



*École Doctorale des Sciences de la Terre
et Environnement*

LIVE

THÈSE

présentée par :

Juliane Wiederkehr

Soutenu le :

23 janvier 2015

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline : Géochimie

**Estimation des incertitudes associées aux indices
macroinvertébrés et macrophytes pour l'évaluation de l'état
écologique des cours d'eau**

Rapporteurs

*Usseglio-Polatera Philippe
Haury Jacques*

Professeur, Université de Lorraine
Professeur, Agrocampus Ouest, Rennes

Examinatrice

Berti-Equille Laure

Directrice de recherche, IRD

Directrices de thèse

*Le Ber Florence
Trémolières Michèle*

ICPF, Université de Strasbourg/ENGEES
Professeure, Université de Strasbourg

*« La plus grande faiblesse de la pensée
contemporaine me paraît résider dans la
surestimation extraordinaire du connu
par rapport à ce qui reste à connaître. »*

André Breton, 1937

INTRODUCTION	15
PARTIE 1 Etat de l'art et stratégies de recherche	19
CHAPITRE 1 Outils de biosurveillance	20
Titre 1 : Contexte réglementaire	21
Titre 2 : La bioindication	23
Titre 3 : Historique des indices biologiques	24
Section 1 : Les indices basés sur les macroinvertébrés	25
§1. En France	25
1. Les premiers indices	25
2. L'Indice Biologique Global Normalisé (IBGN)	26
3. L'Indice Biologique Global DCE (IBG DCE)	27
4. Le nouvel Indice Invertébrés MultiMétrique : l'I2M2	29
§2. En Europe	31
1. Les indices européens	32
2. Les modèles prédictifs	34
Section 2 : Les indices macrophytiques	35
§1. Le Rang Trophique Moyen (MTR)	36
§2. Le Trophie-Index Macrophyten (TIM)	37
§3. L'indice Biologique Macrophytique en Rivières (IBMR)	37
Section 3 : Les nouveaux questionnements engendrés par la DCE	40
§1. Le projet AQEM : un système d'évaluation de la qualité écologique des rivières européennes	41
§2. Le projet STAR : Normalisation de la classification des rivières	43
§3. Le projet Wiser : Integrative Systems to assess Ecological status and Recovery	44
CHAPITRE 2 Incertitudes associées aux deux indices biologiques	46
Titre 1 : Qu'est-ce que l'incertitude	47
Titre 2 : Les sources d'incertitudes	49
Section 1 : Les sources d'incertitudes des indices basés sur les macroinvertébrés	49
§1. L'impact des facteurs environnementaux	49
1. La mosaïque d'habitats	50
2. Le choix de la placette et de la méthode d'échantillonnage	50
2.1. Le choix de la placette d'échantillonnage	50
2.2. Le choix de la méthode d'échantillonnage	51
2.3. Le nombre d'échantillons ou le volume prélevé	52
§2. L'impact de la phase laboratoire	53
1. L'effet opérateur	54
2. L'effet du sous échantillonnage	55
§3. D'autres sources incertitudes	56
1. Le tri sur le vivant	56
2. Le niveau d'identification des taxons	56
3. La prise en compte des taxons rares ou peu abondants	57
§4. Les incertitudes peu ou non traitées	58
Section 2 : Les sources d'incertitudes des indices macrophytiques	58
§1. L'impact des incertitudes associées aux facteurs environnementaux	59
§2. L'impact de l'opérateur	60
§3. Les incertitudes peu ou non traitées	61
CHAPITRE 3 Stratégies de recherche	62

Titre 1 : Le choix des groupes macroinvertébrés et macrophytes et indices associés _____	63
Titre 2 : Les axes de recherches et les outils développés _____	64
Section 1 : Les macroinvertébrés _____	64
§1. La variabilité intra-substrat _____	66
§2. L'effet opérateur cumulé à la variabilité intra-substrat _____	66
§3. L'effet opérateur et des pratiques laboratoires _____	67
Section 2 : Les macrophytes _____	67
<i>PARTIE 2 Méthodes et contributions</i> _____	69
<i>CHAPITRE 1 Méthodes expérimentales de mesure des incertitudes associées aux macroinvertébrés</i> _____	70
Titre 1 : Article: Experimental study of the uncertainty of the intrasubstrate variability on two French index metrics based on macroinvertebrates : Résumé _____	71
Section 1: Introduction _____	72
Section 2: Materials and methods _____	73
§1. Study area _____	73
§2. Site selection design _____	74
§3. Sampling design _____	75
§4. Sorting phase _____	76
§5. Data analysis _____	76
1. Metrics calculation _____	76
2. Distance analysis _____	78
3. Cluster analysis _____	78
4. Similarity indices _____	79
5. Multidimensional scaling _____	80
Section 3: Results _____	81
§1. Substrate variability and metrics robustness _____	81
§2. Communities similarity according to substrate _____	87
§3. Site effect _____	89
Section 4: Discussion _____	90
Section 4: Conclusion _____	94
Titre 2 : Etude des incertitudes associées aux indices macroinvertébrés liées aux variabilités naturelle et opérateur sur la base des essais inter-laboratoires menés par AGLAE _____	95
Section 1 : Introduction _____	96
Section 2 : Présentation de l'essai inter-laboratoire en hydrobiologie _____	96
Section 3 : Mesure de la variabilité opérateur en laboratoire _____	98
§1. Variabilité des métriques de l'I2M2 _____	99
§2. Variabilité des listes faunistiques _____	104
Section 4 : Discussion & Conclusion _____	109
<i>CHAPITRE 2 Procédure de simulation de mesure des incertitudes associées aux macroinvertébrés</i> _____	113
Titre 1 : Introduction _____	114
Titre 2 : Matériel et méthodes _____	115
Section 1. Métadonnées _____	115
Section 2. L'algorithme _____	117
Section 3 : Les expérimentations _____	120
Section 4. Analyses des données _____	121

§1. Calcul des métriques de l'IBG DCE et de l'I2M2	121
§2. Analyse des distances	123
§3. Calcul des Z-scores	124
Titre 3 : Résultats	125
Section 1. Robustesse de l'IBG DCE	125
Section 2. Robustesse de l'I2M2	129
Section 3. Performances de l'opérateur	137
Titre 4 : Discussion	140
Titre 5 : Conclusion	147
<i>CHAPITRE 3 Méthode virtuelle de mesure des incertitudes liées à l'opérateur associées aux macrophytes</i>	149
Titre 1 : Experimental Study of uncertainty in the Macrophyte index (IBMR) based on species identification and cover (Wiederkehr, et al., 2015) : Résumé	150
Titre 2 : Introduction	151
Titre 3 : Materials and methods	152
Section 1: The IBMR formula (AFNOR, 2003)	153
Section 2: Metadata	153
Section 3. Algorithm	155
§1. Taxa modification (TAX)	156
§2. Abundance modification (AC)	157
§3. Taxa and cover class modification (TAX+AC)	158
Section 4. Experiments	158
Section 5. Data analysis	159
§1. Distance analysis	160
§2. Z-Score calculation	161
§3. JUMI	161
Titre 4 : Results	162
Section 1: IBMR robustness	162
Section 2 : Class changes	165
Section 3. Taxa weight	166
Titre 6 : Discussion	168
Titre 7 : Conclusion	170
<i>PARTIE 3 Discussion générale</i>	172
Titre 1 : De la mesure de la variabilité à l'incertitude	173
Section 1 : Les macroinvertébrés	173
Section 2 : Les macrophytes	180
Titre 2 : Critiques et pistes d'améliorations des méthodes d'analyses utilisées	183
Section 1 : Approche expérimentale	183
Section 2 : Approches virtuelles	185
§1. Approche virtuelle d'étude de la variabilité intra-substrat et opérateur au sein des protocoles d'évaluation basés sur les macroinvertébrés	185
§2. Approche virtuelle d'étude de l'effet opérateur au sein de l'indice IBMR	187
Titre 3 : Extension des outils proposés aux autres indices biologiques	188

Titre 4 : Combinaison des incertitudes associées aux indices _____	191
CONCLUSION _____	195
REFERENCES _____	198

Figure 1: Facteurs de variabilité et différentes étapes de mise en œuvre des protocoles d'évaluation des milieux aquatiques basés sur les macroinvertébrés	65
Figure 2: Facteurs de variabilité et différentes étapes de mise en œuvre des protocoles d'évaluation des milieux aquatiques basés sur les macrophytes	68
Figure 3: Location of the 31 sites in western and southwestern France, chosen for the experimentation (each point corresponds to a studied site and numbers correspond to hydroecoregions (as defined by Wasson, et al., 2004)	74
Figure 4: Sampling design used for the experimentation	76
Figure 5 : Average variation coefficient (CV) for IBGN metrics (GI, VT, and VC) and I2M2 metrics (ASPT, TAX, H, PS, and OS) for each substrate. CV < 16% = not significant, 16% < CV < 33.3% = significant, and CV > 33.3% = critical error. (Average of CV calculated for all sites for a given substrate)	84
Figure 6: Cluster dendrograms based on Euclidean distance and associated heatmaps on faunistic lists (A) and I2M2 metrics (B) calculated from replicate samples of bryophytes at the three sampled sites in Massif Central Nord (heatmap : black very close, white completely different)	85
Figure 7: Cluster dendrograms based on Euclidean distance and associated heatmaps on faunistic lists (A) and I2M2 metrics (B) calculated from replicate samples of submerged spermaphytes at the five sampled sites in Massif Central Nord, Causses Aquitains and Massif Armoricaïn (heatmap : black very close, white completely different)	86
Figure 8: Cluster dendrograms based on Euclidean distance and associated heatmaps on faunistic lists (A) and I2M2 metrics (B) calculated from replicate samples of gravels at the four sampled sites in Massif Central Nord, and Massif Armoricaïn (heatmap : black very close, white completely different)	87
Figure 9: Average of three similarity indices (Jaccard index, Sørensen index, and Renkonen index) for each sampled substrate at all sites (in %)(Example: JS average of blocks = (blocks average JS site 1 + blocks average JS site 2 + blocks average JS site 3/ 3))	88
Figure 10: Multi-Dimensional Scaling based on the Sørensen distance matrix on taxonomic lists of gravel substrates for all sampled sites represented in a 2D space. Each point corresponds to a gravel replicate.	90
Figure 11: Multi-Dimensional Scaling based on the Sørensen distance matrix on taxonomic lists of roots for all sampled sites represented in a 2D space. Each point corresponds to a roots replicate.	90
Figure 12: Comparison between intrasubstrate variability and substrates liveability. A: Variability gradient of substrates depending on the Sørensen similarity index, B: Substrates liveability depending on the last French sampling grid (AFNOR 2009) (in boxes mineral substrates)	92
Figure 13: Variability gradient of metrics based on CV (in dotted lines: metrics of biological traits, large dotted lines: Diversity metrics, in boxes: richness taxonomic metrics; in bold: I2M2 metrics, in italic: IBGN metrics) CV <16% = not significant, 16% < CV < 33.3% = significant, and CV >33.3% = critical error.	94
Figure 14: Classification hiérarchique basée sur la distance euclidienne et carte de chaleur associée obtenues sur les métriques de l'I2M2 calculées pour le substrat pierres sur la Vienne (carte de chaleur : rouge = proche, blanc = complètement différent ; le premier numéro = code laboratoire, 16 laboratoires, et le deuxième = code répétition, 4 répétitions de substrats, exemple : laboratoire 10 – répétition 3)	101
Figure 15: Classification hiérarchique basée sur la distance euclidienne et carte de chaleur associée obtenues sur les métriques de l'I2M2 calculées pour le substrat racines sur la Vienne (carte de chaleur : rouge = proche, blanc = complètement différent ; le premier numéro = code laboratoire, 16 laboratoires, et le deuxième = code répétition, 4 répétitions de substrats, exemple : laboratoire 10 – répétition 3)	102
Figure 16: Classification hiérarchique basée sur la distance euclidienne et carte de chaleur associée obtenues sur les métriques de l'I2M2 calculées sur le substrat spermaphytes immergés sur la Vienne (carte de chaleur : rouge = proche, blanc = complètement différent ; le premier numéro = code laboratoire, 16 laboratoires, et le deuxième = code répétition, 4 répétitions de substrats, exemple : laboratoire 10 – répétition 3)	103
Figure 17: Classification hiérarchique basée sur la distance euclidienne et carte de chaleur associée obtenues sur les listes faunistiques des substrats pierres sur la Vienne (carte de chaleur : rouge = proche, blanc = complètement différent ; le premier numéro = code laboratoire, 16 laboratoires, et le deuxième = code répétition, 4 répétitions de substrats, exemple : laboratoire 10 – répétition 3)	105
Figure 18: Classification hiérarchique basée sur la distance euclidienne et carte de chaleur associée obtenues sur les listes faunistiques des substrats racines sur la Vienne (carte de chaleur : rouge = proche, blanc = complètement	

différent ; le premier numéro = code laboratoire, 16 laboratoires, et le deuxième = code répétition, 4 répétitions de substrats, exemple : laboratoire 10 – répétition 3) _____ 106

Figure 19: Classification hiérarchique basée sur la distance euclidienne et carte de chaleur associée obtenues sur les listes faunistiques des substrats spermaphytes immergés sur la Vienne (carte de chaleur : rouge = proche, blanc = complètement différent ; le premier numéro = code laboratoire, 16 laboratoires, et le deuxième = code répétition, 4 répétitions de substrats, exemple : laboratoire 10 – répétition 3) _____ 107

Figure 20: Positionnement multi dimensionnel en 2D basé sur la matrice des distances de Sørensen sur les listes faunistiques des substrats pierres sur la Vienne. Chaque point correspond une liste faunistique et un laboratoire. _____ 108

Figure 21: Positionnement multi dimensionnel en 2D basé sur la matrice des distances de Sørensen sur les listes faunistiques des substrats racines sur la Vienne. Chaque point correspond une liste faunistique et un laboratoire. _____ 108

Figure 22: Positionnement multi dimensionnel en 2D basé sur la matrice des distances de Sørensen sur les listes faunistiques des substrats spermaphytes immergés sur la Vienne. Chaque point correspond une liste faunistique et un laboratoire. _____ 109

Figure 23 : Gradient de variabilité des métriques basé sur les coefficients de variation moyens obtenus pour chaque métriques de l'I2M2 : 1) expérimentation intra-substrat et 2) expérimentation AGLAE. CV <16% = pas significatif, 16% < CV < 33.3% = significatif, and CV >33.3% = erreur critique. _____ 111

Figure 24: Localisation des 25 sites sélectionnés en France _____ 120

Figure 25: Proportion de déclassements de l'IBG DCE pour chaque classe de qualité écologique (la couleur du point correspond à la classe de qualité originale : très bonne - bleu, bonne - vert, moyenne -jaune, médiocre - orange et mauvaise - rouge) après l'exclusion des taxons rares ou peu abondants (R) pour chaque probabilité d'erreur (en %) (Nombre de sites observés n=25 où n=5/ \mathcal{P}_R , 100 tests/ \mathcal{P}_R). Les seuils de classes de qualité sont définis par le Ministère de l'Ecologie (Ministère de l'Ecologie, du Développement durable et de l'Energie, 2012) _____ 126

Figure 26: Coefficients de variation (CV) moyens des notes IBG DCE et du groupe indicateur (GI), pour chaque classe de qualité écologique et chaque probabilité d'erreur \mathcal{P}_R introduite pour tester l'impact de l'exclusion aléatoire des taxons rares ou peu abondants, obtenus avec l'algorithme présenté précédemment. Les moyennes des CV sont calculées sur les 100 tests effectués par probabilité d'erreur et pour chaque classe de qualité (nombre de sites observés : n=25, n_i=5/ \mathcal{P}_R , 100tests/ \mathcal{P}_R /classe de qualité). CV <16% pas d'impact significatif de l'erreur, 16% < CV < 33,3% erreur significative et CV >33,3% erreur critique. _____ 127

Figure 27: Coefficients de variation (CV) moyens des richesses taxonomiques (VT) et des classes de variété (VC), pour chaque classe de qualité écologique et chaque probabilité d'erreur \mathcal{P}_R introduite pour tester l'impact de l'exclusion aléatoire des taxons rares ou peu abondants obtenus avec l'algorithme présenté précédemment. Les moyennes des CV sont calculées sur les 100 tests par probabilité d'erreur pour chaque classe de qualité (nombre de sites observés : n=25, n_i=5/ \mathcal{P}_R , 100tests/ \mathcal{P}_R /classe de qualité). CV <16% pas d'impact significatif de l'erreur, 16% < CV < 33,3% erreur est significative et CV >33,3% erreur critique. _____ 127

Figure 28 : Variations des notes de l'IBG DCE obtenues pour cinq rivières de classe de qualité différente (les couleurs des symboles correspondent à la classe initiale : très bonne - bleu, bonne - vert, moyenne - jaune, médiocre - orange et mauvaise- rouge ; symboles carrés = score original, symboles ronds = scores possibles obtenus après modifications aléatoires) après l'exclusion aléatoire des taxons rares ou peu abondants (R) obtenus avec l'algorithme présenté précédemment en fonction de chaque probabilité d'erreur \mathcal{P}_R (nombre de rivières concernées n=5, nombre de tests 20/ \mathcal{P}_R /rivière) _____ 128

Figure 29: Variation des scores de l'I2M2 obtenus pour cinq rivières de classe de qualité différente (la couleur des symboles correspondant à la classe initiale : très bonne - bleu, bonne - vert, moyenne - jaune, médiocre - orange et mauvaise- rouge ; symboles carrés = score original, symboles ronds = scores après modifications aléatoires) après l'exclusion aléatoire des taxons rares ou peu abondants (R), la modification aléatoire des estimations des abondances (AB) et des deux modifications cumulées (R+AB) par le biais de l'algorithme présenté précédemment pour chaque probabilité d'erreur (nombre de rivières concernées n=5, nombre de tests 20/ \mathcal{P} /rivière) _____ 130

Figure 30: Coefficients de variation (CV) moyens des cinq métriques de l'I2M2 (ASPT, TAX, H, OS et PS) pour chaque classe de qualité écologique et chaque probabilité d'erreur introduite pour tester l'exclusion aléatoire des taxons

rare ou peu abondants (R), la modification aléatoire des estimations des abondances (AB) et des deux modifications (R+AB) obtenus par l'algorithme utilisé. Les moyennes des CV sont calculées sur les 100 tests par probabilité d'erreur pour chaque classe de qualité ($n=5/\mathcal{P}$, $100\text{tests}/\mathcal{P}/\text{classe de qualité}$). $CV < 16\%$: pas d'impact significatif de l'erreur, $16\% < CV < 33,3\%$: l'erreur est significative et $CV > 33,3\%$: erreur critique. _____ 132

Figure 31 : Coefficients de variation (CV) moyens des EQR des cinq métriques de l'I2M2 (ASPT, TAX, H, OS et PS) pour chaque classe de qualité écologique et chaque probabilité d'erreur introduite pour tester l'impact de l'exclusion aléatoire des taxons rares ou peu abondants (R), de la modification aléatoire des estimations des abondances (AB) et des deux modifications cumulées (R+AB) obtenus avec l'algorithme présente précédemment. Les moyennes des CV sont calculées sur les 100 tests par probabilité d'erreur pour chaque classe de qualité ($n=5/\mathcal{P}$, $100\text{tests}/\mathcal{P}/\text{classe de qualité}$). $CV < 16\%$: pas d'impact significatif de l'erreur, $16\% < CV < 33\%$: l'erreur est significative et $CV > 33\%$: erreur critique. _____ 134

Figure 32 : Coefficients de variation (CV) moyens des scores I2M2 pour chaque classe de qualité écologique et chaque probabilité d'erreur introduite pour tester l'impact de l'exclusion aléatoire des taxons rares ou peu abondants (R), de la modification aléatoire des estimations des abondances (AB) et des deux modifications cumulées (R+AB) obtenus par l'algorithme présenté précédemment. Les moyennes des CV sont calculées sur les 100 tests par probabilité d'erreur pour chaque classe de qualité ($n=5/\mathcal{P}$, $100\text{tests}/\mathcal{P}/\text{classe de qualité}$). $CV < 16\%$: pas d'impact significatif de l'erreur, $16\% < CV < 33,3\%$: l'erreur est significative et $CV > 33,3\%$: erreur critique. _____ 136

Figure 33 : Perte maximale de points pour l'IBG DCE, ses métriques (GI, VC) et la diversité taxonomique pour chaque classe de qualité (très bonne, bonne, moyenne, médiocre et mauvaise) obtenue avec l'algorithme développé précédemment due à l'exclusion des taxons rares ou peu abondants (écarts obtenus par rapport aux valeurs initiales) _____ 143

Figure 34 : Variations maximales obtenues après (A) l'exclusion des taxons rares (R), (B) les erreurs d'estimations des abondances (AB) et (C) l'exclusion des taxons rares et les erreurs d'estimations des abondances (R+AB) pour l'I2M2 et ses EQR calculés par métriques (TAX, ASPT, H, PS et OS) pour chaque classe de qualité (très bonne, bonne, moyenne, médiocre et mauvaise) obtenues avec l'algorithme développé précédemment (écarts obtenus par rapport aux valeurs des EQR initiaux) _____ 144

Figure 35 : Gradient de variabilité des métriques basé sur les coefficients de variation moyens obtenus (CV) pour chaque métriques et classe de qualité (en gras métriques de l'I2M2, en italique métriques de l'IBG DCE). $CV < 16\%$: pas d'impact significatif de l'erreur, $16\% < CV < 33,3\%$: l'erreur est significative et $CV > 33,3\%$: erreur critique. _____ 145

Figure 36 : Localisation of the 25 sites, chosen for our experiments, in France _____ 159

Figure 37 : Total number of downgrade or upgrade on each trophic status (Very low, Low, Medium, High, Very high) obtained with the algorithm defined in Section3: after taxa modification (TAX), abundance classes modification (AC), and taxa and abundance classes modifications (TAX+AC) (for each kind of modification: number of original sites $n=25$, $n=5/\text{error rate}$, number of modified lists 100 tests/site/error rate or 500 tests/error rate)} _____ 163

Figure 38 : Variation of IBMR scores obtained with the algorithm defined in Section 3 for three trophic states (Very low, Medium, Very high): after taxa modification (TAX), abundance classes modification (AC), and after taxa and abundance classes modifications (TAX+AC) (for each kind of modification: number of original lists $n=3$, number of modified lists 20 tests/site/error rate or 60 tests/error rate) _____ 164

Figure 39 : Coefficients de variation (CV) moyens obtenus pour chaque substrat et par regroupement de métriques de l'I2M2 et de l'IBG DCE (Traits biologiques = Fréquence des individus polyvoltins PS + fréquence des individus ovovivipares OS; Variété taxonomique = richesse taxonomique TAX + classe de variété VC, H = indice de Shannon et Polluosensibilité = ASPT + Groupe indicateur GI). $CV < 16\%$ pas d'impact significatif de l'erreur, $16\% < CV < 33,3\%$ erreur est significative et $CV > 33,3\%$ erreur critique. _____ 174

Figure 40 : Comparaison des coefficients de variation des 5 métriques de l'I2M2 (ASPT, Richesse TAX, indice de Shannon H, Fréquence des individus polyvoltins PS et ovovivipares OS) obtenus pour les pierres avec l'expérimentation intra-substrat (IS) et l'expérimentation AGLAE. Pour IS $n=3$ et pour AGLAE $n=16$, $CV < 16\%$ pas d'impact significatif de l'erreur, $16\% < CV < 33,3\%$ erreur est significative et $CV > 33,3\%$ erreur critique. _____ 175

Figure 41 : Gradients de variabilité des métriques des indices macroinvertébrés basés sur les coefficients de variation moyens obtenus en fonction des différentes expérimentations réalisées pour toutes classes de qualité

confondues (en gras : métriques de l'IBG DCE) CV <16% pas d'impact significatif de l'erreur, 16% < CV < 33,3% erreur est significative et CV >33,3% erreur critique. _____ 177

Figure 42 : Propagation des variabilités étudiées au sein des protocoles d'évaluation de l'état écologiques des rivières basés sur les macroinvertébrés, sources d'incertitudes, démarches mises en œuvre, effets et résultats _____ 178

Figure 43: Sources de variabilités au sein des protocoles d'évaluation de l'état écologiques des rivières basés sur les macroinvertébrés, solutions correctives et attente d'amélioration des résultats _____ 179

Figure 44: Propagation des variabilités étudiées au sein du protocole d'évaluation de l'état trophique des rivières basé sur l'IBMR, sources d'incertitudes, méthodes d'analyses et erreurs observées _____ 181

Figure 45: Sources de variabilités au sein des protocoles d'évaluation de l'état écologiques des rivières basé sur l'IBMR, solutions correctives et attente d'amélioration des résultats _____ 182

Tableau 1 : Méthodes européennes d'évaluation de la qualité écologique des rivières _____ 32

Tableau 2 : Moyennes des coefficients de variation (CV) et des distances (DS) pour les métriques de l'I2M2 pour les trois substrats échantillonnés dans le cadre de l'essai inter-laboratoire mené par AGLAE et rappel des coefficients de variation moyens obtenus sur les mêmes substrats par l'expérimentation sur la variabilité intra-substrat (cf. Partie 2 Chapitre 1) (en bleu ciel 16% < CV < 33.3%, en bleu ciel et gras CV > 33.3%) CV <16% = pas significatif, 16% < CV < 33% = significatif, and CV >33% = erreur critique. _____ 99

Tableau 3 : Extrait de la matrice des seuils utilisés par chaque type d'erreur en fonction des substrats _____ 117

Tableau 4 : Exemple de liste faunistique validée \mathcal{L} _____ 118

Tableau 5 : Descriptions des notations utilisées pour présenter le fonctionnement de l'algorithme _____ 118

Tableau 6 : Exemple de liste modifiée basée sur la liste faunistique validée \mathcal{L} _____ 119

Tableau 7 : Z-scores minimums, maximums et moyens obtenus sur les notes de l'I2M2 après les trois différentes modifications (R, AB, R+AB) par le biais de l'algorithme présenté précédemment pour chaque classe de qualité (très bonne, bonne, moyenne, médiocre et mauvaise) (moyenne des valeurs absolues). Valeurs minimale et maximales obtenues sur les cours d'eau par classe d'état écologique, $n=5/\text{taux d'erreur/modification pour } 100 \text{ tests/P}$; moyenne sur 100 tests/P/classe d'état. (Z-score <2 : non significatif, $2 \leq \text{Z-score} \leq 3$: significatif, Z-score > 3 : erreur critique) En gras = valeur significative, en bleu clair = valeur critique. _____ 137

Tableau 8 : Proportion totale des erreurs significatives (SE) et des erreurs critiques (CE) pour l'I2M2 calculée sur la base des Z-scores significatifs et critiques (Z-score <2 : non significatif, $2 \leq \text{Z-score} \leq 3$: significatif, Z-score > 3 : erreur critique) obtenus par l'algorithme présenté précédemment pour chaque classe de qualité (très bonne, bonne, moyenne, médiocre et mauvaise), chaque type de modifications (R, AB et R+AB) et pour chaque probabilité d'erreur. Valeurs obtenues par le cumul des erreurs pour les 5 rivières par classe ($n=5/\text{classe}$, 100 tests/ \mathcal{P}) _ 138

Tableau 9 : Z-scores minimums, maximums et moyens obtenus calculés sur les résultats de l'algorithme présenté précédemment sur les notes de l'IBG DCE et les métriques associées (GI, Richesse taxonomique (VT) et classe de variété VC) après l'exclusion aléatoire des taxons rares ou peu abondants (R) pour chaque classe de qualité (très bonne, bonne, moyenne, médiocre et mauvaise) (moyenne des valeurs absolues). Valeurs minimales et maximales obtenues sur les cours d'eau par classe d'état écologique, $n=5/\mathcal{P}/\text{modification pour } 100 \text{ tests}/\mathcal{P}$; moyenne sur 100 tests/ $\mathcal{P}/\text{classe d'état}$. (Z-score <2 : non significatif, $2 \leq \text{Z-score} \leq 3$: significatif, Z-score > 3 : erreur critique) En gras = valeur significative, en bleu clair = valeur critique. _____ 139

Tableau 10 : Proportion totale des erreurs significatives (SE) et des erreurs critiques (CE) pour l'IBG DCE calculé sur la base des Z-scores (Z-score <2 : non significatif, $2 \leq \text{Z-score} \leq 3$: significatif, Z-score > 3 : erreur critique) par le biais de l'algorithme présenté précédemment pour chaque classe de qualité (très bonne, bonne, moyenne, médiocre et mauvaise), après l'exclusion aléatoire des taxons rares ou peu abondants (R) et pour chaque probabilité d'erreur. Valeur obtenue sur le cumul des erreurs pour les 5 rivières par classe ($n=5/\text{classe}$, 100 tests/ \mathcal{P}) _____ 140

Tableau 11 : Mode de calcul du degré d'altération dans le système d'harmonisation EASY pour une rivière alsacienne _____ 193

Table 1: Sampled substrate list with associated velocity range, location (HER) and the number of sampled sites (values in brackets correspond to the hydro ecoregion codes; see Figure 3)	75
Table 2: Acronym summary of the metrics used in this paper	80
Table 3: Average standard deviation (SD) and variation coefficient (CV) of IBGN and I2M2 metrics for each substrate and at all sites (in light blue $16\% < CV < 33.3\%$, in light blue and bold $CV > 33.3\%$) $CV < 16\%$ = not significant, $16\% < CV < 33.3\%$ = significant, and $CV > 33.3\%$ = critical error. (Example CV average of blocks = (blocks CV site 1 + blocks CV site 2 + blocks CV site 3/ 3)).	81
Table 4: Average of standard deviation (SD) and variation coefficient (CV) of supplementary metrics for each substrate. $CV < 16\%$ = not significant, $16\% < CV < 33.3\%$ = significant, and $CV > 33.3\%$ = critical error. (in light blue $16\% \leq CV < 33.3\%$, in light blue and bold $CV > 33.3\%$; example: CV average of blocks =(blocks CV site 1+blocks CV site 2+blocks CV site 3/3))	81
Table 5: Standard deviation (SD) and variation coefficient (CV) for each site and for each metric for block and sand substrates (in light blue $16\% < CV < 33.3\%$, in light blue and bold $CV > 33.3\%$) $CV < 16\%$ = not significant, $16\% < CV < 33.3\%$ = significant, and $CV > 33.3\%$ = critical error.	83
Table 6: Inferior ecological class thresholds by 3 river types and 7 HER (concerned in this study) for the IBGN score (Ministère de l'Ecologie, du Développement durable et de l'Energie, 2012). (thresholds explanation : a-b-c-d : a = very good status inferior threshold,, b = good status inferior threshold, c = moderate status inferior threshold, d = poor status inferior threshold)	91
Table 7: Part of the confusion matrix on macrophytes taxa (CLASPX = Cladophora sp., MICSPX = Microspora sp., RHISPX = Rhizoclonium sp., STISPX = Stigeoclonium sp., STITEN = Stigeoclonium tenuie, VAUSPX = Vaucheria sp).	154
Table 8: Confusion matrix on macrophytes cover classes (C0: absent, C1: covering $< 0.01\%$, C2: $0.1\% \leq$ covering $< 1\%$, C3: $1\% \leq$ covering $< 10\%$, C4: $10\% \leq$ covering $< 50\%$, C5: covering $\geq 50\%$ of stream bed)	155
Table 9: Example of an original floristic list \mathcal{L}	156
Table 10: Modified floristic list on taxa from original floristic list L (Table 14) obtained with the algorithm defined in Section 3 after abundance classes modifications (TAX)	157
Table 11: Modified floristic list on abundance classes from original floristic list \mathcal{L} (Table 14) obtained with the algorithm defined in Section 3 after abundance classes modifications (AC)	158
Table 12: Averages of CVt and DSt for each trophic status (Very low, Low, Medium, High, Very high), error type (TAX, AC and TAX+AC) and each error rate. Averages of CVt and DSt were calculated based on 100 tests at 5 sites per trophic status, error type, and error rate. The average of IBMR scores were obtained based on the 5 sites per trophic status.	165
Table 13: Proportion of downgrading and upgrading (in %) for each trophic status and error type for an experienced surveyor (error rate $\leq 25\%$) for all the 25 sites (600 tests/ trophic status)	166
Table 14: Minimal, maximal and average of Z-score after the three modifications (TAX, AC, TAX+AC) obtained with the algorithm defined in §2 for each trophic status (Very low, Low, Medium, High, Very high) (average of absolute values, Number of original sites $n=25$, $n=5$ /trophic status, number of modified lists 20 tests/site/ error rate, 100 tests/error rate/error type)	166
Table 15: Minimal, maximal and average of Z-score after the three modifications (TAX, AC, TAX + AC) obtained with the algorithm defined in §2 based on a random selection for three trophic states (very low, medium, very high) (average of absolute values). Number of original site $n = 1$ /trophic status, number of modified lists 20 tests/error rate/error type, same sites as Fig. 38).	167
Table 16: Values of major confUSion between Macrophytes Indicator (JUMI) for each taxa able of being confused (74 taxa among the 219 IBMR taxa, in blue $JUMI \geq 7 $)	168

Je profite de ces pages pour remercier toutes les personnes qui ont contribué au bon déroulement de cette thèse et qui m'ont permis de m'épanouir pleinement durant ces trois années.

Tout d'abord, j'adresse mes plus sincères remerciements à mes deux co-directrices de thèse Michèle Trémolières et Florence Le Ber. Vous m'avez apporté un grand soutien durant ces trois années que ce soit par le biais de vos nombreuses relectures ou vos conseils précieux qui m'ont offert la possibilité de mener à bien ce projet ambitieux. Je tiens également à exprimer ma profonde gratitude envers mes deux encadrants Corinne Grac qui, lors de mon stage de fin de Master 2, m'a initiée au monde des macroinvertébrés et s'est montrée d'une très grande patience et générosité, et Frédéric Labat qui m'a transmis sa passion débordante pour les petites bêtes et donne envie de toujours aller plus loin dans la réflexion. Cette thèse n'aurait pu voir le jour sans votre présence et la confiance que vous m'avez accordées.

Je suis également très reconnaissante envers Jacques Haury pour avoir accepté de présider le jury de soutenance de thèse de doctorat, mais également pour son intérêt et ses remarques constructives qui m'ont permis de prendre du recul sur mon travail et d'améliorer de manière significative mon manuscrit.

Je tiens également à remercier Philippe Usseglio-Polatera qui a accepté d'examiner ma thèse en étant l'un des rapporteurs. Je suis extrêmement reconnaissante pour l'attention particulière portée à mon travail que ce soit en participant à mon comité de pilotage, en répondant à mes requêtes diverses et variées mais également pour toutes les remarques qu'il m'a suggérées pour améliorer la qualité de ma thèse.

Merci à Laure Berti-Equille pour avoir accepté d'être examinatrice lors de ma soutenance de thèse.

Je remercie également toutes les personnes qui ont répondu présentes et ont accepté de participer à mon comité de pilotage de thèse : Christian Chauvin, Thierry Leviandier, Yorick Reyjol et Philippe Usseglio-Polatera. Vos connaissances et conseils m'ont guidée tout au long de ce travail de thèse.

Remerciements

Ayant bénéficié d'un co-financement, ces trois années ont été partagées entre Saint-Germain-du-Puch (33) et Strasbourg (67). Je tiens donc remercier les responsables d'équipes, chef de laboratoire et directeur : Bruno Fontan (PDG d'Aquabio), Dominique Badariotti (LIVE) et Jean-François Quéré (ENGEES) pour m'avoir permis de travailler dans des conditions optimales et fourni tous les outils nécessaires que ce soit matériel ou intellectuel au bon déroulement de cette thèse.

Un grand merci à toute l'équipe d'Aquabio, vous avez pris le pari fou de miser sur moi pour mener à bien ce long travail. Vous m'avez toujours soutenue malgré les difficultés rencontrées. Vous m'avez offert un cadre de travail idéal, toujours dans une ambiance agréable et donné tous les moyens de réussir. Un merci particulier à ceux que j'ai emmenés sur le terrain pour leur aide et bonne humeur, Lucky Luke, Jojo et Mich Mich.

Je remercie également Mickaël Fabrègue qui a conçu les algorithmes utilisés dans cette thèse. Grâce à toi, j'ai découvert des langages insoupçonnés enrichissant mes connaissances et étant très pédagogue et à l'écoute tu a été très efficace dans la réalisation de mes requêtes. Merci également, Jérôme Simon qui m'a aidée pour la conception des matrices de confusion sur les macrophytes. Sans tes lumières, je n'aurais pas pu réaliser un tel travail, tu as toujours pris le temps de me parler des plantes, ton aide a été incontestable.

Mickaël et Xavier, mes compagnons de bureau « engeesiens », merci pour ses moments de rigolade, de grande discussion en terrasse et pour nos instants de pause salvatrice.

Une pensée pour Eugénie et Sarah, les dernières doctorantes de Michèle, je vous remercie pour votre soutien et vos encouragements durant les derniers mois.

Je remercie également mes amis qui ont été présents durant ces trois années. Vous avez toujours su me motiver et me reconforter même dans les moments de doutes. Merci à Romain et Céline, ma « famille adoptive » pour les apéro-jeux des plus divertissants et tous les moments passés ensemble, Eva pour les soirées atypiques complètement folles, Fred et Steph pour les apéros improvisés et Titi, Claire, Ben, Dam, Sophie et tous ceux que j'oublie pour m'avoir toujours soutenue. Merci à mes amies d'Alsace Manue's, Gaëlle et Joanne pour nos soirées filles qui m'ont bien changé les idées. Merci à David pour m'avoir fait découvrir la Franche-Comté et offert de grandes bouffées d'oxygène.

Un grand merci à mes parents et ma sœur, vous m'avez encouragée, supportée. Vous vous êtes montrés vraiment très compréhensifs surtout dans la dernière ligne droite, où le stress était à son comble. Merci encore pour toute l'aide que vous m'avez apporté.

Enfin, je remercie l'ANR qui a permis, en partie, le financement des travaux de cette thèse dans le cadre du projet Fresqueau (ANR11 MONU14) dans lequel Aquabio est partenaire.

INTRODUCTION

L'incertitude est-elle notre seule certitude¹ ? Cette pensée philosophique a engendré de nombreuses doctrines en philosophie mais aussi dans le domaine des Sciences. Pyrrhon d'Elis (360-275 av J.-C.), père fondateur de l'école sceptique, est l'un des premiers à s'être penché sur la notion de doute. Il a mis en place la distinction entre, ce que l'on appellerait aujourd'hui, le subjectif et l'objectif (Brochard, 2002). D'autres, bien après lui, ont continué dans cette mouvance et ont apporté d'autres fondements. Ainsi, René Descartes s'est opposé aux sceptiques, qui selon lui étaient irrésolus et ne s'appuyaient sur rien. Descartes aspirait à étendre la certitude mathématique à l'ensemble des connaissances. Selon lui, tous les phénomènes pouvaient être expliqués mathématiquement (Descartes, 1637). Il a développé son discours autour de la méthode de doute avec comme premier principe le « *Cogito, ergo sum* » (je pense, donc je suis), ainsi il posa les premiers jalons de son *Discours de la méthode* (Descartes, 1637).

Dans les années 1920 a émergé un nouveau courant, le positivisme logique (Carnap, et al., 2010), rompant avec les pensées qui tentaient d'expliquer des phénomènes par des croyances mystérieuses. Le positivisme logique essaye d'apporter des lois permettant de décrire et de prédire des phénomènes. Pourtant, il connaît un recul avec la découverte de la physique quantique, puisque celle-ci démontre qu'un phénomène ne peut être simplement interprété par des images de la pensée et ne peut être prédit par des lois (Langevin, 1950; De Broglie, 1983; Planck, 1989). Le positivisme a alors laissé sa place, dans les années 1960, à la théorie du chaos (Lorenz, 1972) qui sous-entend que tout phénomène mathématique est chaotique et que les systèmes sont non linéaires (Ekeland, 1995).

¹ <http://cafes-philo.org/> consulté le 10 octobre 2014

A l'heure actuelle, il est demandé aux scientifiques d'être toujours plus précis et d'évaluer la part d'incertitude dans leurs mesures, comme par exemple en métrologie où de nombreuses définitions sont données (JCGM 200, 2008). En hydrobiologie, il est acquis que de nombreuses incertitudes existent dans les protocoles d'évaluation des milieux aquatiques. En effet, la communauté scientifique connaît la complexité des écosystèmes, d'autant plus lorsque s'y ajoutent les activités humaines. Cette complexité apporte une variabilité importante, difficilement appréciable, qui ne peut être approximée que par le biais d'expérimentations ou de modèles. De plus, la plupart des mesures effectuées en hydrobiologie repose sur des protocoles s'appuyant sur les connaissances et l'expérience des hydrobiologistes. Ceux-ci s'appropriant les normes, leur subjectivité se retrouve au cœur des évaluations. Ainsi, les incertitudes associées à l'évaluation de la qualité des cours d'eau, au travers des indices biologiques, forment une thématique d'actualité majeure, en particulier pour les différents compartiments biologiques (oligochètes, diatomées, poissons, invertébrés et macrophytes).

La prise de conscience de l'existence des incertitudes de mesure au sein des indices biologiques a donc conduit la communauté scientifique à des réflexions sur : (1) la nécessité d'évaluer les incertitudes de mesures (2) sur la possibilité de les réduire, et (3) l'amélioration possible et/ou le développement de protocoles répondant aux exigences de la Directive Cadre Européenne sur l'Eau (Directive 2000/60/CE, 2008), fournissant des données valides et de qualité.

Dans ce contexte, nos travaux de recherche se focalisent sur (1) l'étude de l'impact des incertitudes, définies comme toutes les erreurs induites par les facteurs environnementaux et actions humaines, sur deux indices biologiques basés sur les macroinvertébrés et les macrophytes en rivière, (2) et l'estimation de certaines de ces incertitudes et d'observer leur propagation jusqu'à l'évaluation globale de la qualité écologique des rivières.

Dans une première partie, nous présentons les outils de biosurveillance et leur historique (chapitre 1), l'état actuel des recherches sur les incertitudes (Chapitre 2) qui nous permettra :

(1) d'approfondir et d'évaluer l'ensemble des sources d'incertitudes existantes dans la chaîne de production des données biologiques depuis le terrain (échantillonnage) jusqu'au choix de la classe d'état obtenue (calcul indice, choix des seuils des classes d'état) en passant par la phase laboratoire (tri, détermination) pour les deux compartiments biologiques choisis,

(2) d'approfondir pour ces deux compartiments les incertitudes déjà étudiées, et enfin, une présentation de nos stratégies de recherche pour ces deux groupes (Chapitre 3).

Pour les macroinvertébrés, nous nous sommes particulièrement intéressés :

- (1) à l'impact de la variation intra-substrat, qui n'a jusqu'ici pas encore été étudié dans la construction des indices comme l'Indice Biologique Global Normalisé (IBGN, AFNOR, 2004) et le nouvel indice français l'I2M2 (Mondy, et al., 2012)
- (2) et à l'effet des incertitudes induites par la subjectivité et les compétences de l'opérateur, ainsi qu'à la propagation de celles-ci dans le calcul de l'IBG DCE et de l'I2M2.

Pour les macrophytes, notre objectif est d'analyser les incertitudes basées sur l'effet opérateur, à la fois in situ et au laboratoire, comme les oublis de taxons, les erreurs de détermination et les erreurs d'estimation des classes d'abondance associés à l'Indice Biologique Macrophytes en Rivière (IBMR, AFNOR, 2003).

Une deuxième partie sera consacrée aux protocoles expérimentaux et méthodes virtuelles mis en place pour mesurer les incertitudes définies et associées à chacun de deux groupes, ainsi que les résultats obtenus (Partie 2, Chapitre 1, 2 et 3).

La variabilité intra substrat (Chapitre 1) et son impact sur les métriques des indices basés sur les macroinvertébrés (IBGN et I2M2) ont été testés au travers d'une démarche expérimentale basée sur un échantillonnage répété des habitats sur des rivières françaises. Ce protocole s'appuie sur les règles strictes d'échantillonnage établies par les normes françaises en vigueur, et permet de s'affranchir, en partie, du biais humain variable puisque toutes les étapes (échantillonnage, tri et identification des taxons) ont été réalisées par un opérateur unique.

L'estimation des incertitudes associée à l'opérateur (Chapitre 2) s'appuie sur des données existantes issues des campagnes de terrain menées par Aquabio (bureau d'études expert en hydrobiologie), dans le cadre du suivi du réseau de contrôle de surveillance des eaux. Ces données nous ont permis de développer une approche innovante pour étudier les effets opérateur sur le calcul des indices IBGN et I2M2. Celle-ci repose sur l'élaboration d'une méthode spécifique nous permettant de générer des modifications virtuelles contrôlées, à savoir des suppressions de taxons rares ou peu abondants (moins de 5 individus) et des sous-estimations aléatoires d'abondance. Afin d'avoir un contrôle sur les modifications réalisées, il a été nécessaire :

- (1) de définir des valeurs seuils pour les deux types d'erreurs (déterminés grâce à la littérature

et aux normes) et

(2) de définir des probabilités d'erreurs à introduire dans la méthode (obtenues par le biais de l'étude sur la variabilité intra-substrat).

L'étude de l'effet opérateur (terrain et au laboratoire) sur l'indice IBMR (Chapitre 3) s'appuie, elle aussi, sur le développement d'une méthode spécifique qui génère des modifications virtuelles des déterminations de taxons et des classes de recouvrement. Cette méthode a nécessité la création de deux matrices de confusion s'appuyant :

(1) sur les erreurs d'identification entre les taxons contributifs à l'IBMR et

(2) sur les erreurs d'estimation des classes de recouvrement (valeurs de confusion déterminées par dires d'experts).

Nous avons également défini des taux d'erreur théoriques associés aux compétences des opérateurs.

Enfin, après avoir démontré l'impact des incertitudes environnementales et celles liées à l'opérateur sur les métriques, les indices et sur l'évaluation de l'état des rivières, nous proposerons une synthèse des résultats obtenus et une critique des méthodes utilisées, une extension de nos méthodes sur les autres compartiments biologiques et des perspectives de recherche (Partie 3).

PARTIE 1 Etat de l'art et stratégies de recherche

Cette partie se découpe en trois chapitres. Le premier donne un état des lieux des outils de biosurveillance ainsi que l'historique de leur construction pour deux compartiments biologiques : les macroinvertébrés et les macrophytes. Le second porte sur les incertitudes associées à ces deux compartiments et aux indices biologiques qui seront traités dans cette étude. Enfin, le dernier chapitre présente les axes de recherches privilégiés.

CHAPITRE 1 Outils de biosurveillance



Ce chapitre présente le contexte réglementaire dans lequel ont été produits les outils de biosurveillance et donne un état des lieux des différents indices basés sur les macroinvertébrés et les macrophytes, leur évolution et leurs équivalents européens.

Titre 1 : Contexte réglementaire

Depuis la décennie 1970, la préoccupation pour la protection des ressources naturelles en eau ne cesse de s'accroître. Le réchauffement climatique et le manque de précipitations dans certaines régions du monde oblige la communauté internationale à réfléchir à des moyens de préserver la ressource en eau mais également à lui garantir une bonne qualité pour les générations futures.

En Europe, une politique publique commune a été engagée avec la Charte européenne de l'eau le 26 mai 1967 (Commission Européenne, 1968). Mais, il a fallu attendre la Directive européenne Cadre sur l'Eau (DCE) adoptée le 23 Octobre 2000, pour connaître la première réglementation communautaire relative au bon état écologique des rivières avec la définition d'objectifs par masse d'eau (Directive 2000/60/CE, 2000) et sa mise en application en France en 2004 (Code de l'Environnement, 21 avril 2004). La DCE a mis en évidence l'importance de développer des outils permettant l'évaluation de l'état des masses d'eau et la synthèse des informations multiples et complexes sur la structure et le fonctionnement des hydrosystèmes.

La DCE place la biologie au cœur de l'évaluation de l'état écologique, ce qui a provoqué une généralisation des indices biologiques déjà proposés par certains pays d'Europe. Ces pays, dont la France, ont donc continué à développer leurs indices biologiques ou bioindicateurs pour les généraliser à l'échelle européenne. Dans ce cadre, un élargissement des groupes biologiques utilisés a vu le jour avec une réflexion pour que les indices existants puissent être intercalibrés entre pays et soient compatibles avec la DCE (Birk & Hering, 2006; Birk, et al., 2006; Buffagni, et al., 2006; Buffagni & Furse, 2006; Erba, et al., 2009; Birk & Willby, 2010). Institutions gouvernementales et experts finalisent actuellement ces outils en vue de les adapter aux exigences de la DCE.

La DCE impose plusieurs exigences en matière de bio-évaluation auxquelles les méthodes développées par les pays européens doivent répondre (Conseil Européen, 2007).

Il est nécessaire de définir des conditions de référence prenant en compte les variabilités naturelles. Ainsi, la DCE propose deux systèmes distincts :

- (A) un système A figé qui correspond à une régionalisation se traduisant par des écorégions caractérisées par des paramètres imposés aux limites de classe prédéfinies comme

l'altitude, la géologie et la taille du bassin versant et

(B) un système B plus souple, qui correspond à une classification caractérisée par des paramètres obligatoires (latitude-longitude, altitude, géologie et taille du bassin versant) ou optionnels (distance à la source, débit, substrat, etc.).

De nombreux pays européens, dont la France, ont reconnu le système A comme valide puisque la diversité naturelle est fonction des caractéristiques géographiques des milieux aquatiques. Pourtant, il peut facilement devenir peu pertinent : (1) les écorégions proposées sont liées à la distribution des insectes aquatiques (Illies, 1978) et ne sont donc pas représentatives du fonctionnement du milieu et (2) la prise en compte des caractéristiques géologiques sont insuffisantes et peuvent masquer des discontinuités naturelles. Ainsi, les pays européens ont souvent préféré développer de nouvelles typologies dérivées du système B. En France, l'approche régionalisée n'est pour autant pas abandonnée puisqu'elle reste une typologie simple, robuste et cohérente avec les pressions anthropiques et permet de prendre en compte la grande variabilité climatique, altitudinale et géologique de la France. Elle a donc été transposée à l'espace français et une approche par hydro-écorégion (HER), intégrant bon nombre de facteurs du système B, a été développée (Wasson, 2001; Wasson, et al., 2004). Les HER sont délimitées par des déterminants primaires (géologie, relief et climat) et sur le fait qu'au sein d'une même HER les cours d'eau se distinguent par des caractéristiques physiques, chimiques et biologiques similaires et que les cours d'eau de différentes régions se différencient sur au moins un paramètre abiotique important (Hughes & Larsen, 1988; Gallant, et al., 1989; Warry & Anau, 1993). 22 HER ont donc été définies et validées par des études basées sur les macroinvertébrés (Wasson, et al., 2002) et sur les diatomées (Tison, et al., 2005; Rison, et al., 2007).

La DCE recommande de calculer des ratios de qualité écologique (EQR, Conseil Européen, 2007) ou des écarts à la référence. En effet, pour détecter et comprendre des changements dans les écosystèmes résultant de l'activité humaine, il est nécessaire de pouvoir évaluer les sites par rapport à une référence et donc de mesurer des écarts (Karr & Chu, 1997).

La DCE stipule également la nécessité de prendre en compte les abondances, la diversité et le ratio entre les taxons sensibles et les taxons résistants. Ces variables inhérentes aux futurs indices développés permettraient de mettre en lumière plusieurs pressions anthropiques (Conseil Européen, 2007).

Les outils développés devraient à terme répondre à ces exigences, pourtant à l'heure actuelle

ils peuvent encore se montrer imparfaits (Titre 3).

Titre 2 : La bioindication

La complexité des hydrosystèmes et la multiplicité des sources de dégradations anthropiques ont nécessité, pour évaluer la qualité des eaux, des indicateurs biologiques en plus des paramètres physiques et chimiques. En effet, il est apparu que les variables biologiques étaient de bons intégrateurs des perturbations des milieux aquatiques (Verneaux, 1973; Reyjol, et al., 2012). La bioindication reposait au préalable sur la recherche de la présence/absence d'organismes et éventuellement leur abondance, ceci permettant d'apporter des informations précises sur la qualité du milieu (Blandin, 1986). Selon Bellan, « les indicateurs biologiques ou taxons indicateurs sont des détecteurs révélant l'existence de conditions complexes à interpréter et résultant le plus souvent d'un ensemble de facteurs biotiques ou abiotiques difficiles à mesurer directement » (Bellan, 1984). Blandin stipule que les bioindicateurs sont définis comme des espèces animales ou végétales permettant de donner une indication du niveau de dégradation d'un milieu par le biais de leurs caractéristiques physiologiques, éthologiques ou écologiques. Ils mettent ainsi en évidence des modifications naturelles ou provoquées des milieux (Blandin, 1986).

Ainsi, dès le début du XX^{ème} siècle, en Europe, le développement des premiers indices de saprobie ont permis l'introduction des notions d'indicateurs des conditions environnementales (Kolkwitz & Marsson, 1909). Puis à partir des années 70 en France, le concept d'indices biocénétiques s'est développé et a connu une véritable impulsion, avec notamment la mise au point des premiers indices basés sur les macroinvertébrés (cf titre 3).

Dans les années 80, ont été développées les premières approches multimétriques avec la conception d'un indice multimétrique basé sur les communautés de poissons (IPR) (Oberdorff, et al., 2001; Oberdorff, et al., 2002) et inspiré des premiers travaux américains énonçant la notion d'intégrité biotique (Karr, 1981; Karr, 1986).

Le contexte de la DCE offre de nouvelles perspectives aux méthodes de bioindication, puisque celle-ci préconise d'évaluer les milieux aquatiques à l'aide d'approche indiquant

l'ensemble des impacts biologiques induits par les différentes pressions subies par le milieu. Ainsi, les méthodes doivent évoluer et être améliorées avec notamment la prise en compte de nouveaux paramètres. De plus, la conception d'indices multimétriques présente l'avantage d'intégrer à la fois la variabilité naturelle du milieu dans l'espace et le temps mais aussi les différentes pressions. De ce fait, les exigences imposées par la DCE ont été un véritable moteur et ont entraîné un effort dans le développement des méthodes de bioindication pour chacun des éléments de qualité biologiques.

Les évolutions de la bioindication ont conduit à des choix dans le processus de conception des outils de surveillance.

En effet, il a été nécessaire de construire des méthodes garantissant leur reproductibilité, la comparabilité des résultats obtenus, la fiabilité des données récoltées mais aussi une certaine notion de coût-rentabilité. Ce dernier critère a été fixé afin que les méthodes puissent être applicables en routine (Conseil Européen, 2007; Lorenz, et al., 2004). En France, les méthodes bioindicatrices développées s'inspirent principalement de méthodes européennes déjà existantes (les indices biologiques macrophytes et macroinvertébrés français se sont inspirés des méthodes anglo-saxonnes, l'Indice Biologique Macrophytique en Rivière IBMR du Mean Trophic Rank anglo-saxon et l'Indice Biologique Global Normalisé du Trent Biotic Index).

L'intérêt ici est de présenter un historique des méthodes de bioindication fondées sur les compartiments macroinvertébrés et macrophytes et de comprendre les différents choix qui ont été faits pour répondre à la fois aux critères précités mais également aux exigences de la DCE.

Titre 3 : Historique des indices biologiques

En Europe, plusieurs compartiments biologiques des rivières ont été étudiés et ont permis de créer les différents indices biologiques. Les macroinvertébrés, les diatomées, les poissons, les macrophytes et les oligochètes en sont les cinq principaux.

Nous ne développons ci-après que l'historique des indices sur lesquels est basé ce travail de recherche, à savoir les macroinvertébrés (Section 1) et les macrophytes (Section 2) en France et en Europe.

Section 1 : Les indices basés sur les macroinvertébrés

Cette partie exposera l'évolution des indices biologiques basés sur les macroinvertébrés en France (§1) et les indices existant dans d'autres pays européens (§2).

§1. En France

Dès les années 1960, il y a eu une effervescence autour de l'intérêt de qualifier l'état écologique des rivières. Jean Verneaux (Tuffery & Verneaux, 1968) en est l'un des principaux investigateurs et a permis la création d'un premier indice biologique (1.) qui a ensuite évolué au cours des décennies suivantes (2., 3., 4.).

1. Les premiers indices

Les premiers indices ont été développés en vue de répondre aux problèmes de fortes dégradations des milieux aquatiques, notamment liées à des insuffisances en matière d'assainissement. Les indices basés sur l'étude des macroinvertébrés permettent d'évaluer les pressions organiques, physiques et les modifications de l'habitat, mais ne sont pas des indices de saprobie tels que définis par Sladeczek (1973). Les invertébrés sont de bons intégrateurs de la qualité globale de l'écosystème aquatique (Agences de l'eau, 2000).

Les premiers indices ont vu le jour en France dès 1967 avec l'Indice Biotique (IB) (Tuffery & Verneaux, 1968) inspiré par le Trent Biotic Index (TBI) (Woodiwiss, 1964) utilisé en Angleterre. La note de cet indice est calculée à partir d'une table à double entrée prenant en compte la richesse taxonomique et le groupe indicateur qui correspond au taxon le plus polluo-sensible. Il s'appuie sur des prélèvements réalisés en zone lotique et en zone lentique. Cet indice, créé en Franche-Comté, se montrait encore imprécis (niveau taxonomique : famille, ordre) et ne donnait pas de

résultats suffisamment satisfaisants dans le reste de la France, aussi subit-il plusieurs modifications. Il devient alors, en 1976, l'Indice de Qualité Biologique Globale (IQBG) (Verneaux, et al., 1978). Cet indice se voulant plus sensible et fidèle à la réalité, s'appuie sur la recherche non plus des zones lenticues et lotiques mais de tous les substrats, susceptibles d'accueillir des invertébrés, présents sur la station étudiée. Des rectifications ont néanmoins lieu et l'IQBG devient en 1982, l'indice biologique global (IBG) (Verneaux & coll., 1982). Il se distingue des autres indices grâce à sa méthode d'échantillonnage plus précise, à une liste taxonomique mieux définie mais également grâce à un tableau de calcul de l'indice plus fin (38 taxons indicateurs et 7 classes de qualité). Le protocole repose sur un mode d'échantillonnage de huit habitats, i.e. couples substrat-vitesse, choisis par ordre décroissant de capacité biogène des substrats en se limitant aux plus biogènes lorsque plus de huit habitats sont présents, ou si leur nombre est insuffisant en répétant les substrats les plus biogènes, en faisant varier les classes de vitesse en fonction de leur biogénicité définie pour partie selon l'expertise de l'opérateur. Ce mode d'échantillonnage devait permettre d'obtenir une image de la diversité du site suivie la plus complète possible. L'IBG pose encore un problème de reproductibilité de l'échantillonnage et se verra remanié dans la normalisation par l'AFNOR. Il deviendra l'Indice Biologique Global Normalisé (IBGN).

2. L'Indice Biologique Global Normalisé (IBGN)

L'IBGN a été normalisé en 1992, réactualisé en 2004 (AFNOR, 2004) et un guide technique a été édité (AFNOR, 2006). Reposant sur les mêmes bases fondamentales que les précédents indices, il a subi d'importantes modifications du protocole d'échantillonnage et d'exploitation des listes faunistiques. En effet, il s'appuie toujours sur l'échantillonnage de huit habitats par ordre décroissant de capacité biogène des substrats en se limitant aux plus biogènes si nécessaire (plus de huit habitats présents) ou à l'inverse en répétant les plus abondants ou représentatifs si le nombre de substrats est limité. Ainsi la notion d'expertise de l'opérateur disparaît ce qui rend la méthode reproductible. Le classement des substrats par ordre décroissant de leur capacité biogène, dans le tableau d'échantillonnage, a été modifié par rapport à l'IBG et le tableau à double entrée a été ajusté : il compte désormais quatorze classes de variété contre douze pour l'IBG et neuf groupes faunistiques repères contre huit pour l'IBG.

La détermination de l'IBGN s'appuie donc sur une méthode plus fine et précise que les précédents indices (IB, IQBG et IBG) et offre une meilleure prise en compte de l'ensemble des

niveaux typologiques (amplitude typologique et conditions biogéographiques) des systèmes d'eau courante en s'appuyant sur les travaux de biotypologie de Verneaux (Verneaux, 1973). Il permet, à l'échelle d'une station, d'évaluer la qualité biologique de l'eau mais également d'évaluer l'état des habitats constituant le site. Il donne la possibilité aux gestionnaires de suivre l'évolution temporelle et spatiale (amont/aval) de l'état biologique de l'eau courante d'une station et d'évaluer l'effet d'une perturbation sur la faune invertébrée du milieu.

L'obtention de l'indice s'appuie sur une grille à double entrée basée sur les quatorze classes de variété taxonomique (VC) qui correspondent au nombre de familles présentes et les neuf groupes indicateurs (GFI) qui sont établis en fonction de la polluosensibilité des taxons (AFNOR, 2004). La note de l'indice est obtenue grâce à la formule suivante :

$$IBGN = GFI + (VC - 1)$$

La note alors obtenue est comprise entre 0 et 20 et permet d'établir la classe de qualité biologique. Les classes de qualité biologique sont définies par le SEQ-Bio (Agences de l'Eau, 2000). Pour l'IBGN il existe cinq classes de qualité biologique oscillant entre très mauvaise et bonne qualité.

Le développement des différents indices basés sur les macroinvertébrés au fil du temps, aboutissant à l'IBGN, a permis de normaliser plusieurs critères tels que la taille minimale du prélèvement, le plan d'échantillonnage permettant de donner une image de la mosaïque d'habitats, le nombre de prélèvements ou encore le niveau de détermination des taxons (pour l'IBGN, la famille principalement).

3. L'Indice Biologique Global DCE (IBG DCE)

La DCE préconise des résultats produits par des échantillonnages représentatifs des principaux habitats présents sur le site suivi (Archambault & Dumont, 2010). Or, l'IBGN s'appuie sur l'échantillonnage des habitats les plus biogènes et ne permet pas de mesurer l'écart entre les peuplements observés et les peuplements de référence. Il ne correspond pas aux exigences de la DCE et n'est donc plus adapté au suivi de la qualité des rivières (Archambault & Dumont, 2010). Afin de répondre aux préconisations de la DCE et de conserver une cohérence dans les chroniques

de suivis de l'état écologique, la France a développé dans un premier temps un nouveau protocole d'échantillonnage reposant sur les préceptes de l'IBGN avant de produire un nouvel indice. Ainsi, une circulaire française a vu le jour pour apporter des recommandations et des modifications au protocole utilisé (Conseil Européen, 2007). Par la suite, de nouvelles normes NF XP T90-333 et NF XP T90-388 (AFNOR, 2009; AFNOR, 2010) ont été mises au point : la première est consacrée à l'échantillonnage et la deuxième aux traitements en laboratoire des échantillons comprenant les opérations de tri et d'identification. Ces protocoles appliqués au sein des réseaux de surveillance, donnent un peu plus d'importance à la représentativité des substrats et non plus à leur seule capacité biogène. L'IBG DCE s'appuie sur l'échantillonnage d'habitats marginaux (phase A, superficie strictement inférieure à 5%) et d'habitats dominants (superficie supérieure ou égale à 5% de la station) tout en prenant en compte leur habitabilité et leur représentativité. Dans la continuité de l'IBGN, les phases A et B se rapprochent de son mode d'échantillonnage puisqu'elles prennent en considération les habitats marginaux et dominants prélevés en suivant l'ordre d'habitabilité des substrats. Les phases B et C rentrent dans une logique DCE puisque les habitats dominants sont échantillonnés à la fois selon l'habitabilité de leur substrat (Phase B) mais aussi selon leur représentativité (Phase C). Ainsi, ce mode d'échantillonnage permet de pouvoir utiliser des chroniques en calculant un équivalent « IBG » ou IBG DCE, il associe les avantages de l'IBGN à ceux des méthodes européennes et est comparable aux autres indices européens (Tableau 1).

Au laboratoire, les niveaux requis d'identification des taxons ont également évolué puisque selon les groupes, l'identification est réalisée au genre ou à la famille. La détermination au genre limite les erreurs d'identification des taxons à la famille et donc réduit les incertitudes d'identification.

L'application des deux normes précitées permet toujours l'obtention d'un indice calculé sur le même principe de la norme IBGN (AFNOR, 2004). Cet indice est nommé IBG DCE (appellation utilisée dans la suite du manuscrit) et non l'IBGN.

Pourtant, l'IBG DCE répond de façon imparfaite à la DCE, puisqu'il ne prend pas compte explicitement l'abondance et la diversité des taxons ou encore l'absence de calcul d'écart à un état de référence (Conseil Européen, 2007). De plus, il présente une faible sensibilité à certaines pressions notamment hydromorphologiques (Reyjol, et al., 2012). Aussi, à terme, il devrait être remplacé par un nouvel Indice Invertébrés MultiMétrique : l'I2M2, compatible avec les exigences de la DCE.

4. Le nouvel Indice Invertébrés MultiMétrique : l'I2M2

Le nouvel indice proposé vise donc à combler les faiblesses et lacunes de l'IBGN et de l'IBG DCE par rapport aux exigences DCE, exigences qui ont conduit au développement d'un nouvel outil d'évaluation de la qualité écologique des rivières. Le nouvel indice multimétrique d'évaluation de l'état écologique : I2M2 (Mondy, et al., 2012) s'appuie sur les données faunistiques acquises à l'aide des protocoles normalisés XP T90-333 et XP T90-888. L'I2M2, contrairement à l'IBG DCE se calcule à partir des identifications requises par la norme (ordre, famille et genre). Les niveaux d'identification différents peuvent amener une source d'incertitude supplémentaire. Dans de nombreux modèles, dont l'I2M2, il est recommandé pour limiter cette incertitude de choisir un niveau d'identification constant qui correspond à un nettoyage des listes afin de faciliter la comparaison inter-sites ou avec un site de référence (Moss, et al., 1999; Ostermiller & Hawkins, 2004). De plus, il a été montré que l'identification à la famille et au genre permettent toutes deux de répondre à des impacts, cependant il existe une plus forte corrélation au niveau générique (Pond, et al., 2008).

La construction de l'I2M2 a pris en considération 10 catégories de pressions chimiques (micropolluants, matières phosphorées, etc.) et 7 catégories de pressions hydromorphologiques (instabilité hydrologique, colmatage, etc.). Cet indice permet un ratio de qualité écologique (EQR) qui correspond à l'écart entre la situation observée et la situation de référence.

Pour sa construction, plus de 2500 métriques taxonomiques décrivant les communautés macrobenthiques ont été testées en fonction de différents critères statistiques comme par exemple le caractère positif ou négatif de la réponse des différentes métriques à une catégorie de pression donnée ou le calcul de coefficients de variation. Parmi ces 2500 métriques, seulement cinq ont été retenues. Ces cinq métriques complémentaires apportent des informations sur la communauté en place, il s'agit de l'Average Score Per Taxon (ASPT), la diversité de Shannon (H), la richesse taxonomique (TAX) (selon le niveau requis dans la norme XP T90-388), la fréquence relative des organismes polyvoltins (PS) et la fréquence relative des organismes ovovivipares (OS). Les trois premières (ASPT, H' et TAX) sont couramment utilisées dans d'autres indices européens prouvant ainsi leur caractère bio-indicateur et leur robustesse (Chester, 1980; Alba-Tercedor & Sanchez-Ortega, 1988; Gardeniers, et al., 1996; Gabriels, et al., 2010). Les deux dernières (PS et OS) sont des métriques innovantes montrant l'intérêt de développer une approche plus fonctionnelle.

L'**Average Score Per Taxon** ASPT, (Armitage, et al., 1983) calculé sur les habitats dominants (phases B et C), correspond au niveau de polluo-sensibilité moyen de l'assemblage des macroinvertébrés benthiques (au moins un individu) et varie de 0 à 10.

L'**indice de diversité de Shannon** H' , (Shannon, 1948) est calculé sur les habitats les plus biogènes (phases A et B) avec la formule suivante:

$$H' = - \sum_{i=1}^n (P_i \times \ln P_i) \quad (1)$$

Où P_i correspond à la proportion de l'effectif du taxon i par rapport à l'effectif total des n taxons.

L'indice de diversité de Shannon est un indice de diversité spécifique (taxons identifiés à l'espèce), or le niveau d'identification requis des taxons dans le calcul de l'I2M2 est plus faible, ce qui peut entraîner une part d'incertitude sur les valeurs de l'indice. Guerold a montré que le changement de niveau d'identification entre espèce et genre/genre-famille conduisait à une sous-estimation de la valeur H' . Cependant, il est à noter que lors de cette étude tous les taxons n'ont pas été identifiés à l'espèce (Guerold, 1992; Guerold, 2000).

La richesse taxonomique (TAX) correspond au nombre de taxons identifiés au niveau systématique requis par la norme AFNOR XP T90-388 au moins représenté par un individu dans l'ensemble du prélèvement (phases A, B et C).

La fréquence relative des taxons polyvoltins (PS) (i.e. avec au moins deux générations successives au cours d'une année) (Tachet, et al., 2010) est calculée sur l'ensemble des listes faunistiques (phases A, B et C). Cette métrique révèle une stabilité ou instabilité stationnelle (Tachet, et al., 2010). En effet, lorsqu'un milieu connaît une dégradation, les espèces à cycle long ont tendance à disparaître (espèces monovoltines), tandis que les espèces à cycle courts augmentent (espèces polyvoltines).

La fréquence relative des taxons ovovivipares (OS) (i.e. dont la totalité du développement embryonnaire s'effectue dans l'abdomen de la femelle, l'éclosion des œufs suivant immédiatement leur expulsion dans le milieu aquatique) est calculée sur l'ensemble des listes

faunistiques (phases A, B et C). Cette métrique informe sur les dégradations de l'habitat (communication personnelle, Usseglio-Polatera, P., 2014 ; Van Looy, et al., 2014).

Les fréquences relatives des taxons polyvoltins et ovovivipares sont calculées sur l'ensemble du prélèvement (phases A, B et C) selon la formule suivante (Usseglio-Polatera, et al., 2000) :

$$M = \frac{(\sum_i^S m_i \times \ln(q_i + 1))}{\sum_i^S \ln(q_i + 1)} \quad (2)$$

Où M correspond à la fréquence relative d'utilisation de la modalité considérée sur l'ensemble du prélèvement, S le nombre total de taxons, m_i la fréquence relative d'utilisation de la modalité m du trait biologique considéré pour le taxon i , et q_i l'abondance du taxon i dans l'assemblage faunistique considéré.

Grâce à leur combinaison et leur pondération par des coefficients de discrimination, elles permettent d'intégrer plusieurs types de pressions physico-chimiques et morphologiques. Ainsi, sont calculés 17 sous-indices pour chacune des catégories de pression associée à la qualité de l'eau. La note de l'indice I2M2 correspond à la moyenne arithmétique de ces 17 sous-indices et est comprise entre 0 et 1.

Pour conclure, l'I2M2 est un nouvel indice qui permet le calcul d'écart à la référence ; il prend en compte la typologie des cours d'eau mais également les abondances, la diversité et le ratio entre taxons sensibles et taxons résistants; et montre une sensibilité aux 17 catégories de pression prises en compte dans l'élaboration de l'indice et représentatives des principales pressions subies par les milieux aquatiques. Karr et Chu (1997) ont déjà pu démontrer que les indices multimétriques apportaient des évaluations plus robustes que les indices monométriques. Ainsi, ce nouvel indice, l'I2M2, donne une évaluation plus efficace, plus robuste et plus en adéquation avec les critères de la DCE et devient comparable aux autres méthodes européennes.

§2. En Europe

Si La France a développé plusieurs indices biologiques basés sur les macroinvertébrés, les autres pays européens ont également développé différents indices biologiques (1) et conçu pour

certaines des modèles de prédiction (2).

1. Les indices européens

Tout comme la France, la plupart des pays européens concernés par l'évaluation de qualité biologique des cours d'eau, ont commencé leur réflexion sur le développement d'outils de biosurveillance dans les années 1970-1980. Avec la mise en place de la DCE, ils ont créé leurs propres indices d'évaluation de l'état écologique des rivières. Ces indices reposent sur des protocoles de terrain qui diffèrent de l'IBGN, néanmoins le résultat se calcule souvent sur le même modèle de tableau à double entrée. En Europe, il existait en 2012 environ 70 méthodes d'évaluation appliquées aux rivières, aux eaux côtières ou encore aux lacs (Birk, et al., 2012). Le Tableau 1 présente certaines méthodes référencées d'évaluation de la qualité des eaux européennes.

Tableau 1 : Méthodes européennes d'évaluation de la qualité écologique des rivières

Italie Indice Biotico Estes	Méthode d'échantillonnage : Prélèvements effectués sur un transect reliant les deux berges, le nombre d'échantillons est défini en fonction des habitats rencontrés. 4 saisons d'échantillonnage par an, les périodes de hautes eaux sont à éviter. Tri et identification : Tri sur le vivant et détermination à la famille ou au genre en fonction des groupes Calcul de la note de l'indice : utilisation d'une grille à double entrée (variété et groupes indicateurs)
IBE (Ghetti, 1997)	
Pays Bas Ecologische BEO ordelingsmethode voor Stromend WA ter op basis van macrofauna	Méthode d'échantillonnage : les habitats dominants sont sélectionnés sur 10 à 30 m de longueur dans le cours d'eau. Même effort d'échantillonnage pour tout habitat. Tri et identification : Tri sur le vivant à l'œil et détermination à la famille ou au genre en fonction des groupes Calcul de l'indice : calcul de 4 métriques : l'indice de Shannon, l'ASPT, le DSFI et un indice d'acidité
EBEOSWA (Gardeniers, et al., 1996)	
Danemark Danish Stream Fauna Index	Méthode d'échantillonnage : Procédure normalisée et échantillonnage, en principe, de tous les habitats. 3 transects sont placés à travers le cours d'eau tous les 10m. Prélèvement avec technique du « <i>kick sampling</i> » (le filet est placé au fond du lit face au courant, le pied de l'opérateur se retrouve face au filet et il effectue des mouvements à contre-courant sur 40 cm pour décoller le substrat et les invertébrés qui sont alors entraînés au fond du filet). L'opération est répétée 4 fois sur chaque transect en partant d'une berge. En complément de ces prélèvements sont

DSFI, (Skriver, et al., 2000)	<p>effectués des «<i>hand-picking</i>» pour les invertébrés adhérents au support. Le support est alors frotté 5 min de plus dans le filet. Ces deux types de prélèvement sont conservés séparément. La saison privilégiée pour réaliser cette méthode d'échantillonnage est le printemps.</p> <p>Tri et identification : les prélèvements de «<i>kick sampling</i>» et «<i>hand-picking</i>» sont analysés en laboratoire sans sous-échantillonnage.</p> <p>Détermination des individus à la famille ou au genre en fonction des ordres.</p> <p>Calcul de la note de l'indice : utilisation d'une grille à double entrée (variété, groupe indicateur)</p>
<p>Lettonie Latvian Sampling Method</p>	<p>Méthode d'échantillonnage : sélection d'un site représentatif de la rivière d'une longueur de 20 à 50 m. La station est prospectée et tous les substrats sont recensés ainsi que leur occurrence. Des mesures physico-chimiques sont réalisées (comme par exemple le pH et la température de l'eau).</p> <p>Les invertébrés sont prélevés avec un racleur de fond ou avec une pince sur les branches ou les pierres. 20 échantillons élémentaires sont obtenus en fonction de l'occurrence des biotopes. La saison d'échantillonnage privilégiée est l'automne et le printemps.</p> <p>Tri et identification : tri effectué sur le terrain, les individus sont comptés. Si l'identification est impossible, des échantillons sont conservés pour un temps illimité. L'identification des individus se fait au niveau du genre ou de l'espèce.</p> <p>Calcul de la note de l'indice : calcul d'un indice saprobique</p>
LVS 240, (Latvian Standard LVS 240, 1999)	<p>Méthode d'échantillonnage : L'échantillonnage est réalisé en toutes saisons sauf l'hiver. Tous les habitats sont échantillonnés à l'aide de la technique du «<i>kick sampling</i>» et du «<i>hand picking</i>» sur une bande de 10 à 20m pendant 3 à 5 minutes. Lorsque le cours d'eau est trop profond des substrats artificiels sont placés dans le lit durant 3 semaines.</p> <p>Tri et identification : Tous les échantillons collectés sont examinés au laboratoire. Détermination au niveau requis (famille, genre ou niveau intermédiaire). Les individus sont comptés et lorsqu'il y a plus de 10 individus pour un même taxon les abondances sont estimées.</p> <p>Calcul de la note de l'indice : indice multimétrique basé sur 5 métriques (Nombre total de taxons présents TAX, Indice de Shannon & Wiener, nombre d'Ephéméroptères, Plécoptères et/ou Trichoptères présents EPT, score moyen de tolérance MTS, nombre de taxons sensibles NST), le score est égal à la somme des 5 métriques divisée par 20.</p>
MMIF (Gabriels, et al., 2010)	<p>Méthode d'échantillonnage : Echantillonnage réalisé au printemps et/ou en automne. Tronçon de 10 m de long le plus homogène possible en termes de mosaïque d'habitats. Profondeur maximum 1m et vitesse supérieure à 10cm/s. les</p>
<p>Suède Swedish invertebrate method</p>	<p>Méthode d'échantillonnage : Echantillonnage réalisé au printemps et/ou en automne. Tronçon de 10 m de long le plus homogène possible en termes de mosaïque d'habitats. Profondeur maximum 1m et vitesse supérieure à 10cm/s. les</p>

<p>SIS Method, (SS-EN 27828, 2000)</p>	<p>zones d'échantillonnage sont restreintes aux radiers et au substratum dur. La méthode employée est le « <i>kick sampling</i> » (sur 1m et pendant 1min). Les échantillons sont récoltés sur une zone homogène de 50m long. Cinq répétitions pour chaque échantillonnage. Les échantillons sont conservés séparément.</p> <p>Tri et identification : tri et détermination à l'espèce.</p> <p>Calcul de la note de l'indice : Calcul de l'indice MILA qui s'appuie sur 6 indices (nombre de famille, nombre de mollusques, nombre d'éphéméroptères, le ratio entre l'abondance relative des plécoptères et l'abondance relative des éphéméroptères, l'indice AWIC (acid waters indicator community index) et l'abondance relative des broyeurs. Les valeurs de l'indice sont comprises entre 0 et 100.</p>
<p>Angleterre Biological Monitoring Working Party</p> <p>BMWP, (Chester, 1980)</p>	<p>Méthode d'échantillonnage : les principaux habitats sont échantillonnés en 3 minutes par « <i>kick sampling</i> »</p> <p>Tri et identification : détermination à la famille.</p> <p>Calcul de la note de l'indice : à chaque famille est attribué un score en fonction de sa polluosensibilité, le score de l'indice est la somme des scores de chaque famille présente. Une alternative est le calcul de l'ASPT qui est la moyenne du BMWP.</p>

2. Les modèles prédictifs

Les indices biologiques s'étant bien développés à travers l'Europe, une autre façon de penser a émergé. En effet, jusqu'à présent les méthodes s'appuyaient sur la connaissance de la diversité taxonomique et de la polluosensibilité supposée des taxons qui sont fonction des conditions hydromorphologiques et physico-chimiques des rivières. La nouvelle approche de l'évaluation de la qualité des cours d'eau s'appuie sur des modèles de prédiction.

Le premier modèle prédictif RIVPACS ou *River InVertebrate Prediction And Classification System* a été développé par le CEH (Center for Ecology & Hydrology) au Royaume-Uni (Wright, et al., 1988). Ce modèle, aboutissement d'un projet lancé sur quatre ans, répondait à une demande croissante d'une meilleure connaissance de l'écologie des cours d'eau dans le cadre de leur gestion. L'objectif de ce projet était de déterminer si la communauté des macroinvertébrés sur un site pouvait être prédite à partir des caractéristiques hydromorphologiques et physico-chimiques. Pour cela a été réalisée une classification des sites d'eau courante non pollués basée sur leur faune macrobenthique.

Le modèle d'évaluation RIVPACS d'un cours d'eau repose sur la comparaison des indices

biotiques de la faune observée (obtenue par prélèvements réalisés sur un site en suivant un protocole standardisé), et de la faune attendue (s'il n'y a pas de perturbations ou de pollutions) et permet d'obtenir des indices de qualité environnementale (EQI). Plus la valeur EQI est importante plus la différence entre la faune observée et la faune attendue diminue ce qui signifie que le site est proche des conditions de référence.

Ce modèle s'est répandu à travers l'Europe et s'applique en Autriche, en Allemagne ou encore en Grèce. De plus, il a l'avantage de répondre aux exigences de la DCE puisqu'il permet de mesurer des écarts aux références.

D'autres modèles prédictifs dérivés de RIVPACS ont vu le jour à travers le monde tels que PERLA la variante Tchèque (Kokeš, et al., 2006), AusRivas le modèle australien (Simpson & Norris, 2000), Medpacs le modèle méditerranéen (Poquet, et al., 2009) ou encore le modèle Mondego version portugaise (Feio, et al., 2007).

Les modèles prédictifs sont également adaptés à d'autres groupes comme par exemple les modèles MADPACS (MACrophyte Prediction And Classification System) et MAC (Macrophyte Assessment and Classification) qui s'appliquent aux communautés végétales (Aguilar, et al., 2011).

Section 2 : Les indices macrophytiques

Tout comme les macroinvertébrés, dès les années soixante-dix les macrophytes ont suscité l'intérêt des hydrobiologistes et leurs premières réflexions ont visé à répondre à la même problématique de l'évaluation de la qualité des eaux. Il a été montré un lien fort entre la distribution des macrophytes et l'eutrophisation des cours d'eau liés aux rejets de nitrates et phosphates. Ainsi dès le début des années 1970, Kohler, et al. (1971) ont établi une typologie des communautés de macrophytes en relation avec les paramètres chimiques des cours d'eau. Ces auteurs allemands ont ouvert le chemin à d'autres qui ont conforté leurs résultats et développé des typologies en France (Carbiener, et al., 1990; Robach, et al., 1996; Thiébaud & Muller, 1999) et enfin ont proposé des indices biologiques (Schneider & Melzer, 2003). Contrairement, aux indices basés sur les macroinvertébrés où le calcul s'obtient souvent par le biais d'une grille à double entrée (variété, polluosensibilité), les indices macrophytiques sont calculés par des

formules relativement similaires données ci-après pour les 3 principaux indices.

Nous développerons ici les deux principaux indices européens : le rang trophique moyen (MTR) (Holmes, et al., 1999) (§1) et le Trophie-Index Macrophyten (TIM) (Schneider & Melzer, 2003) (§2), sur lesquels s'est appuyé la conception de l'indice français, l'Indice Biologique Macrophytique en Rivières (IBMR) (Haury, et al., 2006) (§3).

§1. Le Rang Trophique Moyen (MTR)

Le rang trophique moyen ou MTR (Holmes, et al., 1999) est l'indice macrophytique actuellement utilisé au Royaume Uni. Il repose sur les premiers travaux de Harding (Harding, 1981). Il a été créé dans le but de répondre à la nécessité d'améliorer la gestion des rivières en ayant une meilleure connaissance des milieux permettant ainsi une meilleure appréciation du traitement des rejets urbains. Il évalue la situation trophique des communautés installées.

Le protocole de cet indice se base sur un relevé effectué sur des sites d'une longueur de 100m, un croquis de la station suivie peut être également réalisé. A chaque taxon est attribué un coefficient d'abondance et les taxons identifiés sont répertoriés.

Il s'appuie sur le calcul d'un indice basé sur une formule prenant en considération les coefficients d'abondances (CV) et les scores trophiques (TS) de chaque taxon.

Il se calcule par la formule suivante :

$$MTR = \frac{\sum(TS \times CV)}{\sum CV} \times 10 \quad (3)$$

Où :

- TS correspond au score trophique allant de 1 à 10 pour chaque taxon,
- CV correspond au coefficient d'abondance attribué à chaque taxon, exprimé en classe allant de 1 à 9 et
- le coefficient multiplicateur 10 permet d'obtenir un indice MTR variant de 10 à 100.

Plus la valeur est proche de 100 plus la rivière a un faible niveau trophique et plus la valeur est proche de 10 plus la rivière a un fort niveau trophique.

§2. Le Trophie-Index Macrophyten (TIM)

L'Indice Trophique Macrophytique ou Trophie-Index Macrophyten (TIM) (Schneider & Melzer, 2003) est un indice allemand appliqué aux cours d'eau de Bavière en Allemagne. Cet indice est construit à partir d'une valeur bioindicatrice (I_{wa}) basée sur la valeur trophique de différents végétaux et de leur indice de tolérance. Il prend en compte la notion d'abondance d'une espèce par un indice de quantité de l'espèce (Q_a) mais également de son importance grâce à un indice de poids fonction de la tolérance de chaque espèce (G_a). L'indice comporte sept classes de trophie et se calcule suivant la formule suivante :

$$TIM = \frac{\sum_{a=1}^n I_{wa} \times G_a \times Q_a}{\sum_{a=1}^n G_a \times Q_a} \quad (4)$$

Avec

- I_{wa} qui correspond à la valeur indicatrice de l'espèce a ,
- G_a indice de poids de l'espèce a et,
- Q_a qui représente l'indice de quantité de l'espèce a .

Tout comme le MTR, le TIM nous renseigne sur l'état trophique d'une rivière en la classant d'oligotrophe à polytrophe ($1 \leq TIM \leq 4$).

§3. L'indice Biologique Macrophytique en Rivières (IBMR)

En France, la conception d'indices basés sur les macrophytes s'est véritablement développée dans les années 1990. Pourtant, l'utilisation des approches typologiques (Butcher, 1933; Kohler, 1971; Holmes, 1983) et phytosociologiques (Carbiener, et al., 1990) mettaient déjà en avant le lien entre qualité des eaux et groupements aquatiques (Meriaux, 1978; Oberdorfer, 1990; Haury, 1994).

Les premiers indices s'appuyant sur des listes de macrophytes présents sur un site portaient sur des calculs de valeurs moyennes d'indices spécifiques (variant de 1 espèce tolérante, à 10

espèce sensible) des espèces présentes ou d'observation d'écart à la référence (Harding, 1981; Haslam, 1982). Ces démarches ont posé les premiers jalons des méthodes d'évaluation des eaux par les macrophytes mais elles présentaient certaines lacunes comme le manque de pertinence de certaines valeurs de cotes spécifiques et une certaine faiblesse taxonomique (Haury & Peltre, 1993).

La demande des gestionnaires croissante, d'intégrer les végétaux dans les outils d'évaluation a conduit à la conception des premiers indices français. Principalement basés sur les modèles anglo-saxons (Harding, 1981; Holmes, et al., 1999), le premier indice français, l'indice GIS, avait pour objectif de fournir un outil d'évaluation de la qualité de l'eau, de minéralisation et de pH, de trophie et de milieu physique des cours d'eau (Haury, et al., 1996).

L'indice GIS, basé sur des cotes spécifiques CS_i variant entre 0 et 10 reprenant les propositions de Harding (Harding, 1981), prenait en compte la présence/absence des espèces (PA) associé ou non à un coefficient d'abondance-dominance (AD), ou l'ensemble des espèces aquatiques (AQ) ou uniquement les espèces aquatiques sténoèces (AQSt) ou encore les espèces aquatiques et les espèces supra-aquatiques (A+S).

Ainsi, le calcul de l'indice repose sur deux variantes :

- la première sans pondération par les abondances :

$$\text{Indice GIS (PA)} = \sum_i CS_i / n \quad (5)$$

avec pour chaque taxon i la côte spécifique CS.

- la deuxième avec pondération par les recouvrements estimés par les coefficients d'abondance-dominance (AD) :

$$\text{Indice GIS (AQ, A + S)AD} = \left(\sum_i AD_i \times CS_i \right) / \sum_i AD_i \quad (6)$$

avec pour chaque taxon i la côte spécifique CS et les coefficients d'abondance-dominance (AD).

L'indice GIS a fait l'objet de diverses applications expérimentales en France et est une adaptation de l'indice anglais MTR à la France. En effet, il prend en compte une plus grande

diversité d'espèces de macrophytes adaptée au territoire français, les cotes spécifiques attribuées aux espèces résultent d'un croisement d'informations obtenues sur plusieurs écorégions, enfin les deux variantes d'indices permettent d'obtenir des diagnostics complémentaires. Ainsi, l'indice GIS représente les prémices de l'Indice Biologique Macrophytique en Rivière ou IBMR (Haury, et al., 2002; Haury, et al., 2006).

Dans le cadre de la DCE, l'IBMR a été normalisé sous la référence NF T90-395 en octobre 2003 (AFNOR, 2003). L'IBMR traduit essentiellement le degré trophique global du cours d'eau lié aux teneurs en nutriments (azote, phosphore), ainsi qu'aux pollutions organiques fermentescibles (Haury, et al., 2006). Il dépend secondairement des caractéristiques physiques du milieu comme les intensités de la lumière, les écoulements et la sédimentation (Haury, et al., 2006).

Le protocole de calcul de cet indice se base sur l'observation exhaustive in situ des peuplements macrophytiques, l'identification des taxons et l'estimation de leurs recouvrements sur une surface donnée (minimum de 100m²) dans deux faciès de courant lentique et lotique. Un prélèvement d'échantillons est fait pour vérification taxonomique, si nécessaire.

Le calcul de la note IBMR est alors réalisé à partir de la liste floristique, considérant les taxons de la liste des 209 taxons contributifs (fournie dans la norme AFNOR, 2003) et des cotes spécifiques (*CS_i*) qui varient de 1 pour les espèces hyper-eutrophes à 20 pour les espèces oligotrophes.

Cette note s'obtient grâce à la formule suivante :

$$IBMR = \frac{\sum_i^n E_i \times K_i \times CS_i}{\sum_i^n E_i \times K_i} \quad (7)$$

Où pour une espèce *i*:

- *CS_i* correspond à la cote spécifique du niveau trophique,
- *K_i* est le coefficient d'abondance (1 à 5 selon la gamme de recouvrement) et
- *E_i* représente le coefficient de sténoécie (1 euryèce à 3 sténoèce).

L'IBMR se différencie de l'indice anglo-saxon MTR de par les classes d'abondances proposées mais également par l'ajout du coefficient de sténoécie. Diminuer le nombre de classe de recouvrement à 5 alors qu'il est de 9 (pour une longueur de 100m) pour le MTR (Holmes, et al., 1999), réduit l'incertitude découlant de la subjectivité des opérateurs sur le terrain. Le coefficient de sténoécie, en plus de la cote spécifique, donne le degré de bio-indication de sensibilité des

taxons à un niveau trophique donné.

L'indice IBMR ne répond pas aux exigences de la DCE, puisqu'il ne permet pas le calcul d'écart à la référence (EQR). Ainsi, il doit évoluer pour y répondre. Le travail commencé en 2011 par l'IRSTEA, devrait aboutir en 2016 (REEE 2016) à une proposition de notes indicielles de référence par type de cours d'eau. L'EQR correspond au ratio entre note observée et note de référence, est compris entre 0 et 1. Ces valeurs seront également converties en classe de qualité écologique.

Outre l'évolution des indices, la mise en place de la DCE a soulevé différentes problématiques :

- le recensement et l'évaluation de toutes les perturbations par les outils biologiques existants,
- la nécessité d'harmonisation des outils mis en place entre les différents pays européens et,
- les incertitudes liées à ces différents indices.

Des éléments de réponses sont apportés dans la section 3.

Section 3 : Les nouveaux questionnements engendrés par la DCE

Dans le cadre de la DCE, il est nécessaire de pouvoir comparer les modalités d'évaluation de la qualité des rivières européennes entre elles. Trois programmes ont retenu notre attention car ce sont les principaux projets européens menés au cours des dix dernières années : le projet AQEM (§1), le projet STAR (§2) et le projet WISER (§3) qui ont traité de ces questions. Ces projets ont tous en commun l'objectif de définir des pratiques de terrain et de laboratoire harmonisées et d'élaborer une démarche d'évaluation de l'état écologique à l'échelle européenne.

§1. Le projet AQEM : un système d'évaluation de la qualité écologique des rivières européennes

Un des premiers programmes européens, lancé en 2000, est le projet AQEM² (*Assessment system for the ecological Quality of streams and rivers throughout Europe using benthic Macroinvertebrates, 2000-2002*) (Hering, et al., 2004). Ce programme avait pour but la mise en œuvre de la DCE avec le développement d'un système d'évaluation européen harmonisé de la qualité écologique des cours d'eau européens basé sur les macroinvertébrés benthiques.

Le développement d'une méthode européenne unique et fiable a nécessité l'étude des sources potentielles d'incertitudes. Tout d'abord a été testée la cohérence de la typologie des cours d'eau en Europe et notamment si celle-ci répondait aux exigences de la DCE (Moog, et al., 2004; Verdonschot & Nijboer, 2004). Il a été montré que les principales régions géographiques européennes (écorégions, Conseil Européen, 2007) regroupaient efficacement l'ensemble des types de rivière définis et, qu'en général, la distribution des macroinvertébrés des cours d'eau européens répondait aux conditions climatologiques et géomorphologiques que ce soit à large échelle ou à une petite échelle ; conditions sur lesquelles sont basées le découpages des écorégions (Verdonschot & Nijboer, 2004). Cependant à une échelle régionale, il a été constaté que le pouvoir discriminant de plusieurs métriques biologiques pour évaluer l'état écologique augmentait avec l'utilisation d'une classification biorégionale (division d'une écorégion en plusieurs biorégions) (Moog, et al., 2004). Ainsi, même si la typologie de la DCE est utile dans l'élaboration d'un système d'évaluation basé sur les macroinvertébrés, il n'en reste pas moins que la notion d'écorégion est trop large et variable pour fournir des directives en matière de gestion régionale. Il est donc intéressant de prendre en compte ces deux typologies de cours d'eau (écorégions et bio-régions). La réflexion d'AQEM sur les différentes typologies (Système A et bio-régions) souligne également la pertinence du développement de la typologie française en HER. AQEM a ensuite harmonisé et développé des méthodes d'échantillonnage sur le terrain et au laboratoire. Le protocole d'échantillonnage des macroinvertébrés sur le terrain repose sur les préceptes du *Rapid Bioassessment Protocols* (RBA, Barbour et al., 1999), des procédures de l'Agence de l'Environnement (Environment Agency, 1999), des lignes directrices autrichiennes (Moog, et al., 1999) et de la norme ISO 7828 (ISO 7828, 1985). Ce protocole d'échantillonnage

² <http://www.aqem.de/>

AQEM s'appuie sur la conception d'une cartographie multi-habitat et prend seulement en considération les habitats dominants au prorata de leur surface de recouvrement (supérieure ou égale à 5%). Ainsi, vingt échantillons sont récoltés, à savoir de manière générale sept en zone de radier (*riffle*) et treize en zone de mouille (*pool*), lorsque celles-ci sont bien différenciées (Hering, et al., 2004).

Le choix de la méthode d'analyse des échantillons au laboratoire a suscité une recherche de fiabilité des résultats se traduisant par l'étude de certaines sources d'incertitudes, e.g. le nombre d'individus à sortir et à identifier, l'effet de l'exclusion des taxons peu abondants ou du niveau d'identification des taxons sur la qualité des résultats produits (Lorenz, et al., 2004; Nijboer & Schmidt-Kloiber, 2004; Schmidt-Kloiber & Nijboer, 2004). Ces incertitudes et leur impact ont plus largement été traités dans le chapitre 2. Ainsi, les échantillons sont ensuite triés au laboratoire et l'identification est effectuée au meilleur niveau atteignable en privilégiant l'espèce.

Le projet AQEM portait deux objectifs majeurs :

- (1) placer un tronçon d'une rivière au sein d'une classe de qualité allant de 5 (très bon) à 1 (médiocre) en fonction de sa liste taxonomique et
- (2) apporter une information quant aux causes d'une possible dégradation, dans le but d'orienter les gestionnaires des cours d'eau dans les futurs aménagements et/ou la réduction des pressions.

L'originalité de ce projet reposait sur un nouveau jeu de données prenant en compte à la fois la faune et les caractéristiques des cours d'eau communes à vingt-huit types de rivières des pays européens. Par conséquent, pour chaque type de masses d'eau, différentes méthodes de calcul ont été appliquées en fonction des sites de référence. Cependant le même modèle d'évaluation harmonisé a été appliqué à tous les types de rivières et le programme voulait également généraliser des outils applicables à la biosurveillance de l'ensemble des cours d'eau européens (Hering, et al., 2004).

La démarche d'AQEM se divise en deux étapes :

- (1) identification et évaluation de la perturbation subie par chaque type de masses d'eau. Il peut s'agir de l'acidification, d'une pollution organique ou encore d'une pression hydromorphologique (Buffagni, 2001). Néanmoins dans certains cas plusieurs facteurs de stress peuvent être identifiés et évalués séparément, ils sont ensuite réunis dans le cadre de l'évaluation de la « dégradation générale ».

- (2) développement du système multimétrique pour lequel la formule est construite, pour chaque type de cours d'eau, à partir de la combinaison de différentes méthodes de calcul qui indiquent le mieux l'état des sites suivis. A partir d'une liste taxonomique, sont calculées des métriques, telles l'EPT ou l'indice de diversité de Simpson, qui nous permettent d'obtenir un score et donc une classe de qualité. Dans la mesure du possible, les métriques utilisées dans le calcul de l'état écologique d'une rivière sont catégorisées selon leur habilité à détecter l'impact de certains facteurs.

L'indice multimétrique unique permet l'attribution d'un score final allant de 5 (qualité très bonne) à 1 (qualité médiocre).

§2. Le projet STAR : Normalisation de la classification des rivières

Le projet STAR³ (*Standardisation of River Classifications*, 2001-2005) (Clarke & Hering, 2006) démarré en 2001 a réuni, au sein d'un consortium de plusieurs pays européens dont le Royaume-Uni, l'Autriche, la République Tchèque, la France et la Suède.

Ce projet partait du constat qu'il existait une variété de méthodes d'évaluation des cours d'eau en Europe qui a permis la mise en vigueur de la DCE. Pourtant, la diversité de ces méthodes et de leurs résultats pouvaient entraîner de sérieux problèmes stratégiques de gestion des rivières. Le nombre de groupes d'organismes utilisé pour évaluer l'état écologique, et les méthodes disponibles pour le faire étaient si divers que les enjeux de la standardisation des méthodes et de l'inter-calibration paraissaient donc cruciaux.

Le but du projet était de proposer une méthode générale d'utilisation des différents groupes d'organismes (macroinvertébrés, macrophytes, poissons et phytoplancton) pour l'évaluation des cours d'eau. Plusieurs questions, comme par exemple quelles méthodes ou quels groupes taxonomiques faudrait-il utiliser par type de pression, ont été posées. Les erreurs affectant les méthodes d'évaluation ont été étudiées. Ainsi, les sources d'incertitudes éventuelles associées à chaque étape de l'évaluation de l'état écologique, de la phase d'échantillonnage sur le terrain jusqu'à la précision des modèles utilisés, ont été incluses dans le développement de l'outil. Tout

³ <http://www.eu-star.at/frameset.htm>

comme pour le projet AQEM, le projet STAR a donc étudié l'incertitude liée à l'échantillonnage, au tri et à l'identification des taxons (Clarke, et al., 2000; Clarke, et al., 2002; Clarke & Hering, 2006; Clarke, et al., 2006; Lorenz & Clarke, 2006). Des audits ont été menés sur les différentes étapes : ils ont conclu à l'importance de l'entraînement des opérateurs. L'étude des incertitudes devait permettre d'établir les bases d'un modèle conceptuel. (Furse, et al., 2006). Ainsi, ils ont développé un indice de métriques communes inter-étalonnées (ICMi) qui regroupent les métriques en trois catégories : tolérance aux perturbations, abondance/habitat et richesse/diversité. À chaque métrique est associé un poids de robustesse par rapport à la catégorie à laquelle elle appartient. Ce poids a été obtenu par analyse de corrélation, par exemple la métrique ASPT, associée à la catégorie de tolérance, a un poids de 0.333. Dans chaque catégorie, plus le poids de la métrique est élevé plus la métrique est robuste. Le score de l'indice résulte de la somme des scores des six métriques pondérés (Buffagni, et al., 2005).

Un autre aboutissement du projet a été le développement d'un module d'erreur ou *starbugs* (Clarke & Hering, 2006; Staniszewski, et al., 2006). Celui-ci a été produit pour apporter une aide dans l'évaluation de l'impact des différentes sources de variation sur une ou plusieurs mesures de l'incertitude globale dans l'attribution de cours d'eau à des classes de l'état écologique. Il permet de simuler la distribution des ratios de qualité écologique qui sont obtenus après ajout des variations liées aux différentes sources d'incertitude. Ainsi, *starbugs* prend en considération la variation liée à l'échantillonnage, au tri et à l'identification mais également celle liée aux valeurs des conditions de référence. Ce module permet de rappeler qu'il n'existe pas de vérité absolue et que l'incertitude est présente dans toutes les approches mises en place et qu'elle ne peut être évaluée seulement à l'aide d'information disponible limitée (Clarke & Hering, 2006). C'est pourquoi il est plus que nécessaire d'étudier les sources d'incertitudes existantes à chacune des étapes des méthodes d'évaluation de la qualité écologique des rivières.

§3. Le projet Wiser : Integrative Systems to assess Ecological status and Recovery

Un des plus récents projets est le projet Wiser⁴ (2009-2012). Il avait pour objectif de soutenir la mise en œuvre de la DCE en développant des outils d'évaluation de l'état écologique des eaux

⁴ <http://www.wiser.eu/>

de surface européennes en intégrant les lacs, les zones côtières et les eaux de transition et en évaluant le processus de récupération des eaux dégradées.

Le programme Wiser s'inscrit dans un contexte où plusieurs lacunes subsistent. En effet, peu de pays ont déjà achevé l'élaboration de leurs systèmes d'évaluation et de nombreux groupes indicateurs (poissons, phytoplancton, etc.) ne sont pas intégrés dans ces systèmes. De plus, les impacts de certaines pressions, comme la dégradation hydromorphologique sur la faune et la flore, restent largement méconnus. Il existe également peu d'informations quant aux incertitudes associées aux systèmes d'évaluation et à la comparaison des indices européens entre eux. Enfin, les connaissances sur la récupération des différents groupes suite à une pression et leur réponse face au changement climatique sont encore insuffisantes, limitant ainsi la réussite des restaurations futures entreprises par les gestionnaires.

Le programme Wiser s'est focalisé plus particulièrement sur ces lacunes et a essayé de les combler. Contrairement aux projets AQEM et STAR, il ne propose pas de méthode harmonisée d'échantillonnage sur le terrain et au laboratoire et s'appuie sur les différentes méthodes des pays européens participant à ce programme.

Il s'est intéressé aux incertitudes qui affectent la classification des eaux de surface mais également aux métriques associées à chaque indice comme le préconise la DCE (Clarke & Jones, 2009; Clarke, 2012). En effet, la DCE exige que les résultats d'évaluation présentent des intervalles de confiance et une estimation de la précision des déterminations.

Plusieurs sources d'incertitudes peuvent être dues aux variations spatiale et temporelle naturelle, à la méthode d'échantillonnage ou encore à la modélisation prédictive utilisée (Clarke & Hering, 2006).

Pour évaluer l'importance relative des différentes sources de variation, le programme Wiser s'appuie sur les travaux d'incertitudes du projet STAR. Celui-ci a mis en lumière pour les macrophytes par exemple, que la variabilité inter-opérateur était plus faible que les effets des variations temporelles (années et saisons) et de l'ombrage et que la plus forte variation était due à des modifications physiques de l'habitat (e.g. pont, réaménagement du chenal) (Staniszewski, et al., 2006).

Le programme a abouti au développement d'un outil « *Wiserbugs* » ou « *WISER Bioassessment Uncertainty Guidance Software* » qui permet d'obtenir des estimations des classes de confiance basées sur des métriques individuelles, des indices multimétriques et/ou des classifications multimétriques à partir des éléments de la qualité biologique de la DCE (Clarke, 2013).

CHAPITRE 2 Incertitudes associées aux deux indices biologiques

L'évolution des indices biologiques au sein de la DCE a conduit à l'étude des incertitudes induites par l'application des différentes méthodes. Les incertitudes ont été largement abordées dans les programmes de recherche précités plus particulièrement par le programme WISER. Dans ce chapitre, nous donnons tout d'abord quelques définitions des incertitudes (Titre 1), puis nous aborderons les sources d'incertitudes liées aux évaluations de l'état écologique des rivières par les macroinvertébrés et les macrophytes (Titre 2).

Titre 1 : Qu'est-ce que l'incertitude

L'incertitude, ou le caractère de ce qui est incertain (Larousse, 2014), est au cœur de nombreux domaines scientifiques et s'est trouvée introduite dès 1797 par Jean-François de La Pérouse, au cours de son voyage autour du monde (1785-1788). Il mettait en doute la précision des mesures liée à l'utilisation des tables de la Lune⁵, bien que celles-ci aient été réalisées par des opérateurs parfaitement exercés (Milet-Mureau, 1797). En médecine également, Claude Bernard montrait que l'expérimentateur pouvait avoir un rôle dans l'incertitude du fait de ses connaissances et ainsi conduire à une imprécision (Bernard, 1865). En physique quantique, il était impossible, en théorie, de connaître simultanément la position et la vitesse d'une particule avec une précision supérieure à un certain seuil. Heisenberg comprit que toute détermination liée à la position et de la quantité de mouvement de la particule observée serait entourée d'incertitudes (Heisenberg, 1930) et développa les « relations d'incertitudes ».

Ainsi, la notion d'incertitude peut être déclinée dans plusieurs domaines scientifiques thématiques. En sciences économiques et sociales, par exemple, l'incertitude est affiliée à la notion de variable inobservable dépendante d'autres variables observables (Bloom, et al., 2013) ou encore à la notion de probabilité ou de risques (Moureau & Rivaud-Danset, 2004; Zinn, 2008) et donne une limitation des approches économiques basées sur le concept d'efficacité dans le cadre de l'explication et de la prévision des résultats (Beckert, 1996).

Quelque soit le domaine, on s'accorde à penser qu'elle est liée à l'idée de mesure et donc de résultats. En conséquence, la notion d'incertitude s'est développée au travers de concepts de base associés aux différentes sciences et prend en compte différents paramètres variables pouvant entraîner un biais, une erreur. Par exemple, la précision d'un équipement de mesure que ce soit le sextant, un instrument de navigation utilisé au XVI^{ème} siècle, pour la Pérouse, ou une sonde de mesure de pH pour un technicien d'aujourd'hui, a sûrement un effet sur la mesure. L'opérateur qui effectue la mesure, selon son habileté, son expérience, sa formation ou encore son honnêteté

⁵ Les tables de la Lune de Mayer (1752) permettent de déduire la longitude grâce à la distance angulaire vraie entre deux astres mesurée au sextant. Toute erreur sur la distance observée produit une erreur trente fois plus forte sur la détermination de la longitude.

ou sa rigueur, peut influencer sur la qualité de la mesure (ASME, 2007).

En métrologie, les incertitudes sont depuis longtemps appréhendées et mesurées. Des normes existent pour les définir et les quantifier. Ce traitement des incertitudes est celui qui peut le plus s'apparenter à celui utilisé en hydrobiologie. C'est pourquoi nous précisons les définitions des incertitudes proposées en métrologie. Une **erreur de mesure** est définie comme la « *différence entre la valeur mesurée d'une grandeur et une valeur de référence ou valeur vraie* ». L'**incertitude de mesure** caractérise la « *dispersion des valeurs attribuées à un mesurande (grandeur que l'on veut mesurer) à partir des informations utilisées* » (JCGM 200, 2008). Elle est, en général, associée à une valeur déterminée attribuée au mesurande. Une modification de cette valeur entraîne une modification de l'incertitude associée. L'incertitude reflète donc le manque de connaissance précise de la valeur d'un mesurande (JCGM 200, 2008) et reste toujours présente même après la correction d'erreurs systématiques connues.

L'étude des incertitudes nécessite, en métrologie, la connaissance d'une valeur de référence ou d'une valeur prise comme telle, qui se rapproche de la valeur vraie et dont l'incertitude de mesure peut être considérée comme négligeable. Il est donc nécessaire de transposer cette notion à notre domaine. Dans une partie de notre recherche, nous avons considéré que les valeurs de références peuvent être les scores des indices basés sur les macroinvertébrés ou les macrophytes qui ont fait l'objet de vérifications. L'obtention des scores des indices repose sur des procédures complexes basées sur plusieurs étapes (terrain, laboratoire, saisie des données et opération de calculs des métriques et indices) et passées entre plusieurs mains (technicien préleveur, technicien laborantin, opérateur de saisie). En comparant ces valeurs avec des valeurs obtenues après dégradation par le biais d'un modèle virtuel des indices, nous pourrions mesurer des écarts et donc des incertitudes. Les incertitudes peuvent découler de la variabilité de différents facteurs pris en considération lors d'une mesure. Il est donc important de cibler les incertitudes que nous considérons ici, à savoir l'effet de la variabilité intra-substrat, l'impact de l'opérateur réalisant les manipulations sur le terrain, au laboratoire et aux différentes étapes de production. Ces incertitudes peuvent affecter les scores des indices basés sur les macroinvertébrés et les macrophytes. Les stratégies de recherches s'y rattachant ainsi que les choix réalisés seront détaillés ci-après dans le chapitre 3.

Il est tout d'abord primordial de présenter un état des incertitudes de notre domaine déjà

étudiées et démontrées depuis les premières étapes de la mise en place des normes relatives aux indices biologiques (XP T 90-333, NF T 90-395) (AFNOR, 2003; AFNOR, 2009). En effet, elles représentent un véritable sujet de préoccupation des hydrobiologistes. Comme expliqué ci-après, elles font l'objet de programmes de recherche à différents niveaux géographiques, notamment européen, qui s'inscrivent dans le cadre de la Directive européenne Cadre sur l'Eau (Directive 2000/60/CE, 2000) mais traitent essentiellement des indices basés sur les macroinvertébrés.

Titre 2 : Les sources d'incertitudes

Nous présentons ici les incertitudes déjà traités pour les macroinvertébrés (Section 1) et pour les macrophytes (Section 2).

Section 1 : Les sources d'incertitudes des indices basés sur les macroinvertébrés

Les premières études se sont focalisées sur les incertitudes liées à l'impact des facteurs environnementaux (§1). D'autres études ont, plus particulièrement, examiné les incertitudes liées à la phase laboratoire (§2). D'autres incertitudes (§3) ont été traitées comme par exemple, l'effet de l'exclusion des taxons peu abondants ou rares. L'ensemble de ces études a permis de mettre en exergue des incertitudes existantes à chaque étape des protocoles. Néanmoins, elles ne représentent que les prémices d'une réflexion ayant vocation à évoluer de par l'amélioration des connaissances et des normes.

§1. L'impact des facteurs environnementaux

Les premières études s'intéressant aux incertitudes ont plus particulièrement porté sur les protocoles appliqués sur le terrain. Elles ont montré une incidence de la mosaïque d'habitats (1),

de l'emplacement des placettes de prélèvements et de la méthode de prélèvement (2), du volume récolté ou du nombre d'échantillons (3) sur le calcul d'un indice ou les métriques associées à cet indice.

1. La mosaïque d'habitats

Plusieurs travaux français (Beisel, et al., 1998a; Beisel, et al., 2000) ont mis en évidence l'importance de la mosaïque d'habitats grâce au suivi de plusieurs cours d'eau et à des cartographies précises des sites étudiés. En effet, il est apparu que le nombre d'espèces recensées a tendance à augmenter avec l'hétérogénéité de la mosaïque d'habitats. Cela s'explique par le fait que les déplacements d'espèces entre différents habitats est facilité. A contrario lorsqu'un site est homogène, la richesse taxonomique sera plus faible.

Ces études ont aussi démontré que les abondances des espèces varient en fonction de la disponibilité de la nourriture et que la diversité varie avec la stabilité des substrats (Beisel, et al., 1998b). Ainsi, les habitats les plus fluctuants sont les vases, les limons et les sables et les plus stables les autres substrats minéraux que le sable (des graviers aux dalles).

Enfin, il a aussi été démontré qu'à chaque type de mésohabitat (couple substrat-vitesse-hauteur d'eau) correspond une population combinant traits biologiques et stratégies écologiques spécifiques. Par exemple, les espèces sensibles à une contamination organique, plutôt racleuses de substrats, exploitent principalement les microphytes (Usseglio-Polatera, et al., 1999).

2. Le choix de la placette et de la méthode d'échantillonnage

Les auteurs ayant démontré l'importance de la mosaïque d'habitat, ont poursuivi leurs recherches en s'intéressant plus précisément à l'impact du choix de la placette d'échantillonnage (2.1) et de la méthode d'échantillonnage (2.2), ainsi que du volume de substrat récolté (2.3).

2.1. Le choix de la placette d'échantillonnage

L'application des protocoles IBGN NF T90-350 (AFNOR, 2004) et XP T 90-333 (AFNOR, 2009) sur un site défini donne la possibilité aux hydrobiologistes de choisir la placette d'échantillonnage

d'un habitat dans le respect des normes. En effet, il est souvent possible d'échantillonner un même mésohabitat à plusieurs endroits.

L'étude réalisée par les DREAL (Matte, 2005) dans le cadre du Groupe National Qualité des Eaux, a donc porté sur la subjectivité de l'opérateur dans le choix de l'emplacement de la placette et a consisté sur une même station à doubler l'échantillonnage d'un même habitat pour en étudier la variabilité.

Les résultats de cette étude ont montré qu'il existe des variations de richesse taxonomique entre les différentes placettes. Ces variations peuvent s'élever à quatre taxons pour les stations stables et atteindre un maximum de quatorze taxons pour la plus variable, tandis que le groupe indicateur reste dans l'ensemble relativement stable.

Pourtant ces résultats sont à relativiser. En effet, selon la DREAL Lorraine, certains facteurs peuvent minimiser ou maximiser cette variation comme par exemple le fait que les hydrobiologistes préleveurs n'aient pas échantillonné les placettes en limite de protocole ou encore le fait que durant cette expérimentation les stations suivies soient essentiellement de bonne qualité ce qui induit une augmentation des taxons. De plus, l'expérimentation ne repose que sur le doublement de la placette d'échantillonnage ce qui semble limiter la robustesse des résultats et ne permet pas d'identifier la réelle variabilité d'un substrat. Enfin selon la DREAL, d'autres facteurs telle que la vitesse ont vraisemblablement un impact négligeable sur la variation de la richesse taxonomique c'est pourquoi toutes les classes de vitesses n'ont pas été échantillonnées, mais cela serait à vérifier.

Néanmoins, il a été constaté dans le cadre de cette étude, que lors de la réalisation de duplicats non stricts, c'est-à-dire dans le cas où la classe de vitesse ou le substrat sont légèrement différents pour deux placettes homologues, l'écart constaté entre les placettes est bien supérieur à celui constaté entre deux placettes aux caractéristiques exactement identiques pour ces deux paramètres.

2.2. Le choix de la méthode d'échantillonnage

Tout comme le choix de la placette d'échantillonnage, la méthode de prélèvement des échantillons a suscité l'intérêt des hydrobiologistes.

La DREAL Lorraine (Mazuer, et al., 2007) a mené une étude visant à comparer deux pratiques d'échantillonnage des macroinvertébrés aquatiques en rivière, à savoir l'agitation (agité et frotté,

ou peigné avec les doigts pour les végétaux) et la récolte du support.

Les résultats de cette étude démontrent qu'il existe un réel impact du mode de prélèvement sur les listes taxonomiques obtenues. En effet, il a été constaté que lorsque seule l'agitation est prise en compte, de nombreux taxons se retrouvent absents des résultats. L'échantillonnage avec récolte du substrat apporte plus de 20% de taxons en plus. Cette étude propose donc de privilégier la récolte du support plutôt que l'agitation.

Cependant, il est à regretter que les deux techniques de prélèvements aient été réalisées sur la même placette d'échantillonnage et non pas sur deux placettes différentes, ce qui aurait permis une comparaison plus objective, mais qui aurait ajouté à la variabilité de l'échantillonnage la variabilité intra-substrat.

En frottant, une partie du substrat est emportée dans le filet avec l'aide du courant. La deuxième manipulation peut engendrer la mobilisation d'une quantité plus importante de substrat que pour une récolte sans "frottage" préalable, donc potentiellement plus de faune, tant en termes de diversité que d'abondance.

Néanmoins, le fait de froter /agiter et récolter le substrat au même emplacement peut induire un biais important, telle que l'estimation des taxons apportés en plus par la mobilisation de substrat avoisinant la placette de prélèvement lors de la récolte de l'habitat. Il pourrait être intéressant de proposer de nouvelles investigations pour annuler ce biais, comme par exemple d'augmenter le nombre de répétitions.

2.3. Le nombre d'échantillons ou le volume prélevé

En France, l'IBGN repose sur un échantillonnage des taxons à partir de huit habitats mais ces habitats peuvent varier au sein d'un même substrat à un instant t , compte-tenu de la variabilité liée à la mosaïque d'habitats. Une étude a porté sur l'analyse de la variabilité des listes faunistiques à l'échelle d'une station en fonction des différentes combinaisons de mésohabitats possibles (Beisel, et al., 1998b).

Il a été constaté que cette variabilité existe, qu'elle est directement liée aux mésohabitats et à leur combinaison lors de l'application des protocoles d'échantillonnage. A chaque mésohabitat est associé un cortège de taxons spécifiques. En fonction de la combinaison des mésohabitats réalisée sur le terrain, il peut s'opérer une compensation des taxons absents dans un mésohabitat

par ceux présents dans les autres. Ainsi, l'image du peuplement peut être complètement différente en termes de taxons présents mais constante en termes de richesse taxonomique. Les abondances des espèces trouvées dépendent des mésohabitats échantillonnés.

En ce qui concerne l'évaluation de l'IBGN, il a été montré qu'en fonction des mésohabitats échantillonnés, il pouvait y avoir d'importantes variations, de 1 à 3 points, et que celles-ci se répercutaient directement sur le diagnostic de la qualité biologique des milieux. En effet, le groupe indicateur peut facilement varier puisque son abondance est souvent faible et qu'il n'est présent que dans un nombre très limité d'habitats.

De plus, l'effort d'échantillonnage sur le terrain n'est pas équivalent pour chacune des méthodes d'échantillonnage européennes. Une étude a donc porté sur la taille de l'échantillon et son impact sur l'évaluation biologique des rivières (Vlek, et al., 2006). Cette recherche s'est effectuée dans le cadre du projet AQEM (cf : chapitre 1, Section 3, §1), qui visait à développer une évaluation biologique standardisée des eaux de surface en Europe. La méthode AQEM repose sur un échantillonnage et une identification des taxons à l'espèce, protocole très lourd en termes de temps. Vlek et al. ont donc analysé les conséquences de la réduction de la taille de l'échantillon sur le coût et l'évaluation de la qualité des rivières. Cette étude a montré que la précision et la variabilité dépendent du type d'habitat et de la métrique examinés. Les auteurs ont montré que pour l'ASPT et l'indice de saprobie par exemple, un échantillonnage d'une longueur inférieure à 5m donnait des coefficients de variation inférieurs à 10%. Pour les métriques associées au nombre d'individus (EPT, nombre de taxons), l'échantillonnage devait être réalisé sur une longueur supérieure à 5m pour obtenir de faibles coefficients de variation. La taille de l'échantillon doit donc être basée sur les habitats spécifiques présents sur la station et sur les métriques utilisées pour l'évaluation biologique. Ils proposent une stratégie sur le terrain d'échantillonnage multiple en fonction des habitats afin d'obtenir les meilleurs résultats et des coefficients de variation inférieurs à 10%.

§2. L'impact de la phase laboratoire

Suite à l'échantillonnage sur le terrain se déroule la phase laboratoire qui, elle aussi, peut comporter plusieurs points d'incertitudes. Un des premiers biais, à prendre en compte est l'effet opérateur qui induit une incertitude dans l'analyse de par son niveau d'expérience lors du tri et

de la détermination des macroinvertébrés (1.). Un deuxième biais est lié à une technique possible effectuée lors du tri : le sous échantillonnage défini ci-dessous (2.).

1. L'effet opérateur

Le processus d'analyse des échantillons au laboratoire se compose de la phase de tri puis de la phase de détermination des individus présents. Ces deux phases peuvent être réalisées par des opérateurs différents et reposent sur des protocoles d'extraction et d'identification différents en fonction des pays européens.

Haase (2006) a évalué l'impact des erreurs d'extraction et d'identification des macroinvertébrés en comparant différentes méthodes européennes de tri telles que celles proposées dans les programmes AQEM, RIVPACS ou encore MAS (méthode de tri STAR-AQEM modifiée). Ces auteurs donnent une première vision de la qualité des données produites et du degré d'erreur existant lors des opérations d'extraction et d'identification en fonction des opérateurs. Ainsi, il a été mis en lumière la double nécessité de standardiser les différentes méthodes de tri et de confier cette étape à des opérateurs expérimentés, à moins de mettre en place une démarche qualité reposant sur des audits pour suivre des novices (Haase, et al., 2006). Ceci a été confirmé par une étude australienne (Metzeling, et al., 2003) qui s'est plus particulièrement intéressée à l'impact de l'expérience de l'opérateur sur l'identification des espèces et notamment entre des opérateurs novices (étudiants) et des experts (hydrobiologistes confirmés). Ils ont constaté des oublis de taxons lors des phases de tri au laboratoire, mais également des erreurs de détermination, notamment sur les taxons les moins abondants. De plus, ils ont noté des erreurs de retranscription de la fiche laboratoire à la fiche de résultat finale. Cette étude n'a cependant pas quantifié ces erreurs relatives au degré d'inexpérience.

Pourtant, la détermination correcte des taxons est également très importante surtout lorsque les indices sont basés sur des métriques fonctionnelles, qui reposent sur les données écologiques des espèces. Pour diminuer les erreurs de détermination, il est nécessaire de mettre en place des programmes d'entraînement rigoureux et des vérifications systématiques.

2. L'effet du sous échantillonnage

Ces études se fondent sur deux façons de traiter les échantillons au laboratoire en fonction des protocoles en application. Les auteurs se sont soit intéressés à l'influence du nombre d'individus extraits et déterminés (Lorenz, et al., 2004), soit à l'influence du traitement de fractions de l'échantillon (Petkovska & Urbanic, 2010).

Les premières approches (Lorenz, et al., 2004) se sont plus particulièrement orientées vers la recherche du nombre nécessaire d'individus pour une évaluation valide de l'état des rivières étudiées. Ces études portent sur des sous échantillonnages électroniques, c'est à dire créés de façon aléatoire par une méthode statistique à partir d'un échantillon de référence. Cette technique permet de générer automatiquement différentes tailles de sous échantillons (100, 200, 300 ou encore 700 individus) à partir du même échantillon. L'objectif de cette étude était de déterminer le nombre suffisant d'individus à extraire afin d'obtenir une image valide du peuplement tout en prenant en compte les notions de temps passé et de rentabilité au niveau du coût de l'analyse des échantillons. Les auteurs ont conclu que, pour obtenir des résultats probants, au moins 300 individus doivent être extraits et déterminés. Il faut néanmoins garder à l'esprit que la variabilité et la sensibilité des paramètres sont différentes en fonction du sous-échantillonnage réalisé.

La question de l'impact du sous échantillonnage sur les métriques associées aux macroinvertébrés a été abordée par Petkovska (op.cit). Ces expérimentations portaient sur la comparaison des résultats de l'analyse de différentes portions de l'échantillon avec les résultats obtenus par l'analyse complète de l'échantillon. Elles montrent qu'en règle générale la variabilité des métriques diminue avec l'augmentation de la fraction de l'échantillon analysée ou le nombre d'individus déterminés. Cependant, ceci est à nuancer car la variabilité des métriques dépend d'une part des métriques suivies (abondance, richesse taxonomique, etc.) et d'autre part du type de rivière. Les métriques pour lesquelles la plus grande variation a été observée sont les métriques liées à la richesse spécifique et la diversité.

Cette étude a montré qu'à partir du quart de l'échantillon traité, on constate une diminution de la variabilité des métriques pour la plupart des rivières. En ce qui concerne les cours d'eau naturellement faiblement peuplés, il est suggéré de traiter une part plus importante de l'échantillon (la moitié).

Cette étude révèle que le sous échantillonnage a bien une incidence sur les résultats obtenus. Il est donc préférable de traiter entièrement un échantillon afin d'avoir un minimum

d'incertitudes, ce que préconisent les normes IBGN (AFNOR, 2004) et XP T90-388 (AFNOR, 2010) en France.

§3. D'autres sources d'incertitudes

D'autres sources d'incertitudes ont été étudiées par les hydrobiologistes comme l'impact du tri sur le vivant (Haase, et al., 2004; Nichols & Norris, 2006)(1.) ou l'effet du niveau de détermination (famille, genre, espèce) sur la qualité des résultats (Schmidt-Kloiber & Nijboer, 2004)(2.) ou bien encore la présence des taxons peu abondants ou rares (Nijboer & Schmidt-Kloiber, 2004; Clarke & Murphy, 2006)(3.).

1. Le tri sur le vivant

La variabilité qui est testée, se fonde ici entre le tri sur le vivant sur le terrain et le tri au laboratoire (Nichols & Norris, 2006). Plus précisément, ces auteurs ont comparé ces deux modalités à deux périodes différentes : printemps et automne. Ils se sont aperçus qu'il existait de grandes disparités de comptage des individus en fonction des saisons pendant lesquelles l'expérimentation a été menée. En effet, en automne les opérateurs identifient plus d'individus au laboratoire que sur le terrain, il existe en particulier des oublis de taxons de petites tailles, tandis qu'au printemps, le tri sur le vivant donne de meilleurs résultats. Ceci s'explique par le stade de croissance des larves d'invertébrés, en effet au printemps de nombreuses larves sont sur le point d'émerger donc elles sont de taille plus importante qu'en automne. Cette étude a montré que le tri au laboratoire est à privilégier lors d'évaluation biologique.

2. Le niveau d'identification des taxons

Une autre source d'incertitude ayant fait l'objet d'une étude est l'impact du niveau de détermination sur l'évaluation des classes écologiques des rivières dans le cadre de l'application du système AQEM Assessment Software (AAS) (Schmidt-Kloiber & Nijboer, 2004). Le système AQEM a été construit à partir de listes de taxons identifiés à l'espèce, la question que se sont posés

Schmidt-Kloiber et Nijboer était de savoir si l'outil AAS serait aussi performant avec un niveau de résolution taxonomique supérieur. Ils ont vérifié, dans un premier temps, si la modification du niveau de résolution taxonomique induisait des changements de classes de qualité par le biais de l'AAS, et dans un deuxième temps, si des variations observées pouvaient être importantes et si elles étaient unidirectionnelles ou non.

Ils ont constaté que l'attribution d'une classe de qualité à un site était fonction des différents niveaux taxonomiques utilisés. Dans la plupart des cas, il a été constaté le gain ou la perte d'une seule classe et occasionnellement de plus grandes déviations. Pour comprendre les causes de ces changements, ils ont analysé chaque métrique séparément et ont ainsi observé que plusieurs d'entre elles étaient directement reliées à des informations autoécologiques spécifiques et donc qu'elles ne pouvaient pas être appliquées à un niveau de détermination hiérarchique supérieur (famille), comme par exemple le pourcentage de taxons rhéophiles ou limnophiles et l'indice de saprobie. En revanche, d'autres métriques calculées à partir d'un niveau supérieur, comme par exemple la proportion des trichoptères ou l'indice de Shannon, ont conduit à montrer de plus ou moins bonnes distinctions entre les classes de qualité écologique.

Ainsi, ils ont conclu que le système AAS ne pouvait être utilisé qu'à un rang taxonomique spécifique et qu'il n'était pas possible de corriger cette erreur du fait de sa non uni-directionnalité.

3. La prise en compte des taxons rares ou peu abondants

D'autres recherches se sont plus particulièrement tournées vers l'étude de l'impact de l'exclusion de certains taxons peu abondants ou rares.

Les taxons rares ou peu abondants sont ceux qui ont des abondances faibles dans un échantillon et ceux avec de faibles distributions (Nijboer & Schmidt-Kloiber, 2004). Il a été démontré que leur exclusion entraîne une perte réelle de l'information et que les échantillons étaient mis dans des classes de qualité écologiques plus faibles. En effet, souvent les taxons peu abondants ou rares sont des indicateurs de bonne qualité écologique. Les auteurs proposent pour pallier cette incertitude de créer une métrique ne prenant en compte que ces taxons rares dans les systèmes d'évaluation de la qualité écologique des hydrosystèmes.

Clarke & Murphy (2006) ont abordé ce sujet d'une façon différente en réfléchissant à comment limiter la variabilité liée à la prise en compte des taxons rares dans le modèle de prédiction RIVPACS. Ils se sont orientés vers une recherche de l'effet de seuil de probabilité entre

valeurs observée et attendue pour chaque taxon.

Exclure les taxons ayant le moins de probabilité RIVPACS d'apparition dans les résultats diminue la variation des estimations pour les sites de référence et au sein des échantillons. Il a été également montré que lors d'un stress environnemental, les premiers taxons qui disparaissent sont ceux avec les plus faibles probabilités RIVPACS de présence sur site et les plus faibles abondances. De plus, les taxons qui ont les plus fortes probabilités RIVPACS d'être présents sur un site sont généralement les taxons les plus abondants.

Ces deux approches montrent que les effets des taxons rares ou peu abondants sont complexes puisqu'il s'agit souvent des taxons les plus polluo-sensibles et donc les plus bioindicateurs, mais que leur présence entraîne également le plus de variation dans les systèmes d'évaluation de la qualité écologique des rivières.

§4. Les incertitudes peu ou non traitées

Pour les indices concernant les macroinvertébrés, de nombreuses sources d'incertitudes ont été abordées. Il en reste toutefois encore quelques-unes qui sont encore peu étudiées et surtout quasiment jamais quantifiées notamment en France. En effet, la plupart des études précitées ne convertissent pas les incertitudes en valeurs numériques ce qui apporterait une information sur la qualité et la validité des résultats. Parmi les sources d'incertitudes peu ou pas étudiées, nous pouvons citer la variabilité intra-substrat, la variabilité saisonnière ou encore l'effet opérateur sur le tri et la détermination. Ce sont ces incertitudes que nous proposons d'étudier dans ce travail et que nous tenterons de quantifier.

Section 2 : Les sources d'incertitudes des indices macrophytiques

La DCE prévoit d'évaluer et de suivre au moins un groupe animal et un groupe végétal. Nous avons précédemment détaillé les incertitudes liées aux macroinvertébrés, nous allons donc nous intéresser aux macrophytes. Ce groupe n'a pas fait l'objet d'autant d'études que les macroinvertébrés.

Les études réalisées jusqu'à présent ont principalement concerné l'impact des facteurs environnementaux sur la distribution des macrophytes et sur les différentes métriques associées aux indices (§1). Très peu d'entre elles se sont intéressées à l'effet opérateur (§2). Nous ajouterons quelques mots sur les incertitudes peu ou non traitées (§3).

§1. L'impact des incertitudes associées aux facteurs environnementaux

Les quelques études menées sur la variabilité des macrophytes présents sur un site donné, se sont appuyées sur des expérimentations basées sur des répétitions d'échantillonnage (Staniszewski, et al., 2006). Ces auteurs ont analysé l'impact de quatre sources d'incertitudes sur les métriques du MTR (Mean Trophic Rank, présenté en Partie 1, Chapitre 1 section 2 §1) : la variation temporelle, l'influence de facteurs stressants comme l'ombrage et la dégradation hydromorphologique, et secondairement l'effet inter-opérateur. Les modifications de l'habitat, l'ombrage mais également la saisonnalité sont autant de facteurs qui jouent un rôle sur la variabilité du nombre d'espèces, l'indice de Shannon ou encore l'abondance des plantes. Cette étude a mis en évidence que le facteur le plus important serait la modification des habitats. L'influence des variations temporelles (années et saisons) et l'ombrage auraient un effet moindre sur la variation des métriques du MTR. C'est l'effet opérateur qui aurait le moins d'impact.

Le rôle des conditions hydromorphologiques et hydrologiques sur la distribution des plantes a été démontré notamment par Breugnot, et al. (2004, 2008). Ces auteurs se sont intéressés à la variation de la distribution des plantes en grands cours d'eau. Pour mener cette recherche ils ont réalisé sur le terrain des échantillonnages par le biais de la méthode point par point, des relevés de paramètres physiques (profondeur, distance depuis la berge, etc), et des analyses physico-chimiques. Ils ont ainsi montré que la distribution des macrophytes en grand cours d'eau dépend d'une combinaison de variables physiques interdépendantes et que les mécanismes de colonisation sont liés aux traits des espèces. Par exemple, dans les zones peu profondes (inférieures à 1,5m de profondeur) proches des bancs, se trouvent les taxons adaptés à de faibles contraintes hydrodynamiques et à de forts ensoleillements (*Lemna minor*, *Spirodela polyrhiza*).

Les effets de la variabilité des conditions hydromorphologiques et hydrologiques ont été testés sur les métriques des macrophytes et des indices (Szozkiewicz, et al., 2009). Il est apparu

que certaines métriques, comme celles associées au MTR, sont peu sensibles aux différentes sources de variabilité : cela confirme leur utilité dans le cadre du suivi de rivières. Les indices MI et TIM sont sensibles à certaines conditions hydromorphologiques : ainsi les métriques de la diversité spécifique sont influencées par l'ombrage, l'hydromorphologie et la longueur de la station suivie et ne sont pas utilisées dans l'évaluation de l'état écologique des rivières. Les indices testés (MTR, IBMR, TIM et MI, indice macrophytique des lacs) ont montré leur robustesse quant aux potentielles sources de variations causées par l'ombre et les modifications hydromorphologiques. Ces indices sont recommandés comme indicateur trophique et dans l'évaluation des états écologiques des rivières par l'auteur Szoszkiewicz (2009).

§2. L'impact de l'opérateur

Quelques études se sont intéressées à l'impact de l'opérateur sur le terrain mais celui-ci reste encore souvent considéré comme négligeable (Staniszewski, et al., 2006; Mascaro, et al., 2013).

Ces études ont traité l'effet opérateur en parallèle de l'étude de l'impact des facteurs environnementaux. Ce n'était donc pas leur principal objectif, elles ont pu néanmoins apporter des observations.

Ainsi, Staniszewski et al. (2006), ont comparé les scores des indices macrophytes obtenus sur 26 sites par trois opérateurs expérimentés et indépendants. Ils ont évalué la probabilité pour un site donné de changer de classe pour trois indices (indice d'Ellenberg, le MTR et l'IBMR). Ainsi, ils ont observé que l'indice d'Ellenberg était le plus variable. Le MTR et l'IBMR donnaient des résultats similaires et ne sont pas très impactés par l'effet opérateur dans le cadre de cette étude. Malgré une plus forte variabilité observée pour l'indice d'Ellenberg celle-ci n'est pas suffisamment importante pour être significative.

L'étude de Mascaro et al. (2013) s'est focalisée sur l'effet de la spatialité, de la temporalité mais également de l'effet opérateur sur l'évaluation des eaux côtières. Les auteurs se sont appuyés sur des données issues de plusieurs indices et chroniques ; ainsi ces données reflètent les trois aspects étudiés. Pour étudier l'effet de ces facteurs, ils ont utilisé les ratios de qualité écologique (EQR) et le système *WISERBUGS*. Les méthodes utilisées sont robustes face aux erreurs d'échantillonnage inter-opérateur. Ils expliquent cette faible variabilité, entre 0 et 7% au sein des masses d'eau, par le fait que les indices macrophytiques des eaux côtières reposent sur des

identifications de taxons plus faciles que pour les indices basés sur les diatomées ou les macrophytes en rivières par exemple. De plus l'utilisation d'un site d'échantillonnage fixé au préalable et d'un protocole d'échantillonnage précis aide à réduire la variabilité entre opérateurs.

§3. Les incertitudes peu ou non traitées

Nous constatons que les études basées sur les incertitudes liées aux macrophytes sont moins nombreuses que celles portant sur les macroinvertébrés. De ce fait, il existe encore de nombreuses parts d'ombre, notamment en ce qui concerne les incertitudes liées à l'application du protocole français (IBMR).

Les auteurs cités ont déjà identifié plusieurs types d'incertitudes essentiellement liées au milieu et, jusqu'à présent, l'effet opérateur a été jugé comme minime. Ils n'ont traité que l'effet inter-opérateur sur les indices, ils ne se sont pas intéressés à son impact à chaque étape du protocole d'échantillonnage (terrain et laboratoire) et à sa propagation sur l'évaluation finale. Enfin, ces études n'ont pas donné d'ordre de grandeur à ces incertitudes et n'ont pas proposé de démarches correctives.

CHAPITRE 3 Stratégies de recherche



Si l'existence de nombreux biais a d'ores et déjà été démontrée comme ceux liés aux conditions naturelles ou à l'action humaine, tous n'ont pourtant pas encore été étudiés. Notre objectif est donc de poursuivre les recherches déjà menées en explorant ces incertitudes biologiques non étudiées induites aux cours de la production d'évaluation générées par les protocoles basés sur les macroinvertébrés et les macrophytes.

Dans ce chapitre, nous justifierons nos choix des deux groupes biologiques étudiés : les macroinvertébrés et les macrophytes (Titre 1) et nous présenterons ensuite les deux axes de recherches et méthodologies associées pour ces deux groupes (Titre 2).

Titre 1 : Le choix des groupes macroinvertébrés et macrophytes et indices associés

Notre étude porte uniquement sur les incertitudes liées à l'application des protocoles d'évaluation de l'état des rivières et il nous était impossible de traiter les incertitudes associées à tous les compartiments biologiques potentiellement utilisés. Nous avons donc décidé d'étudier les incertitudes sur un groupe animal et un groupe végétal pour rester dans le cadre de la DCE.

Nous avons choisi deux groupes visibles à l'œil nu et les plus étudiés. De plus, notre objectif étant d'étudier la variabilité induite par l'échantillonnage : ces deux compartiments présentent des protocoles différents et apportent des informations complémentaires de l'état écologique des rivières. Les macroinvertébrés informent de l'état écologique global de la rivière, ainsi la communauté macrobenthique d'un site traduit les facteurs d'influence écologiques et les pressions lorsqu'elles existent. La présence ou l'absence de certaines espèces peut montrer par exemple, une mauvaise oxygénation du milieu. Les macrophytes quant à eux, permettent de déterminer l'état trophique de la rivière et donc la charge en nutriments du milieu.

Concernant le groupe animal, nous avons choisi d'étudier les incertitudes liées aux indices basés sur les macroinvertébrés. Ceux-ci sont les plus anciens et sont utilisés dans plusieurs pays européens comme indicateurs de la qualité des eaux et une large bibliographie leur est associée. Ainsi, de nombreuses recherches ont déjà été menées et ont mis en lumière des incertitudes existantes à différentes étapes de l'application des protocoles. Enfin, la facilité d'établissement de la liste faunistique et du calcul des indices ainsi que nos connaissances propres et l'expertise du bureau d'étude d'accueil ont justifié ce choix.

Concernant le groupe végétal, nous nous sommes tournés vers les macrophytes. Ils font partie, eux-aussi, des compartiments principaux. Leur utilisation en tant que bio-indicateur est plus récente et moins courante à l'échelle européenne. Ils n'ont donc pas fait l'objet d'autant d'études au sujet des incertitudes les concernant que les macroinvertébrés. Mais, le protocole d'échantillonnage de l'IBMR permet de prendre en considération les incertitudes liées à la phase terrain et à la phase laboratoire. Comme pour les macroinvertébrés, la facilité d'acquisition de listes floristiques et du calcul de l'indice ainsi que l'expertise du bureau d'étude d'accueil ont justifié ce choix.

Les trois autres groupes (oligochètes, diatomées et poissons) n'ont pas été retenus car leur

appréhension sur le terrain et au laboratoire demande des connaissances particulières tant en ce qui concerne l'application des protocoles d'échantillonnage (pour les trois indices) et de préparation des échantillons (pour les diatomées et les oligochètes) que l'identification des taxons.

Titre 2 : Les axes de recherches et les outils développés

Nous présentons ici les incertitudes choisies et les méthodes développées pour les macroinvertébrés (Section 1) et les macrophytes (Section 2).

Section 1 : Les macroinvertébrés

La Figure 1 construite à partir de l'état de l'art présenté dans le chapitre 2 représente les incertitudes pouvant exister à chaque étape de l'application des normes françaises NF T90-350, XP T90-333 et XP T90-388 (AFNOR, 2004; AFNOR, 2009; AFNOR, 2010). Certaines ont déjà pu faire l'objet de recherche et d'autres ont encore été très peu, voire non étudiées.

La Figure 1 présente les principales étapes de mises en œuvre des protocoles, les différentes sources de variabilité identifiées, les résultats intermédiaires et finaux, les incertitudes s'y propageant et, enfin, sont mises en valeur les sources de variabilité et leurs incertitudes associées que nous traiterons dans cette étude.

Rappelons que l'effet de la variabilité naturelle par l'étude de l'impact de la mosaïque d'habitats a été démontré ($\delta 1$, Figure 1) (Beisel, et al., 1998a; Barnes, et al., 2013), l'évaluation de l'incertitude liée à l'application de la norme XP T90-333 est toujours en cours ($\delta 3$, Figure 1) (Archambault, et al., 2011) et l'effet opérateur a déjà été étudié dans l'application de protocoles internationaux (Metzeling, et al., 2003). Dans ce contexte, nous avons décidé de nous intéresser : 1) à la variabilité naturelle liée à l'hétérogénéité des substrats eux-mêmes ($\delta 2$, Figure 1), 2) à l'effet opérateur sur les indices français (IBG DCE et I2M2) lors des étapes de tri et d'identification des taxons au laboratoire ($\delta 4$, $\delta 4'$ et $\delta 4''$, Figure 1). Nous ne traiterons pas ici les incertitudes liées à

la retranscription des résultats (δ_5) ni celles liées aux choix des métriques et des indices (δ_6), car ces incertitudes sont difficilement appréciables et nous ne discuterons pas ici du choix des indices. Cependant, nous mènerons une réflexion sur la propagation des incertitudes choisies ($I=f(\delta_2 ; \delta_4)$) à chaque étape du processus d'évaluation jusqu'à l'obtention des indices et de l'état écologique ainsi que sur le comment limiter ces incertitudes.

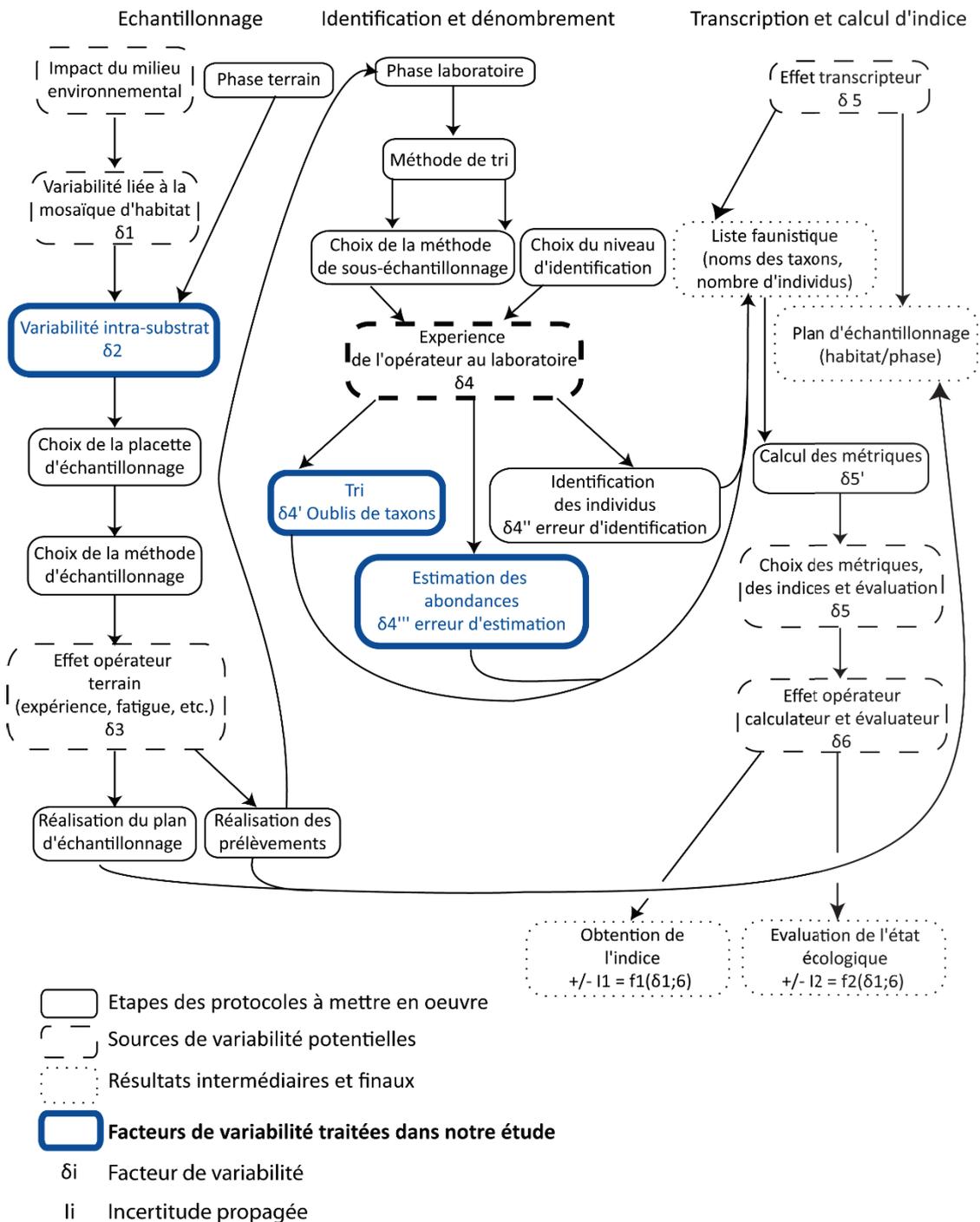


Figure 1: Facteurs de variabilité et différentes étapes de mise en œuvre des protocoles d'évaluation des milieux aquatiques basés sur les macroinvertébrés

Pour étudier les deux facteurs de variabilités sélectionnés (δ_2 et δ_4 , Figure 1) et les incertitudes engendrées, nous avons opté pour trois démarches différentes. Les deux premières reposent sur des expérimentations réelles avec la mise en place d'un protocole d'échantillonnage sur le terrain. Nous avons entièrement réalisé la première démarche qui nous a permis d'étudier la variabilité intra-substrat. Pour la seconde, nous avons utilisé les données produites dans le cadre d'essais inter-laboratoires que l'Association Générale des Laboratoires d'Analyse de l'Environnement (AGLAE) a bien voulu mettre à notre disposition. La troisième se base sur un lot d'expérimentations virtuelles se basant sur des jeux de données existant au sein d'Aquabio et sur le développement d'un algorithme spécifique. Les deux dernières approches nous ont permis d'étudier les effets cumulés des opérateurs mettant en œuvre les protocoles et de la variabilité intra-substrat.

§1. La variabilité intra-substrat

L'incertitude liée à la variabilité intra-substrat se base sur une démarche expérimentale réalisée dans le cadre de l'application des normes XP T90-333 et XP T90-390 (AFNOR, 2009; AFNOR, 2010). L'expérimentation repose sur un protocole d'échantillonnage sur le terrain des différents substrats présents. Nous avons privilégié cette démarche plutôt que l'utilisation de données déjà existantes car elle nous permettait d'avoir une incertitude constante associée à l'opérateur unique, et de nous affranchir de celui lié à la saisonnalité.

§2. L'effet opérateur cumulé à la variabilité intra-substrat

L'étude de l'effet de l'opérateur au laboratoire repose, quant à elle, sur deux types de données. Tout d'abord, nous disposons de données produites par AGLAE lors d'un essai inter-laboratoires. Le but de cette expérimentation est d'étudier la variabilité inter-opérateurs cumulée à la variabilité intra-substrat et d'observer les écarts de notes obtenus pour les indices IBGN et IBG DCE.

§3. L'effet opérateur et des pratiques laboratoires

Nous avons également développé un algorithme spécifique basé sur des modifications virtuelles de listes faunistiques afin de mesurer l'effet opérateur au laboratoire pendant les opérations de tri et d'identification. Cette expérimentation repose sur l'étude de l'impact de la prise en compte des taxons peu abondants ou rares et les erreurs d'estimations des abondances des individus sur l'I2M2 et l'IBG DCE.

Section 2 : Les macrophytes

Certaines incertitudes associées à l'application des protocoles liés aux macrophytes ont déjà été étudiées. Cependant, elles concernent principalement l'impact des facteurs environnementaux ($\delta 1$ et $\delta 2$, Figure 2) comme nous avons pu le montrer précédemment (Chapitre 2, section 3) et ne portent que très rarement sur l'effet opérateur ($\delta 4$, Figure 2). Nous allons donc nous intéresser essentiellement aux incertitudes liées aux actions de l'opérateur sur le terrain et au laboratoire ($\delta 4$ et $\delta 5$, Figure 2). Le schéma ci-dessous présente les incertitudes déjà étudiées et celles qui nous intéressent (Figure 2). Il est basé sur l'état de l'art présenté dans le chapitre 2 et représente les incertitudes pouvant exister à chaque étape de l'application de la norme française NF T90-395 (AFNOR, 2003), les différentes sources de variabilité identifiées, les résultats intermédiaires et finaux et les incertitudes s'y propageant. Y sont mises en valeur les sources de variabilité et leurs incertitudes associées que nous traiterons dans cette étude.

Nous ne nous intéresserons pas ici à l'incertitude liée à la retranscription des résultats et au choix des métriques et indices. Cependant, nous mènerons une réflexion sur la propagation des incertitudes engendrées par les variabilités choisies ($\delta 4$ et $\delta 5$, Figure 2) à chaque étape du processus d'évaluation jusqu'à l'obtention de l'indice IBMR et de l'état trophique ainsi des propositions pour limiter ces incertitudes.

Le protocole d'échantillonnage des macrophytes selon la norme XP T90-395 (AFNOR, 2003) demande plus de temps de mise en œuvre que le protocole d'échantillonnage des macroinvertébrés sur le terrain, de plus l'identification des taxons se fait à l'espèce occasionnant des erreurs de déterminations dans des genres difficiles (cas des Callitriches). De ce fait, pour

étudier la variabilité liée à l'opérateur lors de l'application du protocole d'échantillonnage nous avons décidé d'opter pour une démarche virtuelle basée sur le développement d'un algorithme spécifique permettant la modification des listes floristiques existantes à Aquabio fondée sur les confusions d'espèces et les erreurs d'estimations d'abondances.

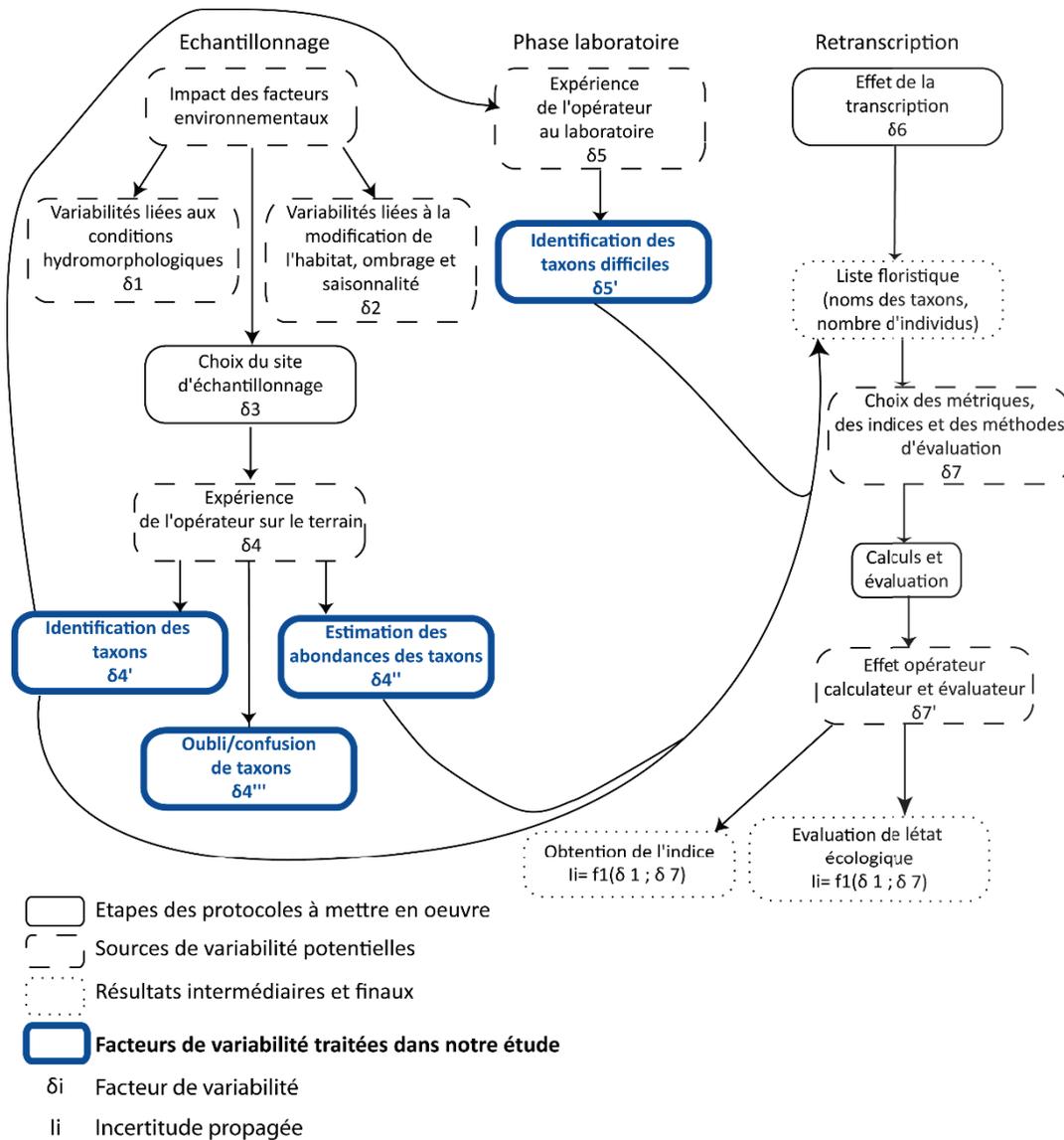


Figure 2: Facteurs de variabilité et différentes étapes de mise en œuvre des protocoles d'évaluation des milieux aquatiques basés sur les macrophytes

PARTIE 2 Méthodes et contributions

Cette partie est consacrée à la réalisation des différentes approches expérimentales et virtuelles utilisées pour l'étude des variabilités naturelles et opérateur pour les deux compartiments biologiques retenus, leurs résultats et les différents éléments de solutions correctives que nous pouvons proposer.

Les deux premiers chapitres concernent uniquement les stratégies de recherche construites pour l'étude des incertitudes associées aux macroinvertébrés. Le premier chapitre se partage en deux volets : le premier porte sur la démarche terrain élaborée pour étudier la variabilité intra-substrat, le second sur les variabilités intra-substrat et opérateur par le biais des données fournies par l'Association Générale des Laboratoires d'Analyse de l'Environnement (AGLAE). Le deuxième chapitre présente les résultats obtenus grâce à la démarche virtuelle élaborée pour estimer l'impact de l'expérience de l'opérateur.

Le dernier chapitre porte sur l'étude des incertitudes associées à l'indice IBMR et sur la démarche virtuelle mise en place pour mesurer l'effet opérateur.

CHAPITRE 1 Méthodes expérimentales de mesure des incertitudes associées aux macroinvertébrés

Ce chapitre présente les différents protocoles expérimentaux mis en place pour étudier les incertitudes de mesure liées à la variabilité naturelle intra-substrat à l'application des protocoles d'échantillonnage des macroinvertébrés.

La première partie de ce chapitre est rédigée sous la forme d'un article en cours de soumission (Hydrobiologia) et présente les principaux résultats obtenus par notre démarche expérimentale de mesure de la variabilité intra-substrat (Titre1).

La deuxième partie traite des variabilités naturelle et opérateur mesurées sur les essais inter-laboratoires menés par AGLAE.

Titre 1 : Article: Experimental study of the uncertainty of the intrasubstrate variability on two French index metrics based on macroinvertebrates : Résumé

De nombreuses recherches portant sur les macroinvertébrés ont mis en lumière l'impact de la mosaïque d'habitats sur la distribution des macroinvertébrés, du type de substrat et de l'hétérogénéité au sein d'un substrat, comme sources d'incertitudes.

Nous avons décidé d'étudier l'effet des incertitudes liées à l'impact du choix de la placette d'échantillonnage sur les métriques de l'IBGN et de l'I2M2 (deux indices biologiques français) en utilisant une démarche expérimentale sur le terrain, basée sur l'échantillonnage répété de substrats. Neuf types de substrats ont été étudiés, sur 31 sites, dans sept hydro écorégions (HER). Dix répétitions par substrat et par site ont été collectées, 305 listes faunistiques associées à 305 répétitions de substrat ont été obtenues. Douze métriques, telles que l'indice de Shannon et la classe de variété, ont été calculées par liste. Nous avons utilisé une classification des données basée sur la méthode des plus proches voisins, un positionnement multi dimensionnel et des indices de similarité pour analyser l'effet de la variabilité du substrat.

Les résultats obtenus nous ont permis d'observer que les différentes échelles spatiales imbriquées génèrent une variation : il existe une relation entre les variations observées sur les différentes HER, celles observées intersites, mais aussi celles sur le même site. Les blocs et les graviers sont les substrats les moins variables, tandis que les racines et les sables sont les plus variables. La variabilité des substrats affecte différemment les métriques des indices. Les métriques de l'I2M2, calculées sur chaque répétition, sont plus robustes que les métriques de l'IBGN, et utiliser les métriques de l'I2M2 gomme, pour partie, la variabilité intra-substrat.

Les estimations des abondances des taxons et la diversité sont deux paramètres qui ont une incidence sur les métriques calculées sur les substrats. Ils sont liés aux méthodes utilisées en laboratoire et aux compétences de l'opérateur. Une approche de contrôle de la qualité et des processus normalisés pourraient être développés pour améliorer le tri et les estimations des abondances des taxons.

Section 1: Introduction

The Water Framework Directive (WFD) established a consistent water policy within the European Union (Directive 2000/60/CE, 2000). It introduced the principle of protection and management of surface, transitional, coastal and ground waters.

The use of standardized biological indices is therefore needed in monitoring programs to assess the ecological quality of freshwaters. Furthermore, the WFD requires reliability, reproducibility and comparability of ecological data produced by European countries. Whereas high variability in metrics associated with biological indices has been reported (Springe, et al., 2006; Johnson, et al., 2006). WFD highlighted the need for estimating the confidence and precision level in monitoring programs. The uncertainty measurement concept was thus developed.

Since the 1970s, macrobenthos aquatic system assessment has constantly attracted the European scientific community's interest. In recent decades, protocols have been continuing to evolve to generate valid information and fulfil WFD requirements.

Uncertainty measurements associated with macroinvertebrate protocols has become a real concern for hydrobiologists and this issue has been dealt with in several studies. Some of them were based on the study of the influence of habitat on macrobenthic population distribution (Beisel, et al., 1998a; Beisel, et al., 2000), substrate heterogeneity in the streambed (Boyero, 2003; Barnes, et al., 2013), and the relationship between the mesohabitat type and macrobenthic community by combining biological traits and specific ecological strategies (Usseglio-Polatera, et al., 1999). Other uncertainties were linked with the sampling phase (Mazuer, et al., 2007; Matte, 2005) of the French standardized global biological index method (IBGN, Indice Biologique Global Normalisé), NF T90-350, and XP T 90-333 standards (AFNOR, 2006; AFNOR, 2009). Likewise, few studies have focused on the surveyor's skills and on the different sampling methods used (Haase, et al., 2006; Metzeling, et al., 2003; Lorenz, et al., 2004; Lorenz & Clarke, 2006; Petkovska & Urbanic, 2010). The efficiency of the live-sorting has thus been tested on metrics (Nichols & Norris, 2006), taxonomic resolution (family, genus, species) on the quality of results (Schmidt-Kloiber & Nijboer, 2004), and on the impact of rare taxa on ecological assessment (Nijboer & Schmidt-Kloiber, 2004). All of these studies helped to highlight uncertainties at each protocol step. Research on the habitat mosaic (Beisel, et al., 2000) and the impact of habitat heterogeneity (Barnes, et al., 2013; Boyero, 2003; Erman & Erman, 1984) revealed the importance of the substrate as an uncertainty source. However, the impact of intrasubstrate variations in the index

calculation has not been yet studied.

The present research was focused on uncertainties in the calculation of the IBGN (AFNOR, 2006; AFNOR, 2009) and the new French multimetric index I2M2 (Mondy, et al., 2012), and more specifically on intrasubstrate variability in a same habitat mosaic. The sampling process was based on strict rules established by standards (AFNOR, 2004; AFNOR, 2009). However, the choice of sampling plots at a monitoring site could be very diverse. The aim of this study was to estimate uncertainties due to intrasubstrate variability on the metrics of French macroinvertebrate indices (IBGN, I2M2). We used an experimental approach based on substrate replicate sampling in French rivers. We measured the variation in metrics according to each replicate.

Section 2: Materials and methods

The sampling protocol was established in different sizes of river (small, medium and large). Sampling sites were selected in: (1) homogenous geographical areas, and (2) when they had a very good ecological status. Furthermore, (3) the site had to offer enough of substrate area to allow substrate sample replication, while being easily accessible.

§1. Study area

We sampled thirty one reference sites mainly located in central and western areas of France (Figure 3). These areas included seven hydroecoregions (HER) (as defined by Wasson, et al., 2004): “Massif Central Nord”, “Massif Central Sud” (north and south of the Massif Central), “Massif Armoricaïn” (Armorican Massif, Britain, west of France), “Coteaux Aquitains” (Aquitaine slopes, near Toulouse), “Causses Aquitains”, (Aquitaine Causses, Dordogne), “Landes” (southwestern of France), and “Pyrénées” (Pyrenean region).

Field surveys were carried out in summer 2012.

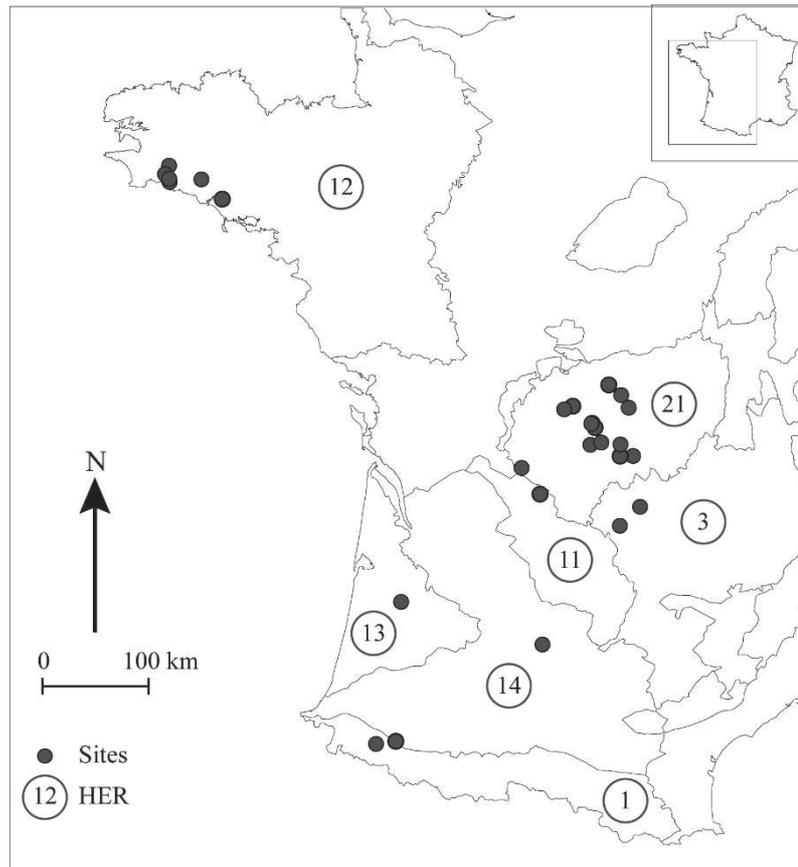


Figure 3: Location of the 31 sites in western and southwestern France, chosen for the experimentation (each point corresponds to a studied site and numbers correspond to hydrocoregions (as defined by Wasson, et al., 2004)

§2. Site selection design

The sites were selected: (1) depending on a highly pollution sensitive indicator group (GI) from the IBGN table, such as Perlodidae, Chloroperlidae or Perlidae, to ensure the high ecological quality of the sites, and (2) with a percentage substrate cover of over 10% to ensure the replicate sampling. GI and cover percentages were from the Aquabio database (Aquabio is a consultant agency specialized in hydrobiological data expertise) and were measured and recorded during river monitoring by experienced surveyors between 2007 and 2011.

Moreover, for each substrate, we selected three sites with different wetted widths to obtain a robust dataset and to not limit the study to a single stream type.

The Armorican Massif was a special case. This region is characterized by high taxonomic diversity despite the strong anthropogenic pressure. We thus decided to sample three

supplementary sites with an organic substrate and a mineral substrate in this region.

Based on these rules, we sampled 9 substrates, i.e. block, bryophyte, submerged spermaphyte (only Callitriche), pebble and stone, gravel, litter, root, hard bottom and sand substrates (Table 1). 75% of the I2M2 substrate types were sampled (12 substrates in the XP T 90-333 standard, AFNOR 2009, used for the I2M2 sampling grid).

To perform replicates, we sampled each substrate (1) in its representative current velocity class, e.g. litter and roots were sampled in a current velocity range of under 5 cm/sec, (2) at the same water depth, and (3) within the same habitat mosaic (Table 1). These three conditions were fulfilled in the field through environmental monitoring and whenever possible by sampling the same substrate patch.

During the sampling surveys, we had to deal with hazards linked to changing habitat mosaics in comparison to the database information. For example, some substrates (litter, root and hard bottom substrates) were not present on site or not in sufficient quantity to perform 10 replicates.

Table 1: Sampled substrate list with associated velocity range, location (HER) and the number of sampled sites (values in brackets correspond to the hydro ecoregion codes; see Figure 3)

Substrates	Speed range	Hydroecoregion (HER)	Number of sampled sites
Blocks	$25 \leq S < 75$	Massif Central Nord (21), Massif Central Sud (3)	3
Bryophytes	$25 \leq S < 75$	Massif Central Nord (21)	3
Submerged spermaphytes	$25 \leq S < 75$	Causses Aquitains (11), Massif Armoricaïn (12), Massif Central Nord (21)	5
Pebbles	$25 \leq S < 75$	Coteaux Aquitains (14), Causses Aquitains (11), Pyrénées (1)	3
Gravels	$5 \leq S < 25$	Massif Armoricaïn (12), Massif Central Nord (21)	4
Litters	$S < 5$	Massif Armoricaïn (12), Massif Central Nord (21)	2
Stones	$5 \leq S < 25$	Massif Central Nord (21)	3
Roots	$S < 5$	Massif Central Nord (21), Landes (13)	3
Hard bottoms	$25 \leq S < 75$	Massif Central Nord (21)	2
Sands	$S < 5$	Massif Central Nord (21), Massif Central Sud (3)	3

§3. Sampling design

The protocol was based on the sampling of 10 substrate replicates by the same surveyor using a Surber Sampler (20 X 25 cm frame, 500 μ m mesh). We followed the sampling rules for each substrate, in accordance with the XP T 90-333 standard (AFNOR, 2009) (Figure 4).

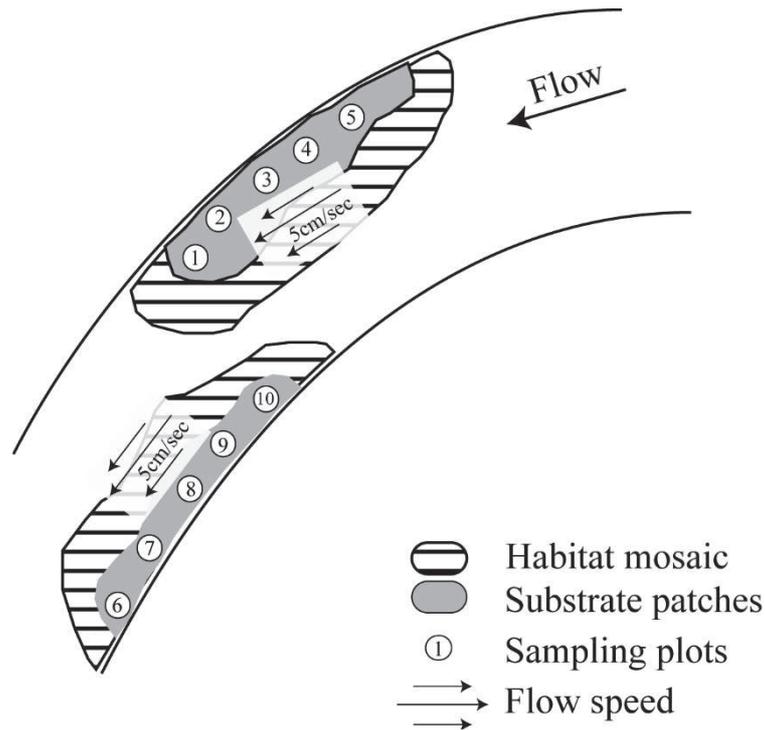


Figure 4: Sampling design used for the experimentation

§4. Sorting phase

All samples were identified in accordance with the XP T 90-388 standard (AFNOR, 2010). Sorting phase and taxa identification were performed by a single surveyor. Identifications were done at the family or genus level based on the required level mentioned in the standard.

§5. Data analysis

1. Metrics calculation

We calculated three IBGN (AFNOR, 2006) metrics obtained from each sampled plot and each substrate based on the IBGN table. Firstly, the group indicator class (GI) corresponded to the indicator taxa represented by at least 3 individuals or 10 based on the group. GI ranged from 1, for pollution tolerant taxa, to 9 for pollution sensitive taxa (AFNOR, 2006). Secondly, the

taxonomic richness (VT) was equal to the taxon number picked up and identified at the family level, and thirdly the IBGN variety class (VC) depending on the VT.

We calculated the five I2M2 metrics (Mondy, et al., 2012) for each replicate substrate.

The metrics calculation was adapted to our dataset but this did not allow us to calculate the IBGN. The I2M2 score could not be calculated because the metrics calculation was performed on a single substrate and not on complete a sampling. The values obtained thus provided an indication but not an ecological assessment.

We identified the taxonomic richness (TAX) of the Mulimetric Macroinvertebrate Index Flanders (MMIF) corresponding to the present taxa number (Gabriels, et al., 2010), the original average score per taxon (ASPT), which was the average of the tolerance scores of all macroinvertebrate families found, and ranged from 0 to 10.

The Shannon diversity Index H' corresponded to the specific diversity index and was expressed with the following formula:

$$H' = - \sum_{i=1}^n (P_i \times \ln P_i) \quad (8)$$

P_i corresponds to the relative abundance of each taxon i , n is the number of taxa.

The relative abundance of the ovoviviparous species OS corresponded to the sum of products of the coefficients C_i assigned to each taxon i based on its affinity to the biological trait OS and the taxon frequency P_i .

$$OS = \sum_{i=1}^n C_i \times P_i \quad (9)$$

The relative abundance of polyvoltine species PS corresponded to the sum of the products of the coefficients C_i assigned to each taxon i based on its affinity to the biological trait PS and the taxon frequency P_i .

$$PS = \sum_{i=1}^n C_i \times P_i \quad (10)$$

We used four supplementary metrics: the taxonomic richness at the genus level (GV), the number of Ephemeroptera, Plecoptera, and Trichoptera (EPT), the Simpson index (SI) and the Equitability (E) corresponding to the two following formulae:

$$SI = 1 - H \quad (11)$$

$$E = \frac{H}{\ln(S)} \quad (12)$$

Where S corresponded to the taxonomic richness, and H to the Shannon index (see Formula 1).

2. Distance analysis

The variation coefficient CV_r is the ratio between the standard deviation σ_r and the average of the metric values m_u , for n samples, for a river r , with very good trophic status. It is expressed in percentage.

$$CV_r = \frac{\sigma_r}{\frac{1}{n} \sum_{u=1}^n m_u} \times 100 \quad (13)$$

The CV values can be interpreted as follows (Gendron & Martin, 2004): when the CV is less than 16% the variation is low and the average estimation is reliable; between 16% and 33.3%, the variation is high and average could contain errors, and when CV is over 33.3% the variation is high and the average is not reliable.

3. Cluster analysis

Single linkage agglomerative clustering, also called nearest neighbour sorting with Euclidean distance, was used to agglomerate replicates on the basis of their shortest pairwise distance or greatest similarities, implemented within R Software (Legendre & Legendre, 1998; Borcard, et al., 2011). This approach allowed us to easily compare replicates and how samples agglomerated with each other. The cluster analysis was based on faunal lists and metric matrices, so we were able to

measure similarities between substrate replicates within each site in terms of taxa presence and metric values.

We applied this analysis on all replicates i and sites r , for each substrate s . We implemented this statistical approach on taxonomic lists and on I2M2 metrics, and we obtained two clusters per substrate. Then we built heatmaps producing color images associated with dendrograms (black very close, white completely different).

4. Similarity indices

Three similarity indices were calculated from the macroinvertebrate replicates: (1) the Jaccard index, (2) the Sorensen index, and (3) the Renkonen index (Jaccard, 1901; Sørensen, 1948; Renkonen, 1938). These indices are generally used in ecology to compare populations of sample pairs.

Jaccard (J_s) and Sorensen (S_s) indices based on the presence/absence of taxa in a sample pair, are calculated with the following formulas:

$$J_s = \frac{a}{a + b + c} \quad (14)$$

$$S_s = \frac{2a}{2a + b + c} \quad (15)$$

Where a corresponds to the number of species shared by samples, b corresponds to the number of species belonging only to the first sample, and c is the number of species belonging only to the second sample.

The Renkonen index (R_s) accounted for relative species abundance and was calculated as

$$R_s = \sum_{i=1}^n \text{minimum}(P_i^1, P_i^2) \quad (16)$$

Where P_i^1 is the frequency of species i in sample 1 and P_i^2 is the frequency of species i in sample 2.

5. Multidimensional scaling

To observe how samples are distant in a 2D space, we used multidimensional scaling analysis (MDS) based on the Sorensen distance matrix D_m , and implemented within R software (Cox & Cox, 2001).

We calculated S_s on each pair of faunistic lists per substrate and site. We obtained a similarity matrix per substrate and transformed this matrix in a distance matrix D_m . To calculate the distance d between each replicate, we used the following formula:

$$d = 1 - S_s \quad (17)$$

Then we obtained an MDS projection for each substrate. Each point corresponded to a replicate for a given substrate at a given site.

Table 2 is an acronym summary of the metrics used in this paper.

Table 2: Acronym summary of the metrics used in this paper

Acronyms	Metrics
GI	Indicator group class
VT	Taxonomic richness
VC	Variety class
TAX	MMIF taxonomic variety
ASPT	Average score per taxon
H	Shannon index
OS	Relative abundance of ovoviviparous species
PS	Relative abundance of polyvoltine species
GV	Genus richness
EPT	Number of Ephemeroptera, Plecoptera and Trichoptera
SI	Simpson index
E	Equitability
CV	Variation coefficient
Js	Jaccard index
Ss	Sørensen index
Rs	Renkonen index

Section 3: Results

§1. Substrate variability and metrics robustness

Table 3 and Table 4 gave the average standard deviations (SD) and variation coefficients (CV) for IBGN and I2M2 metrics (Table 3) and supplementary metrics (Table 4) for each substrate.

Table 3: Average standard deviation (SD) and variation coefficient (CV) of IBGN and I2M2 metrics for each substrate and at all sites (in light blue $16\% < CV < 33.3\%$, in light blue and bold $CV > 33.3\%$) $CV < 16\%$ = not significant, $16\% < CV < 33.3\%$ = significant, and $CV > 33.3\%$ = critical error. (Example CV average of blocks = (blocks CV site 1 + blocks CV site 2 + blocks CV site 3/ 3)).

Metrics Substrates	GI		VT		VC		ASPT		TAX		H		PS		OS	
	SD	CV	SD	CV	SD	CV	SD	CV	SD	CV	SD	CV	SD	CV	SD	CV
Blocks	0.54	6.87	2.82	13.63	0.76	11.57	0.45	6.43	2.81	13.80	0.20	12.33	0.29	44.77	0.04	70.43
Bryophytes	0.38	5.07	3.83	22.20	1.03	18.10	0.57	8.20	3.69	21.77	0.25	16.08	0.34	48.60	0.15	61.17
Submerged spermaphytes	1.60	34.10	1.93	17.94	0.62	15.40	0.38	6.80	2.01	19.36	0.32	24.02	1.10	30.79	0.21	87.10
Pebbles	0.58	8.57	2.78	18.60	0.81	15.97	0.45	6.53	2.79	22.53	0.21	10.07	0.26	30.17	0.02	42.57
Gravels	1.26	20.58	2.03	14.93	0.61	12.63	0.52	8.05	2.02	14.98	0.16	8.50	0.17	29.28	0.09	50.75
Litters	1.14	33.35	1.97	17.55	0.69	17.55	0.46	8.85	1.85	16.90	0.14	8.70	0.21	32.85	0.10	50.75
Stones	0.22	2.77	3.06	18.37	0.80	14.33	0.28	4.07	3.38	19.60	0.21	9.10	0.29	41.40	0.02	64.13
Roots	1.24	19.23	3.49	18.27	0.95	15.47	0.41	6.40	3.46	19.07	0.29	21.80	0.30	48.00	0.08	50.03
Sands	1.59	46.40	2.55	25.73	0.85	23.30	0.88	15.87	2.41	26.10	0.31	19.37	0.48	43.03	0.08	57.60
Hard bottoms	2.00	35.30	3.49	25.45	1.08	23.00	0.51	8.00	3.49	25.45	0.25	12.00	0.32	56.35	0.04	87.90

Table 4: Average of standard deviation (SD) and variation coefficient (CV) of supplementary metrics for each substrate. $CV < 16\%$ = not significant, $16\% < CV < 33.3\%$ = significant, and $CV > 33.3\%$ = critical error. (in light blue $16\% \leq CV < 33.3\%$, in light blue and bold $CV > 33.3\%$; example: CV average of blocks =(blocks CV site 1+blocks CV site 2+blocks CV site 3/3))

Metrics Substrates	GV		EPT		SI		E	
	SD	CV	SD	CV	SD	CV	SD	CV
Blocks	3.08	12.93	2.05	18.24	0.04	5.23	0.07	13.34
Bryophytes	3.97	21.46	2.42	25.63	0.10	15.84	0.08	15.03
Submerged spermaphytes	2.11	18.76	1.55	31.90	0.13	22.69	0.11	19.83
Pebbles	3.24	17.78	2.15	24.87	0.05	5.95	0.06	8.33
Gravels	2.15	13.46	1.68	25.45	0.05	6.91	0.06	8.27
Litters	2.19	18.08	1.00	35.26	0.04	5.63	0.05	7.75
Stones	3.99	20.59	2.14	23.36	0.04	5.10	0.06	7.04
Roots	4.32	20.87	2.11	27.39	0.12	23.60	0.09	21.64
Sands	2.87	26.62	1.40	47.92	0.10	14.93	0.09	13.53
Hard bottoms	3.84	26.38	1.72	25.60	0.06	7.27	0.07	9.69

The intrasubstrate variability depended on the substrate type. We observed small metrics variability on the mineral substrates, i.e. block, pebble, stone and gravel substrates. For example, on blocks, only two metrics had a high CV (over 33.3%): the relative abundance of polyvoltine species (PS) was 44.77% and the relative abundance of ovoviviparous species (OS) was 70.43% (see Table 3). Sand and hard bottom substrates were highly variable, excepted for ASPT in both substrates and for the Shannon index (H) in sand substrates.

In contrast, all organic substrates (bryophyte, submerged spermaphyte, litter and root substrates) were more variable. For example, 6 of the 8 metrics associated with the submerged spermaphyte replicates had a high CV (above 16%).

Some metrics, independently of the substrate, had a similar variability. For example, the two metrics associated with the biological traits (PS and OS) presented the highest variability (CV >33.3%) and the lowest one concerned the ASPT (CV <16%) regardless of the substrate (Table 3). Figure 5 showed that some substrates were usually above the threshold of 16%, e.g. sand or hard bottom substrates.

Considering the four additional metrics (Table 4), we observed that the mineral substrates, i.e. block, pebble, gravel and stone substrates, were generally less variable than all organic substrates, and the two mineral substrates, i.e. sand and hard bottom.

The substrates responded differently depending on the replicate sampling, but also on the metrics.

Metrics associated with taxonomic richness (VT, TAX, GV and EPT) had a high CV and a SD ranging from 1.93 to 3.83 for family richness. Concerning the IBGN index, family richness variation (VT) did not necessarily lead to a modification in richness class (VC). For example, the SD of litter VT was equal to plus or minus 3.06 taxa, while the SD of VC was less than 1, so the class change was not systematic.

Metrics linked to the biological traits (polyvoltinism and ovoviviparity) showed high CV variability even with small deviations ranging from 0.17 to 1.10 for polyvoltinism and 0.02 to 0.20 for ovoviviparity, but the SD depended on the metrics amplitude (for polyvoltinism and ovoviviparity the metrics amplitude ranged from 0 to 1).

Metrics associated with the macroinvertebrate diversity (H, SI and E) were not very sensitive to the intrasubstrate variability. The ASPT did not show significant variability. Depending on the substrate, the least variable metrics were ASPT, Shannon diversity (H), richness class (VC), Simpson

index (SI) and Equitability (E).

These observations had to be moderated because the results were derived from an average coefficient of variation between different sites, which should be kept in mind regarding the intersite differences. Indeed, when considering the CV for all the metrics for each site, we observed that for some substrates a site seems to be an outlier. For example, the individual CVs (Table 5) of blocks which were sampled at 3 sites showed that for one site the CV was lower than at the two other sites. This difference led to a decrease in the average CV. In contrast, for sand replicates at 3 sites, one of them had a very high CV that led to an increase in the CV average.

Table 5: Standard deviation (SD) and variation coefficient (CV) for each site and for each metric for block and sand substrates (in light blue $16\% < CV < 33.3\%$, in light blue and bold $CV > 33.3\%$) $CV < 16\%$ = not significant, $16\% < CV < 33.3\%$ = significant, and $CV > 33.3\%$ = critical error.

Substrates Metrics	SD		CV		Blocks		SD		CV		Sands		SD		CV	
	SD	CV	SD	CV	SD	CV	SD	CV	SD	CV	SD	CV	SD	CV		
GI	0.57	8	0.63	7.2	0.42	5.41	1.84	34.03	0.82	10.21	2.10	95.35				
VT	4	13.66	1.5	5.82	3.75	19.32	1.76	12.6	1.96	11.2	4.9	56.06				
GV	3.98	16.3	1.5	7.1	3	17.53	1.35	10.89	1.69	11.40	4.60	54.76				
VC	1.07	14.53	0.52	7.82	0.70	12.48	0.52	11.74	0.42	8.11	1.62	50.6				
ASPT	0.60	8.40	0.32	4.76	0.42	6.10	0.59	10.29	0.54	7.99	1.5	29.26				
TAX	4.16	17.18	1.44	6.87	2.84	17.29	1.40	11.85	1.57	10.96	4.27	55.45				
EPT	2.63	23.36	1.26	11.93	2.27	21.42	1.62	33.74	0.74	12.51	1.85	97.52				
H	0.25	20.05	0.15	7.91	0.20	9.09	0.16	8.85	0.25	15.80	0.52	33.48				
PS	0.36	51.88	0.12	27.05	0.40	55.5	0.33	36.25	0.33	30.62	0.77	62.31				
OS	0.01	64.01	0.10	76.64	0.02	70.75	0.13	57.30	0.04	43.94	0.08	71.46				
SI	0.05	5.33	0.04	4.76	0.05	5.60	0.05	6.97	0.10	16.78	0.14	21.03				
E	0.08	21.60	0.05	8.50	0.07	9.91	0.06	8.25	0.08	13.74	0.15	18.6				

Despite these intersite differences, the mineral substrates were the most stable, except for sand and hard bottom substrates, and the organic substrates were the most variable.

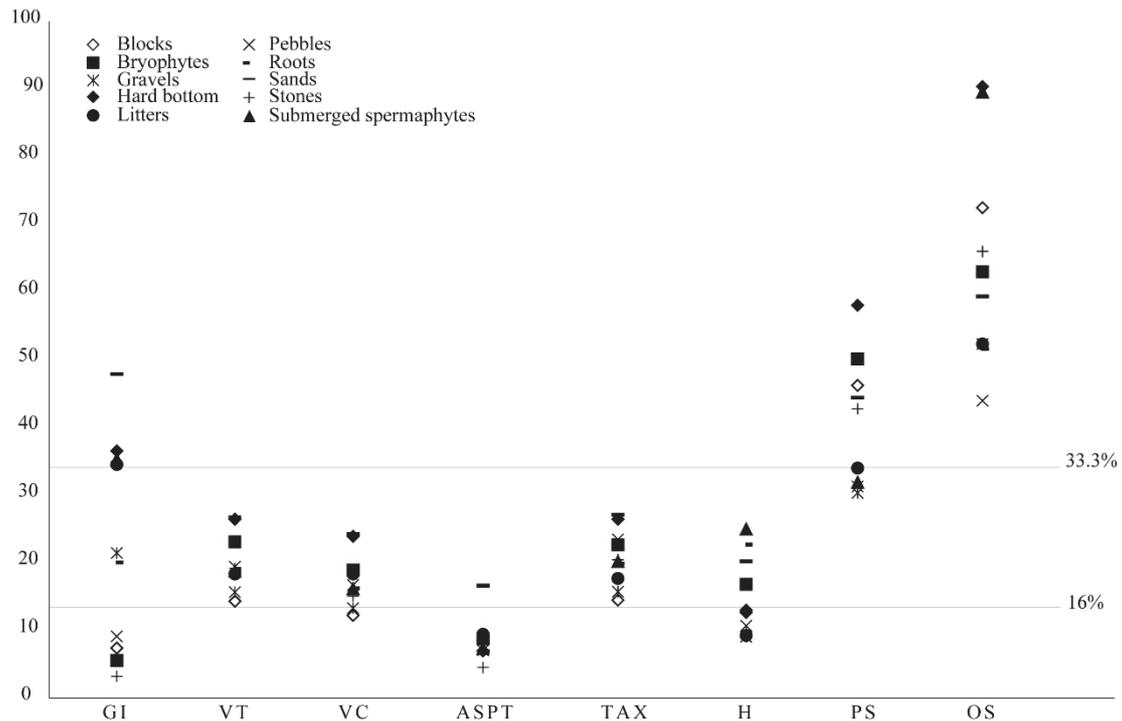


Figure 5 : Average variation coefficient (CV) for IBGN metrics (GI, VT, and VC) and I2M2 metrics (ASPT, TAX, H, PS, and OS) for each substrate. CV < 16% = not significant, 16% < CV < 33.3% = significant, and CV > 33.3% = critical error. (Average of CV calculated for all sites for a given substrate)

The nearest neighbour sorting on faunistic lists showed the Euclidean distance between metrics for each substrate replicate of each site. For simplicity, we chose to present here the bryophyte, submerged spermaphyte and gravel clusters (Figure 6, Figure 7, and Figure 8, on the left).

Intrasite substrate replicates were generally clustered. This observation was more or less pronounced depending on the substrate. For example, the three bryophyte (Figure 6) sites were clearly differentiated while this was not the case for submerged spermaphytes (Figure 7).

The distances between intrasite replicates were usually reduced. For example, the distance between all bryophyte replicates was between 0.06 and 1.05 and that between bryophyte replicates of site 3 did not exceed 0.29. However, some intrasite replicate samples diverged and were distant from other samples, e.g. replicates 5 and 6 of site 1 for gravel (see Figure 8) or spermaphyte replicate 1 of site 2.

These results are illustrated by the heatmap for each substrate, which very clearly shows the site clustering and distance between replicates and sites. Some replicates of the same station were completely dissociated. We also noted the same pattern for submerged spermaphyte, litter, block, pebble, stone, sand and hard bottom substrates.

The same patterns were not observed, however, with the nearest neighbour sorting on the I2M2 metrics (Figures 6, 7, and 8, on the right). There was no distinction between sites. The Euclidean distances between replicates were considerably reduced and ranged from 0.003 to 0.36. However, as for the single linkage agglomerative clustering on faunistic lists, some replicates were outliers, such as replicates 6 and 3 of site 1, and replicates 5 and 3 of site 4 for gravel or for bryophyte replicate 6 of site 3 (Figure 6).

For all substrates, the sites were differentiated based on the faunistic list cluster and not on that of the metrics, except for sand, root and submerged spermaphyte substrates, whose sites were not clearly identified with faunistic lists or metrics.

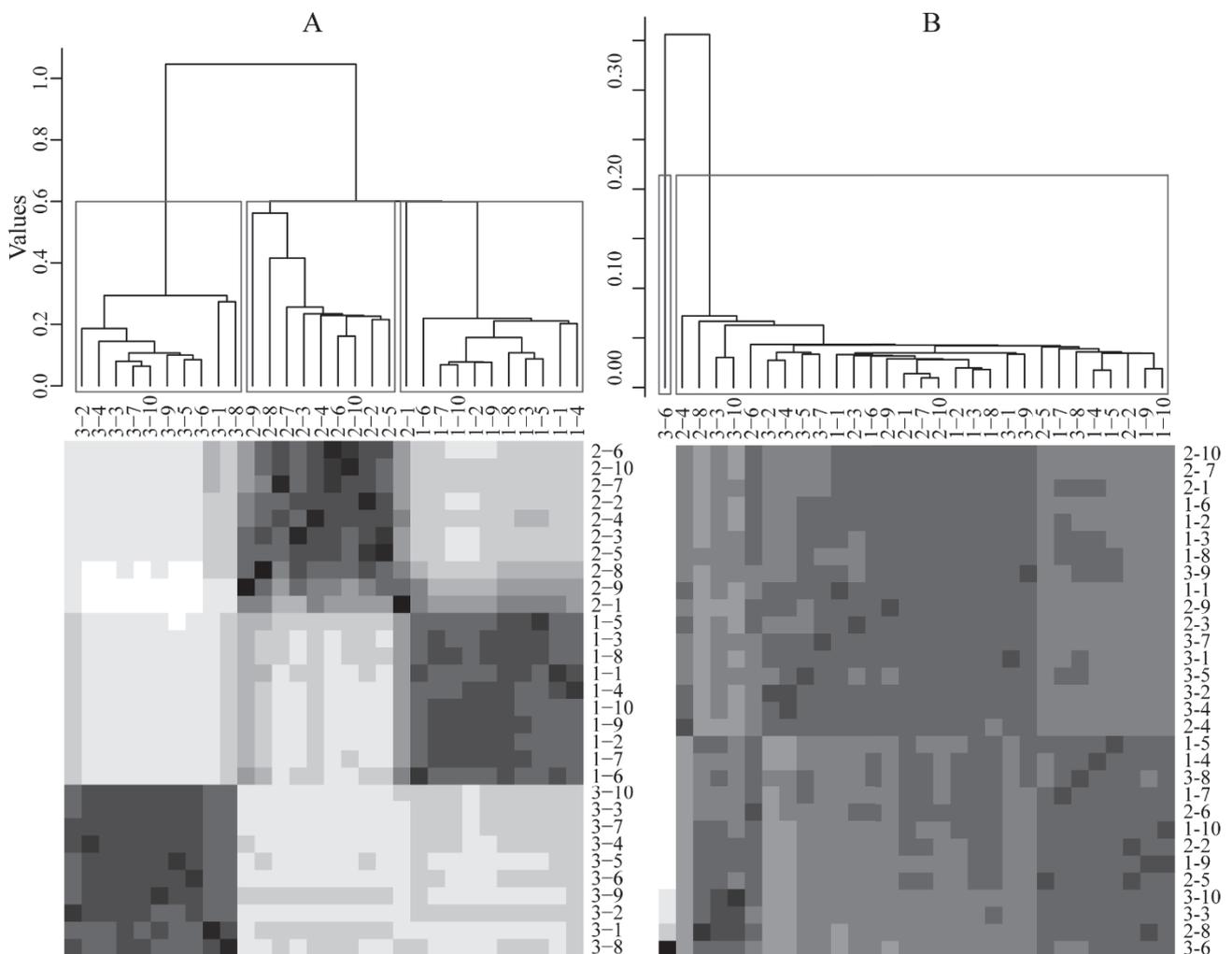


Figure 6: Cluster dendrograms based on Euclidean distance and associated heatmaps on faunistic lists (A) and I2M2 metrics (B) calculated from replicate samples of bryophytes at the three sampled sites in Massif Central Nord (heatmap : black very close, white completely different)

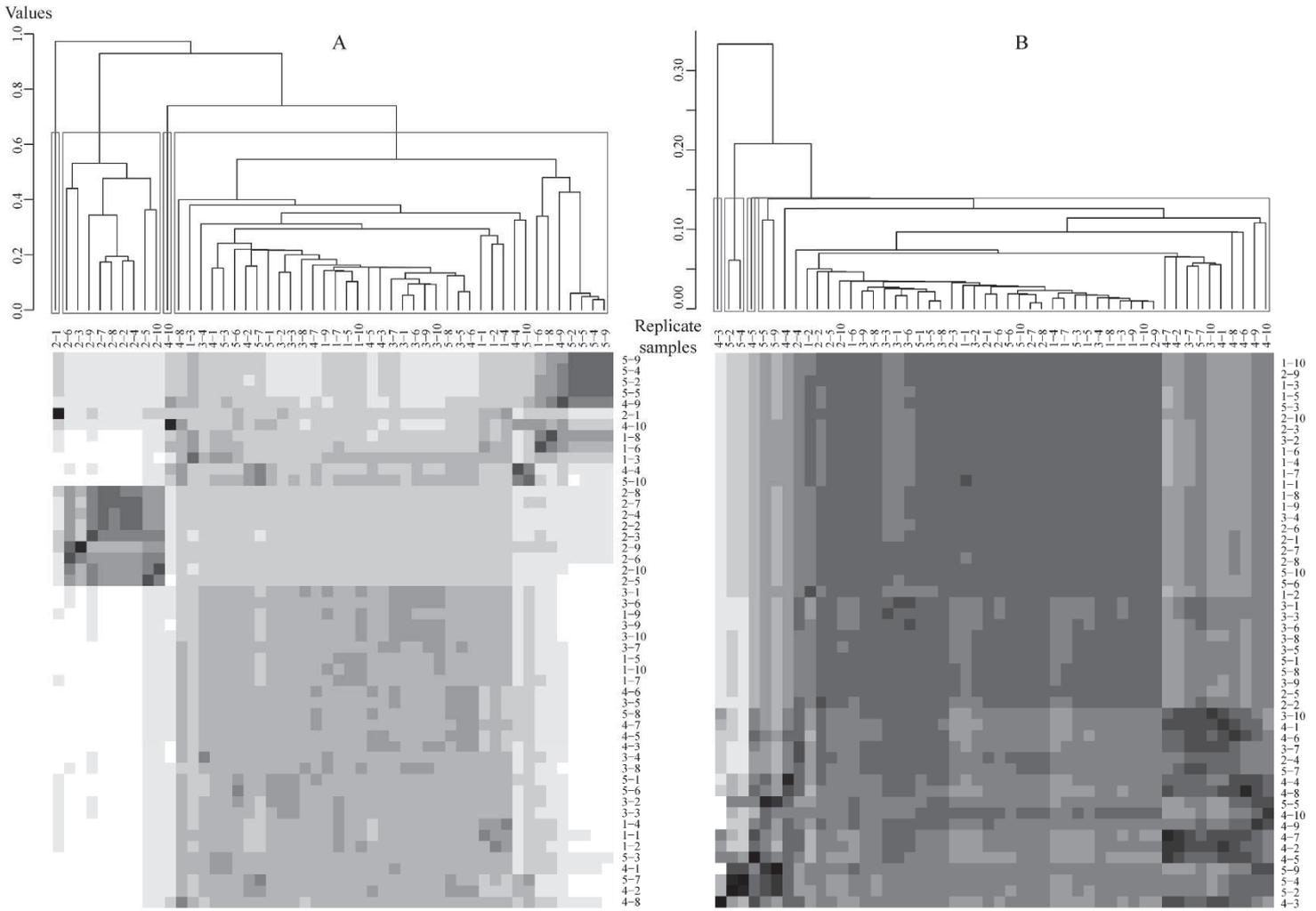


Figure 7: Cluster dendrograms based on Euclidean distance and associated heatmaps on faunistic lists (A) and I2M2 metrics (B) calculated from replicate samples of submerged spermaphytes at the five sampled sites in Massif Central Nord, Causses Aquitains and Massif Armoricaïn (heatmap : black very close, white completely different)

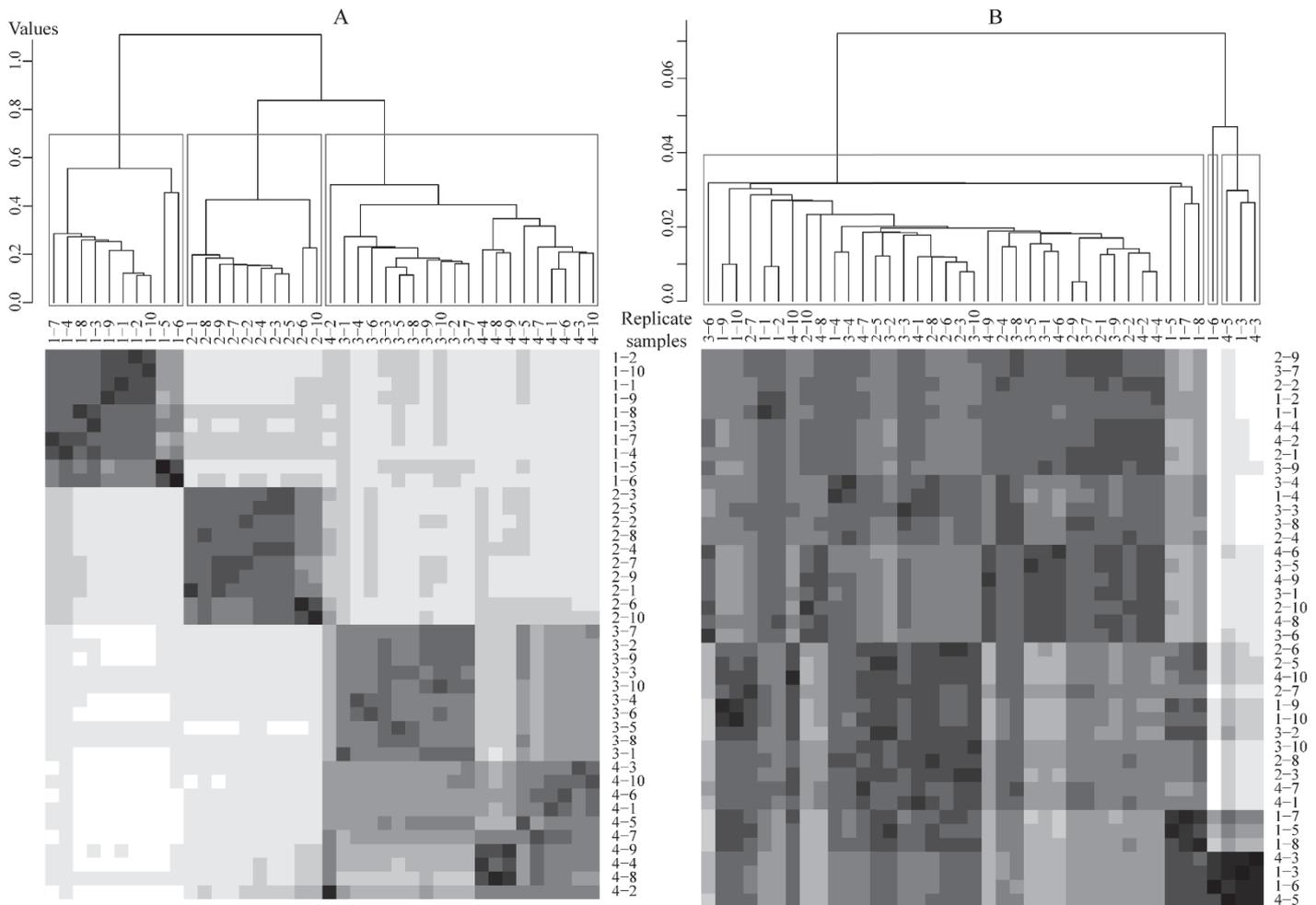


Figure 8: Cluster dendrograms based on Euclidean distance and associated heatmaps on faunistic lists (A) and I2M2 metrics (B) calculated from replicate samples of gravels at the four sampled sites in Massif Central Nord, and Massif Armoricain (heatmap : black very close, white completely different)

§2. Communities similarity according to substrate

Figure 9 shows the average similarity calculated with the Sørensen, Jaccard and Renkonen indices between all pairs of replicates at a site, for each substrate and for all sites.

The Jaccard and Sørensen similarity indices based on the presence or absence of taxa had the same distribution. They highlighted the proportion of common species between replicates of each substrate—the replicate similarity increased as the index increased. The Jaccard index ranged from 45 to 59% (Figure 9). Block, gravel and pebble substrates had the highest similarity index. The number of taxa shared by all replicates was higher than in other substrates.

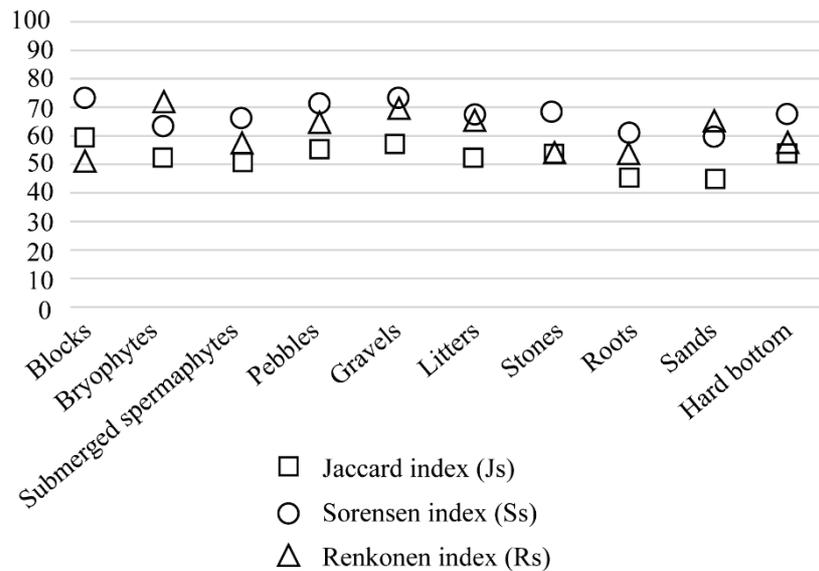


Figure 9: Average of three similarity indices (Jaccard index, Sørensen index, and Renkonen index) for each sampled substrate at all sites (in %)(Example: JS average of blocks = (blocks average JS site 1 + blocks average JS site 2 + blocks average JS site 3/ 3))

The Sørensen index gave more weight to common species, so we found higher similarity values between samples. It ranged from 60 to 73%. Compared with the Jaccard index, the Sorensen index showed which substrate had more common species. Again the mineral substrates were very similar as litter and submerged spermaphyte substrates. Root and sand substrates had the lowest similarity.

The Renkonen index compared the proportional dominance of taxa in replicates and was based on the sum of the minimum number of species detected in replicates, so it was based on the dominance structure. The response of the Renkonen index differed more than that of the Jaccard and Sørensen indices. Renkonen indices ranged from 51 to 71%. When the Renkonen index was higher than the Jaccard and Sørensen indices, this meant that the taxa shared between the sample pairs were dominant in terms of frequency with respect to non-shared taxa. When the Renkonen index was low, this meant that shared taxa were less frequent in a sample. Bryophyte, sand, pebble and gravel substrates had a high Renkonen index, whereas block, stone and root substrates had the lowest.

§3. Site effect

Multidimensional scaling (MDS) based on the Sørensen index represented, in a 2D space, the approximate distance between each substrate replicate according to the site distribution. For simplification, we present the MDS for root and gravel substrates (Figures 10 and 11). In most cases, all intrasite replicates were clustered, such as root or gravel substrates.

The sites were clearly identified, even though replicates of several sites were sometimes very close, depending on the number of taxa shared by the different replicates. Hence, the HER effect was quite clear, e.g. for gravel substrates (Figure 10), site 2 corresponded to the “Massif Central Nord” region (21) and the three other sites to the “Massif Armoricaïn” region (12). We noted that these HER were distinct from each other. The HER sampled for roots could also be distinguished. Site 1 for roots corresponded to “Massif Central Nord” while sites 2 and 3 corresponded to “Landes” (Figure 11). This HER distinction was valid for all substrates.

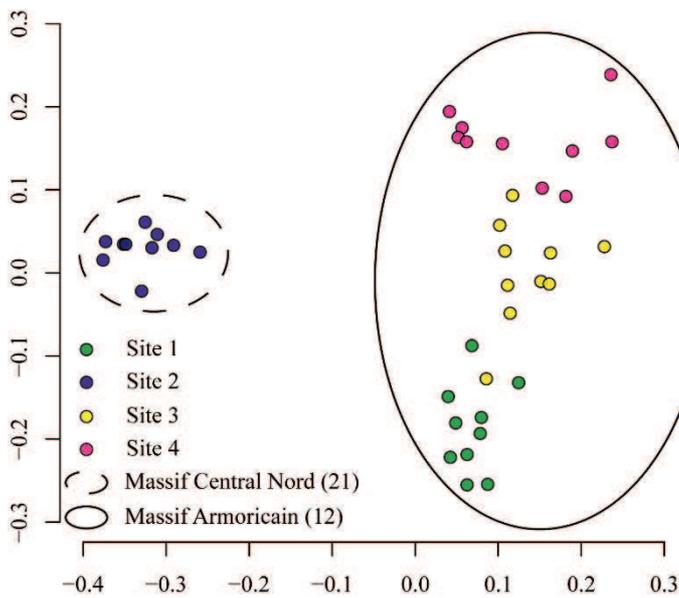


Figure 10: Multi-Dimensional Scaling based on the Sørensen distance matrix on taxonomic lists of gravel substrates for all sampled sites represented in a 2D space. Each point corresponds to a gravel replicate.

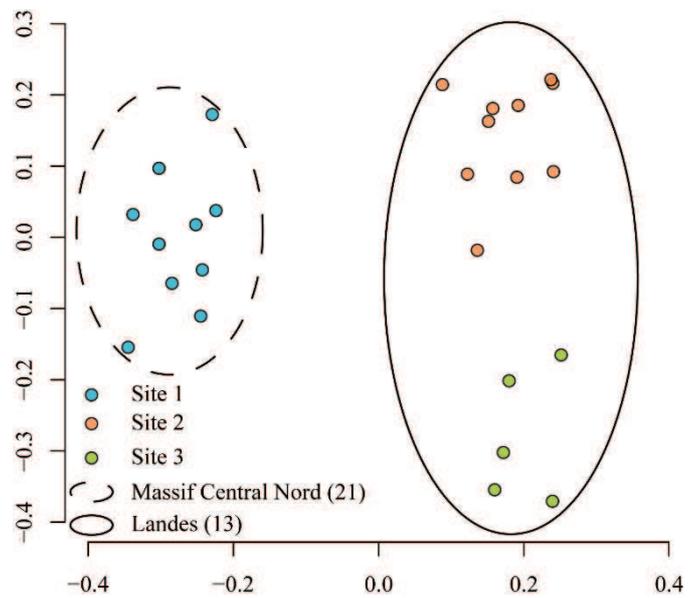


Figure 11: Multi-Dimensional Scaling based on the Sørensen distance matrix on taxonomic lists of roots for all sampled sites represented in a 2D space. Each point corresponds to a roots replicate.

Section 4: Discussion

Substrate variability has been clearly documented (Barnes, et al., 2013; Beisel, et al., 2000) but the impact on biocenotic metrics has not yet been measured. Furthermore, the spatial arrangement of some substrates such as litter and sand patches has explained invertebrate variability (abundance of Chironomidae and Copepods) (Palmer, et al., 2000). Our experiment was carried out to estimate the intrasubstrate variability, the extent of site effect, and their impacts on biocenotic metrics, IBGN and I2M2 metrics robustness.

The cluster and multidimensional scaling analysis revealed the site effect through the significant differences between HER. It was not possible to compare substrates from different sites or HER because the macroinvertebrate taxa lists differed. These differences were due to the macroinvertebrate ecoregional distribution depending on the geology, topography and climate (Wasson, et al., 2004). In fact the definition of ecological class limits based on IBGN and I2M2 was linked to HER (Table 6) (Directive 2000/60/CE, 2008; Ministère de l'Ecologie, du Développement durable et de l'Energie, 2012; Mondy, et al., 2012).

Table 6: Inferior ecological class thresholds by 3 river types and 7 HER (concerned in this study) for the IBGN score (Ministère de l'Ecologie, du Développement durable et de l'Energie, 2012). (thresholds explanation : a-b-c-d : a = very good status inferior threshold,, b = good status inferior threshold, c = moderate status inferior threshold, d = poor status inferior threshold)

HER	General case or HER 2	Medium	Small & Very small
Massif Central Nord (21)	General case	18-15-11-6	18-15-11-6
Massif Central Sud (3)	General case	18-15-11-6	18-15-11-6
Causses Aquitains (11)	General case		15-13-9-6
Coteaux Aquitains (14)	General case	15-13-9-6	15-13-9-6
Massif Armoricaïn (12)	A-Center-South	15-13-9-6	15-13-9-6
	B-West-North East	16-14-10-6	16-14-10-6
Pyrénées (1)	General case	16-14-10-6	16-14-10-6
Landes (13)	General case	15-13-9-6	15-13-9-6

Our results confirmed the DCE choice of thresholds associated with the different HERs. Beyond the site effect, intrasite substrate variations were observed depending on the substrate type, so it would make sense to measure the intrasubstrate variability effect on the biocenotic metrics.

Substrates were classified on the basis of their taxonomic variability and similarity indices (Figure 12-A). Some substrates were more stable within the same site. Reconsidering the faunistic lists could thus help to understand how and why the replicates were similar or not. For example, within the same site, blocks represented the least variable substrate, whereas sand and bryophyte substrates were the most variable. This variability could thus be linked to the substrate biogenic capacity (Figure 12-B). The biogenic capacity of substrates was defined by an empirical method and could vary according to standard variations. For example, the sampling grid varied between IBGN and the new standard (AFNOR, 2006; AFNOR, 2009): blocks were added, litter and root substrates were divided into two substrates, and algae and hard bottom substrates were inverted in the last standard (AFNOR, 2009). We observed that the least variable substrates were the least biogenic, except for sand, and we noted the succession of mineral substrates and organic substrates (Figure 12).

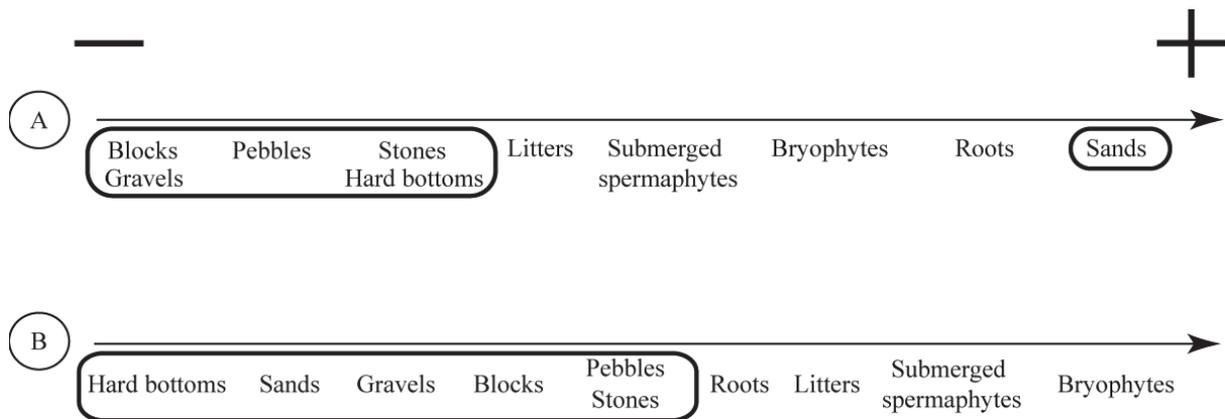


Figure 12: Comparison between intrasubstrate variability and substrates liveability. A: Variability gradient of substrates depending on the Sørensen similarity index, B: Substrates liveability depending on the last French sampling grid (AFNOR 2009) (in boxes mineral substrates)

Root and sand substrates were the most variable according to the CV and similarity indices. Roots were an extremely heterogeneous substrate, which could be explained by the variable spatial density of roots, i.e. locally more or less charged with coarse organic matter, and thus containing different habitats. The current velocity between the sampled plot and the root hair was difficult to measure and led to some variations within this substrate. All of these differences might influence taxa diversity and abundance. Bryophytes were also found to be a variable substrate. The choice of bryophyte species to be sampled was crucial because bryophyte species were an extremely heterogeneous substrate. For example, a patch of *Fontinalis antipyretica* which has a sparse morphology, was not as biogenic as a patch of *Rhynchostegium riparioides* which is denser, grows rapidly, and consequently represents a shelter habitat. Field protocols do not take the sampled bryophyte species into account and an inexperienced operator may bias the sampling. In sand substrates, the heterogeneity might be due to the nature and degree of clogging and the presence of secondary substrates such as litter. The proportion of secondary substrates could vary between sampled plots and modify the presence and abundance of taxa. The same could apply regarding the variation in hard bottom substrate, e.g. some bryophytes could cover a hard bottom, as a secondary substrate, and bring some new taxa.

This has been recognised by experienced hydrobiologists and the last French sampling standard (AFNOR, 2009), for some substrates the presence of a secondary substrate led to variability. Field operator subjectivity has an impact on the sampling, since sampled substrates must be described to explain part of the habitat variation. Therefore, standards provided details on contextual informations to collect during sampling such as facies (pool, riffle, channel bottom, etc.), current

velocity, water depth, degree of clogging, secondary substrate, and tertiary substrate (AFNOR, 2006; AFNOR, 2009).

The above observations suggest that variations in three nonsampled substrates, i.e. clay, helophytes and algae, could be predicted. Clay could have the same behaviour as hard bottom without any secondary substrate—it is a very dense and poor substrate, so low variability could be expected. Helophytes could be compared to roots, i.e. they have varied density and plant structures, so there could be highly variable. Algae, as well as submerged spermaphytes, trap matter and invertebrates depending on their biomass, suggesting that algae substrate could be variable. In our study, the variability in submerged spermaphytes was perhaps overestimated because we only sampled water-starwort plants (*Callitriche* sp.), i.e. a very heterogeneous substrate (personal communication (Association Générale des Laboratoires d'Analyse de l'Environnement, 2013)). The observed substrate variation only concerned dominant substrates (cover area $\geq 5\%$ of the studied site). So it is not certain that the marginal habitats (cover area $< 5\%$ of the studied site) had the same variability pattern. Moreover, the assessed substrate variability concerned only very good sites, so it is possible that the variations were not the same for other ecological quality classes. Likewise, we observed differences between substrates coming from different HERs, so the observed substrate variation in a given HER could thus differ in another.

Metrics clustering indicated that the replicates were closer than taxonomic list replicates but the metrics varied, so the ecological interpretation of the results could change and not be valid. Here again, the metrics were classified according to their CV (Figure 13). Hence, ASPT and diversity metrics were the least variable, while metrics associated with the biological traits and GI were the most variable. ASPT in block substrates had a CV of 6.43% and 15.86% in sand. ASPT was based on identifications at the family level. Improvements in the standard, which called for genus level identification, could reduce family identification errors. However, the CV of the relative abundance of polyvoltine species (PS) was 44.77% in block substrates and 43.03% in sand. The biological trait metrics (PS, OS) were based on abundance estimation. Biological trait variability was a consequence of the abundance variation. Laboratory abundance estimation is not yet a standardised process (AFNOR, 2010). For example, Aquabio used its own standardised estimation method based on intersurveyor tests and high variability between surveyors was noted (ongoing study, Aquabio 2014). Finally, these two metrics were more variable than others.

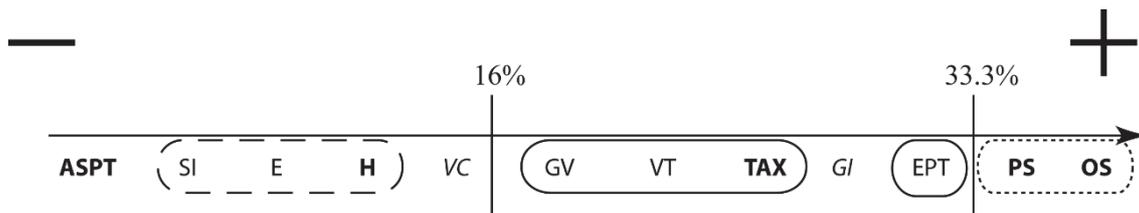


Figure 13: Variability gradient of metrics based on CV (in dotted lines: metrics of biological traits, large dotted lines: Diversity metrics, in boxes: richness taxonomic metrics; in bold: I2M2 metrics, in italic: IBGN metrics) CV <16% = not significant, 16% < CV < 33.3% = significant, and CV >33.3% = critical error.

We worked on the metrics of the two French indices calculated for only one substrate per site, but our study did not permit the index calculation. The two IBGN metrics GI and VT were based on pollution-sensitive species and richness and further depended on the substrate combination. We showed that these two metrics had high degrees of variability. We thus expected that the IBGN could be more sensitive to substrate variations than the I2M2. Beisel et al. (1998) already showed that uncertainty in the IBGN score was highly based on randomisation of the mesohabitat combination. They estimated that the difference between the measured and simulated IBGN values could be as high as 3 points. Intrasubstrate variability in addition to the substrate combination could lead to a high IBGN variation.

By contrast, intrasubstrate variation had less impact on the I2M2 metrics, i.e. the distances between replicates were smaller. The I2M2 metrics calculation took the taxa abundance, richness, and taxon identification into account. Moreover, the I2M2 calculation depended on the discrimination efficiency based on 17 pressures for each tested metric (Mondy, 2012), so so the I2M2 metrics were likely more robust and less impacted by the habitat combination.

Section 4: Conclusion

We assessed the intrasubstrate variability and its impacts on intrasite metrics using experimental field sampling based on substrate replicates. We observed that different nested scales, as defined by Wasson (2004), generated variability: this study highlighted the significant relationship with the HER and intersite variation, but also intrasubstrate variability within the same site. Block and gravel substrates were the least variable substrates while root and sand substrates were the most variable. Substrate variability impacted the index metrics differently.

I2M2 metrics calculated for each replicate were more robust than IBGN metrics, and actually the I2M2 metrics eliminated the intrasubstrate variability.

Taxa abundance estimation and diversity were two parameters impacting the metrics calculated for substrates. They were linked to the laboratory method used and to the surveyor's skills. A quality control approach and normalized process could be developed to monitor the sorting and abundance estimation. As further research, it could be interesting to measure index score uncertainty due to surveyor based on the presence and/or absence of rare or barely abundant taxa and the abundance estimation error. We do not yet know how their aggregation impacts the IBGN and I2M2 scores and if there is taxa compensation between different substrates of a sample or error accumulation.

Titre 2 : Etude des incertitudes associées aux indices macroinvertébrés liées aux variabilités naturelle et opérateur sur la base des essais inter-laboratoires menés par AGLAE

Le substrat, sa nature et son hétérogénéité, apportent une part importante d'incertitude dans l'évaluation de la qualité des rivières (Beisel, et al., 1998a; Barnes, et al., 2013, cf Titre 1). Cependant, la variabilité naturelle liée aux substrats n'est pas la seule composante pouvant influencer les résultats. En effet, toute application des protocoles dépend également des opérateurs, responsables de l'échantillonnage, du terrain jusqu'à la saisie des listes faunistiques. Ainsi, après avoir démontré la variabilité intra-substrat, nous nous sommes intéressés à la variabilité liée à l'opérateur. Pour prolonger la démarche expérimentale, nous exploitons les données issues de l'essai inter-laboratoire mené par l'Association Générale des Laboratoires d'Analyse de l'Environnement (AGLAE).

Section 1 : Introduction

De nombreuses études se sont penchées sur les pratiques mises en œuvre, comme l'échantillonnage sur le terrain, mais également sur le rôle de l'expérience de l'opérateur. Ainsi, il a été montré que si l'effort d'échantillonnage sur le terrain jouait un rôle sur les résultats finaux (Vlek, et al., 2006), l'appropriation des normes par l'opérateur avait également son importance dans la construction du plan d'échantillonnage (Archambault, 2014). La méthode d'échantillonnage utilisée au laboratoire peut également mener à des différences notoires d'évaluation des rivières (Haase, et al., 2006), de même que l'expérience de l'opérateur (Metzeling, et al., 2003).

Le poids de l'expérience de l'opérateur et les pratiques en laboratoire sont donc des sources d'incertitude reconnues pour l'évaluation des rivières. C'est dans ce contexte qu'intervient AGLAE qui a pour objectif de contribuer à l'amélioration des analyses grâce à des essais inter-laboratoires dans les domaines de la chimie comme de l'environnement. Les essais inter-laboratoires en hydrobiologie, et plus particulièrement sur les protocoles macroinvertébrés sont parmi les premiers existants et AGLAE nous a accordé l'autorisation d'utiliser les données et résultats produits en 2011.

Dans un premier temps, nous allons donner une brève présentation du programme mené par AGLAE et des premiers résultats obtenus (§1). Dans un deuxième temps, nous présenterons les analyses statistiques réalisées dans le cadre de notre étude (§2). Enfin, nous proposerons quelques pistes d'explication des variations observées (§3).

Section 2 : Présentation de l'essai inter-laboratoire en hydrobiologie

Le programme lancé par AGLAE a démarré en 2010 et concerne 37 laboratoires privés ou publics. Son objectif est de mettre en avant les compétences des laboratoires pour le tri, la détermination et le dénombrement des individus dans un IBGN ou un IBG DCE. Il est à noter que l'ensemble des résultats transmis par AGLAE dans le cadre de ce travail préserve l'anonymat des laboratoires participants.

L'essai de 2011 a été suivi par seize laboratoires et reposait sur un échantillonnage, répété 64 fois pour chacun de trois substrats différents sur une même station de très bonne qualité par le même bureau d'étude, afin de limiter le biais induit par l'échantillonnage. Les trois substrats échantillonnés sont les racines, les pierres et les spermaphytes immergés dans la rivière Vienne. Les échantillons associés aux indices IBGN et IBG DCE sont envoyés à chaque participant. Ils sont composés, pour chacun, de quatre échantillons élémentaires sélectionnés aléatoirement par substrat, afin de reformer un échantillonnage complet de douze échantillons (3 substrats X 4 échantillons élémentaires). Chaque laboratoire pratique le tri et la détermination des individus en respectant la norme XP T90-388 (AFNOR, 2010), qui fixe le niveau de la détermination dans la plupart des cas au genre ou la norme NT T90-350 (AFNOR, 2004), qui fixe la détermination à la famille. AGLAE demande à chaque participant de fournir les listes faunistiques associées à chaque échantillon de substrat et la liste globale, ainsi que de renseigner la variété taxonomique et le groupe faunistique indicateur trouvés et, éventuellement, les notes IBGN et IBG DCE obtenues, sur le lot réduit de substrat suivant la norme choisie. AGLAE nous a fourni les listes faunistiques par laboratoire et répétitions et les résultats des analyses menées par AGLAE décrites ci-dessous. A partir de ces données, AGLAE a réalisé des analyses statistiques descriptives (moyenne, intervalle de confiance et écart-type) ainsi qu'une analyse de performance des participants avec le calcul du Z-score (Association Générale des Laboratoires d'Analyse de l'Environnement, 2013). Ils ont ainsi pu mettre en évidence que la plupart des laboratoires évaluait la qualité biologique des eaux de manière équivalente. En effet, tous les laboratoires ont obtenu le même groupe faunistique indicateur. Cependant, il a également été montré que certains laboratoires se distinguaient de façon singulière des autres laboratoires. Ils n'avaient pas obtenu le même taxon indicateur ou avaient des diversités taxonomiques plus réduites sur la base des intervalles de confiance, mais également des identifications douteuses de taxons mises en évidence par les Z-scores et confirmées par dire d'expert.

Cette divergence de résultat entre laboratoires n'est pas expliquée par l'étude mise à disposition par AGLAE. Elle peut être due à la variabilité intra-substrat, se répercutant dans l'échantillonnage, mais également au degré de compétences dans la phase de traitement des échantillons au laboratoire des laboratoires concernés, ou aux deux, c'est ce qui nous intéresse dans ce jeu de données. Nous considérons ici que l'effet opérateur sur l'échantillonnage lors de la phase terrain est constant car il a été réalisé par un seul laboratoire et nous allons essayer de

mieux comprendre l'effet opérateur en réalisant d'autres analyses statistiques sur ce jeu de données.

Section 3 : Mesure de la variabilité opérateur en laboratoire

Grâce à AGLAE, nous disposons d'un jeu de données obtenu après des pratiques de tri et d'identification menées par des opérateurs différents, dont nous pouvons supposer qu'ils travaillent avec des méthodes différentes (élutriation, différents grossissements de loupes binoculaires, etc.) et ont des expériences différentes. Ainsi, ce jeu de données reflète à la fois la variabilité intra-substrat et les variabilités associées aux pratiques de laboratoire et de l'expérience de l'opérateur au laboratoire, dans le cas où nous considérons la variabilité due à l'opérateur sur le terrain comme négligeable.

Afin de mesurer les effets des différentes variabilités, nous avons à notre disposition un jeu de données regroupant les listes faunistiques associées à chacune des répétitions de substrat et pour chacun des seize laboratoires. Ainsi, nous avons quatre listes par substrat et par laboratoire, soit un total de 192 listes faunistiques. Afin de pouvoir comparer les variabilités liées à l'opérateur au laboratoire pour chaque substrat nous avons décidé de réaliser des analyses statistiques par substrat et non pas sur la reconstitution d'un échantillonnage fictif (4 répétitions par substrat soit 12 échantillons par laboratoire). Nous avons produit les analyses sur trois jeux de données composés de 64 listes faunistiques chacun.

Nous avons décidé d'opter pour les mêmes analyses statistiques que celles que nous avons utilisées pour déterminer la variabilité liée aux substrats. Nous avons calculé les coefficients de variation (CV) et les écarts types (DS) sur les cinq métriques de l'I2M2, elles-mêmes calculées sur chaque répétition par substrat. Nous avons également réalisé un positionnement multidimensionnel, basé sur l'indice de similarité de Sørensen, construit à partir des listes faunistiques pour chaque substrat par laboratoire, ainsi qu'une classification des données, basée sur la méthode des plus proches voisins, à partir des listes faunistiques et des métriques de l'I2M2 calculées sur chaque répétition de substrat.

Les coefficients de variation (CV) et les distances (DS) ainsi que les classifications sur les métriques nous renseignent sur la variabilité des métriques de l'I2M2 en lien avec les variabilités

intra-substrat, opérateur et celles dues aux pratiques au laboratoire. Les positionnements multidimensionnels et les classifications réalisées à partir des listes faunistiques nous renseignent sur la variation de ces listes en fonction des mêmes facteurs cités précédemment.

§1. Variabilité des métriques de l'I2M2

Le Tableau 2 présente les CV et DS moyens calculés sur les 5 métriques de l'I2M2 (Cf Partie 2, Chapitre 1, §2, 5.1) pour les trois types de substrats prélevés dans le cadre des essais inter-laboratoire.

Tableau 2 : Moyennes des coefficients de variation (CV) et des distances (DS) pour les métriques de l'I2M2 pour les trois substrats échantillonnés dans le cadre de l'essai inter-laboratoire mené par AGLAE et rappel des coefficients de variation moyens obtenus sur les mêmes substrats par l'expérimentation sur la variabilité intra-substrat (cf. Partie 2 Chapitre 1) (en bleu ciel $16\% < CV < 33.3\%$, en bleu ciel et gras $CV > 33.3\%$) $CV < 16\%$ = pas significatif, $16\% < CV < 33\%$ = significatif, and $CV > 33\%$ = erreur critique.

Métriques	Richesse			H			ASPT			PS			OS		
	DS	CV	CV IS	DS	CV	CV IS	DS	CV	CV IS	DS	CV	CV IS	DS	CV	CV IS
Pierres	10,10	27,41	19,6	0,90	18,26	9,10	0,65	9,90	4,07	0,06	15,58	41,40	0,03	76,76	64,13
Racines	9,78	30,95	19,07	0,74	24,68	21,80	0,65	10,41	6,40	0,09	17,45	48,00	0,11	42,65	50,03
Spermaphytes immergés	6,54	21,65	19,36	0,59	10,94	24,02	0,68	10,48	6,80	0,08	9,32	30,79	0,10	23,57	87,10

Nous constatons que la plupart des métriques de l'I2M2 ont une variation significative ($CV > 33,3\%$) quelque soit le substrat concerné. Il est également intéressant de noter que des métriques qui étaient peu variables ou à l'inverse très variables pour un type de substrat, dans le cadre de notre expérimentation sur la variabilité intra-substrat, ont ici des CV qui dépassent les 16% ou au contraire sont en deçà ou proches du seuil de 16%. Si nous comparons les CV moyens obtenus pour l'étude sur la variabilité intra-substrat et ceux obtenus ici, nous constatons des variations plus ou moins importantes. Par exemple, les variations de la fréquence des individus polyvoltins est ici proche des 16%, alors que celles obtenues avec l'expérimentation intra-substrat étaient proches ou supérieures à 33,3%.

Les figures 14, 15 et 16 présentent les classifications hiérarchiques selon la méthode du plus proche voisin sur les métriques de l'I2M2 calculées pour les trois substrats testés (pierres, racines et spermaphytes immergés). Nous constatons que la distance est majoritairement réduite entre les ensembles de répétitions par laboratoire, excepté les répétitions singulières. Elle n'excède pas 0,07 point pour les pierres, 0,06 point pour les racines et 0,05 point pour les spermaphytes immergés. Certains laboratoires sont marginaux et se démarquent par une plus grande distance. En effet, nous constatons que les laboratoires 5, 10, 14 et 15, quelque soit le substrat, forment une classe à part bien distincte du reste des laboratoires ; ainsi que les laboratoires 12 et 16 pour les racines, 13 pour les spermaphytes immergés et 12 pour les pierres. Pour les laboratoires 14 et 15 toutes les répétitions se retrouvent distinctes de celles des autres. Il se pourrait donc que l'effet laboratoire soit dominant. Les autres laboratoires singuliers n'ont qu'une seule répétition qui se distingue des autres, elle pourrait donc démontrer un effet lié au substrat.

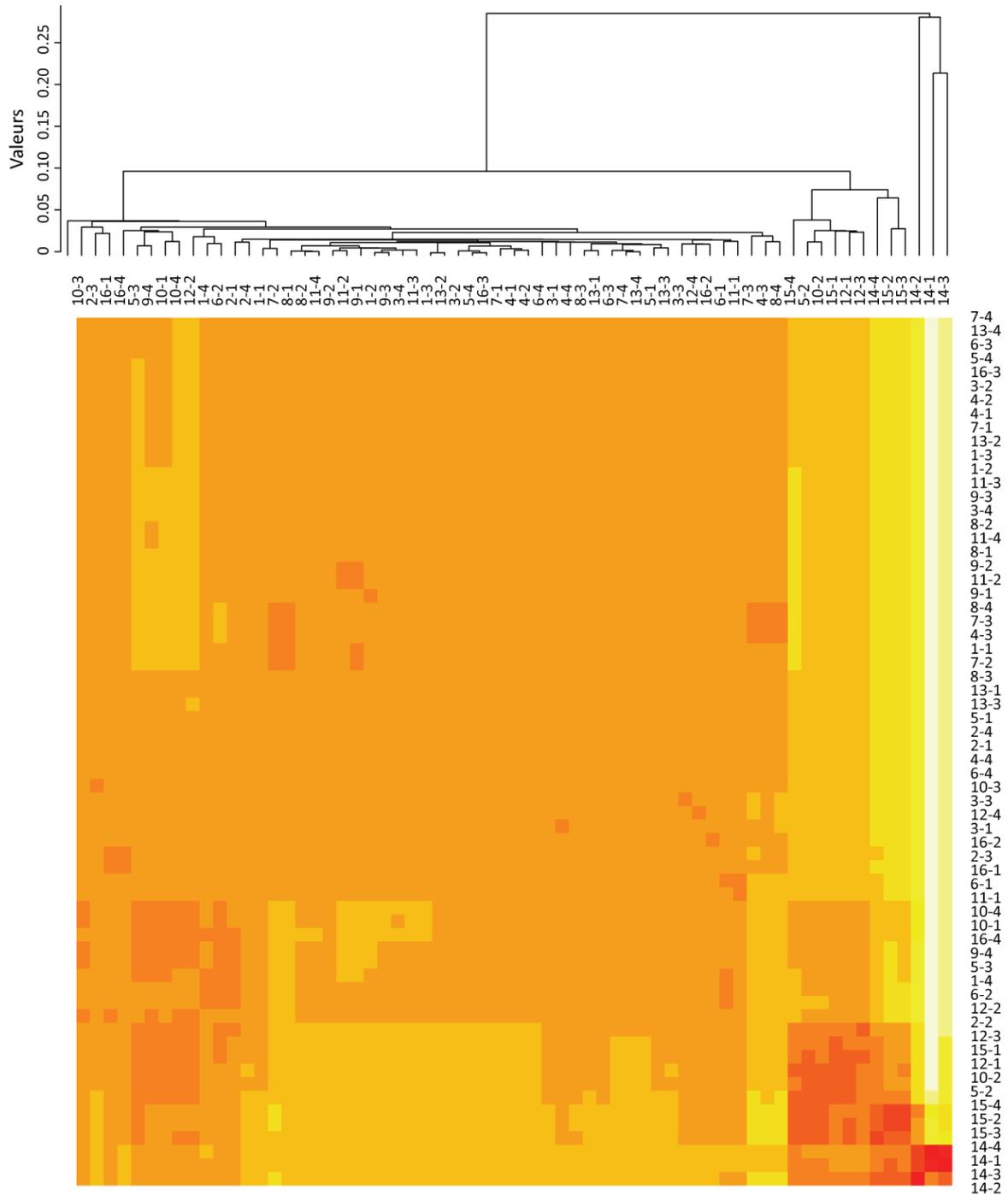


Figure 14: Classification hiérarchique basée sur la distance euclidienne et carte de chaleur associée obtenues sur les métriques de l'I2M2 calculées pour le substrat pierres sur la Vienne (carte de chaleur : rouge = proche, blanc = complètement différent ; le premier numéro = code laboratoire, 16 laboratoires, et le deuxième = code répétition, 4 répétitions de substrats, exemple : laboratoire 10 – répétition 3)

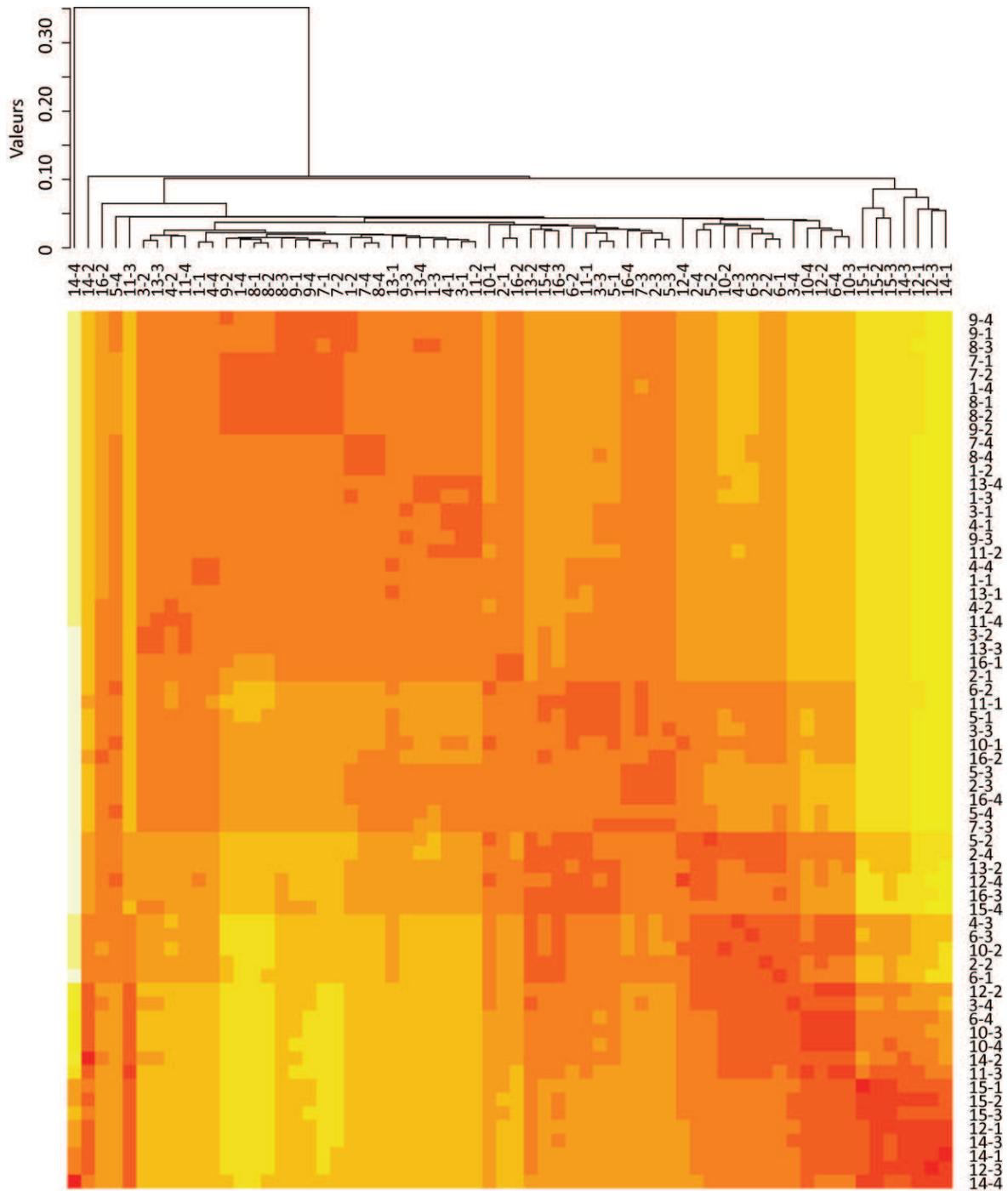


Figure 15: Classification hiérarchique basée sur la distance euclidienne et carte de chaleur associée obtenues sur les métriques de l'I2M2 calculées pour le substrat racines sur la Vienne (carte de chaleur : rouge = proche, blanc = complètement différent ; le premier numéro = code laboratoire, 16 laboratoires, et le deuxième = code répétition, 4 répétitions de substrats, exemple : laboratoire 10 – répétition 3)

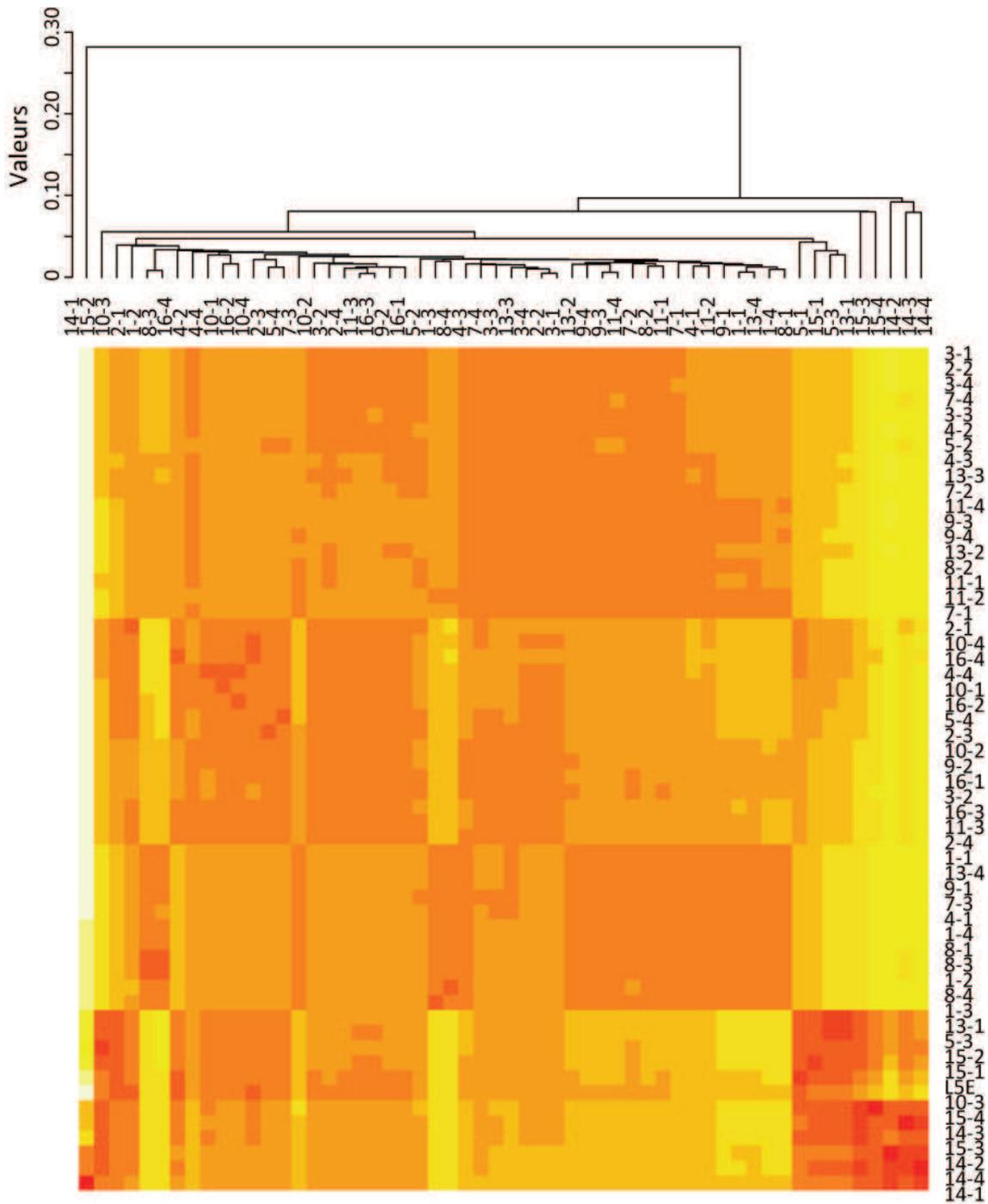


Figure 16: Classification hiérarchique basée sur la distance euclidienne et carte de chaleur associée obtenues sur les métriques de l'I2M2 calculées sur le substrat spermaphytes immergés sur la Vienne (carte de chaleur : rouge = proche, blanc = complètement différent ; le premier numéro = code laboratoire, 16 laboratoires, et le deuxième = code répétition, 4 répétitions de substrats, exemple : laboratoire 10 – répétition 3)

§2. Variabilité des listes faunistiques

Les classifications hiérarchiques obtenues à partir des listes faunistiques et selon la distance euclidienne présentent des distances entre les laboratoires plus importantes que celles observées sur les classifications effectuées à partir des métriques de l'I2M2. Elles oscillent entre 0,07 et 1,06 pour les pierres, 0,11 et 0,91 pour les racines et 0,06 et 0,83 pour les spermaphytes immergés. Nous notons à nouveau des laboratoires qui apparaissent de manière singulière. Nous retrouvons cette singularité pour les laboratoires 5, 14 et 15 pour tous les substrats, le laboratoire 10 pour les pierres et les spermaphytes immergés, le 13 pour les pierres et les racines, et les laboratoires 3, 4, 6, 8, 9, 11 et 12 pour les pierres.

Toutes les répétitions de substrat affectées aux laboratoires 14 et 15 sont indépendantes des autres répétitions des substrats, comme ce que l'on a pu observer au niveau des métriques. Les autres laboratoires concernés n'ont qu'une ou deux répétitions qui se différencient des autres répétitions de substrat.

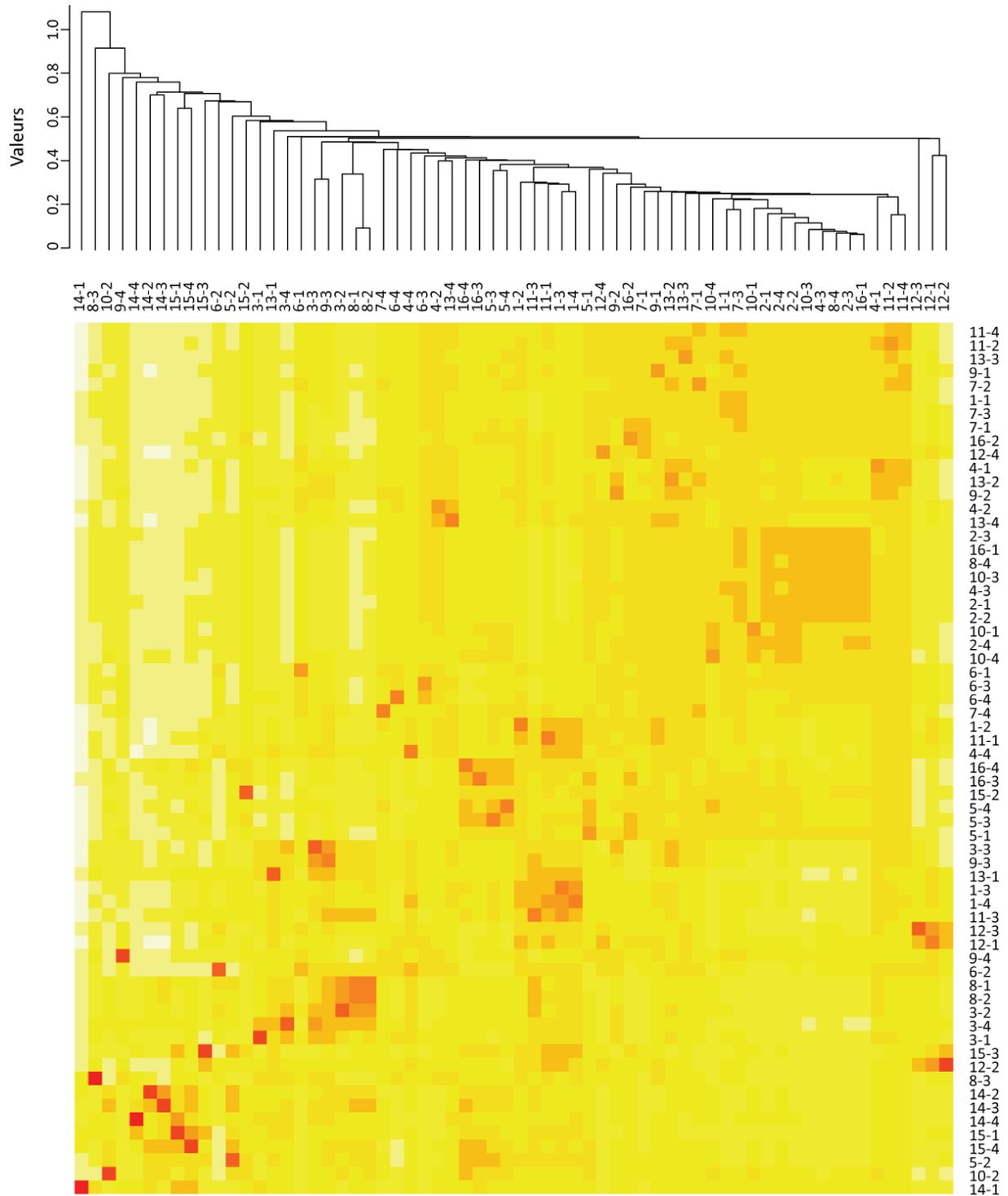


Figure 17: Classification hiérarchique basée sur la distance euclidienne et carte de chaleur associée obtenues sur les listes faunistiques des substrats pierres sur la Vienne (carte de chaleur : rouge = proche, blanc = complètement différent ; le premier numéro = code laboratoire, 16 laboratoires, et le deuxième = code répétition, 4 répétitions de substrats, exemple : laboratoire 10 – répétition 3)

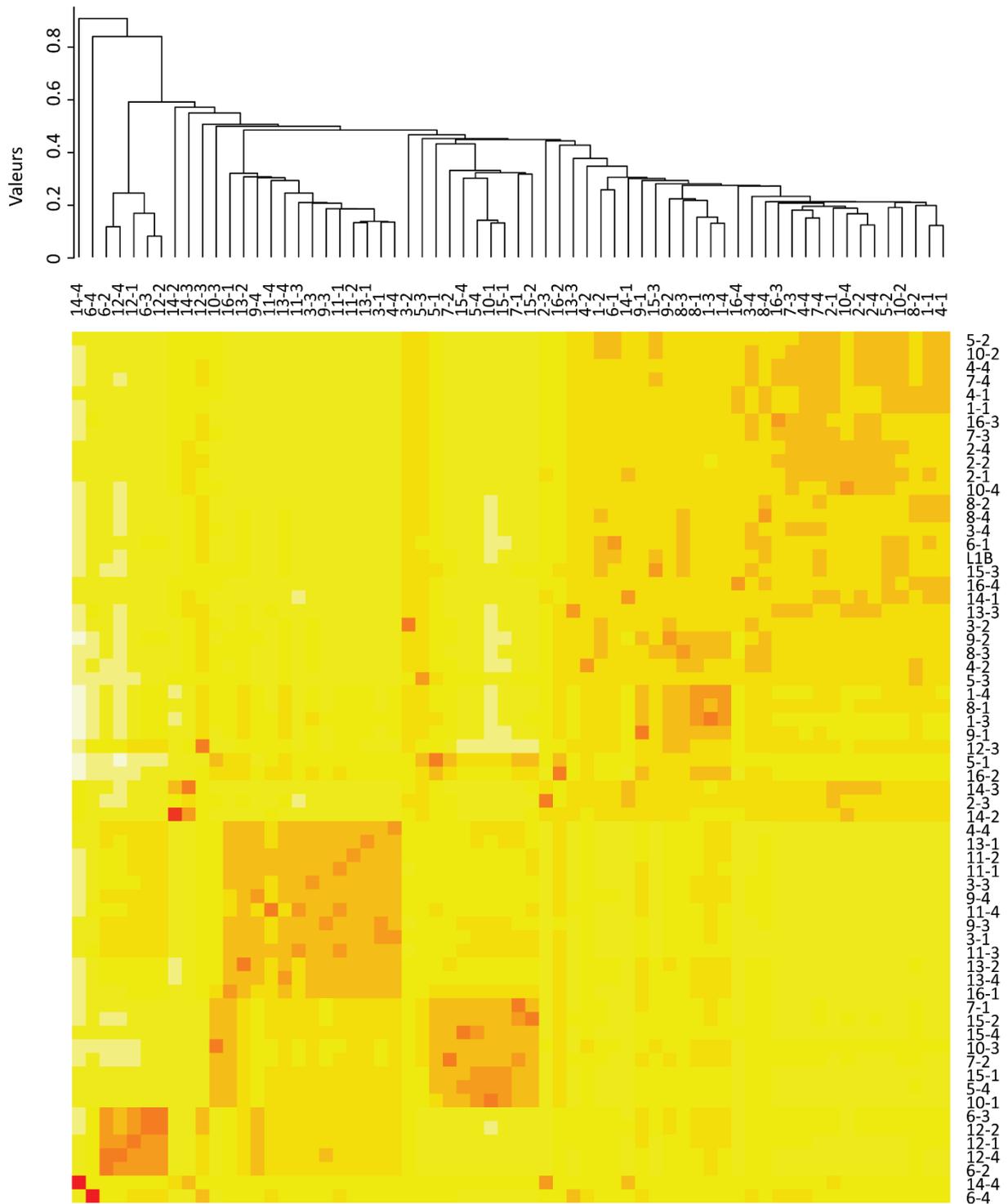


Figure 18: Classification hiérarchique basée sur la distance euclidienne et carte de chaleur associée obtenues sur les listes faunistiques des substrats racines sur la Vienne (carte de chaleur : rouge = proche, blanc = complètement différent ; le premier numéro = code laboratoire, 16 laboratoires, et le deuxième = code répétition, 4 répétitions de substrats, exemple : laboratoire 10 – répétition 3)

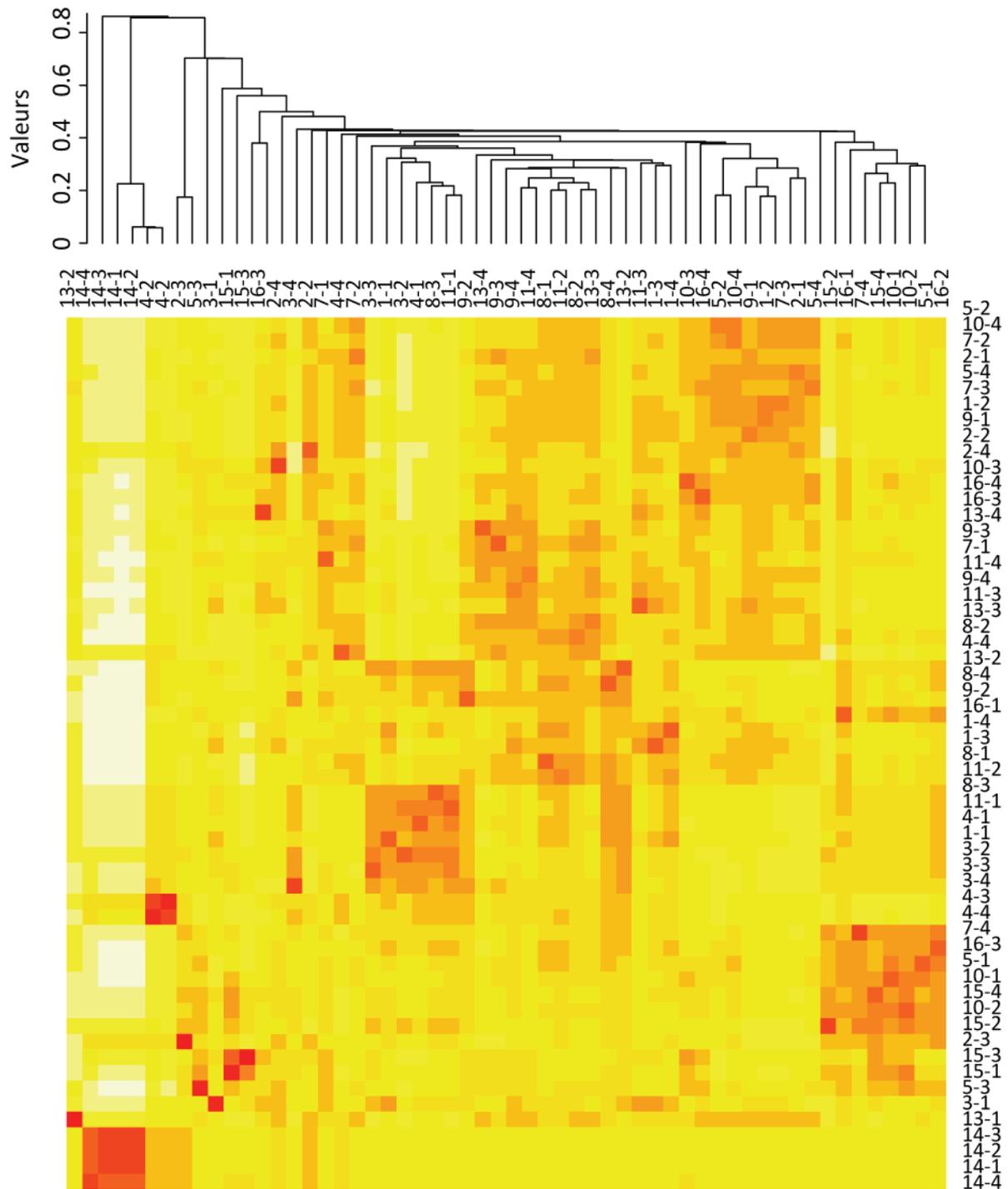


Figure 19: Classification hiérarchique basée sur la distance euclidienne et carte de chaleur associée obtenues sur les listes faunistiques des substrats spermaphytes immergés sur la Vienne (carte de chaleur : rouge = proche, blanc = complètement différent ; le premier numéro = code laboratoire, 16 laboratoires, et le deuxième = code répétition, 4 répétitions de substrats, exemple : laboratoire 10 – répétition 3)

Les positionnements multidimensionnels donnent une projection 2D des distances de similarité selon l'indice de Sørensen entre chaque liste faunistique par type de substrat. A chaque point est associée une liste faunistique et un laboratoire.

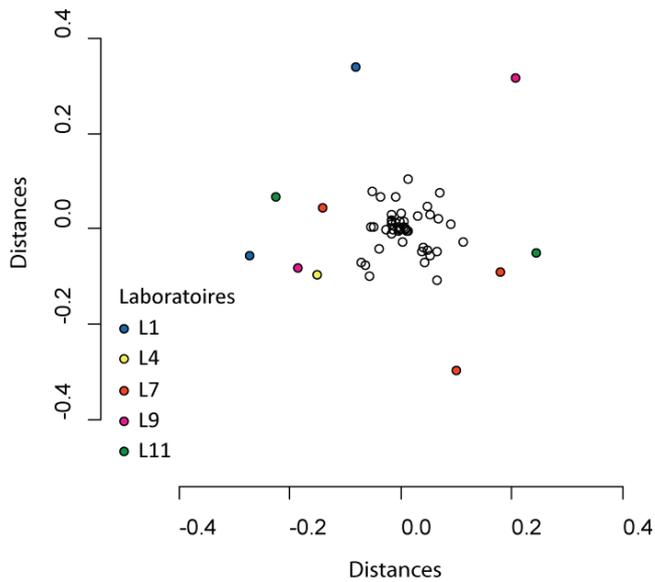


Figure 20: Positionnement multi dimensionnel en 2D basé sur la matrice des distances de Sørensen sur les listes faunistiques des substrats pierres sur la Vienne. Chaque point correspond une liste faunistique et un laboratoire.

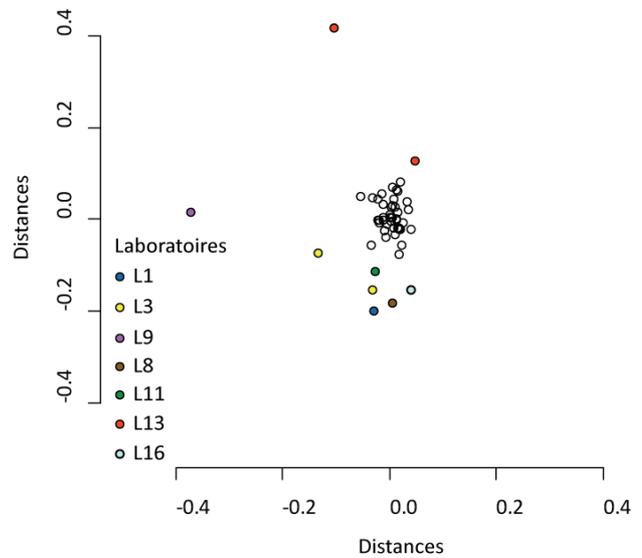


Figure 21: Positionnement multi dimensionnel en 2D basé sur la matrice des distances de Sørensen sur les listes faunistiques des substrats racines sur la Vienne. Chaque point correspond une liste faunistique et un laboratoire.

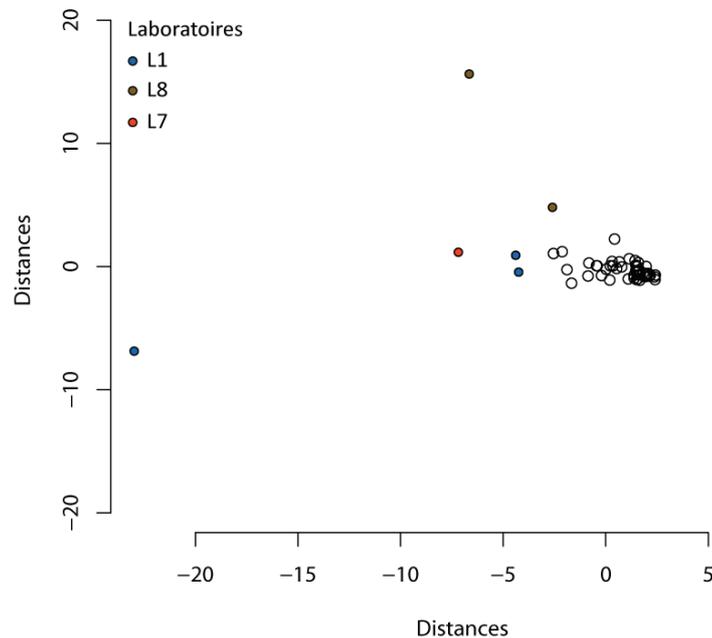


Figure 22: Positionnement multi dimensionnel en 2D basé sur la matrice des distances de Sørensen sur les listes faunistiques des substrats spermaphytes immergés sur la Vienne. Chaque point correspond une liste faunistique et un laboratoire.

Nous constatons, pour les trois substrats ; que les listes sont généralement regroupées entre elles, excepté pour les pierres où les répétitions sont plus dispersées même si les distances sont faibles. Cependant, nous constatons à nouveau quelques marginaux.

Pour les pierres, nous constatons que les répétitions des laboratoires 1, 4, 7, 9 et 11 ont des répétitions qui se distinguent des autres (Figure 20); pour les racines, il s'agit des laboratoires 1, 3, 8, 9, 11, 13 et 16 (Figure 21); pour les spermaphytes immergés des laboratoires 1, 7 et 8 (Figure 22). Les Laboratoires 1, 3, 7 et 11 se remarquent pour plusieurs substrats.

Ces répétitions de substrats se distinguent de la majorité, par des taxons supplémentaires ou au contraire par l'absence de certains taxons trouvés dans les autres répétitions d'un même substrat.

Section 4 : Discussion & Conclusion

L'effet opérateur, dans le cadre de l'application des protocoles basés sur les macroinvertébrés, fait l'objet de plusieurs études en cours. Archaimbault (2014) a montré que les erreurs les plus fréquentes, associées à la phase laboratoire, étaient liées aux différences d'identification entre

opérateur et auditeur mais également aux oublis de taxons représentés par un seul individu. La force de l'étude menée par AGLAE repose sur le fait que seize laboratoires aient participé à l'essai en 2011. Nous savons que ces seize laboratoires sont caractérisés par des types de prétraitement des échantillons différents (pas de prétraitement ou élutriation, tamisage sur colonne, lavage, etc.), par des grossissements utilisés pour le tri sur les petits tamis différents (X7, X2, X4, etc.) et aussi des grossissements utilisés pour l'identification qui peuvent varier d'un laboratoire à un autre (X40, X80, X100, etc.) (Association Générale des Laboratoires d'Analyse de l'Environnement, 2013). Ces différentes techniques et manipulations peuvent avoir un impact sur les opérations de tri et d'identification et expliquer une partie des divergences entre laboratoires. Une uniformisation des pratiques de laboratoires pourraient permettre de réduire la variabilité associée aux techniques et matériels employés.

Les analyses statistiques réalisées ainsi que les résultats fournis par AGLAE ont mis en évidence une déviance de certains laboratoires par rapport à la majorité. Cette déviance s'exprime par des identifications douteuses de taxons (identifiées par un expert et incluses dans le rapport, AGLAE, 2013), mais également par des valeurs de variétés taxonomiques relativement basses. Les erreurs d'identification traduisent un manque d'expérience des laboratoires, tandis que les variations de variété taxonomique peuvent être liées à la fois à l'effet opérateur et à la variabilité intra-substrats. Ainsi, les coefficients de variation (CV), sur ce jeu de données, sont de bons indicateurs du cumul de la variabilité opérateur et de la variabilité intra-substrat. Nous avons constaté que les métriques de l'I2M2 calculées sur les trois substrats (pierres, racines et spermaphytes immergés) ne variaient pas de manière similaire à celles calculées sur les répétitions de substrats dans le cadre de l'étude de la variabilité intra-substrat. En effet, nous avons constaté que l'effet opérateur s'ajoutant à l'effet substrat pouvait entraîner une diminution de la variabilité ou à l'opposé une augmentation de la variabilité. A nouveau, nous pouvons proposer une classification des métriques en fonction des coefficients de variation moyens (Figure 23).

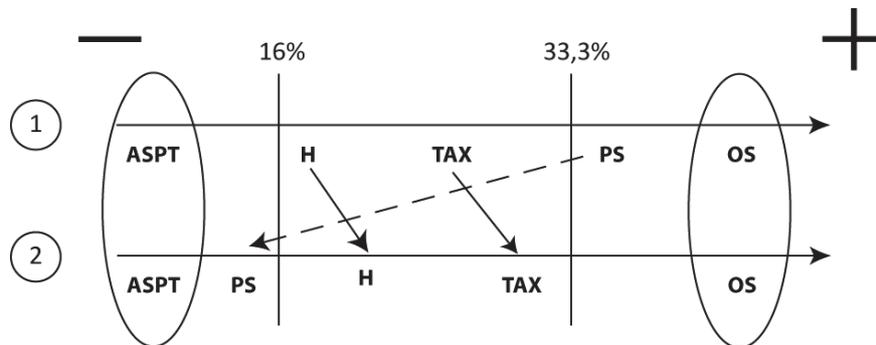


Figure 23 : Gradient de variabilité des métriques basé sur les coefficients de variation moyens obtenus pour chaque métriques de l'I2M2 : 1) expérimentation intra-substrat et 2) expérimentation AGLAE. $CV < 16\%$ = pas significatif, $16\% < CV < 33,3\%$ = significatif, and $CV > 33,3\%$ = erreur critique.

Ainsi, nous confirmons que l'ASPT est une métrique faiblement variable quels que soient les différents laboratoires et que la fréquence des individus ovovivipares est toujours la métrique la plus variable.

Les métriques associées à la diversité et à la richesse taxonomique (H, TAX) sont plus variables dans le jeu de données AGLAE que lorsqu'un seul opérateur a réalisé les opérations de tri et d'identification. Nous observons donc ici un effet combiné de la variabilité intra-substrat et opérateur. Enfin, la fréquence des individus polyvoltins est nettement moins variable : ce serait une métrique pour laquelle les effets opérateurs et les pratiques de laboratoire réduiraient la variabilité intra-substrat, du moins sur les trois substrats étudiés. Contrairement aux résultats obtenus avec la méthode expérimentale précédente étudiant les seules variabilités intra-substrat, les spermaphytes immergés sont moins variables que les deux autres substrats, selon les coefficients de variation moyens obtenus. Les pierres sont le substrat le plus variable ainsi que les racines. La variabilité supérieure des pierres par rapport à l'expérimentation sur la variabilité intra-substrat peut s'expliquer par les pratiques différentes en laboratoires de tri. En effet, les racines et les spermaphytes immergés sont plus difficiles à trier au laboratoire que les pierres et nécessitent un prétraitement et une exploration systématique plus longue. Les pierres peuvent être facilement passées sur une colonne de différentes mailles de tamis ou éluutriées ou non prétraitées et simplement partagées en différents bacs de tri.

Il existe bien un effet inter-opérateur, qui dépend vraisemblablement des pratiques laboratoires, liées au tri et à l'identification, et de la compétence et de l'expérience de l'opérateur observés dans cet essai mené par AGLAE. Néanmoins, cet effet est indissociable de la variabilité

observée sur le substrat. Une homogénéisation des pratiques aux laboratoires et un entraînement des opérateurs pourraient permettre de réduire les variations observées et de ce fait réduire les écarts sur les notes.

CHAPITRE 2 Procédure de simulation de mesure des incertitudes associées aux macroinvertébrés

Ce chapitre présente la procédure de simulation mise en place pour étudier les incertitudes de mesure liées à l'effet opérateur dans l'application des protocoles d'échantillonnage des macroinvertébrés.

Dans ce chapitre, les figures présentées sont en anglais en vue de la préparation d'un article.

Titre 1 : Introduction

L'étude menée précédemment nous a permis de mettre en lumière la variabilité intra-substrat pour différents types d'habitat dans des rivières de très bonne qualité et les effets liés aux opérateurs et aux pratiques en laboratoires (AGLAE). Ainsi, nous avons observé qu'il existait un impact de cette variabilité intra-substrat sur les métriques des indices français et que certains habitats étaient plus hétérogènes que d'autres. Pourtant, il ne nous a pas été possible de calculer la propagation de cette variabilité au sein des indices et son impact sur l'évaluation des rivières du fait que nous n'avions pas d'échantillonnages complets.

La variabilité intra-substrat et l'effet opérateur génèrent donc des variations sur la diversité taxonomique et les estimations des abondances, plus ou moins marquées, en fonction des habitats. La variabilité de ces deux paramètres se propage dans les indices. C'est pourquoi nous proposons de mesurer l'effet de l'opérateur et son expérience associé aux variations de la diversité taxonomique (exclusion de taxons), mais également l'effet des pratiques en laboratoire associé aux estimations des abondances, par le biais d'une expérimentation virtuelle permettant d'altérer des listes faunistiques.

Il est à noter que, dans la littérature, de précédentes recherches sur l'IBGN ont déjà pu montrer que les abondances sont dépendantes des habitats échantillonnés et que l'effet de la variabilité de la richesse taxonomique était modéré, du fait de la transformation en classe de variété (Beisel, et al., 1998a). Cependant, la combinaison de différents échantillons, correspondant à différents habitats, conduisait à une variation de l'indice IBGN, l'incertitude sur la note pouvait varier de 1 à 3 points (Beisel, et al., 1998a). De plus, Ramade stipule que « le nombre d'espèces rares conditionne la diversité du peuplement » (Ramade, 2005), il n'est donc pas étonnant de voir que de nombreux auteurs se sont déjà intéressés à l'influence de la prise en compte des taxons rares et/ou peu abondants dans le calcul d'indices et ont montré leur importance ou leur redondance (Cao, et al., 1998; Arscott, et al., 2006). A l'inverse, les effets des erreurs des estimations des abondances sur les indices ou l'évaluation des rivières n'ont pas vraiment été traités comme tels. Jusqu'à présent la problématique liée aux abondances a plutôt été envisagée sous l'angle des différentes techniques d'échantillonnage ou de sous échantillonnages (Barbour, et al., 1999; Doberstein, et al., 2000; Lorenz, et al., 2004; Ligeiro, et al.,

2013) lors du tri au laboratoire, ou sous l'angle de la structure et de la complexité des habitats végétaux, ou encore de la distance entre certains habitats (Thomaz, et al., 2008; Burton, et al., 2002; McAbendroth, et al., 2005; Trigal-Domínguez, et al., 2009; Taniguchi, et al., 2003; Mormul, et al., 2011).

Le poids de la variabilité de la diversité taxonomique et des sous-estimations des abondances dans les indices français (IBG DCE, I2M2) et sa conséquence sur l'évaluation de la qualité des rivières représentent un enjeu pour la compréhension de l'agrégation des incertitudes associées à ces deux facteurs et l'éventuelle compensation des taxons entre substrats.

Pour tester l'impact de l'exclusion des taxons rares ou peu abondants et les erreurs d'estimations d'abondance des taxons, nous nous appuyons sur un algorithme modifiant aléatoirement les listes faunistiques réelles en fonction de différents taux d'erreurs. Trois types d'erreurs peuvent être générés : l'erreur résultant de l'exclusion des taxons rares (R), l'erreur résultant de la modification des estimations des abondances (AB) et ces deux erreurs cumulées (R+AB).

Le but de notre recherche est de tester la sensibilité aux deux expérimentations virtuelles des deux indices français (IBG DCE et I2M2), et de leurs métriques, et d'observer si les erreurs produites sont négligeables ou significatives.

Titre 2 : Matériel et méthodes

Nous présentons ici la méthode virtuelle : les différentes matrices conçues pour réaliser nos différents tests virtuels et le fonctionnement de l'algorithme, le jeu de données utilisé, ainsi que les calculs des différentes métriques et les différentes méthodes statistiques choisies pour mesurer les effets des erreurs testées.

Section 1. Métadonnées

Chacun des tests repose sur une matrice de seuils. La réflexion pour définir les valeurs des

seuils a été menée par substrat pour chacun des deux tests (R et AB) réalisés en fonction de différents critères.

Les seuils attribués pour définir la valeur maximale à partir de laquelle nous ne considérons plus un taxon comme rare ou peu abondant ont été obtenus par le biais de notre expérimentation précédente, fondée sur la répétition d'échantillonnage de substrats. Nous avons, pour chaque substrat, identifié le nombre de taxons présents uniquement sur une seule répétition de substrat. Les valeurs obtenues pouvaient osciller entre 2 et 8 individus en fonction d'un substrat, de ce fait dans un souci d'homogénéisation, nous avons ainsi fixé qu'un taxon pouvait être considéré comme peu abondant ou rare en dessous de 5 individus, quelque soit le substrat considéré. Cette valeur est également le seuil retenu par Nijboer pour identifier les taxons considérés comme très rares (Nijboer & Schmidt-Kloiber, 2004). Ainsi, l'algorithme que nous avons conçu recherche les taxons dont l'abondance est inférieure ou égale à 5 individus et, en fonction de la probabilité d'erreur introduite, supprime tout ou partie des taxons concernés (Tableau 3). Ce test simule les erreurs pouvant être induites par l'inexpérience d'un opérateur.

Le seuil choisi pour tester l'impact de la variabilité des abondances a été défini en prenant en considération la norme XP T90-388 (AFNOR, 2010). Celle-ci stipule que, si l'identification des taxons au genre ne peut pas être réalisée à l'œil nu lors de l'opération de tri, 40 individus doivent être sortis pour être identifiés afin de réaliser un sous échantillonnage, mais également que, lorsque l'abondance d'un taxon identifié au genre est supérieure à 40, il est possible de réaliser une estimation de l'abondance de ce taxon. Nous avons donc gardé la limite de 40 individus comme valeur seuil. L'algorithme recherche les taxons dont les abondances sont strictement supérieures à 40 individus et, en fonction de la probabilité d'erreur introduite, réalise une transformation des abondances en les réduisant à 40 individus. Nous avons décidé de réduire les abondances à une valeur fixe car nous sommes partis du postulat que si l'opérateur avait pu réaliser une estimation de l'abondance des taxons, il était sûr de sa détermination et avait donc compté au moins les individus jusqu'à la valeur prescrite par la norme (Tableau 3). Ce test simule les erreurs produites par les pratiques en laboratoire. Celui-ci sous-estime brutalement les abondances et pourrait permettre de tester la sensibilité des métriques de l'indice I2M2 aux dégradations d'abondance, ainsi que de mesurer l'importance de la précision d'un comptage exhaustif.

Tableau 3 : Extrait de la matrice des seuils utilisés par chaque type d'erreur en fonction des substrats

Tests / Substrats	Blocs	Bryophytes	Galets	Graviers	Pierres
Seuil taxon rare	5	5	5	5	5
Seuil taxon abondant	41	41	41	41	41
Nouvelle valeur des abondances	40	40	40	40	40

Afin de réaliser une modification aléatoire des listes faunistiques sélectionnées, nous avons fixé des probabilités d'erreurs différentes pour chacun des deux types d'erreurs. Ces probabilités d'erreurs ont été définies par le biais de l'étude menée sur la variabilité intra-substrat et, notamment, à l'aide des coefficients de variation (CV) calculés sur les métriques pouvant être reliées aux deux paramètres que nous testons ici. Les coefficients de variation retenus concernent la richesse taxonomique au genre et les abondances. Ainsi, pour déterminer ces valeurs, nous avons choisi de prendre les CV minimaux, moyens et maximaux, tous substrats confondus. Pour l'erreur associée à l'exclusion des taxons (R), nous avons donc obtenu comme valeur minimale 10%, valeur moyenne 20% et valeur maximale 40%. Pour l'erreur associée aux estimations d'abondances (AB) : la valeur minimale était égale à 25%, la valeur moyenne à 50% et la valeur maximale à 85%. Enfin pour les deux types d'erreurs cumulées (R+AB) nous avons opté pour une combinaison des probabilités d'erreurs ; ainsi il n'y a pas seulement trois séries de tests mais neuf (10% et 25%, 10% et 50%, 10% et 85%, 20 et 25%, 20 et 50%, 20 et 85%, etc.)

Section 2. L'algorithme

Nous avons implanté, à l'aide du langage Ruby, un algorithme qui permet de modifier aléatoirement une liste faunistique complète et validée, composée de douze listes élémentaires associées chacune à un habitat. Nous avons choisi ce type de langage de programmation car il est adapté à la conception de scripts. Il permet une manipulation très facile de fichiers de données sous forme de matrice. En considérant la matrice de seuils (Table 8) et une liste faunistique donnée, l'algorithme peut de façon aléatoire : (1) supprimer les taxons rares ou peu abondants de la liste faunistique en accord avec les seuils définis (2) baisser l'estimation des abondances des taxons en accord avec les seuils définis et (3) combiner les deux types d'erreurs. Nous allons

maintenant présenter un exemple de données fictives consistant en une liste faunistique basée, sur uniquement quatre habitats (Tableau 4) et une matrice de seuils associée (Tableau 3).

Tableau 4 : Exemple de liste faunistique validée \mathcal{L}

	Graviers	Pierres	Racines	Litières
Hydropsyche	56	23	7	0
Chironomidae	256	4	13	54
Gammarus	153	108	2	23
Baetis	26	41	9	3
Oligochètes	43	20	3	13

Quelque soit l'habitat, nous considérons un taxon comme rare ou peu abondant si son abondance est inférieure ou égale à 5. A l'inverse, un taxon est identifié comme abondant lorsque son abondance est supérieure ou égale à 41. Dans cet exemple, 108 *Gammarus* ont été comptés dans les pierres et 26 *Baetis* dans les graviers (Tableau 4).

Pour des raisons de clarté, nous avons décidé de synthétiser dans le Tableau 5 les différentes notations utilisées ci-dessous.

Tableau 5 : Descriptions des notations utilisées pour présenter le fonctionnement de l'algorithme

Notations	Descriptions
\mathcal{L}	Liste faunistique
$\mathcal{T} \in \mathcal{L}$	Un taxon appartenant à la liste faunistique
$S \in \mathcal{L}$	Un substrat appartenant à la liste faunistique
Abondance (\mathcal{T}, S)	Abondance du taxon \mathcal{T} dans le substrat S
Abondant $_{\theta}(S)$	Seuil des taxons abondants pour un substrat S
Rare $_{\theta}(S)$	Seuil des taxons rare pour un substrat S
\mathcal{P}_{AB}	Probabilité de modifier l'abondance d'un taxon abondant
\mathcal{P}_R	Probabilité de supprimer un taxon rare

En fonction d'une liste faunistique \mathcal{L} , un taxon $\mathcal{T} \in \mathcal{L}$, un substrat $S \in \mathcal{L}$ et de la matrice des seuils, trois cas sont possibles :

1. Si l'abondance (\mathcal{T}, S) \leq Rare $_{\theta}(S)$, le taxon est considéré comme un taxon rare ou peu abondant ;
2. Si l'abondance (\mathcal{T}, S) \geq Abondant $_{\theta}(S)$, le taxon est considéré comme un taxon abondant ;
3. Si Rare $_{\theta}(S) <$ Abondance (\mathcal{T}, S) $<$ abondant $_{\theta}(S)$, le taxon est considéré comme un taxon non concerné pas les erreurs choisies.

Nous allons principalement nous focaliser sur les cas 1 et 2. En effet, le but est de modifier les abondances des taxons abondants et de supprimer les taxons rares selon une probabilité (voir Tableau 5). Le type d'erreur et la probabilité sont deux paramètres de l'algorithme fixés par l'utilisateur. Ainsi, pour chaque taxon \mathcal{T} dans une liste faunistique \mathcal{L} :

1. si nous voulons sous-estimer les abondances des taxons \mathcal{T} pour un substrat S , la probabilité de changer l'abondance (\mathcal{T}, S) en la valeur arbitraire de 40, est donnée par \mathcal{P}_{AB} qui est égale à 25, 50 ou 85% ;
2. si nous voulons supprimer les taxons \mathcal{T} rares pour un substrat S , la probabilité de changer l'abondance (\mathcal{T}, S) en 0 est donnée par le paramètre \mathcal{P}_R qui correspond à 10, 20 ou 40%.

La probabilité d'erreur introduite dans l'algorithme permet de simuler un taux d'erreur moyen commis par l'opérateur, i.e. celui-ci a un certain pourcentage de chances de se tromper. Ainsi, le taux d'erreur réel peut varier et être supérieur ou inférieur à la probabilité d'erreur théorique. L'algorithme et la multiplicité des tests permettent de simuler une probabilité d'erreur moyenne. Plus les expérimentations seront répétées, plus la valeur du taux d'erreur réel moyen tendra vers la probabilité d'erreur théorique.

Le Tableau 6 est un exemple de liste faunistique modifiée par les deux erreurs cumulées (R+AB) à partir de la liste faunistique \mathcal{L} présentée dans le Tableau 4 et sur les seuils retenus (Tableau 3) et une probabilité d'erreur \mathcal{P}_R de 10% de \mathcal{P}_{AB} de 50%. Les valeurs des abondances modifiées sont sur fond bleu clair. Dans cet exemple, certains taxons abondants ont vu leurs abondances décroître, comme les *Chironomidae* et les *Oligochètes* dans les graviers et les *Baetis* dans les pierres, et un taxon rare a été supprimé : les *Baetis* dans les litières.

Tableau 6 : Exemple de liste modifiée basée sur la liste faunistique validée \mathcal{L}

	Graviers	Pierres	Racines	Litières
Hydropsyche	56	23	7	0
Chironomidae	40	4	13	54
Gammarus	153	108	2	23
Baetis	26	40	9	0
Oligochètes	40	20	3	13

Section 3 : Les expérimentations

Nous avons utilisé un jeu de données composé de 25 listes faunistiques issues du réseau de surveillance des rivières et de la base de données d'Aquabio (Bureau d'études spécialisé dans l'expertise des données hydrobiologiques). Pour étudier l'impact des erreurs testées en relation avec la qualité écologique des rivières, nous avons sélectionné cinq listes par classe de qualité (très bonne, bonne, moyenne, médiocre et mauvaise).

Les listes faunistiques choisies proviennent de sites situés dans les régions ouest, centre et nord-est de la France (Figure 24). Trois hydroécorégions (HER) sont concernées par notre zone d'étude : le Massif Armoricain (12), le Massif Central Nord (21) et l'Alsace (18). Les listes faunistiques ont été obtenues en suivant les règles des normes XP T90-333 et XP T90-388 (AFNOR, 2009; AFNOR, 2010) et validées par des experts internes à Aquabio.

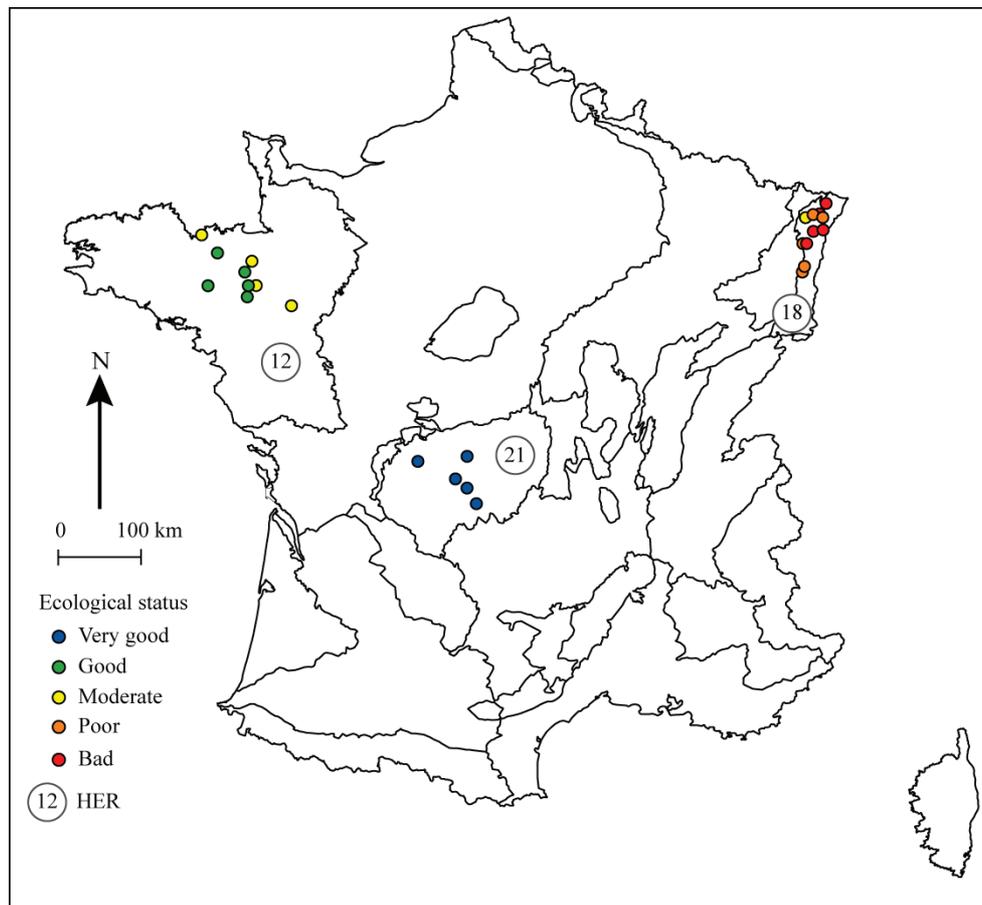


Figure 24: Localisation des 25 sites sélectionnés en France

Nous avons donc Q qualité écologique (très bonne, bonne, moyenne, médiocre et mauvaise) et n sites ($n=5$) par classe de qualité écologique q . Nous avons généré P répétitions ($P=20$) pour chaque type d'erreur (R, AB ou R+AB) et probabilité d'erreur (\mathcal{P}_R et \mathcal{P}_{AB}). Ceci nous permet de déterminer l'impact des erreurs engendrées par R, AB et R+AB. Nous avons répété chaque test (R, AB, et R+AB) 20 fois pour chaque probabilité d'erreur (pour R : 10, 25 et 40% et pour AB : 20, 50 et 85%, pour R+AB 9 combinaisons possibles des taux d'erreurs associés à R et AB, soit un total de 15 probabilités d'erreurs testées). Nous avons ainsi obtenu 1500 nouvelles listes erronées pour les deux premiers tests (R et AB, 25 listes \times 3 probabilités d'erreur \times 20 répétitions) et 4500 listes erronées pour le dernier test (R+AB, 25 listes \times 9 probabilités d'erreur \times 20 répétitions). Chaque liste erronée est caractérisée par une liste faunistique \mathcal{L} , sa classe de qualité associée q , mais aussi par la probabilité d'erreurs ou combinaison de probabilités d'erreurs introduites (\mathcal{P}_{AB} , \mathcal{P}_R ou \mathcal{P}_{R+AB}) et le type de modification (R, AB ou R+AB).

Section 4. Analyses des données

Pour étudier les incertitudes impactant les métriques, les indices IBG DCE et I2M2 ainsi que l'évaluation de la qualité écologique, nous avons utilisé une analyse des distances ainsi qu'un test de performance basé sur les Z-scores (cf. Partie 2, Chapitre 1, Titre 1, §2).

§1. Calcul des métriques de l'IBG DCE et de l'I2M2

Les listes sélectionnées pour réaliser nos modifications aléatoires ont été obtenues à l'aide des normes expérimentales de l'IBG DCE (AFNOR, 2009; AFNOR, 2010).

Les métriques ont été calculées uniquement pour le test concernant l'exclusion des taxons rares ou peu abondants (R), puisque seule cette erreur peut affecter l'IBG DCE.

Nous avons ainsi obtenu, pour chaque liste modifiée selon R, le groupe indicateur faunistique (GI) qui correspond au taxon indicateur le plus polluosensible (AFNOR, 2004) représenté par au

moins 3 ou 10 taxons en fonction de la famille, la richesse taxonomique (VT) qui est égale au nombre de taxons sortis et identifiés au niveau requis (famille) par la norme (AFNOR, 2010) et la classe de variété (VC) associée à la richesse taxonomique (VT).

Nous avons calculé les indices I2M2 (Mondy, et al., 2012), les cinq métriques ainsi que les ratios de qualité écologiques ou Ecological Quality Ratio (EQR), associés à chaque liste modifiée, pour les trois types d'erreurs expérimentés : l'exclusion des taxons rares ou peu abondants (R), les erreurs d'estimation des abondances (AB) et les deux erreurs cumulées (R+AB). Nous avons utilisé la version du script R du 03 février 2014 (Mondy C.P. & Usseglio-Polatera P., 2014, version fournie par les auteurs). Nous ne détaillerons ici que le mode de calcul des cinq métriques de l'I2M2, le calcul de l'indice I2M2 et des EQR étant présentés par Mondy (Mondy, et al., 2012)

Les cinq métriques de l'I2M2 sont les suivantes :

L'indice de diversité de Shannon H' , (Shannon, 1948) calculé sur les habitats les plus biogènes (phases A et B) avec :

$$H' = - \sum_{i=1}^n (P_i \times \ln P_i) \quad (19)$$

Où P_i correspond à la proportion de l'effectif du taxon i par rapport à l'effectif total des n taxons.

L'Average Score Per Taxon ASPT, (Armitage, et al., 1983) calculé sur les habitats dominants (phases B et C), correspond au niveau de polluo-sensibilité moyen de l'assemblage des macroinvertébrés benthiques (au moins un individu) et varie de 0 à 10.

La fréquence relative des taxons polyvoltins (PS) (i.e. avec au moins deux générations successives au cours d'une année) (Tachet, et al., 2010) est calculée sur l'ensemble des listes faunistiques (phases A, B et C).

La fréquence relative des taxons ovovivipares (OS) (i.e. dont la totalité du développement embryonnaire s'effectue dans l'abdomen de la femelle, l'éclosion des œufs suivant immédiatement leur expulsion dans le milieu aquatique) est calculée sur l'ensemble des listes faunistiques (phases A, B et C).

Les fréquences relatives des taxons polyvoltins et ovovivipares sont calculées sur l'ensemble du prélèvement selon la formule suivante (Usseglio-Polatera, et al., 2000) :

$$M = \frac{(\sum_i^S m_i \times \ln(q_i + 1))}{\sum_i^S \ln(q_i + 1)} \quad (20)$$

Où M correspond à la fréquence relative d'utilisation de la modalité considérée sur l'ensemble du prélèvement, S le nombre total de taxons, m_i la fréquence relative d'utilisation de la modalité m du trait biologique considéré pour le taxon i , et q_i l'abondance du taxon i dans l'assemblage faunistique considéré.

La richesse taxonomique (TAX) correspond au nombre de taxons identifiés au niveau systématique requis par la norme AFNOR XP T90-388 au moins représenté par un individu dans l'ensemble du prélèvement (phases A, B et C).

§2. Analyse des distances

Pour chacune des simulations p , sur une liste faunistique i , nous avons calculé la distance $D_{p,i}$ entre la note simulée p_i et la note de référence u_i .

$$D_{p,i} = p_i - u_i \quad (21)$$

Nous avons considéré la dispersion ds_i pour la liste i à savoir la différence entre la note simulée p et la note de référence u à l'aide de la formule suivante :

$$ds_{i,q} = \sqrt{\sum_{p=1}^{|P|} (D_{p,i})^2} \quad (22)$$

La moyenne de la dispersion DS_q est considérée pour n listes faunistiques i par classe de qualité écologique q , telle que:

$$DS_q = \frac{\sum_{i=1}^n ds_{i,q}}{n} \quad (23)$$

Le coefficient de variation CV_q est le ratio entre DS_q et la moyenne des notes de référence u_i , pour n listes originales, par classe de qualité q . Il est exprimé en pourcentage.

$$CV_q = \frac{DS_q}{\frac{1}{n} \sum_{i=1}^n u_i} \times 100 \quad (24)$$

Selon Gendron & Martin (2004), quand le coefficient de variation est inférieur à 16% : la variation observée est faible et l'estimation de la moyenne est fiable; quand le coefficient de variation se situe entre 16 et 33,3% : la variation est considérée comme important et la moyenne peut contenir des erreurs ; enfin quand le coefficient de variation est supérieur à 33,3% la variation est forte et la moyenne n'est pas fiable.

§3. Calcul des Z-scores

Le calcul des Z-scores est une méthode classique utilisée pour évaluer la performance d'un opérateur lors d'essais inter-laboratoire.

Dans notre cas, pour chaque probabilité d'erreur, nous avons considéré P répétitions, pour n rivières i , par classe de qualité écologique q .

Un Z-score i est calculé pour chaque répétition p : il est le ratio entre les $D_{p,i}$ et l'écart type, σ_i , de la note p pour P répétitions.

$$\sigma_i = \sqrt{\frac{1}{n} \sum_{i=1}^n (p - \bar{p})^2} \quad (25)$$

(26)

$$Z - score_{p,i} = \frac{D_{p,i}}{\sigma_i}$$

Cette méthode nous permet d'obtenir une valeur qui met en évidence la relation entre la note erronée et la note de référence. Quand le Z-score est inférieur à 2, les erreurs n'ont pas d'impact significatif sur la note obtenue. Quand il est situé entre 2 et 3, les erreurs ont un effet significatif. S'il est supérieur à 3, les erreurs ont un effet critique sur la note (AFNOR, 2005). Le signe négatif ou positif du Z-score indique soit une perte soit un gain de points.

Titre 3 : Résultats

Nous présentons ici les résultats obtenus en fonction des trois types d'erreurs (R, AB et R+AB) appliquées. Tout d'abord, nous montrons comment l'IBG DCE a répondu à l'erreur associée à l'exclusion des taxons uniquement (R). En effet, seule cette erreur peut avoir un impact, car le calcul de l'indice ne prend en compte que le nombre de familles et le groupe indicateur faunistique. Ensuite, nous présentons la réponse de l'IZM2, de ses métriques et des EQR associés aux métriques aux trois types d'erreurs (R, AB et R+AB). Enfin, nous observons comment ces différents tests peuvent traduire la compétence de l'opérateur et comment celle-ci peut affecter les indices.

Section 1. Robustesse de l'IBG DCE

La Figure 25 présente le nombre de déclassements après l'exclusion des taxons rares (R) par classe de qualité écologique et par probabilité d'erreur. Le nombre de déclassements augmente avec la probabilité d'erreur et nous observons des déclassements uniquement pour les classes médianes (bonne, moyenne et médiocre), la classe mauvaise ne pouvant évidemment pas être déclassée.

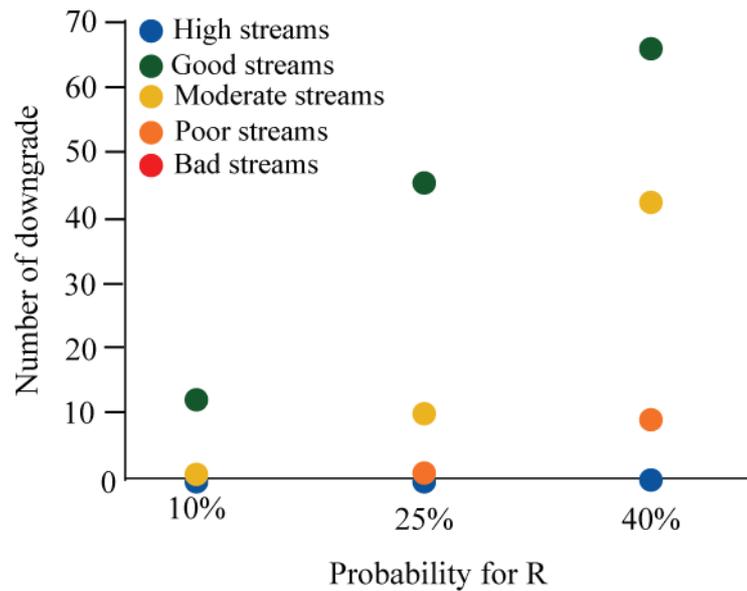


Figure 25: Proportion de déclassements de l'IBG DCE (en %) pour chaque classe de qualité écologique (la couleur du point correspond à la classe de qualité originale : très bonne - bleu, bonne - vert, moyenne -jaune, médiocre - orange et mauvaise - rouge) après l'exclusion des taxons rares ou peu abondants (R) pour chaque probabilité d'erreur (en %) (Nombre de sites observés $n=25$ où $n=5/\mathcal{P}_R$, 100 tests/ \mathcal{P}_R). Les seuils de classes de qualité sont définis par le Ministère de l'Ecologie (Ministère de l'Ecologie, du Développement durable et de l'Energie, 2012)

La classe de bonne qualité (en vert sur la figure) est la plus affectée par les oublis de taxons : le taux de déclassement est supérieur à 10%, quelque soit la probabilité d'erreur et pour une probabilité d'erreur de 40%, le risque de placer la rivière dans la classe de qualité inférieure est de 65%. La classe moyenne, elle aussi, est affectée par cette erreur. Pour une probabilité d'erreur de 40%, il existe un risque de 42% de se tromper de classe. Les oublis de taxons affectent également la classe de qualité médiocre mais dans une moindre mesure, puisque pour une probabilité d'erreur de 40%, le risque de déclassement n'atteint pas 10%. Nous constatons également que la classe de très bonne qualité n'est jamais déclassée.

Les figures 26 et 27 nous présentent les CV des notes de l'IBG DCE et ses métriques (GI, richesse taxonomique).

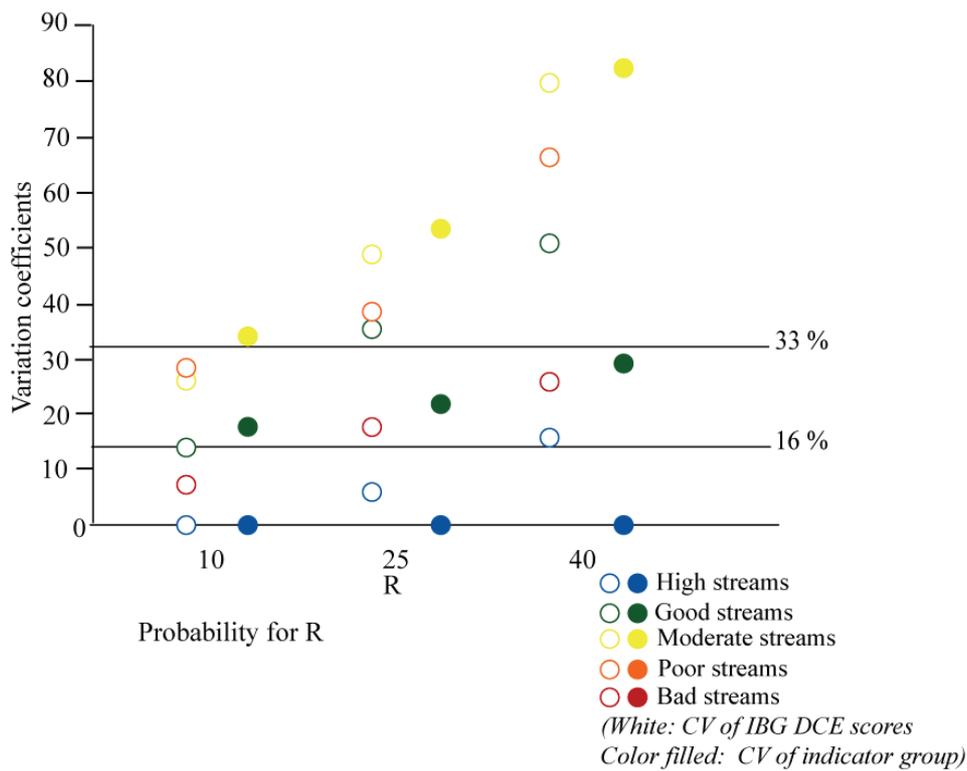


Figure 26: Coefficients de variation (CV) moyens des notes IBG DCE et du groupe indicateur (GI), pour chaque classe de qualité écologique et chaque probabilité d'erreur \mathcal{P}_R introduite pour tester l'impact de l'exclusion aléatoire des taxons rares ou peu abondants, obtenus avec l'algorithme présenté précédemment. Les moyennes des CV sont calculées sur les 100 tests effectués par probabilité d'erreur et pour chaque classe de qualité (nombre de sites observés : $n=25$, $n_i=5/\mathcal{P}_R$, 100tests/ \mathcal{P}_R /classe de qualité). $CV < 16\%$ pas d'impact significatif de l'erreur, $16\% < CV < 33,3\%$ erreur significative et $CV > 33,3\%$ erreur critique.

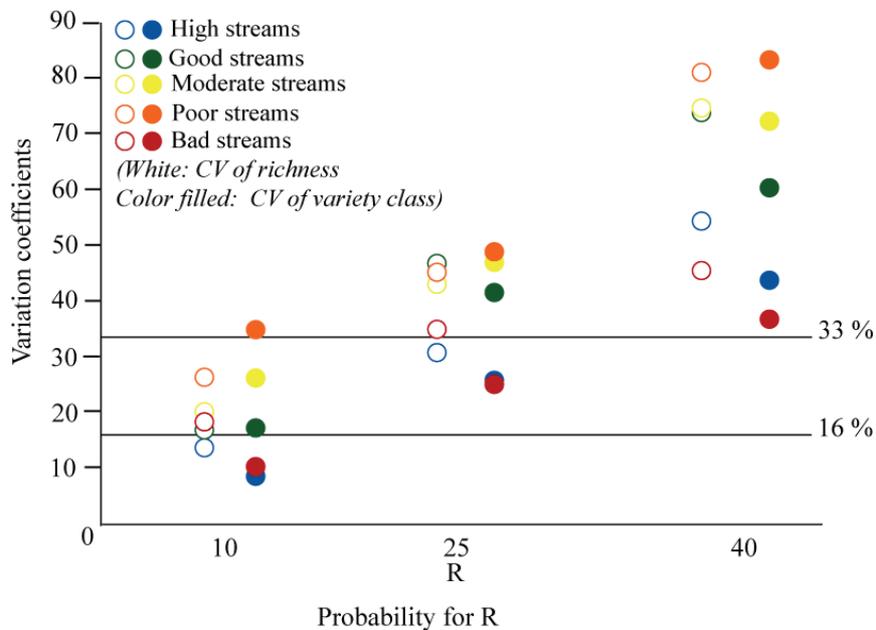


Figure 27: Coefficients de variation (CV) moyens des richesses taxonomiques (VT) et des classes de variété (VC), pour chaque classe de qualité écologique et chaque probabilité d'erreur \mathcal{P}_R introduite pour tester l'impact de l'exclusion aléatoire des taxons rares ou peu abondants obtenus avec l'algorithme présenté précédemment. Les moyennes des CV sont calculées sur les 100 tests par probabilité d'erreur pour chaque classe de qualité (nombre de sites observés : $n=25$, $n_i=5/\mathcal{P}_R$, 100tests/ \mathcal{P}_R /classe de qualité). $CV < 16\%$ pas d'impact significatif de l'erreur, $16\% < CV < 33,3\%$ erreur est significative et $CV > 33,3\%$ erreur critique.

Les coefficients de variation moyens (CV), calculés sur la note de l'IBG DCE, augmentent de manière linéaire avec la probabilité d'erreur introduite quelque soit la classe de qualité observée. Pour une probabilité d'erreur de 10% d'oublis de taxon, la variation observée pour toutes les classes de qualité écologique est inférieure à 33,3% et ; pour les cours d'eau de très bonne ou de mauvaise qualité, la variation est inférieure à 16% (Figure 26). Pour des probabilités d'erreur de 25 et 40%, la variation observée pour les cours d'eau de très bonne ou de mauvaise qualité est inférieure à 33,3%, tandis que pour les cours d'eau de qualité bonne, moyenne ou médiocre, les CV sont supérieurs à 33,3%. À nouveau, les classes médianes sont les plus affectées par les oublis de taxons.

Les CV calculés sur les groupes faunistiques indicateurs (GI) ne répondent pas comme les CV des notes. En effet, nous observons que, quelque soit la probabilité d'erreur introduite, la variation est nulle pour les classes de qualité très bonne, médiocre ou mauvaise. Tandis que pour les classes moyenne et bonne, les CV du GI fluctuent : ils sont supérieurs à 33,3% pour les cours d'eau de qualité moyenne et entre 16% et 33,3% pour les cours d'eau de bonne qualité (Figure 27).

Les CV calculés sur la richesse augmentent avec les probabilités d'erreur introduites : nous constatons à nouveau que les classes les plus affectées sont les classes médianes (Figure 27).

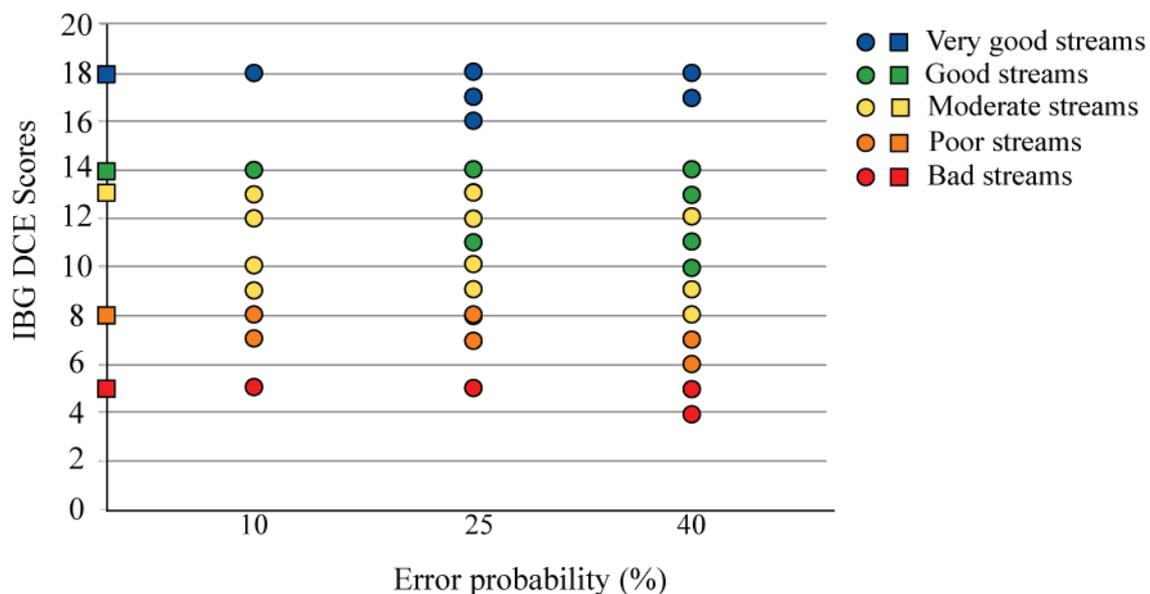


Figure 28 : Variations des notes de l'IBG DCE obtenues pour cinq rivières de classe de qualité différente (les couleurs des symboles correspondent à la classe initiale : très bonne - bleu, bonne - vert, moyenne - jaune, médiocre - orange et mauvaise - rouge ; symboles carrés = score original, symboles ronds = scores possibles obtenus après modifications aléatoires) après l'exclusion aléatoire des taxons rares ou peu abondants (R) obtenus avec l'algorithme présenté précédemment en fonction de chaque probabilité d'erreur \mathcal{P}_R (nombre de rivières concernées $n=5$, nombre de tests $20/\mathcal{P}_R$ /rivière)

La figure 28 présente la dispersion des notes pour cinq rivières de classes de qualité différentes ($n = 5$; 1 rivière par classe de qualité). Ces cinq rivières ont été sélectionnées car les écarts de notes observés y étaient les plus importants. Les résultats détaillés ici ne concernent que ces cinq rivières. Nous constatons que la perte de points existe principalement pour les classes médianes (bonne, moyenne et médiocre) et, que la perte de point augmente avec la probabilité d'erreur. Pour une probabilité d'erreur de 10% d'oublis de taxons rares seules les classes médianes sont concernées : les deux cours d'eau de classe très bonne et mauvaise voient leurs notes inchangées. Les notes des cours d'eau de bonne qualité et de qualité moyenne peuvent diminuer de 2 points et celle du cours d'eau de qualité médiocre d'un point.

Pour une probabilité d'erreur de 25%, les classes médianes (bonne, moyenne et médiocre) sont à nouveau les plus affectées. La note de bonne qualité peut diminuer de 3 points tandis que celle de moyenne qualité peut diminuer de 5 points. Nous constatons ici que la note du cours d'eau de très bonne qualité peut diminuer de 2 points.

La plus importante perte de points s'observe pour la probabilité d'erreur de 40% d'oublis de taxons rares. La dispersion est très étendue : les notes des cours d'eau de bonne qualité et de qualité moyenne peuvent décroître de 4 points, celle du cours de qualité médiocre de 2 points. La perte de points peut mener à un changement de classe de qualité. Ici les classes de qualité ne sont pas représentées car les limites de chaque classe dépendent de l'hydroécocorégion (HER) et du type des rivières (large, très large, etc.). Les rivières que nous avons sélectionnées sur ce graphique n'appartiennent ni au même type de cours d'eau ni à la même HER donc nous ne pouvons représenter les limites de classe. Cependant, il est à noter que pour les cours d'eau de classe bonne, moyenne ou médiocre, les pertes de points ont conduit à un déclassement, contrairement aux cours d'eau de très bonne qualité ou aux cours d'eau de mauvaise qualité, ces derniers ne pouvant pas être déclassés.

Section 2. Robustesse de l'I2M2

La Figure 29 présente la dispersion des notes de l'I2M2 obtenus pour cinq stations de qualité différente, après application des trois types d'erreurs (R, AB, R+AB) en fonction des probabilités

d'erreur associées. Ces cinq stations sont les mêmes que celles sélectionnées précédemment. Elles ne présentent donc pas les écarts les plus importants pour l'I2M2.

Ici, nous ne parlerons pas de changements de classe associés aux effets de l'inexpérience de l'opérateur et des pratiques en laboratoires car nous avons constaté des écarts de classe initiales entre l'I2M2 et l'IBG DCE. La qualité écologique des milieux s'est vu souvent être sous-estimée et déclassante par rapport à l'IBG DCE.

Nous constatons, de manière générale, que le test portant sur les oublis des taxons rares ou peu abondants (R) mène à une diminution des notes tandis que le test sur les modifications d'abondances (AB) conduit à une augmentation de l'I2M2. Lorsque les deux types d'erreurs sont cumulés (R+AB), nous observons un effet de compensation. La perte de points due au test des oublis de taxons (R) constatée auparavant serait atténuée par l'augmentation des notes observées avec le test des sous estimations des abondances (AB).

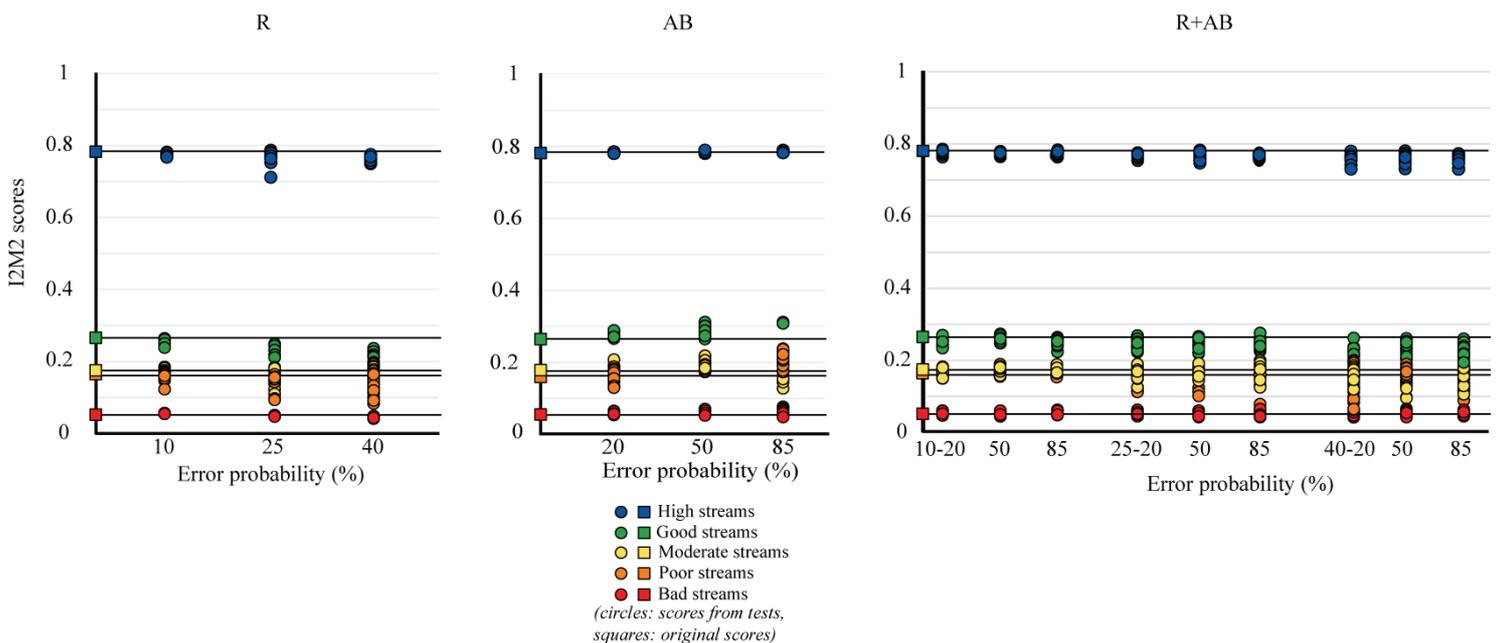


Figure 29: Variation des scores de l'I2M2 obtenus pour cinq rivières de classe de qualité différente (la couleur des symboles correspondant à la classe initiale : très bonne - bleu, bonne - vert, moyenne - jaune, médiocre - orange et mauvaise - rouge ; symboles carrés = score original, symboles ronds = scores après modifications aléatoires) après l'exclusion aléatoire des taxons rares ou peu abondants (R), la modification aléatoire des estimations des abondances (AB) et des deux modifications cumulées (R+AB) par le biais de l'algorithme présenté précédemment pour chaque probabilité d'erreur (nombre de rivières concernées $n=5$, nombre de tests $20/P$ /rivière)

Ainsi, nous constatons que le gain ou la perte de points se compte à partir de la troisième ou deuxième décimale et très rarement à la décimale près. La rivière de très bonne qualité voit son

indice varier de +/- 0.009 point, tandis que l'indice de la rivière de qualité médiocre peut varier de +/- 0.02 point pour le test portant sur l'exclusion des taxons rares ou peu abondants. Tout comme pour la dispersion des scores de l'IBG DCE, nous constatons que les écarts sont plus importants pour les classes médianes (bonne, moyenne et médiocre). Les écarts de notes augmentent avec les probabilités d'erreurs introduites pour chaque série de tests et sont, en général, plus importants pour l'exclusion des taxons rares (R) que pour les modifications des estimations des abondances (AB) du moins en ce qui concerne les classes médianes.

Nous avons également noté que les sites de bonne qualité initiale pour l'IBG DCE étaient déclassés avec l'I2M2 avant même d'avoir procédé aux différents tests d'erreurs. L'indice I2M2 qui prend en compte des critères plus larges (pressions hydromorphologiques et physico-chimiques, 5 métriques : ASPT, TAX, H, PS et OS) que ceux de l'IBG DCE, serait plus fidèle et sensible que l'IBG DCE. C'est pourquoi nous n'avons pas dénombré les changements de classes pour l'I2M2. De plus, à ce jour les bornes de classes de qualité de l'I2M2 provisoires, reposent sur des opérations d'intercalibration européennes et le système ne devrait être stable qu'à la fin 2015.

Les coefficients de variation moyens sont donnés pour chacune des métriques de l'I2M2 (ASPT, TAX, indice de Shannon H, fréquence des individus polyvoltins PS et fréquence des individus ovovivipares OS), pour chaque probabilité d'erreur et type d'erreur (R, AB ou R+AB) et par classe de qualité écologique (Figure 30).

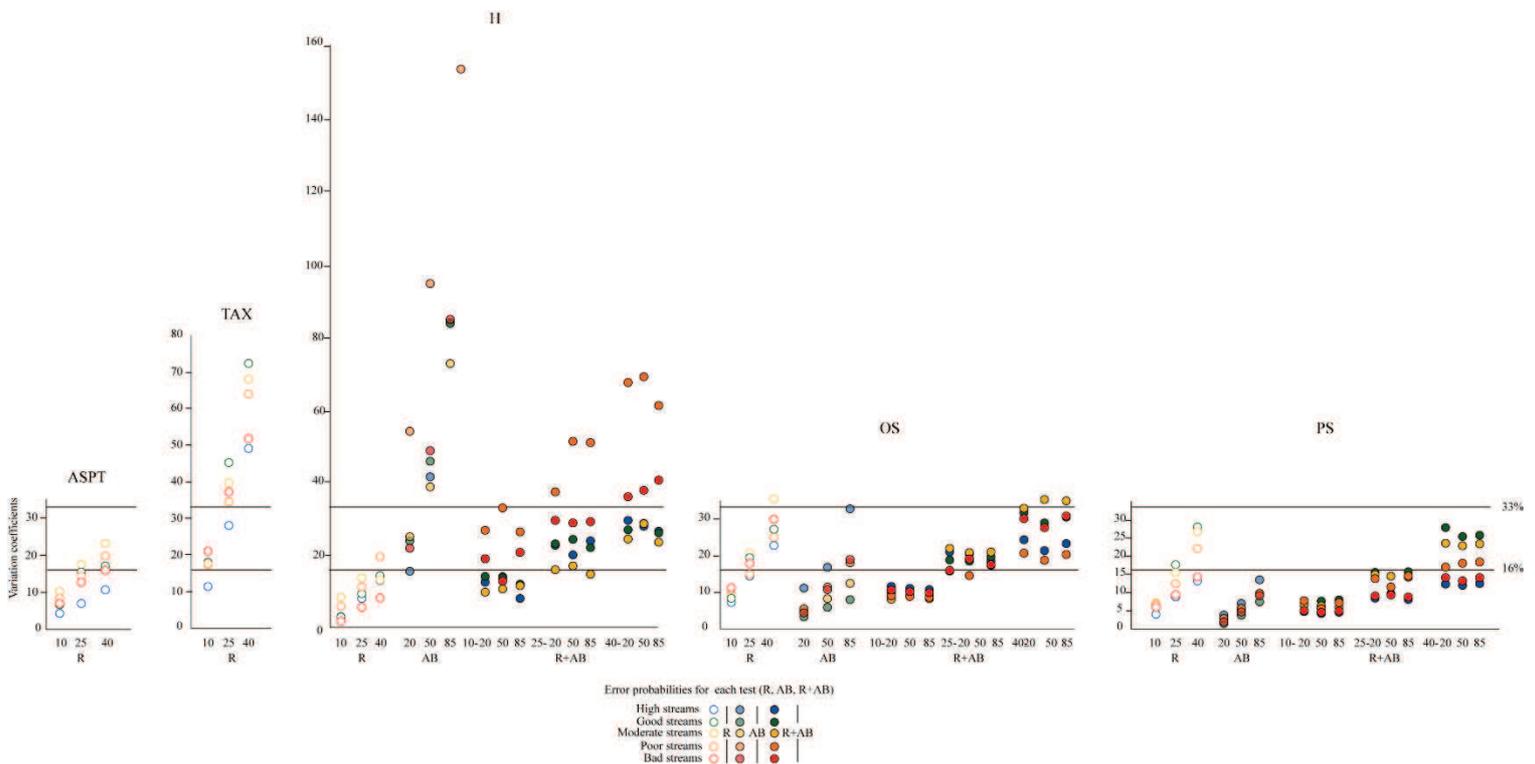


Figure 30: Coefficients de variation (CV) moyens des cinq métriques de l'I2M2 (ASPT, TAX, H, OS et PS) pour chaque classe de qualité écologique et chaque probabilité d'erreur introduite pour tester l'exclusion aléatoire des taxons rares ou peu abondants (R), la modification aléatoire des estimations des abondances (AB) et des deux modifications (R+AB) obtenus par l'algorithme utilisé. Les moyennes des CV sont calculées sur les 100 tests par probabilité d'erreur pour chaque classe de qualité ($n=5/\mathcal{P}$, $100\text{tests}/\mathcal{P}$ /classe de qualité). $CV < 16\%$: pas d'impact significatif de l'erreur, $16\% < CV < 33,3\%$: l'erreur est significative et $CV > 33,3\%$: erreur critique.

En général, nous constatons que les CV moyens augmentent avec la probabilité d'erreur, quelque soit le type d'erreur appliqué ou la classe de qualité. Les métriques des cours d'eau de très bonne et mauvaise qualités sont moins sensibles aux oublis de taxons et aux erreurs d'estimation des abondances que les classes médianes (bonne, moyenne et médiocre). Les cinq métriques ne répondent pas de manière similaire aux trois types d'erreurs. L'ASPT et les fréquences des individus polyvoltins (PS) et ovovivipares (OS) sont les trois métriques dont les coefficients de variation moyens sont globalement inférieurs à 33,3%, quelque soit la probabilité d'erreur et le type d'erreur. La richesse (TAX) et l'indice de Shannon (H) sont plus variables et les CV moyens peuvent être largement supérieurs à 33,3%.

Les CV moyens des deux métriques basées sur la diversité, l'ASPT et la richesse (TAX), ne sont donnés que pour les oublis de taxons (R) car l'abondance des taxons n'est pas prise en compte dans leur calcul. Les CV moyens pour l'ASPT oscillent entre 4,3% et 10% pour une probabilité d'erreur de 10%, entre 7,2 et 17,2% pour une probabilité d'erreur de 25% et entre 10,6 et 23,2%

pour une probabilité d'erreur de 40% d'oublis de taxons. Ces CV restent tous inférieurs à 33,3% : l'erreur est donc à surveiller mais n'est pas significative. Les CV moyens les plus faibles concernent les cours d'eau de très bonne qualité et les plus élevés correspondent aux cours d'eau de classe moyenne.

Les CV calculés sur les richesses taxonomiques varient entre 11,5 et 21,2% pour 10% d'erreur, entre 28,3 et 45,3% pour 25% d'erreur et entre 49 et 72% pour 40% d'erreur. Les cours d'eau de très bonne qualité sont les moins sensibles à l'exclusion aléatoire des taxons rares ou peu abondants et l'erreur est significative pour une probabilité d'erreur de 40%. Les cours d'eau de bonne qualité sont les plus sensibles aux oublis de taxons et sont les plus variables.

Les variations observées sur les indices de Shannon (H) diffèrent en fonction des types d'erreur. En effet, nous constatons que la variation est faible pour les oublis de taxons rares ou peu abondants. Les CV n'excèdent pas le seuil des 16% pour les probabilités d'erreur de 10 et 25%, et seules les stations de qualité médiocre voient leur CV dépasser les 16% : l'erreur mesurée n'est donc pas critique pour les métriques basées sur la diversité. Les variations sont plus importantes pour le type d'erreur concernant les modifications des estimations des abondances des taxons. Les sites de qualité médiocre ont toujours, en moyenne, des CV supérieurs à 33,3%. Nous constatons que les CV augmentent avec la probabilité d'erreur introduite et qu'ils sont très fortement supérieurs à 33,3%.

Les CV des fréquences des individus polyvoltins (PS) et ovovivipares (OS) répondent de façon analogue aux trois types d'erreur. La variation mesurée est plus faible pour le type d'erreur basé sur les modifications des abondances que pour celui basé sur les oublis de taxons. Les CV restent généralement inférieurs à 33,3% : l'erreur induite par ces trois types d'erreur n'est donc pas significative. Nous constatons que, pour l'erreur basée sur les oublis de taxons, les CV des métriques des traits biologiques, pour les cours de très bonne qualité, sont les moins élevés (entre 7,35% et 23% pour OS et entre 4,04 et 13,4 pour PS) et que les CV sont plus élevés pour les cours d'eau appartenant aux classes bonne, moyenne ou médiocre (entre 11,25 et 35,5 pour OS et entre 6,69 et 27,9% pour PS). Pour le type d'erreur concernant les estimations des abondances, nous constatons que les CV calculés sur les cours d'eau de très bonne qualité sont les plus élevés (11,14 à 32,8% pour OS et 3,82 et 13,53% pour PS) et que ceux calculés sur les cours d'eau de bonne qualité sont les moins élevés (3,27 à 8,07% pour OS et 1,5 et 7,5% pour PS).

Après avoir constaté une variation plus ou moins significative des valeurs des métriques nous

allons nous intéresser à la variation des ratios de qualité écologique (EQR). Les valeurs des EQR sont utilisées lors du diagnostic de l'état des rivières, d'où l'importance de mesurer leur variabilité liée aux erreurs de l'opérateur et intra-substrat.

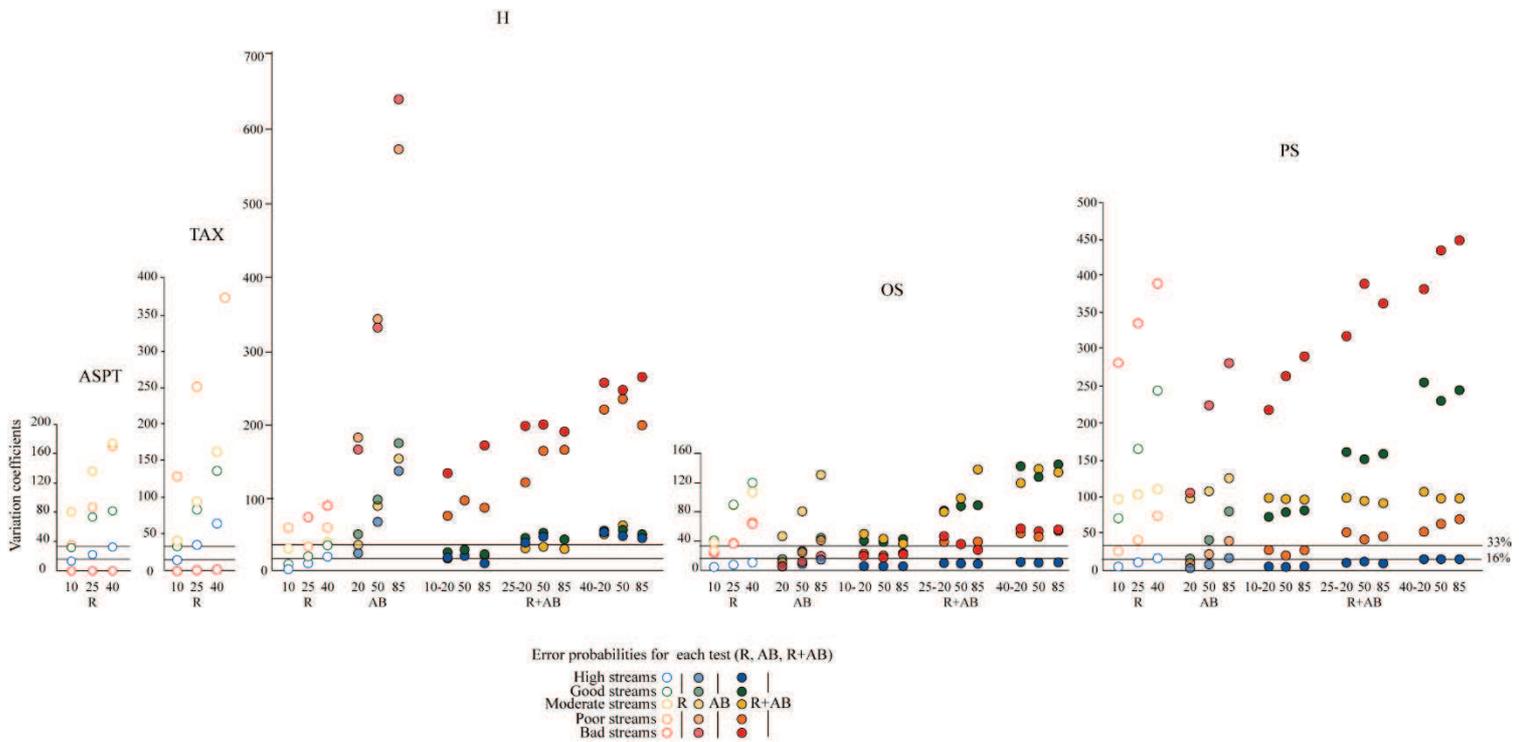


Figure 31: Coefficients de variation (CV) moyens des EQR des cinq métriques de l'I2M2 (ASPT, TAX, H, OS et PS) pour chaque classe de qualité écologique et chaque probabilité d'erreur introduite pour tester l'impact de l'exclusion aléatoire des taxons rares ou peu abondants (R), de la modification aléatoire des estimations des abondances (AB) et des deux modifications cumulées (R+AB) obtenus avec l'algorithme présente précédemment. Les moyennes des CV sont calculées sur les 100 tests par probabilité d'erreur pour chaque classe de qualité ($n=5/P$, $100\text{tests}/P/\text{classe de qualité}$). $CV < 16\%$: pas d'impact significatif de l'erreur, $16\% < CV < 33\%$: l'erreur est significative et $CV > 33\%$: erreur critique.

Pour rappel, les EQR ou ratios de qualité écologique, sont des valeurs permettant de rendre compte d'un écart à la référence et de comparer directement des valeurs de métriques pour tous les cours d'eau relevant du protocole normalisé.

Nous constatons que les CV des EQR ne se comportent pas de façon similaire suivant les métriques et n'ont pas la même sensibilité aux types d'erreur et probabilités d'erreurs (Figure 31). De manière générale, les CV des EQR sont plus élevés que ceux observés sur les valeurs des métriques brutes. Les CV de l'ASPT et des fréquences des individus polyvoltins et ovovivipares n'étaient pas supérieurs à 33,3%, indiquant aussi une erreur non significative. Tandis que les CV des valeurs des EQR, même pour de faibles probabilités d'erreur, peuvent être largement supérieurs à 33,3%. Les classes médianes (bonne, moyenne et médiocre) sont généralement plus variables que les classes

très bonne et mauvaise. D'autre part, nous pouvons également observer que, contrairement aux CV des métriques brutes, il est possible que les EQR ne varient pas et que les CV restent à 0.

Les CV de l'ASPT présentent une valeur nulle pour les stations de très mauvaise qualité et une valeur inférieure à 33,3% pour les stations de très bonne qualité (les CV sont compris entre 13 et 32,5%). Les stations de qualité moyenne ont des CV supérieurs 33,3%.

Considérant les CV calculés sur les EQR de la richesse taxonomique (TAX), nous constatons à nouveau qu'il n'y a pas de variation observée pour les stations de mauvaise qualité et que les stations de très bonne qualité ne sont pas sensibles à une probabilité d'erreur de 10% (CV=13,3%). Les CV calculés sur les stations de qualité moyenne (bonne, moyenne et médiocre) sont supérieurs à 33,3% quelque soit la probabilité d'erreur introduite.

Les CV des EQR de Shannon (H), de l'ovoviviparité (OS) et du polyvoltinisme (PS) sont sensibles aux trois types d'erreur mais ne répondent pas de manière analogue.

Pour l'indice de Shannon (H), nous constatons que la variation est plus importante pour l'erreur liée aux modifications d'abondances que pour celle fondée sur les oublis de taxons. Cette même remarque avait été formulée lors de l'observation de la variation des valeurs brutes de l'indice. Les CV sont très élevés et s'étendent sur un intervalle compris entre 30 et 678 % pour une probabilité d'erreur de 85%. Les stations de qualités très bonne ou moyenne ont des CV inférieurs à 33, 3% pour une probabilité d'erreur de 20%. Concernant les oublis de taxons, nous constatons que les CV, quelque soient les EQR considérés, ont des valeurs moins étendues et que les stations de très bonne et bonne qualité sont les moins sensibles avec des CV compris entre 6 et 23% pour les très bons cours d'eau, et entre 9 et 40% pour les bons cours d'eau. En revanche, les stations de qualité médiocre ou mauvaise ont des CV supérieurs à 60%. La variation observée des valeurs des CV est plus faible pour les stations de très bonne, bonne et moyenne qualités et pour des probabilités d'erreur de 10% d'oublis de taxons (R) ainsi que des probabilités d'erreur de 20, 50 et 85 % sur les estimations des abondances (AB).

Les CV calculés sur les EQR des traits biologiques répondent de manière analogue aux trois types d'erreur et, à nouveau, nous constatons une réponse similaire à celle observée pour les CV obtenus avec les valeurs brutes. En effet, il y a une sensibilité plus importante aux oublis de taxons qu'aux erreurs d'estimation des abondances (AB). Cela se traduit par des CV inférieurs à 33,3% pour les classes de qualité très bonne ou mauvaise pour l'ovoviviparité (OS) et les classes très bonne et médiocre pour le polyvoltinisme (PS).

Les variations des métriques et des EQR établies, il est intéressant d'observer la variation de l'indice I2M2 lui-même. La Figure 32 présente les coefficients de variation moyens obtenus pour chaque classe de qualité écologique, pour chaque type d'erreur et chaque probabilité d'erreur.

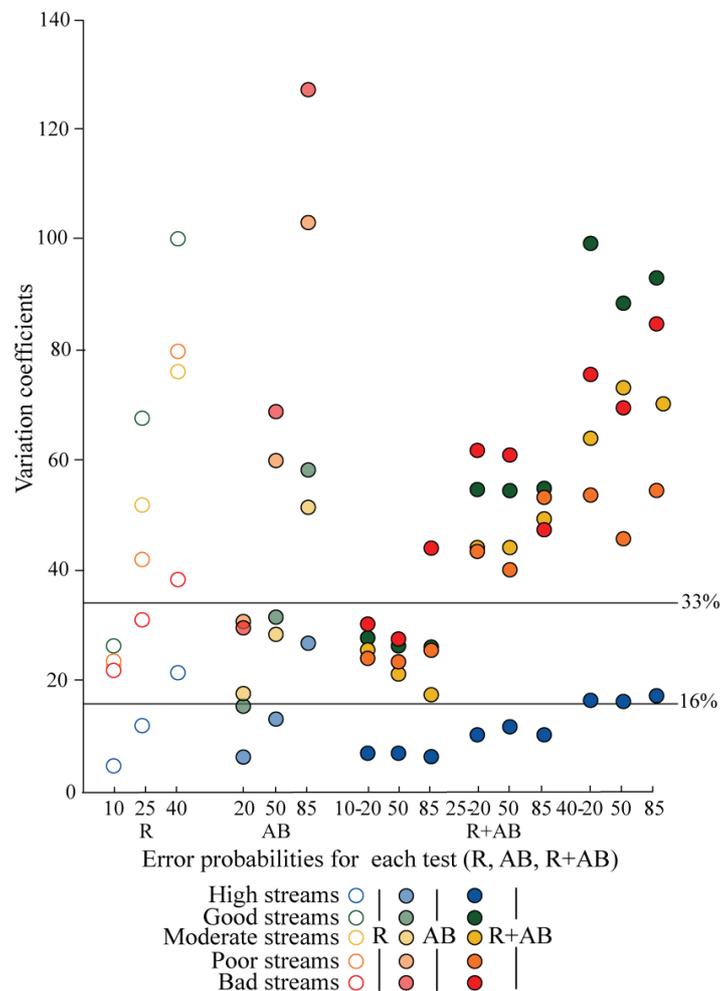


Figure 32 : Coefficients de variation (CV) moyens des scores I2M2 pour chaque classe de qualité écologique et chaque probabilité d'erreur introduite pour tester l'impact de l'exclusion aléatoire des taxons rares ou peu abondants (R), de la modification aléatoire des estimations des abondances (AB) et des deux modifications cumulées (R+AB) obtenus par l'algorithme présenté précédemment. Les moyennes des CV sont calculées sur les 100 tests par probabilité d'erreur pour chaque classe de qualité ($n=5/\mathcal{P}$, $100\text{tests}/\mathcal{P}/\text{classe de qualité}$). $CV < 16\%$: pas d'impact significatif de l'erreur, $16\% < CV < 33,3\%$: l'erreur est significative et $CV > 33,3\%$: erreur critique.

Nous pouvons ainsi constater l'effet des trois modifications sur le nouvel indice I2M2, mais également le comportement différent suivant les différentes classes de qualité.

Les CV moyens associés aux très bons cours d'eau ont une valeur inférieure à 33,3%, cela rejoint les faibles variations observées pour les métriques et les EQR. Nous constatons également que pour des probabilités d'erreurs inférieures à 40%, pour les oublis de taxons (R), la variation n'est

pas significative. Il en est de même pour des probabilités d'erreur inférieures à 85% pour les modifications des estimations d'abondance (AB).

Section 3. Performances de l'opérateur

Nous avons pu observer que dans la plupart des tests réalisés nous sommes en présence d'au moins une erreur significative voire critique. Les Z-scores nous permettent d'identifier l'impact des différentes erreurs en fonction des probabilités d'erreur introduites dans l'algorithme. Les trois tests réalisés montrent un effet sur les valeurs de l'I2M2.

Tableau 7 : Z-scores minimums, maximums et moyens obtenus sur les notes de l'I2M2 après les trois différentes modifications (R, AB, R+AB) par le biais de l'algorithme présenté précédemment pour chaque classe de qualité (très bonne, bonne, moyenne, médiocre et mauvaise) (moyenne des valeurs absolues). Valeurs minimale et maximales obtenues sur les cours d'eau par classe d'état écologique, $n=5$ /taux d'erreur/modification pour 100 tests/P ; moyenne sur 100 tests/P/classe d'état. (Z-score < 2 : non significatif, $2 \leq$ Z-score \leq 3 : significatif, Z-score > 3 : erreur critique) En gras = valeur significative, en bleu clair = valeur critique.

Ecological status		Very good			Good			Moderate			Poor			Bad		
Error types	Error rates	min	Max	Avg	min	Max	Avg	min	Max	avg	min	Max	avg	min	max	avg
R	10%	0,01	-3,80	0,99	0,00	-4,48	1,41	0,01	-3,98	0,97	0,00	-4,64	0,96	0,00	-2,96	0,95
	25%	0,01	-4,71	1,48	0,34	-4,93	2,32	0,01	-6,92	1,77	0,00	-3,85	1,48	0,00	-4,01	1,16
	40%	0,05	-4,46	2,13	1,13	-6,08	3,13	0,08	-4,81	2,04	0,06	-5,19	2,19	0,00	-4,11	1,57
AB	20%	0,00	3,10	0,85	0,00	3,37	0,85	0,00	4,49	0,94	0,00	3,94	0,95	0,00	3,62	0,66
	50%	0,02	4,37	1,65	0,00	4,09	1,40	0,00	4,21	1,50	0,01	4,11	1,48	0,00	2,91	0,98
	85%	0,20	6,60	4,19	0,43	7,04	4,21	0,00	6,00	2,69	0,22	6,19	3,49	0,00	5,21	2,35
R+AB	10-20%	0,00	2,88	0,81	0,04	-3,82	1,13	0,01	-3,32	0,94	0,00	-3,05	0,87	0,00	3,39	0,64
	10-50%	0,01	-3,05	0,79	0,04	-3,62	1,29	0,01	-3,70	0,99	0,00	-2,89	0,82	0,00	-3,17	0,57
	10-85%	0,01	-3,16	0,86	0,03	-4,01	1,33	0,03	-3,41	0,90	0,00	3,43	0,81	0,00	3,74	0,65
	25-20%	0,01	-4,18	1,02	0,00	-3,81	1,59	0,02	-4,17	1,44	0,04	-3,02	0,90	0,00	2,81	0,84
	25-50%	0,01	-3,10	1,01	0,07	-4,25	1,71	0,12	-4,84	1,38	0,02	-3,29	0,77	0,00	2,98	0,71
	25-85%	0,06	-4,29	1,08	0,22	1,71	2,03	0,01	-4,79	1,39	0,00	-3,21	0,93	0,00	3,35	0,68
	40-20%	0,01	-3,88	1,27	0,03	-4,81	2,57	0,01	-5,73	1,95	0,04	-2,97	0,99	0,00	2,51	0,79
	40-50%	0,01	-3,67	1,17	0,14	-5,59	2,57	0,02	-4,64	1,89	0,00	-3,34	0,82	0,00	2,54	0,84
	40-85%	0,00	-4,24	1,40	0,09	-5,35	2,59	0,09	-5,04	1,68	0,00	-3,29	0,88	0,00	3,59	0,81

Le Tableau 7 montre que, pour chaque type d'erreur (R, AB et R+AB) et quelque soit la probabilité d'erreur, il y a des erreurs significatives ($2 \leq$ Z-score < 3) ou critiques (Z-score > 3). Les Z-scores moyens nous informent qu'il existe logiquement moins d'erreurs pour les faibles probabilités d'erreurs (R=10%, AB=20% et R+AB=10-20%, 10-50%, 10-85%) que pour les probabilités d'erreurs élevées (R=25%,40%, AB=50%,85%) ou les deux tests cumulés (R+AB=25-50%,25-85%,40-20%,40-50%,40-85%). Il est également intéressant de noter que les sous-estimations d'abondances entraînent plutôt des hausses du score de l'I2M2 (Z-scores positifs) et

les exclusions de taxons rares ou peu abondants, une diminution de la note de l'indice (Z-scores négatifs). Nous observons également que l'erreur cumulée (R+AB) entraîne une augmentation des notes de l'I2M2 pour les rivières de mauvaise qualité.

Le Tableau 8 nous renseigne plus précisément sur la présence d'erreurs significatives et critiques, obtenues sur la base des Z-scores, ainsi que sur leur part au sein de chaque état écologique, chaque type de modification et pour les différentes probabilités d'erreur.

Tableau 8 : Proportion totale des erreurs significatives (SE) et des erreurs critiques (CE) pour l'I2M2 calculée sur la base des Z-scores significatifs et critiques (Z-score < 2 : non significatif, $2 \leq$ Z-score \leq 3 : significatif, Z-score > 3 : erreur critique) obtenus par l'algorithme présenté précédemment pour chaque classe de qualité (très bonne, bonne, moyenne, médiocre et mauvaise), chaque type de modifications (R, AB et R+AB) et pour chaque probabilité d'erreur. Valeurs obtenues par le cumul des erreurs pour les 5 rivières par classe ($n=5/\text{classe}$, 100 tests/2)

Ecological status		Very good		Good		Moderate		Poor		Bad	
Error types	Error rates	SE	CE	SE	CE	SE	CE	SE	CE	SE	CE
R	10%	12	3	19	8	11	5	6	4	12	0
	25%	17	10	29	25	28	16	20	9	11	5
	40%	35	19	30	49	23	24	24	23	19	10
AB	20%	13	1	16	3	10	4	15	3	8	4
	50%	31	9	22	8	25	8	21	8	16	1
	85%	26	62	8	86	27	41	24	67	35	29
R+AB	10-20%	7	0	12	5	15	9	8	1	9	3
	10-50%	10	1	20	5	9	3	6	0	6	2
	10-85%	7	1	16	7	8	1	5	1	10	2
	25-20%	7	2	19	12	18	10	10	1	11	0
	25-50%	8	2	28	8	25	5	7	2	5	1
	25-85%	8	3	27	18	18	10	12	1	6	2
	40-20%	15	7	37	31	21	25	11	0	10	1
	40-50%	14	4	30	36	26	16	3	1	10	1
	40-85%	18	10	35	34	19	18	6	2	9	1

Nous constatons que dans la plupart des cas la proportion d'erreurs significatives (SE) est supérieure à celle des erreurs critiques (CE), mais également que la part d'erreurs non significatives domine pour les probabilités d'erreurs faibles (R=10% et 25%, AB= 20%, 50%). Plus la probabilité d'erreur introduite augmente, plus le nombre d'erreurs significatives et critiques est élevé.

Le Tableau 9 présente les Z-scores obtenus pour l'IBG DCE et ses métriques, nous pouvons ainsi

observer quelles sont les métriques les plus affectées par l'exclusion des taxons rares et ainsi apprécier sa propagation de l'incertitude liée aux oublis de taxons rares sur l'indice.

Tableau 9 : Z-scores minimums, maximums et moyens obtenus calculés sur les résultats de l'algorithme présenté précédemment sur les notes de l'IBG DCE et les métriques associées (GI, Richesse taxonomique (VT) et classe de variété VC) après l'exclusion aléatoire des taxons rares ou peu abondants (R) pour chaque classe de qualité (très bonne, bonne, moyenne, médiocre et mauvaise) (moyenne des valeurs absolues). Valeurs minimales et maximales obtenues sur les cours d'eau par classe d'état écologique, $n=5/\mathcal{P}$ /modification pour 100 tests/ \mathcal{P} ; moyenne sur 100 tests/ \mathcal{P} /classe d'état. (Z-score < 2 : non significatif, $2 \leq$ Z-score \leq 3 : significatif, Z-score > 3 : erreur critique) En gras = valeur significative, en bleu clair = valeur critique.

Ecological status	Probability	Very good			Good			Moderate			Poor			Bad		
		Min	Max	Avg	Min	max	Avg	Min	Max	avg	Min	Max	Avg	Min	max	avg
IBG DCE	10%	0	0	0	0	-3,25	0,36	0	-3,13	0,58	0	-4,47	0,77	0,00	-4,47	0,09
	25%	0	-4,47	0,16	0	-5,08	0,93	0	-4,62	1,34	0	-6,16	1,46	0,00	-4,47	0,16
	40%	0	-2,91	0,89	0	-4,01	2,16	0	-5,25	2,55	0	-4,96	1,96	0,00	-4,47	0,36
GI	10%	0	0	0	0	-2,17	0,16	0	-4,47	0,14	0	0	0	0	0	0
	25%	0	0	0	0	-1,96	0,18	0	-4,47	0,29	0	0	0	0	0	0
	40%	0	0	0	0	-2,44	0,39	0	-2,73	0,74	0	0	0	0	0	0
VT	10%	0	-3,67	1,00	0	-3,50	1,08	0	-3,85	0,90	0	-3,41	0,95	0	-4,47	0,40
	25%	0	-4,69	1,65	0	-5,29	2,30	0	-4,23	1,71	0	-4,36	1,43	0	-3,60	0,72
	40%	0	-5,36	2,54	0	-4,62	2,62	-0,91	-5,54	2,88	0	-5,50	2,15	0	-3,60	0,96
VC	10%	0	-4,47	0,25	0	-3,25	0,28	0	-2,44	0,56	0	-4,47	0,79	0	-4,47	0,09
	25%	0	-4,09	0,98	0	-5,08	1,06	0	-4,47	1,70	0	-6,16	1,51	0	-4,47	0,24
	40%	0	-4,25	1,94	0	-4,09	2,02	0	-6,50	2,86	0	-4,96	1,96	0	-4,47	0,37

Il existe une possibilité d'erreur significative ou critique quelque soient la classe de qualité observée ou la probabilité d'erreur introduite pour l'IBG DCE (Tableau 9). Nous constatons également que, si le GI ne connaît des erreurs significatives et critiques que pour les classes bonne et moyenne, la classe de variété liée à la richesse taxonomique (VT) est, quant à elle, plus sensible aux oublis de taxons. Ainsi nous trouvons au moins une erreur significative ou critique quelque soient la classe de qualité ou la probabilité d'erreur introduite. Le nombre d'erreurs augmente avec la probabilité d'erreur : nous constatons pour les classes médianes (bonne, moyenne et médiocre) que, pour 40% d'oublis de taxons rares ou peu abondants, la proportion d'erreurs significatives domine. L'exclusion des taxons rares ou peu abondants entraîne donc une diminution des notes de l'IBG DCE.

Tableau 10 : Proportion totale des erreurs significatives (SE) et des erreurs critiques (CE) pour l'IBG DCE calculé sur la base des Z-scores (Z-score < 2 : non significatif, $2 \leq$ Z-score ≤ 3 : significatif, Z-score > 3 : erreur critique) par le biais de l'algorithme présenté précédemment pour chaque classe de qualité (très bonne, bonne, moyenne, médiocre et mauvaise), après l'exclusion aléatoire des taxons rares ou peu abondants (R) et pour chaque probabilité d'erreur. Valeur obtenue sur le cumul des erreurs pour les 5 rivières par classe ($n=5/\text{classe}$, 100 tests/2)

Ecological status	Error rates	Very good		Good		Moderate		Poor		Bad	
		SE	CE	SE	CE	SE	CE	SE	CE	SE	CE
IBGN	10%	0	0	12	2	22	1	23	1	0	2
	25%	0	4	20	2	22	5	31	21	3	1
	40%	21	0	3	39	36	36	28	22	12	3
GI	10%	0	0	6	0	4	1	0	0	0	0
	25%	0	0	0	0	3	1	0	0	0	0
	40%	0	0	16	0	32	0	0	0	0	0
Richness	10%	9	5	20	6	11	2	18	3	2	3
	25%	28	11	38	21	27	11	23	5	11	3
	40%	40	31	35	36	21	52	25	25	18	6
VC	10%	9	1	9	2	25	0	23	2	0	2
	25%	1	5	20	2	46	3	31	23	8	1
	40%	33	18	15	27	31	50	28	22	12	3

Nous constatons que, dans la plupart des cas, la proportion d'erreurs significatives (SE) est supérieure à celle des erreurs critiques (CE), mais également que la part d'erreur non significatives domine pour les probabilités d'erreurs faibles ($R=10\%$, 25%). Plus la probabilité d'erreur introduite augmente, plus le nombre d'erreurs significatives et critiques est important. Il y a plus d'erreurs observées pour les IBG DCE appartenant aux classes médianes (bonne, moyenne et médiocre). Très peu d'erreurs concernent le GI. La classe de variété, par contre, est plus impactée par l'exclusion des taxons rares et peu abondants.

Titre 4 : Discussion

Des études ont montré les effets de la prise en compte des taxons rares ou peu abondants, de leurs oublis, et des erreurs d'estimations d'abondance (Cao, et al., 1998; Marchant, 2002; Lavoie, et al., 2009). Les facteurs de variations de ces deux paramètres sont fonction de la variabilité de l'échantillonnage, elle-même dépendant de la mosaïque d'habitats, de la variabilité intra-substrat

et de celle de l'opérateur. Elles sont également fonction de la variabilité des opérations de tri et d'identification au laboratoire. Échantillonnage, tris et identification sont sujets à l'évaluation visuelle réalisée par les opérateurs. Or, il résulte de l'évaluation visuelle des divergences entre opérateurs comme par exemple pour les estimations de recouvrement de végétaux ou l'estimation des abondances des invertébrés. L'évaluation visuelle reste, néanmoins, considérée comme peu coûteuse et fournissant les informations nécessaires à l'évaluation des rivières (Kilroy, et al., 2013; Cheshmedjiev, 2013; Barbour, et al., 1999). En ce qui concerne la partie laboratoire, ces deux paramètres (tri efficace des taxons rares et peu abondants et estimations des abondances) peuvent être extrêmement variables d'un opérateur (trieur et identificateur) à l'autre en fonction de son expérience.

Il existe très peu de travaux qui ont porté sur l'étude de l'effet des erreurs d'estimation des abondances sur les indices ou sur l'évaluation des rivières. Ceux existants concernant les macroinvertébrés, se sont plus focalisés sur l'étude des effets de l'environnement, de la distribution des macrophytes et des méthodes d'échantillonnage sur les abondances des macroinvertébrés (Taniguchi, et al., 2003; Burton, et al., 2002; Lorenz, et al., 2004; McAbendroth, et al., 2005). A contrario, la prise en compte des taxons rares ou peu abondants a plus largement été étudiée. Notamment, l'exclusion ou l'inclusion des taxons rares a soulevé une divergence entre auteurs. Certains considèrent que leur prise en compte dans les analyses multivariées et dans l'évaluation biologique des milieux aquatiques génère une information redondante et peut produire du bruit inutile dans ces analyses (Arscott, et al., 2006; Marchant, 2002). D'autres mettent en avant le fait que les présences/absences de certains taxons permettent de détecter les perturbations d'un milieu (Cao, et al., 1998; King & Richardson, 2002). Ces études ont porté sur différents compartiments comme les poissons (Pritt & Frimpong, 2014), les macroinvertébrés (Cao, et al., 2001; Cao, et al., 1998) ou encore les diatomées (Lavoie, et al., 2009) et concernent essentiellement des indices étrangers : ils reposent donc sur des protocoles d'échantillonnage et de calculs de métriques qui leur sont propres. C'est pourquoi nous avons étudié les effets de l'exclusion des taxons rares ou peu abondants et des erreurs d'estimation des abondances sur les indices macroinvertébrés français à l'aide d'une méthode innovante, peu coûteuse et facilement applicable.

La rareté doit être définie. Les auteurs qui s'en sont préoccupés, se basent soit sur les occurrences ou les proportions des taxons rares dans un échantillon (5%, etc.) (Cao, et al., 1998; Cao, et al., 2002; Cao, et al., 2001; Poos & Jackson, 2012; Wan, et al., 2010), soit sur leurs abondances

relatives (<1%, etc.) (Arscott, et al., 2006; Lavoie, et al., 2009; Pritt & Frimpong, 2014), soit sur des critères mathématiques tels que le quartile de Gaston (Gaston, 1994) ou le quartile inférieur (25%) (Siqueira, et al., 2012). Enfin, Nijboer (2004) a défini différents niveaux de rareté en fonction des abondances des taxons et a attribué une valeur fixe pour définir les taxons peu abondants en fonction de la moyenne des individus présents dans un échantillon (≤ 5). Notre méthode de définition des seuils présentée plus haut (§2) est proche de celle de Nijboer. Elle prend en compte les espèces difficiles à échantillonner (§2) puisqu'elles apparaissent uniquement dans certaines répétitions de substrats avec des abondances faibles et traduit l'inexpérience d'un opérateur. Dans la littérature, il n'existe pas de définition précise pour identifier un taxon abondant, aussi, pour définir notre seuil nous nous sommes appuyés sur la norme XP T90-388 (AFNOR, 2010) ainsi que sur les expérimentations intra-substrats (Titre 1). Nous avons considéré qu'au-delà de 40 taxons, il s'agissait d'estimations d'abondance donc nous avons utilisé cette valeur comme seuil. Les effets des pratiques de laboratoire sont induites avec les sous-estimations des abondances des taxons.

Notre démarche se rapproche des méthodes reposant sur la permutation d'échantillons élémentaires (Archambault, et al., 2014) ou sur la suppression systématique des taxons rares et la recombinaison de listes (Cao, et al., 1998). Mais elle s'en distingue originalement en essayant de rester plus fidèle à la réalité en simulant les variabilités humaine par le biais d'expérimentations aléatoires basées sur deux tests d'erreurs progressifs en fonction de différentes probabilités d'erreurs : (1) la suppression probable aléatoire des taxons rares, et (2) la sous-estimation aléatoire des abondances. Ainsi, les nouvelles listes faunistiques dégradées obtenues peuvent être le reflet de la variabilité de l'opérateur, de son manque d'expérience et des effets des pratiques de laboratoire. Lors de l'échantillonnage sur le terrain des habitats (cf titre 1), nous avons pu mesurer une variabilité plus ou moins marquée en fonction des habitats, mais cette variabilité s'ajoute à celle liée à l'opérateur et ses pratiques lors de l'étape du tri et de l'identification des taxons. Cependant, les tests que nous avons menés ici, nous ont seulement permis d'estimer l'efficacité et la compétence des opérateurs en fonction des oublis de taxons et des pratiques de laboratoires par des sous-estimations d'abondances. Nous avons pu observer dans quelle mesure les métriques sont plus ou moins sensibles aux trois types d'erreurs : l'exclusion des taxons rares, la sous-estimation des abondances et les deux cumulés. Les métriques de l'IBG DCE sont touchées uniquement par l'exclusion des taxons rares ou peu abondants (GI et VC, figure 33), tandis, que

les métriques de l'I2M2 sont soit touchées par l'exclusion des taxons rares (TAX et ASPT, figure 34-A) soit par les sous-estimations des abondances (indice de Shannon, figure 34-B), soit à la fois par l'exclusion et les sous-estimations des abondances (indice de Shannon, fréquences des individus polyvoltins et ovovivipares, figure 34-C). La Figure 33 présente la variabilité maximale de l'IBG DCE et ses métriques pour l'exclusion des taxons rares et peu abondants et la Figure 34 présentent les variations maximales observées de l'I2M2 et ses métriques pour chaque type d'erreurs.

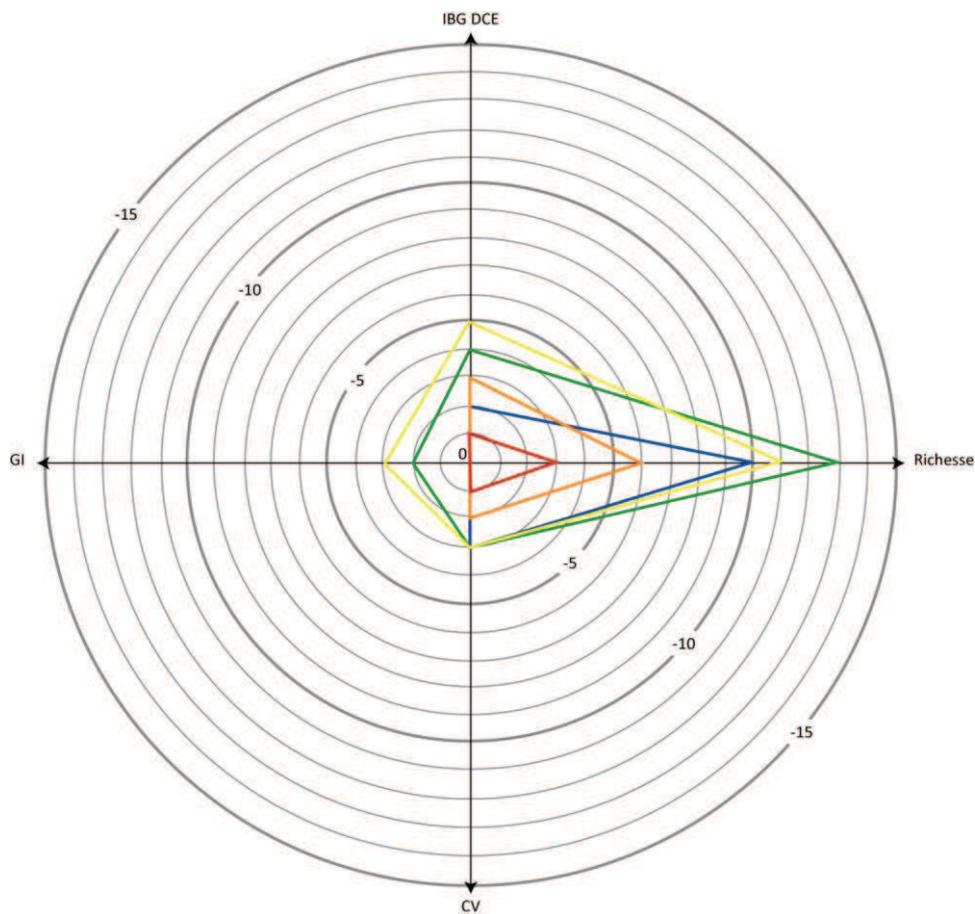


Figure 33: Perte maximale de points pour l'IBG DCE, ses métriques (GI,VC) et la diversité taxonomique pour chaque classe de qualité (très bonne, bonne, moyenne, médiocre et mauvaise) obtenue avec l'algorithme développé précédemment due à l'exclusion des taxons rares ou peu abondants (écarts obtenus par rapport aux valeurs initiales)

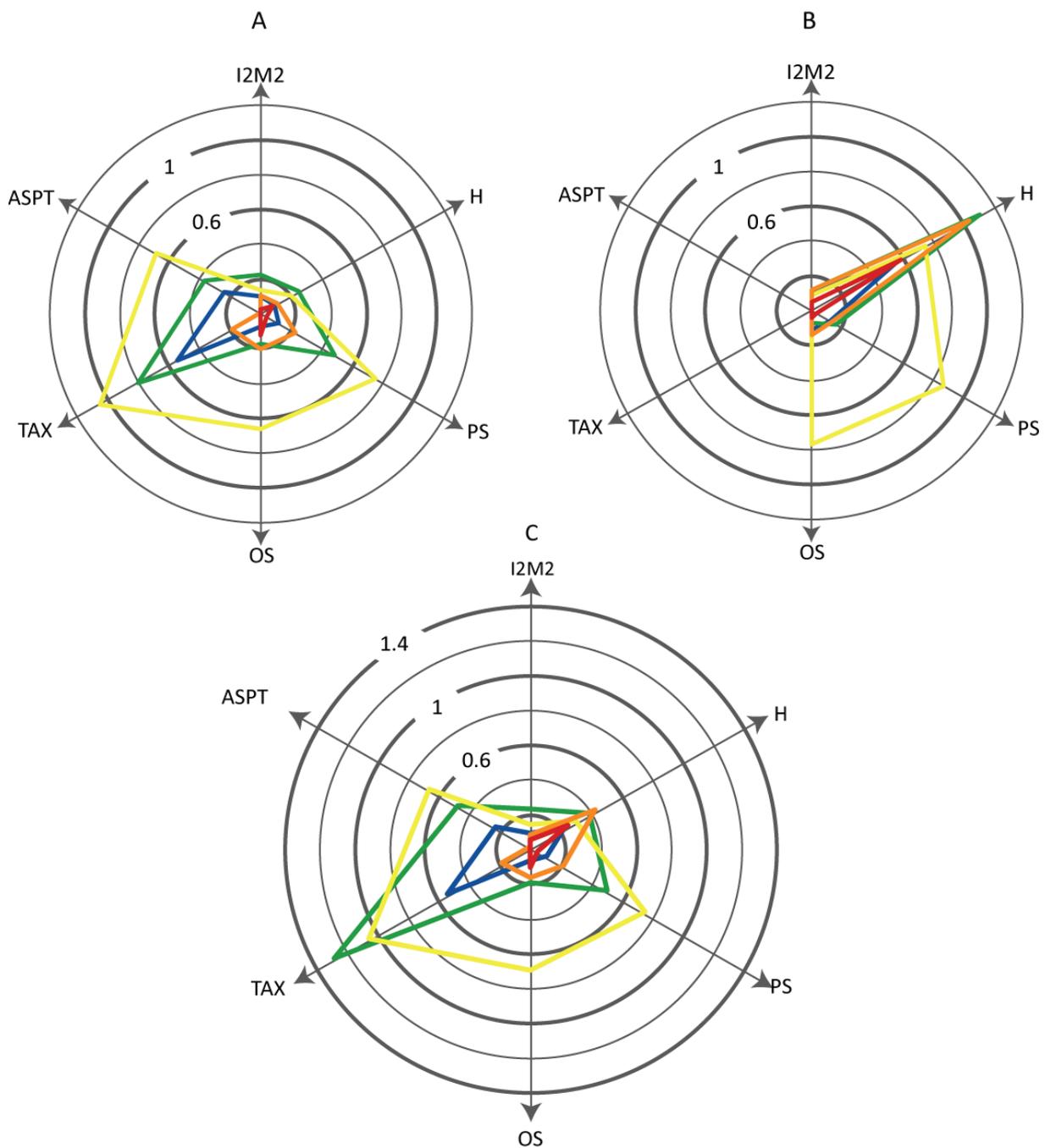


Figure 34 : Variations maximales obtenues après (A) l'exclusion des taxons rares (R), (B) les erreurs d'estimations des abondances (AB) et (C) l'exclusion des taxons rares et les erreurs d'estimations des abondances (R+AB) pour l'I2M2 et ses EQR calculés par métriques (TAX, ASPT, H, PS et OS) pour chaque classe de qualité (très bonne, bonne, moyenne, médiocre et mauvaise) obtenues avec l'algorithme développé précédemment (écarts obtenus par rapport aux valeurs des EQR initiaux)

Les taxons rares ou peu abondants fréquemment oubliés correspondent aux taxons de petites tailles (Archaibault, et al., 2014; Metzeling, et al., 2003) et aux plus polluosensibles (Archaibault, et al., 2014). Leur exclusion entraîne une diminution de la diversité taxonomique et donc impacte les métriques associées à la richesse. Ainsi, nous avons constaté que les métriques

les plus sensibles à cette erreur (R) sont la classe de variété (CV) et le groupe indicateur (GI) (Figures 33 et 35) pour l'IBG DCE et les EQR de l'ASPT et TAX (figures 34- A et 35) pour l'I2M2. Nous avons néanmoins constaté que même si la richesse taxonomique diminuait, selon la classe de qualité observée, les variations du groupe indicateur et de l'ASPT pouvaient être faibles voire nulles ; c'est le cas pour les rivières de très bonne qualité. Celles-ci sont caractérisées par un nombre important de taxons très polluosensibles qui même après l'exclusion d'une partie, permettent le maintien du groupe indicateur ; ainsi un score de polluosensibilité équivalent est possible. Les taxons polluosensibles sont également associés aux taxons semi ou monovoltins et à pontes libres, leur exclusion impacte les métriques liées aux deux traits biologiques (PS et OS). En effet, celles-ci augmentent avec les oublis de taxons et peuvent alors dénoter une mosaïque d'habitat diversifiée mais également une certaine instabilité du milieu (figures 34 et 35). L'indice de Shannon (H) est, quant à lui, plus sensible à la modification des estimations des abondances. En effet, celui-ci augmente lorsque les fortes abondances sont réduites à 40, ceci a donc tendance à traduire un peuplement plus équilibré, tandis que l'exclusion des taxons aura un effet contraire puisque l'indice va diminuer traduisant la dominance de certains taxons (Figures 34 et 35).

En général, l'ensemble des métriques de l'I2M2 sont plus sensibles à l'exclusion des taxons qu'aux erreurs des estimations des abondances. Pour les métriques qui prennent également en compte l'abondance des taxons, lorsque les deux erreurs sont cumulées, un effet compensatoire existe et la variation des métriques est moindre. Une classification des métriques selon leur variabilité peut être proposée de la moins variable à la plus variable en fonction des coefficients de variation moyens et des différents types d'erreurs (Figure 35).

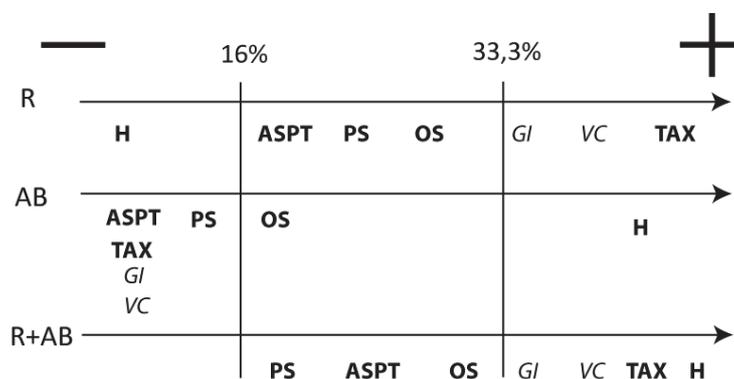


Figure 35 : Gradient de variabilité des métriques basé sur les coefficients de variation moyens obtenus (CV) pour chaque métriques et classe de qualité (en gras métriques de l'I2M2, en italique métriques de l'IBG DCE). CV < 16% : pas d'impact significatif de l'erreur, 16% < CV < 33,3% : l'erreur est significative et CV > 33,3% : erreur critique.

L'IBG DCE n'est sensible qu'à l'exclusion des taxons qui entraîne une diminution des notes. La diminution de la note est plus importante pour les classes moyennes que pour les classes très bonne et mauvaise. Cela est dû, pour les classes moyennes, au double effet des variations des deux métriques (GI et VC) prises en compte dans l'indice, tandis que les classes extrêmes n'ont que la classe de variété (VC) qui est modifiée. En ce qui concerne l'I2M2, les trois tests exercent une influence différente sur la note. En effet, l'exclusion des taxons rares mène à une diminution des notes, les sous-estimations d'abondances augmentent la note, et il existe une compensation lorsque les deux erreurs sont cumulées.

Les indices sont donc sensibles aux deux incertitudes testées et les classes les plus impactées sont les classes intermédiaires (bonne, moyenne et médiocre). D'autres études ont également montré que les classes intermédiaires étaient plus sensibles aux erreurs et que les incertitudes avaient plus d'impact sur ces classes que sur les autres (Cao, et al., 1998; Archambault, et al., 2014). Les rivières dégradées ont un nombre de taxons polluosensibles plus réduit, leur suppression mène à une diminution de la note et éventuellement à un déclassement. Pour la probabilité d'erreur maximale (40%) les écarts de note les plus importants constatés sont pour l'IBG DCE de -5 points pour les classes moyennes, et pour l'I2M2 de +/- 0,22 points pour les bonnes qualités. Ces diminutions de note ne reflètent que l'impact dû à un manque de compétence d'un opérateur, puisque nous n'avons testé que la disparition de taxons. Beisel & al (1998a) ont montré que l'indice IBGN pouvait varier de 1 à 3 points selon les différentes combinaisons d'habitats possibles. Ainsi, il serait possible d'observer un effet compensatoire entre la variabilité intra-substrat et celle liée à l'opérateur et donc une variation de la note moins importante.

Les milieux de très bonne qualité ont une faune diversifiée et comptent de nombreux taxons polluosensibles, donc malgré les différents tests, les chances d'un déclassement sont moindres. De même pour les rivières de très mauvaise qualité, le nombre de taxons résistants domine largement, les différents tests menés n'entraînent pas une diminution significative des notes (l'exclusion des taxons rares entraîne une perte de 1 point maximum pour l'IBG DCE et une variation de 0,02 point pour l'I2M2). Tout comme Cao (op.cit), nous constatons que le nombre d'espèces dites rares ou peu abondantes est supérieur dans les sites peu impactés. L'impact de l'exclusion de taxons sur la note IBG DCE serait lié à la dégradation observée d'un site, excepté pour les sites très dégradés.

Titre 5 : Conclusion

La démarche que nous avons mise en œuvre est originale par rapport aux travaux basés sur des permutations d'échantillons ou des créations artificielles de nouvelles listes qui traduisent la variabilité liée à l'habitat. En effet, notre démarche porte sur des altérations virtuelles de listes faunistiques reproduisant les erreurs dues à l'inexpérience d'un opérateur et des pratiques de laboratoires. Elle pourrait être utilisée sur d'autres compartiments. Ainsi, nous l'avons adaptée pour l'étude sur les incertitudes associées à l'IBMR (Chapitre 3). Concernant les macroinvertébrés, nous avons axé notre étude sur les oublis de taxons rares ou peu abondants et les sous-estimations d'abondances, celles-ci traduisent l'effet de l'inexpérience de l'opérateur (R) et l'effet des pratiques en laboratoires (AB). Les oublis de taxons impactent plus les indices et leurs métriques que les sous-estimations d'abondances. Les métriques associées à la richesse taxonomique sont très sensibles à l'exclusion des taxons, ainsi pour les écarts maximums générés : l'IBG DCE peut diminuer jusqu'à moins 5 points pour une classe moyenne et pour une probabilité d'erreur de 40% et l'I2M2 de 0,02 point pour une classe bonne et une probabilité d'erreur maximale de 40% (inexpérience maximale testée).

Les indices sont donc sensibles à l'inexpérience de l'opérateur et aux pratiques en laboratoire. Pour limiter l'effet opérateur et éventuellement les effets dus aux différentes pratiques en laboratoire, des solutions correctives peuvent être mises en place. L'entraînement des opérateurs ainsi que le développement de leur connaissances des protocoles issus des normes sont nécessaires pour obtenir des résultats fiables et de meilleure qualité (Haase, et al., 2006; Archambault, et al., 2014). Des études portant sur les incertitudes liées aux indices macroinvertébrés ont montré l'importance de développer des démarches qualité, comme des entraînements rigoureux et des audits afin de réduire les incertitudes. Il peut être également recommandé la recherche systématique des individus de petites tailles afin d'éviter les oublis de taxons rares, notamment dans les substrats les plus difficiles à trier (les litières par exemple). Les estimations des abondances des taxons restent le facteur le plus difficile à corriger même avec l'entraînement et l'expérience puisqu'il est extrêmement variable d'un opérateur à un autre (Aquabio, étude en cours 2014). Une solution peut être la confrontation des estimations des

abondances entre plusieurs opérateurs, notamment pour les opérateurs débutants.

CHAPITRE 3 Méthode virtuelle de mesure des incertitudes liées à l'opérateur associées aux macrophytes

Ce chapitre a pour but de présenter la démarche mise en œuvre pour évaluer l'incertitude liée à l'opérateur lors de l'application du protocole basée sur l'indice IBMR NF T90-395 (AFNOR, 2003). Il est présenté sous la forme d'un article accepté qui est en cours de parution dans *Ecological Indicators* (Wiederkehr, et al., 2015).

Titre 1 : Experimental Study of uncertainty in the Macrophyte index (IBMR) based on species identification and cover (Wiederkehr, et al., 2015) : Résumé

Notre objectif est d'étudier le rôle des incertitudes, définies comme toutes les actions entraînant une erreur, sur l'Indice Biologique des Macrophytes de Rivière (IBMR) basé sur les macrophytes. L'IBMR évalue le niveau trophique des rivières.

Nous avons choisi d'analyser les incertitudes basées sur l'effet opérateur, à la fois *in situ* et au laboratoire, comme les oublis de taxons, les erreurs de détermination et les erreurs d'estimation des classes d'abondance. Nous avons utilisé une approche innovante (la même que celle utilisée dans le chapitre précédant pour les indices macroinvertébrés) proche d'une analyse de sensibilité, basée sur des modifications virtuelles contrôlées des déterminations de taxons et des classes de recouvrement par le biais de deux matrices de confusion.

La création de nouvelles listes floristiques expérimentales et le calcul des métriques en fonction des erreurs spécifiques aléatoires nous permettent de mesurer les effets de ces erreurs sur l'IBMR et le niveau trophique.

Nos résultats ont montré que les confusions de taxons et les deux erreurs combinées (erreurs d'identification des taxons et d'estimations des classes d'abondance) ont toujours un effet plus important que les confusions sur les classes d'abondance. Afin de limiter leurs impacts, des solutions peuvent être apportées comme par exemple l'entraînement des opérateurs, la confrontation entre opérateurs des déterminations et estimations des recouvrements et le développement d'une démarche qualité.

Titre 2 : Introduction

The Water Framework Directive (WFD) established a coherent water policy within the European Union (Directive 2000/60/CE, 2000). It introduced the principle of protection and management of surface, transitional, coastal and ground waters.

Therefore, the use of standardized biological indices is necessary in monitoring programs to assess the ecological quality of freshwater. Furthermore, the WFD requires reliability, reproducibility and comparability of ecological data produced by European countries and asking them to estimate the confidence and precision levels of the results provided by the monitoring programs. These levels, such an estimation, necessarily implied to study uncertainties at each step of the process. This process involved measurement of biological indices by humans that induces high variability on obtained metrics whatever the spatial scale or used biological groups (Springe, et al., 2006; Johnson, et al., 2006).

Most research focused on the uncertainties of the benthic invertebrate metrics and indices such as sampling or sorting phases (Beisel, et al., 1998a; Beisel, et al., 2000; Mazuer, et al., 2007; Matte, 2005). Otherwise, visual assessment and rapid biological assessment (RBA) (Cheshmedjiev, 2013; Barbour, et al., 1999) efficiency was investigated in several studies on periphyton, invertebrates and fish. Thus, through a comparison with quantitative sampling, it has been shown that RBA and visual assessment are cost effective and provide the necessary information for streams assessment (Gillies, et al., 2009; Metzeling, et al., 2003; Kilroy, et al., 2013). These studies did not take into account the operator experience but it may have an effect on the ecological assessment. Further researches focused on the operators' skills and to the different sampling methods used (Lorenz & Clarke, 2006; Lorenz, et al., 2004; Metzeling, et al., 2003).

The variability of macrophytes - another group of organisms for which the WFD prescribes such assessments - has been dealt with in few studies, some of which were based on replicate sampling experiments (Staniszewski, et al., 2006). Likewise, few studies have focused on the field inter-surveyor impact, still considered to be negligible (Mascaro, et al., 2013). The role of hydrological and morphological conditions on plant distribution was highlighted (Breugnot, et al., 2008), the effects of their variability were tested on macrophyte metrics and indices (Szoszkiewicz,

et al., 2009). Furthermore, habitat modification, shading as well as seasonality have been shown to play an important role in the variability of the species number, Shannon-Weaver index or plant abundance (Staniszewski, et al., 2006). However, the impact of uncertainty factors such as the surveyor effect in index calculation and assessment of ecological status, has not been studied yet.

The Macrophyte Biological Index for Rivers (IBMR, (Haury, et al., 2006; AFNOR, 2003) is based on three metrics: field cover percentage, the species trophic score, and a coefficient of ecological amplitude, the two latter metrics relying on an estimation made by various experts involved in the guidance standard (AFNOR, 2003).

As the surveyor's skills may impact these three metrics, the aim of this study was to estimate uncertainties due to species identification and cover percentage errors, and show how sampling, metrics of IBMR, index and trophic status assessment could vary with these errors. We proposed an innovative approach close to sensitivity analysis, relying on virtual changes based on two confusion matrices and a specific algorithm. As a result, the uncertainty is represented by confusion coefficients based on field knowledge. We calculated the variation of metrics according to random specified errors and we measured the effect of these errors on the IBMR robustness and trophic status.



Titre 3 : Materials and methods

To estimate the variability associated with identification and cover percentage errors, we chose a mathematical method close to sensitivity analysis, which created virtual identification errors or taxa omissions and cover percentage errors. These errors directly affect the IBMR metrics.

Section 1: The IBMR formula (AFNOR, 2003)

The IBMR technical process is based on in-situ observation of macrophytes to identify each taxa i , and estimate the cover percentage transformed into a cover scale K_i . The calculation of IBMR also relies on the definition of a species' trophic score CS_i and a coefficient of ecological amplitude E_i , both relying on an estimation of various experts involved in the standard (AFNOR, 2003).

$$IBMR = \frac{\sum_i^n E_i \times K_i \times CS_i}{\sum_i^n E_i \times K_i} \quad (27)$$

Where i is the contributory taxa, n is the total number of contributory taxa, CS_i is the trophic score⁶ of i ranging from 0 to 20, E_i is the coefficient of ecological amplitude¹ from 1 to 3, and K_i is the cover scale⁷ from 1 to 5.

The IBMR score is translated into five trophic states decreasing from oligotrophic to hypertrophic: $IBMR > 14$ corresponds to a very low trophic status, $14 \geq IBMR > 12$ to a low trophic status, $12 \geq IBMR > 10$ to a medium trophic status, $10 \geq IBMR > 8$ to a high trophic status, and $IBMR \leq 8$ corresponds to a very high trophic status.

Identification errors and taxa omissions impact the CS_i and E_i . Cover class errors impact the K_i .

To create random errors a list of contributory taxa and two confusion matrices were built.

Section 2: Metadata

The contributory taxa matrix was based on the list of the 208 contributory taxa from the standard (AFNOR, 2003). This matrix comprised the CS_i and E_i and allowed us to calculate the index

⁶ Values fixed by the judgments of various experts (AFNOR, 2003)

⁷ Values estimated by the surveyor in the field

score.

The taxa confusion matrix (Table 7) was based on all contributory (208) and non-contributory taxa with which confusions were possible (11 taxa). Altogether based on real field knowledge, 82 taxa among the 219 taxa could be confused. These taxa were the most difficult to identify and/or the rarest. For example, specific differentiation of the genus *Callitriche* or *Potamogeton* is based on their development stage, such as flowering and fruiting, which are not always observable at the sampling time.

For each pair of taxa we determined a confusion value depending on the difficulty to identify species. The confusion values were fixed by the experts' judgments based on their field and laboratory experiences.

Table 7: Part of the confusion matrix on macrophytes taxa (CLASPX = *Cladophora* sp. MICSPX = *Microspora* sp., RHISPX = *Rhizoclonium* sp., STISPX = *Stigeoclonium* sp., STITEN = *Stigeoclonium tenuie*, VAUSPX = *Vaucheria* sp).

Taxa	CLASPX	MICSPX	RHISPX	STISPX	STITEN	VAUSPX
CLASPX	80	5	15	0	0	0
MICSPX	5	80	15	0	0	0
RHISPX	15	15	70	0	0	0
STISPX	0	0	0	50	50	0
STITEN	0	0	0	50	50	0
VAUSPX	0	0	0	0	0	100

For example, *Microspora* sp. can be mistaken for *Cladophora* sp. with 5% error rate whereas confusion with *Vaucheria* sp. was not possible (Table 7).

A cover class confusion matrix (Table 8) was built on the same principle as the taxa confusion matrix. It is composed of different classes established in the IBMR. We also created a new class (class 0) allowing taxa omissions. The confusion values were fixed by the experts' judgments based on field experience and errors related to contiguous classes.

Table 8 : Confusion matrix on macrophytes cover classes (C0: absent, C1: covering < 0.01%, C2: 0.1% ≤ covering < 1%, C3: 1% ≤ covering < 10%, C4: 10% ≤ covering < 50%, C5: covering ≥ 50% of stream bed)

	C0	C1	C2	C3	C4	C5
C0	85	15	0	0	0	0
C1	15	70	15	0	0	0
C2	0	15	70	15	0	0
C3	0	0	15	70	15	0
C4	0	0	0	15	70	15
C5	0	0	0	0	15	85

In these two matrices, the confusion values did not depend on the surveyors' skills. Whatever skills, the potential error was the same. The skill level was taken into account later in the process by introducing the error rate in the algorithm.

Complete matrices are available as supplementary material (see).

Section 3. Algorithm

The algorithm was designed to randomly modify an observed species list, based on the confusion matrices described above, and was implemented in the Ruby language⁸ which proved its robustness in many applications (Katoh, et al., 2009; Aerts & Law, 2009).

The algorithm runs as follows. It takes a floristic list and an error rate as input, and produces a randomly modified taxa or cover class list, by considering either the taxa confusion matrix, or the cover class confusion matrix. Each process can be performed separately or together.

Three parts compose the algorithm: (1) the random selection of lines to modify, according to the error rate; (2) the application of confusion values according to the two matrices and (3) the calculation of the IBMR index.

More formally, let \mathcal{L}_i be a floristic list, with a number $n = |\mathcal{L}_i|$ of lines, let \mathcal{E}_t be a theoretical error rate, \mathcal{E}_t in $[0,100]$, the number of randomly selected lines is $n_s = n \mathcal{E}_t / 100$.

The final list \mathcal{L}_m has n lines, among which $n_m \leq n_s$ lines are different from those of \mathcal{L}_i . A real

⁸<https://www.ruby-lang.org/en/>

error rate \mathcal{E}_r can then be defined as follows:

$$E_r = 100 n_m/n \quad (28)$$

Since some taxa cannot be confused with others, and all taxa / cover classes can always be confused with themselves (see Tables 12 and 13) the real error rate is always lower than \mathcal{E}_t .

The floristic list \mathcal{L} in Table 9 illustrates this. It represents four taxa, their corresponding cover class and the obtained IBMR score equal to 6.83. With $\mathcal{E}_t = 75\%$, the method randomly selects three lines (in light blue in Table 9) representing taxa STIPX, VAUSPX and RHISPX and their associated cover class.

Table 9: Example of an original floristic list \mathcal{L}

Taxa	Abundance class
STIPX	Class 1
MICSPX	Class 1
VAUSPX	Class 2
RHISPX	Class 3
IBMR : 6,83	

After the line selection process, we can now modify taxa or/and the cover class based on the confusion values in the taxa confusion matrix and in the cover class confusion matrix.

§1. Taxa modification (TAX)

In this module, the taxa \mathcal{T} of each selected line has a probability to be replaced by another taxa \mathcal{T}' in the confusion matrix (see Table 7), where the confusion value of \mathcal{T} with \mathcal{T}' is greater than 0. The higher the confusion value, the higher the probability of \mathcal{T} being replaced by \mathcal{T}' . For example, the taxa RHISPX has a 15% probability to be replaced by the taxa CLASPX. It is important

to note that it also has a 70% probability to be replaced by itself. A taxa can be selected to be replaced, but there is a probability that the taxa remains unchanged. Given the selected taxa in Table 9, Table 10 provides a possible modified floristic list. STISPX taxa was replaced by STITEN with a 50% probability, RHISPX taxa was replaced by CLASPX with a 15% probability while VAUSPX taxa remained unchanged since it was replaced by itself with a 100% probability, which means that this taxa cannot be confused with another taxa according to the taxa confusion matrix. With $\mathcal{E}_t = 75\%$ and considering the probability of taxa being replaced by themselves, two taxa were changed so we had a real error rate \mathcal{E}_r of 50%. With these modifications the IBMR score decreased by 1.5 (Table 15) but there was no change in trophic class.

Table 10 : Modified floristic list on taxa from original floristic list L (Table 14) obtained with the algorithm defined in Section 3 after abundance classes modifications (TAX)

Taxa	Abundance class
STITEN	Class 1
MICSPX	Class 1
VAUSPX	Class 2
CLASPX	Class 3
IBMR : 5,30	

§2. Abundance modification (AC)

This part of the algorithm consists in modifying cover classes using the cover class confusion matrix (Table 8). Cover class C of each selected line has a given probability to be replaced by another cover class based on the cover class confusion matrix. For example cover class 1 has a 15% probability to be both replaced by class 0 or class 2 and a 70% probability to be replaced by itself, that is to say cover class 1 remained unchanged. To illustrate this process we use the same example as in Table 9. Table 11 presents a possible modified floristic list. The cover class of the taxa STITEN changed from cover class 1 to cover class 0 (a decrease in the cover class), the cover class of the taxa RHISPX changed from cover class 3 to cover class 4 (an increase in the cover class), while the cover class of the taxa VAUSPX remained unchanged. With $\mathcal{E}_t = 75\%$ and considering the cover class probability to be replaced by themselves, two cover classes were changed so we had

a real error rate \mathcal{E}_r of 50%. Thus, the IBMR score decreased from 6.83 to 5.33.

Table 11: Modified floristic list on abundance classes from original floristic list \mathcal{L} (Table 14) obtained with the algorithm defined in Section 3 after abundance classes modifications (AC)

Taxa	Abundance class
STITEN	Class 0
VAUSPX	Class 2
RHISPX	Class 4
IBMR : 5.33	

§3. Taxa and cover class modification (TAX+AC)

The algorithm gave the possibility to modify taxa and cover classes together. To perform this combined error, we used the same confusion matrices and process, as explained previously. The two modifications were carried out one after the other. Firstly we modified taxa based on confusion values and secondly, in the new experimental list, we modified cover classes, according to the same theoretical error rate.

Section 4. Experiments

We used a data set composed of 25 checked floristic lists from river monitoring and the Aquabio's database (a consultancy specialized in hydrobiological data expertise). In order to validate these lists, ambiguous samples were sent to external experts who checked them.

To determine the impact of studied errors in relation with the trophic status of streams, we selected 5 lists in each trophy level (very low, low, medium, high and very high). The selected floristic lists came from sites located in the central and western areas of France (see Figure 36). These areas include five hydroecoregions (HER): "Massif Armoricain" (Armorican Massif, west of France), "Massif Central nord" (north of the Massif Central, north of Clermont-Ferrand), "Tables Calcaires" (calcareous zone, Parisian Basin), "Dépôts Argilo-sableux" (sandy-clay soils, south of

Orleans) and "Dépressions Sédimentaires" (sedimentary basin, east of the Massif Central). The floristic lists were produced in accordance with the NF T 90-395 (AFNOR, 2003).

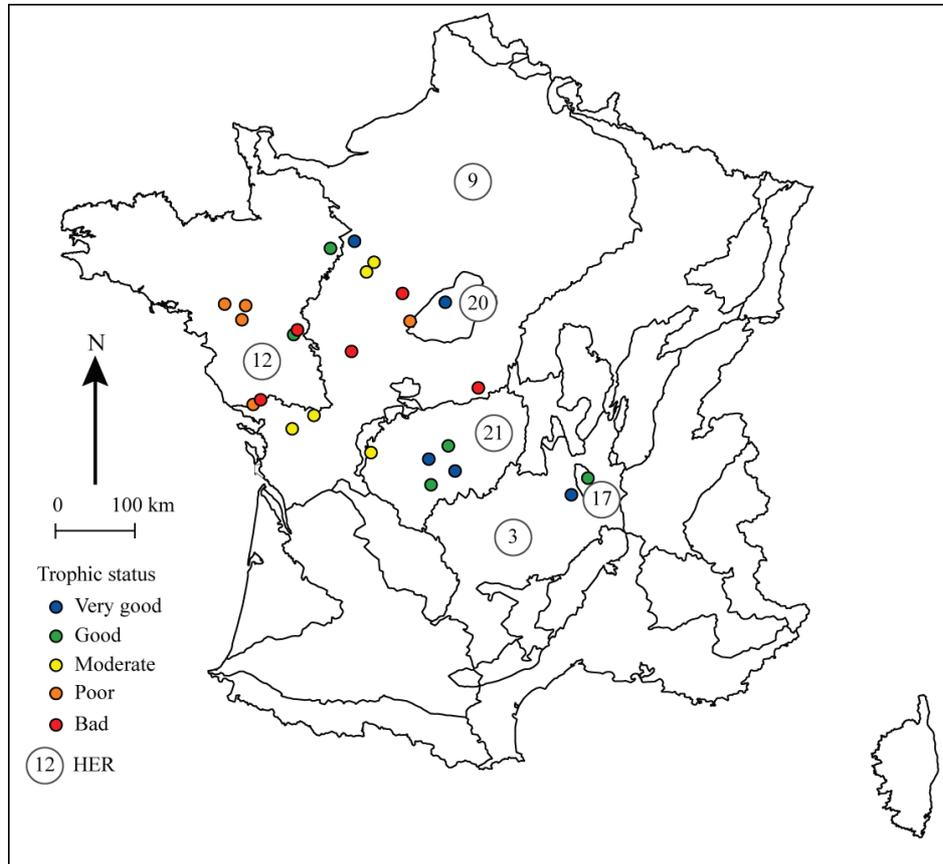


Figure 36: Localisation of the 25 sites, chosen for our experiments, in France

We developed three sets of tests for each checked floristics lists: (1) A first one with only the taxa modifications, (2) another one with the cover class modifications (AC) and (3) the last with both modifications (TAX+AC). We ran each test 20 times for each theoretical error rate (10, 25, 50, 75 and 100%). We obtained 2500 new degraded lists for each test (25 lists \times 5 error rates \times 20 times). Degraded lists were characterized by the floristic list \mathcal{L} and the associated trophic status t , but also by the introduced theoretical error rate \mathcal{E}_t and the modification (TAX, AC or TAX+AC). So, we had five trophic states (very low, low, medium, high, very high) and n streams ($n=5$) in each trophic status t . We generated P simulations ($P=20$) for each error rate \mathcal{E}_t and error type (TAX, AC or TAX+AC), allowing us to determine the impact of errors due to TAX, AC and TAX+AC.

Section 5. Data analysis

To study the metrics uncertainties, the IBMR index and trophic status assessment we used a distance analysis and a Z-score analysis. We developed a new indicator, namely the major confusion between Macrophyte species Indicator, JUMI, to appraise the role of taxa in IBMR variability.

§1. Distance analysis

For each simulation p we calculated the distance $D_{p,i}$ between the simulated IBMR score $_p$ and the reference score $_i$.

$$D_{p,i} = score_p - score_i \quad (29)$$

We considered the dispersion ds_i for the list i as the sum of distances between all score $_p$ values and score $_i$, with the following formula:

$$ds_i = \sqrt{\sum_{p=1}^P (D_{p,i})^2} \quad (30)$$

We considered the average dispersion DS_t in each trophic status t :

$$DS_t = \frac{\sum_{i=1}^n ds_{i,t}}{n} \quad (31)$$

The variation coefficient CV_t is the ratio of DS_t to the average of the reference scores for n list in each trophic status t . It is expressed in percentage.

$$CV_t = \frac{DS_t}{\frac{1}{n} \sum_{i=1}^n score_i} \times 100 \quad (32)$$

CV values can be interpreted as follows (Gendron & Martin, 2004) : when the CV is less than 16.0% the variation is low and the average estimation is reliable; between 16.0% and 33.3%, the variation is important and the average could contain errors, and when the CV is over 33.3% the variation is high and the average is not reliable.

§2. Z-Score calculation

Computing Z-score is a classical method to assess the surveyor performance in inter-laboratory trials.

Here, for each error rate, we considered P simulations for n river i in each trophic status t . A $Z\text{-score}_{p,i}$ for each p simulation was the ratio of the distance $D_{p,i}$ to the standard deviation σ_i of the score_p values for P simulation.

$$\sigma_i = \sqrt{\frac{1}{n} \sum_{i=1}^n (p - \bar{p})^2} \quad (33)$$

$$Z\text{-score}_{p,i} = \frac{D_{p,i}}{\sigma_i} \quad (34)$$

This method allowed us to obtain a value which highlighted the relationship between simulated score and reference score. When the Z-score was lower than 2, the observed errors were not significant. When it was between 2 and 3, the error had a significant effect. Over 3 the error was critical (AFNOR, 2005). The negative or positive Z-score sign respectively indicated the loss or gain of points.

§3. JUMI

Let s be a taxa and C_s the set of all taxa $i \neq s$ that can be confused with s (see example in Table 7 and complete matrix in appendix). Let k be the cardinality C_s , i.e., the number of taxa that can be confused with s . Let CC_i^s be the confusion coefficient of s with a taxa i . The specificity S_s of s is based on the values of the trophic score CS_s and the coefficient of ecological amplitude E_s .

$$S_s = CS_s \times E_s \quad (35)$$

The JUMI for each taxa s was defined as the difference between the taxa s specificity S_s weighted by the sum of all confusion coefficients CC_i^s and the sum of all taxa i specificities S_i weighted by CC_i^s .

$$JUMI(s) = \sum_{i=1}^k (S_i \times CC_i^s) - S_s \times \sum_{i=1}^k CC_i^s \quad (36)$$

Titre 4 : Results

Section 1: IBMR robustness

All the experiments on floristic lists showed that the IBMR in trophic status varied differently depending on the error type (TAX, AC or TAX+AC) but also on the error rate.

The results for each tested error showed which factor had the highest impact on IBMR.

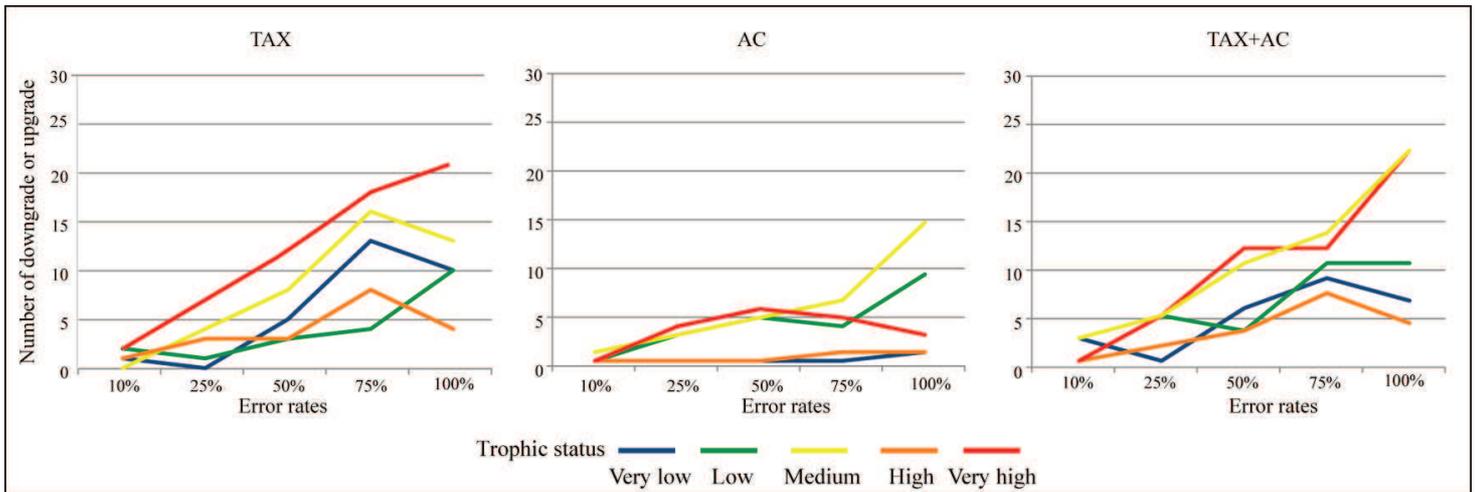


Figure 37: Total number of downgrade or upgrade on each trophic status (Very low, Low, Medium, High, Very high) obtained with the algorithm defined in Section 3: after taxa modification (TAX), abundance classes modification (AC), and taxa and abundance classes modifications (TAX+AC) (for each kind of modification: number of original sites $n=25$, $n=5/\text{error rate}$, number of modified lists 100 tests/site/error rate or 500 tests/error rate)}

Figure 37 shows that the number of class changes increases with the error rate but it is not proportional. Identification and cover class errors (TAX and AC) did not have the same effect on the IBMR score. Indeed, identification errors (TAX) caused 178 class changes (97 downgrades and 81 upgrades) on our data set, corresponding to approximately 7% of the total number of results for this error. The error on cover classes (AC) caused only 98 class changes (77 downgrades and 21 upgrades), which represented two times fewer than the identification errors, and only 4% of the total number of simulated scores for this error. Finally, the two combined errors caused 239 class changes (141 downgrades and 98 upgrades) or roughly 10% of the total number of results for this test. So, in 90% of the cases we did not observe a class change. Moreover, there was no rule about the most impacted trophic status: streams with a very low trophic status were not the most affected by downgrades. In contrast, we found that streams with a very high trophic status can be upgraded.

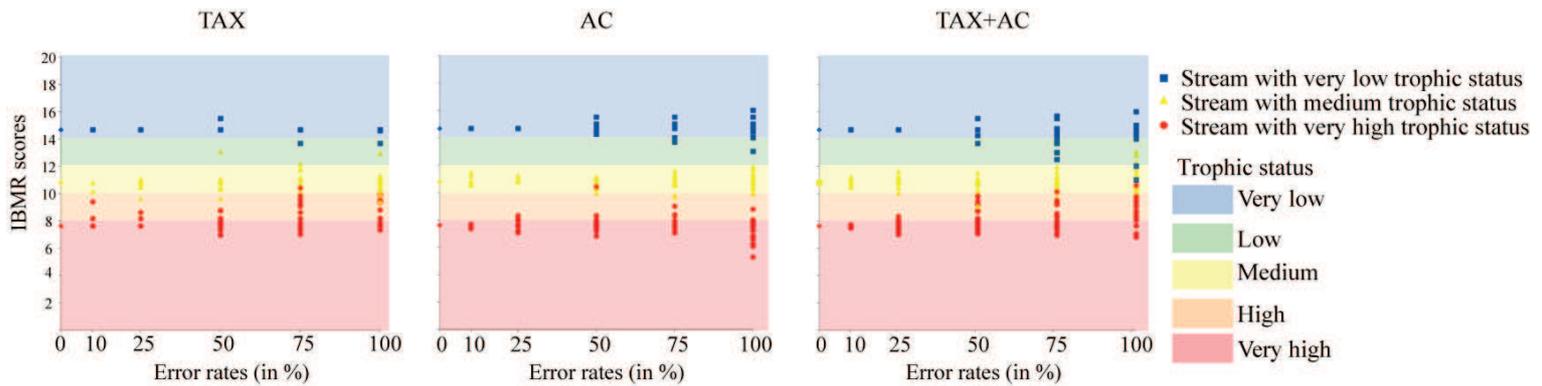


Figure 38 : Variation of IBMR scores obtained with the algorithm defined in Section 3 for three trophic states (Very low, Medium, Very high): after taxa modification (TAX), abundance classes modification (AC), and after taxa and abundance classes modifications (TAX+AC) (for each kind of modification: number of original lists $n=3$, number of modified lists 20 tests/site/error rate or 60 tests/error rate)

The spread of scores (Figure 38) concerned three rivers with three trophic states (very low, medium and very high). The scores of these three rivers are near the class bounds. Indeed, the difference between the class bounds and the three IBMR scores did not exceed a point. The oligotrophic stream had a score of 14.67, the mesotrophic one of 10.79 and the eutrophic one of 7.61.

We observed class changes. For identification errors (TAX), we observed a maximum loss of two trophic status and the gain of one class. One trophic status with the cover classes error (AC) could be lost or gained. For both combined errors (TAX+AC), a loss of two classes and a gain of one class were observed.

The CV_t variation coefficient and the DSt dispersion are given for each error rate, each error type and each trophic status (.

Table 12). CV_t values did not exceed 18% for 10 and 25% error rates. They usually increased, and sometimes irregularly, with error rates reaching a maximum of almost 48% for a 100% error rate (TAX+AC and very high trophic status). Furthermore, for a same error rate CV_t increased with the trophic: for confusion on cover classes and a 50% error rate, CV_t varied from 5.46% for a very low trophic status to 20.41% for a very high trophic status.

CV_t were higher for taxa error (TAX) or the cumulated errors (TAX+AC) than for cover class errors (AC).

DSt responded in the same way. The loss or gain of points increased with the error rate. For an

error rate between 10 and 25% and whatever the trophic status, the observed distance with the reference score is on average 0.81 for TAX, 0.52 for AC and 0.95 for TAX+AC. DS_t increased with the error rate, reaching up to 2.91 for TAX+AC.

Table 12 : Averages of CVt and DSt for each trophic status (Very low, Low, Medium, High, Very high), error type (TAX, AC and TAX+AC) and each error rate. Averages of CVt and DSt were calculated based on 100 tests at 5 sites per trophic status, error type, and error rate. The average of IBMR scores were obtained based on the 5 sites per trophic status.

Error rate		Score average	10%		25%		50%		75%		100%	
Error type	Trophic status		DSt	CVt								
TAX	Very low	14.87	0.63	4.25	0.39	2.60	1.55	10.39	1.98	13.34	1.88	12.67
	Low	12.72	0.36	2.82	1.44	11.35	1.69	13.31	1.90	14.91	2.00	15.72
	Medium	10.67	0.60	5.66	1.17	10.94	1.91	17.85	2.31	21.67	2.56	24.00
	High	8.73	0.53	6.08	0.89	10.15	1.03	11.83	1.31	15.00	1.22	13.99
	Very high	7.21	0.90	12.49	1.17	16.26	1.92	26.63	3.11	43.15	3.08	42.66
AC	Very low	14.87	0.25	1.66	0.36	2.43	0.81	5.46	1.09	7.34	1.44	9.66
	Low	12.72	0.48	3.80	0.93	7.33	1.26	9.90	1.50	11.83	2.12	16.65
	Medium	10.67	0.38	3.60	0.54	5.06	0.93	8.71	1.17	10.93	1.55	14.52
	High	8.73	0.37	4.24	0.64	7.29	0.82	9.38	0.94	10.77	1.02	11.74
	Very high	7.21	0.44	6.09	0.77	10.74	1.47	20.41	1.22	16.99	1.62	22.48
TAX+AC	Very low	14.87	1.12	7.54	0.63	4.23	1.72	11.54	2.31	15.56	3.01	20.22
	Low	12.72	0.41	3.24	1.56	12.23	2.27	17.84	2.78	21.89	3.30	25.95
	Medium	10.67	0.69	6.49	1.37	12.79	2.23	20.93	2.48	23.28	3.11	29.18
	High	8.73	0.40	4.53	1.30	14.91	1.31	15.02	1.52	17.42	1.71	19.61
	Very high	7.21	0.77	10.65	1.29	17.90	1.99	27.59	3.00	41.59	3.40	47.11

Section 2 : Class changes

The probability for streams to be downgraded or upgraded increased with error rate whatever tested errors (TAX, AC, TAX+AC). They are more likely to change classes if their score is near class limit. The probability for a stream to be downgraded was greater than the probability to be upgraded (Table 13). When an experienced surveyor (error rate $\leq 25\%$) performed sampling, oligotrophic streams had a 6.66% likelihood of being downgraded. Mesotrophic streams had a 20.00% likelihood to be downgraded against a 3.33% likelihood of being upgraded. We noted that the streams belonging to the low and medium trophic classes were more likely to be downgraded than the others, which is probably linked to the small interval of these classes.

Table 13: Proportion of downgrading and upgrading (in %) for each trophic status and error type for an experienced surveyor (error rate $\leq 25\%$) for all the 25 sites (600 tests/ trophic status)

Trophic status	Downgrading	Upgrading
Very low	6.66	-
Low	11.52	3.04
Medium	20	3.33
High	4.16	-
Very high	-	11.66

Section 3. Taxa weight

We found that, in most tests, we always had at least one significant or critical error. For each series of tests, we observed one monitoring error in two thirds of the cases. The Z-scores allowed us to identify which tests are problematic and hence the error source (Table 14). Here, it is mainly the confusion on taxa.

Table 14: Minimal, maximal and average of Z-score after the three modifications (TAX, AC, TAX+AC) obtained with the algorithm defined in §2 for each trophic status (Very low, Low, Medium, High, Very high) (average of absolute values, Number of original sites $n=25$, $n=5$ /trophic status, number of modified lists 20 tests/site/ error rate, 100 tests/error rate/error type)

Error rate	Error type	Trophic status	10%			25%			50%			75%			100%		
			Min	Max	Avg												
TAX	Very low		-4.47	4.39	0.18	-2.76	3.56	0.23	-3.86	4.47	0.46	-3.83	3.16	0.53	-3.80	1.48	0.45
	Low		-2.50	4.47	0.22	-4.41	3.67	0.45	-3.49	2.93	0.61	-3.41	3.45	0.57	-2.39	2.68	0.73
	Medium		-4.47	4.01	0.25	-3.30	2.95	0.55	-2.02	3.24	0.68	-2.33	3.20	0.79	-2.21	3.12	0.86
	High		-4.45	3.63	0.37	-2.90	4.21	0.47	-3.63	2.81	0.71	-3.39	3.70	0.71	-2.84	2.79	0.81
	Very high		-4.24	3.66	0.35	-2.05	3.49	0.55	-2.21	4.21	0.61	-2.33	3.27	0.74	-1.78	3.44	0.73
AC	Very low		-4.24	3.23	0.24	-3.14	2.64	0.46	-2.87	3.15	0.65	-2.18	2.33	0.71	-2.64	1.92	0.75
	Low		-3.12	4.47	0.32	-3.49	2.53	0.60	-2.65	2.87	0.72	-2.58	2.42	0.74	-3.26	1.95	0.77
	Medium		-3.32	3.40	0.47	-3.68	3.33	0.57	-2.62	2.53	0.75	-2.45	3.34	0.72	-2.58	2.40	0.81
	High		-3.42	3.45	0.55	-2.76	2.49	0.70	-2.43	1.93	0.76	-2.44	2.99	0.77	-2.52	2.17	0.79
	Very high		-3.11	2.23	0.59	-2.19	3.21	0.66	-2.48	4.24	0.68	-2.02	2.67	0.79	-3.00	1.93	0.83
TAX+AC	Very low		-4.23	3.95	0.26	-2.82	2.40	0.58	-3.46	2.72	0.61	-2.73	2.08	0.74	-3.38	1.50	0.67
	Low		-3.12	4.29	0.34	-3.28	3.24	0.61	-2.68	3.09	0.74	-2.80	3.31	0.76	-3.01	2.51	0.83
	Medium		-4.19	3.66	0.49	-2.67	3.31	0.60	-2.81	3.17	0.78	-2.26	2.40	0.83	-2.17	2.68	0.88
	High		-2.49	2.83	0.62	-3.12	2.75	0.71	-3.00	3.08	0.74	-3.54	2.77	0.76	-2.40	2.93	0.76
	Very high		-3.70	4.29	0.52	-3.17	3.39	0.64	-1.94	3.11	0.81	-2.42	3.08	0.76	-2.59	3.77	0.83

Table 15: Minimal, maximal and average of Z-score after the three modifications (TAX, AC, TAX + AC) obtained with the algorithm defined in §2 based on a random selection for three trophic states (very low, medium, very high) (average of absolute values). Number of original site $n = 1/\text{trophic status}$, number of modified lists 20 tests/error rate/error type, same sites as Fig. 38).

Trophic status	Very low					Medium					Very high				
	10%	25%	50%	75%	100%	10%	25%	50%	75%	100%	10%	25%	50%	75%	100%
Min TAX	0.00	0.00	0.00	-2.73	-2.74	-4.47	-3.04	-1.61	-1.66	-1.86	0.00	0.00	-1.58	-0.60	-0.33
Max TAX	0.00	0.00	4.47	0.00	0.00	0.00	0.74	3.18	2.18	2.82	3.22	3.49	2.69	2.75	2.77
Avg TAX	0.00	0.00	0.22	0.41	0.42	0.22	0.48	0.57	0.73	0.71	0.37	0.46	0.63	0.93	0.77
Min AC	0.00	0.00	-1.51	-2.16	-2.41	-1.15	-0.14	-2.38	-2.45	-1.51	-3.11	-2.04	-2.06	-1.08	-3.00
Max AC	0.00	0.00	2.98	1.79	1.92	3.40	3.33	1.09	1.75	1.86	0.84	2.60	1.81	2.67	1.51
Avg TAX	0.00	0.00	0.54	0.58	0.66	0.44	0.51	0.73	0.67	0.78	0.40	0.55	0.72	0.74	0.71
Min TAX+AC	0.00	0.00	-2.63	-2.69	-3.38	-1.82	-2.41	-2.81	-2.26	-1.86	-2.67	-2.05	-0.70	-0.76	-0.69
Max TAX+AC	0.00	0.00	2.19	1.24	1.22	2.52	2.66	1.21	1.80	2.06	1.91	2.36	2.75	2.85	2.82
Avg TAX+AC	0.00	0.00	0.43	0.76	0.57	0.52	0.57	0.68	0.75	0.79	0.44	0.58	0.83	0.85	1.07

Table 14 shows that for each error type (TAX, AC and TAX+AC) and regardless of the error rate, there were significant or critical errors. From the results of Z-Scores on IBMR scores in each stream for the three trophic states in Table 15, we can conclude that we find less significant errors for low error rates (10 and 25%) than for high error rates (50, 75 and 100%). Regarding the Z-scores on specific richness, taxa omissions had an impact on the IBMR index assessment. For a 10-25% error rate, omitting a single taxa is enough to indicate a significant error. The number of taxa omissions leading to critical error (Z-score >3) increased with the increase of error rate.

JUMI provided the specificity of each confusable taxa (Table 16) and indicated the impact of identification error (TAX) on IBMR score. The JUMI sign showed if the misidentification led to an increase or a decrease in the score. We can mention as an example the case where confusion was possible between two taxa only (Table 16), when *Sparganium emersum* (SPAEML) was confused with *Sparganium erectum* (SPAERE), JUMI value was 0.75, meaning that IBMR score increased. On the contrary, when SPAERE was confused with SPAEML, the IBMR score decreased.

All the taxa misidentifications did not have the same effect on the IBMR score. Some of them had limited impact while others caused high variability of IBMR. Among the most impactful (JUMI > |7|) we can cite *Cratoneuron filicinum* (CRAFIL), *Ranunculus trichophyllus* (RANTRI), *Stigeoclonium sp.* (STISPX), *Stigeoclonium tenue* (STITEN), *Potamogeton polygonifolius* (POTPOL), and *Rhynchostegium riparioides* (RHYRIP).

Table 16 : Values of major confusion between Macrophytes Indicator (JUMI) for each taxa able of being confused (74 taxa among the 219 IBMR taxa, in blue JUMI $\geq |7|$)

Taxa	JUMI	Taxa	JUMI	Taxa	JUMI	Taxa	JUMI
CINRIP	-0.40	CALSTA	-2.90	RANCAL	-3.0	MOUSPX	-1.60
CINFON	0.40	CALPLA	2.78	RANPEL	-1.00	MICSPX	-3.30
BRARIV	-4.60	CALOBT	-0.38	RANOMI	-3.50	CARNC	6.9
AMBRIP	4.90	CALHAM	-2.54	RANOLO	-3.50	CALBRUNC	3.04
AMBFLU	3.20	BERERE	-3.60	RANFLU	-1.60	RHYRIP	8.60
PORPIN	0.60	APINOD	3.60	POTPOL	-7.80	CYANNC	2.85
CHIPOL	-0.60	AGRSTO	-5.0	POTPEC	1.65	CHAMAJ	3.0
STITEN	11.5	ACOCAL	-1.10	POTNOD	3.90	CHAPOL	3.0
STISPX	-11.5	SPISPX	1.60	POTBER	-1.95	PHAARU	0.80
RANPEN	4.0	RANTRI	7.0	POTNAT	3.90	PHAAUS	-0.80
LEMMIN	-2.0	IRIPSE	1.10	ELONUT	1.0	ELOCAN	-1.0
CARVES	-0.30	CARROS	-6.6	SCHRIV	0.90	RHAACI	-1.80
ORTRIV	0.90	FONSQU	-3.80	FONANT	3.80	FISVIR	1.80
DREFLU	-0.90	CRAFIL	-11.20	RHISPX	2.10	FISRUF	-1.80
SPAERE	-0.75	SPAEML	0.75	PHOSPX	-3.40	FISPUS	-1.80
LYNSPX	-0.85	DIASPX	-4.80	OSCSPX	1.40	FISCRA	1.80
POLHYD	-1.60	TYPANG	-0.80	TYPLAT	0.80	ZANPAL	1.90
DIANC	4.80	LMNC	2.0	PTNC	5.0	PNC	1.6
CLASPX	1.20	CHAHIS	-6.00				

Titre 6 : Discussion

In ecology, the study of uncertainties is usually focused on field experiments or on tests by surveyors (Furse, et al., 1981; Prygiel, et al., 2002; Kilroy, et al., 2013; Kolada, et al., 2014).

Our study highlighted the uncertainty sources associated to the surveyor's experience (taxa misidentification and abundance class error) and based on a theoretical approach and how each error type affected the IBMR index. We assigned each theoretical error rate ($\mathcal{E}_t = 10, 25, 50, 75$

and 100%) to a theoretical surveyor' skill level. Thus, the 10-25% range corresponded to skilled surveyors, 50% moderately experienced and between 75 and 100% inexperienced (e.g. a student). The index robustness and the weight of some taxa included in the index calculation were analysed.

The results showed that the more inexperienced the surveyor (error rate $\mathcal{E}_t > 25\%$), the more the index varied regardless of the error type and the trophic class. The confusion on taxa (TAX) and the combined errors (TAX+AC) always had a stronger impact than the confusion on abundance classes (AC), which confirmed the observation made in field experiments by some hydrobiologists (Prygiel, et al., 2002; Haase, et al., 2006). For a theoretical error rate, \mathcal{E}_t , lower than or equal to 25%, the differences could reach 0.6 points on the index, whereas for 50, 75 and 100% they could exceed one point.

When considering the real error rate, \mathcal{E}_r , we had the following results (\mathcal{E}_r being always lower than \mathcal{E}_t by construction). For an inexperienced surveyor, \mathcal{E}_r was on average higher than 16%, for a moderately experienced \mathcal{E}_r equal to 11%, and for an expert \mathcal{E}_r lower than 5%.

The level of surveyor experience plays a major role during protocol application as it has already been shown by Kilroy (2013), and Prygiel (2002) in studies based on a field approach. But in contradiction with Staniszewski (2006) which showed the smallest role of surveyor as compared with habitat variation, but their study was based only on experienced surveyor.

We observed that, regardless of the error rate, the IBMR index increased or decreased by more than 0.5 points, which is considered as indicating a real change in environmental conditions by some experts (personal communication, Haury, J., 2012). In class limits, it could lead to a status modification. This phenomenon can be observed for mesotrophic streams which are more often misclassified than other trophic types. Staniszewski (2006) showed that the probability of misgrading a site in a moderate trophic status was higher for the Ellenberg index (50%) than for the IBMR (15%), and the variation depending on several factors such as shadows or experienced surveyor effect. The IBMR scores for selected sites with a medium trophic status were closed to the inferior limit boundary, so they were more likely to be downgraded than to be upgraded.

Returning to floristic lists allowed us to understand significant differences in IBMR scores and class changes. The taxa identification step in-situ and/or in laboratory seems to be crucial. Misidentification caused class changes. However some taxa had a bigger impact than others. For example confusion between *Ranunculus trichophyllus* (RANTRI) and *Ranunculus ololeucos*

(RANOLO) as well as between *Potamogeton polygonifolius* (POTPOL) and *Potamogeton nodosus* (POTNOD) systematically induced an IBMR score decrease and even a downgrade. We proposed the major confusion between Macrophytes Indicator (JUMI) as an indicator for measuring the weight of taxa misidentification based on the taxa confusion matrix, and on the trophic score CS_i and coefficient of ecological amplitude E_i of each taxon. JUMI highlighted these taxa as having the highest impact on IBMR scores. To understand why these confusions caused IBMR score changes, it is necessary to look at the values of CS_i and E_i associated with each taxa: e.g., for *Ranunculus ololeucos* (RANOLO) and *Ranunculus trichophyllus* (RANTRI), CS_i ranged from 19 to 11 and E_i from 3 to 2. The decrease or increase in the IBMR score is directly related to the difference between CS_i values. It is thus necessary to pay special attention to the taxa which can be mistaken and whose coefficients are distant. In the field, these species and others such as the algae *Rhizoclonium sp.* (RHISPX), the Bryophyte *Amblystegium riparium* (AMBRIP) or the Spermatophyte *Potamogeton polygonifolius* (POTPOL) must be sampled systematically to ensure the accuracy of their identification.

Taxa omissions might also cause a class change and in consequence a significant change in IBMR scores, as the IBMR calculation is based on a logarithmic formula which favours low cover species, often the better bio-indicators. Pondweeds with filiform leaves such as *Potamogeton panormitanus*, *Potamogeton berchtoldii* or *Zannichellia* are a good example. Therefore it is necessary, to be careful in the field, and make sure low cover and high CS_i taxa are included in the survey.

Titre 7 : Conclusion

Index sensitivity to the surveyor's skills requires not only control rules to limit the impact of mistakes but also a certain level of training for surveyors before their surveys can be used in IBMR calculation. Adequate surveyor training seems to be the best way to obtain reliable results (Haase, et al., 2006; Kilroy, et al., 2013).

Numerous uncertainty studies on benthic macroinvertebrates e.g. (Haase, et al., 2006) have highlighted the importance of developing a quality control approach, such as rigorous training programs and auditing schemes in order to minimize the impact of the uncertainty. Studies on diatoms e.g., (Prygiel, et al., 2002), proposed the development of inter comparison identification exercises and the creation of reference collections for difficult species.

In a tool based on macrophytes, a quality control approach should involve a systematic search for mistaken taxa in the field and sending pre-identified samples to experts for verification. The different tests which we carried out showed that efforts in terms of quality and control should focus on taxa identification by favoring taxa with distant CS_i . Abundance class errors were less impacting and more difficult to correct even with training and experience. A solution might be an in-situ confrontation between two surveyors and their abundance class estimations, especially when they are novices.

PARTIE 3 Discussion générale

L'objectif de cette thèse est double, puisqu'il s'agit : (1) d'étudier l'impact de certaines incertitudes sur deux catégories d'indice biologique, l'une concernant les macroinvertébrés et l'autre les macrophytes en rivière et (2) de quantifier ces incertitudes et d'estimer leur effet sur l'évaluation globale de la qualité écologique des rivières.

Nous avons choisi, pour répondre à ce double objectif, de mettre en place deux stratégies qui nous ont permis de mettre en évidence les variabilités liées aux substrats et à l'opérateur pour quatre indices français : l'Indice Biologique Global Normalisé (IBGN), l'Indice Biologique Global DCE, l'I2M2 et l'Indice Biologique Macrophytique en Rivière (IBMR) (Titre 1). Dans le titre 2, nous critiquons les méthodes mises en œuvre. Puis, nous discutons de l'extension et du transfert possible des démarches proposées à d'autres compartiments biologiques (Titre 3). Enfin, nous nous interrogeons sur l'impact des incertitudes sur la combinaison des deux indices étudiés et sur celle de l'ensemble des cinq indices biologiques français (IBG DCE, IBMR, IOBS, IBD et IPR) (Titre 4).

Titre 1 : De la mesure de la variabilité à l'incertitude

Section 1 : Les macroinvertébrés

Des variabilités existent à chaque étape du processus de l'évaluation de l'état écologique des rivières. Elles n'ont, cependant, pas toutes la même influence sur l'évaluation et certaines peuvent même être minimisées.

Pour les macroinvertébrés, les variabilités liées à l'environnement (Erman & Erman, 1984; Beisel, et al., 1998a) ou à la saisonnalité (Murphy, 1978; Armitage, et al., 1983) sont connues et peuvent être limitées à l'aide des différentes méthodes d'échantillonnage des indices biotiques. En effet, dans le cadre d'un réseau de surveillance par exemple, l'effet de la saisonnalité peut être atténué par un échantillonnage annuel toujours réalisé durant la même période de l'année. La variabilité liée à la mosaïque d'habitats est difficilement maîtrisable, seuls la bonne pratique et le respect de normes peuvent garantir une meilleure régularité dans les résultats.

Notre première approche expérimentale basée sur la répétition des substrats confirme une réelle variabilité associée aux types de substrats, distincte de la variabilité opérateur. Cette variabilité n'est pas équivalente entre substrats et elle se propage différemment au travers des différentes métriques des indices français (IBGN et I2M2) (Figure 39). Nous observons ainsi que certains substrats présentent une variabilité supérieure à d'autres, que cette variabilité peut être en partie reliée à la capacité biogène des substrats mais également à leur complexité, à savoir des granulométries hétérogènes ou la présence de substrats secondaires comme des litières dans les sables. Nous avons pu observer, grâce aux indices de similarité, qu'en général les substrats organiques sont plus variables que les substrats minéraux, à l'exception des sables. Les sables et les surfaces dures sont deux substrats particuliers puisqu'ils peuvent être associés à un substrat secondaire, les variations observées reflètent en réalité la variation de deux substrats associés. D'autres effets peuvent expliquer la variabilité des sables, comme par exemple la stabilité du substrat ou l'effet de la mosaïque d'habitat. La complexité provient donc de l'hétérogénéité d'un substrat, puisque ceux-ci sont rarement homogènes notamment du fait de la part plus ou moins

importante d'un substrat secondaire. La complexité des substrats joue un rôle majeur dans la distribution des peuplements, l'abondance des individus et la diversité taxonomique. Cette variabilité se traduit également à l'échelle des métriques des deux indices français (Figure 39). Sur cette figure, dans un souci de lisibilité, nous avons regroupé les métriques en fonction des résultats précédents (Partie 2, Chapitre 1). Les métriques des deux traits biologiques (polyvoltinisme et ovoviviparité) sont très variables (>33%), quelque soit le substrat observé. Les métriques relatives à la richesse taxonomique (TAX et VC), à la polluosensibilité (ASPT et groupe indicateur) et l'indice de Shannon montrent une variation inférieure à 33,3% quelque soit le substrat, mais présentent tout de même une plus grande variabilité pour les substrats organiques que pour les substrats minéraux, excepté les sables et surfaces dures pour les deux traits biologiques.

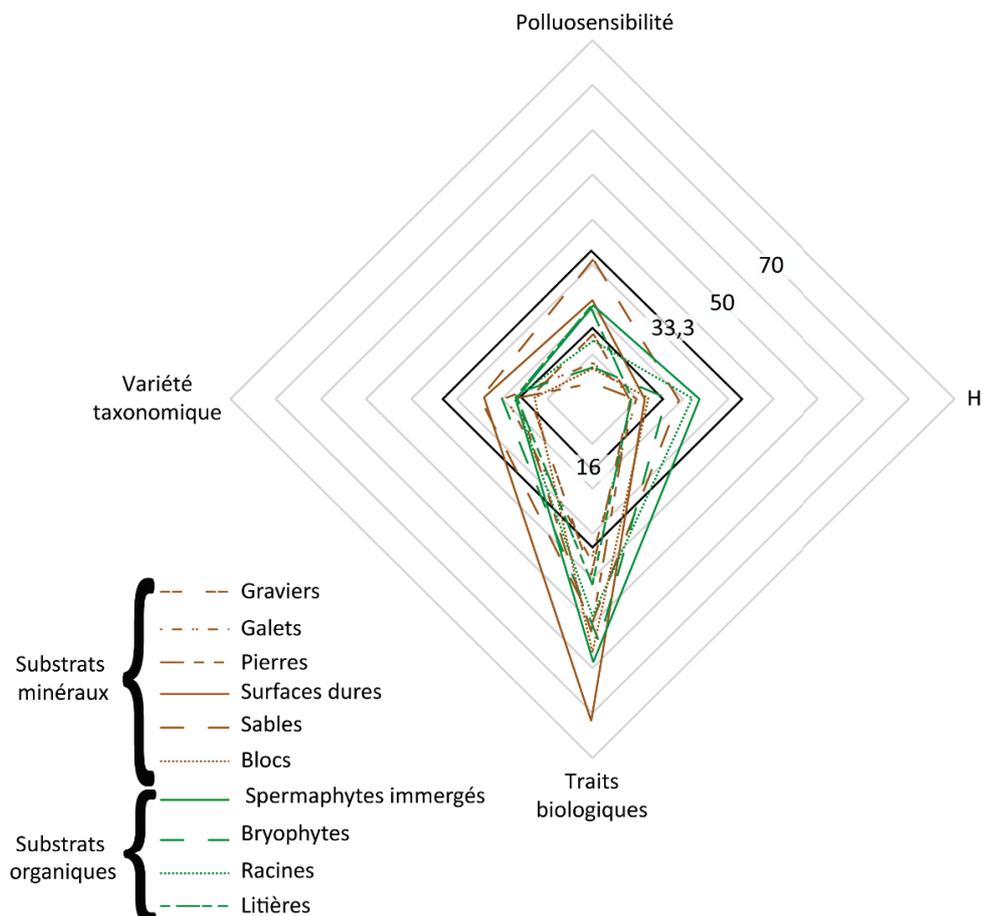


Figure 39 : Coefficients de variation (CV) moyens obtenus pour chaque substrat et par regroupement de métriques de l'I2M2 et de l'IBG DCE (Traits biologiques = Fréquence des individus polyvoltins PS + fréquence des individus ovovivipares OS; Variété taxonomique = richesse taxonomique TAX + classe de variété VC, H = indice de Shannon et Polluosensibilité = ASPT + Groupe indicateur GI). CV <16% pas d'impact significatif de l'erreur, 16% < CV < 33,3% erreur est significative et CV >33,3% erreur critique.

Les listes faunistiques et les différentes métriques indiquent chacune, l'opposition entre substrats minéraux, peu variables, et organiques, très variables. Néanmoins, l'expérimentation menée sur la variabilité intra-substrat ne nous a pas permis de prédire leur propagation dans le cadre d'un échantillonnage complet, dans les indices.

La variabilité intra-substrat n'est pas la seule à gouverner les écarts de mesure puisque la variabilité associée à l'opérateur intervient à chaque étape des protocoles d'évaluation. Les traitements réalisés sur les données de l'essai inter-laboratoires d'AGLAE, montre que pour trois substrats différents (racines, pierres et spermaphytes immergés), les variations taxonomiques observées peuvent être supérieures à celles estimées lors de notre expérimentation. Ainsi, à la variabilité intra-substrat s'ajoute la variabilité liée aux différents opérateurs, qui se répercute différemment selon les métriques. Il peut y avoir un effet de cumul des deux variabilités, comme par exemple pour l'ASPT, la richesse (TAX), l'indice de Shannon (H) et la fréquence des individus ovovivipares (OS) pour les pierres, ou bien un effet de compensation partielle, puisque les coefficients de variation des métriques, comme par exemple pour la fréquence des individus polyvoltins (PS) pour les pierres, peuvent être inférieurs à ceux de l'expérimentation intra-substrat (Figure 40).

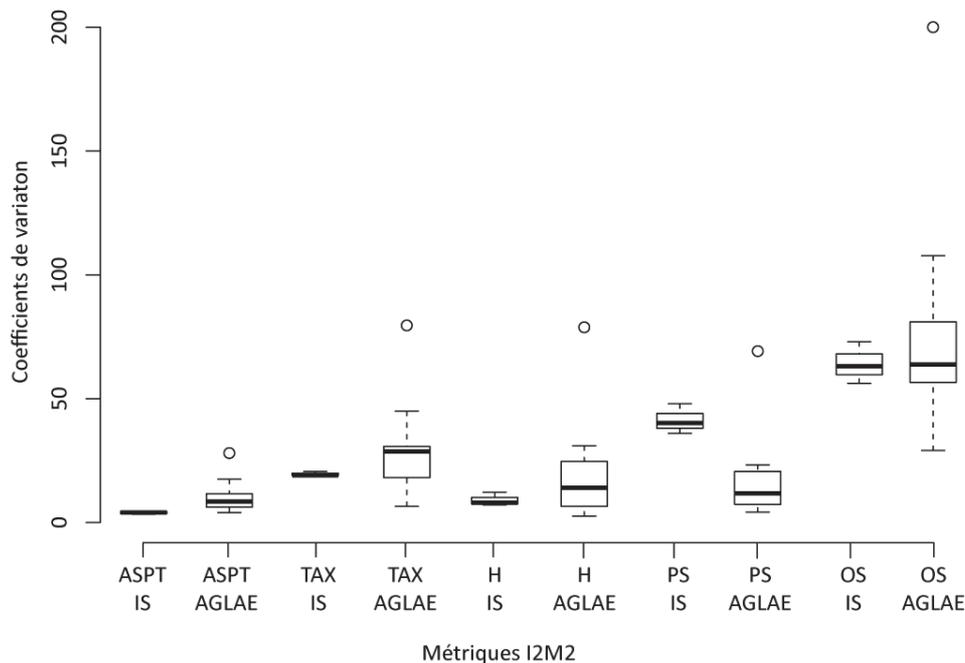
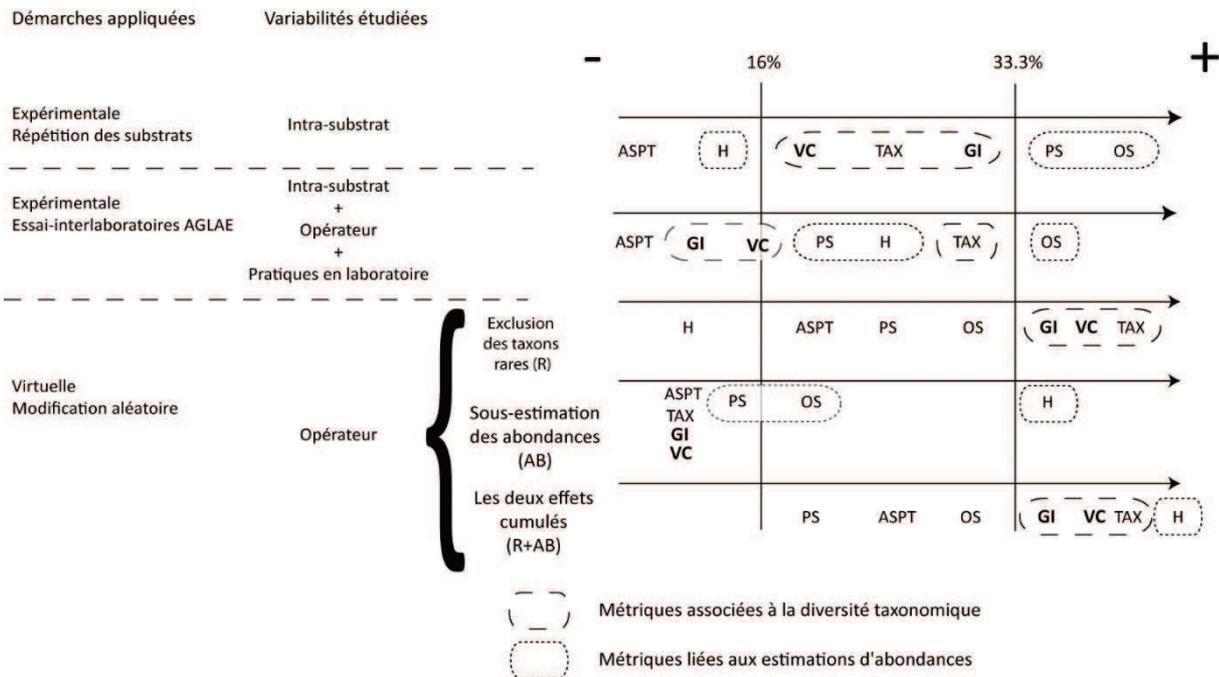


Figure 40 : Comparaison des coefficients de variation des 5 métriques de l'I2M2 (ASPT, Richesse TAX, indice de Shannon H, Fréquence des individus polyvoltins PS et ovovivipares OS) obtenus pour les pierres avec l'expérimentation intra-substrat (IS) et l'expérimentation AGLAE. Pour IS $n=3$ et pour AGLAE $n=16$, $CV < 16\%$ pas d'impact significatif de l'erreur, $16\% < CV < 33,3\%$ erreur est significative et $CV > 33,3\%$ erreur critique.

La variabilité opérateur peut être due aux différentes pratiques au laboratoire et aux matériels utilisés. En effet, la norme XP T90-388 (AFNOR, 2010) ne précise pas comment doivent être estimées les abondances et ne donne pas une démarche stricte pour réaliser le tri. Ainsi, certains laboratoires trient sous une lampe loupe, tandis que d'autres trient la petite fraction (500 microns) sous une loupe binoculaire. Il pourrait être envisagé d'harmoniser les pratiques afin de réduire cette variabilité associée à l'opérateur du tri. Mais même en uniformisant les protocoles, il existera toujours une incertitude liée à l'opérateur. Aquabio a testé, dans une approche limitée (sur 6 IBGN et 2 opérateurs), l'incertitude liée à l'effet opérateur au laboratoire. Il a été montré que même avec l'application d'une même technique de tri au laboratoire et une expérience équivalente entre opérateurs, il existe une incertitude de 10% sur les estimations des abondances et 5% sur la diversité taxonomique (Tartare, 2004).

La deuxième expérimentation que nous avons menée, est basée sur des modifications virtuelles, ce qui a permis de tester l'introduction aléatoire d'erreurs simulant l'effet lié à l'expérience de l'opérateur, pouvant affecter les notes des indices français pour différentes qualités écologiques. Nous constatons que certaines métriques sont relativement stables et peu variables, tandis que d'autres sont très variables. Les oublis de taxons et les sous estimations des abondances des taxons sont deux facteurs qui jouent un rôle important et s'imbriquent dans le calcul des métriques et des indices eux-mêmes. Les métriques étant plus ou moins sensibles en fonction des types d'erreurs testées, nous distinguons deux catégories : (1) les métriques les plus affectées, ce sont celles qui sont dépendantes de la diversité taxonomique (TAX, VC) et celles prenant en compte des estimations des abondances (H, PS et OS) (2) des métriques peu sensibles (ASPT, GI) (Figure 41) classées selon les coefficients de variation moyens. Une synthèse de la variabilité des métriques peut être proposée grâce à nos différentes expérimentations (Figure 41).



D'autres études menées sur la variabilité opérateur, ont obtenu des résultats similaires aux nôtres (Archambault, et al., 2014) et ont montré qu'il existait une distinction entre les métriques et leurs réponses face aux différentes sources de variabilité.

Nous avons également observé que les classes de qualité médiane sont les plus sensibles aux erreurs et l'incertitude mesurée pour ces classes est plus importante que pour les autres, ceci a également été observé par d'autres auteurs (Cao, et al., 1998; Archambault, et al., 2014).

La présence d'incertitude n'est pas limitée aux seules étapes techniques du terrain et du laboratoire, mais elle peut également être fonction de la retranscription entre la liste faunistique établie au laboratoire et la liste faunistique finale (Hose, et al., 2004; Metzeling, et al., 2003; Stribling, et al., 2008; Haase, et al., 2010). Les erreurs de retranscription vont également affecter les résultats et l'évaluation finale de l'état écologique des rivières, mais nous ne les avons pas traitées dans cette thèse.

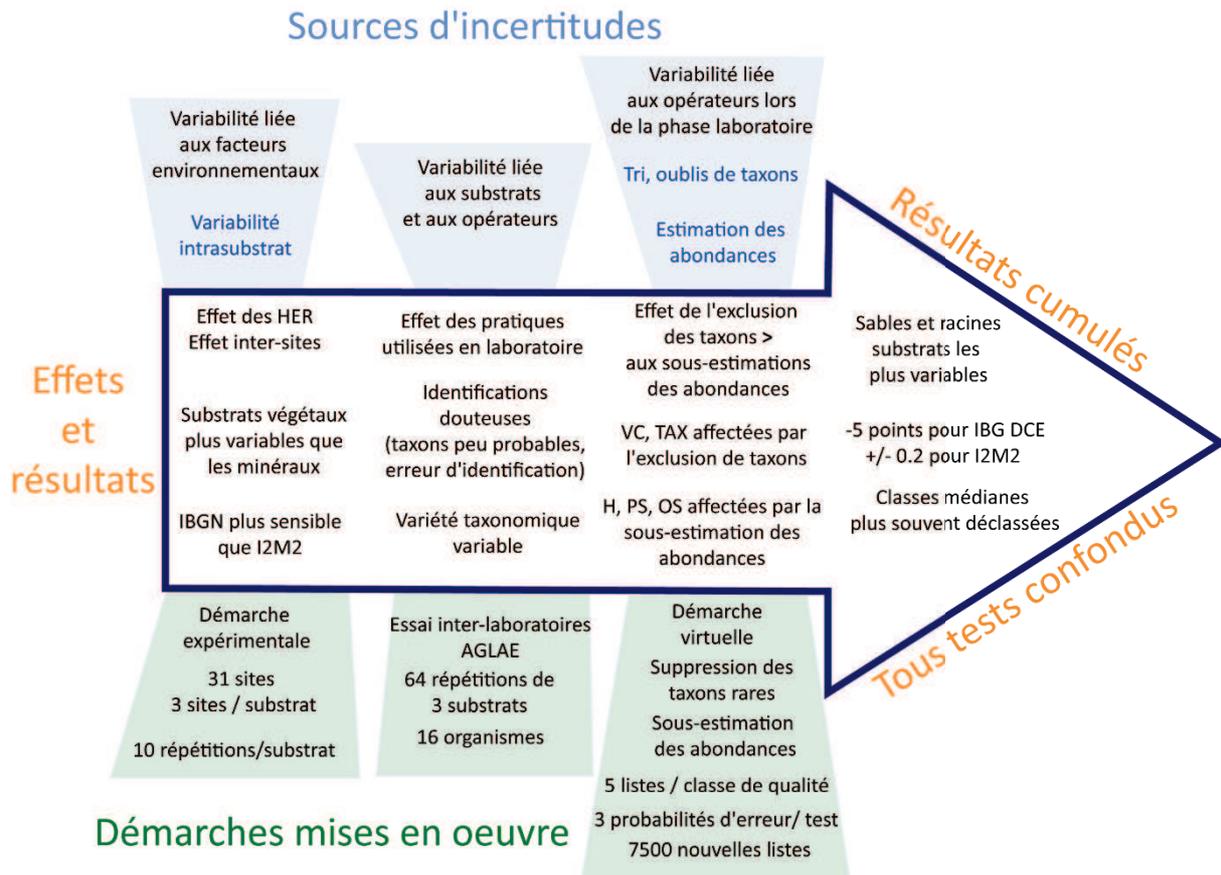


Figure 42 : Propagation des variabilités étudiées au sein des protocoles d'évaluation de l'état écologique des rivières basés sur les macroinvertébrés, sources d'incertitudes, démarches mises en oeuvre, effets et résultats

La figure 42 schématise les différentes sources de variabilités analysées, les démarches mises en place pour les traiter, les principaux effets de ces variabilités et leur impact sur les indices et leurs métriques. Chaque source de variabilité engendre une incertitude qui se répercute sur les indices : des écarts de notes peuvent se produire ainsi que des erreurs d'évaluation de l'état écologique. La démarche virtuelle basée sur les erreurs associées à l'expérience de l'opérateur, a montré, au maximum, une variation de l'IBG DCE de 5 points d'écart, soit 25% de la note et une variation de l'I2M2 de 0,2, soit 20% de la note. La variation maximale observée est donc équivalente pour les deux indices. Les classes intermédiaires sont les plus concernées et les plus affectées par la variabilité liée à l'opérateur.

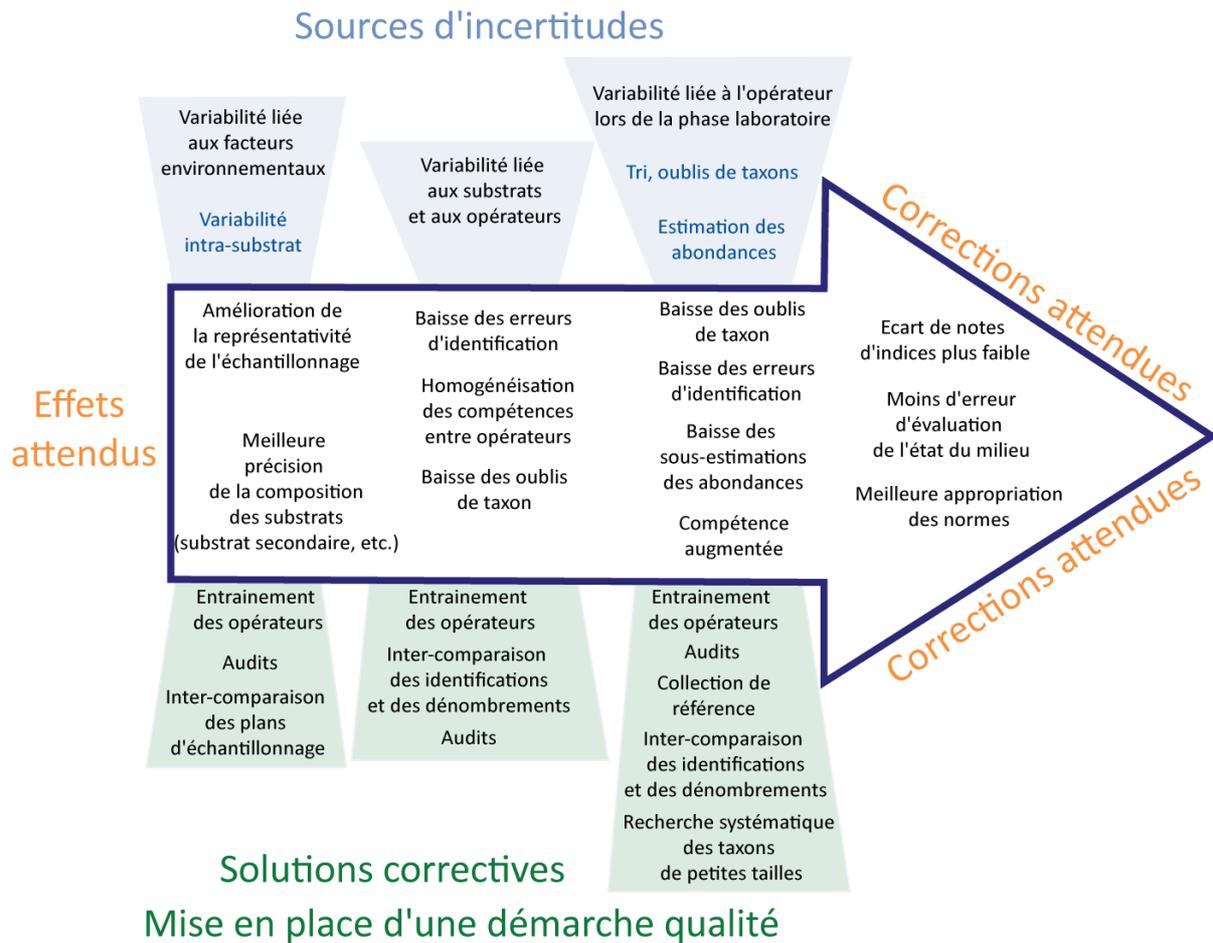


Figure 43: Sources de variabilités au sein des protocoles d'évaluation de l'état écologiques des rivières basés sur les macroinvertébrés, solutions correctives et attente d'amélioration des résultats

Les incertitudes découlant des variabilités observées, hormis les variabilités naturelles, peuvent être réduites grâce à différentes solutions correctives (Figure 43). En effet, l'entraînement et le suivi des opérateurs de terrain peuvent permettre une meilleure appropriation des protocoles d'échantillonnage et ainsi garantir une amélioration de la réalisation des échantillonnages et une meilleure description des habitats. De même, l'entraînement des opérateurs au laboratoire, la mise en place d'inter-comparaisons des estimations des abondances, la recherche systématique des taxons de petite taille et la vérification des identifications devraient permettre de réduire les effets de ces facteurs et ainsi générer une amélioration de la qualité et de la fiabilité des résultats. Ces solutions ont également été proposées par différents auteurs dans les études d'incertitudes (Metzeling, et al., 2003; Archaimbault, et al., 2014). Ces moyens correctifs reposent essentiellement sur le perfectionnement des opérateurs mais également sur l'appropriation des normes afin de limiter les erreurs.

Section 2 : Les macrophytes

Tout comme pour les protocoles basés sur les macroinvertébrés, les protocoles macrophytiques sont affectés par des variabilités qui s'expriment à chaque étape du processus de l'évaluation de l'état trophique des rivières. Ces variabilités correspondent à différents facteurs, qui n'ont pas le même impact sur l'indice IBMR et peuvent également être minimisées.

Jusqu'à présent les principales études de variabilité et d'incertitudes menées sur les indices macrophytiques se sont intéressées préférentiellement aux effets des facteurs environnementaux sur la distribution et la présence des macrophytes. Le nombre d'espèces dépend de la saison mais également de l'ombrage et de la dégradation hydromorphologique (Breugnot, et al., 2008; Staniszewski, et al., 2006; Szoszkiewicz, et al., 2009). Les métriques des indices basés sur les macrophytes ont, jusqu'à présent, montré une faible variation face aux modifications des facteurs environnementaux. Cependant, il est à noter que dans le cadre d'un suivi écologique des rivières basé sur les macrophytes, les échantillonnages, en principe, sont réalisés en période de développement de la végétation et en basses eaux (AFNOR, 2003). Ainsi, l'impact des facteurs environnementaux est réduit dans le cadre d'un suivi pluriannuel.

Les effets du milieu ainsi écartés, il reste cependant l'impact de l'opérateur et de son niveau d'expérience. Cet effet a déjà pu être traité dans certaines études, qui concernaient des opérateurs chevronnés. De ce fait, l'effet n'a jusqu'ici été considéré comme minime et ne représentant pas une réelle source de variabilité sur les métriques et les indices (Staniszewski, et al., 2006; Mascaro, et al., 2013). Or l'expérience de l'opérateur semble être déterminante dans l'évaluation trophique des milieux aquatiques. C'est pourquoi nous nous sommes intéressés à cette question et avons utilisé une approche virtuelle analogue à celle développée sur les macroinvertébrés, afin d'appréhender la propagation de la variabilité liée aux compétences de l'opérateur sur l'IBMR. Notre approche, bien que virtuelle, tend à refléter la réalité puisqu'elle s'appuie sur deux matrices de confusion établies à l'aide d'experts. La première décrit les erreurs d'identification des taxons concernant les 208 taxons contributifs à l'IBMR et onze taxons supplémentaires non contributifs avec lesquels une confusion est possible. Au total 82 taxons sont concernés et les taux de confusion attribués résultent de dires d'experts. La deuxième porte sur les erreurs d'estimation des cinq classes de recouvrement : des taux de confusion entre classes

ont été attribués, les confusions possibles étant cantonnées aux classes adjacentes. Les mauvaises identifications et erreurs d'estimations des classes de recouvrement générées sont donc gouvernées par un modèle réaliste qui repose sur les connaissances de terrain et de laboratoire d'un « macrophytiste » expérimenté (6 années d'expérience, Aquabio).

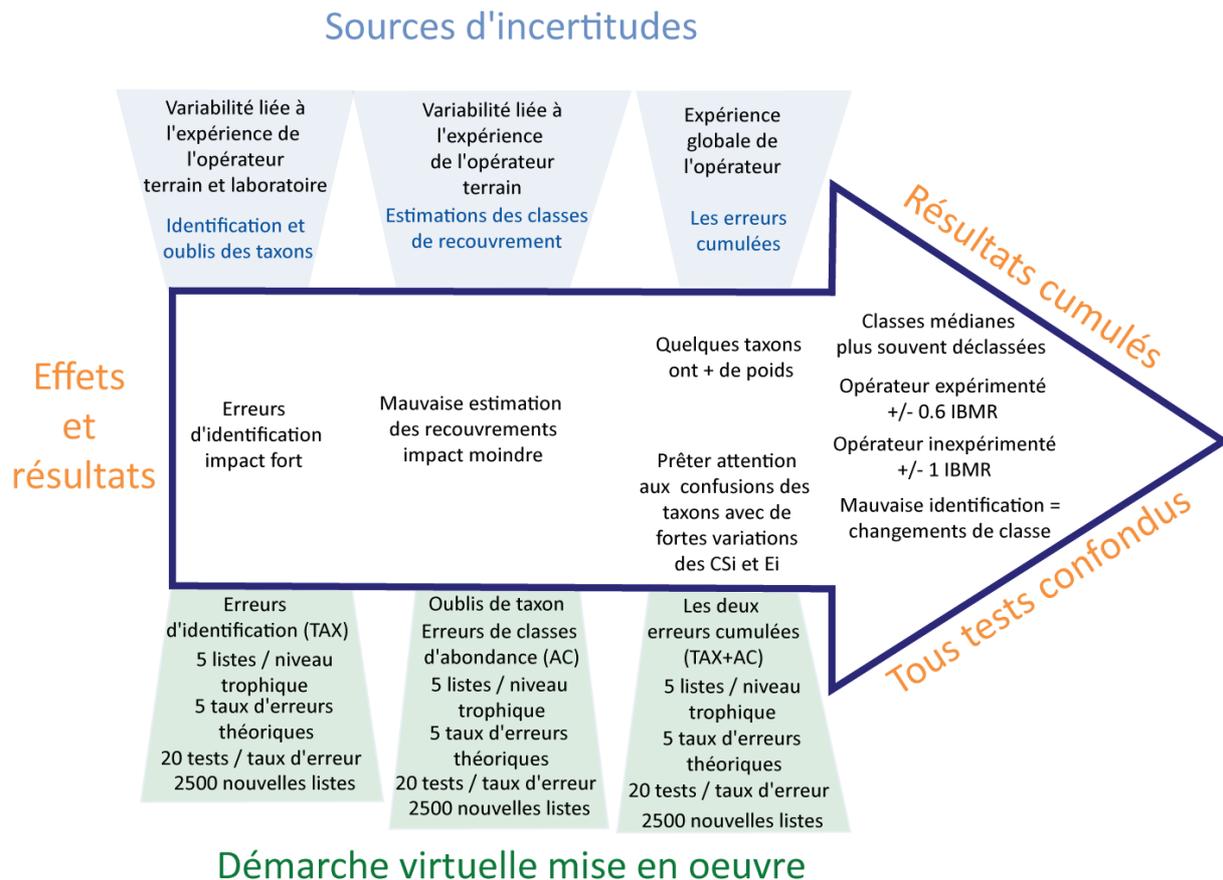


Figure 44: Propagation des variabilités étudiées au sein du protocole d'évaluation de l'état trophique des rivières basé sur l'IBMR, sources d'incertitudes, méthodes d'analyses et erreurs observées

La figure 44 schématise le processus de propagation des variabilités liées à l'opérateur. Suite à nos différents tests, nous constatons que la source majeure de variabilité est liée aux erreurs d'identification des taxons. Celle-ci peut affecter l'IBMR, voire mener à un changement de classe de qualité trophique, lorsque les deux types d'erreurs sont cumulées : nous avons constaté un changement de classe dans 10% des cas. Ces erreurs d'identification peuvent également induire des interprétations douteuses et faussées, celles-ci ne se basant pas uniquement sur la note IBMR mais aussi sur les différentes espèces présentes. L'expérience de l'opérateur joue un rôle

important, puisque ses compétences sont au cœur de l'application de la norme NF T90-395 (AFNOR, 2003). Nous avons modélisé cette notion grâce aux différents taux d'erreurs théoriques. La note IBMR peut osciller de plus ou moins 1 point lorsque le protocole de relevé est appliqué par un opérateur inexpérimenté (taux d'erreur théorique > 25%) et d'environ 0,6 point lorsqu'il s'agit d'un opérateur expérimenté (taux d'erreur théorique < 25%).

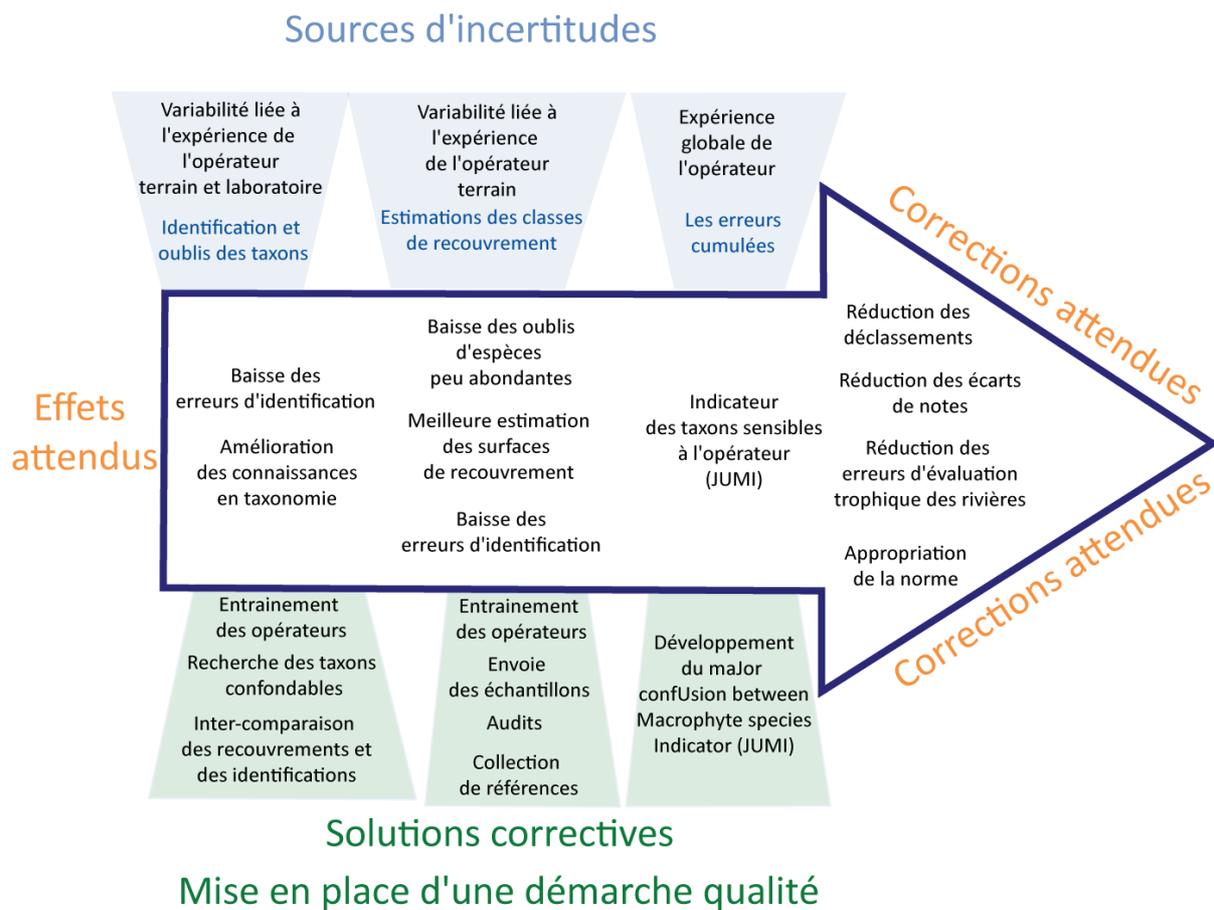


Figure 45: Sources de variabilités au sein des protocoles d'évaluation de l'état écologiques des rivières basé sur l'IBMR, solutions correctives et attente d'amélioration des résultats

A nouveau, des solutions correctives pourraient être mises en place afin d'améliorer la qualité des résultats obtenus (Figure 45). En effet, l'entraînement et le suivi des opérateurs sur le terrain et au laboratoire devraient permettre de réduire les erreurs d'identification, d'estimation des classes de recouvrement mais également prévenir l'oubli de taxons. Le maJor confUsion between Macrophyte species Indicator (JUMI), développé dans le cadre de cette thèse, pourrait également être utilisé pour donner aux opérateurs une indication quant aux taxons susceptibles de créer une

plus forte variation en cas d'erreurs d'identification. La diminution des erreurs d'identification des taxons, principalement, devrait mener à une diminution du risque de changement de classe et à une réduction des écarts de notes.

Tout comme pour les macroinvertébrés, les erreurs de retranscription entre les listes floristiques établies sur le terrain ou au laboratoire et les listes floristiques finales sont possibles. Celles-ci peuvent dans une moindre mesure, comme pour les macroinvertébrés, ajouter un biais au résultat final. Mais nous ne les avons pas traitées ici.

Titre 2 : Critiques et pistes d'améliorations des méthodes d'analyses utilisées

Les démarches mises en place bien qu'efficace et apportant des résultats intéressants, pourraient être améliorées ou perfectionnées.

Section 1 : Approche expérimentale

Nous avons établi un protocole d'échantillonnage de répétitions des différents substrats pris en compte dans les indices IBGN, IBG DCE et I2M2. Ce protocole a pour particularité d'échantillonner des substrats en condition de dominance uniquement, tandis que la norme XP T90-333 (AFNOR, 2009) prévoit également la considération des substrats marginaux, c'est-à-dire ceux qui n'occupent qu'une surface inférieure à 5% du site d'étude. Dans le cadre du réseau de surveillance, ces substrats soulèvent un point d'intérêt car, du fait de leur faible recouvrement, d'une année sur l'autre, ils peuvent disparaître ou selon l'opérateur ne pas être pris en considération dans les habitats marginaux et donc être absents du plan d'échantillonnage. Ainsi, nous pouvons supposer que ces habitats apporteraient une nouvelle source de variabilité et aussi une faune associée diversifiée et des taxons supplémentaires. De ce fait, ils pourraient présenter

une variabilité différente de celle observée sur les substrats dominants. Du fait des surfaces réduites qu'ils occupent, un échantillonnage répété est difficilement envisageable et réalisable. De plus, trouver des placettes de prélèvements similaires (hauteur d'eau, vitesse du courant, mosaïque d'habitat) sur un même site est très complexe. Le choix d'échantillonner les substrats dominants uniquement semblaient donc être la meilleure alternative pour étudier la variabilité intra-substrat et garantir la fiabilité des résultats. Les résultats obtenus sont difficilement transférables aux mêmes substrats en situation marginale.

De plus, il est nécessaire de rappeler que tous les substrats n'ont malheureusement pas pu être échantillonnés, en particulier les hélophytes, les algues et les vases. Leur variabilité n'a donc pas pu être étudiée et nous n'avons pu que formuler des hypothèses quant à leur comportement. Cependant, il est important de signaler que ces substrats, notamment les hélophytes, peuvent apporter des taxons supplémentaires comme par exemple des hétéroptères. Ainsi, leur prise en considération dans le plan d'échantillonnage pourrait ajouter une variabilité se propageant dans les indices.

Par ailleurs, la méthode d'échantillonnage ne nous a malheureusement pas permis de calculer de vrais indices mais seulement leurs métriques. Il serait intéressant de réaliser un échantillonnage complet. Cette piste a été écartée car le nombre d'échantillons, les temps d'échantillonnage sur le terrain et d'analyse en laboratoire auraient considérablement été augmentés, rendant difficilement envisageable cette stratégie de recherche dans le temps imparti de la thèse. Nous avons ainsi échantillonné et analysé 305 échantillons unitaires. Si nous avons décidé de réaliser un échantillonnage complet par site, soit un total de 324 échantillons supplémentaires, cela aurait représenté le double de temps d'analyse, sans compter le temps de retranscription des listes et leur traitement statistique. Dans le temps imparti, cela semblait être difficilement faisable, mais cela reste une perspective d'amélioration de cette démarche expérimentale qui rendrait alors possible la permutation d'échantillons, méthodes déjà utilisées par d'autres auteurs (Beisel, et al., 1998a; Cao, et al., 1998; Archambault, et al., 2014), l'obtention des notes des indices et ainsi la possibilité de suivre la propagation de la variabilité intra-substrat au sein de l'évaluation des milieux aquatiques.

Un dernier point important doit être souligné. L'expérimentation que nous avons menée, ne concerne que des substrats associés à des stations de très bonne qualité écologique. Pour aller plus loin, nous avons imaginé un protocole exploratoire plus large et mis en place un échantillonnage de deux substrats différents (un minéral et un organique) sur différentes classes

de qualité. Ceci nous semble d'autant plus intéressant que nous avons constaté que les effets des variabilités étudiées étaient plus importants pour les classes intermédiaires. Cependant, cette partie de l'expérimentation n'a pas abouti, du fait de l'absence des substrats en quantité suffisante sur site au moment des campagnes de terrain. Néanmoins, cette piste n'est pas à écarter et reste à mettre en œuvre.

La démarche expérimentale proposée par AGLAE dans le cadre des essais inter-laboratoires présente également certaines limites identifiables. En effet, tout comme notre approche expérimentale, les substrats échantillonnés sont en situation de dominance. Ce choix a été fait afin de garantir la reproductibilité de l'échantillonnage et aussi le nombre suffisant d'échantillons envoyés aux participants de l'essai. Du fait du nombre important d'échantillons à récolter, la combinaison de substrats offerte est limitée à trois substrats, ce qui est faible pour être représentatif d'un IBGN ou d'un IBG DCE. Néanmoins, cette solution semble la plus efficace, car réaliser un échantillonnage complet semble quasiment impossible à mettre en pratique tant en nombre de placettes d'échantillonnage pour des habitats dominants et marginaux que dans des mosaïques d'habitats similaires. De plus appliquer une telle méthode est très chronophage sur le terrain puisque depuis 2014, le nombre de laboratoires participant aux essais ayant augmenté (38 participants), il est nécessaire d'échantillonner 152 fois un même substrat.

Section 2 : Approches virtuelles

§1. Approche virtuelle d'étude de la variabilité intra-substrat et opérateur au sein des protocoles d'évaluation basés sur les macroinvertébrés

L'approche que nous avons développée est innovante puisqu'elle repose uniquement sur des altérations virtuelles générant des erreurs contrôlées. Ces erreurs dépendent des seuils que nous avons fixés dans la matrice utilisée par l'algorithme. Notre étude ne s'intéresse qu'à l'exclusion des taxons rares ou peu abondants et à la sous-estimation des abondances. Le choix des seuils peut être remis en question et notamment le seuil désignant les taxons dits abondants égal à 40

individus.

En effet, la norme XP T90-388 (AFNOR, 2010) stipule que « si les individus sont reconnus au niveau A (famille) de manière certaine mais ne sont pas reconnus au niveau B (genre), il est demandé une extraction représentative de 20 (cas des familles à diversité bioécologique faible) ou 40 (cas des familles à diversité bioécologique forte) individus minimum pour les taxons du niveau A ». Dans tous les cas, il est demandé un comptage exhaustif jusqu'à 20 individus, puis des estimations d'abondance peuvent être réalisées. Nous avons décidé de rester sur la valeur de 40 individus minimum afin de ne pas réduire les abondances de façon trop drastique et de tester uniquement des erreurs de sous-estimations. Des erreurs de surestimations des taxons seraient également envisageables.

Le seuil définissant un taxon rare ou peu abondant a été fixé à 5 individus. Il peut également être discuté. En effet, dans la plupart des cas, nous constatons, par expérience en laboratoire, que les oublis de taxons en laboratoire concernent des taxons représentés par un seul individu. Nous avons cependant décidé de choisir ce seuil de 5 individus d'une part, du fait des valeurs établies dans les références bibliographiques (Nijboer & Schmidt-Kloiber, 2004) et d'autre part pour permettre une meilleure prise en considération des variabilités cumulées liées au substrat et à l'opérateur.

De plus, notre méthode s'appuyant sur des matrices de valeurs seuils, présente l'avantage de pouvoir évoluer : même si nous n'avons testé que quelques seuils donnés, il est facilement imaginable d'en tester d'autres et ainsi de différencier les tests.

La démarche mise en œuvre présente une première approche pour estimer la sensibilité des métriques des indices invertébrés et présente certaines limites liées aux différents choix des tests menés. En effet, nous n'avons été capables d'observer que le biais engendré par l'expérience de l'opérateur. Il serait envisageable pour tester les effets cumulés des variabilités naturelles intra-substrat et opérateur de développer d'autres tests. Par exemple, l'algorithme pourrait générer non seulement des disparitions de taxons rares mais également des apparitions de taxons. Nous pourrions, ainsi, observer les réponses des métriques et des indices face à ces nouvelles variations et les possibles compensations ou aggravations liées à ces deux variabilités. Nous avons également fait le choix de tester uniquement des sous-estimations des abondances des taxons. D'autres tests peuvent être imaginés comme des surestimations d'abondances ou des sous-estimations moins radicales. Ceci nous permettrait de confirmer la sensibilité des métriques (indice de Shannon et fréquence des individus ovovivipares) aux erreurs d'estimations des abondances.

D'autre part, il avait été envisagé un temps de créer une matrice de confusion d'identification des taxons, comme celle construite pour les macrophytes, afin de générer des erreurs aléatoires de détermination. Ces erreurs d'identification sont une des principales causes de variabilité des indices, comme de nombreuses études ont pu le montrer à l'aide d'expérimentations réelles (Metzeling, et al., 2003). Pourtant, nous avons dû nous résoudre à écarter cette piste, face au nombre de taxons pris en compte dans les indices français basés sur les macroinvertébrés : 465, contre 218 taxons pour les macrophytes. La réalisation d'une telle matrice de confusion n'était donc pas possible dans le cadre de cette thèse. Cependant, cette piste de recherche pourrait être abordée dans des travaux futurs.

§2. Approche virtuelle d'étude de l'effet opérateur au sein de l'indice IBMR

Tout comme pour les macroinvertébrés, la démarche innovante développée pour étudier l'effet opérateur sur l'IBMR produit des erreurs aléatoires en fonction de deux matrices de confusion. Bien que la démarche soit virtuelle, ces matrices s'appuient sur des valeurs de confusion entre taxons ou classes de recouvrement basées sur une expérience de terrain et de laboratoire. L'identification de taxons et leur recouvrement sont les deux facteurs primordiaux pris en compte dans le calcul de l'indice IBMR. La matrice de confusion des identifications des taxons représente un travail important de synthèse et de formalisation et prend en compte toutes les confusions possibles et leur degré d'apparition. Elle est, de ce fait, difficilement améliorable, mais reste toutefois ajustable en fonction de l'évolution de l'indice IBMR. La matrice de confusion des classes de recouvrement est également un outil facilement modifiable.

L'algorithme créé s'appuie sur les nerfs centraux de l'IBMR : les erreurs d'identification et les erreurs de classe de recouvrement qui jouent sur les métriques de l'IBMR. La seule autre possibilité aurait été de tester les modifications aléatoires, non pas des classes de recouvrement, mais des valeurs en pourcentages réelles des surfaces de recouvrement. Nous n'avons pas opté pour ce choix car il n'est pas obligatoire de noter les pourcentages de recouvrement lors du relevé floristique, et donc de nombreux laboratoires traduisent directement la surface de recouvrement en classe. Cependant, en perspectives de futurs travaux il pourrait être intéressant de modifier aléatoirement les valeurs réelles de pourcentage de recouvrement à l'aide à nouveau d'une

matrice de confusion basée sur les compétences et du retour terrain d'un « macrophytiste » expérimenté.

Finalement, compte tenu du mode de construction et bien que notre démarche s'appuie sur des tests virtuels, nous pouvons considérer qu'ils reflètent la réalité.

Titre 3 : Extension des outils proposés aux autres indices biologiques

Les démarches virtuelles que nous avons développées pour les macrophytes et les macroinvertébrés, afin d'estimer l'effet de l'opérateur sur les indices associés, pourraient éventuellement être transformées et adaptées pour être appliquées aux trois autres indices français non étudiés (Indice Biologique Diatomées IBD, Indice Poissons Rivière IPR et Indice Oligochète de Bioindication des sédiments IOBS) dans le cadre de cette thèse. Les indices biologiques tels que ceux basés sur les poissons et les diatomées et leurs incertitudes ont déjà été étudiés par de nombreux auteurs (Kahlert, et al., 2012; Kelly, et al., 2009; Prygiel, et al., 2002; Pritt & Frimpong, 2014).

Les notes indicielles françaises associées aux diatomées, poissons et oligochètes sont, tout comme pour les indices macroinvertébrés et macrophytes, calculées grâce à différents modes de calcul plus ou moins complexes. À nouveau, nous pouvons identifier les facteurs qui pourraient affecter les notes et ainsi essayer de proposer une adaptation de nos méthodes à ces indices, dans le but de mesurer les effets des différentes sources de variabilité. Rappelons que les méthodes virtuelles, que nous avons mises en place, nous ont permis de tester principalement l'effet opérateur:

- (1) par la confusion des identifications des taxons et des modifications de leur abondance (changements de classes pour les macrophytes et sous-estimations des abondances pour les macroinvertébrés),
- (2) et par les oublis de taxons rares ou peu abondants dans les deux groupes étudiés.

L'objectif serait d'adapter ces différents types d'erreurs aux indices diatomées, oligochètes et

poissons.

Pour les indices diatomées, il a été montré, au Royaume-Uni, que la variabilité temporelle était un facteur entraînant une variation importante des EQR des métriques des indices, notamment pour les classes intermédiaires. Pour limiter le risque de déclassement, les auteurs conseillent l'augmentation du nombre d'échantillons (Kelly, et al., 2009). D'autres sources de variabilité ont été mises en évidence comme le choix du site, le type de substrat et les erreurs d'identification. La standardisation des techniques de comptage des valves des diatomées devrait considérablement réduire les incertitudes sur les métriques (Furse, et al., 2006; Besse-Lototskaya, et al., 2006; Kahlert, et al., 2012). Une étude canadienne a également testé l'effet de l'exclusion des taxons rares et le changement de niveau de détermination des diatomées (IDEC: Indice Diatomées de l'Est du Canada). Dans cette étude, les auteurs montrent que l'exclusion des taxons avec une abondance relative maximale inférieure à 2% n'affecte pas les résultats. Au total, 125 taxons sur 311 auront été exclus de la matrice originale. Cependant, l'exclusion de taxons, comme présenté dans cette étude, n'affecte pas la distinction entre les sites impactés et de référence, bien qu'il existe une perte d'information pour l'interprétation de l'état écologique (Lavoie, et al., 2009).

Pour la mise en œuvre de l'indice biologique diatomées français (IBD) (AFNOR, 2007), l'échantillonnage sur le terrain consiste à frotter un support à l'aide d'une brosse à poils durs pour récolter les diatomées présentes. Si c'est possible, l'opérateur doit échantillonner un substrat dur naturel, au minimum cinq pierres ou l'équivalent d'une surface minimale de 100 cm². En cas d'impossibilité de respecter ces recommandations le rapport d'essai doit en faire état. Le choix du support et de sa localisation dans la zone de courant dépend de la compétence de l'opérateur. De plus, il existe également une variabilité naturelle comme l'a démontré une étude de la réponse des algues benthiques aux points d'échange de l'eau hyporhéique et de l'eau de surface dans le canal principal (Wyatt, et al., 2008). Cette étude a montré que les assemblages de diatomées variaient selon la nature des échanges des eaux. Les variabilités naturelle et opérateur pourraient être étudiées grâce à des expérimentations menées sur le terrain. Au laboratoire, l'identification repose sur un nombre fixe de 400 individus minimum à identifier sur lamelles. Ainsi, il nous semblerait adéquat de tester les erreurs d'identification en créant une matrice de confusion des taxons. Cette variabilité est d'autant plus intéressante que l'indice diatomée évolue. En effet, entre 2000 et 2007, les taxons contributifs sont passés de 209 à plus de 800 espèces. Cette

évolution permanente nécessite que les diatomistes se maintiennent à jour, c'est pourquoi l'incertitude associée aux erreurs d'identification doit être maîtrisée.

L'indice oligochète (IOBS) (AFNOR, 2002) et ses incertitudes sont très peu étudiés à ce jour. En France, l'IOBS n'a pas été retenu pour suivre l'état des rivières dans le cadre de la DCE, sauf pour le cas particulier des canaux. Ailleurs, en Europe, il n'existe pas d'équivalent de cet indice. Pourtant, lui aussi, peut présenter différentes sources de variabilité intéressantes à estimer. Le protocole d'échantillonnage de terrain de l'IOBS requiert la réalisation de 3 prélèvements de sédiments fins représentatifs correspondant à une surface de 100 cm². La notion de sédiments fins représentatifs du site d'étude reste floue et la subjectivité de l'opérateur entre en jeu. Ce manque de précision quant au substrat à prélever peut amener une source de variabilité et peut se propager dans le calcul d'indice. Cette source de variabilité exprime un biais opérateur qui peut être difficilement appréciable avec un algorithme, une expérimentation réalisée sur le terrain semblerait plus adéquate. L'échantillon au laboratoire est traité et seuls les 100 premiers individus sortis sont identifiés. Sur ces 100 individus, il est courant qu'une portion soit non identifiable car les individus sont immatures ou abîmés. En effet, la norme NF T90-390 (AFNOR, 2002) donne la liste des taxons contributifs à l'indice IOBS et distingue les espèces reconnaissables sur des caractères sexuels de ceux reconnaissables sur des critères morphologiques quand ils sont à l'état immature. Environ 41% des taxons pris en compte ne sont pas reconnaissables à l'état immature, la question posée alors, est comment cette part d'individus non reconnaissables peut-elle affecter l'indice ? Cette problématique pourrait être étudiée, à partir de listes faunistiques d'origine, en affectant aléatoirement le caractère mature ou immature aux oligochètes et en respectant, par exemple, la proportion de 41%. Il nous semblerait également intéressant de tester les erreurs d'identification qui peuvent intervenir en créant une matrice de confusion des identifications de taxons.

En revanche, il est à signaler que, pour l'IOBS et l'IBD, il n'y a pas d'intérêt à étudier l'impact de l'exclusion des taxons rares. En effet, comme nous l'avons dit précédemment, les protocoles s'appuient sur des comptages fixes des individus donc il ne peut y avoir d'oubli de taxons rares. De plus, l'IBD supprime automatiquement les espèces représentées par une abondance inférieure à 7,5 %, soit 3 diatomées pour un total de 400 diatomées (Agences de l'eau, IRSTEA et Ministère de l'Aménagement du Territoire et de l'Environnement, s.d.). Pour les deux indices, l'oubli de taxons, s'il a lieu, pourrait être dû à l'effort du sous-échantillonnage pour obtenir la fraction

d'échantillon à analyser. Cet effet peut être étudié par voie expérimentale comme l'ont fait d'autres auteurs pour les indices macroinvertébrés (King & Richardson, 2002; Lorenz, et al., 2004).

Pour les indices basés sur les poissons les auteurs se sont plus particulièrement intéressés à l'impact de la prise en compte ou non des espèces rares dans les analyses multivariées, mais ne l'ont pas traité comme une source de variabilité (Pritt & Frimpong, 2014; Poos & Jackson, 2012). L'indice poisson est construit sur une structure plus complexe qui prend en compte différents paramètres (AFNOR, 2011) : il s'appuie d'une part sur des données récoltées sur le terrain (identification des poissons, densité et certains paramètres environnementaux) et d'autre part sur des données obtenues à partir des cartographies des sites (pente, altitude, surface du bassin versant, etc.). Toutes ces données sont obtenues par les opérateurs, de ce fait plusieurs facteurs de variabilité peuvent être étudiés. Trois types d'erreurs semblent possibles : les oublis d'espèces peu abondantes, les erreurs d'identification et le mauvais renseignement des différents paramètres environnementaux. Les matrices de confusion pourraient être adaptées pour correspondre à ces types d'erreurs. Il pourrait être également envisagé de générer des dégradations des abondances des espèces, afin de simuler des poissons échappant à l'échantillonnage par pêche électrique. En effet, l'efficacité de la pêche électrique peut varier en fonction de la position du poisson par rapport à l'anode, mais également en fonction de sa taille ainsi qu'aux conditions environnementales du site (variabilité d'accessibilité des zones refuges) (Arrignon, 1970). Dans ce sens, Aquabio réalise, à titre d'expérimentation, plusieurs passages sur une même station afin d'estimer la proportion de poissons oubliés (Etude interne, Aquabio, 2014).

Titre 4 : Combinaison des incertitudes associées aux indices

L'étude que nous avons menée nous a permis d'estimer une part de la variation des indices associée à des facteurs naturels mais également humains. Une des perspectives les plus intéressantes serait d'étudier la combinaison de toutes les incertitudes associées aux différents indices et leur propagation jusqu'à l'évaluation globale des milieux aquatiques. Jusqu'ici, nous

n'avons été en mesure que d'apprécier le comportement de certaines incertitudes pour deux compartiments biologiques. Adapter nos outils aux autres indices pour estimer les effets de différentes sources de variabilités peut être une première étape de l'analyse de l'effet global de l'incertitude dans un système complexe. De plus, la DCE préconise, pour évaluer l'état global des eaux, de ne retenir que la plus mauvaise classe obtenue (Ministère de l'Ecologie, du Développement durable et de l'Energie, 2012). Ainsi, il pourrait être intéressant de connaître l'impact des différentes variabilités sur cette évaluation, s'il existe une compensation ou au contraire un cumul des incertitudes au niveau des indices et donc une sous-évaluation ou une surévaluation de la qualité des milieux. En effet, nous avons observé que les deux indices étudiés ne répondaient pas de manière similaire aux variabilités testées. Selon nos résultats et les sources de variabilité choisies, l'indice IBMR peut être surclassé ou déclassé, tout comme l'I2M2. Mais l'IBG DCE ne peut qu'être déclassé. Comment ces écarts se répercutent-ils sur une évaluation globale ? Cette question reste en suspens et nécessiterait de réaliser des études d'incertitudes sur les autres indices retenus par la DCE (IBD et IPR) pour l'évaluation de l'état des eaux courantes en France afin d'avoir un regard complet sur les effets des variations des indices.

Le système d'harmonisation « EASY » (Ecological Ambiance System) proposé par Lafont (Lafont, 2001) pourrait être un outil utile pour comparer les effets des incertitudes sur la combinaison des indices. Il part du postulat que les biocénoses d'un écosystème sont liées aux apports de matières organiques et minérales et aux processus physico-chimiques et biologiques. A partir de ce concept, le système EASY intègre quatre compartiments biologiques associés aux bio-indicateurs. Ainsi, l'IBGN est relié à la qualité biologique globale, la qualité des sédiments à l'IOBS, la qualité biologique des eaux à l'IBD, la qualité biologique poisson à l'IPR et le niveau trophique à l'IBMR (Lafont, 2001; Perrin, et al., 2004) ; d'autres bio-indicateurs peuvent éventuellement être pris en considération. Ce système se distingue du système utilisé actuellement qui vise à définir la qualité d'une masse d'eau par le plus mauvais indicateur observé (Ministère de l'Ecologie, du Développement durable et de l'Energie, 2012), puisqu'il propose, pour un site, le calcul d'un degré d'altération qui correspond à la moyenne des classes de qualité biologique obtenues pour tous les indices. Le tableau ci-dessous présente le mode d'obtention du degré d'altération pour une rivière alsacienne.

Tableau 11 : Mode de calcul du degré d'altération dans le système d'harmonisation EASY pour une rivière alsacienne

Indices	Très bon (S1)	Bon (S2)	Moyen (S3)	Médiocre (S4)	Mauvais (S5)
IBGN	X				
IOBS			X		
IBMR			X		
IBD			X		
IPR	X				

Le degré d'altération, dans cet exemple, est égal 2,2 $((1+3+3+3+1)/5)$ et nous indique qu'il y a un début de dégradation des sédiments (IOBS = S3) et un début de dégradation physico-chimique liée à un niveau trophique moyen (IBD et IBMR = S3) mais hors matières organiques (IBGN et IPR = S1). Ici, il y a bon potentiel de résilience générale et une possibilité de résilience sur les sédiments.

Grâce à ce degré d'altération, nous pourrions envisager d'observer dans quelle mesure les variations peuvent se compenser ou non, mais également si les incertitudes étudiées ont un réel impact sur l'interprétation de l'état global du milieu, i.e. l'état des habitats, le niveau trophique, l'état des sédiments, état de l'eau, et aussi sur la capacité de résilience de ces différents compartiments.

Nous avons pu voir dans la première partie de cette thèse que des programmes européens étaient menés dans le but d'harmoniser les systèmes d'évaluation des états écologiques et de mieux cerner leurs limites. La DCE demande d'estimer les intervalles de confiance des outils d'évaluation utilisés. Cette préconisation ambitieuse, comme ont pu le montrer les travaux réalisés dans cette thèse et ceux réalisés par d'autres, pousse les hydrobiologistes à développer de nouveaux outils. Ainsi, le projet WISER propose l'outil *Wiserbugs* (Clarke, 2011; Clarke, 2013). *Wiserbugs* est un outil qui suite à l'insertion des valeurs de variances des métriques connues par les utilisateurs, permet aux gestionnaires d'étudier la propagation de ces variations au sein des indices. La France en est encore au stade d'estimer les variances de chaque métrique des indices, *Wiserbugs*. Pour cela, il s'appuie sur la connaissance de la variance et de l'écart-type des métriques en relation avec différentes sources de variabilités (échantillon, saison, spatiale et inter-annuelle).

Chaque source de variabilité représente une part de la variance totale qui varie selon les métriques (Clarke, 2013). L'outil, après que l'utilisateur ait renseigné les métriques et leurs variances, donne un intervalle de confiance autour de chaque valeur de métrique et propose une gamme d'incertitude qui permet aux gestionnaires de connaître la validité et la robustesse de l'évaluation de la classe de qualité des masses d'eaux. Nos travaux ont tenté d'estimer l'incertitude au niveau de la métrique et se situent dans en amont des travaux du projet WISER.

L'outil *Wiserbugs* est une méthode d'évaluation de la variation globale des métriques qui, associée aux algorithmes développés dans cette thèse, pourrait permettre de quantifier l'incertitude globale dans le calcul des indices français.

CONCLUSION

À chaque résultat scientifique peut être associée une part plus ou moins importante d'incertitude. Que cette incertitude soit quantifiable ou non, elle soulève de nombreux problèmes et les chercheurs sont toujours en quête d'amélioration des mesures par la minimisation de cette inconnue.

En hydrobiologie, l'incertitude, inhérente à l'évaluation de systèmes complexes, est induite par différentes sources de variabilité qui la rendent difficilement quantifiable. Deux aspects majeurs se retrouvent au cœur de toutes les études d'incertitudes, i.e. la variabilité naturelle et la variabilité humaine. Nous nous sommes, comme d'autres avant nous, intéressés à ces deux volets en les réduisant à des facteurs bien définis. Nous avons abordé la variabilité naturelle affectant les indices invertébrés en étudiant la variabilité intra-substrat à l'aide d'une démarche expérimentale de terrain. Celle-ci a mis en évidence son existence et l'association à la capacité biogène et à la complexité de la structure de chaque substrat. Cette étude complète ainsi les observations portant sur la variabilité associée à la mosaïque d'habitats et à l'hétérogénéité des substrats (Beisel, et al., 1998a; Boyero, 2003; Barnes, et al., 2013).

Nous avons également pu, grâce aux données fournies par l'Association Générale des Laboratoires d'Analyse de l'Environnement, montrer l'effet cumulé, substrat et opérateur, sur les métriques des deux indices invertébrés français actuels, l'Indice Biologique Global DCE et de l'I2M2. Les pratiques utilisées en laboratoire, l'expérience de l'opérateur et la variabilité naturelle sont autant de facteurs qui peuvent affecter les indices et entraîner une mauvaise évaluation de l'état des rivières. Les variabilités intra-substrat et opérateurs peuvent soit se cumuler dans certains cas, en

fonction des substrats et des métriques (l'ASPT, la richesse (TAX), l'indice de Shannon (H) et la fréquence des individus ovovivipares (OS) pour les pierres) soit se compenser partiellement dans d'autres cas (la fréquence des individus polyvoltins (PS) pour les pierres).

Enfin, nous avons testé l'impact de l'expérience de l'opérateur dans les mesures des indices basés sur les macroinvertébrés et macrophytes grâce à deux démarches virtuelles innovantes. Chaque processus d'évaluation est effectué par des opérateurs plus ou moins expérimentés, qui affectent la qualité des résultats, les erreurs pouvant être induites par le niveau d'une compétence de l'opérateur. Les indices répondent différemment en fonction du type d'erreurs testé. En effet, l'IBMR est très sensible aux erreurs d'identification des taxons surtout ceux ayant des cotes spécifiques (CS_i) et des coefficients de sténocécie (E_i) très éloignés. Le major confusion between Macrophyte species Indicator (JUMI), combinant ce deux facteurs, développé dans le cadre de cette thèse, fournit une connaissance des taxons susceptibles de présenter une plus grande source de variabilité et sur lesquels il est nécessaire de porter une attention particulière. L'effet des erreurs testées sur les indices basés sur les macroinvertébrés est plus complexe et dépend des métriques sur lesquelles ils sont fondés. Les métriques associées au nombre de taxons et à leur degré de polluosensibilité sont plus affectés par l'exclusion de taxons rares, alors que les métriques dépendantes des abondances des taxons (Indice de Shannon, Fréquence des individus polyvoltins et ovovivipares), prises en compte dans le nouvel indice I2M2, sont sensibles aux sous-estimations des abondances.

Chaque facteur testé mène à une variabilité des deux indices plus ou moins importante, mais qui s'est avérée être plus importante pour les classes de qualité intermédiaire (bonne, moyenne et médiocre). Pourtant, ce sont justement les classes intermédiaires qui préoccupent en priorité les gestionnaires pour la mise en place de programmes de gestion en vue de la reconquête du bon état préconisé par la réglementation européenne actuelle (Conseil Européen, 2007; Directive 2000/60/CE, 2000). La capacité de résilience d'un milieu est donc estimée, en partie, par les résultats donnés par les différents indices. Or, si les différentes sources de variabilités entraînent une mauvaise classification, cela peut avoir des répercussions, non seulement sur l'interprétation de l'évaluation globale, mais également sur les travaux mis en œuvre pour garantir un bon état écologique des masses d'eau par les gestionnaires. C'est pourquoi, il est nécessaire de proposer des mesures correctives permettant de réduire l'effet des différentes sources de variabilités mais également de mieux les connaître, afin d'avoir une vision de l'état écologique la plus fiable possible et la plus proche de la réalité.

Pour conclure, rappelons que Descartes énonçait que du doute naissait la connaissance. Le doute s'oppose à la certitude ; l'idée de la possibilité d'avancer, de se poser des questions ou au contraire l'idée de blocage et d'obstacle à la prise de décision en découlent. Pourtant, nous avons besoin de certitudes pour agir même si un doute peut toujours être présent. D'ailleurs lorsqu'un opérateur réalise son plan d'échantillonnage, sur le terrain ou ses identifications au laboratoire, la certitude l'anime et le pousse à agir. Cette prise de décision peut être accompagnée d'un contrôle sur les différentes étapes réalisées pour maximiser la fiabilité des données récoltées. Cependant, comme nous avons pu le voir tout au long de cette thèse, des incertitudes d'origine naturelle, technique ou humaine demeurent à différents paliers du processus d'évaluation des rivières. Avoir conscience de ces incertitudes, les étudier et les prendre en considération ont conduit à un mouvement positif qui s'est traduit par la proposition de normes (AFNOR, 2004; AFNOR, 2010, etc.) qui continuent d'évoluer. Ces différents travaux dont ceux présentés dans cette thèse ont permis de mieux connaître l'incertitude existant à certaines étapes du processus, et ainsi de rendre les méthodes utilisées plus robustes grâce à des solutions correctrices.

RÉFÉRENCES

-A-

- Aerts, J. & Law, A., 2009. *An introduction to scripting in Ruby for biologists*. 1e éd. s.l.:BMC bioinformatics.
- AFNOR, 2002. NF T90-390 - Détermination de l'Indice Oligochètes de Bioindication des Sédiments (IOBS).
- AFNOR, 2003. NF T90-395 Qualité de l'eau - Détermination de l'indice biologique macrophytique en rivière (IBMR).
- AFNOR, 2004. NF T90-350 Qualité de l'eau - Détermination de l'Indice Biologique Global Normalisé (IBGN).
- AFNOR, 2005. NF ISO 13528 - Méthodes statistiques utilisées dans les essais d'aptitude par comparaisons interlaboratoires. p. 64.
- AFNOR, 2006. GA T90-374 Qualité de l'eau - Guide d'application de la norme NF T90-350:2004, IBGN (Détermination de l'indice biologique global normalisé).
- AFNOR, 2007. NF T90-354 - Détermination de l'Indice Biologique Diatomées (IBD).
- AFNOR, 2009. XP T90-333 Qualité de l'eau - Prélèvement des macro-invertébrés aquatiques en rivières peu profondes.
- AFNOR, 2010. XP T90-388 Qualité de l'eau - Traitement au laboratoire d'échantillons contenant des macro-invertébrés de cours d'eau.
- AFNOR, 2011. NF T90-344 - Détermination de l'Indice Poissons Rivière (IPR).
- Agences de l'eau, IRSTEA et Ministère de l'Aménagement du Territoire et de l'Environnement, s.d. Guide méthodologique pour la mise en oeuvre de l'Indice Biologique Diatomées NF T90-354. p. 134.
- Agences de l'eau, 2000. Indice biologique global normalisé (IBGN) - NF T 90-350 - Guide technique. *Etudes des agences de l'eau*, Volume 00, p. 37.
- Agences de l'Eau, 2000. Système d'évaluation de la qualité biologique es cours d'eau - SEQ-Bio (Version 0) Rapport de présentation. *Les Etudes des Agences de l'Eau*, Volume 77, p. 27.
- Aguiar, F., Feio, M. & Ferreira, M., 2011. Choosing the best method for stream bioassessment using macrophyte communities: Indices and predictive models. *Ecological Indicators*, Volume 11, pp. 379-388.
- Alba-Tercedor, J. & Sanchez-Ortega, A., 1988. Un método rapido y simple para evaluar la calidad biologica de las aguas corrientes basado en el de Hellawell (1978). *Limnetica*, Volume 4, pp. 51-56.
- Archambault, V. & Dumont, B., 2010. L'indice biologique global normalisé (IBGN) : principes et évolution dans le cadre de la directive cadre européenne sur l'eau. *Revue SET*, Volume 01, pp. 36-39.

Archambault, V., Josset, A. & Prieto-Montes, M., 2014. *Evaluation de l'incertitude liée à l'application du protocole d'échantillonnage des invertébrés en cours d'eau selon la norme XP T 90-333 - Exemples de variabilité lors de la phase d'acquisition des données*. La Paraclat: IRSTEA.

Archambault, V., Josset, A. & Roger, M.-C., 2011. Evaluation de l'incertitude liée à l'application du protocole d'échantillonnage des macro-invertébrés benthiques en cours d'eau peu profonds selon la norme XP T 90-333. *Rapport AQUAREF 2011*, Volume Bilan d'avancement 2011, p. 28.

Armitage, P., Moss, D., Wright, J. & Furse, M., 1983. The performance of a new biological water quality score system based on macroinvertebrates over a wide range of unpolluted runningwater sites. *Water Research*, 17(3), pp. 333-347.

Arrignon, J., 1970. Aménagement piscicole des eaux intérieures. p. 643.

Arscott, D., Jackson, J. K. & Kratzer, E., 2006. Role of rarity and taxonomic resolution in a regional and spatial analysis of stream macroinvertebrates. *Journal of the North American Benthological Society*, 25(4), pp. 977-997.

ASME, 2007. B89.7.3.2 - Guidelines for the Evaluation of Dimensional Measurement Uncertainty (Technical Report). p. 28.

Association Générale des Laboratoires d'Analyse de l'Environnement, 2013. *Rapport de comparaison interlaboratoires - Hydrobiologie*, Association AGLAE.

-B-

Barbour, M., Gerritsen, J., Snyder, B. & Stribling, J., 1999. *Rapid bioassessment protocols for use in streams and wadeable rivers*. 2e éd. Washington: USEPA.

Barnes, J., Vaughan, I. & Ormerod, S., 2013. Reappraising the effects of habitat structure on river macroinvertebrates. *Freshwater Biology*, Volume 58, pp. 2154-2167.

Beckert, J., 1996. What is sociological about economic sociology? Uncertainty and the embeddedness of economic action. *Theory and society*, 25(6), pp. 803-840.

Beisel, J.-N., Usseglio-Polatera, P. & Moreteau, J.-C., 2000. The spatial heterogeneity of a river bottom: a key factor determining macroinvertebrate communities. *Hydrobiologia*, Volume 422-423, p. 163–171.

Beisel, J.-N., Usseglio-Polatera, P., Thomas, S. & Moreteau, J.-C., 1998(a). Influence de l'échantillonnage des communautés macrobenthiques sur l'évaluation de la qualité biologique d'un cours d'eau. *Annals Limnol.*, 34(4), pp. 445-454.

Beisel, J.-N., Usseglio-Polatera, P., Thomas, S. & Moreteau, J.-C., 1998(b). Stream community structure in relation to spatial variation: the influence of mesohabitat characteristics. *Hydrobiologia*, 389(1-3), p. 73–88.

Bellan, G., 1984. Indicateurs et indices biologiques dans le domaine marin. *Bulletin d'Ecologie*, 15(1), pp. 13-20.

Bernard, C., 1865. Introduction à l'étude de la médecine expérimentale. p. 480.

- Besse-Lototskaya, A., Verdonshot, P. & Sinkeldam, J., 2006. Uncertainty in diatom assessment: Sampling, identification and counting variation. *Hydrobiologia*, Volume 566, pp. 247-260.
- Birk, S. et al., 2012. Three hundred ways to assess Europe's surface waters : An almost complete overview of biological methods to implement the Water Framework Directive. *Ecological Indicators*, Volume 18, pp. 31-41.
- Birk, S. & Hering, D., 2006. Direct comparison of assessment methods using benthic macroinvertebrates: a contribution to the EU Water Framework directive intercalibration exercise. *Hydrobiologia*, Volume 566, pp. 401-415.
- Birk, S., Korte, T. & Hering, D., 2006. Intercalibration of assessment methods for macrophytes in lowland streams: direct comparison and analysis of common metrics. *Hydrobiologia*, Volume 566, pp. 417-430.
- Birk, S. & Willby, N., 2010. Towards harmonization of ecological quality classification: establishing common grounds in European macrophyte assessment for rivers. *Hydrobiologia*, Volume 652, pp. 149-163.
- Blandin, P., 1986. Bioindicateurs et diagnostic des systèmes écologiques. *Bull. Ecol.*, 17(4), pp. 215-307.
- Bloom, N., Kose, M. & Terones, M., 2013. Le poids de l'incertitude. *Finances & Développement*, pp. 38-41.
- Borcard, D., Gillet, F. & Legendre, P., 2011. *Numerical Ecology with R*. Springer Science.
- Boyero, L., 2003. The quantification of local substrate heterogeneity in streams and its significance for macroinvertebrate assemblages. *Hydrobiologia*, Volume 499, pp. 161-168.
- Breugnot, E., Dutartre, A., Laplace-Treyture, C. & Haury, J., 2004. Variabilité des peuplements des macrophytes aquatiques en grands cours d'eau: premiers résultats sur l'hydrosystème Adour-Garonne. *Ingénieries*, Volume 37, pp. 39-52.
- Breugnot, E., Dutartre, A., Laplace-Treyture, C. & Haury, J., 2008. Local distribution of macrophytes and consequences for sampling methods in large rivers. *Hydrobiologia*, Volume 610, pp. 13-23.
- Brochard, V., 2002. *Les sceptiques grecs*. Références éd. Livre de Poche.
- Buffagni, A. et al., 2005. Towards European Intercalibration for the Water Framework Directive: Procedures and examples for different river types from the E.C.. *project STAR*, Volume 11th STAR deliverable, p. 468.
- Buffagni, A. et al., 2006. The STAR common metrics approach to the WFD intercalibration process: Full application for small, lowland rivers in three European countries. *Hydrobiologia*, Volume 566, pp. 379-399.
- Buffagni, A. & Furse, M., 2006. Intercalibration and comparison - major results and conclusions from the STAR project. *Hydrobiologia*, Volume 566, pp. 357-364.
- Burton, T., Stricker, C. & Uzarski, D., 2002. Effects of plant community composition and exposure to wave action on invertebrate habitat use of Lake Huron coastal wetlands. *Lakes & Reservoirs: Research & Management*, 7(3), pp. 255-269.

Butcher, R., 1933. Studies on the ecology of rivers - I : On the distribution of macrophytic vegetation in the rivers of Britain. *J. Ecol.*, Volume 21, pp. 58-91.

-C-

Cao, Y., Larsen, D. & Thorne, R., 2001. Rare species in multivariate analysis for bioassessment: some considerations. *Journal of the North American Benthological Society*, 20(1), pp. 144-153.

Cao, Y., Williams, D. & Larsen, D., 2002. Comparison of ecological communities: the problem of sample representativeness. *Ecological Monographs*, 72(1), pp. 41-56.

Cao, Y., Williams, D. & Williams, N., 1998. How important are rare species in aquatic community ecology and bioassessment?. *Limnology and Oceanography*, 43(7), pp. 1403-1409.

Carbiener, R., Trémolières, M., Mercier, J.-L. & Ortscheit, A., 1990. Aquatic macrophyte communities as bioindicators of eutrophication in calcareous oligosaprobe stream waters (Upper Rhine plain, Alsace). *Vegetatio*, 86(1), pp. 71-88.

Carnap, R., Hahn, H., Neurath, O. & Schlick, M., 2010. *Manifeste du Cercle de Vienne et autres écrits*. Vrin.

Cheshmedjiev, S., 2013. Rapid Biological Assessment (RBA) based on analysis of benthic macroinvertebrate communities. p. 24.

Chester, R., 1980. Biological Monitoring Working Party. *The 1978 national testing exercise. Technical Memorandum*, Volume 19.

Clarke, R., 2011. *Deliverable D6. 1-3: WISERBUGS (WISER Bioassessment Uncertainty Guidance Software) tool for assessing confidence of WFD ecological status class - User Manual and Software: Release 1.2 (November 2011)*.

Clarke, R., 2012. *Deliverable D6.1-2: Manuscript reviewing components of uncertainty and their assessment, including guidelines for estimation and quality assurance*, Bournemouth University.

Clarke, R., 2013. Estimating confidence of European WFD ecological status class and WISER Bioassessment Uncertainty Guidance Software (WISERBUGS). *Hydrobiologia*, 1(704), pp. 39-56.

Clarke, R. et al., 2006. Estimates and comparisons of the effects of sampling variation using "national" macroinvertebrate sampling protocols on the precision of metrics used to assess ecological status. *Hydrobiologia*, Volume 566, pp. 477-503.

Clarke, R. et al., 2002. Sampling variation in macroinvertebrate data and implications for river quality indices. *Freshwater Biology*, Volume 47, pp. 1735-1751.

Clarke, R. & Jones, I., 2009. *Deliverable D6.1-1: Report on a workshop to bring together experts experienced with tool development and uncertainty estimation*, s.l.: Center for Ecology and Hydrology.

Clarke, R. & Murphy, J., 2006. Effects of locally rare taxa on the precision and sensitivity of RIVPACS bioassessment of freshwaters. *Freshwater Biology*, 51(10), pp. 1924-1940.

Clarke, R. T., 2013. Estimating confidence of European WFD ecological status class and WISER Bioassessment Uncertainty Guidance Software (WISERBUGS). *Hydrobiologia*, Volume 704, pp. 39-56.

Clarke, R. T. & Hering, D., 2006. Errors and uncertainty in bioassessment methods – major results and conclusions from the STAR project and their application using STARBUGS. *Hydrobiologia*, Volume 566, pp. 433-439.

Clarke, R., Wright, J. F., Sutcliffe, D. W. & Furse, M. T., 2000. Uncertainty in estimates of biological quality based on RIVPACS. In *Assessing the biological quality of fresh waters: RIVPACS and other techniques*. Proceedings of an International Workshop held in Oxford, UK, on 16-18 September 1997.. *Freshwater Biological Association*, pp. 39-54.

Code de l'Environnement, 21 avril 2004. Loi n° 2004-338. Dans: *Transposition de la directive 2000/60/CE du Parlement européen et du Conseil du 23 octobre 2000 établissant un cadre pour une politique communautaire dans le domaine de l'eau (1)*.

Commission Européenne, 1968. *Charte Européenne de l'eau*. Strasbourg.

Conseil Européen, 2007. Circulaire DCE 2007/22. *Bulletin officiel du Ministère de l'Ecologie, du Développement et de l'Aménagement Durable*.

Cox, T. & Cox, M., 2001. *Multidimensional scaling*. Second edition éd. Chapman and Hall.

-D-

De Broglie, L., 1983. *Nouvelles perspectives en microphysique*. Flammarion.

Descartes, R., 1637. *Discours de la méthode*.

Directive 2000/60/CE, 2000. Framework directive of the European Parliament and of the Council of 23 October 2000. *Official journal*.

Directive 2000/60/CE, 2008. Décision de la commission du 30 octobre 2008. *Journal officiel de l'Union Européenne*, Volume L.332/20.

Doberstein, C., Karr, J. & Conquest, L., 2000. The effect of fixed-count subsampling on macroinvertebrate biomonitoring in small streams. *Freshwater Biology*, 44(2), pp. 355-371.

-E-

Ekeland, I., 1995. *Le chaos*. Dominos éd. s.l.:Flammarion.

Environment Agency, 1999. Procedures for collecting and analysing macroinvertebrates samples.. *Environment Agency Document*, Volume BT001.

Erba, S. et al., 2009. The validation of common European class boundaries for river benthic macroinvertebrates to facilitate the intercalibration process of the Water Framework Directive. *Hydrobiologia*, Volume 633, pp. 17-31.

Erman, D. & Erman, N., 1984. The response of stream macroinvertebrates to substrate size and heterogeneity. *Hydrobiologia*, Volume 108, pp. 75-82.

-F-

Feio, M., Reynoldson, T., Ferreira, V. & Graça, M., 2007. A predictive model for freshwater bioassessment (Mondego River, Portugal). *Hydrobiologia*, 589(1), pp. 55-68.

Furse, M. et al., 2006. Macrophytes and diatoms - major results and conclusions from the STAR project. *Hydrobiologia*, Volume 566, pp. 175-178.

Furse, M. et al., 2006. The STAR project: context, objectives and approaches. *Hydrobiologia*, Volume 566, pp. 3-29.

Furse, M., Wright, J., Armitage, P. & Moss, D., 1981. An appraisal of pond-net samples for biological monitoring of lotic macro-invertebrates. *Water Research*, Volume 15, pp. 679-689.

-G-

Gabriels, W., Lock, K., De Pauw, N. & Goethals, P., 2010. Multimetric Macroinvertebrate Index Flanders (MMIF) for biological assessment of rivers and lakes in Flanders (Belgium). *Limnologica*, 40(3), pp. 199-207.

Gallant, A. et al., 1989. *Regionalization as a tool for managing environmental resources*. Corvallis, Oregon: U.S EPA Environmental Research Laboratory.

Gardeniers, J., Peeters, E. & Tolkamp, H., 1996. EBEOSWA: Ecologische BEOordelingsmethode voor Stromend Water op basis van macrofauna.

Gaston, K., 1994. *Rarity*. Chapman and Hall.

Gendron, A. & Martin, L., 2004. *Méthodes statistiques appliquées à la psychologie - Traitement des données avec Excel*. SMG.

Ghetti, P., 1997. Manuale di Applicazione: Indice Biotico Esteso – I macroinvertebrati nel controllo della qualità degli ambienti di acque correnti. *Provincia Autonoma di Trento, Servizio Protezione Ambiente*.

Gillies, C., Hose, G. & Turak, E., 2009. What do qualitative rapid assessment collections of macroinvertebrates represent? A comparison with extensive quantitative sampling. *Environmental Monitoring and Assessment*, 149(1-4), pp. 99-112.

Guerold, F., 1992. *L'acidification des cours d'eau: impact sur les peuplements de macroinvertebrés benthiques : application au Massif Vosgien*. Thèse de Doctorat de l'Université de Metz.

Guerold, F., 2000. Influence of taxonomic determination level on several community indices. *Wat. Res.*, 34(2), pp. 487-492.

-H-

- Haase, P. et al., 2006. Assessing the impact of errors in sorting and identifying macroinvertebrate samples. *Hydrobiologia*, Volume 566, pp. 505-521.
- Haase, P., Pauls, S., Sundermann, A. & Zenker, A., 2004. Testing different sorting techniques in macroinvertebrate samples from running waters. *Limnologica*, Volume 34, pp. 366-378.
- Haase, P., Pauls, S. U., Schindehütte, K. & Sundermann, A., 2010. First audit of macroinvertebrate samples from an EU Water Framework Directive monitoring program: human error greatly lowers precision of assessment results. *Journal of the North American Benthological Society*, 4(29), pp. 1279-1291.
- Harding, J., 1981. Macrophytes as monitors of river quality in the southern NWWA Area.. *NWW Authority Rivers Division*, Volume TS-BS-81-2, p. 54.
- Haslam, S., 1982. A proposed method for monitoring river pollution using macrophytes. *Envir. Technol. Letters*, Volume 3, pp. 19-34.
- Haury, J., 1994. Les associations macrophytiques vasculaires en tant que descripteurs des caractéristiques d'habitat des cours d'eau à saumons : exemple du Scorff. Dans: J. Gehu, éd. *La Syntaxonomie et le Synsystème Européennes, comme Base typologique des Habitats*. Berlin Stuttgart: Cramer Ed., pp. 31-54.
- Haury, J. & Peltre, M.-C., 1993. Intérêts et limites des "indices macrophytes" pour qualifier la mésologie et la physico-chimie des cours d'eau : exemples armoricains, picards et lorrains. *Annls Limnol.*, 29((3-4)), pp. 239-253.
- Haury, J. et al., 1996. Des indices macrophytiques pour estimer la qualité des cours d'eau français : premières propositions. *Ecologie*, 27(4), pp. 233-244.
- Haury, J. et al., 2002. *Une méthode pour mettre en évidence la trophie de l'eau et la pollution organique avec les macrophytes: l'indice biologique macrophytique en rivière (IBMR) - Application à différents types de rivières et de pollutions*. Moliets et Maa (France), 11e Symposium EWRS.
- Haury, J. et al., 2006. A new method to assess water trophy and organic pollution - the Macrophyte Biological Index for Rivers (IBMR): its application to different types of river and pollution. *Hydrobiologia*, Volume 1, pp. 153-158.
- Heisenberg, W., 1930. *The physical principles of the quantum theory*. New York.
- Hering, D., Moog, O., Sandin, L. & Verdonshot, P., 2004. Overview and application of the AQEM assessment system. *Hydrobiologia*, Volume 516, pp. 1-20.
- Holmes, N., 1983. *Typing British rivers according to their flora. Focus on nature conservancy 4*. Huntingdon, U.K.: Nature Conservancy Council.
- Holmes, N. et al., 1999. Mean Trophic Rank: A user's Manual. *Environmental Agency*.
- Hose, G., Turak, E. & Waddell, N., 2004. Reproducibility of AUSRIVAS rapid bioassessments using macroinvertebrates. *Journal of the North American Benthological Society*, 23(1), pp. 126-139.

Hughes, R.-M. & Larsen, D.-P., 1988. Ecoregions: an approach to surface water protection. *Journal of Water Pollution Control Federation*, 60(4), pp. 486-493.

-I-

Illies, J., 1978. *Limnofauna Europaea*. Stuttgart: Fischer Verlag.

ISO 7828, 1985. *Water quality - Methods of biological sampling guidance on handnet sampling of aquatic benthic macroinvertebrates*.

-J-

Jaccard, P., 1901. Etude comparative de la distribution florale dans une portion des Alpes et du Jura. *Bulletin de la société Vaudoise des sciences naturelles*, Volume 37, pp. 547-579.

JCGM 200, 2008. International vocabulary of metrology - Basic and general concepts and associated terms (VIM). p. 90.

Johnson, R., Hering, D., Furse, M. & Clarke, R., 2006. Detection of ecological change using multiple organism groups: metrics and uncertainty. *Hydrobiologia*, 566(1), pp. 115-137.

-K-

Kahlert, M. et al., 2012. Identification versus counting protocols as sources of uncertainty in diatom-based ecological status assessments. *Hydrobiologia*, 685(1), pp. 109-124.

Karr, J., 1981. Assessment of biotic integrity using fish communities. *Fisheries*, Volume 6, pp. 21-27.

Karr, J., 1986. Assessing biological integrity in running waters: a method and its rationale. *Illinois Natural History Survey*, Volume 5, pp. 1-28.

Karr, J. & Chu, E., 1997. *Biological Monitoring and Assessment: Using Multimetric Indexes Effectively*. EPA 235-R97-001, Seattle: University of Washington.

Katoh, K., Kuma, K., Toh, H. & Miyata, T., 2009. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic acids research*, 33(2), pp. 511-518.

Kelly, M. et al., 2009. Uncertainty in ecological status assessments of lakes and rivers using diatoms. *Hydrobiologia*, Volume 633, pp. 5-15.

Kilroy, C. et al., 2013. Estimating periphyton standing crop in streams: a comparison of chlorophyll a sampling and visual assessments. *New Zealand Journal of Marine and Freshwater Research*, 47(2), pp. 208-224.

- King, R. & Richardson, C., 2002. Evaluating subsampling approaches and macroinvertebrate taxonomic resolution for wetland bioassessment. *Journal of the North American Benthological Society*, 21(1), pp. 150-171.
- Kohler, A., 1971. Zur Ökologie submerser Gefäß-Makrophyten in Fließgewässern. *Ber.Dtsch.Bot.Ges*, Volume 84, pp. 713-720.
- Kohler, A., Vollrath, H. & Beisl, E., 1971. Zur Verbreitung, Vergesellschaftung und Ökologie der Gefäß-Makrophyten im Fließgewässersystem Moosach (Münchener Ebene). *Arch. Hydrobiol.*, Volume 69, pp. 333-365.
- Kokeš, J. et al., 2006. The PERLA system in the Czech Republic: a multivariate approach for assessing the ecological status of running waters. *Hydrobiologia*, Volume 566, pp. 343-354.
- Kolada, A., Ciecierska, H., Ruszczynska, J. & Dynowski, P., 2014. Sampling techniques and inter-operator variability as sources of uncertainty in Polish macrophyte metric for lake ecological status assessment. *Hydrobiologia*, Volume 737, pp. 265-279.
- Kolkwitz, R. & Marsson, M., 1909. Ökologie der tierischen Saprobien. Beiträge zur Lehre von der biologischen Gewässerbeurteilung. *Internationale Revue des gesamten Hydrobiologie und Hydrographie*, 2(1-2), pp. 126-152.

-L-

- Lafont, M., 2001. A conceptual approach to the biomonitoring of freshwater : the ecological ambience system. *Journal of Limnology*, 60(1), pp. 17-24.
- Langevin, P., 1950. *La pensée et l'action. Textes recueillis et présentés par Paul Laberrenne*. Français Réunis.
- Latvian Standard LVS 240, 1999. Water quality - Operative Evaluation of Small Streams Biological Quality by Sprobit Index of Macroinvertebrates. *VSIA Latvijas Standarts*.
- Lavoie, I., Dillon, P. & Campeau, S., 2009. The effect of excluding diatom taxa and reducing taxonomic resolution on multivariate analyses and stream bioassessment. *Ecological Indicators*, 9(1), pp. 213-225.
- Legendre, P. & Legendre, L., 1998. *Numerical Ecology*. Elsevier.
- Ligeiro, R., Ferreira, W., Hughes, R. & Callisto, M., 2013. The problem of using fixed-area subsampling methods to estimate macroinvertebrate richness: a case study with Neotropical stream data. *Environmental monitoring and assessment*, 185(1), pp. 4077-4085.
- Lorenz, A. & Clarke, R., 2006. Sample coherence a field study approach to assess similarity of macroinvertebrates samples. *Hydrobiologia*, Volume 188, pp. 461-476.
- Lorenz, A., Kirchner, L. & Hering, D., 2004. "Electronic subsampling" of macrobenthic samples: how many individuals are needed for a valid assessment result?. *Hydrobiologia*, Volume 516, pp. 299-312.

-M-

- Marchant, R., 2002. Do rare species have any place in multivariate analysis for bioassessment?. *Journal of the North American Benthological Society*, 21(2), pp. 311-313.
- Mascaro, O. et al., 2013. Exploring the robustness of macrophyte-based classification methods to assess the ecological status of coastal and transitional ecosystems under the Water Framework Directive. *Hydrobiologia*, Volume 704, pp. 279-291.
- Matte, J.-L., 2005. *Evaluation de l'influence du choix des placettes de prélèvement sur l'indice IBGN*. Direction Régionale de l'Environnement Lorraine.
- Mazuer, P., Kieffer, S., Matte, J.-L. & Heudre, D., 2007. *Comparaison de deux pratiques d'échantillonnage des macroinvertébrés aquatiques en rivière (agitation et récolte du support)*. Direction Régionale de l'Environnement Lorraine.
- McAbendroth, L. et al., 2005. Does macrophyte fractal complexity drive invertebrate diversity, biomass and body size distributions?. *Oikos*, 111(2), pp. 279-290.
- Meriaux, J., 1978. Etude analytique et comparative de la végétation aquatique d'étangs et de marais du Nord de la France. *Doc. Phytosociol.*, Volume 3, pp. 1-244.
- Metzeling, L., Chessman, B., Hardwick, R. & Wong, V., 2003. Rapid assessment of rivers using macroinvertebrates: the role of experience, and comparisons with quantitative methods. *Hydrobiologia*, Volume 510, pp. 39-52.
- Milet-Mureau, M., 1797. *Voyage de la Perouse autour du monde*. Imprimerie de la République éd. s.l.
- Ministère de l'Ecologie, du Développement durable et de l'Energie, 2012. Guide relatif à l'évaluation de l'état des eaux de surface continentales (cours d'eau, canaux, plans d'eau). p. 84.
- Mondy, C., 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*, Volume 14, pp. 452-467.
- Moog, O., Chovanec, A., Hinterregger, J. & Römer, A., 1999. *Richtlinie zur Bestimmung des saprobiologischen Gewässergüte von Fließgewässern*. Wien: Bundesministerium für Land und Forstwirtschaft.
- Moog, O., Schmidt-Kloiber, A., Ofenböck, T. & Gerritsen, J., 2004. Does the ecoregion approach support the typological demands of the EU "Water Framework Directive"?. *Hydrobiologia*, Volume 516, pp. 21-33.
- Mormul, R., Thomaz, S., Takeda, A. & Behrend, R., 2011. Structural complexity and distance from source habitat determine invertebrate abundance and diversity. *Biotropica*, 43(6), pp. 738-745.
- Moss, D., Wright, J., Fure, M. & Clarke, R., 1999. A comparison of alternative techniques for prediction of the fauna of running water sites in Great Britain. *Freshwater Biology*, Volume 41, pp. 167-181.
- Moureau, N. & Rivaud-Danset, D., 2004. *L'incertitude dans les théories économiques*. Repères éd. La découverte.

Murphy, P., 1978. The temporal variability in biotic indices. *Environmental Pollution*, 17(3), pp. 227-236.

-N-

Nichols, S. & Norris, R., 2006. River condition assessment may depend on the sub-sampling method: field live-sort versus laboratory sub-sampling of invertebrates for bioassessment. *Hydrobiologia*, Volume 572, pp. 195-213.

Nijboer, R. & Schmidt-Kloiber, A., 2004. The effect of excluding taxa with low abundances or taxa with small distribution ranges on ecological assessment. *Hydrobiologia*, Volume 516, pp. 347-363.

-O-

Oberdorfer, E., 1990. *Pflanzensoziologische Exkursionsflora*. Ulmer, Stuttgart: 6.

Oberdorff, T., Pont, D., Hugueny, B. & Chessel, D., 2001. A probabilistic model characterizing fish assemblages of French rivers: a framework for environmental assessment. *Freshwater Biology*, 46(3), pp. 399-415.

Oberdorff, T., Pont, D., Hugueny, B. & Porcher, J.-P., 2002. Development and validation of a fish-based index for the assessment of 'river health' in France. *Freshwater Biology*, 47(9), pp. 1720-1734.

Ostermiller, J. & Hawkins, C., 2004. Effects of sampling error on bioassessments of stream ecosystems: application to RIVPACS-type models. *Journal of the North American Benthological Society*, 23(2), pp. 363-382.

-P-

Palmer, M. et al., 2000. Streambed landscapes: evidence that stream invertebrates respond to the type and spatial arrangement of patches. *Landscape Ecology*, Volume 15, pp. 263-576.

Perrin, J.-F., Lafont, M. & Pery, A., 2004. Outils de diagnose fonctionnelle pour l'ingénierie écologique en rivière. *Ingénieries N° Spécial*, pp. 57-67.

Petkovska, V. & Urbanic, G., 2010. Effect of fixed-fraction subsampling on macroinvertebrate bioassessment of rivers. *Environ Monit Assess*, Volume 169, pp. 179-201.

Planck, M., 1989. *Initiations à la physique*. 2ème éd. Flammarion.

Pond, G. et al., 2008. Downstream effects of mountaintop coal mining: comparing biological conditions using family - and genus - level macroinvertebrate bioassessment tools. *Journal of the North American Benthological Society*, 27(3), pp. 717-737.

Poos, M. & Jackson, D., 2012. Addressing the removal of rare species in multivariate bioassessments: the impact of methodological choices. *Ecological Indicators*, Volume 18, pp. 82-90.

Poquet, J. et al., 2009. The MEDiterranean Prediction And Classification System (MEDPACS): an implementation of the RIVPACS/AUSRIVAS predictive approach for assessing Mediterranean aquatic macroinvertebrate communities. *Hydrobiologia*, 623(1), pp. 153-171.

Pritt, J. & Frimpong, E., 2014. The effect of sampling intensity on patterns of rarity and community assessment metrics in stream fish samples. *Ecological Indicators*, Volume 39, pp. 169-178.

Prygiel, J. et al., 2002. Determination of the biological diatom index (IBD NF T 90–354): results of an intercomparison exercise. *Journal of Applied Phycology*, Volume 14, pp. 27-39.

-R-

Ramade, F., 2005. *Eléments d'écologie - Ecologie appliquée*. Sciences Sup éd. Dunod.

Renkonen, O., 1938. Statistisch-Skologische Untersuchungen tiber die terrestrische Kfiferwelt des finnischen Bruchmoore. *Zool. Soc. Zool. Bot. Fenn Vanamo*, Volume 6, pp. 1-231.

Reyjol, Y., Spyrtos, V. & Basilico, L., 2012. *Bioindication : des outils pour évaluer l'état écologique des milieux aquatiques - Perspectives en vue du 2ème cycle DCE - Eaux de surface continentales*. Paris: Les rencontres de l'ONEMA.

Rison, J. et al., 2007. Predicting diatom reference communities at the French hydrosystem scale: A first step towards the definition of the good ecological status. *Ecological Modelling*, 203(1-2), pp. 99-108.

Robach, F., Thiébaud, G., Trémolières, M. & Muller, S., 1996. A reference system for continental running waters: plant communities as bioindicators of increasing eutrophication in alkaline and acidic waters in north-east France. *Hydrobiologia*, Volume 340, pp. 67-76.

-S-

Saldecek, V., 1973. System of water quality from the biological point of view. *Arch. Hydrobiol. Beih. Ergebn. Limnol.*, 7(4), p. 218.

Schmidt-Kloiber, A. & Nijboer, R., 2004. The effect of taxonomic resolution on the assessment of ecological water quality classes. *Hydrobiologia*, Volume 516, pp. 269-283.

Schneider, S. & Melzer, A., 2003. The Trophic Index of Macrophytes (TIM) - a New Tool for Indicating the trophic State of Running Waters. *International Review of Hydrobiology*, 88(1), pp. 49-67.

Shannon, C., 1948. A mathematical theory of communication. *The Bell System Technical Journal*, 17(1), pp. 379-423 & 623-656.

Simpson, J. & Norris, R., 2000. Biological assessment of water quality: development of AUSRIVAS models and outputs. *Freshwater Biological Association*, pp. 125-142.

- Siqueira, T. et al., 2012. Common and rare species respond to similar niche processes in macroinvertebrate metacommunities. *Ecography*, 35(2), pp. 183-192.
- Skriver, J., Friberg, N. & Kirkegaard, J., 2000. Biological assessment of running waters in Denmark: introduction of the Danish Stream Fauna Index (DSFI). *Verh. Internat. Verein. Limnol*, Volume 27, pp. 1822-1830.
- Sörensen, T., 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content. *Biol. Kriffter. Bd*, Volume 4, pp. 1-34.
- Springe, G., Sandin, L., Briede, A. & Skuja, A., 2006. Biological quality metrics: their variability and appropriate scale for assessing stream. *Hydrobiologia*, Volume 188, pp. 153-172.
- SS-EN 27828, 2000. Vattenundersökningar – Metoder för biologisk provtagning – Riktlinjer för vatten - Del 2: Vägledning vid presentation av biologiska kvalitetsdata från undersökningar av bottenfauna (ISO 8689-2).
- Staniszewski, R. et al., 2006. Assessment of sources of uncertainty in macrophyte surveys and the consequences for river classification. *Hydrobiologia*, Volume 566, pp. 235-246.
- Stribling, J., Pavlik, K., Holdsworth, S. & Leppo, E., 2008. Data quality, performance, and uncertainty in taxonomic identification for biological assessments. *Journal of the North American Benthological Society*, 27(4), pp. 906-919.
- Szoszkiewicz, K., Zbierska, J., Staniszewski, R. & Jusik, S., 2009. The variability of macrophyte metrics used in river monitoring. *International Journal of Oceanography and Hydrobiology*, 38(4), pp. 117-126.

-T-

- Tachet, H., Richoux, M., Bournaud, M. & Usseglio-Polatera, P., 2010. *Invertebrés d'eau douce: Systématique, Biologie, Écologie*. Paris, France: CNRS éditions. p. 607.
- Taniguchi, H., Nakano, S. & Tokeshi, M., 2003. Influences of habitat complexity on the diversity and abundance of epiphytic invertebrates on plants. *Freshwater Biology*, 48(4), pp. 718-728.
- Tartare, J., 2004. *Contribution à l'amélioration de la démarche qualité déjà en place sur l'Indice Biologique Global Normalisé (IBGN)*, Saint-Germain-du-Puch.
- Thiébaud, G. & Muller, S., 1999. A macrophyte communities sequence as an indicator of eutrophication and acidification levels in weakly mineralised streams in north-eastern France. *Hydrobiologia*, Volume 410, pp. 17-24.
- Thomaz, S. M. et al., 2008. Influence of aquatic macrophyte habitat complexity on invertebrate abundance and richness in tropical lagoons. *Freshwater Biology*, 53(2), pp. 358-367.
- Tison, J. et al., 2005. Typology of diatom communities and the influence of hydro-ecoregions: A study on the French hydrosystem scale. *Water Research*, 39(14), pp. 3177-3188.
- Trigal-Domínguez, C., Fernández-Aláez, C. & García-Criado, F., 2009. Habitat selection and sampling design for ecological assessment of heterogeneous ponds using macroinvertebrates.. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 19(7), pp. 786-796.

Tuffery, G. & Verneaux, J., 1968. Méthode de détermination de la qualité biologique des eaux courantes. Exploitation codifiée des inventaires de la faune du fond. *Ministère de l'Agriculture (France) Centre National d'Etudes techniques et de recherches technologiques pour l'agriculture, les forêts et l'équipement rural 'C.E.R.A.F.E.R.'*, Volume Section Pêche et Pisciculture, p. 23.

-U-

Usseglio-Polatera, P., Bournaud, M., Richoux, P. & Tachet, H., 2000. Biomonitoring through. *Hydrobiologia biological traits of benthic macroinvertebrates: how to use species trait databases?*, Volume 422-423, pp. 153-162.

Usseglio-Polatera, P., Thomas, S., Beisel, J.-N. & Moreteau, J.-C., 1999. Illustration de la valeur indicatrice des caractéristiques biologiques des macroinvertébrés d'une communauté benthique à différentes échelles d'observation. *Annls Limnol.*, 35(1), pp. 71-80.

-V-

Van Looy, K., Tormos, T. & Souchon, Y., 2014. Disentangling dam impacts in river networks. *Ecological Indicators*, Volume 37, pp. 10-20.

Verdonschot, P. & Nijboer, R., 2004. Testing the European stream typology of the Water Framework Directive for macroinvertebrates. *Hydrobiologia*, Volume 516, pp. 35-54.

Verneaux, J., 1973. *Cours d'eau de Franch-Comté (Massif du Jura). Recherches écologiques sur le réseau hydrographique du Doubs. Essai de biotypologie*, Université de Besançon: Thèse de Doctorat d'Etat.

Verneaux, J. & coll., 1982. Une nouvelle méthode pratique d'évaluation de la qualité des eaux courantes. Un indice biologique de qualité générale (IBG). *Ann. Sc. Univ. Franche-Comté Besançon , Biol. anim.*, 4(3), pp. 11-21.

Verneaux, J., Faessel, B. & Malesieux, G., 1978. Note préliminaire à la proposition de nouvelles méthodes de détermination de la qualité des eaux courantes. *Internal report. Centre Hydrobiol. Univ. Besançon et Lab. Hydroécol. C.T.G.-R.E.F.*

Vlek, H. E., Sporka, F. & K, I., 2006. Influence of macroinvertebrate sample size on bioassessment of streams. *Hydrobiologia*, Volume 566, pp. 523-542.

-W-

Wan, H. et al., 2010. The impact of rare taxa on a fish index of biotic integrity. *Ecological Indicators*, 10(4), pp. 781-788.

Warry, N.-D. & Anau, M., 1993. The use of terrestrial ecoregions as a regional-scale screen for selecting representative reference site for water quality monitoring. *Environmental Management*, 17(2), pp. 267-276.

Wasson, J.-G., 2001. Les questions de recherche posées par la Directive Cadre Européenne sur l'Eau : problématique pour les eaux de surface continentale. *Hydroecol. Appl.*, 13(1), pp. 1-19.

Wasson, J.-G., Chandesris, A., Pella, H. & Blanc, L., 2002. *Les hydro-écorégions de France métropolitaine approche régionale de la typologie des eaux courantes et éléments pour la définition des peuplements de référence d'invertébrés*, Lyon: CEMAGREF.

Wasson, J.-G., Chandesris, A., Pella, H. & Blanc, L., 2004. Les hydro-écorégions: une approche fonctionnelle de la typologie des rivières pour la Directive cadre européenne sur l'eau. *Ingénieries*, Issue 40, pp. 3-10.

Wiederkehr, J. et al., 2015. Experimental study of uncertainties on the macrophyte index (IBMR) based on species identification and cover. *Ecological Indicators*, Issue 50, pp. 242-250.

Woodiwiss, F. S., 1964. The biological system of stream classification used by the River Trent Board. *Chem. Indust*, Volume 14, pp. 443-447.

Wright, J. F. et al., 1988. Analysis of natural and polluted river communities. Report for the period April 1984 - March 1988.. *A report to the Department of the Environment, Scottish Development Department and Welsh Office by the Freshwater Biological Association*.

Wyatt, K. H., Hauer, F. R. & Pessoney, G. F., 2008. Benthic algal response to hyporheic-surface water exchange in an alluvial river. *Hydrobiologia*, 607(1), pp. 151-161.

-Z-

Zinn, J., 2008. *Social theories of risk and uncertainty: an introduction*. Blackwell Pub..

**Estimation des incertitudes associées aux indices
macroinvertébrés et macrophytes pour l'évaluation de l'état
écologique des cours d'eau**

Résumé

L'évaluation de l'état écologique des rivières est régie par l'application de protocoles sur le terrain et au laboratoire. Ceux-ci induisent de nombreuses incertitudes tant naturelles qu'humaines. L'objectif de cette thèse est (1) d'étudier l'impact de certaines incertitudes sur les indices biologiques basés sur les macroinvertébrés et les macrophytes en rivière, (2) de les estimer et (3) d'observer leur propagation jusqu'à l'évaluation de la qualité écologique des rivières. Pour les macroinvertébrés, nous nous sommes intéressés à l'impact de la variation intra-substrat et à l'effet des variabilités liées aux compétences de l'opérateur dans les indices IBGN, IBG DCE et I2M2. Pour les macrophytes, nous nous sommes focalisés sur l'effet opérateur dans l'IBMR. Des variations des indices ont été observées jusqu'à conduire à une évaluation globale erronée. Des solutions correctives peuvent être proposées afin de réduire les effets des sources de variabilités étudiées.

Mots clefs : incertitudes, macroinvertébrés, macrophytes, effet opérateur, variabilité intra-substrat, risques d'évaluation erronée

Abstract

Ecological status assessment in waterbodies depends on standards application in the field and in the laboratory. They induce many uncertainties associated to natural or human variabilities. The aim is (1) to study the effect of some uncertainties on french indices based on macroinvertebrates and macrophytes, (2) to estimate them, and (3) to observe their propagation on final ecological status.

Concerning macroinvertebrates, we are interest in the intrasubstrate variability and the surveyor effect in IBGN, IBG DCE and I2M2. Concerning the macrophytes, we focus on surveyor impact in IBMR. Indices variation appears and could lead to ecological status changes. Control quality approach could be proposed to reduce uncertainties effects.

Keywords: uncertainty, macroinvertebrates, macrophytes, surveyor effect, intrasubstrate variability, missclassification risks