes d'IBD2007 selon les opérateurs

Au niveau global du jeu de données, les analyses statistiques n'ont pas révélé de différence significative de répartition des notes d'IBD₂₀₀₇ entre types d'opérateurs, quel que soit le sous-bloc de comparaison considéré (le 3^{ème} entre opérateur A et C, basé sur une seule comparaison sans réplicat, ne pouvant faire l'objet d'aucune analyse de variance).

L'IBD₂₀₀₇ affiche donc une certaine stabilité de ses résultats malgré des communautés observées parfois bien différentes. A titre d'exemple, c'était d'ailleurs le cas pour les 2 relevés intercomparés du bloc 3 qui, malgré une forte différence de biodiversité observée entre les 2 opérateurs A et C (de 11 à 24 taxons, soit du simple à plus du double –cf. Figure 11, sous-bloc en bas à gauche–, obtient dans ce cas le même résultat indiciel révélant des communautés en TBE.

Même s'ils peuvent présenter des différences taxonomiques et morphologiques parfois importantes, notamment au hasard des caractéristiques physiques des micro-habitats où ils ont été prélevés, les cortèges prélevés puis observés sont le plus souvent composés d'espèces bien adaptées au niveau intégré de qualité des eaux caractéristique de leur milieu de vie, et vont avoir de bonnes chances de présenter sur ce plan un profil écologique assez comparable selon l'IBD₂₀₀₇ (notion d'homologie de la valeur bio-indicatrice de taxons, qui peuvent par ailleurs être adaptés à des conditions de vie différentes sur le plan du fonctionnement des milieux naturels).

En d'autres termes, l'attribution de profil de qualité des taxons dans l'IBD a été calculée en tenant compte de 7 descripteurs influents représentatifs des effets de l'altération anthropique sur la qualité des eaux, et le profil de chacun d'entre eux résume donc à titre principal le niveau intégré de qualité des eaux tel que décrit par ces 7 descripteurs.

On pourra donc trouver des taxons très différents sur le plan de leur adaptation à des conditions environnementales locales (par exemple, milieu plus ou moins hydrodynamique, milieu plus ou moins éclairé, qualité et rugosité des substrats...), mais qui réagissent à peu près de la même façon aux caractéristiques d'enrichissement organique ou trophique de la colonne d'eau sous influence des altérations anthropiques.

Même si des co-évolutions sous-jacentes d'autres facteurs non pris en compte sont toujours possibles à la marge, c'est la réaction des taxons contributifs de l'IBD₂₀₀₇ au degré d'enrichissement anthropique de la colonne d'eau qui va très principalement déterminer la réponse de cette méthode indicielle.

Il est intéressant de préciser ici que des communautés très dissimilaires ne conduisent pas forcément à de forts écarts de la note d'IBD (cf. Figure 14 ci-dessous).

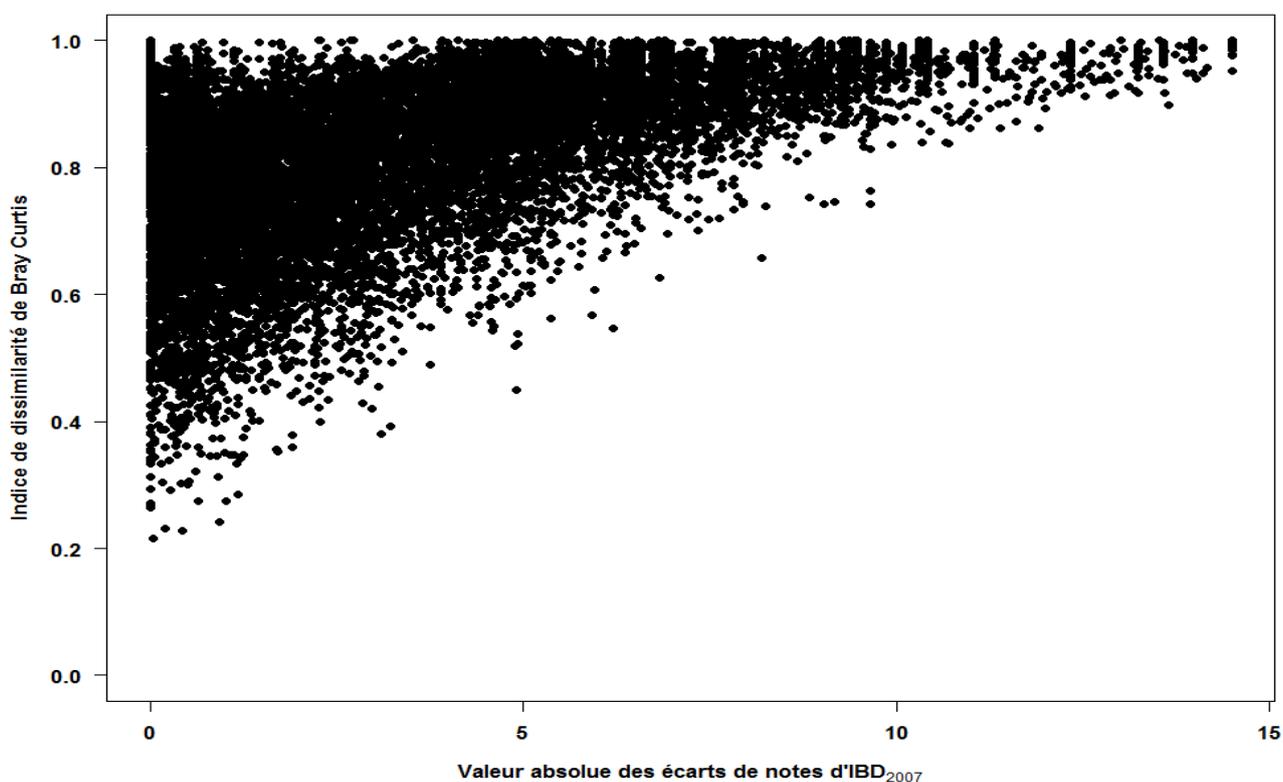


Figure 14 : Distance de Bray-Curtis sur la matrice spécifique en fonction des écarts de notes d'IBD2007

A l'appui de l'exemple commenté un peu plus haut, cette figure présente en effet la relation entre dissimilarité taxonomique des communautés de notre jeu de données d'étude mesurée par l'indice de Bray-Curtis, tous relevés confondus, et écart de note obtenu (plus l'indice de dissimilarité de Bray-Curtis est proche de 1, plus les communautés observées sont différentes).

On peut y repérer en partie gauche de la figure que certaines communautés parfois très semblables (proches de 0), parfois très dissemblables au niveau taxonomique (proches de 1), peuvent se révéler très homogènes du point de vue de la représentation de la qualité écologique

du milieu et conduire à des notes d'indices comparables (partie gauche de la figure, en zone proche de l'axe des y). Ce cas de figure peut se présenter pour les communautés prélevées par des opérateurs différents au même site, les communautés en place pouvant manifester une faible dissimilarité ou une dissimilarité plus importante sans conduire forcément à des écarts dans l'évaluation de la qualité du milieu (via les écarts de note d'IBD).

En revanche, de forts écarts de note d'indice (zone plus à droite de la figure) correspondent toujours à des structures de communautés très différentes (valeurs de dissimilarité de Bray-Curtis élevées, s'approchant de 1).

Cela confirme que la méthode indicielle de détermination de l'état diatomique des cours d'eau basée sur l'IBD₂₀₀₇ est rustique et apte à tamponner, jusqu'à un certain point, les écarts locaux de communautés prélevées, dans la mesure cependant que les différents taxons trouvés présentent des profils de qualité comparables dans l'IBD.

Après avoir vérifié en Figure 13 l'absence d'effet-opérateur significatif sur la statistique globale de notes d'IBD obtenues sur toute l'étude, la Figure 15 ci-dessous permet de se concentrer sur l'analyse de la dispersion inter-opérateurs des notes d'IBD pour chaque site d'étude.

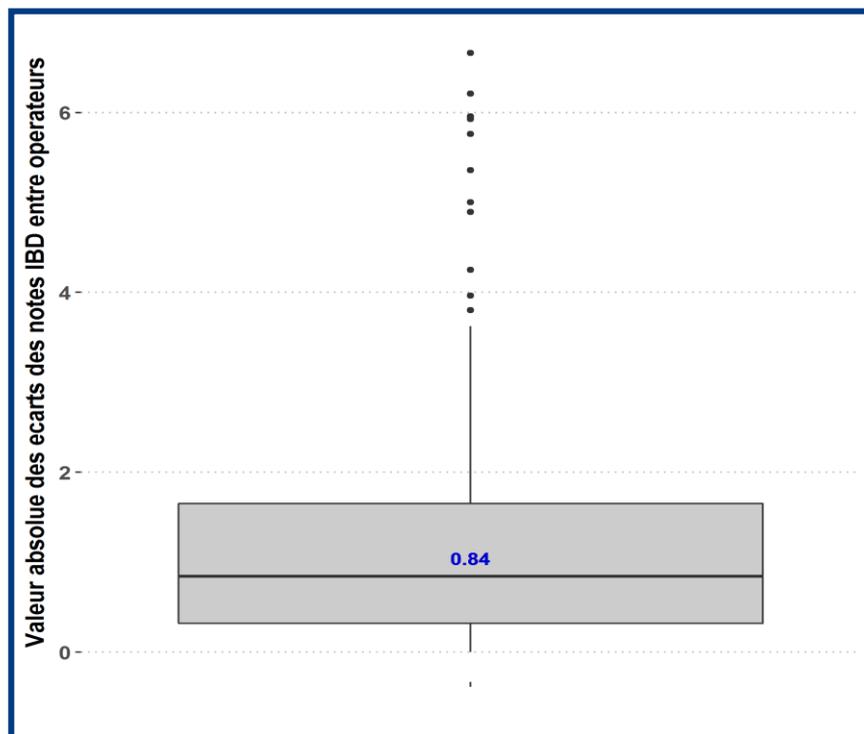


Figure 15 : Valeurs absolues des écarts 2 à 2 obtenus entre opérateurs sur les notes d'IBD2007 pour chaque site d'étude

Cette façon de représenter l'effet inter-opérateur rend compte à la fois de la distribution des plus faibles écarts et des plus forts écarts 2 à 2 obtenus par 2 types d'opérateurs au même site, y compris sur des sites où il existe un doute assez important, mais d'origine non prouvée, sur la reproductibilité des opérations de prélèvement. L'écart maximum 2 à 2 rencontré dans l'étude a pu atteindre 6,6 points d'IBD.

Avant de commenter ces résultats, il convient de rappeler au préalable que les seuls relevés invalidés pour l'analyse l'ont été sur problème avéré et objectif, à savoir : 1) des relevés où

l'écart de dates de passage était trop important et ne respectait pas la fenêtre temporelle maximale tolérée (> 14 jours), rajoutant à la variabilité inter-opérateur, qui était la cible centrale de ce volet d'étude, un effet indissociable de variabilité temporelle ; 2) un nombre très réduit de relevés avec erreur de site manifeste et confirmée ; ou 3) dans 3 cas, des relevés où la fenêtre temporelle était respectée mais où une forte crue d'orage était intervenue entre le passage des 2 premiers opérateurs et le 3^{ème}, invalidant le relevé réalisé par le dernier opérateur.

Dans le jeu de relevés diatomiques intercomparés, il reste donc des suspicions de relevés à comparabilité défectueuse (certains relevés étant très divergents des 2 autres), un opérateur ayant par exemple pu échantillonner un site différent ou un faciès sensiblement différent des 2 autres. Faute d'implémentation systématique sur toutes les fiches-terrain de tout l'assortiment de descripteurs annexes demandés dans le cadre de l'étude (notamment données physico-chimiques ou coordonnées géoréférencées captées sur le terrain qui étaient souvent manquantes, dans le désordre.), qui auraient permis de porter une expertise certaine et de confirmer ou d'infirmer d'écart de comparabilité sur les triplets de relevés diatomiques en cause, il a souvent été impossible de caractériser avec certitude qu'un problème de ce type s'était posé, et lequel.

En fonction de l'objectif opérationnel de cette étude nationale, l'enjeu n'était pas de contenir un peu artificiellement la variabilité de la méthode d'évaluation diatomique en supprimant les relevés avec écarts douteux ou gênants, mais de **situer un intervalle de confiance robuste associé au résultat** qui puisse intégrer de façon globale différents types de problèmes susceptibles de se poser dans la pratique habituelle des réseaux.

Il a donc été décidé de maintenir dans les intercomparaisons certains relevés où, **même si un problème de comparabilité était assez fortement suspecté de par la composition biologique**, sa réalité sur le plan d'écarts ou d'erreurs de protocoles **n'a pas pu être démontrée sur la base de critères objectifs**. Ce faisant, l'intervalle de confiance attribué au final à la classification d'état écologique est un peu plus large, mais aussi plus fiable par rapport à l'incorporation de problèmes qui peuvent arriver au cas par cas dans la vie réelle des réseaux de surveillance.

En fonction de cette sélection prudente (*i.e.* sur problème confirmé...) intervenue sur le jeu de relevés préalablement à l'intercomparaison, les écarts les plus fréquents obtenus entre les opérateurs aux mêmes sites peuvent varier entre 0 et 3,8 points d'IBD hors cas très exceptionnels (outliers vrais dans la distribution), la médiane s'établissant à 0,84 point d'IBD par doublets de relevés.

Quelques relevés outliers, possiblement associés à un défaut de comparabilité de pratique ou peut-être même de site de prélèvement échantillonné, ont atteint un écart de note indicelle situé entre 4 et un peu plus de 6 points d'IBD. Il a aussi été tenu compte de ces écarts, bien que d'un niveau un peu exceptionnel par rapport à l'expérience des praticiens, dans l'attribution de l'intervalle de confiance associé à la méthode d'évaluation diatomique basée sur l'IBD.

la Figure 16 ci-dessous permet cette fois de se concentrer sur l'analyse de la dispersion inter-opérateurs des notes d'IBD obtenues au même site, calculée à partir des écarts à la note moyenne par site.

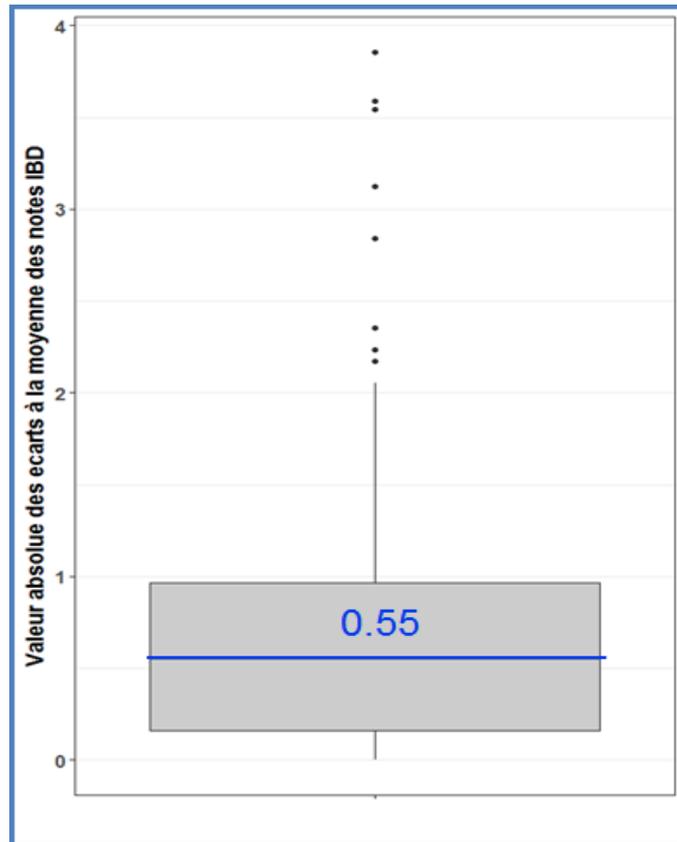


Figure 16 : Valeur absolue des écarts à la moyenne des notes d'IBD par relevé au même site

Selon ce principe de représentation, les écarts inter-opérateurs peuvent varier entre 0 et un maximum de 3,8 points d'IBD par site dans le cas le plus extrême.

La médiane des écarts par site, tous sites de l'étude inter-opérateur confondus, s'établit ici à **0,55 point d'IBD₂₀₀₇** et **plus de 75% des cas** présentent un écart à la moyenne **inférieur à 1 point**. Ces 2 valeurs-guides correspondent en gros à la fourchette de variation classiquement évoquée par les praticiens ayant acquis un recul assez important sur la mise en œuvre de l'IBD.

La limite de distribution usuelle peut s'étendre jusqu'à un peu plus de 2 points d'IBD, hors outliers vrais.

Les 4 à 5 sites plus inhabituels, sur lesquels s'est probablement posé un problème particulier (fort écart souvent constaté entre 2 intervenants plutôt convergents et le 3^{ème}), génèrent 9 doublets avec des écarts à la moyenne s'échelonnant entre 2,2 et 3,8 points entre relevés.

2.7. Etat écologique

Dans ce chapitre, il fallait passer des notes d'IBD au relevé à l'évaluation d'état écologique au relevé, en tenant compte du niveau de référence adopté pour chaque ensemble naturel homogène de notre territoire national (notions d'HER regroupées en régions diatomiques).

Les classes d'état écologique obtenues pour chaque site d'étude à partir de la composition floristique de chaque relevé diatomique (2 à 3 relevés comparables selon le site) ont été dérivées des notes d'IBD₂₀₀₇ en s'appuyant sur les grilles d'évaluation d'état écologique au relevé

contenues dans l'Arrêté Evaluation en vigueur au moment des interprétations finales de l'étude (Arrêté du 27 Juillet 2015).

Pour cette analyse (cf. Figure 17 ci-dessous), sur les **61 sites d'étude**, 45 sites différents ont fait l'objet d'analyses de triplets, auxquels se sont rajoutés 2 triplets supplémentaires (2 sites échantillonnés plusieurs fois par certains opérateurs), soit **47 triplets**.

16 doublets de relevés disponibles ont aussi été intégrés dans l'analyse. La principale raison de manque d'un relevé a été liée à une élimination pour raison d'écart temporel trop important (12 relevés). Mais il a aussi pu aussi intervenir une erreur avérée de site chez un intervenant (2 cas), un relevé non fait par un opérateur (1 cas), un doublet suivi d'un triplet (un site ayant été fait par un doublet dans le bon créneau temporel, puis ayant pu être ré-échantillonné plus tard par les 3 types d'intervenants...).

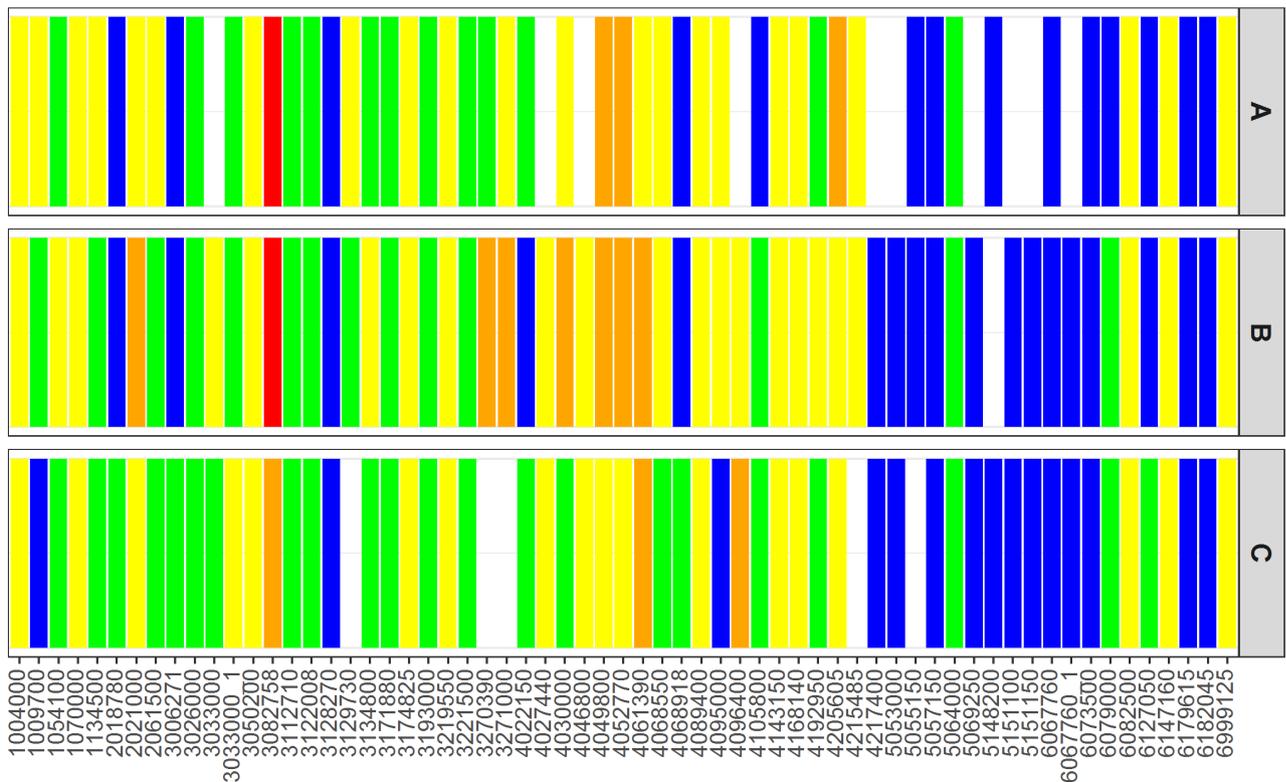


Figure 17 : Comparaison des classes d'état écologique obtenues sur les 61 sites d'étude (soit 63 inter-comparaisons d'état).

Nb : Les codes stations figurent en abscisse, les classes d'état en ordonnée, selon le type d'opérateur (bleu : très bon état ; vert : bon état ; jaune : état moyen ; orange : état médiocre ; rouge : mauvais état).

Le diagnostic d'écart de dates jugé acceptable (maximum 2 semaines) ne garantit d'ailleurs pas à tout coup qu'il n'ait pas pu y avoir dans certains cas, en fonction de renseignements ou d'éléments de consolidation utiles mais pas toujours renseignés chez les 3 intervenants (physico-chimie de l'eau, coordonnées géographiques saisies sur le terrain et reportées sur la fiche de notation etc...), une différence non complètement avérée introduisant une comparabilité un peu défectueuse de quelques relevés supplémentaires.

Dans ces conditions, les opérateurs, quel qu'en ait été le nombre (3 ou 2), ont obtenu un **classement écologique exactement convergent** dans la majorité des cas, soit sur **35 sites (55%)**, qui sont évalués dans la même classe de qualité.

Ils ont assez souvent attribué un état écologique **avec une classe d'écart**, soit sur **24 sites (38%)**. Le plus souvent, dans ce cas, l'écart entre les valeurs d'EQRs est contenu dans une fourchette modeste et fait basculer un site proche d'une limite de classe dans 2 classes de qualité contigües selon le résultat du relevé.

Dans quelques cas cependant, l'écart ne reste pas cantonné au voisinage de la jonction entre 2 classes mais s'étend un peu plus largement, révélant souvent 2 évaluations plutôt homogènes qui présentent le niveau habituel de reproductibilité attaché à l'application de cette méthode, et une évaluation un peu plus éloignée probablement révélatrice d'un écart de pratique qui, quelle qu'en soit l'origine, exerce déjà une répercussion non-négligeable sur le niveau d'évaluation du site au relevé. S'agissant d'une méthode hydrobiologique qui présente une part intrinsèque de variabilité dans son application et ses résultats, et qui doit d'autre part évaluer une part assez conséquente de sites jouxtant des limites de classes de qualité, les résultats précédents sont logiques et étaient pressentis.

Au bilan, **93 % des sites prospectés** lors de cette étude inter-opérateurs **semblent évalués d'une façon raisonnable à assez raisonnable** ne laissant pas suspecter de problème majeur de reproductibilité.

Ce qui se révèle un peu plus problématique est que 4 sites (représentant 6% des sites d'étude et 7% des intercomparaisons de relevés réalisées), pourtant prélevés dans des intervalles de dates corrects, ont un résultat d'évaluation atteignant 2 classes d'écart entre intervenants, laissant suspecter un problème de comparabilité d'une autre nature. Ces sites sont bien sûr responsables des plus gros écarts de notes observés entre intervenants 2 à 2, qui peuvent atteindre à l'extrême plus de 6 points d'IBD (cf. Figure 15).

Même si la cause de ces écarts anormaux n'a pas pu être diagnostiquée avec certitude, les possibles raisons des écarts sous-jacents à ces cas particuliers nécessitent d'être analysées avec le plus d'acuité, car les problèmes qu'ils illustrent ou laissent subodorer vont devoir être résolus en priorité afin de contenir de façon plus satisfaisante les incertitudes liées à la méthode IBD. Les hypothèses sous-tendant ces écarts feront donc l'objet d'un développement spécifique au **Chapitre 2-8** « Discussion - conclusion du volet inter-opérateur ».

Les résultats de classement écologique comparé des sites pour chaque type d'opérateur ont ensuite été projetés géographiquement sur trois cartes, une par type d'opérateur (**Figure 18** en page suivante).

Même s'il peut arriver qu'un site soit évalué de façon assez divergente par un opérateur par rapport aux 2 autres, la physionomie générale de ces cartes et des états écologiques par bassin est globalement assez similaire.

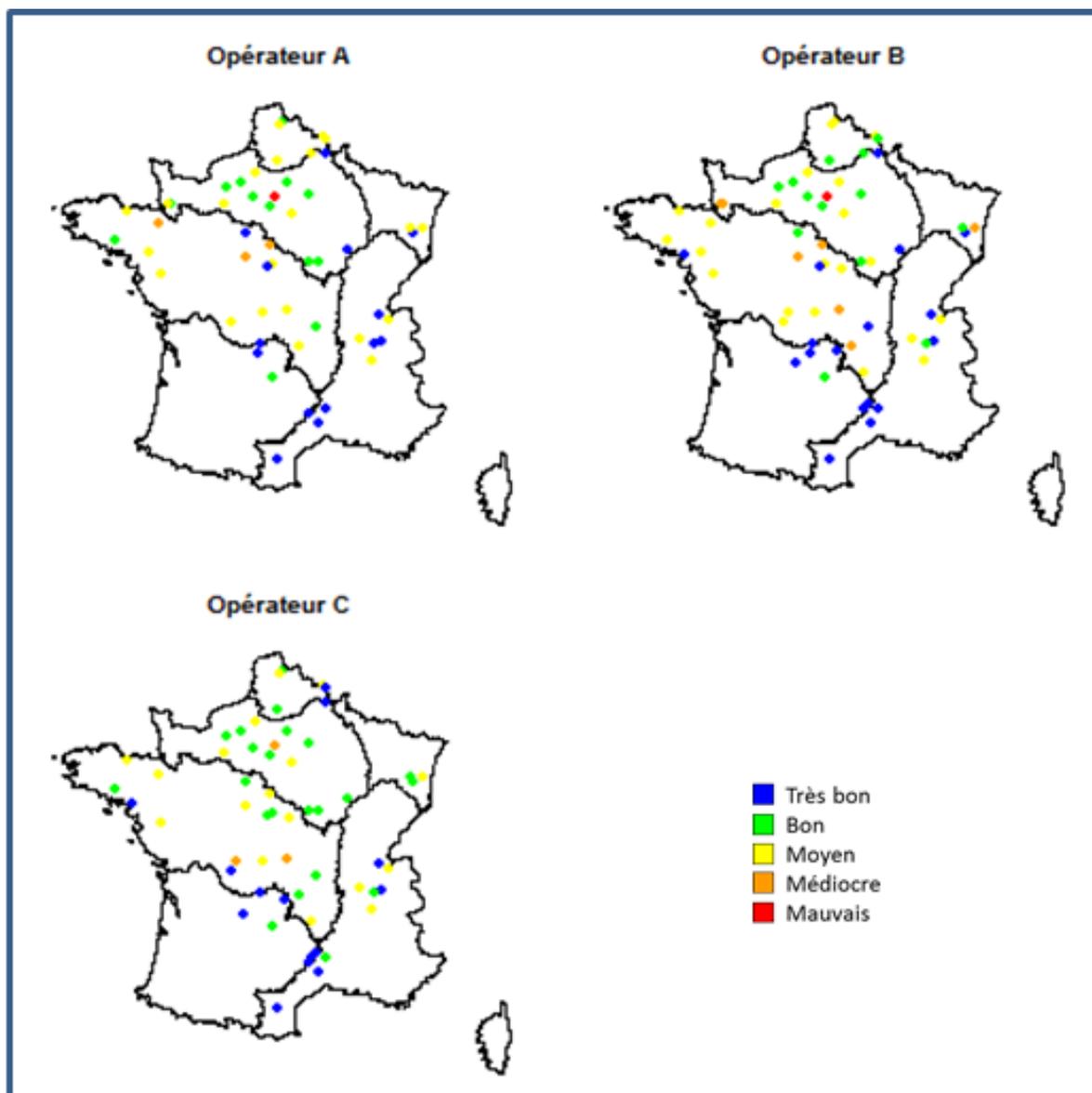


Figure 18 : Cartes d'état écologique par opérateur

Les 2 bassins Loire-Bretagne et Artois-Picardie, aux eaux peu minéralisées et sujets à l'eutrophisation, voire à des pollutions toxiques (Bassin AP), atteignent assez rarement le bon état, alors que le bassin Seine-Normandie, avec des eaux plus minéralisées et carbonatées, l'atteint plus souvent.

Les points du bassin Rhône-Méditerranée font aussi l'objet d'une évaluation proche par les 3 intervenants, un point particulier pouvant être évalué ici ou là avec une classe d'écart. Une divergence plus importante d'évaluation se dégage sur les 3 points du bassin Rhin-Meuse, avec un écart systématique d'une voire 2 classe de qualité (problème de localisation de site par certains intervenants, ou problème lié à des faciès prospectés différents ? Flores de ce bassin moins bien maîtrisées par certains des intervenants ?). Il en a été de même sur 2 points situés sur la Loire et l'Allier sur leur cours intermédiaire (possible erreur de localisation de site ? de rive prospectée ?).

La distribution des proportions de classement d'état écologique assise sur les 173 relevés validés disponibles a déjà été détaillée dans le texte en page 47 concernant l'exploitation de la Figure 17. Si l'on sélectionne uniquement les sites disposant de trois relevés, la variabilité de la classification d'état va un peu augmenter par rapport à ce qu'elle est sur la base de 2 relevés. La convergence d'évaluation va donc forcément diminuer un peu, ce qui est logique.

Sur la base des relevés disponibles en triplets, 46,7 % des sites sont classés par les trois opérateurs dans la même classe d'état écologique. 44,4 % des sites présentent un écart d'une classe d'état, ce qui est le plus souvent le signe d'un site à statut écologique assez proche d'une limite de classe (gros aléa de classification dans ce cas), mais peut aussi recouvrir dans quelques cas un écart de pratique un peu plus conséquent entre 2 opérateurs et le 3^{ème}.

Des raisons pouvant expliquer des écarts sensibles inter-opérateurs, pouvant parfois conduire à un écart d'évaluation de plus d'une profondeur de classe entre 2 intervenants et le 3^{ème}, ont parfois été repérées.

Les faciès prospectés ont pu être assez sensiblement différents en fonction d'une signalétique de station perfectible et de la complexité de terrain, notamment sur des hydrosystèmes d'assez grande taille aux caractéristiques morphologiques et hydrodynamiques fortement variables selon localisation de la station par l'intervenant (Eure, Lys canalisée, Allier...).

Plusieurs sites, qui présentaient des conditions locales sensiblement artificialisées ou très particulières, ont requestionné chaque intervenant de terrain vis-à-vis de la bonne application de la norme (priorisation du type de faciès et du type de substrat par rapport à la situation terrain, par exemple berge artificialisée *versus* rive naturelle, prélèvement au racloir sur paleplanches *versus* prélèvement classique sur substrats durs rocheux etc...). De tels sites laissant place à plusieurs éventualités de pratique ont ainsi pu être échantillonnés de façon différente selon l'intervenant et donner des résultats plus variables. Enfin, 8,9 % des sites (soit 4 sites) présentent un écart important, atteignant 2 classes d'écart, entre les évaluations réalisées. Ces écarts anormaux ont certainement été liés à un problème assez important ayant fait diverger la pratique et/ou l'échantillonnage entre les 3 opérateurs.

Pour le 4^{ème} site, au vu des coordonnées géoréférencées de la feuille de terrain, il semble que l'opérateur C ait prélevé à une cinquantaine de mètres de distance des 2 autres, et la composition floristique de son prélèvement a été sensiblement différente (très forte abondance de l'espèce *Achnanthydium pyrenaicum*, qui a conduit à une meilleure note d'indice).

Ce fort écart traduit peut-être une forte hétérogénéité de ce site, de la façon de réaliser le prélèvement de terrain, et/ou l'existence d'influences parasites de bordure d'incidence inégale...

La valeur absolue des écarts à la moyenne obtenue pour chaque relevé (cf Figure 16), qui semble la façon la plus raisonnable et robuste d'exprimer la variabilité inter-opérateur (*i.e.* qui se réfère à l'écart statistique de chaque opérateur par rapport à la note moyenne obtenue par tous les opérateurs sur un même relevé), peut également être exprimée en fonction du classement d'état écologique du relevé (voir Figure 19 ci-dessous).

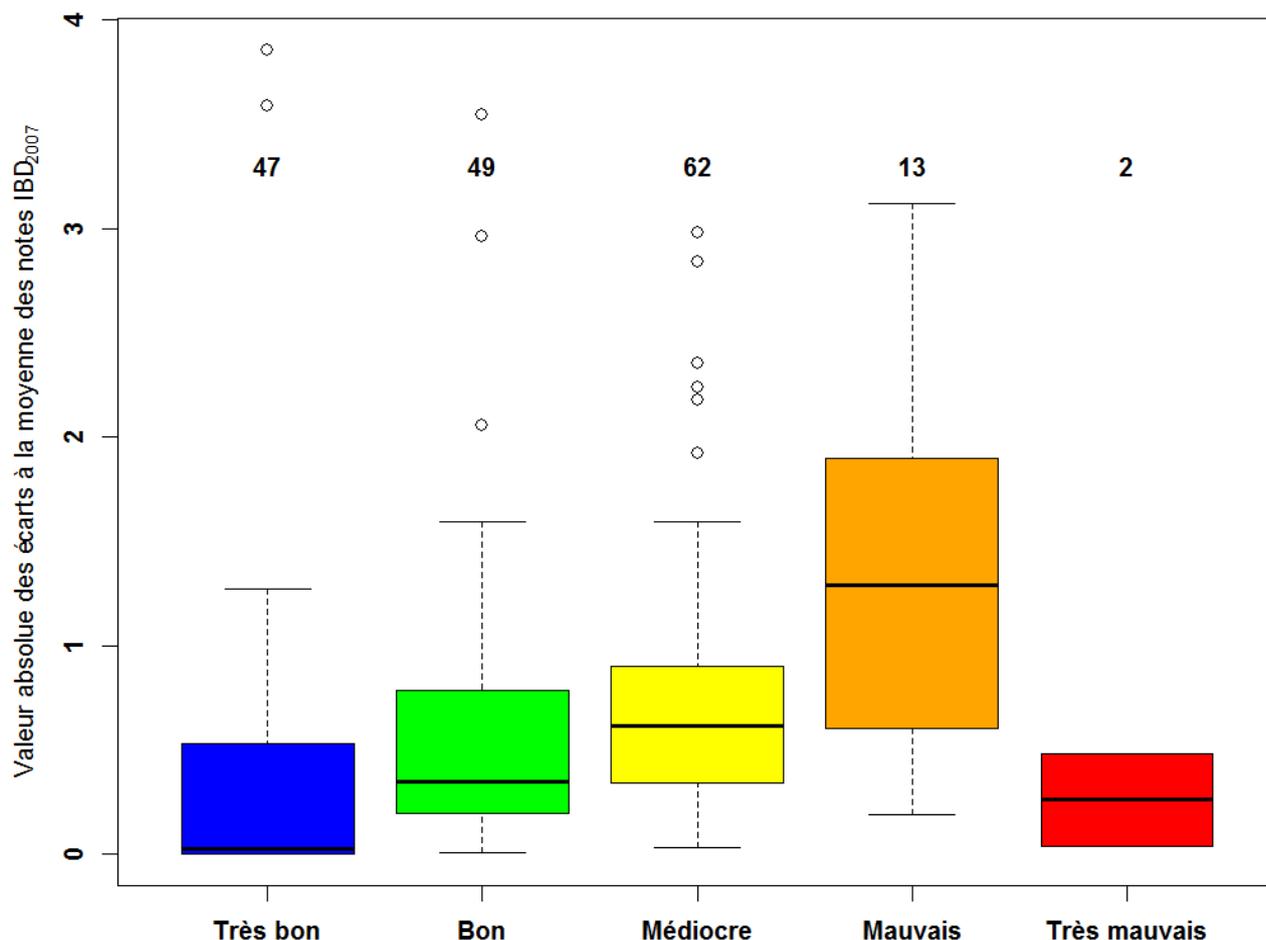


Figure 19 : Valeur absolue des écarts à la moyenne des notes IBD₂₀₀₇ par classe d'état

Nb : Les chiffres en gras représentent le nombre de relevés par classe.

Les nombres de relevés par classe d'état obtenus au cours de cette étude sont relativement inégaux, mais ils correspondent assez bien à la réalité statistique d'état écologique tel que représentée actuellement dans les réseaux de surveillance.

En effet, compte-tenu de la stratégie nationale promouvant la collecte des rejets en tout-à-l'égout et les stratégies d'assainissement collectif, la plupart des sites antérieurement les plus pollués (aval de municipalités, aval de grosses STEPs, de grosses industries...) ont fait progressivement l'objet de stratégies d'investissement visant à re-dimensionner l'équipement collectif et/ou à améliorer la performance du procédé utilisé, avec l'aide des Agences de l'Eau, permettant progressivement de résoudre les principaux points noirs péri-urbains ou industriels et de réduire efficacement, au niveau national, la statistique de points gravement impactés.

Il est par contre plus difficile de restaurer l'état écologique sous l'influence de pollutions diffuses de plus large échelle, notamment d'origine agricole ou domestique plus diffuse, qui procurent des altérations moins intenses et se traduisent par une dominance de sites dans la classe d'état moyen, ou avoisinant la limite entre état moyen et bon état.

Compte-tenu de ce qui précède, notre jeu de données d'étude constitue une photographie assez représentative de l'état relatif des stations dans les réseaux nationaux de surveillance. En effet,

les effectifs de relevés par classe d'état y sont relativement déséquilibrés et présentent une forte sous-représentation des classes d'état médiocre (couleur orange) et de mauvais état (couleur rouge), en bonne correspondance avec ce qui est actuellement observé à l'échelle nationale. Cependant, en fonction de cet état de fait, il ne serait pas très robuste d'entrer dans une interprétation trop approfondie des résultats, et en particulier sur ces classes sous-dotées.

Au niveau de la tendance générale, plusieurs critères (niveau de variabilité médiane par classe de qualité, profondeur du box-plot par classe, occurrence de valeurs outliers avec forts écarts...) convergent pour nous indiquer que la variabilité inter-opérateurs est statistiquement plus réduite pour les classes de très bon état et de bon état que pour les classes plus altérées. Il s'agit probablement d'un constat basé sur une réalité (plus un site est considéré comme de bonne à très bonne qualité, plus il risque d'être homogène au niveau de ses micro-habitats, de sa qualité des eaux et de la qualité de ses flores, et moins les prélèvements et analyses de relevés effectués par divers opérateurs ont des chances de diverger au niveau de la note d'IBD et du classement de qualité).

De fait, la plus grande étendue des box-plots d'Etat Moyen (jaune) et d'Etat Médiocre (orange), ainsi que l'augmentation de leur valeur médiane, semblent confirmer un accroissement des variabilités regroupées sous l'effet inter-opérateur sur des sites plus altérés (probablement en correspondance avec une variabilité plus forte de la qualité des flores selon les micro-habitats de la station).

Enfin, la classe de mauvais état étant très fortement sous-dotée en relevés, donc aussi en représentativité de la distribution des variabilités constatées par rapport à l'étendue des possibles, il n'est pas possible d'en tirer d'autres enseignements que la relative convergence factuelle du calcul d'IBD sur les 2 relevés ayant conduit à cette évaluation en mauvaise qualité (seulement 0,4 point d'écart à la moyenne, respectivement, dans le cadre du triplet mis en correspondance, qui comportait 2 relevés en Mauvais Etat et un relevé en classe d'Etat Médiocre). Ces faibles valeurs d'écarts illustrent simplement que ces 3 relevés ont donné des valeurs d'IBD proches entre eux, et que l'état du site était proche de la limite de basculement entre ces 2 classes.

2.8. Discussion - conclusion du volet inter-opérateur

En fonction du temps nécessaire pour rapatrier tous les résultats et homogénéiser les éléments envoyés sous une forme multisupports disparate, la totalité des données des campagnes 2013 et 2014 a pu être bancarisée sous une forme standardisée au 2ème semestre 2015. Les analyses d'étape présentées dans le rapport intermédiaire de fin 2015 avaient permis une première approche de la notion de variabilité inter-opérateur des notes d'IBD₂₀₀₇, qui est étudiée de façon plus complète et plus aboutie dans le présent rapport final.

La présente discussion passe en revue différents aspects ayant pu influencer la qualité de la donnée collectée et la variabilité inter-opérateurs.

Certains aspects, repérés au cas par cas, sont évoqués de manière **plutôt qualitative**, sur la base de cas d'espèce effectivement rencontrés en cours d'étude et cités en exemples. S'ils ont pu jouer une influence ponctuelle sur les écarts concernant certains triplets de données, le poids qu'ils exercent sur la variabilité globale inter-opérateurs est plus difficile à quantifier. Cependant, ils contribuent à leur niveau à générer certains écarts indésirables, et leur résolution ne constitue pas une quantité négligeable vis-à-vis de la qualité globale des référentiels obtenus. Celle-ci passe

à la fois par l'amélioration de certains documents aidant à trouver les bons sites de façon plus reproductible en inter-intervenants (ex : amélioration de certaines fiches signalétiques de stations) et par l'amélioration des pratiques d'un intervenant donné (ex : saisie plus systématique de données d'accompagnement, meilleure prise en compte des aspects normatifs relatifs à la réalisation de l'échantillonnage de terrain, du comptage, amélioration de la qualification et de l'expertise-maillon individuelle, etc...).

D'autres aspects de variabilité exploités sur le **plan quantitatif**, basés sur l'analyse des résultats procurés par les relevés biologiques prélevés et analysés séparément par les 3 catégories d'intervenants, rentrent réellement au cœur de l'exercice d'intercomparaison inter-opérateurs réalisé.

Ces exploitations, réalisées à l'échelle du jeu de données complet d'étude, permettent au final de quantifier un niveau d'incertitude global inter-opérateur attaché à un résultat d'évaluation biologique au relevé et de le propager sur l'évaluation de qualité écologique d'un site donné.

2.8.1. Aspects qualitatifs :

Un premier aspect important pour faciliter la réalisation reproductible d'une opération de prélèvement est la mise à disposition aux participants à cette étude ou aux opérateurs de réseaux d'une **signalétique de station satisfaisante**, c'est-à-dire nécessaire et suffisante pour permettre à un opérateur ne connaissant pas la station par avance de la trouver et d'y trouver des repères pérennes dans le paysage aidant à conforter rapidement son bon positionnement.

En effet, en fonction d'impératifs organisationnels, il faut avoir en tête que ces interventions de terrain se font habituellement en contexte temporel tendu.

Toute facilitation permettant de conforter l'intervenant de terrain vis-à-vis de sa localisation de site sera donc susceptible de limiter, voire d'éviter presque totalement les erreurs de lieu d'échantillonnage, qui n'ont pas grand-chose à voir avec la précision de la méthode de bio-indication diatomique.

Dans le cadre de cette étude, il était prévu que les DREAL fournissent à l'avance des fiches-stations permettant aux autres intervenants de trouver facilement la station, ce qui a été bien fait dans la plupart des cas, et avec l'anticipation suffisante pour permettre l'intervention de terrain dans le schéma d'organisation temporelle propre à chaque type de participant.

Un exemple d'écart lié à une réception trop tardive de fiche-station a résulté d'une différence d'appréciation de l'application de la norme de prélèvement par les 3 opérateurs en cause, qui ont commencé par analyser comment appliquer au mieux la norme de prélèvement en fonction des caractéristiques du site.

En vue de l'homogénéisation de la pratique sur une station présentant de des particularités pour le prélèvement, il serait utile de préciser sur la fiche-station correspondante le **type de prélèvement spécifique** qui doit y être appliqué pour une bonne reproductibilité des opérations de suivi. Ce serait aussi le cas, par exemple, quand le prélèvement doit se faire sur piles de pont ou substrats bétonnés, sur tiges de macrophytes, ou autres substrats un peu inhabituels imposés par la situation locale et décalés par rapport à la pratique normative la plus classique (ces autres cas d'espèce n'ayant pas été rencontrés dans le cadre de la présente étude).

Sur le plan de la qualité des éléments informatifs sur les stations, la plupart des signalétiques et/ou des supports envoyés étaient assez explicites et se sont révélés satisfaisants à très satisfaisants pour l'usage.

Cependant, un petit nombre de signalétiques (4 ou 5 sur 61 stations, soit entre 6 et 8 %) seraient très aisément perfectibles dans l'optique de facilitation indiquée précédemment, ce qui peut indiscutablement participer à la qualité générale de la donnée collectée et à la limitation d'erreurs inter-opérateurs accidentelles (qui peuvent ensuite se répercuter sur une variation interannuelle quand les opérateurs ont changé du fait d'un changement de marché public).

A titre de simple exemple, une fiche-station plutôt localisée en partie Nord de la France s'est révélée particulièrement confondante, voire a pu être génératrice d'erreurs (ce qui va à l'inverse de l'objectif poursuivi par la fourniture de signalétiques-stations) par manque de qualité et d'actualisation des supports visuels. Elle présentait comme un élément caractéristique et central du paysage un pont mobile (aspect non précisé sur la fiche-station) en structure métallique rivetée (Figure 20 ci-dessous).



Figure 20 : Exemple de signalétique de station à réactualiser (code de site anonymé)

Or, sauf erreur toujours possible des préleveurs de terrain, en fonction de la qualité médiocre et de la pixélisation de ce cliché n'aidant pas à la prise de bons repères de terrain, ce pont aurait été démolé et enlevé, et n'existerait plus dans le paysage depuis 15 à 20 ans (information obtenue en dernier recours auprès des services municipaux de l'agglomération après une importante perte de temps à sillonner le secteur).

Pour un intervenant ne connaissant pas le site et son historique, une telle photo, si elle présente un ouvrage qui n'existe plus, n'est pas une aide mais au contraire, une information confondante, cette difficulté ayant été amplifiée par la qualité plutôt moyenne du cliché et par le fait que, dans le périmètre d'1 ou 2 kilomètres autour de cet endroit ou le pont a été enlevé, au moins 5 ponts rivetés sont encore présents sur ce cours d'eau ou sur le réseau de canaux jouxtant immédiatement ce cours d'eau, voire connectés à ce cours d'eau. Enfin, cette photo de qualité très perfectible présentait des éléments de paysage qui ont beaucoup évolué depuis (grands arbres au bord de la rivière, des bâtiments ont été rasés et d'autres construits, l'état de la voirie présentée a beaucoup changé).

Ce site présenté sans indication de la berge à échantillonner et, qui plus est, sans précision du type de substrat habituellement prélevé (paleplanches ou au contraire recherche de substrats naturels) est l'un de ceux ou les 3 opérateurs, à l'évidence, n'ont certainement pas procédé à des échantillonnages comparables.

Par rapport au repérage de stations, malgré l'amélioration progressive de l'environnement national « données », un premier problème qui peut se poser est que les signalétiques présentes sur le site du Sandre sont uniquement décrites en texte et carte, sans photos. Un second est que plusieurs systèmes de repérage des coordonnées géographiques peuvent cohabiter selon l'origine de l'information, les fiches-stations fournies n'étant déjà pas toutes homogènes de ce point de vue, selon les bassins ou entités les ayant constituées, et un tout petit nombre d'entre elles (2 ou 3) étant manifestement douteuses.

Enfin, les systèmes de coordonnées géographiques des différents dispositifs permettant de se repérer sur le terrain (GPS routier, GPS de terrain, GPS d'appareil photo ...) ne sont pas forcément dans le même système de géoréférencement que la fiche, selon le fabricant et son pays d'origine.

Ces hétérogénéités génèrent donc différentes difficultés pratiques face à laquelle l'argument « il a été fourni les coordonnées géoréférencées du site, c'est une information suffisante pour bien localiser la station », bien qu'acceptable dans l'absolu, peut aussi contribuer, par manque de facilitation, à générer des erreurs évitables qui se répercuteront dans la qualité des référentiels sans nécessairement pouvoir être repérées et mises à l'écart *a posteriori* (c'est-à-dire pendant l'étape de validation des données de terrain).

En pratique, les opérateurs n'ont souvent pas beaucoup de temps pour étudier à l'avance les fiches-stations en détail, au-delà de diagnostiquer rapidement en travers si les éléments fournis leur semblent suffisants *a priori* pour pouvoir localiser et reconnaître la station. De plus, dans le cadre de l'étude, certaines signalétiques de stations ont été envoyées avec trop peu d'anticipation pour permettre cet examen (réception le dernier jour de la semaine précédant la mission, voire en cours de la semaine de mission, opérateurs déjà itinérants).

Ce n'est donc qu'une fois arrivés sur le terrain que les opérateurs pourront réellement juger si la fiche permet de trouver facilement la station, s'il se pose une divergence de système de coordonnées géographiques entre certaines fiches et leur appareil, ou s'il y a un problème éventuel de coordonnées sur la fiche-station... Enfin, les coordonnées géographiques peuvent ne pas être suffisantes pour indiquer l'accès le plus facile vers le site (celui qu'emploient habituellement les opérateurs-terrain déjà familiarisés avec ce site), voire pour garer le véhicule en opération.

Une bonne fiche-station est donc plus qu'un élément facultatif, c'est un élément important qui participe à la facilitation de la réalisation conforme et rapide de la mission de terrain, sachant que celle-ci est souvent exécutée sous contrainte de temps par des intervenants économiques variés au fil du temps, au gré de l'attribution des Marchés. L'établissement d'une bonne fiche-station et son amélioration en fonction de retours-utilisateurs n'est donc pas un élément négligeable dans la génération ou l'évitement d'erreurs de localisation.

Il apparaît important qu'outre les coordonnées géographiques qui, en fonction de ces aléas de systèmes référentiels, ne permettent quelquefois qu'un contrôle *a posteriori*, figurent aussi d'autres éléments concrets de repérage réellement utilisables par l'intervenant pour faciliter l'approche et obtenir rapidement une confirmation de la bonne localisation de la station.

Les repères solides et durables, lorsqu'ils sont donnés, jouent ce rôle et sont précieux pour acquérir rapidement une certitude (pont sur telle rivière, sur telle voie, voire dans la traversée de telle agglomération ou village, station à l'amont ou à l'aval de l'ouvrage, à telle distance, sur quelle rive pour l'accès etc...).

Les petits extraits cartographiques fournis dans bon nombre de fiches sont aussi intéressants dans ce but, même s'il n'est pas toujours aisé de repérer leur localisation dans une carte beaucoup plus large, ni de bien appréhender l'échelle de distance dans un support qui ne la présente pas et ne la respecte plus (zoom plus ou moins important d'un petit extrait de carte).

Certaines photos signalétiques ne présentant pas de repères solides dans le paysage peuvent prêter à confusion, voire contribuer à provoquer des erreurs. Par exemple, les photos centrées sur de la ripisylve, sur tel ou tel arbre caractéristique etc... sont très évolutives en bordure de rivière (rivière initialement dégagée suite à une coupe qui se recouvre en 5 à 8 ans, ou au contraire rivière recouverte d'un corridor arboré que les riverains coupent à certains moments, aspect hiver-été sur de tels types de repères végétaux, etc...).

Pour rendre les signalétiques plus utiles et opérationnelles à peu de frais, il serait important de faire le postulat que l'opérateur ne connaît pas *a priori* le site et que la fiche doit l'aider à le trouver facilement et avec certitude (elle est surtout faite dans cet objectif et son utilité est maximale dans ce cas).

En effet, dans le cadre des réseaux de surveillance, où des intervenants de terrain venant parfois de loin ont souvent une contrainte de temps, l'erreur de site n'est pas complètement à exclure pour un intervenant n'y étant jamais venu auparavant, et risque de se rajouter inutilement à une variabilité-opérateur classique.

La présente étude d'intérêt national témoigne de cette réalité car elle a visiblement comporté un petit nombre de situations (entre 4 et 6) où les signalétiques fournies pour les stations pouvaient être gradées de peu explicites à plutôt confondantes. Or c'est souvent dans de tels cas que 2 intervenants en combinaison aléatoire ont prélevé un assemblage visiblement très différent du troisième, générant des écarts sensibles d'assemblages et de notes.

Les quelques fiches les moins explicites étant très facilement perfectibles, il serait possible à peu de frais de réduire fortement ce type de risque et d'améliorer ainsi la qualité des référentiels collectés dans les réseaux ou dans de telles études, en repartant du principe de base que la fiche signalétique de terrain est essentiellement destinée à aider un intervenant lambda à trouver avec un bon degré de certitude un site de prélèvement qu'il ne connaît pas.

2.8.2. Réalisation et rapportage des opérations de terrain :

Pour quelques **fiches de notation de terrain**, comme déjà évoqué, certaines données (géographiques, physico-chimiques, substrats...) n'ont pas été systématiquement renseignées malgré l'enjeu collectif de l'étude nationale et les recommandations organisationnelles données. Certains croisements d'informations ont donc fait défaut pour permettre d'évaluer de façon objective la bonne comparabilité des données inter-opérateurs collectées au même site, notamment en cas d'écart de dates de visite, d'assemblages biologiques apparemment assez différents, de survenue d'un épisode hydrologique...

Ces données de contrôle de contexte auraient d'une part pu être utiles pour aider à « racheter » certains relevés acquis avec un seuil d'écart de temps important (plus de 2 semaines), mais dont les éléments abiotiques et biologiques collectés indiquent plutôt une stabilité des conditions et résultats. Dans quelques cas, comme il n'était pas possible de vérifier la comparabilité des conditions abiotiques rencontrées par les 3 types d'intervenants en cas de dépassement du seuil de décalage temporel, un seuillage automatique basé sur ce critère a été appliqué, limitant ainsi l'assise de triplets de données éventuellement « rachetables » et conduisant malheureusement à une mobilisation affaiblie de certains relevés effectués, faute de possibilités simples de recroisement autorisant certaines consolidations ou validations *a posteriori*.

D'autre part, leur mobilisation aurait aussi pu être précieuse pour diagnostiquer plus avant certains écarts d'évaluation d'ampleur *a priori* anormale pour les praticiens réguliers de la méthode IBD.

Or, dans quelques cas où une évaluation de qualité d'un site divergeait assez fortement des 2 autres, suggérant assez fortement une erreur de localisation du faciès à échantillonner, voire du site ou même peut-être du cours d'eau, il a souvent manqué de façon décroisée entre les 3 intervenants, soit les coordonnées géoréférencées acquises sur le terrain, soit la physico-chimie de l'eau (la conductivité électrique, mesure très robuste, étant ici un descripteur particulièrement utile pour juger de la comparabilité de l'hydrochimie associée à chaque prélèvement), empêchant de faire les recoupements nécessaires et d'acquiescer une certitude sur la réalité et l'origine du problème rencontré dans ces cas un peu extrêmes.

L'enjeu national de cette étude, qui ne visait pas à l'application normative minimale de la méthode IBD dans le cadre d'application routinière d'un réseau, mais à **établir et conforter l'estimation de la variabilité inter-opérateurs de l'évaluation d'état diatomique à partir de l'IBD₂₀₀₇**, en appui au domaine réglementaire, a pu être mal capté par certains intervenants.

Il est donc arrivé qu'ils ne mesurent pas la physico-chimie de terrain au prétexte qu'elle est facultative dans le cadre d'application des réseaux, les opérations normatives de **prélèvement biologique** et de **mesure physico-chimique** ne représentant pas la même opération élémentaire et n'étant pas nécessairement réalisées ensemble et par les mêmes opérateurs au même site dans la pratique des réseaux.

Or la réalisation couplée des 2 opérations avait ici pour but de mieux décrire les conditions locales d'accompagnement au moment du prélèvement, et de se donner quelques moyens de valider ou infirmer la comparabilité de résultats d'inventaires conduisant à des différences sensibles d'évaluation : ces écarts pouvant être par exemple sous l'influence d'une variabilité temporelle, de la survenue d'un épisode hydrologique interposé entre des relevés, d'un écart de de faciès de prélèvement, voire d'une erreur de site... (liste non exhaustive).

Dans ces conditions un peu trop soumises à des manques de descripteurs d'accompagnement, il a donc dû être appliqué de façon parfois abrupte, sans expertise complémentaire, une fenêtre temporelle maximum de 2 semaines pour pouvoir intercomparer des relevés, le but de cette mesure étant d'éviter d'amplifier l'effet inter-opérateur par un effet de variabilité temporelle. D'autre part, même en cas d'écart un peu anormal d'évaluation procurée par des relevés au même site, les seuls relevés invalidés pour l'inter-comparaison ont été ceux pour lesquels une erreur et son origine étaient établies sur une base objectivée. Pour d'autres relevés douteux produisant des évaluations atteignant parfois 2 classes d'écart (7 à 8% des relevés), faute d'éléments tangibles matérialisant la réalité du problème rencontré, ces relevés ont été conservés et ont participé à l'établissement de l'intervalle de confiance de la méthode d'évaluation basée sur l'IBD₂₀₀₇.

Cette façon de procéder apporte la garantie que l'incertitude globale de cette méthode n'a pas été sous-dimensionnée, et qu'elle peut être considérée comme robuste.

2.8.3. Aspects quantitatifs (inventaires, évaluation) :

Les caractéristiques habituelles de la distribution des taxons diatomiques dans les assemblages naturels (cf. Figure 12), constitués de quelques taxons dominants en occurrences et en abondances relatives, puis d'une très longue queue de distribution de détection-comptage beaucoup plus aléatoire entre types d'intervenants comme en répétition de l'opération par le même intervenant, marquent sensiblement les résultats des analyses comparatives faites dans le cadre de la présente étude inter-opérateurs.

Ainsi, les comparaisons de résultats basées sur la **présence-absence des taxons dans les assemblages** (Figure 6, 7 et 8) et sur la comparaison de **richesse spécifique** entre les relevés des 3 types d'opérateurs (Figure 11) conduisent à des différences sensibles liées à la composition des assemblages. C'est un état de fait normal, en fonction de l'aléa de prise en compte ou non de ces taxons peu représentés avec la pression d'observation appliquée (400 individus), qui font diverger largement les relevés sur le plan de leur composition spécifique (seulement 33,3% de taxons en commun entre les 3 opérateurs).

Cependant cette forte dissimilarité taxonomique qui préside à la composition des assemblages diatomiques naturels ne conduit pas systématiquement à des écarts importants de note d'indice. En effet, des communautés très dissimilaires sur le plan des taxons constitutifs et de leurs effectifs en abondances relatives (cf. Figure 15) peuvent se révéler très homogènes du point de vue de l'indication de qualité des eaux qu'elles procurent et correspondre à des notes d'indices comparables. En revanche, de forts écarts d'indice correspondent toujours à des communautés très différentes.

Les analyses mettant en jeu l'**abondance relative des taxons** vont déjà donner des résultats sensiblement plus comparables, puisque les espèces dominantes d'un assemblage seront mieux prises en compte en fonction de leurs effectifs et contribueront à une meilleure convergence des résultats inter-opérateurs (cf. Figure 9).

L'intervention des abondances relatives se rapproche du mode de construction de l'outil de bio-indication IBD₂₀₀₇, dont la note résulte d'un calcul accordant notamment un poids au profil de qualité proportionnel aux effectifs de chaque taxon contributif à cette méthode. Aussi, malgré des résultats de richesse des assemblages parfois significativement différents entre les

intervenants (cf. Figure 11, 1^{er} bloc), **les résultats exprimés en notes d'IBD₂₀₀₇ révèlent l'absence de différence significative globale** en inter-opérateur.

Au final et compte-tenu de ces principes de sélection prudente des relevés, la médiane des écarts à la moyenne entre opérateurs pour un même site s'établit à **0,55 point d'IBD₂₀₀₇** sur l'ensemble de ce jeu de données.

55% des intercomparaisons conduisent à une évaluation homogène de qualité. Dans 45% des cas, cet écart-standard s'accompagnait d'une divergence de classe d'état écologique chez au moins un opérateur. La plupart du temps, ce phénomène reste dans le domaine du normal, du fait qu'un bon nombre de stations ne sont pas centrées sur une classe de qualité mais sont proches d'une limite inter-classes (beaucoup de sites de ce jeu de données, comme du RCS national, étant notamment proches de la limite Bon Etat –Etat Moyen).

Cependant, dans quelques cas, une différence d'évaluation d'une classe de qualité au site peut cacher des écarts d'évaluation d'une profondeur proche de 2 classes entre les 2 relevés donnant les écarts les plus forts. Enfin, les écarts générés sur l'évaluation se sont parfois étendus jusqu'à 2 classes d'état (dans 7 à 8 % des cas), laissant le plus souvent apparaître un relevé douteux par rapport aux 2 autres et induisant un doute sur la bonne comparabilité des pratiques inter-opérateurs dans ce cas.

Au final, l'assise de relevés qui a été utilisée pour l'analyse des écarts inter-opérateurs est réaliste pour évaluer la confiance d'un résultat obtenu dans le cadre des réseaux de surveillance de routine, à savoir qu'elle n'est pas sous-dimensionnée et doit amener un peu plus de robustesse à l'évaluation réalisée.

3.1. *Problématique*

En complément de l'évaluation de l'incertitude de l'IBD au relevé lié à l'effet inter-opérateur global, étudiée au chapitre précédent, un autre aspect de variabilité qui mérite d'être cerné pour aborder l'incertitude d'évaluation d'état sur un site ou un tronçon donné de cours d'eau est l'évaluation de la **variation inter-annuelle de la note d'IBD** en liaison avec la **variabilité des conditions naturelles**. En effet, même dans des conditions de milieu non ou faiblement anthropisées, la variation interannuelle des conditions abiotiques qui s'exercent sur le cours d'eau (année chaude *versus* année à températures plus modérées, année à tendance sèche *versus* année à pluviométrie abondante) est susceptible d'exercer, dans les mêmes environnements naturels ou anthropisés, une variabilité naturelle de la réponse des métriques biologiques (voir bloc en bas à gauche de la Figure 1).

Il est donc important d'appréhender cette variabilité d'origine naturelle et de cerner dans quelle fourchette elle peut contribuer à faire varier l'évaluation de qualité attribuée au même site de réseau sans changement particulier de la pression d'anthropisation. En effet, cette variabilité d'origine naturelle doit être correctement cernée afin d'être répercutée sur l'intervalle de confiance de la méthode et participer ainsi à la fiabilisation de l'évaluation d'état écologique réalisée sur le site.

Ce volet d'étude n'a pas fait appel à des campagnes expérimentales spécifiques mais s'est appuyé sur des **données historiques des réseaux de surveillance** progressivement bancarisées dans la base de données Pandore.

L'avantage de cette façon de procéder a été de pouvoir s'appuyer sur l'étude temporelle de chroniques déjà acquises, qui permettait d'en tirer assez rapidement des éléments de variabilité et de rendre les résultats de l'étude dans un délai raisonnable après sa contractualisation.

Un inconvénient d'ordre qualitatif vis-à-vis des données mobilisées est qu'il est parfois difficile d'effectuer une validation a posteriori de certains paramètres, de certaines unités employées, d'unités non renseignées ou de valeurs paraissant anormales.

Il a de plus été nécessaire de procéder à un post-traitement adéquat des descripteurs arrivant en-dessous des seuils analytiques de détection ou de quantification (les méthodes multivariées ne s'accommodent pas de cases vides au contenu vide ou non quantitatif).

Un autre inconvénient non-négligeable, portant cette fois sur le jeu de données biologique, est lié à l'évolution dans le temps de l'outil indiciel (IBD₂₀₀₀ puis IBD₂₀₀₇), dont la liste taxonomique des taxons contributifs a changé très sensiblement entre les 2 versions (elle est passée de 216 à 846, et les formes tératologiques ont été introduites dans le calcul indiciel), ainsi que l'environnement normatif correspondant (Norme AFNOR révisée 2007).

Ces changements, ainsi que la capitalisation progressive par beaucoup d'opérateurs de nouveautés taxonomiques apparues dans les documents d'accompagnement (guide IBD, atlas des flores régionales...), ont pu jouer une influence non-négligeable sur le contenu des bases de données d'inventaires capitalisées dans le cadre des réseaux de surveillance successifs.

Préalablement aux requêtes et à la sélection de relevés, il a donc dû être mené une réflexion spécifique afin de pouvoir quantifier une variabilité interannuelle sur une base suffisamment comparable pour en répercuter l'incidence de façon adéquate vis-à-vis de l'objectif d'évaluation.

3.2. Assise de données, méthodologie

Afin d'étudier la variabilité interannuelle des notes d'IBD sous influence de la variabilité des conditions naturelles sur des sites donnés, il est nécessaire de travailler sur des chroniques suffisamment longues permettant à la fois de caractériser les données de chimie et de s'assurer de la stabilité de la qualité abiotique des milieux aquatiques au site, puis de dégager la variabilité des notes d'IBD obtenues en fonction de cette variabilité naturelle.

Le but étant d'isoler la part de variation naturelle dans la notation d'IBD, il faut autant que possible éviter de travailler sur une assise de sites dont les variations de qualité seraient principalement dues à des **conditions de pression anthropique très variables dans le temps**. En effet, dans un tel cas, il est normal qu'un outil indiciel fait pour évaluer les effets d'altérations anthropiques procure une réponse variable en fonction de l'intensité de ces altérations, qui n'ont aucune raison de s'exercer de façon stable dans le temps. Et il devient du même coup impossible d'isoler la variabilité d'origine naturelle de celle d'origine anthropique, vu que l'outil indiciel a été conçu en premier lieu pour répondre de façon sensible à l'altération anthropique en fonction de l'intensité avec laquelle elle s'applique.

L'idée de départ était donc de travailler sur des sites naturels à pseudo-naturels suffisamment documentés pour pouvoir s'assurer de leur relative **stabilité hydrochimique**, ce qui aurait permis d'aborder directement la **gamme de variation de leurs notes d'IBD** due à la **variabilité naturelle interannuelle**.

Cette intention initiale, qui aurait nécessité de faire porter le travail de façon dominante ou exclusive sur des **stations de référence** ou **pseudo-naturelles** (*i.e.* faiblement anthropisées), s'est cependant heurtée concrètement à plusieurs écueils pratiques.

Tout d'abord, le premier référentiel d'ampleur intéressante acquis sur sites de référence ou de pseudo-référence au niveau national l'a été dans le cadre du RREF, représentant environ 400 sites suivis pendant 3 ans de suite, dont la plupart étaient situés dans des milieux non ou très peu impactés.

Or ces suivis de sites de référence ont surtout été réalisés entre 2005 et 2007, voire de façon résiduelle sur quelques bassins en 2008 (marchés publics ayant démarré un peu plus tard que les autres).

Une première difficulté rencontrée a été qu'une nouvelle version d'outil indiciel diatomique prenant en compte des modifications conséquentes, l'IBD₂₀₀₇, est entrée en application de routine dans le cadre des réseaux à partir de début 2007, venant créer une discontinuité méthodologique assez importante dans cette première chronique acquise sur réseaux de référence.

D'autre part, les acquisitions biologiques sur le RREF étaient accompagnés de suivis de chimie plus légers que pour les autres réseaux (3 à 4 campagnes d'analyse d'eau par an seulement), permettant difficilement de donner une vision représentative de la chimie au site et de sa stabilité sur le trimestre correspondant à l'échantillonnage biologique. En effet, une seule analyse d'eau, en correspondance plus ou moins bonne avec l'échantillonnage biologique, était en général disponible sur le créneau temporel visé, ne permettant pas du tout de consolider l'assise des données hydrochimiques au site.

Cet aspect n'est pas forcément prédominant pour les nutriments sur ces sites pas ou peu impactés mais, sachant que certains d'entre eux, adossés à des reliefs importants, ont des environnements naturels par nature très instables et à très forte dynamique, il aurait été bon de procéder malgré tout à une sélection visant à éliminer des sites dont la variabilité se révélait excessive.

Enfin, après la durée de vie de ce RREF, le devenir de ces sites de référence a été très variable et, si certains ont été intégrés au RCS ou à des réseaux patrimoniaux de bassin, beaucoup d'entre eux n'ont été suivis qu'épisodiquement voire pas du tout jusqu'à la reprise annoncée de certains d'entre eux dans le réseau de référence pérenne en cours de mise en place. Pour une bonne partie d'entre eux, les chroniques ont été lacunaires à très lacunaires et n'ont pas forcément atteint les seuils quantitatifs de prise en compte pour la sélection de sites stables (voir plus loin).

Afin de rendre possible le travail sur les variabilités interannuelles, il a donc fallu décider d'une adaptation de la stratégie visant à effectuer une sélection de sites dans une gamme plus large de qualités, allant de sites de référence (ceux qui ont fait l'objet de suivis suffisants pour pouvoir être sélectionnés) jusqu'à des sites plus ou moins impactés, mais dont il a été vérifié la **stabilité relative des descripteurs chimiques** et **l'absence de tendance d'évolution de l'état** vis-à-vis de l'altération anthropique.

La Figure 21 en page suivante fait l'état des relevés diatomiques disponibles dans la base Pandore. Cette base a notamment été alimentée avec les prélèvements diatomiques nationaux, dont beaucoup ont été acquis dans le cadre des réseaux de surveillance, ainsi que les données de surveillance chimique qui peuvent leur être associées.

Les plus anciens relevés diatomiques enregistrés datent de 1977. Les plus récents, datant de 2013, ont pu être complétés au niveau de la France entière vers la mi-2016. C'est donc sur cette assise temporelle et quantitative que les requêtes et analyses pouvaient être envisagées.

Sur cette figure, on constate qu'une quantité relativement faible de relevés ont été bancarisés entre les années 1977 et 1995, correspondant à des relevés capitalisés à l'occasion d'études scientifico-techniques diverses et à de premières mises en place expérimentales de réseaux de surveillance diatomique sur certains bassins-pilotes (notamment Rhin-Meuse, Artois-Picardie, RMC). Il s'agissait de campagnes annuelles d'échantillonnage au coup par coup, et non de suivis chroniques.

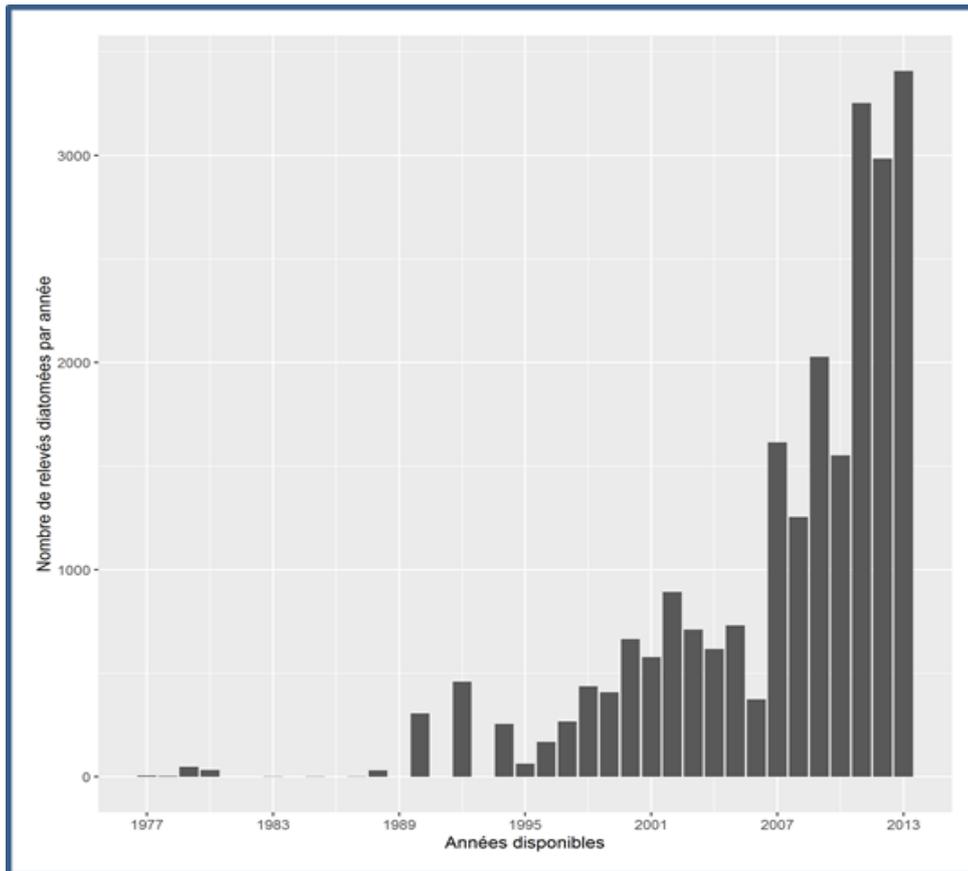


Figure 21 : Nombre de relevés diatomiques disponibles par année dans toute la base Pandore

Entre 1995 et 2007, des données plus nombreuses ont été acquises, correspondant à un objectif de mise en place d'une surveillance régulière dans le cadre de réseaux de surveillance de la qualité des milieux (ex : RNB). Les méthodes de prélèvements sont devenues plus cadrées, l'analyse et le comptage des taxons sont devenus plus reproductibles (matériels d'observation plus performants, taxonomie simplifiée et guide d'aide à l'identification correspondant à l'IBD₁₉₉₆) et permettaient de calculer cet indice, normalisé AFNOR en 2000, sur une base plus homogène.

C'est cependant à partir de 2007 que l'assise nationale de relevés dans le cadre des réseaux de surveillance a atteint sa vitesse de croisière, avec une assise de relevés représentative des cours d'eau de France pour le rapportage européen d'Etat Ecologique (notamment dans le cadre du Réseau de Contrôle et de Surveillance ou RCS, dont c'est l'objectif principal), et selon une

méthode indicielle ayant subi entretemps une évolution sensible (passage à l'IBD₂₀₀₇ appliqué selon la norme AFNOR 2007), dont le cadre d'application est resté stable depuis.

Afin d'obtenir des résultats les plus cohérents entre eux, il a été convenu de commencer à travailler sur des prélèvements et les analyses taxinomiques à partir de 2007. En effet, ce sont ceux acquis selon la même liste d'espèces, sensiblement augmentée, et sur la base d'un guide d'application substantiellement remis à jour à l'époque. Au moment de la réalisation de ces analyses (2^{ème} semestre 2016), les relevés diatomiques étaient disponibles jusqu'à 2013 sur presque toute la France, mais la dernière année correctement couverte au niveau des données de chimie était 2011.

Pour l'obtention de résultats plus robustes, seuls les sites comprenant des relevés effectués sur **au moins trois années différentes de 2007 à 2011** ont donc été retenus. De plus, seuls les prélèvements effectués sur substrat dur et entre les mois de Mai et Septembre (période correspondant plutôt à la saison chaude et à un régime stable de basses-eaux) ont été gardés. Enfin, dans le même souci de travailler sur une base homogène, toutes les notes indicielles ont été recalculées à partir d'une même version d'IBD, en utilisant la même table de transcodage.

Les sites choisis pour la suite de l'étude doivent présenter une composition chimique de l'eau stable et sans tendance d'évolution, afin qu'une forte variabilité d'application d'impacts anthropiques ne soit pas la raison dominante d'une variabilité interannuelle des flores et des résultats de la méthode diatomique. Une étape préliminaire de sélection des sites suivant des critères de stabilité des données environnementales a donc été menée dans la suite de l'analyse.

Le pas de temps d'influence intégrée des conditions hydrochimiques en bonne correspondance avec un cycle de croissance des biofilms naturels est de l'ordre de 2 à 3 mois. D'autre part, plus le créneau temporel visé est court, moins on a de chances de faire intervenir plusieurs analyses d'eau pour conforter les données hydrochimiques et diminuer ainsi l'aspect aléatoire de leur représentativité sur le créneau temporel concerné.

Les données chimiques ont donc été sélectionnées **entre -75 et +15 jours** autour des prélèvements biologiques (durée couverte sur le plan hydrochimique : 3 mois, en période en principe assez stable d'étiage estival).

Dans les cas où plusieurs prélèvements chimiques étaient disponibles dans cet intervalle, la médiane des prélèvements était retenue comme valeur du paramètre (= la moyenne lorsque 2 analyses chimiques étaient disponibles).

Les paramètres chimiques sélectionnés pour déterminer la stabilité des conditions chimiques, certains plutôt représentatifs de l'hydrochimie naturelle de l'eau et la plupart d'entre eux représentatifs du degré d'altération anthropique, sont les suivants :

- Phosphore total (phosTot, mg(P)/L)
- Nitrites (NO₂⁻, mg/L)
- Orthophosphates (PO₄, µg/L)
- Azote Kjeldahl (NKJ, mg de N/l)
- Conductivité électrique (25°C, µS/cm)
- Carbone organique (Corg, mg de C/L), réalisé sur eau filtrée (correspond au COD).
- pH
- Matières en suspension (MES, mg/L)

- Taux de saturation en oxygène (TxSatOx)
- Nitrates (NO_3 , mg/L)
- Ammonium (NH_4 , mg/L)
- Demande biochimique en oxygène (DBO_5 , $\text{mg}(\text{O}_2)/\text{L}$)
- Température (Temp, °C)

Un nettoyage préliminaire des données chimiques a été effectué pour pouvoir réaliser les analyses statistiques :

- sélection des codes remarques
- traitement des valeurs non quantifiées (< limite de quantification) afin de permettre la réalisation de l'ACP, qui ne s'accommode pas de valeurs manquantes (valeurs post-traitées = $\frac{1}{2}$ LQ),
- élimination des doublons.

Pour déterminer la stabilité de l'état chimique des sites sur les différentes années, une Analyse en Composantes Principale (ACP) a été effectuée (Figure 22 ci-dessous).

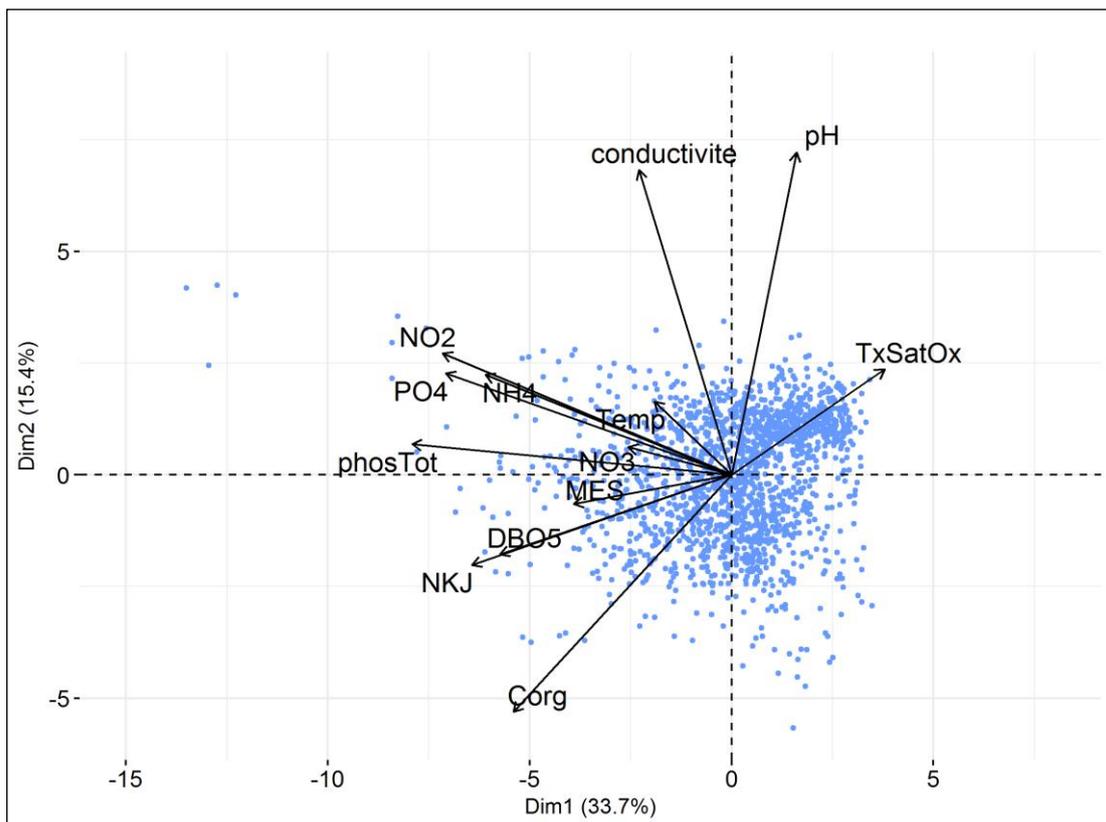


Figure 22 : Analyse en Composante Principale (ACP) de la chimie associée aux relevés diatomiques

Certains paramètres avaient d'emblée une distribution proche d'une loi normale. Cependant, d'autres paramètres (NH_4 , NKJ, carbone organique dissous, DBO_5 , MES, NO_3 , NO_2 , PO_4 , phosphore total) ont dû faire l'objet d'une transformation selon la méthode Box-Cox dans le but de se rapprocher d'une distribution normale pour chaque variable abiotique prise en compte dans l'analyse.

Après transformation des paramètres qui le nécessitaient, l'ACP résultante présente une inertie sur le premier axe d'environ 33%, de 15% sur l'axe 2 et de 10% sur l'axe 3.

L'inertie cumulée sur ces 3 axes principaux étant de 53 % et décroissant très rapidement sur les axes suivants, les 3 premiers axes ont été retenus pour expliquer la structuration du jeu de données et analyser la contribution de chaque paramètre à la formation des axes.

Le premier axe oppose un gradient de type organique et trophique à l'oxygénation de l'eau. L'axe 2 correspond principalement aux données de géochimie-hydrochimie naturelle (faibles conductivités électriques et conditions acides vers le bas, eaux à plus forte conductivité électrique et pH plutôt basique vers le haut de la figure, en liaison avec des substrats plus carbonatés. L'axe 3 (non présenté ici) représente principalement le taux de matières en suspension et la T°C, en correspondance avec le gradient amont-aval des cours d'eau.

Tableau 1 : Contribution des variables de l'ACP à chacun des axes principaux

	Axe 1 (33%)	Axe 2 (15%)	Axe 3 (10%)
Phosphore total (phosTot)	17.77	0.29	1.46
Nitrites (NO2)	14.53	4.58	3.94
Orthohosphates (PO4)	14.23	3.20	1.46
Azote Kjeldahl (NKJ)	11.74	2.52	1.03
Conductivité	1.5	28.89	3.66
Carbone organique (Corg)	8.28	17.46	7.04
pH	0.74	32.31	4.58
Matières en suspension (MES)	4.33	0.26	16.7
Taux de saturation en oxygène (TxSatOx)	4.07	3.46	6.61
Nitrates (NO3)	1.85	0.24	4.3
Ammonium (NH4)	10.55	3.12	12.32
Demande biologique en oxygène (DBO5)	9.34	1.96	10.46
Température (Temp)	1.03	1.67	22.84

Le Tableau 1 ci-dessus formalise la contribution de chaque variable à chaque axe, les variables ressortant en bleu étant les variables les plus explicatives d'un axe donné.

Les 2 premiers axes ne justifient aucun commentaire additionnel par rapport à ceux de la Figure 22.

Concernant l'axe 3, tous les descripteurs les plus influents sont en relation, dans un sens ou dans l'autre, avec le gradient amont-aval. On peut noter que le Corg exerce en fait un poids de structuration sur les 3 axes (sachant que l'inertie portée par les 2 premiers axes est dominante), que l'ammonium contribue à la fois à l'axe 1 et 3. Les autres variables du 3^{ème} bloc en bas de tableau revêtent beaucoup de sens au sens du continuum amont-aval. Il est à noter que le descripteur NO3, plutôt considéré comme un marqueur de pollutions diffuses d'aval de bassin versant, est porté à titre principal par cet axe 3. Mais ce paramètre exerce une contribution faible sur tous les axes, certainement du fait que la teneur en NO3 associée aux rejets de STEP, qui peuvent se localiser un peu partout au long des hydrosystèmes, vient contrebalancer la présence de nitrates à l'aval des bassins versants sous influence des pollutions diffuses et en minimiser la

signature dans cette analyse. Comme c'est l'axe 3 qui reçoit malgré tout la contribution la plus forte des nitrates, cette variable a été portée en bleu dans la 3^{ème} colonne qui lui correspond.

Cette analyse a permis de vérifier la structure du jeu de données utilisé, habituelle pour des données issues de réseaux de surveillance français, et la cohérence des poids contributifs des variables selon les axes et les gradients environnementaux et anthropiques qu'ils prennent en charge.

La représentation dominante des principaux gradients d'anthropisation sur l'axe 1 et le caractère le plus structurant de cet axe lié aux pollutions sont des phénomènes tout-à-fait classiques sur des données provenant de cette origine.

Le principe mobilisé pour la réalisation de l'analyse suivante a été de vérifier l'homogénéité des conditions abiotiques rencontrées sur les sites suivis un nombre de fois suffisant sur la chronique, matérialisée par la taille de l'enveloppe de variation de la position des relevés annuels au site dans cette ACP. Les données chimiques acquises lors de diverses années sur un site stable sur le plan de ses conditions chimiques seront réparties dans un nuage multidimensionnel de taille modérée. Mais, si le site est instable et sa chimie plus variable, les données de la chronique pluri-annuelle seront plus dispersées et occuperont un volume plus important. Rappelons ici que l'étude de la dispersion des relevés acquis au même site au cours de la série interannuelle s'est restreinte à l'exploitation des 3 premiers axes de cette ACP.

Le mode opératoire utilisé est décrit en suivant. A chaque site correspondent plusieurs relevés, ceux-ci étant prélevés au cours de différentes campagnes annuelles. Les seuls sites retenus comme candidats pour cette étude de stabilité sont ceux ayant fait l'objet d'au moins 3 suivis annuels de 2007 à fin 2011.

Dans chaque plan de l'ACP formé par un croisement d'axes (axe1-axe2, axe1-axe3 et axe2-axe3), il est possible de récupérer les coordonnées de chaque relevé. Les distances 2 à 2 entre chacun de ces relevés ont ensuite été calculées pour chaque site, afin de vérifier leur proximité suivant les paramètres de l'ACP. Cette notion de proximité intra-site a été mesurée via le calcul d'un coefficient de variation (CV):

$$CV_i = \frac{\sigma_i}{\mu_i}$$

Équation 1 : Calcul du coefficient de variation des distances entre les relevés (années) d'un même site

- où μ_i représente la moyenne des distances du site i,
- et σ_i représente l'écart-type des distances du site i.

Cette mesure, souvent utilisée en statistique descriptive, est un indicateur sans unité de la dispersion d'une variable. Plus le CV est grand, plus les relevés sont distincts (ici, sur le plan de la dispersion de leurs conditions hydrochimiques).

En vue de déterminer la limite de la valeur de CV pour laquelle la stabilité des variables chimiques d'un site sera considérée comme suffisante, il est important de s'appuyer sur 2 types d'informations : 1) tout d'abord, la dispersion des CV sur les différents plans de l'ACP pris en compte (Figure 23 en page suivante) ; mais aussi 2) la loi de distribution des sites en fonction du CV (Figure 24 en page 68).

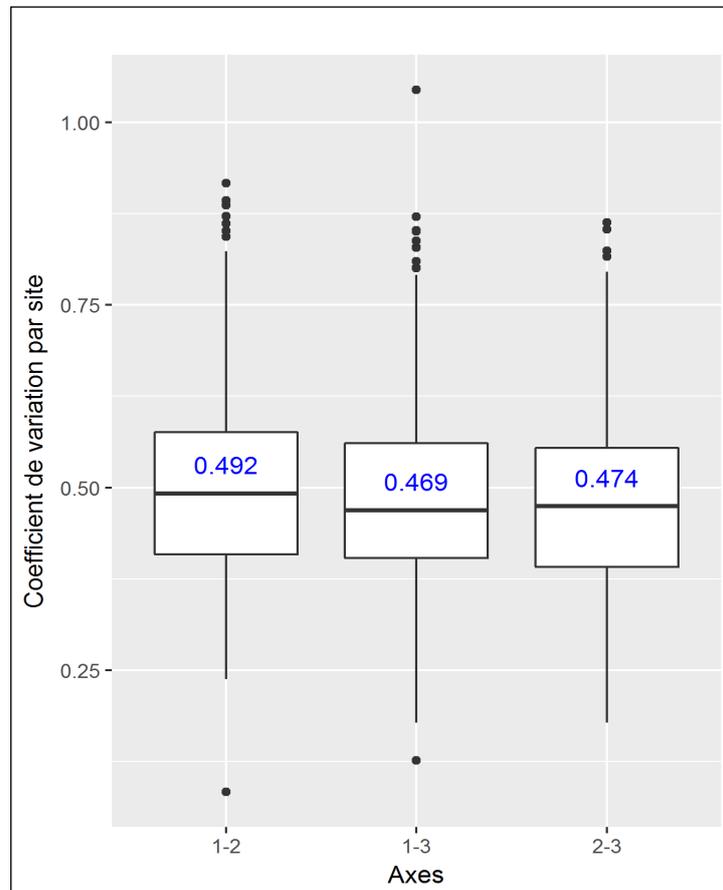


Figure 23 : Boxplot des coefficients de variations de chaque site, sur les 3 plans pour chaque site

Le niveau de coupe choisi au niveau du CV va déterminer le nombre de sites conservés.

Comme on peut le voir sur la Figure 24 en page suivante, le seuillage du CV peut rapidement devenir très contraignant sur les effectifs de sites validés comme stables si l'on applique un seuil trop sélectif. En effet, l'objectif étant de pouvoir appréhender et distribuer le niveau de variabilité naturelle de l'IBD à partir d'une assise représentative sur le plan de la diversité des conditions d'habitats naturels, et non à partir de quelques cas particuliers, l'effectif de sites à conserver, bien que devant représenter des conditions vérifiées comme suffisamment stables site par site, doit constituer un échantillon suffisamment large et représentatif pour asseoir la robustesse des analyses.

La médiane des coefficients de variation des distances pour chaque plan se situant entre 0.4 et 0.5 (Figure 23), nous avons donc étudié les répercussions, en termes d'effectifs de sites, de seuils-candidats correspondant à des valeurs limites de CV de 0.4, 0.45 et 0.5 (Figure 24).

Cet examen de seuils candidats nous a amenés à fixer la limite maximale du CV à la **valeur de 0.5**, garantissant un nombre suffisant de relevés stables.

En effet, en dessous de ce seuil, on peut constater que le nombre de sites s'appauvrit considérablement (seulement 47 sites seraient conservés pour un seuil de CV établi à 0,45).

Pour la suite des analyses, **122 sites** réputés présenter une stabilité hydrochimique suffisante ont donc été préalablement sélectionnés, correspondant à **491 relevés diatomiques annuels**.

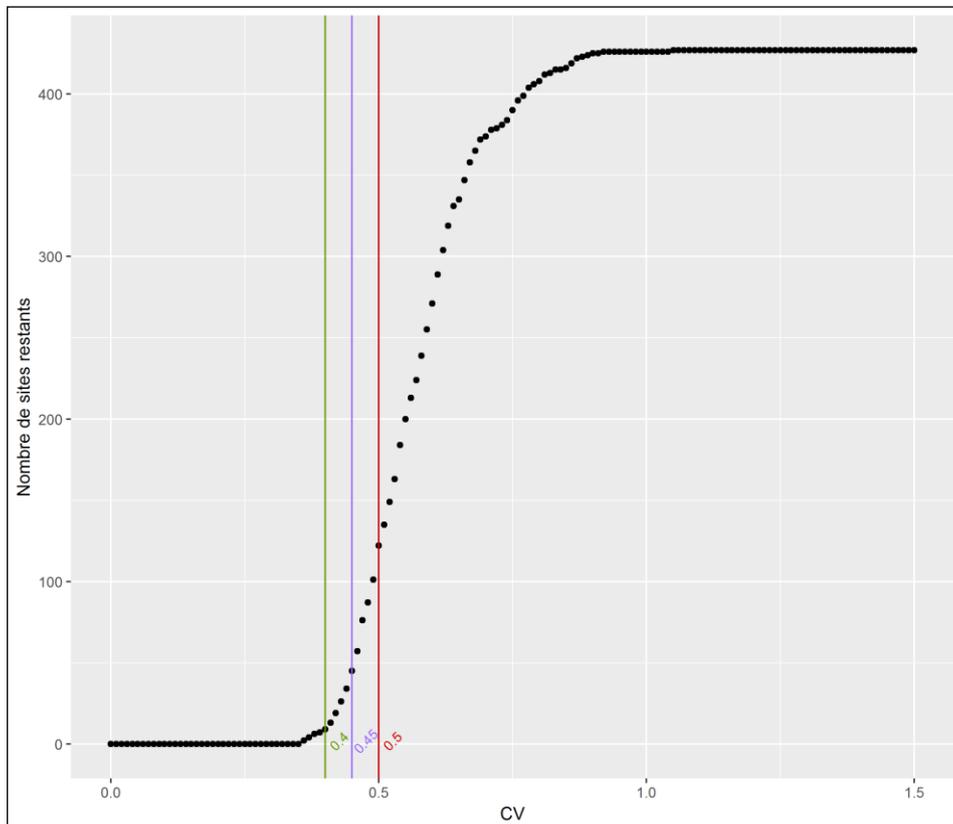


Figure 24 : Influence de la valeur limite du coefficient de variation sur le nombre de sites sélectionnés

Une deuxième étape de contrôle des données a ensuite été menée qui, tout en veillant à mobiliser au mieux l'assise de données disponible, visait à trouver un compromis permettant d'asseoir la variabilité naturelle déterminée sur chaque site sur une base la plus comparable possible sur le plan de l'assise temporelle.

La sélection de sites stables réalisée ci-dessus conduit à une distribution d'assise temporelle de relevés disponibles par site présentée en Figure 25 page suivante. Cette figure permet de montrer que 22 sites sont suivis sur 5 années, à savoir toutes les années de la chronique. La majorité des sites (81) ont bénéficié de 4 années de suivi, et 19 sites n'ont été échantillonnés que sur 3 années sur le plan diatomique.

Des divergences assez conséquentes existent dans le nombre de relevés disponibles selon les sites (une variation de 3 à 5 relevés par site conduit à un écart de 167% entre ces extrêmes), pouvant influencer de façon inégale sur la gamme de variabilité appréhendée au site.

En effet, même si des exceptions peuvent se présenter au cas par cas, augmenter la longueur de la chronique sur un site va conduire à étendre le champ des possibles et, en tendance statistique, à augmenter la variation standard des notes d'IBD.

Dans ces conditions et au vu de la distribution présentée, il nous a semblé préférable d'opter pour le compromis permettant à la fois d'étendre et de standardiser le plus possible l'assise temporelle prise en compte pour chaque site sélectionné, afin de rendre compte de la variabilité naturelle de la méthode sur une assise la plus comparable possible entre sites.

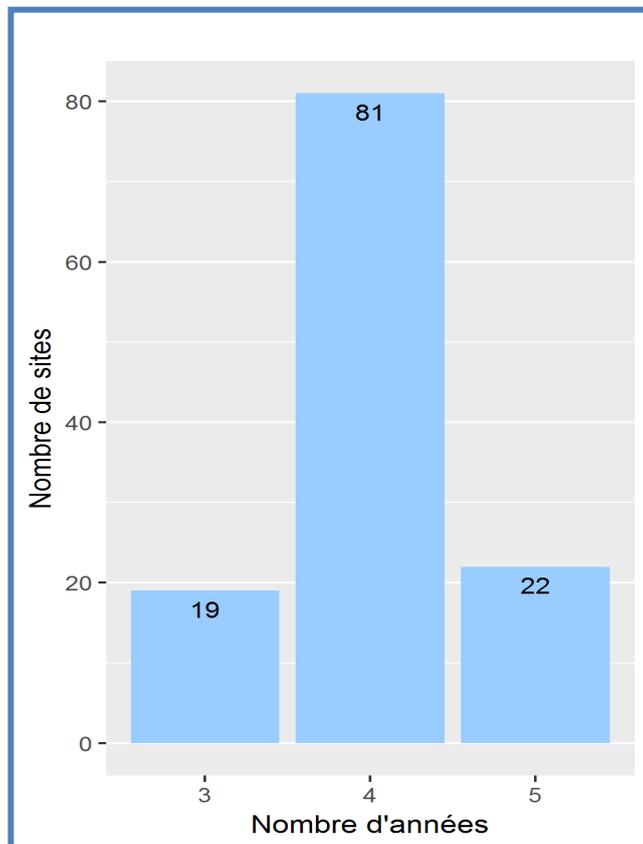


Figure 25 : Nombre de suivis par site à hydrochimie stable

Comme une nette majorité des sites ont bénéficié de chroniques d'au moins 4 ans, la suite des analyses a porté sur l'assise de sites ayant eu un minimum de 4 années de suivi. Pour les stations ayant eu 5 années de suivi, il a été convenu de choisir les 4 dernières années.

Le jeu de données sélectionné au final est composé de **412 relevés** répartis sur **103 sites**, réparties dans 15 Hydro-écorégions de type 1.

La Figure 26 en page suivante présente l'implantation géographique des 103 sites composant la sélection finale de sites considérés comme stables sur le plan de leurs caractéristiques hydrochimiques.

Ces sites ont fait l'objet de la collecte de 4 ou 5 relevés diatomiques sur les 5 années retenues du fait d'un suivi suffisamment homogène sur le plan méthodologique et taxonomique (années 2007 à 2011).

Le bassin Adour-Garonne a subi l'annulation d'un marché public de suivi diatomique dans cette période pour cause de recours contentieux. C'est la raison pour laquelle ce bassin complet ne figure pas dans la sélection d'au moins 4 relevés. D'autres bassins éventuellement peu représentés l'ont été du fait que leurs sites comportant un nombre suffisant d'échantillons

diatomiques présentaient une hydrochimie dont la variabilité dépassait le coefficient de variation-seuil choisi (ex : Alpes internes, Artois-Picardie...).

L'assise de sites, qui mixe des zones peu minéralisées et des zones à hydrochimie calcaire dans un contexte altitudinal varié paraît satisfaisante pour l'objectif de détermination de la variabilité interannuelle.

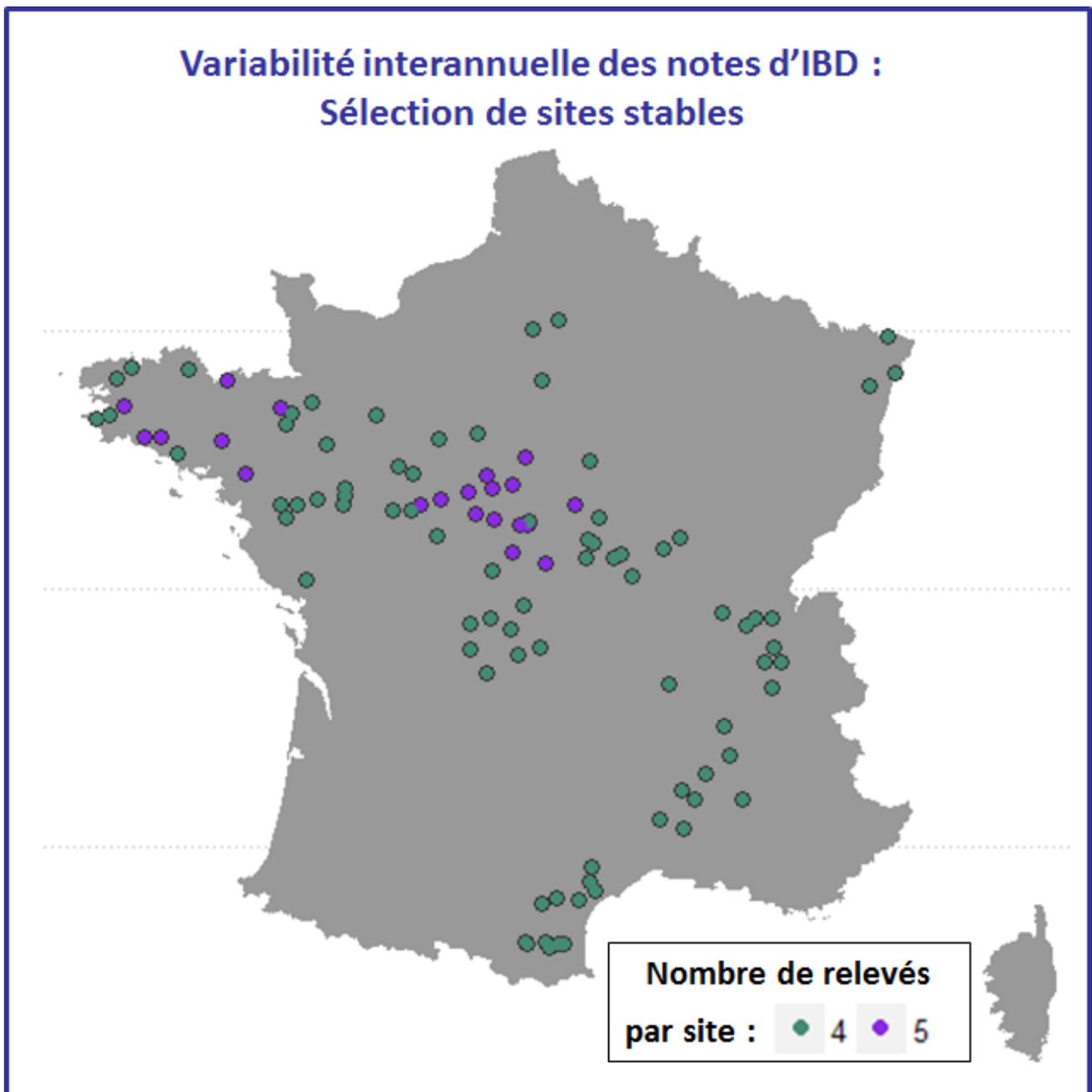


Figure 26 : Projection cartographique des 103 sites "stables" retenus pour l'étude de la variabilité interannuelle des notes d'IBD₂₀₀₇

3.3. Etude des résultats de variabilité interannuelle, discussion

3.3.1. Classes de qualités des relevés du jeu de données sélectionné

La Figure 27 ci-dessous présente la distribution entre les différentes classes d'état écologique des relevés issus des sites sélectionnés en fonction de leur apparente stabilité hydrochimique.

Il est à noter que 16 relevés sur les 412, en fonction de leur typologie nationale et du contenu de l'Arrêté d'évaluation actuellement en vigueur (Arrêté du 27 Juillet 2015), ne disposent pas encore de dispositif officiel d'évaluation valide et ne peuvent pas être pris en compte dans cette statistique (on ne dispose pas d'un système de référence pour classifier l'état des sites). Cette figure est donc construite à partir des **396 relevés classifiables** selon leur état diatomique.

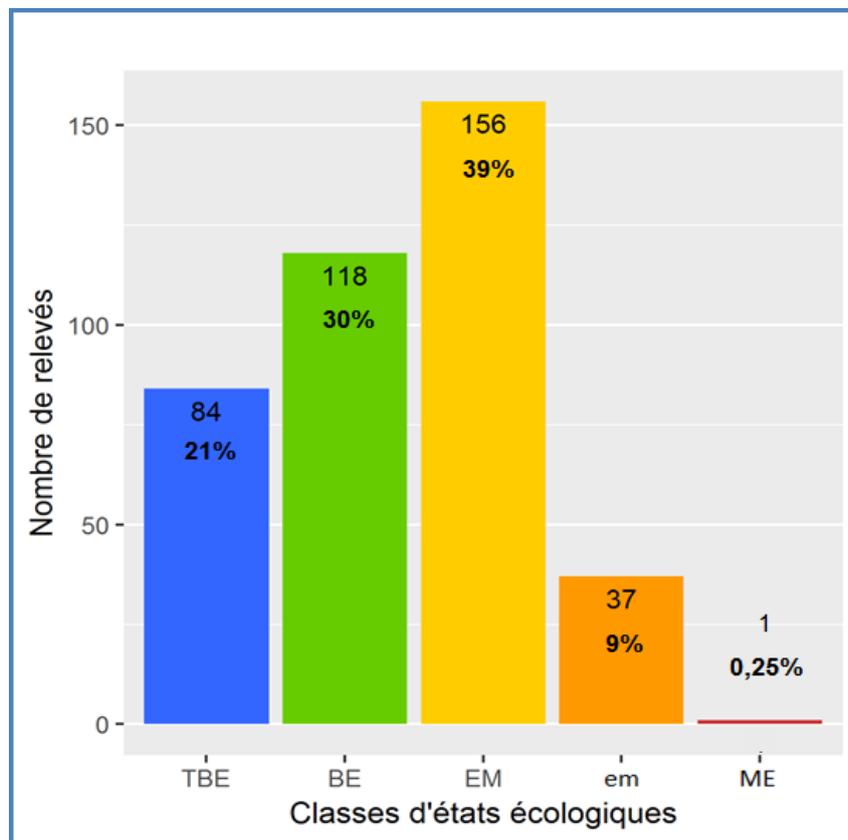


Figure 27 : Effectif et pourcentage de relevés par classe d'état écologique (TBE : Très bon état, BE : Bon état, EM : Etat moyen, em : état médiocre, ME : Mauvais état)

Cette figure permet de vérifier que, comme c'était déjà le cas pour le jeu de données collecté dans le cadre de l'étude inter-opérateurs (cf. chapitre précédent), cette sélection de relevés sur sites considérés comme stables est assez représentative de la distribution en classes de qualité observée pour les jeux de données nationaux collectés dans le cadre des réseaux de surveillance.

Outre la large assise en matière d'HER1 (15 sur 22) et de diversité des conditions naturelles représentées par cette sélection de relevés, cette caractéristique de distribution relativement conforme à celle des réseaux de surveillance de l'état écologique est *a priori* un élément favorable vis-à-vis de l'objectif de cette approche.

En effet, elle va aider à **cerner de façon représentative l'intervalle de confiance** attaché à l'évaluation réalisée à partir de l'IBD2007 en fonction de la **variabilité interannuelle des conditions naturelles**.

Une large majorité des relevés se retrouve concentrée entre la classe d'Etat Moyen (156) et la classe de Bon Etat (118), ce qui correspond bien à une réalité nationale. Un seul relevé arrive dans la classe de Mauvais Etat (c'est-à-dire ici 1 seul relevé sur 4 acquis au même site) et 37 se situent dans la classe d'état médiocre (em).

Cette sélection de relevés donne donc une image relativement fidèle à la distribution actuelle de qualité des sites dans les réseaux de surveillance (résorption à peu près complète des sites en mauvais état et régression forte des sites en état médiocre). En effet, la plupart des points noirs signalés de pollution ponctuelle ont fait l'objet de mesures importantes et efficaces d'amélioration de traitement dans le cadre des politiques nationales de l'eau en place depuis une quarantaine d'années, avec l'intervention déterminante des Agences de Bassins.

On note la présence de 84 relevés classifiés en Très Bon Etat soit environ 21% des relevés. Même si ces relevés en très bonne qualité ne sont pas tous associés à des sites de référence, cela illustre malgré tout un certain niveau de contribution de sites du RREF couverts en 2007, qui ont continué ensuite de faire l'objet de suivis dans le cadre d'autres réseaux. Là aussi, même si cette statistique est éminemment variable selon les bassins hydrographiques et les caractéristiques de leurs environnements naturels, cette proportion de relevés en TBE est assez fidèle à la statistique obtenue à l'échelle globale des réseaux nationaux.

En creux, la statistique commentée ci-dessus signifie aussi qu'une grande majorité des relevés des sites sélectionnés par leur stabilité physicochimique apparente sont déjà plus ou moins fortement sous l'effet d'impacts anthropiques, procurant des réponses biologiques ayant plus ou moins dévié des conditions de référence.

Le classement en Bon Etat, surtout lorsque l'on s'approche de la partie inférieure de la classe de Bon Etat, correspond déjà à une déviation modérée mais tangible par rapport à la condition de référence correspondante. La classe d'état moyen, la plus représentée, subit déjà des impacts anthropiques affirmés, la moitié des relevés est en état moyen ou inférieur, subissant des impacts anthropiques de nets à très nets.

3.3.2. Analyse des écarts 2 à 2

Afin d'analyser la variabilité interannuelle pour chaque site, les écarts des notes IBD₂₀₀₇ et des EQR entre les années prises deux à deux ont été calculés pour chaque site, et leurs distributions ont été représentées sous forme de boîtes à moustaches (voir Figure 28 en page suivante).

Une première remarque préalable au commentaire des résultats présentés dans cette figure renvoie sur une difficulté méthodologique à laquelle nous avons dû faire face en cours de démarche, et qui a conditionné ici l'expression d'un **niveau de variabilité plus important** que l'influence de la seule **variabilité naturelle**.

Dans l'intention initiale (cf pages 59 et 60), la façon la plus appropriée que nous envisagions pour isoler et évaluer l'effet de la seule variabilité naturelle aurait été de ne considérer que des sites de référence ou dans un état pseudonaturel, situés dans des ensembles naturels différents, et de mesurer la variabilité naturelle du résultat de l'IBD sur plusieurs années.

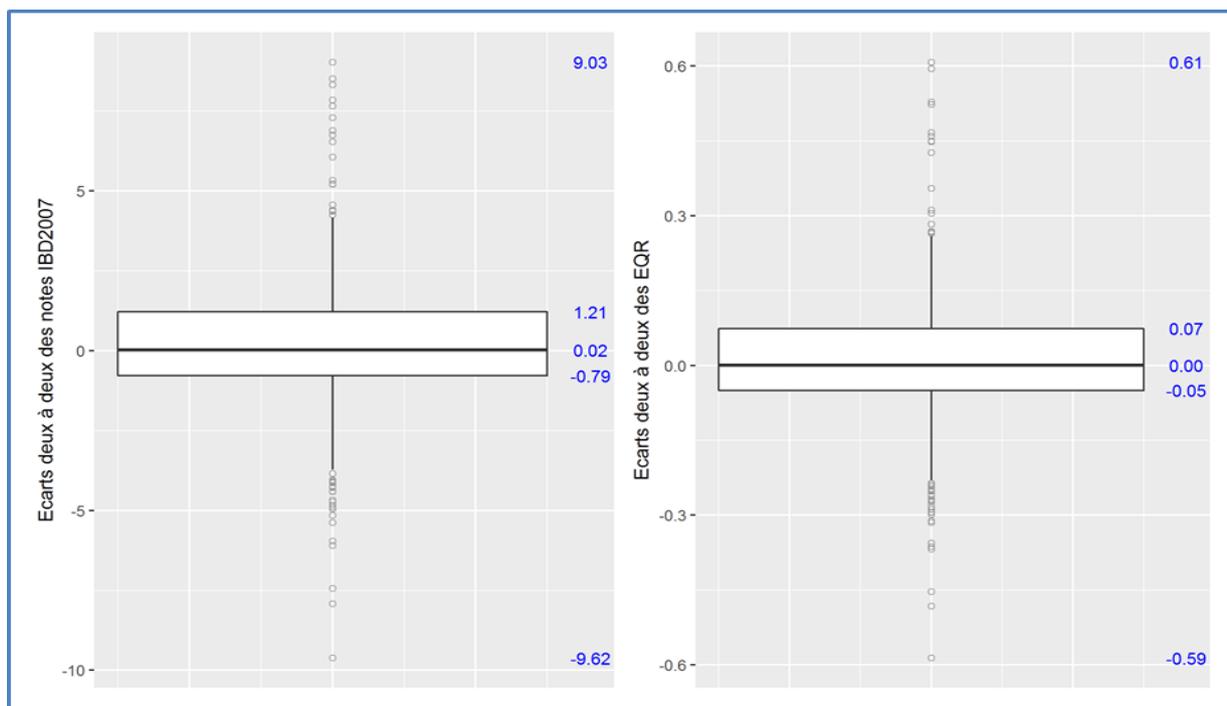


Figure 28 : Distribution des écarts au même site, années prises 2 à 2, des notes d'IBD2007

En sites de référence, sauf évènement inhabituel et localisé de pollution accidentelle, cette variabilité ne serait conditionnée que par la variabilité interannuelle des caractéristiques hydrochimiques et hydroclimatiques naturelles à ce site.

Cette façon de faire nous semblait *a priori* la pratique la plus correcte pour une estimation directe de la variabilité naturelle interannuelle car elle aurait garanti, dans des conditions proches de situations de référence, l'influence nulle ou plutôt marginale d'effets d'altérations anthropiques sur l'expression de cette variabilité interannuelle.

Elle aurait bien correspondu à l'objectif poursuivi de dimensionnement de l'intervalle de confiance de l'IBD sous variabilité naturelle interannuelle des conditions d'habitat, sachant que les prélèvements diatomiques mobilisés pour cette analyse, ainsi que l'acquisition des données hydrochimiques correspondantes, ont bien été réalisés selon les conditions normatives en vigueur dans le cadre des réseaux de surveillance. En effet, les prélèvements sélectionnés ici étaient centrés sur la saison de basses-eaux estivales, hors influence d'un épisode de crue détectable au moment du prélèvement et/ou ayant provoqué un remaniement visible des substrats.

Cependant, les discontinuités intervenues dans la méthodologie de suivi (notamment changements conséquents à partir de 2007, liés au passage à l'IBD₂₀₀₇) et la non-pérennisation du suivi de beaucoup de sites de référence ayant appartenu au RREF entre 2005 et 2007 ont conduit à ne pas disposer d'une assise suffisante et suffisamment comparable de données issues de sites à naturalité confirmée (sachant qu'il fallait en plus bénéficier d'un nombre minimal de répliqués interannuels à chaque site).

C'est donc par force et suite à ce déficit de données permettant de situer les variations interannuelles sur une assise suffisante de sites validés comme non impactés qu'il a été décidé de ré-orienter la sélection de sites sur la base de leur **relative stabilité des conditions abiotiques**, quelle que soit leur gamme de qualité.

La sélection finalement faite comporte bien sûr certains sites de référence, visiblement très minoritaires (21 % de relevés se classent en TBE, mais un pourcentage assez substantiel d'entre eux provient de sites n'ayant pas le statut de référence). En effet, il est tout-à-fait habituel de trouver à certaines dates des relevés qui se classent en TBE sur des sites globalement en BE voire parfois même proches de la limite supérieure d'EM, selon les caractéristiques de pression de pollution et les conditions hydroclimatiques de l'année.

Cependant, malgré la réalisation de cette sélection initiale de sites semblant, à l'analyse, présenter un certain niveau de stabilité chimique interannuelle, l'incorporation de **sites sous altération anthropique** génère manifestement un risque plus grand de **variabilité de la physico-chimie de l'eau et de la réponse biologique**, non forcément perceptible sur la base de la chronique d'hydrochimie disponible au site sur le trimestre (on a entre 1 et, au maximum très rarement atteint, 3 analyses de chimie associables sur l'épisode...), mais pouvant s'exercer à d'autres moments, en fonction de la conjonction d'usages anthropiques et de la survenue d'évènements hydrologiques (pluies ou orages, petites crues...).

Une seconde remarque préalable est que, à l'instar de la Figure 15 relative à l'analyse des valeurs absolues des écarts 2 à 2 obtenus entre opérateurs sur les notes d'IBD2007 pour chaque site d'étude, le mode de représentation des résultats et de leur variabilité tel qu'utilisé en Figure 28 maximise visuellement l'impression générale de dispersion des résultats par rapport à leur assise statistique. En effet tous les écarts d'évaluation, même les plus importants, constatés au site dans la chronique d'années disponibles sont représentés dans cette distribution et élargissent fortement les extrêmes du box-plot sur le plan de son aspect visuel, alors que leur nombre est plutôt modéré dans l'effectif total des intercomparaisons réalisées.

L'assise des valeurs outliers contribue encore plus fortement à l'élargissement de l'échelle d'écarts représentée. Ainsi, dans la construction de cette figure, bâtie sur la base de tous les écarts entre 4 valeurs annuelles acquises à chaque site donné, sur une assise de **103 sites** dont l'état diatomique est évaluable, soit (103 sites X 6 combinaisons 2 à 2 = **618 écarts analysés**), seulement une **trentaine de valeurs outliers**, correspondant à coup sûr à des sites dont l'impact anthropique varie assez fortement dans le temps sans qu'on ait pu le percevoir sur la base des données chimiques disponibles, élargit de façon très conséquente (environ par un facteur de *2,5) la représentation du box-plot par rapport à la zone contenue dans les moustaches, alors qu'elles représentent moins de 5% des écarts mesurés dans tout le jeu de données.

Ces 2 remarques préalables étant faites, **l'examen des résultats de la figure 28**, en écarts de notes d'IBD comme en écarts d'EQR, illustre assez clairement qu'une bonne proportion des écarts présentés **ne reste pas cantonnée à la seule variabilité d'origine naturelle**. Il ressort nettement qu'elle englobe aussi, probablement sur assez nettement plus de la moitié des intercomparaisons (cf statistique d'état des relevés en Figure 27, dont seulement 21%, sous influence quasi-exclusive des conditions naturelles, se classent en TBE), une autre composante de variabilité, de modérée à très importante selon les cas, qui est plutôt à relier à une **variabilité de répercussion d'impacts anthropiques** sur la réponse du compartiment diatomique, en fonction de la conjonction entre la pression anthropique au site et le profil hydroclimatique de l'année.

Le niveau élevé de certains écarts à la médiane en atteste, qui peut atteindre de 4 à 4,5 points en plus ou en moins dans la distribution normale des écarts, et s'élever jusqu'à plus de 9 points pour les quelques valeurs outliers.

Au vu de l'assise d'état des relevés sur notre sélection de sites considérés comme stables (Figure 27), ce constat est logique. Il signifie que, même si notre pratique de sélection de sites par l'analyse de la stabilité des conditions physico-chimiques a dû permettre de contenir en partie la statistique des variabilités interannuelles d'état et les écarts au site par rapport à un jeu de données tout-venant, l'assise (le plus souvent, seulement 1 à 2 relevés ponctuels) et la représentativité temporelle insuffisante de ces données de chimie ne permet pas forcément de détecter efficacement et à tout coup les variabilités et l'instabilité de certains sites, en particulier de sites sous altération anthropique, dont l'état biologique est forcément plus impacté et plus variable que dans des sites naturels.

La part de relevés **sous influence très dominante d'une variabilité naturelle** porte sur l'assise des **relevés en TBE** et assez probablement aussi, sur la partie des relevés couvrant **la moitié haute de la classe de BE**. Par contre, la plupart des relevés dans la moitié basse de la classe de BE subit déjà une baisse de note assimilée par l'IBD à un impact biologique mesurable, dont la variabilité sous effet anthropique se cumule avec la variabilité d'origine naturelle et devient rapidement dominante.

Aussi, il apparaît possible de réaliser une estimation réaliste et même probablement un peu surévaluée de l'influence de la variabilité naturelle sur l'IBD en s'appuyant sur la base la plus habituelle des écarts, estimée à partir de la moitié des intercomparaisons (soit l'assise statistique des intercomparaisons mettant en jeu la totalité des effectifs des classes de [TBE + BE], qui représente ici 51% des effectifs totaux de relevés du jeu de données).

A ce stade, l'examen de la Figure 28 apporte une confirmation claire que la pratique de sélection de sites apparemment stables, s'appuyant sur l'assise des descripteurs chimiques disponibles dans la base de données abiotiques, n'a pas toujours été en mesure d'éliminer l'influence de la variabilité de l'impact anthropique dans la variation de note d'IBD au site, qui peut devenir très influente sur certains sites à certains moments. C'est le cas pour certaines valeurs basses de doublets (relevés sur site manifestement impacté à la date du relevé biologique) dont les écarts peuvent atteindre 3 à 4 points, en distribution relativement habituelle, par rapport à un autre relevé prélevé au même site, voire dépasser largement ces valeurs pour quelques relevés outliers (l'écart maximum d'évaluation constaté au même site en comparaison 2 à 2 ayant pu atteindre jusqu'à 9,3 points d'IBD₂₀₀₇).

De tels niveaux d'écarts apportent la démonstration claire que ces sites à évaluation biologique si variable, malgré leur apparente stabilité sur la base des données hydrochimiques ponctuelles disponibles, sont sous l'influence patente de variation d'effet des altérations anthropiques, le statut anthropisé de beaucoup d'entre eux étant connu dès le départ (cf. référentiels de valeurs des paramètres abiotiques, cf. classement d'état au relevé présenté en Figure 27).

Cependant, même imparfaite, cette pratique de sélection a très probablement assez fortement réduit cette variabilité, sur le plan de la statistique fréquentielle, par rapport à un jeu de données tout-venant incluant beaucoup de sites instables à dynamique interne certainement encore plus variable. Il reste donc intéressant d'utiliser au mieux cette assise de sites et de relevés pour tenter d'approcher la part de **variation naturelle interannuelle** de la note d'IBD.

Comme déjà discuté plus haut, une façon raisonnable d'approcher cette variabilité naturelle repose sur la quantification du niveau de bruit de fond présenté par la méthode dans les situations les plus habituelles, en s'appuyant sur la variabilité montrée par le noyau central du jeu de données mais en évitant d'intégrer les valeurs fortes ou extrêmes d'écarts.

Il est certain que celles-ci sont rencontrées sur des sites nettement anthropisés et qu'elles incorporent manifestement une part de plus en plus importante liée à la **variabilité interannuelle du niveau d'impact anthropique** (sachant que l'IBD a été créé pour donner des réponses sensibles et fortes à ce type d'altérations), qui **se surajoute** et **vient rapidement dominer la variabilité interannuelle d'origine naturelle**.

Ainsi, en partie gauche de la Figure 28, l'assise de la partie centrale du box-plot contenue entre le percentile 25 et le percentile 75, qui représente la statistique de variabilité la plus habituelle, constatée sur 50 % des sites et relevés non extrêmes, permet d'estimer de façon **robuste** et **même certainement un peu surévaluée** une **variabilité d'origine très majoritairement naturelle**, obtenue sous condition hydrochimique et d'anthropisation plutôt stables, contenue ici entre - 0,79 et +1,21 point d'IBD, correspondant à une profondeur de fourchette de variation de **+/-1 point d'IBD**.

Le référentiel d'écart 2 à 2 décliné en Figure 28 a été obtenu à partir de données historiques de réseaux qui présentent certaines limites connues et signalées.

La première limite tient dans la faible représentativité temporelle des données de chimie issues de prélèvements ponctuels qui, du même coup, vont difficilement permettre de garantir la bonne stabilité du niveau d'impact anthropique sur des sites altérés, notoirement plus variables que des sites naturels, en fonction de l'année.

D'autre part, ces données peuvent contenir leur lot d'autres erreurs ou imprécisions liées par exemple à des évolutions taxonomiques, à des changements de marchés publics et d'opérateurs au long de la chronique acquise au même site, à une éventualité probablement rare, mais existante, d'erreur de localisation de site ou de faciès à une date donnée, etc...

Compte-tenu de ce qui précède, le référentiel d'écarts 2 à 2 de **+/-1 point d'IBD** de variabilité interannuelle d'origine naturelle, probablement un peu surévalué par construction (50 % de l'effectif de relevés correspond à **tous** les relevés en TBE et **tous** les relevés en BE, dont environ la moitié se situe déjà sous une variabilité d'origine anthropique mesurable), est finalement assez proche et plutôt cohérent avec celui trouvé antérieurement sur le jeu de données inter-opérateur. En effet (voir Figure 15), celui-ci était de **+/- 0,84** selon le même mode de construction et d'expression (écart-standard 2 à 2 déterminé entre le percentile 25 et le percentile 75).

L'écart standard 2 à 2 estimé de cette façon semble donc consistant et doit permettre d'isoler de façon assez correcte et non sous-évaluée la part de variation interannuelle des écarts de notes d'IBD sous influence naturelle.

La partie droite de la Figure 28 présente les mêmes résultats d'écarts selon une échelle exprimée en EQR (notes d'IBD₂₀₀₇ ramenées en équivalents de qualité écologique par rapport au niveau de référence adéquat ; valeur de 1 = niveau de référence adéquat pour ce site, valeur 0 = plus basse note d'IBD atteignable dans cet ensemble naturel).

L'analyse qui peut en être faite est équivalente et ne suggère pas de commentaire particulier. Sous ce mode d'expression des grandeurs, la part des écarts 2 à 2 qui pourrait être attachée à la variation interannuelle naturelle (*i.e.* non due à une variation d'impact anthropique) serait dans ce cas contenue entre [-0,05] et [+0,07], matérialisant un écart d'une profondeur de **+/-0,06 EQR**.

3.3.3. Analyse des écarts de classification d'état au relevé

La Figure 29 ci-dessous présente les écarts de classification écologique obtenus entre les 4 relevés collectés au même site et retenus pour l'étude de la variation interannuelle (cf. règle de sélection des 4 relevés indiquée en page 62).

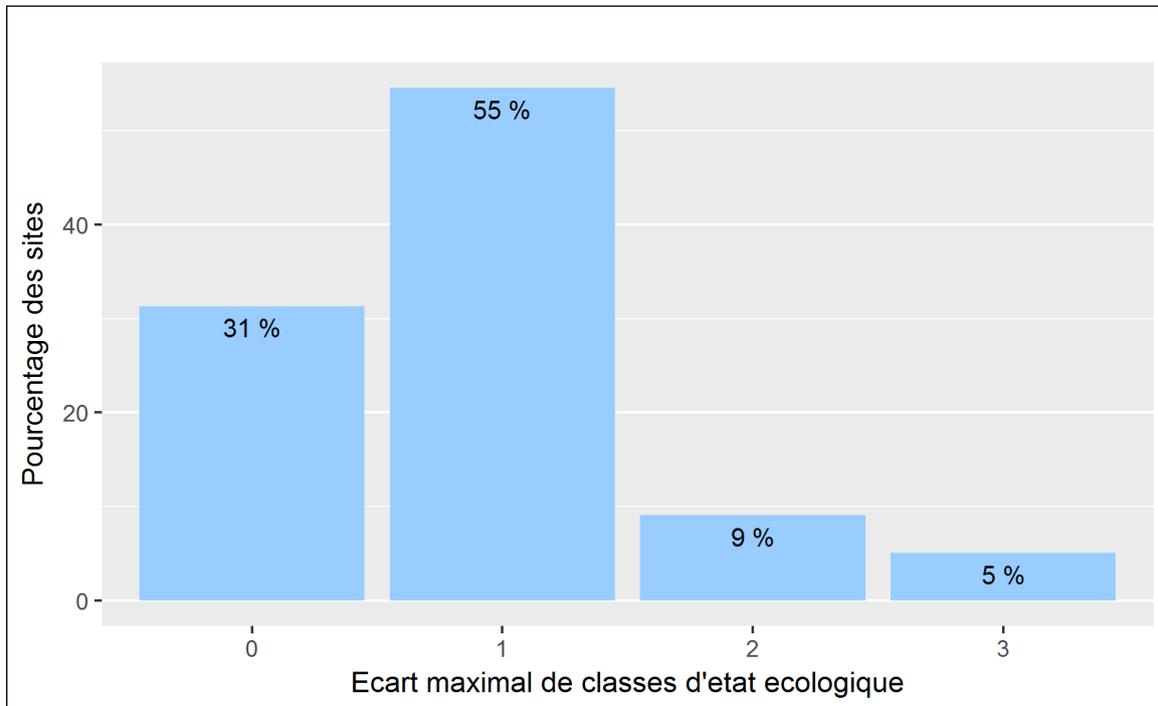


Figure 29 : Nombre de sites suivant le nombre d'écart de classe d'état écologique

Sur ce jeu de données, dont une large majorité de sites est soumise à une pression anthropique non-négligeable, le pourcentage de sites régulièrement évalués dans la **même classe de qualité** selon les années, plutôt modeste (**seulement 31 %**), matérialise le fait que les conditions locales dans les cours d'eau sont éminemment variables selon les années et que les résultats de l'IBD sont assez sensiblement influencés par ces variabilités.

Pour une **majorité de sites** (55% du total), ces évaluations annuelles vont différer **d'une classe de qualité** selon les années. 2 cas de figure peuvent conduire à ce type de résultat :

- Une partie de ces sites peut avoir un état écologique proche d'une limite de classe. Une variation modeste du résultat indiciaire, qu'il soit dû à une fluctuation habituelle du résultat de la méthode (*i.e.* sa propre marge d'imprécision si l'on répète plusieurs fois le processus allant du prélèvement au résultat du comptage) ou à une variation des conditions naturelles selon le profil hydroclimatique particulier de chaque année, suffira à faire basculer certains des résultats indiciaires d'une classe à l'autre. Nous sommes bien ici dans la notion d'intervalle de confiance du résultat de la méthode indiciaire, liée uniquement à une variabilité intrinsèque de sa réponse et/ou à une variabilité des conditions naturelles *in situ*.
- Beaucoup de sites parmi cette sélection de sites stables se situent déjà dans des conditions où s'exerce une pression anthropique non-négligeable (cf Figure 25 montrant que la classe dominante de qualité de cette sélection de relevés est la classe d'Etat Moyen), conduisant forcément à une plus grande variabilité des possibles en matière de qualité d'eau.

Selon les caractéristiques et la temporalité de l'usage anthropique d'une part, selon les conditions hydroclimatiques de l'année d'autre part (chaudes ou froides, sèches ou humides, avec les différences à en attendre sur le plan des transferts et des effets de nutriments et d'autres substances polluantes), la réponse du compartiment diatomique va intégrer une part de variabilité d'origine naturelle comme ci-dessus, mais aussi l'influence de non-négligeable à sensible d'une variabilité d'impact anthropique. Ce 2^{ème} type de fluctuation, qui peut rapidement et largement dépasser la variation naturelle, est normal et même recherché, puisque l'IBD a spécialement été conçu pour évaluer de façon sensible les réponses biologiques liées à l'altération de la qualité des eaux. Dans cette assise de sites manifestant ce basculement d'une classe de qualité, il convient de souligner qu'il existe des sites où, malgré que la fluctuation d'évaluation globale ne dépasse pas une classe d'écart, certains relevés les plus éloignés peuvent présenter un écart 2 à 2 approchant 2 classes...Au bilan, sauf site naturel particulier aux conditions très variables, où les communautés biologiques pourront manifester une perturbation biologique plutôt forte lors de certaines séquences interannuelles inhabituelles, la plupart des sites dont les écarts d'évaluation au relevé vont avoisiner ou dépasser la profondeur d'une classe complète de qualité subissent déjà une part sensible de changement de notation lié à une variation de l'impact anthropique.

Tous les autres cas présentés dans les 2 colonnes de droite de la Figure 29 correspondent à de forts écarts de notation entre relevés, conduisant à un fort écart de classification d'état écologique des sites au relevé.

Les 9% de sites ayant subi un écart de classification dépassant 2 classes de qualité et les 5% de sites révélant un écart de plus de 3 classes de qualité au relevé sont indéniablement des sites où, malgré la vérification de la relative stabilité des conditions chimiques disponibles dans notre jeu de données, le processus de sélection mis en œuvre n'a visiblement pas permis de maîtriser des variabilités de conditions abiotiques qui interviennent *in situ*. Ce problème est principalement lié au défaut de représentativité temporelle des stratégies d'échantillonnage chimique basées sur des prélèvements ponctuels à faible cadence. Dans ces cas, les impacts anthropiques s'exercent avec une sévérité variable, mais leur effet domine la variabilité d'origine naturelle et peut provoquer un fort impact certaines années, faisant fortement à très fortement varier à la baisse le classement d'état au relevé.

Les données de base ayant été acquises dans le cadre des réseaux de routine et dans un contexte multi-intervenants (cf. processus d'attribution de marchés publics), il est aussi imaginable que de tels niveaux de différence puissent être liés, dans quelques cas isolés, à une erreur de localisation de la station par un intervenant sur les 4 ou 5 passages de terrain constituant la chronique sélectionnée ; sachant que, en dépit de son caractère un peu emblématique pour les intervenants, ce cas de figure a parfois été détecté dans le cadre de l'étude nationale inter-opérateurs rapportée au Chapitre 3.

Compte-tenu de ce qui précède, la pratique réalisée au chapitre précédent, à savoir situer la variation d'origine naturelle en se basant sur le domaine de fluctuation le plus habituel, calé sur 50% des écarts rencontrés entre le percentile 25 et le percentile 75, apparaît donc correctement dimensionnée et, en tout état de cause, non sous-évaluée dans le cas d'un tel jeu de données pas uniquement centré sur des relevés de référence.

Ici, seulement 10 à 15% des sites sont dans ce statut ou proches de ce statut, et une nette majorité de sites subit de façon indéniable des pressions anthropiques temporaires ou permanentes, à l'impact plus ou moins variable selon le profil de l'année.

L'examen de la Figure 30 ci-dessous, au regard de la statistique de variabilité de la classification écologique des sites illustrée en Figure 29, vient conforter la discussion sur le statut d'anthropisation des sites en fonction des variations de l'évaluation d'état au relevé.

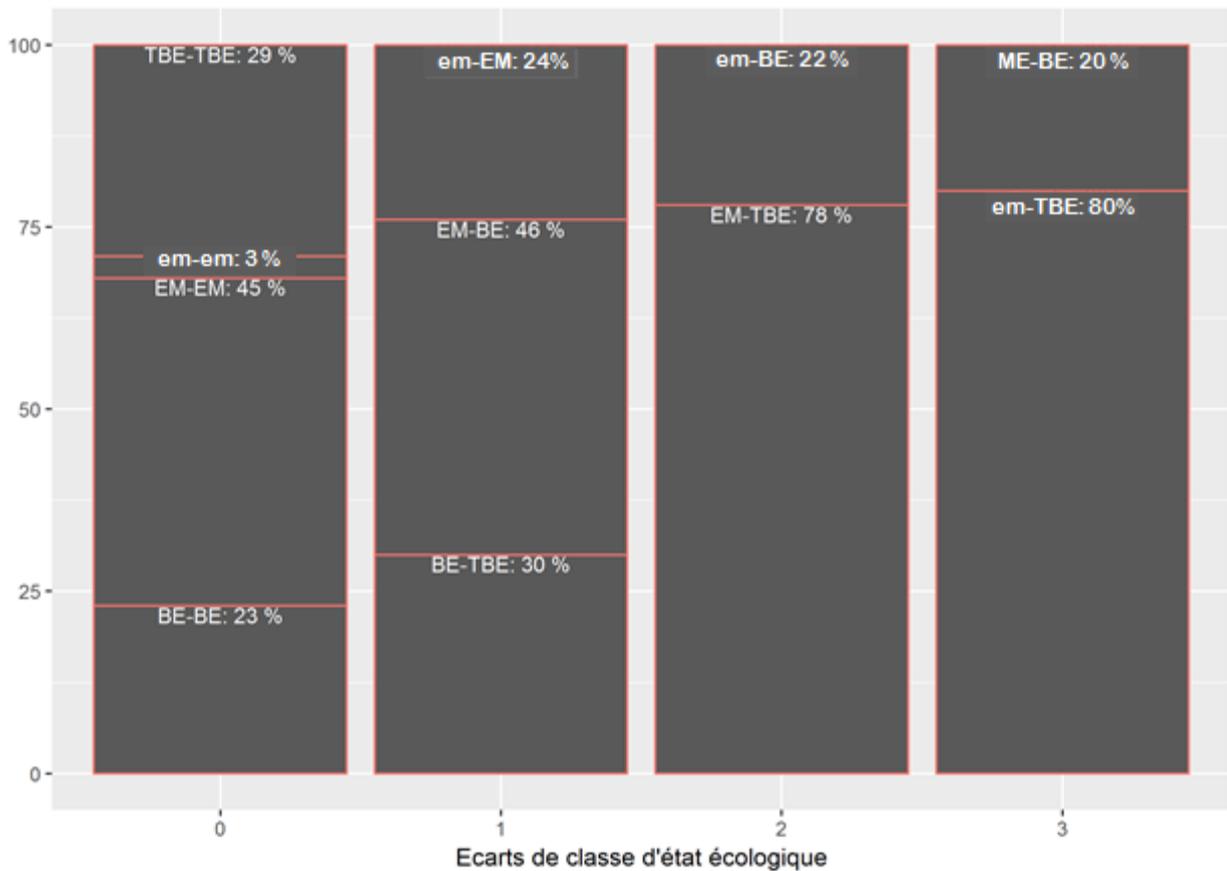


Figure 30 : Variabilité naturelle interannuelle de l'IBD₂₀₀₇ : Statistique des cas rencontrés par écart de classes d'état

L'examen de l'histogramme le plus à gauche, correspondant aux sites dont tous les relevés indiquent le même état, suscite les commentaires suivants :

- Les 31% des sites qui ne changent pas de classification au relevé sur la série de relevés retenus doivent être **assez bien centrés sur leur classe de qualité**. Leur statistique d'état est **relativement comparable** à celle qui se dégage au niveau de la **statistique d'état de tous les relevés sélectionnés** (cf. Figure 27), montrant une dominance des sites en état moyen.
- Dans ces sites sans variation d'état, on peut aussi noter **une certaine sur-représentation de la classe de TBE (29% de sites stables en TBE pour seulement 21% du total des relevés en TBE)**. Une première hypothèse explicative, de nature statistique, peut tenir au fait que ces sites ne peuvent manifester un glissement de qualité que vers un côté de la classe (glissement à la baisse, de TBE vers BE), ce qui pourrait faire baisser la probabilité de changement à partir de cette classe de TBE.

Une autre piste d'explication correspond à un constat qui se vérifie souvent sur le terrain, à savoir qu'un site de référence, ou subissant un faible niveau de faible pression anthropique, est souvent soumis à des fluctuations hydrochimiques (notamment pour les nutriments et les micropolluants toxiques) et d'état biologique **d'ampleur plus modérée que des sites sous pression anthropique avérée**, conduisant à une variabilité plus faible des résultats indiciaires.

- Ce constat de plus grande stabilité d'état des sites régulièrement en TBE (qui correspondent à des sites de référence ou en conditions pseudo-naturelles) vient donc **conforter la pratique d'estimation de la variation naturelle** telle que mise en œuvre au chapitre précédent, à savoir d'en baser l'estimation sur la partie des écarts la plus habituelle (assise sur 50 % des écarts les plus habituels et les plus faibles), et pas sur la variabilité totale. Les situations à écarts plus importants correspondent le plus souvent à des sites anthropisés, où une variation importante d'état selon les années est selon toute vraisemblance occasionnée par une variabilité interannuelle des effets d'impacts anthropiques qui vient se superposer aux écarts habituels, d'ampleur modeste, liés à la variabilité naturelle.

Le 2^{ème} histogramme en partant de la gauche, correspond à la catégorie dominante des sites qui peuvent présenter un écart d'**une classe de qualité** selon les relevés (55% des sites, cf. Figure 29). Comme déjà indiqué, l'appartenance d'un site à cette catégorie peut correspondre à 2 typologies de cas assez différentes :

- 1) soit l'état du site est proche de la jonction de 2 classes de qualité, ce qui augmente l'aléa de classification de part et d'autre en fonction d'une fluctuation modeste de note d'IBD ;
- 2) soit les écarts de classification au site, plus importants et pouvant même à l'extrême s'approcher de la profondeur de 2 classes de qualité, sont dûs à des conditions nettement plus variables des conditions abiotiques selon les années et en particulier, à une importante composante de variabilité d'impact anthropique selon les usages et le profil hydroclimatique de l'année.

Dans une nette majorité des cas (70 %), l'analyse fréquentielle basée sur les limites entre classes concernées montre bien que ces sites à variabilité de classification d'état d'une classe se situent déjà sous une pression anthropique modérée à moyenne (limite EM-BE = 46% des cas), voire forte (limite em-EM = 24%). Dans ce dernier cas de figure, la composante de variabilité d'origine anthropique devient à coup sûr très nettement dominante par rapport à la variabilité d'origine naturelle et vient accroître sensiblement la variabilité du résultat de la méthode sur ces sites impactés.

Les 30 % de cas restants (fluctuation entre le TBE et le BE), doivent correspondre à des situations où l'altération anthropique n'est probablement pas complètement absente (auquel cas le site aurait le plus de chances d'appartenir aux 29 % de sites restant en TBE dans l'histogramme de gauche) mais, dans la plupart des cas, est certainement modérée et/ou intermittente.

Les 2 histogrammes en partie droite de la Figure 29, correspondant aux sites à grosse variation d'état écologique atteignant ou pouvant dépasser la profondeur de de 2 à 3 classes d'évaluation selon les relevés, sont bâtis à partir d'une minorité de sites (en tout, 14%) indéniablement sous l'influence d'une pression anthropique fortement variable et parfois intense (cf les sites qui peuvent descendre jusqu'à l'état médiocre -em-, voire le mauvais état -ME-). Dans un tel cas de figure, si l'état le plus éloigné à la hausse ou à la baisse ne concerne qu'un relevé sur la série, il n'est pas non plus complètement exclu qu'il ait pu intervenir une erreur d'un opérateur sur le site de terrain à échantillonner, générant une autre nature de variabilité.

Quoi qu'il en soit, on perçoit bien que le processus de sélection appliqué à partir de l'apparente stabilité des conditions abiotiques n'a pas bien fonctionné sur les sites avec de tels écarts, complètement inappropriés pour isoler la **part de variabilité naturelle** de la méthode d'évaluation. Celle-ci, probablement d'ampleur modeste, est complètement noyée dans une très forte variation du niveau d'impact anthropique selon les années. Cela confirme bien que l'estimation de la variabilité naturelle interannuelle doit être pratiquée en écartant de l'assise prise en compte les sites sous influence nette d'altérations anthropiques.

3.3.4. Analyse des écarts à la moyenne au site

La Figure 31 ci-dessous présente la statistique de dispersion des écarts à la note moyenne d'IBD2007 obtenue au même site pour chaque relevé de la série pluriannuelle. Cette dispersion rend compte de la variabilité de notation IBD obtenue au même site selon les années.

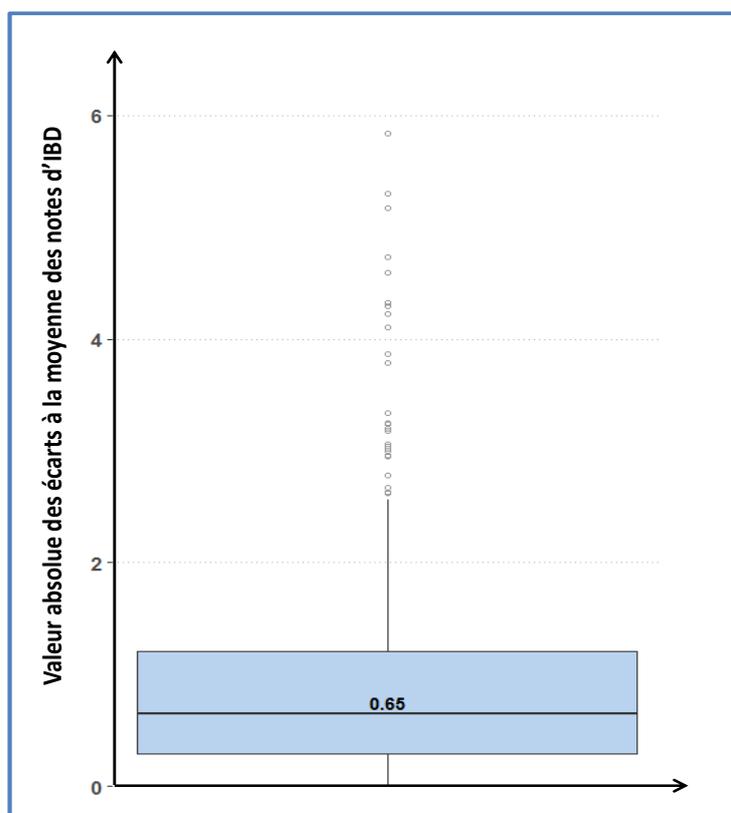


Figure 31 : Variabilité interannuelle naturelle de l'IBD2007 : Valeur absolue des écarts à la note moyenne obtenue au site (sur 103 sites)

Si l'analyse de la Figure 28 avait permis de repérer qu'une partie de la variation interannuelle était d'origine naturelle (c'est la part que l'on cherche à estimer ici), elle faisait aussi clairement ressortir que, sur les sites subissant un certain niveau d'anthropisation, venait se surajouter une part de variabilité interannuelle due à la variabilité de l'effet des pressions anthropiques susceptible de générer de gros écarts d'évaluation (pouvant s'élever à l'extrême jusqu'à 9,3 points d'IBD sur le même site selon l'année) et de dominer très nettement la part de variation naturelle de la méthode.

Cette forte variation d'effet interannuel des impacts anthropiques est la conséquence directe d'une plus forte variabilité des conditions hydrochimiques sur sites altérés, qui résulte d'une conjonction complexe entre les caractéristiques des pratiques et nuisances humaines du moment et des profils hydroclimatiques spécifiques à chaque année, qui vont provoquer un niveau très variable de transfert de polluants et de concentrations dans l'eau. Nous avons pensé maîtriser correctement cette variabilité par la sélection de sites à conditions hydrochimiques ressortant comme stables, cependant la faible représentativité temporelle d'échantillons ponctuels d'eau pour l'analyse chimique et le contenu hydrochimique insuffisamment assis de la chronique trimestrielle à mettre en correspondance avec chaque relevé diatomique (seulement 1 à 2 données seulement, au maximum, par paramètre et par site pour tout le trimestre concerné) n'ont pas permis de maîtriser correctement cette variabilité des effets interannuels de l'anthropisation.

Dans ces conditions, on a déjà pu établir au Chapitre 3.3.2 que l'assise de variabilité basée sur la totalité de la classe de TBE et de la classe de BE (qui représentent en tout 51% des relevés, dont les relevés dans la moitié inférieure de la classe de BE subissent déjà très probablement un niveau d'impact anthropique mesurable) **permettait d'appréhender de façon réaliste, voire un peu surestimée**, la part de variabilité du résultat de la méthode IBD la plus directement liée à la **variabilité interannuelle d'origine naturelle**. Par rapport à l'analyse des écarts 2 à 2 étudiée sur la base de la Figure 28, une présentation sous forme d'écart de chaque relevé par rapport à la moyenne au site (Figure 31) semble la base la plus correcte pour isoler et exprimer la variabilité interannuelle de la note d'IBD d'origine naturelle (de combien varie le résultat obtenu au site par un relevé collecté une année donnée par rapport au résultat moyen obtenu au site sur toute la chronique utilisée pour la comparaison ?).

En partant sur la même base méthodologique que pour l'analyse de la Figure 28 au Chap. 3.3.2, les écarts modérés les plus habituels, correspondant à 50 % des effectifs des relevés (valeur médiane), exprimés sous forme d'écart par rapport à l'évaluation moyenne au site sur la chronique de 4 relevés disponibles, permettent d'estimer une **variabilité interannuelle naturelle de l'IBD₂₀₀₇, qui s'établirait à +/- 0,65 point d'IBD.**

En élément de discussion, ce résultat est déjà probablement entaché d'une **surestimation modérée** due à la prise en compte de la variabilité totale basée sur les effectifs de relevés des classes complètes de TBE et de BE. En effet, une partie des sites classés en moyenne en BE subissent déjà des variabilités en partie occasionnées par une variation de l'impact anthropique, leurs relevés ayant dans ce cas de grandes chances de fluctuer entre la classe de BE et d'EM. En effet, la Figure 29 indique que 55% des sites fluctuent d'une classe de qualité, dont 46% (Figure 30) correspondent nettement au cas évoqué en fluctuant entre le BE et l'EM (soit en tout, 25% des 103 sites). En outre, certains sites restant en permanence en BE, ou fluctuant entre le TBE et le BE selon les années, peuvent aussi déjà subir une variation d'évaluation d'état dont une partie probablement modérée résulte d'une variabilité d'effet anthropique.

Donc le référentiel de variabilité interannuelle de l'IBD tel que quantifié ci-dessus **apparaît judicieux et en tout cas non sous-évalué**. Sans qu'il soit vraiment possible d'évaluer précisément la marge de sécurité qui a été prise, il a même tendance à un peu surestimer la variabilité interannuelle d'origine naturelle, du fait qu'au moins 25 % des sites sur lesquels s'appuie cette statistique (en particulier ceux pouvant fluctuer entre l'Etat Moyen et le Bon Etat) subissent aussi, selon les années, une partie de variabilité probablement non-négligeable liée à un effet différencié des impacts anthropiques.

3.4. Conclusions de l'étude de variabilité temporelle de l'IBD et mise en comparaison avec les résultats de variabilité inter-opérateur

L'intention initiale pour la réalisation de ce 2^{ème} volet d'étude était de baser cette approche sur la variabilité naturelle constatée, lors de chroniques interannuelles, sur des sites de référence validés, voire sur certains sites subissant des conditions essentiellement naturelle (sites de pseudo-référence), pas ou peu soumis à des incidences d'origine anthropique.

Cependant cette approche nécessitait une chronique suffisante de relevés acquis sur un panel représentatif des différentes conditions naturelles de notre territoire national. Or le réseau de référence mis en place au niveau national (RREF) n'a été suivi dans cette géométrie que sur 3 ans (2005 à 2007) et malheureusement, une évolution méthodologique sensible (passage de l'IBD₂₀₀₀ à l'IBD₂₀₀₇ en 2007), est venue couper en 2 la chronique acquise. D'autre part, trop peu de sites du RREF ont été repris en continuité immédiate dans d'autres réseaux type de RCS (cet état de fait va peut-être changer à partir de la mise en place du réseau de références pérennes...).

Au final, même si nous pensions au départ disposer d'une assise de données suffisante dans la base de données Pandore, l'assise temporelle disponible et la nécessité de s'appuyer sur une acquisition de données réalisée selon une base méthodologique comparable (2007-2011) ont conduit au constat que les chroniques acquises sur des sites de référence sur une base homogène étaient très insuffisantes pour évaluer de façon robuste une variabilité de la notation IBD due à une variabilité interannuelle d'origine naturelle.

Il a donc fallu ré-orienter cette démarche en procédant, à partir des données disponibles dans Pandore, à une sélection de sites présentant à la fois des critères de stabilité de leurs conditions chimiques et surveillés au moins 4 fois sur le créneau temporel [2007-2011], garantissant une certaine comparabilité des relevés diatomiques et de leur détermination taxonomique. Cette pratique a conduit au final à sélectionner 103 sites considérés comme stables sur le plan hydrochimique, présentant une chronique de 4 à 5 relevés diatomiques associables à un référentiel d'analyse de chimie de l'eau *in situ*.

Cette sélection contenant une proportion modérée de sites de référence à plutôt naturels (environ 25%) et une nette majorité de sites sous impact anthropique de modéré à fort (environ 75%), il a été nécessaire d'adopter une pratique spécifique assise sur la statistique des écarts afin d'isoler et d'évaluer de façon raisonnable la part de variabilité interannuelle d'origine naturelle.

Celle-ci, probablement un peu surévaluée, a été située à **+/- 0,65 point d'IBD** par rapport à la moyenne interannuelle au site. Ce référentiel de valeurs situe le niveau de **variabilité naturelle interannuelle de l'IBD** à un niveau assez comparable à l'estimation de la **variabilité interopérateur** obtenue au chapitre précédent (qui, exprimée sur la même base et elle aussi non sous-dimensionnée, a été située à **+/- 0,55 point d'IBD** par rapport à la moyenne interopérateur au site.

Ces estimations réalisées sans sélection drastique de données, donc incluant une certaine marge de sécurité, arrivent donc à des résultats proches de ceux obtenus lors de précédentes études (*e.g.* Roussel et al, 2011), et aussi de ce que les praticiens expérimentés avaient en tête (variation la plus habituelle de l'ordre de [**+/- 0,5 point d'IBD**], mais pouvant quelquefois atteindre [**+/- 1 point d'IBD**], selon particularités des situations locales).

Afin de mesurer l'impact du type de variabilité sur la méthode d'évaluation de l'état écologique des cours d'eau par le protocole « IBD₂₀₀₇ », il apparaît ensuite intéressant d'étudier de façon comparative les deux types de variabilité étudiées ici (*i.e.* Volet 1 et Volet 2 de notre programme d'étude), puis d'en utiliser le résultat afin de repercuter et propager l'incertitude qui en découle au niveau de l'évaluation au relevé et de l'évaluation intégrée de l'état écologique.

Pour cela, les distributions des différences des notes en EQR (prises deux à deux) observées en fonction du type de variabilité sur l'assise maximale de données d'écart disponible pour chaque étude ont été représentées en Figure 32.

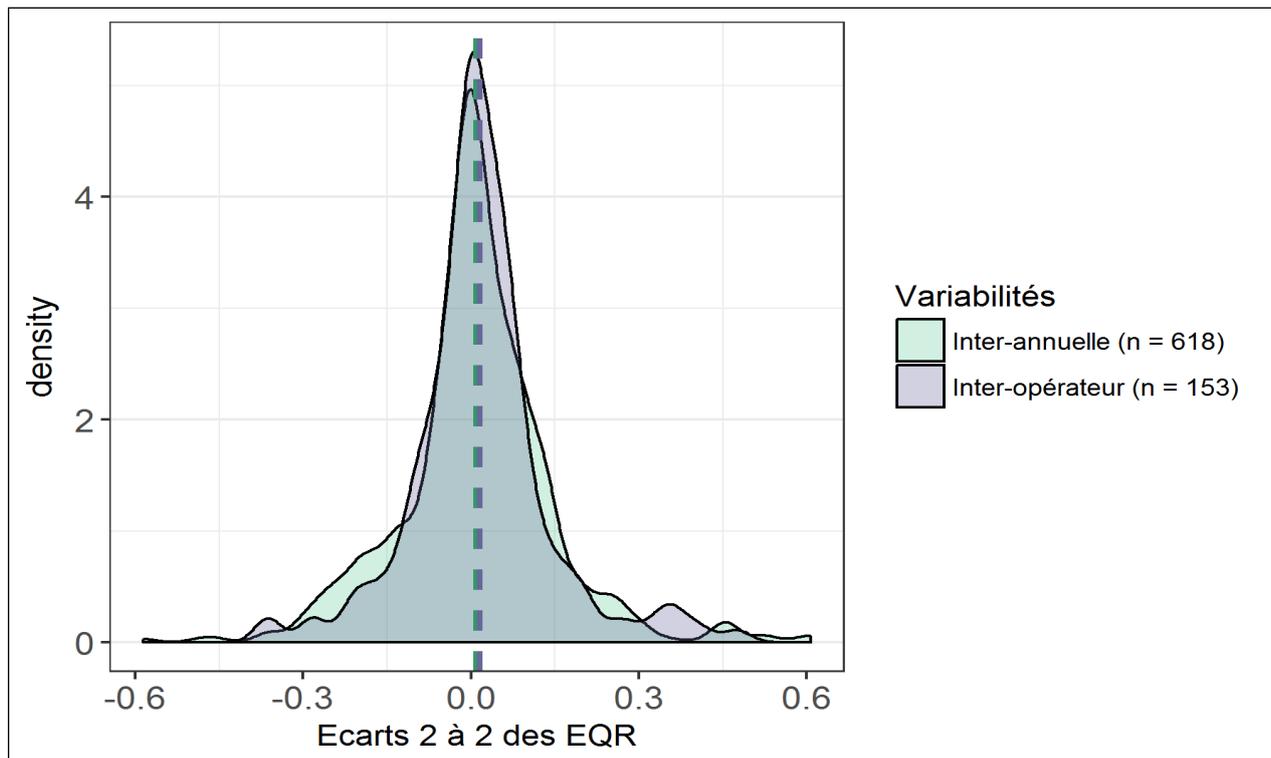


Figure 32 : Courbe de densité des écarts 2 à 2 des EQR, pour la variabilité inter-opérateurs et la variabilité interannuelle.

L'allure des courbes de densité est comparable pour les deux types de variabilités, avec une moyenne proche de 0 dans les 2 cas, une amplitude de -0.36 à +0.48 pour la variabilité liée à l'opérateur, et une amplitude s'étalant de -0.59 à +0.61 pour la variabilité interannuelle.

Les données ne suivant pas une loi normale ($p.value < 0.05$ pour les deux type de variabilités, test de normalité de Shapiro-Wilk) et l'effectif des deux échantillons étant différent, le test non paramétrique de Kruskal-Wallis ([REF. 7 : Kruskal & Wallis, 1952](#)) a été employé pour tester l'effet du type de la variabilité sur les écarts en note EQR. Ce test confirme qu'il n'y a pas de différence significative entre ces 2 sources de variabilité ($p.value = 0.6574$). Ces deux lois de distribution d'écarts sont donc comparables.

Les cas extrêmes de distribution des écarts, très peu nombreux dans l'assise statistique globale, correspondent à des cas particuliers qui ont été maintenus dans l'analyse. Il pouvait s'agir :

- d'une possibilité d'erreur de site ou de faciès par un opérateur, cependant non absolument démontrée, dans l'étude de la variabilité inter-opérateur des notes d'IBD ;

- d'une forte influence de variabilité d'effet anthropique venant se surajouter à la variabilité d'origine naturelle sur certains sites manifestement fortement anthropisés, au volet d'étude de la variabilité interannuelle naturelle des notes d'IBD).

L'assise des données représentées en Figure 32 mérite un commentaire. La stratégie prudente visant à ne pas toiler trop hâtivement des sites à écarts importants et suspects (sauf **erreur de site vraiment démontrée**) contribue à l'élargissement de la distribution jusqu'à quelques fortes valeurs d'EQR qui n'ont probablement plus rien à voir avec une **variabilité inter-opérateur réellement au même site**, ou avec une variabilité interannuelle d'**origine strictement naturelle**. Ces sites ou écarts extrêmes étant en effectifs marginaux, ils contribuent à un surdimensionnement modéré des domaines de fluctuation estimés dans ces 2 volets d'étude, le point positif étant que ce surdimensionnement procure une marge de sécurité et amène un peu plus de robustesse dans le dimensionnement des variabilités étudiées (variation inter-opérateur de l'IBD, variabilité interannuelle d'origine naturelle de l'IBD) et du même coup, dans l'intervalle de confiance attribué au final à la méthode d'évaluation.

Les structures de distribution des écarts 2 à 2 en EQR liés à la variabilité interannuelle et à la variabilité inter-opérateurs sont présentées en Figure 33 ci-dessous.

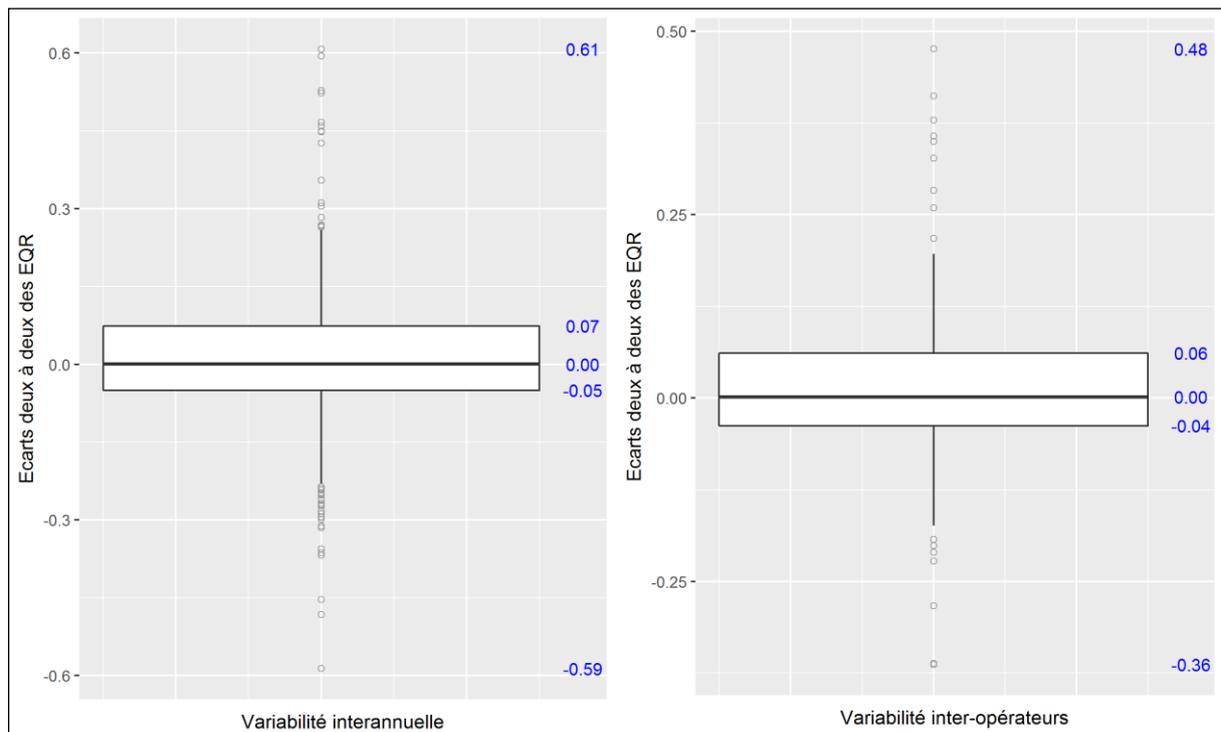


Figure 33 : Comparaison des box-plots de distribution des écarts 2 à 2 en EQR pour la variabilité inter-annuelle (à gauche) et pour la variabilité inter-opérateurs (à droite)

Ces résultats d'écarts, ici remis en correspondance avec le niveau de référence adéquat pour chaque site, ont été renormés de façon comparable et exprimés en EQR.

Ils donnent des caractéristiques de distribution des écarts très proches de celles produites avec les valeurs indicielles d'IBD₂₀₀₇, l'échelle d'expression étant bien sûr différente mais ne conduisant pas à une différence perceptible de structuration des écarts.

Remarque 1 : Dans chaque graphique, les échelles sont auto-adaptées par rapport à l'étendue propre des valeurs contenues, donc différentes et avec un positionnement du zéro un peu décalé entre ces 2 sous-figures. En tenant compte de cet aspect pour l'examen et l'intercomparaison, on peut constater que ces 2 lois de distribution d'écarts sont très comparables, en particulier sur les écarts hors outliers.

Remarque 2 : La signification des forts écarts contenus dans ces 2 lois de distribution n'a pas la même portée par rapport à la variabilité-cible que l'on veut étudier :

- **1) Pour la variabilité inter-opérateur**, il s'agit de sites où il est suspecté un gros défaut de comparabilité de pratique, la raison la plus souvent suspectée mais non complètement éclaircie en tenant probablement à une erreur de localisation de site ou de faciès, voire même de cours d'eau par un opérateur sur les 3 (2 relevés étant souvent plutôt comparables en pareil cas mais divergeant fortement du 3ème). Lorsqu'une erreur de localisation était clairement prouvée, cas qui a pu être effectivement rencontré lors de cette étude, le relevé en cause a été sorti de la comparaison. Mais, lorsque l'origine et la caractérisation de l'écart n'étaient pas certaines, le relevé douteux a été maintenu et conduit à de gros écarts de distribution figurés dans ce box-plot, notamment aux valeurs outliers.
- **2) Pour la variabilité naturelle interannuelle**, malgré le test préalable pratiqué pour sélectionner des sites stables sur le plan de leur hydrochimie, il est clair que certains sites sous altération anthropique forte et/ou fortement variable subissent une variabilité d'évaluation interannuelle beaucoup plus forte que des sites sous conditions naturelles. De tels sites anthropisés participent donc au renforcement des valeurs de certains écarts illustrés dans la distribution de droite, qui est plus large au niveau de l'assise des écarts possibles et surtout, qui inclut une proportion plus forte d'écarts importants. Dans ces cas s'éloignant du niveau d'écart le plus fréquent, une partie devenant très dominante de la variabilité n'est plus due qu'à une variation interannuelle des conditions naturelles du site (variabilité-cible de l'étude), mais à une variabilité d'expression du niveau d'impact anthropique sous l'effet des conditions particulières de l'année, qui se surajoute à la précédente et vient très nettement la dominer. Ces situations conduisent à des écarts forts, très souvent liés à des sites franchement altérés ou en tout cas, dont le niveau d'expression de l'altération est très contrasté selon les années (certains de ces sites pouvant être évalués avec 2 ou 3 classes de qualité d'écart selon les années). Il est donc logique que la distribution des écarts de la loi de distribution interannuelle soit plus étendue que dans l'étude des variabilités inter-opérateurs (on a des sites naturels, des sites altérés, des variabilités naturelles et des variabilités beaucoup plus fortes qui viennent se surajouter aux précédentes, en fonction de l'impact anthropique subi dans les conditions de l'année). Rappelons que le domaine de la **variabilité interannuelle d'origine naturelle** a été estimé en se limitant aux écarts les plus habituels présentés par 50 % de la population des relevés d'étude (soit, sur la base du diagramme de distribution situé en partie gauche de la Figure 33, l'étendue de la partie centrale du box-plot située entre -0,05 et +0,07 EQR, en écart de relevés 2 à 2).

Les écarts plus forts de cette distribution émanent de sites subissant manifestement des impacts anthropiques notables, qui augmentent très sensiblement la variabilité interannuelle de l'évaluation.

Comme indiqué à l'analyse de la Figure 32, la comparaison de ces 2 distributions de variabilités présente la même physionomie et ne révèle pas de différences significatives à l'analyse.

On peut ajouter que la distribution des écarts est très comparable sur la partie la plus habituelle des écarts constatés, qui est une partie de distribution bien établie et robuste. Mais il est important de considérer les référentiels avec plus de précaution lorsqu'on s'approche des limites externes de ces distributions et des valeurs extrêmes d'écarts. En effet, la distribution y est beaucoup moins consolidée et fait intervenir un risque accru d'intervention de phénomènes non-cibles, ou indésirables, ou atypiques (cas particuliers, écarts anormaux, erreurs diverses...).

Au bilan, l'ampleur de la **variabilité inter-annuelle d'origine naturelle** telle qu'établie (et probablement modérément surestimée) selon la méthodologie mise en œuvre dans ce volet d'étude, s'établit à **+/-1 point d'IBD d'écart entre relevés pris 2 à 2**. Sous un autre mode d'expression probablement plus correct et plus facilement manipulable, elle s'établit à **+/-0,65 point d'IBD par rapport à la moyenne interannuelle d'évaluation au site**.

Cette variabilité interannuelle d'origine naturelle pourrait être considérée comme une approximation *a maxima* de l'incertitude liée à la note d'IBD₂₀₀₇. En effet, elle est légèrement plus élevée que la variation inter-opérateur obtenue dans le cadre du 1^{er} volet d'étude, mais elle englobe de fait 2 sources de variabilités, à savoir : 1) une variabilité inter-opérateur globale (différents opérateurs ayant pu intervenir d'une année sur l'autre sur le même site en fonction de l'année et du marché public en vigueur), et 2) les variations inter-annuelles des conditions environnementales naturelles du site, imparfaitement reflétées par le jeu de paramètres physico-chimiques dont nous disposons.

Finalement, les résultats obtenus pour l'étude de ces deux types de variabilité, inter-opérateurs et inter-annuelle, ressortent comme plutôt comparables et d'un ordre de grandeur cohérent. Dans le cas où la mise en place d'un plan expérimental dédié à l'étude de la variabilité inter-opérateurs de la note, plan expérimental particulièrement lourd à mettre en œuvre, ne serait pas envisageable, une étude de la variabilité inter-annuelle des notes selon la méthodologie proposée et mise en œuvre ici peut en représenter une bonne approximation.

4 Calcul de la précision et de la confiance de l'évaluation faite à partir de l'IBD₂₀₀₇

4.1 Eléments de contexte sur l'évaluation

La DCE demande aux Etats-Membres d'intégrer dès que possible le concept de confiance et de précision des outils (notamment hydrobiologiques) dans la mise en application de l'évaluation d'état écologique, afin de pouvoir raisonner l'évaluation d'un site en termes de probabilités d'appartenance à telle ou telle classe d'état écologique (cf Figure 34 ci-dessous, issue de la Guidance Ecostat N°6).

Le résultat obtenu par un outil hydrobiologique dédié-maillon à partir de la composition spécifique d'un relevé biologique unitaire conduit à l'obtention d'une note donnée. Ce résultat converti en EQR régionalisé doit pouvoir ensuite être mis en correspondance avec une grille d'évaluation et décliné en probabilités d'appartenance à telle ou telle classe de qualité écologique.

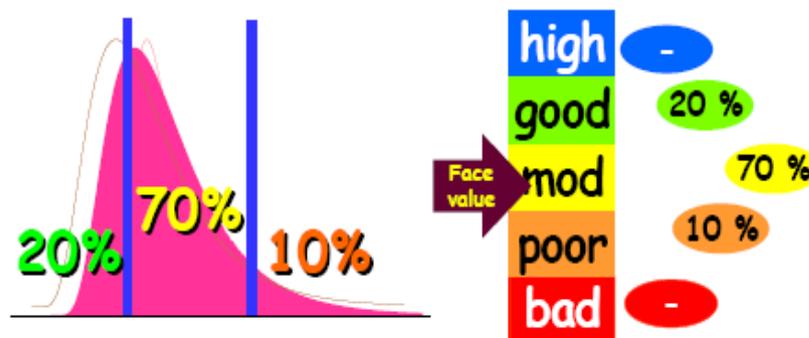


Figure 34 : Probabilité d'appartenance à une classe de qualité
(Guidance EC « Common Implementation Strategy N°7)

La mise en œuvre de ce concept nécessite d'établir de façon rigoureuse et robuste un intervalle de confiance associé à la production d'une note ou d'un EQR à partir d'une méthode d'évaluation donnée (dans notre cas, le système d'évaluation diatomique des cours d'eau de France, construit à partir des valeurs d'IBD₂₀₀₇).

En fonction du niveau de notation obtenu à partir d'un relevé et de l'intervalle de confiance attribué à la méthode (cf approches menées aux 2 chapitres précédents, qui visaient à préciser un intervalle de confiance global lié à l'effet interopérateur d'une part, à la variabilité interannuelle d'origine naturelle d'autre part), il est possible de décliner une statistique de probabilité d'Etat Ecologique attaché à cette notation, via la probabilité d'appartenance du couple (site x relevé) à telle ou telle classe de qualité (cf. partie droite de la Figure 34).

Contrairement aux apparences, la prise en compte de cet intervalle de confiance n'affaiblit pas mais au contraire consolide l'évaluation réalisée, en intégrant la variabilité non-maîtrisée de la méthode dans le processus d'évaluation et en confortant sur cette base l'interprétation qui peut vraiment être faite en terme de probabilité d'état écologique d'un site donné.

Cette probabilité d'affectation d'état doit déjà être pratiquée sur la base d'un seul relevé (évaluation annuelle), mais va aussi devoir l'être sur la base d'une série temporelle de relevés.

Dans le contexte Français, en effet, l'Arrêté Evaluation du 27-07-2015 prévoit ainsi de construire le rapportage européen d'état écologique sur la base des résultats collectés à un site, **moyennés par demi-plan de gestion des masses d'eau** (sur la base d'une moyenne de tous les résultats acquis sur les 3 ans). Ainsi, la méthodologie mise au point dans la suite de ce chapitre sera applicable en premier lieu à chaque relevé biologique individuel, et pourra ensuite sans problème être propagée afin de porter une évaluation moyennée d'état écologique du site de réseau au demi-PGME (soit sur 3 ans).

4.2 Etat de l'Art et approches méthodologiques

4.2.1. Références bibliographiques

Le processus que nous avons à mettre en œuvre à ce stade consiste à étudier et à propager l'incertitude de la méthode IBD2007 depuis la note obtenue au relevé jusqu'à l'attribution de la classe d'état écologique au site de réseau, en établissant un modèle de probabilité de classement à intégrer dans les règles d'évaluation.

Pour nous aider à réaliser cette étape, nous nous sommes basés sur les publications disponibles traitant de propagation d'incertitude et de détermination d'un intervalle de confiance dans les résultats de méthodes hydrobiologiques d'évaluation.

A l'examen, nous avons pu sélectionner deux références intéressantes, publiées par des scientifiques européens qui ont traité de ces sujets dans une optique de mise en œuvre imprégnée du cadre d'application de la DCE, *e.g.* (**REF. 8** : Ellis, J., Andriaenssens, V., 2006) et (**REF. 9** : Kelly et al., 2009), susceptibles de nous aider à mettre assez directement en place une démarche adaptée à notre contexte national de cours d'eau. Ces publications sont en assez bonne adéquation avec la problématique générale que nous avons à traiter et en particulier, mettent en œuvre des méthodologies facilement transposables dans notre cas.

La première référence est un rapport Européen de l'Agence de l'Environnement du Royaume-Uni qui traitait de l'incertitude attachée à l'évaluation de la qualité des cours d'eau à partir de l'indice poisson européen EFI. La seconde référence reprend et approfondit les mêmes réflexions et concepts pour les appliquer à la détermination de l'incertitude d'évaluation de l'état écologique de cours d'eau et de plans d'eau, en liaison avec l'utilisation de méthodes phytobenthiques.

Ces 2 valorisations, les vocables et les réflexions méthodologiques qu'elle contiennent nous ont été utiles pour réfléchir, mettre au point et réaliser, à partir des données diatomiques françaises, une démarche finalisée permettant de traiter de l'incertitude de la méthode IBD et de lui calculer un intervalle de confiance utilisable dans la procédure d'évaluation de l'état écologique (EEE) des cours d'eau de France.

D'après ces auteurs, la confiance associée à la note EQR d'un relevé peut être définie à la fois comme la **probabilité que cette note appartienne à une classe écologique donnée**, mais aussi comme une **approche du risque de se tromper dans le classement d'état écologique** que l'on attribue à un site, à une masse d'eau... Ces 2 notions aident à mesurer la confiance qu'il est possible d'accorder à un classement d'état écologique. Dans cette évaluation d'état, il est également important de connaître comment le positionnement des limites de classes peut venir impacter le classement du cours d'eau.

En effet, le fait qu'un site se classe à proximité d'une zone de basculement de classe de qualité rend forcément plus incertain le classement d'état écologique obtenu que si son état ressort bien centré sur une classe de qualité.

Dans notre propre démarche, nous aurons donc à combiner l'imprécision résultant d'une marge de fluctuation incontrôlée de la méthode diatomique IBD d'une part, et de l'influence statistique de la proximité des limites de classes d'autre part, pour attribuer à un site donné (ou à une masse d'eau donnée), un classement d'état écologique exprimé en termes de probabilité d'appartenance à une classe d'état.

4.2.2. Méthodologie et terminologie associée

La Figure 35 ci-dessous, qui reprend la Figure 15 de la publication de (Kelly et al., 2009), présente un diagramme schématisé permettant de présenter de façon théorique un exemple de calcul de la confiance dans le classement et d'illustrer la terminologie spécifique qui y est liée.

Sur la base de l'exemple théorique traité dans cette figure, *i* représente la valeur d'EQR obtenue pour le relevé considéré. L'aire grisée sous la courbe représente la distribution de la confiance que l'on peut associer à cette valeur de l'EQR, basée sur l'écart-type prédit *ii*.

Les lignes verticales représentent les limites entre les classes d'état (de la gauche vers la droite : B : Bad ou Mauvais Etat ; P : Poor ou état médiocre ; M : Moderate ou Etat Moyen ; G : Good ou Bon Etat ; H : High ou Très Bon Etat).

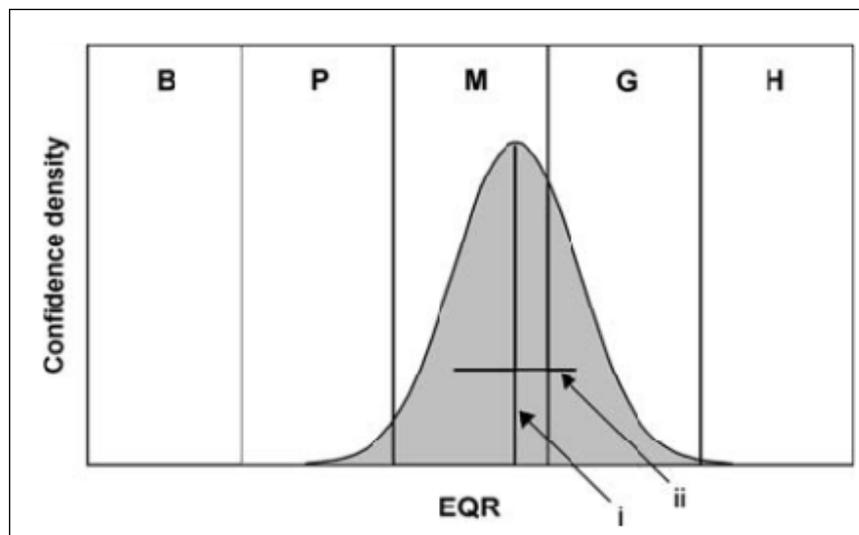


Figure 35 : Diagramme schématisé illustratif de la base de calcul de la confiance du classement et du risque de mauvaise classification d'un site.

Le mode de la distribution de probabilité d'état (*i.e.* probabilité d'état la plus forte) est centré sur la valeur d'EQR observée. Dans l'exemple présenté, la distribution de l'écart-type prédit (*ii*) chevauche la classe d'Etat Moyen et la classe de Bon Etat. La valeur observée de l'EQR suggère donc plutôt une qualité moyenne, assez proche de la limite EM-BE, mais il existe aussi une probabilité, qui est à prendre en considération, pour que le vrai état du site soit le Bon Etat.

Afin de pouvoir établir l'écart-type de la variabilité d'une méthode biologique (le ii de la figure précédente) et donner un intervalle de confiance sur une évaluation d'état écologique qui en découle, il est nécessaire de disposer de données de base permettant de travailler sur une répétition d'évaluations et de cerner une variabilité standard qui ait du sens par rapport à la variabilité intrinsèque de cette même méthode. Une discussion à ce sujet sera menée plus loin, au moment de la réflexion visant à décliner de la façon la plus adéquate possible l'incertitude de la méthode IBD₂₀₀₇ en intervalle de confiance et en probabilités de classement d'état de sites français à évaluer, situés sur des hydrosystèmes « cours d'eau ».

A ce stade, continuons d'examiner l'intérêt potentiel des approches méthodologiques déployées par ces auteurs en vue de nous aider à situer un intervalle de confiance au classement d'état écologique (diatomique) de sites de cours d'eau de France, dans le cadre de réseaux de surveillance de type RCS visant le rapportage européen d'Etat Ecologique.

Dans ces 2 références bibliographiques, qui concernaient 2 maillons biologiques différents (poissons, phytobenthos), le calcul de la confiance s'est effectué selon une démarche assez analogue, faisant se succéder les étapes distinctes suivantes :

- Etablissement de la relation entre écart-type et moyenne des notes d'EQR suivant une fonction polynomiale
- Définition de la loi de probabilité associée à cette fonction et aux limites des classes d'état écologiques
- Calcul de la confiance d'appartenance aux différentes classes d'état écologique

Les jeux de données de travail dont ont disposé ces auteurs pour disposer d'une variabilité de résultats de leur méthode hydrobiologique au même site ont le mérite d'exister et de représenter un travail de capitalisation antérieur très conséquent. Cependant, fait souligné par Ellis pour le jeu de données EFI, comme ils n'avaient pas été acquis spécifiquement dans le but de situer et préciser les différentes sources d'incertitudes, leurs protocoles d'acquisition n'avaient pas été stratifiés en vue d'isoler les diverses sources de variabilités impliquées dans la variation globale du résultat au même site. Ils présentaient donc l'inconvénient de ne pas être pas forcément suffisamment assis et adéquats pour établir proprement la part des diverses variabilités (variabilité naturelle temporelle, variabilité inter-opérateur, variabilité d'effet des altérations anthropiques...). Ces différentes sources de variabilités ont donc été un peu amalgamées dans la détermination de l'intervalle de confiance attaché à la méthode.

Les jeux de données disponibles pour Ellis (2006) concernaient respectivement l'indice poisson EFI sur 6 cours d'eau UK, portant sur le maillon biologique poisson (très intégrateur, à la fois en dimension spatiale et en dimension temporelle (de 3 à 5 ans).

Les jeux de données disponibles pour Kelly et al (2009) portaient sur les résultats de l'indice diatomique TDI, sur la base de jeux de données UK centrés sur le maillon phytobenthos (dont la capacité d'intégration temporelle est beaucoup plus courte que pour le maillon précédent (de l'ordre de 2 à 3 mois maximum). Les 2 jeux de données phytobenthiques ont été acquis **sur lacs d'une part**, systèmes très intégrateurs du fait de leur volume, de leur stock, de leur temps de renouvellement, et **sur cours d'eau d'autre part**, systèmes ouverts beaucoup plus fluctuants sur le plan spatial et temporel.

L'utilisation a été faite de ces jeux de données pourrait susciter divers sujets de discussion vis-à-vis des types de variabilités qu'ils contiennent et sur l'objectif poursuivi pour l'affectation d'un intervalle de confiance (sur la méthode hydrobiologique vs sur la classification d'état), ce qui n'a pas semblé clairement posé par ces auteurs.

Plus que la présentation d'une démarche passe-partout qui serait fournie « clés en mains » pour établir l'intervalle de confiance attaché à une méthode hydrobiologique d'évaluation lambda, ce qui n'était d'ailleurs pas l'objectif affiché de ces auteurs, l'intérêt de ces papiers réside donc dans les **méthodologies mises en œuvre et dans l'intérêt qu'elles peuvent revêtir** pour venir appuyer les différentes étapes de notre propre démarche et nous aider à la réaliser. En effet, sans réflexion spécifique de relativisation par rapport au type de système et aux caractéristiques du maillon, il semble délicat de traiter de l'intervalle de confiance de méthodes biologiques portant sur des maillons biologiques si différents, dans le cadre de types d'hydrosystèmes (rivières, lacs) eux aussi bien différents dans leur mode de fonctionnement et leur dynamique propre.

Les approches méthodologiques utilisées par les auteurs pour aborder les phases listées en page précédente ont été très comparable entre ces 2 références, au maillon biologique près. Elles sont donc présentées et décrites dans les 2 paragraphes ci-dessous.

4.2.3. Relation entre écart-type et moyenne

Sur la base de la distribution des variations de résultats de la méthode biologique au même site, l'idée intéressante mise en œuvre par ces auteurs a été d'établir une relation entre moyenne des notes d'EQR (axe x) et écart-type des EQR (standard deviation ou SD, en axe des y), basée sur les écarts effectivement constatés au même site dans leur jeu de données (voir Figure 36 ci-dessous).

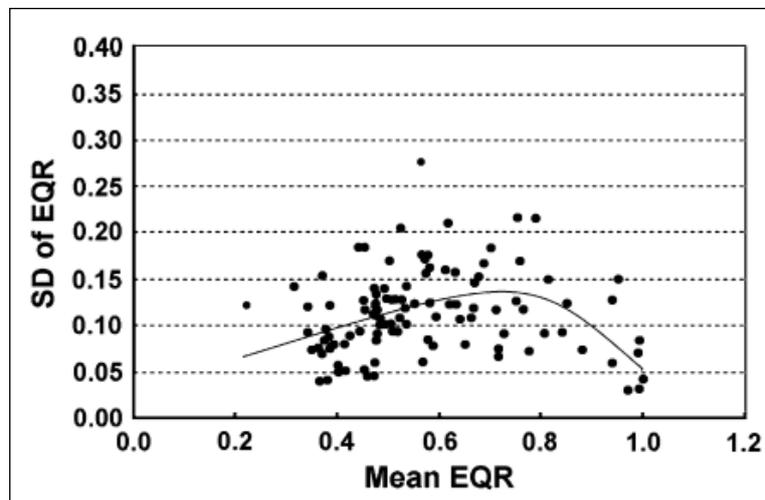


Figure 36 : Variabilité intra-site : mise en relation de l'écart-type de la variation au site par rapport à la moyenne des EQR au site. (Kelly et al., 2009).

La courbe représentée sur cette figure, établie à partir des variabilités site par site, représente la fonction polynomiale de la relation entre la moyenne des EQR et l'écart-type des EQR. Elle a été déterminée suivant la méthode de régression locale non paramétrique Loess (REF. 10 : Cleveland and Devlin, 1988), afin de calculer l'écart-type attendu pour chaque valeur d'EQR et plus précisément, pour les valeurs d'EQR correspondant aux limites de chaque classe d'état écologique (si).

Elle conduit logiquement les auteurs à conclure que les plus grandes variabilités d'état interviennent dans le milieu du gradient d'altération, et à souligner que cela rend plus incertaine l'évaluation d'état dans cette zone proche de la limite basse de Bon Etat et au moins sur toute la classe d'Etat Moyen, alors que cette zone est stratégique pour l'application de la DCE (statistique nationale d'état écologique et son évolution, mise en place des programmes de mesures correctives, dont le coût important est évoqué...).

Outre ces aspects qui sont importants aussi à faire ressortir dans le contexte Français, une conclusion qu'ils n'ont pas vraiment dégagée est que la variabilité des résultats et, par contrecoup, la largeur de l'intervalle de confiance, sont sensiblement plus étroites dans la classe de TBE et dans une bonne partie de la classe de BE. Indirectement, cela montre bien le rôle important des altérations anthropiques et de la variabilité des effets de l'altération anthropique, du moins sur des sites de modérément à moyennement pollués, selon les caractéristiques hydroclimatiques et d'usages anthropiques liés à une année donnée. C'est l'ajout de ces variabilités d'origine anthropique qui est à la fois la cause et l'explication principale de cet accroissement de la variabilité globale du résultat de la méthode entre les sites naturels ou pseudo-naturels et les sites plus altérés.

A notre sens, cela confirme nettement qu'il faut éviter de considérer de la même façon :

- 1) la variabilité intrinsèque à l'application d'une méthode (effets des variations naturelles, divers effets liés aux pratiques d'opérateurs différents, aux effets intra-opérateurs...) qui conduisent à une marge de variation incontrôlée du résultat sans relation avec les effets de l'anthropisation (ces fluctuations non-maîtrisables constituant une **faiblesse de la méthode**),
- et 2) les variations de résultats liées aux conséquences sur le compartiment diatomique des impacts anthropiques, que la méthode est chargée de mesurer de façon la plus précise et sensible possible. Si l'application des effets de l'anthropisation est variable selon le moment (saison, année) et la climatologie, ce qui est tout-à-fait logique et bien connu dans le cas des cours d'eau (l'impact anthropique n'a aucune raison d'être constant et égal...), l'aptitude de la méthode à détecter des niveaux variés d'impact anthropique est normale (ils ont été construits pour y répondre le mieux possible) et constitue au contraire une **qualité** à mettre à l'actif de la **performance de la méthode**.

Dans notre propre démarche, à l'instar de ce qui a déjà été fait lors de la démarche d'évaluation de la variabilité interannuelle d'origine naturelle de l'IBD 2007 (cf Chapitre 3 du présent rapport), il faudra donc éviter d'amalgamer ces 2 types de variabilités, dont l'une est indésirable et l'autre correspond au contraire à une bonne sensibilité de réponse de la méthode biologique utilisée (variation voulue du niveau de réponse en fonction de l'intensité de l'altération anthropique sur l'épisode concerné), dans l'établissement de l'intervalle de confiance servant à conforter l'évaluation d'état écologique.

Dans le contexte du Royaume-Uni que nous cernons probablement mal, il est par contre plus difficile d'argumenter de façon fondée la tendance à la baisse de variabilité du résultat biologique pour les situations très altérées. Nous commenterons donc cet aspect au chapitre des résultats sur la base de la loi de relation telle qu'elle s'est établie en France, à partir d'un contexte d'altération des cours d'eau que nous connaissons mieux.

4.2.4. Loi de probabilité d'Etat en fonction de la valeur d'EQR

Cette étape permet de déterminer le niveau de confiance pour une note en EQR d'appartenir à une classe de qualité donnée. Autrement dit, pour chaque classe i , c'est la probabilité $p(i)$ d'observer une note en EQR appartenant à cette classe.

La probabilité p_i d'appartenance à la classe i est dérivée de la loi de probabilité dont les paramètres sont l'EQR de la limite de classe i (noté L_i) et l'écart-type correspondant obtenu à l'étape précédente (noté S_i). La probabilité $p(i)$ est donc calculée comme suit :

$$p(i) = P(X \geq x \mid \mu = L_i)$$

$$p(i) = 1 - \Phi \left\{ \frac{(x - \mu)}{s_i} \right\}$$

Avec :

- i la classe d'état écologique
- x la note EQR
- μ la moyenne (ici la moyenne a été considérée comme la valeur de l'EQR à la limite L de la classe i , L_i)
- s_i l'écart-type à la limite de la classe i
- Φ la probabilité normale cumulée

Ainsi, il est possible de calculer la loi de probabilité d'appartenance à chaque classe d'état pour tous les EQR obtenus :

- Confiance TME = $100 p_{TME}$
- Confiance ME = $100(p_{ME} - p_{TME})$
- Confiance EM = $100(p_{EM} - p_{ME})$
- Confiance BE = $100(p_{BE} - p_{EM})$
- Confiance TBE = $100(1 - p_{BE})$

Un résultat tel qu'illustré en Figure 37 (page suivante) est alors obtenu. Précisons que cette figure adaptée est extraite de la Figure 3 (2^{ème} partie) de la référence *Kelly et al, 2009*.

Cette figure illustre la probabilité de prédiction de l'état écologique correct en fonction de la note d'EQR obtenue, sur la base ici de 6 échantillons différents pour une période de classification d'état de 3 ans.

Si les réplicats sont moins nombreux (par exemple 3 échantillons pour une période de 3 ans), cela va conduire à plus de chevauchement entre des classes moins différenciées, et les probabilités d'appartenir à l'une ou à l'autre seront définies de façon moins précise.

Ce que l'auteur souhaite mettre ici en avant est l'importance de disposer d'une plus grande répétition d'échantillonnages si l'on veut classifier l'état intégré du site avec plus de chances d'être dans la bonne classe sur la même durée temporelle.

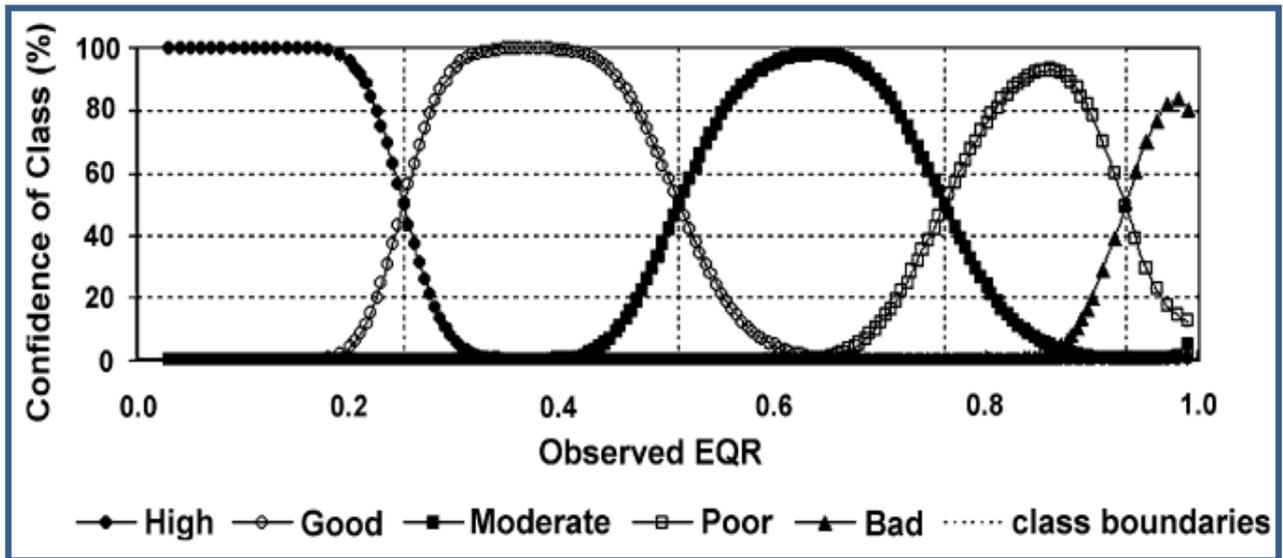


Figure 37 : Confiance dans la prédiction de classe correcte de qualité pour les cours d'eau du Royaume-Uni en fonction de l'EQR observé (d'après Kelly et al, 2009, Figure 3)

4.2.5. Confiance d'appartenance aux classes de qualité

La Figure 38 qui suit (reprise de la Figure 4 de Kelly et al, 2009) présente le résultat du calcul de la confiance d'appartenance aux différentes classes d'état écologique en fonction du nombre de réplicats de la mesure faits sur un pas de temps donné au même site (cours d'eau du Royaume-Uni).

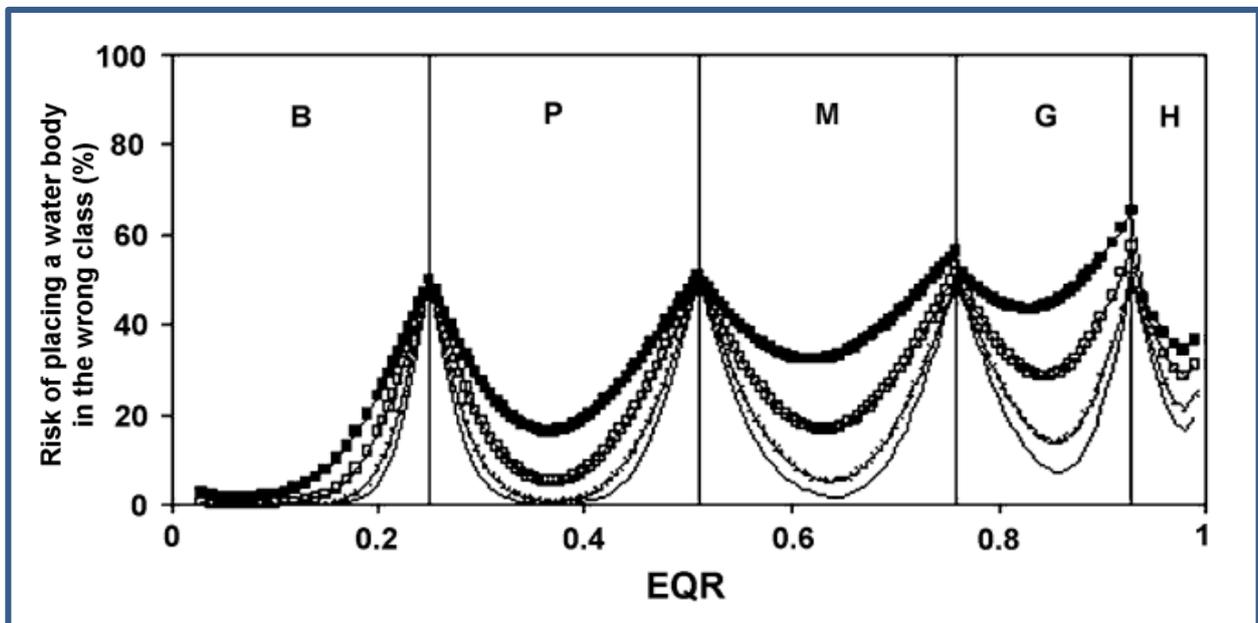


Figure 38 : Confiance dans la prédiction de classe correcte de qualité : Risque correspondant de mauvaise classification pour des rivières du Royaume-Uni en fonction du nombre de réplicats

Dans tous les cas, lorsqu'on se situe précisément à une jonction de classes d'état, la probabilité d'appartenance à une classe de qualité ou à l'autre est de 50 %. Plus on va avoir de réplicats sur le même pas de temps et au même site, plus l'appartenance à une classe donnée va se préciser et se renforcer (courbes du bas, qui se creusent et dont la probabilité de recouvrement de classes diminue). Si peu de réplicats sont disponibles, la probabilité alternative d'appartenance du site à une classe de qualité ou à une autre sera moins bien définie. Le propos de l'auteur, là aussi, vise à renforcer le nombre de suivis annuels ou trisannuels du maillon diatomique si l'on veut améliorer la confiance dans le classement d'état écologique décerné, une des raisons étant que le maillon diatomique a un temps d'intégration biologique relativement court et ne pourra pas être complètement représentatif, par exemple, d'une année complète de chimie de l'eau et d'effet des pressions anthropiques.

4.3 Résultats de l'approche menée dans le cas de la France

4.3.1. Réflexions préalables sur la manière d'aborder la confiance de classement d'état

Si le contenu des 2 références précitées permet de théoriser la façon d'aborder la confiance dans le classement d'état écologique et de proposer des outils méthodologiques permettant de la décliner à partir de données exprimant une variabilité de résultats d'une méthode biologique, une question à bien cerner au départ est : quelle variabilité de la méthode souhaite-t-on maîtriser dans le cadre de l'utilisation de l'intervalle de confiance ?

Les 2 références bibliographiques précitées n'ont pas été réalisées sur le même maillon biologique, parfois pas sur le même type d'hydrosystème (cours d'eau et maillon poisson pour EFI, cours d'eau et lacs pour TDI), et n'étaient pas toujours focalisées sur le même type de variabilité.

Ellis et al 2006 a abordé, sur la base et dans la limite des jeux de données dont il disposait, plusieurs types de variabilités d'évaluation d'état à partir de l'EFI, en particulier spatiale entre cours d'eau, spatiale au sein d'un même cours d'eau, spatiale sur la base de variations locales au site, temporelle (saisonnaire, interannuelle), relatives à l'erreur d'échantillonnage et de détermination (effet intra et inter-opérateur).

Afin de travailler sur les variabilités de la méthode diatomique liée au TDI (Trophic Diatom Index), *Kelly et al* 2008 a travaillé à partir de chroniques interannuelles obtenues durant 6 années sur les mêmes sites en lacs d'une part, en cours d'eau d'autre part, représentant tout l'éventail de qualités possibles dans le contexte du pays (qui est différent du nôtre : il faut notamment relever que l'agriculture est nettement moins intensifiée que chez nous sur le plan de l'irrigation, des intrants fertilisants et phytosanitaires, et laisse beaucoup plus de place à de l'élevage sur prairies naturelles).

Pour ces 2 approches, le point éventuellement un peu criticable est que toutes les sources de variabilité des résultats ont été amalgamées et ont contribué globalement à la définition d'un intervalle de confiance sur le classement d'état écologique, qui était l'objectif final unique de l'approche (raisonner sur la confiance dans le classement d'état écologique obtenu à ce site).

Cette approche est probablement défendable en lacs, systèmes fermés plus ou moins en équilibre interne et dont l'état écologique interne moyen va varier assez lentement sur des pas de temps longs, du fait du fort volume intégrateur du lac, même si une variabilité interannuelle du résultat de la méthode biologique peut intervenir du fait de la variabilité naturelle des conditions de l'année. Dans ce cas, cerner la variabilité naturelle de la méthode a du sens par rapport à la problématique de fixation d'un intervalle de confiance en fonction de la variabilité de son résultat et aussi, sur la probabilité de bonne classification d'état écologique en fonction de la variabilité interannuelle des résultats de la méthode sous forçage naturel. En effet, l'état écologique moyen d'un lac ou d'un autre plan d'eau ne va en principe pas changer rapidement d'une année sur l'autre en fonction d'une dégradation nouvelle et rapide liée à une évolution de la pression anthropique à l'échelle de toute la masse d'eau (qui représente un volume et un volant d'inertie).

Sur de tels hydrosystèmes, c'est donc plutôt une **variabilité naturelle annuelle d'évaluation d'état** qui est très principalement en jeu sur de tels intervalles de temps, et même sur le plan de gestion de 6 ans. En d'autres termes, malgré le gradient d'anthropisation ou le gradient trophique sur lequel était assis le jeu de données en lacs, c'est de façon très dominante la variabilité naturelle, ici interannuelle ou plus largement temporelle (les saisons provoquent aussi des variabilités sensibles sur le plan biologique), qui va être cernée par le type de démarche appliquée.

Cette façon de poser l'analyse pose déjà nettement plus question dans le cadre de cours d'eau, systèmes ouverts soumis à une variation forte des effets de la pression anthropique en fonction des usages et pratiques humaines ayant lieu l'année donnée, ainsi qu'à des variations climatiques et hydrologiques provoquant, selon le cas, très peu ou beaucoup de transferts de nutriments et de polluants, pas de dilution ou de fortes dilutions, des effets anthropiques faibles ou forts selon l'année ou même l'épisode considéré... ; sachant de plus que c'est le cumul des concentrations instantanées qui va déterminer le niveau de l'effet trophique ou toxique sur un maillon tel que les diatomées. Cette très forte variabilité annuelle de qualité des eaux inhérente aux cours d'eau est une réalité très prégnante sur le territoire Français, qui subit une pression anthropique omniprésente d'origine agricole, domestique, urbaine, de STEP..., depuis les stations de montagne jusqu'au littoral, avec une incidence déterminante du faciès hydroclimatique de l'année sur l'intensité des effets de pollution subis *in situ*.

Même sur le réseau de cours d'eau, le contexte est certainement moins fortement anthropisé en Angleterre, notamment sur le plan de l'agriculture intensive et d'implantations urbaines en zones d'altitude (notamment, pas ou très peu de stations et villages de ski, même si le cas existe en Ecosse). Mais, du moins pour la France, il ne semble pas judicieux ni raisonnable de fixer un intervalle de confiance mélangeant à la fois une variabilité naturelle incontrôlée et non maîtrisable intrinsèque à la méthode biologique (qui en constitue une faiblesse), et une variabilité interannuelle de résultat liée à une variation des pollutions, ou à une variabilité d'expression des effets de cette pollution en fonction des usages anthropiques et du profil hydroclimatique des années (qui au contraire, en constitue une qualité et une force).

En effet, sur des territoires anthropisés de France, il n'y a aucune raison que l'état écologique d'un cours d'eau, qu'il soit évalué à partir de descripteurs abiotiques de pollution ou par des maillons biologiques-clés, doive être constant dans le temps. Cette façon un peu simpliste d'aborder le problème est peut-être liée à l'expérience capitalisée sur le compartiment poisson, beaucoup plus intégrateur dans le temps et dans l'espace que le compartiment diatomique.

En cours d'eau, nous sommes en système ouvert, avec des usages variables, une climatologie et des transferts variables, des conditions favorables à la concentration des polluants (étiages sévères) ou au contraire, dilutives (forts débits)...

Enfin, les méthodes diatomiques indicielles dont l'IBD ont été spécifiquement élaborées pour répondre de façon sensible et la plus forte possible à des effets variables de l'anthropisation sur la qualité et le niveau d'enrichissement en nutriments des eaux. La réponse de ce maillon est donc dynamique en fonction du degré d'effet de pollutions anthropiques responsable d'un enrichissement en nutriments.

Sur des stations anthropisées, ce qui s'est d'ailleurs une nouvelle fois bien vérifié dans le volet d'étude des variabilités interannuelles des notes d'IBD figurant dans le présent rapport (cf Chapitre 3), il est normal que l'évaluation d'état soit variable et même, beaucoup plus variable que sur des sites de référence ou pseudo-naturels, du fait d'une forte variabilité interannuelle d'effet des pollutions anthropiques, qui va dominer très largement la variabilité intrinsèque de la méthode liée à la variabilité des conditions de forçage naturel qui s'appliquent au site à évaluer.

Pour la détermination d'un intervalle de confiance, il nous semble donc important d'isoler ce qui résulte de l'imprécision difficilement maîtrisable intrinsèque à la méthode et/ou à sa mise en œuvre par un opérateur donné, qui en constitue une **faiblesse** (marge de fluctuation mal maîtrisée qui peut résulter d'un **effet inter-opérateur**, de la variation liée aux facteurs de forçage **naturels**, saisonniers ou interannuels ...), des fluctuations de résultat occasionnés par la **variation d'intensité d'effet d'impacts anthropiques**, qui constituent à la fois **une qualité et la raison d'être d'une méthode hydrobiologique performante**.

Selon notre analyse, les aspects de **variabilité non maîtrisée** d'une méthode biologique et leurs conséquences sur l'incertitude de l'évaluation doivent être gérés par un intervalle de confiance assorti aux résultats de la méthode et du classement. Par contre, les **variabilités de résultats de la méthode biologique** occasionnées par des **variations interannuelles d'intensité d'impacts anthropiques**, qui sont un phénomène normal en systèmes de type ouvert comme les cours d'eau et dont la méthode biologique a justement été élaborée pour rendre compte, **ne sont pas à gérer par une logique d'intervalle de confiance**, qui met en doute la **pertinence des fluctuations d'évaluation** observées sur une série temporelle. Celles-ci sont normales et font partie du monde réel, elles doivent plutôt être prises en charge par une stratégie d'agrégation adéquate prévue dans la méthode d'évaluation d'état.

C'est dans ce sens que, pour l'évaluation diatomique et, plus largement, hydrobiologique des cours d'eau de France, le dispositif national d'évaluation prévoit une agrégation des résultats triennaux par 1/2 plan de gestion, sur la base du calcul d'une moyenne trisannuelle, pour statuer sur l'état écologique du site intégré à cette maille temporelle. L'intervention de **l'intervalle de confiance** ne doit pas avoir pour but de gérer les variabilités interannuelles dues à des variations d'effets anthropiques, mais à modérer le résultat d'évaluation obtenu en fonction de fluctuations incontrôlées de la méthode pour d'autres raisons que l'impact anthropique (variabilités temporelles naturelles, effet inter-opérateur...).

En vue de situer l'intervalle de confiance de la méthode IBD sur l'évaluation d'état écologique (EEE), nous avons donc finalement choisi de nous appuyer sur les travaux et les référentiels concernant **l'étude de la variabilité inter-opérateurs 2013-2014** (cf. Chapitre 2), qui ont été acquis par 3 types d'opérateurs différents en contexte temporel contraint (au maximum 2 semaines de décalage).

L'intérêt est que, sauf erreur de site éventuelle, parfois suspectée mais très peu fréquente et non absolument certaine, la variabilité exprimée résulte bien de **fluctuations non maîtrisées inhérentes à l'application de la méthode**, mais pas à une variation temporelle importante susceptible de s'accompagner de l'application d'un niveau d'impact anthropique très différencié.

Ce référentiel capitalisé en 2013 et 2014 nous paraît donc **particulièrement intéressant** pour cerner la **marge de variation aléatoire et non maîtrisable de la méthode**, liée à plusieurs catégories d'effets liés à la mise en œuvre de la méthode (variation naturelle, effet inter-opérateur, effet intra-opérateur...), mais en tout cas, **non liés à la faculté de la méthode à rendre compte du niveau d'altération anthropique**, qui constitue sa raison d'être en même temps qu'un gage de sa qualité.

Pour notre étude de détermination d'un intervalle de confiance autour de l'évaluation faite à partir de l'IBD, nous nous sommes donc appuyés sur les résultats et les écarts d'EQR-IBD2007 obtenus pour chacune des 3 catégories d'opérateurs participants sur chacun des 61 sites de la campagne expérimentale dédiée (cf. les 2 campagnes de collecte de terrain organisées en 2013 et 2014).

4.3.2. Relation entre écart-type et moyenne

Chacun des 61 sites d'étude a été prélevé par 3 opérateurs appartenant à des types différents (organisme scientifique, spécialiste-maillon des DREAL, Bureau d'étude). il est donc caractérisé par 3 EQR-IBD. Pour chaque site on peut donc calculer une moyenne et un écart-type des notes obtenues, puis mettre en relation ces 2 paramètres de la même façon que dans la référence Kelly et al, 2009 (cf. **Figure 36**).

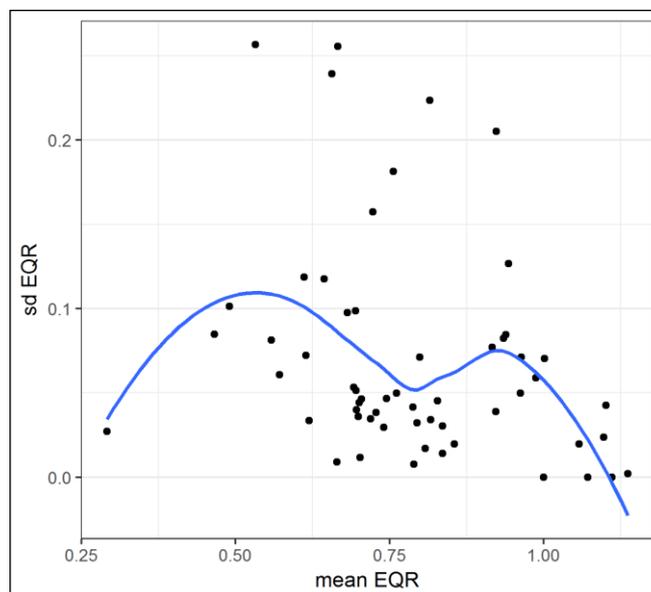


Figure 39 : Relation entre l'écart-type et les moyennes des EQR des sites et la fonction polynomiale associée (courbe en bleue) pour la variabilité inter-opérateurs.

En suivant cette méthodologie, la moyenne des EQR estimés par chaque opérateur a été calculée par site, et la relation avec son écart-type a pu être établie par une fonction polynomiale suivant la méthode non paramétrique Loess (cf. **Figure 39**), sur la base de 61 répliquats d'observations.

La variabilité pour un EQR donné est non négligeable, cependant l'écart-type des EQR observé à chaque site n'excède pas 0.3 et la plupart des sites obtiennent un écart-type inférieur à 0.1.

Pour les sites avec un écart-type de plus de 0,1 (10 sites sur 61), il s'est posé un pb de répétabilité pour un intervenant par rapport à 2 autres qui obtenaient un résultat assez homogène, sans avoir la preuve formelle du problème rencontré. Les hypothèses, malheureusement non traçables dans ce cas, pourraient résider dans l'échantillonnage d'un site différent, ou d'un faciès différent peut-être affecté d'une particularité locale associée à un rejet ponctuel, ou dans l'intervention d'un épisode hydroclimatique fugace, non repérable *a posteriori* au niveau national, entre le passage de 2 intervenants et du 3^{ème} (de type orages localisés), d'une inversion d'échantillons chez un opérateur, de différences liées à la qualification taxonomique des opérateurs...

Lorsque la raison n'a pas pu être clairement objectivée, ce qui a été fait dans quelques autres cas (analyse de doublets au lieu de triplets), le relevé douteux a été maintenu dans l'analyse des écarts et contribue ainsi à conforter l'intervalle de confiance, intégrant ainsi ce qui est susceptible d'intervenir dans une pratique routinière de réseau (il peut survenir quelques erreurs...)

Rejoignant le constat réalisé par les auteurs précités, on peut repérer qu'aux 2 extrémités du gradient d'altération, l'écart standard d'EQR est plus faible mais que la zone subissant des altérations de modérées à assez fortes est plus critique sur le plan des écarts d'évaluation. Dans le cas de notre jeu de données d'étude, elle est bimodale. C'est peut-être une réalité correspondant à notre contexte franco-Français. Cela pourrait aussi être dû à quelques problèmes de répétabilité intégrés dans le jeu de données d'étude, mais non prouvés (dans ce cas, conviendrait-il lisser un peu cette relation pour faire disparaître ce double-mode ???).

Cependant, au final, la variabilité des résultats inter-opérateurs de l'IBD2007 et par conséquent, l'incertitude d'évaluation d'état, se répercute plus vers le milieu du gradient d'altération couvert par le dispositif de stations qu'à ses extrémités, et ce constat rejoint celui réalisé par les 2 auteurs précédents.

Du côté des situations en zones de référence ou pseudo-naturelles (partie droite de la courbe), on voit bien que l'évaluation est plus stable que lorsque les sites subissent des altérations anthropiques. Dans notre cas, vu qu'il n'y a que peu de variation temporelle entre les échantillonnages, l'accroissement des variabilités dans la zone de manifestation des effets de pollutions pourrait être interprété comme le signe d'une difficulté plus importante à obtenir un échantillonnage représentatif de la station en fonction de petits écarts de localisation, de faciès de prélèvement, de pratique des intervenants...). En tout cas, les peuplements y sont plus variables... Les solutions opérationnelles auxquelles on pourrait penser pour remédier à cet état de fait viseraient à mieux spécifier les signalétiques de stations, afin que des acteurs successifs interviennent bien sur la même localisation, et à ré-augmenter l'assise d'un échantillonnage intégré de station, qui a été tiré vers le minimum nécessaire lors du processus de compromis associé à la rédaction de la norme NF-T-90-354 (le minimum d'échantillonnage de 5 galets par station, un peu limite pour bien représenter une station variable, étant adopté comme la règle en même temps que la solution la plus rapide en temps passé, notamment pour les intervenants économiques).

Du côté des plus fortes altérations, si une décroissance des variabilités semble aussi ressortir sur la courbe des écarts standards, il n'est pas possible d'en tirer des conclusions solides, du fait de la diminution de l'occurrence de sites aussi pollués dans le contexte Français actuel (les points noirs à l'aval d'agglomérations, de STEPs, de stations de traitements industriels ou agro-alimentaires, qui étaient les endroits les plus sévèrement pollués par les nutriments et la matière organique de tout notre réseau hydrographique, ont été les premiers à faire l'objet d'un effort intensif des industriels de traitement de l'eau, en fonction des fortes incitations à la fois réglementaires et financières émanant des Agences de Bassins).

Concernant notre échantillon d'étude, seulement un site est à l'origine de cette particularité graphique. Il s'agit de la station « Retit Rosne » à Garges-les –Gonesse. L'explication plausible vis-à-vis de ce cas est qu'outre l'homogénéité du faciès physique et des conditions hydrodynamiques (écoulement lisse, peu profond et peu turbulent sur un fond de canal en béton parsemé de substrats de type morceaux de ciment et d'agglomérés, morceaux de tuiles...), on se trouve ici sous l'influence de pollutions péri-urbaines et de STEP, et pas de transferts diffus d'origine agricole. L'altération est ici de niveau intense et assez constante dans le temps, et les flores y sont probablement plus sélectionnées et plus adaptées à ce niveau de forte pollution... :

4.3.3. Loi de probabilité et calcul de la confiance des classes

En utilisant les données des EQR observés par chaque opérateur à chaque site (Figure 40) et en répercutant l'erreur-standard associée à l'évaluation de chaque site en fonction des écarts inter-opérateurs observés, les courbes associées à la loi de probabilité précédemment décrite (cf. Figure 36 page 92) ont été représentées pour chaque classe d'état écologique. On obtient donc, tel que présenté sur la Figure 40 ci-dessous, les lois de probabilité d'appartenance à telle ou telle classe de qualité en fonction des EQR observés.

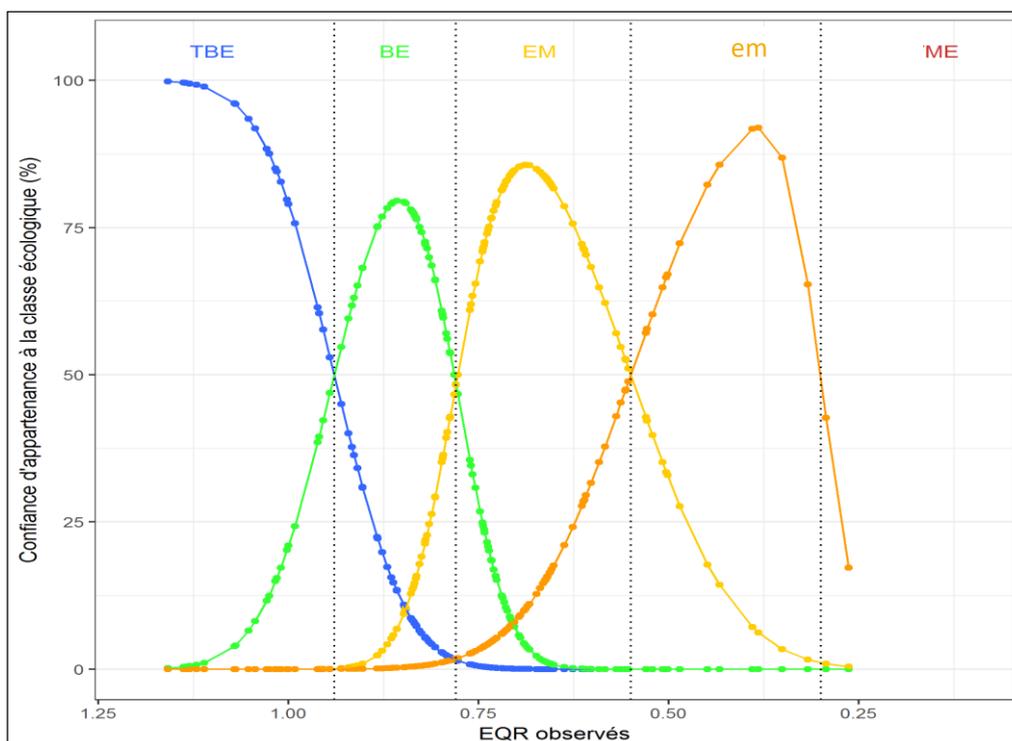


Figure 40 : Loi de probabilité des EQR dans le cadre de la variabilité inter-opérateurs

Par exemple, pour un EQR de 0.94 (limite de classe TBE-BE), on obtient une confiance de 50% que le site soit classé comme « TBE » et de 50% en classement BE. A la limite de classe BE-EM (0,76), il y a des chances égales de chances que le site soit en BE comme en EM (48,5%), mais aussi 3% de chances d’être en état médiocre (em). Cette probabilité d’appartenance à plusieurs classes de qualité peut ainsi se décliner tout au long du référentiel de valeurs en EQR. Faute d’un référentiel de valeurs observées suffisant, il n’a pas été possible de modéliser et de tracer graphiquement une courbe de probabilité précise d’appartenance à la classe de Mauvais Etat (classe rouge).

Toutefois, hors domaine représenté, soit pour un EQR <0,25, la probabilité d’appartenance à la classe de mauvais état se calculera par défaut comme :

$$p(\text{ME}) = 1 - [p(\text{em}) + p(\text{EM})]$$

4.4.4.Exemples illustratifs

Pour illustrer le type de résultats procuré par ce modèle sur des cas réels, trois sites ont été choisis à titre d’exemples pour appliquer la loi de probabilité d’appartenance à une classe d’état écologique à la note EQR attribuée pour chaque site (Figure 41 ci-dessous).

En fonction des valeurs d’EQR-IBD obtenues, les 2 sites situés le plus à gauche (1054100) et le plus à droite (4049800) de la figure ont une probabilité d’appartenance substantiellement plus grande à une classe d’état diatomique qu’à une autre, respectivement classe de Bon Etat (en vert) et classe d’Etat Moyen (en jaune). La probabilité que le site soit dans la classe à probabilité dominante est forte, même s’il existe les possibilités limitées indiquées que ce site appartienne en fait à une autre classe de qualité contigüe à la classe la plus probable.

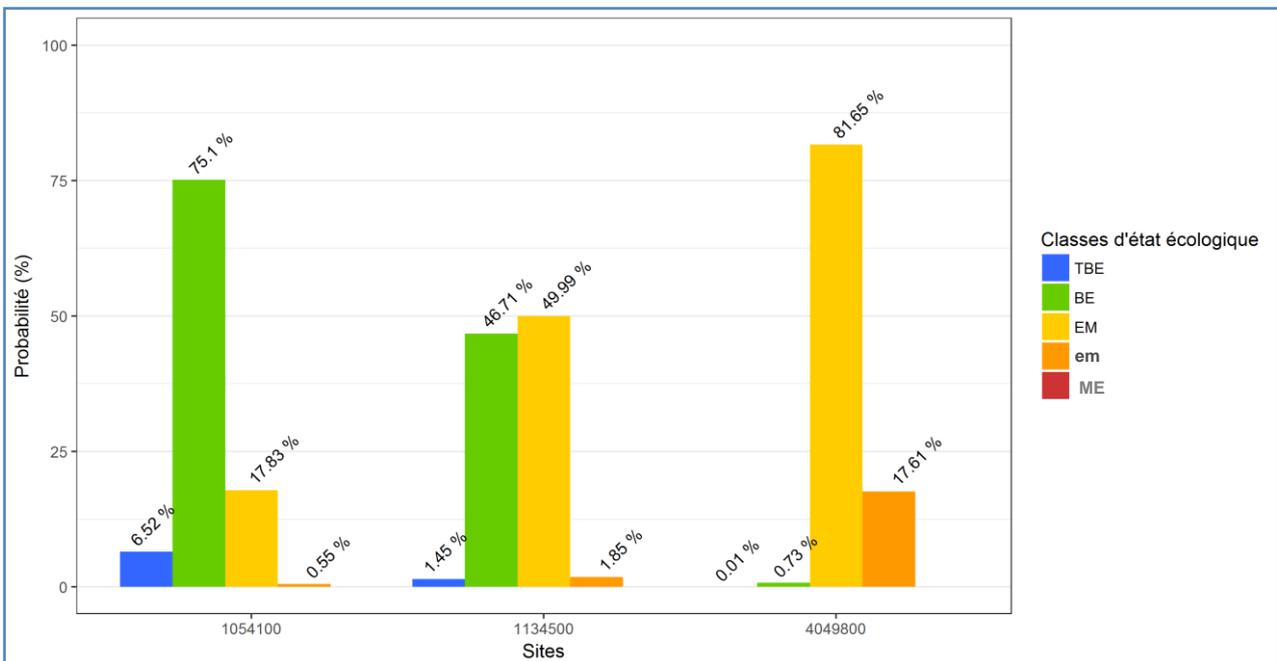


Figure 41 : Probabilité d’appartenance à une classe écologique pour 3 sites différents

Lorsque la note en EQR se situe proche de la limite entre deux classes d'état (site 1134500 dans notre exemple), le risque de se tromper dans l'évaluation, en particulier sur la base d'un seul relevé annuel, devient sensiblement plus importante.

4.4 Propagation de la méthodologie à l'évaluation trisannuelle

La propagation de cette méthodologie pour passer de l'état diatomique annuel (et son intervalle de confiance) à l'état diatomique trisannuel (et son intervalle de confiance) ne pose aucun problème particulier.

L'Arrêté Evaluation en vigueur (actuellement, celui du 27 Juillet 2015) prévoit, pour le maillon diatomique des cours d'eau, une évaluation de l'état écologique moyennée sur 3 ans (soit un demi-plan de gestion). Ce calcul de moyenne vise à donner une vision plus stable et intégrée de l'état écologique du site, sachant que la méthode d'évaluation biologique en cause peut subir une **variation interannuelle** dont une partie peut être occasionnée par la simple variabilité naturelle du contexte hydro-climatique (**fluctuation naturelle interannuelle**), et que l'autre partie, que l'on cherche à évaluer, est dûe à la **variabilité d'impact des altérations anthropiques** en fonction du **contexte spécifique de pollution de l'année**, directement influencé par :

- les usages anthropiques et leurs conséquences sur la qualité du milieu sur l'épisode temporel représenté par l'échantillonnage biologique,
- les conditions pluvio-climatiques conditionnant notamment les transferts vers les milieux aquatiques, leur dilution et leurs effets polluants).

Pour l'évaluation moyenne d'état écologique trisannuel, il suffit de procéder :

- au calcul d'état diatomique moyen (moyenne des EQR au site obtenus par les relevés des années 1, 2 et 3),
- à la probabilité d'affectation par classe de qualité de l'EQR moyen obtenu sur la base des 3 résultats annuels.

5. CONCLUSIONS

Le présent programme d'évaluation de l'incertitude de l'IBD₂₀₀₇ a comporté 3 volets d'étude différents :

- Un premier volet (2013-2015) qui a déjà fait l'objet d'un rapport d'étape fin 2015, était centré sur l'étude de l'**incertitude globale inter-opérateurs**, à partir de 2 campagnes de prélèvements de terrain réalisées en 2013 ou 2014 sur 61 sites d'étude prospectés, dans une fenêtre temporelle courte (moins de 2 semaines), par 3 types d'opérateurs différents.
- Le 2^{ème} volet, réalisé sur la base de données historiques collectées dans Pandore, était destiné à cerner la **variabilité interannuelle de l'IBD₂₀₀₇** et en particulier, sa **composante naturelle** (vu que c'est elle qui impacte le résultat de la méthode biologique d'une fluctuation indésirable, que l'on va chercher à contrôler par la notion d'intervalle de confiance).
- le 3^{ème} volet, lui aussi réalisé sur 2016-2017, visait à la propagation de l'incertitude globale de la méthode d'évaluation IBD₂₀₀₇ sur l'évaluation d'état diatomique d'un site donné de réseau de surveillance de type RCS, selon le cadre d'application édicté par l'Arrêté Evaluation du 27-07-2015 (évaluation annuelle, évaluation trisannuelle moyennée).

Les 2 premiers volets d'étude donnent au final des résultats assez proches :

- **Dans le 1^{er} volet « Incertitude inter-opérateur »**, en incluant quelques variabilités constatées allant assez largement au-delà d'un effet inter-opérateur classique, mais dont les hypothèses explicatives n'ont pas pu être éclaircies de façon certaine, **l'intervalle de confiance de la méthode d'évaluation diatomique s'établit à +/- 0,55 point d'IBD₂₀₀₇** par rapport à la moyenne inter-observateur établie sur un même site (cf. Figure 16), dans un contexte de variabilité temporelle maîtrisée (prélèvements faits dans un créneau temporel le plus court possible, seuillé à un maximum de 2 semaines).
- Dans le 2^{ème} volet **« Variabilité interannuelle de l'IBD₂₀₀₇ »**, qui visait surtout à cerner les fluctuations indésirables du résultat de la méthode liées à la variabilité interannuelle des conditions naturelles (= « bruit de fond » de la méthode), l'analyse de données historiques effectuée à partir de requêtes faites dans Pandore n'a malheureusement pas pu porter uniquement sur des sites de référence présentant un nombre suffisant de réplicats interannuels pour atteindre le seuil nécessaire (seuillage initial visé, sur une chronique temporelle liée à une pratique méthodologique homogène : 5 résultats annuels d'IBD₂₀₀₇, ramenés au final à 4 résultats). L'analyse des variabilités a donc dû être faite sur une sélection de 103 sites naturels ou altérés présentant une relative stabilité chimique apparente en fonction des analyses d'eau disponibles (basées sur des échantillonnages d'eau ponctuels, avec les défauts de représentativité temporelle qui y sont liés...). Ce jeu de données finalement utilisé comporte : **1) une part de variabilité incontrôlée d'origine naturelle** dans le résultat de la méthode biologique, qu'il est nécessaire d'isoler et de maîtriser par un **intervalle de confiance**, et **2) une part de variabilité liée à l'expression biologique d'impacts de pollutions**, que la méthode est au contraire destinée à évaluer de façon précise et sensible.

Les sites subissant manifestement des altérations anthropiques non-négligeables sont aussi ceux où se manifeste la variabilité inter-annuelle la plus forte du résultat d'IBD, les impacts biologiques pouvant se révéler fortement variables en fonction de l'intensité des usages

humains accompagnant l'épisode temporel considéré mais aussi, en fonction du contexte hydro-climatique de l'année, plus ou moins favorable aux transferts de polluants et à leur degré de dilution (qui conditionne aussi leur concentration instantanée ou intégrée dans la colonne d'eau). Outre une variabilité de résultat d'IBD pouvant être reliée à la **variabilité temporelle des conditions naturelles**, en général d'ampleur modeste, de tels sites sous l'effet d'altérations anthropiques avérées subissent surtout une forte variabilité des résultats de bio-indication diatomique sous l'influence dominante, voire très largement dominante, de la **variabilité interannuelle de manifestation biologique des effets d'altérations anthropiques**, qui peut atteindre 5 ou 6 points d'écart à la moyenne au même site (cf. Figure 31) et, sous une autre forme d'expression, jusqu'à 9,5 points d'écart par rapport au relevé le moins impacté de la même série temporelle acquise au même site (cf. partie gauche de la Figure 28). Dans ces conditions, afin d'isoler les **variabilités sous influence naturelle dominante**, qui constituent la part indésirable de fluctuation des résultats qu'il convient de maîtriser par un **intervalle de confiance assorti à la méthode**, les sites sous variabilité de résultats diatomiques dominés par la **variabilité d'effet des altérations anthropiques** ont été écartés de l'étude de la variabilité naturelle, sur une base fréquentielle plutôt sous-évaluée (percentile 50 des plus gros écarts, alors que les sites sous altération anthropique avérée représentent certainement une part encore plus importante des sites et des relevés). En procédant de cette façon, dans le contexte où interviennent à la fois des variabilités naturelles inter-annuelles et aussi, de fait, une part mal cernée de variabilité liée à un effet inter-opérateur (au gré des attributions de marchés annuels et inter-annuels), **l'estimation de la variabilité interannuelle d'origine principalement naturelle a ainsi pu être estimée à +/- 0,65 points d'IBD.**

Les estimations d'intervalle de confiance de l'IBD 2007 qui ressortent de ces 2 volets d'étude, obtenus sans sélection excessivement drastique de données, donc incorporant une certaine marge de sécurité, procurent des résultats plutôt proches qui d'une part, se confortent mutuellement [$\pm 0,55$ vs $\pm 0,65$ points d'IBD], et d'autre part se rapprochent de celles obtenues lors de précédentes études (e.g. Roussel et al, 2011), ainsi que de ce que les praticiens expérimentés avaient en tête (variation la plus habituelle de l'ordre de [$\pm 0,5$ point d'IBD], mais pouvant parfois atteindre [± 1 point d'IBD], selon particularités des situations locales).

Dans le but de satisfaire la requête de la DCE, qui demande d'assortir les résultats des méthodes participant à l'évaluation d'Etat Ecologique du calcul d'un intervalle de confiance, le **troisième volet d'étude** a visé à élaborer une méthodologie permettant de **répercuter l'incertitude inter-opérateur globale** de l'IBD 2007, ainsi que **l'intervalle de confiance qui en découle**, sur **l'Evaluation d'Etat Ecologique (volet diatomique)** telle que mise en œuvre dans le cadre des réseaux institutionnels de surveillance de type Réseau de Contrôle et de Surveillance (RCS), Réseau de Référence Pérenne (RRP)...

Dans cette optique, nous nous sommes appuyés sur des éléments de démarches méthodologiques intéressantes proposés par *Cleveland et al*, 2008 et *Kelly et al*, 2009, et avons conduit une réflexion spécifique sur le cas à traiter dans notre contexte national.

Ainsi, le jeu de données d'étude produit dans le cadre du Volet 1 « Incertitude inter-opérateur » a tout d'abord été utilisé pour mettre au point un **modèle de calcul d'un intervalle de confiance**, en s'appuyant sur la distribution des écarts inter-opérateurs réellement observés tout au long du gradient de qualité de notre jeu de données d'étude.

Puis, comme le requiert la DCE, nous avons pu mettre en place et utiliser les **lois de probabilité d'appartenance du site à telle ou telle classe de qualité** en fonction de la valeur d'EQR obtenue.

L'état calculé sur la base de chaque relevé diatomique annuel peut ainsi se décliner :

- en probabilité de classement du site concerné dans les classes d'état pour chaque relevé diatomique annuel collecté (**détermination d'état diatomique annuel au relevé**),
- mais aussi, afin de s'adapter à la pratique de la détermination d'état moyen trisannuel telle qu'entrée en vigueur *via* l'Arrêté national Evaluation du 27-07-2015, dans l'optique du rapportage Européen d'Etat Ecologique moyen effectué par demi-Plan de Gestion, à la **probabilité de classement moyen trisannuel d'état diatomique d'un site donné**. Cette dernière étape ne comporte pas de difficulté méthodologique particulière, l'état moyen étant la moyenne des états annuels obtenus au site sur le demi-Plan de Gestion, exprimés en EQR, et la probabilité moyenne de classement du site dans chaque classe d'état étant calculée sur la base de l'EQR moyen trisannuel obtenu.

Suite à ce rapport final, le dispositif tel que proposé pourrait rapidement passer dans la pratique du rapportage européen d'état, pour peu que le Ministère en charge de l'Environnement/ Direction de l'Eau souhaite l'officialiser (via un avenant de l'Arrêté Evaluation) et le voir entrer en application.

Dans un tel cas, 2 points mériteraient probablement une discussion scientifico-technique complémentaire avec les tutelles et l'AFB, ainsi qu'une validation collective, préalablement à la mise en application :

- le niveau de la marge de sécurité introduite dans l'estimation de l'intervalle de confiance qui, dans l'état, est probablement un peu surestimé du fait du très faible toilettage de données douteuses (n'ont été enlevés qu'un faible nombre de relevés entachés d'un problème certain ayant entraîné un défaut de reproductibilité). Cette façon de faire élargit un peu cet intervalle de confiance mais en contrepartie, procure une robustesse accrue à l'évaluation d'état.
- l'éventuelle intervention *a minima* d'une fonction de lissage de la fonction polynomiale représentant la relation entre l'écart-type et les moyennes des EQR des sites (cf. Figure 39), dans le seul but de résorber son caractère bimodal (caractéristique pas forcément attendue ni scientifiquement explicable sur la base d'un argumentaire logique...).

Références

- REF. 1 :** DCE, Directive Cadre sur l'Eau du 23 Octobre 2000 (JOCE L 327 du 20 /12/2000). Directive 2000/60/EC du Parlement Européen et du Conseil établissant le cadre de l'action Communautaire dans le domaine de l'eau. http://ec.europa.eu/environment/water/water-framework/index_en.html.
- REF. 2 :** Arrêté du 27 juillet 2015 modifiant l'arrêté du 25 janvier 2010 relatif aux méthodes et critères d'évaluation de l'état écologique, de l'état chimique et du potentiel écologique des eaux de surface pris en application des articles R. 212-10, R. 212-11 et R. 212-18 du code de l'environnement. Journal Officiel de la République Française n°0198 du 28 août 2015.
- REF. 3 :** MONDY, C.P., VILLENEUVE, B., ARCHAIMBAULT, V., USSEGLIO POLATERA, P. - 2012. A new macroinvertebrate-based multimetric index (I2M2) to evaluate ecological quality of French wadeable streams fulfilling the WFD demands: A taxonomical and trait approach. *Ecological Indicators*, vol. 18, p. 452-467
- REF. 4 :** MARZIN, A., DELAIGUE, O., LOGEZ, M., BELLIARD, J., PONT, D. - 2014. Uncertainty associated with river health assessment in a varying environment: the case of a predictive fish-based index in France. *Ecological Indicators*, vol. 43, p. 195-204
- REF. 5 :** REYJOL, Y., CHAUVIN, C. - 2012. Les interactions scientifiques-gestionnaires dans le domaine de l'écologie aquatique : exemple de la mise en œuvre de la DCE. Conférence Invitée. JILO - Journées Internationales de Limnologie 21/05/2014-23/05/2014, Marseille, FRA. 45 p.
- REF. 6 :** ROUSSEL, F., TISON ROSEBERY, J., COSTE, M. - 2010. Etude de la variabilité inter et intra-opérateur de la note d'Indice Biologique Diatomée 2007 : Approche statistique des résultats d'exercices d'intercomparaison. Rapport final d'étude. Irstea Eds. 24 p.
- REF. 7 :** Kruskal, W.H., and Wallis, W.A. (1952). Use of Ranks in One-Criterion Variance Analysis. *J. Am. Stat. Assoc.* 47, 583–621.
- REF. 8 :** Ellis, J., Andriaenssens, V. (2006). Uncertainty estimation for monitoring results by the WFD biological classification tools. WFD Report (Environment Agency, Rio House, Waterside Drive, Aztec West, Almondsbury, Bristol, BS32 4UD: Environment Agency).
- REF. 9 :** Kelly, M., Bennion, H., Burgess, A., Ellis, J., Juggins, S., Guthrie, R., Jamieson, J., Adriaenssens, V., and Yallop, M. (2009). Uncertainty in ecological status assessments of lakes and rivers using diatoms. *Hydrobiologia* 633, 5–15.
- REF. 10 :** Cleveland, W.S., and Devlin, S.J. (1988). Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *J. Am. Stat. Assoc.* 83, 596–610.