N°d'ordre NNT : 2016LYSEM015

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée au sein de
l'École des Mines de Saint-Etienne

École Doctorale N° 488
Sciences, Ingénierie, Santé

Spécialité de doctorat : Sciences et génie de l'Environnement

Soutenue publiquement le 22/09/2016, par :
Damien Evrard

Approche intégrée pour une évaluation multicritère des meilleures techniques disponibles

Devant le jury composé de :

Gourdon, Rémy	Professeur	INSA Lyon	Président
Azzaro-Pantel, Catherine	Professeure	Université de Toulouse	Rapporteur
Glaus, Mathias	Professeur	ETS Montréal	Rapporteur
Cikankowitz, Anne	Enseignante Chercheuse	EME Rennes	Examinatrice
Laforest, Valérie	Directrice de recherche	ENSM-SE	Directrice de thèse
Villot, Jonathan	Maître assistant	ENSM-SE	Co-encadrant de thèse
Gaucher, Rodolphe	Responsable d'unité Technologies Propres	INERIS	Invité
Darmon, Natacha	Responsable du Groupe Réglementation et Installations	EDF - DPNT -DIPDE	Invitée

Spécialités doctorales	Responsables :	Spécialités doctorales	Responsables
SCIENCES ET GENIE DES MATERIAUX	K. Wolski Directeur de recherche	MATHEMATIQUES APPLIQUEES	O. Roustant, Maître-assistant
MECANIQUE ET INGENIERIE	S. Drapier, professeur	INFORMATIQUE	O. Boissier, Professeur
GENIE DES PROCEDES	F. Gruy, Maître de recherche	IMAGE, VISION, SIGNAL	JC. Pinoli, Professeur
SCIENCES DE LA TERRE	B. Guy, Directeur de recherche	GENIE INDUSTRIEL	X. Delorme, Maître assistant
SCIENCES ET GENIE DE L'ENVIRONNEMENT	D. Graillot, Directeur de recherche	MICROELECTRONIQUE	Ph. Lalevée, Professeur

EMSE : Enseignants-chercheurs et chercheurs autorisés à diriger des thèses de doctorat (titulaires d'un doctorat d'Etat ou d'une HDR)

ABSI	Nabil	CR	Génie industriel	CMP
AUGUSTO	Vincent	CR	Image, Vision, Signal	CIS
AVRIL	Stéphane	PR2	Mécanique et ingénierie	CIS
BADEL	Pierre	MA(MDC)	Mécanique et ingénierie	CIS
BALBO	Flavien	PR2	Informatique	FAYOL
BASSEREAU	Jean-François	PR	Sciences et génie des matériaux	SMS
BATTON-HUBERT	Mireille	PR2	Sciences et génie de l'environnement	FAYOL
BEIGBEDER	Michel	MA(MDC)	Informatique	FAYOL
BLAYAC	Sylvain	MA(MDC)	Microélectronique	CMP
BOISSIER	Olivier	PR1	Informatique	FAYOL
BONNEFOY	Olivier	MA(MDC)	Génie des Procédés	SPIN
BORBELY	Andras	MR(DR2)	Sciences et génie des matériaux	SMS
BOUCHER	Xavier	PR2	Génie Industriel	FAYOL
BRODHAG	Christian	DR	Sciences et génie de l'environnement	FAYOL
BRUCHON	Julien	MA(MDC)	Mécanique et ingénierie	SMS
BURLAT	Patrick	PR1	Génie Industriel	FAYOL
CHRISTIEN	Frédéric	PR	Science et génie des matériaux	SMS
DAUZERE-PERES	Stéphane	PR1	Génie Industriel	CMP
DEBAYLE	Johan	CR	Image Vision Signal	CIS
DELAFOSSE	David	PR0	Sciences et génie des matériaux	SMS
DELORME	Xavier	MA(MDC)	Génie industriel	FAYOL
DESRAYAUD	Christophe	PR1	Mécanique et ingénierie	SMS
DJENIZIAN	Thierry	PR	Science et génie des matériaux	CMP
DOUCE	Sandrine	PR2	Sciences de gestion	FAYOL
DRAPIER	Sylvain	PR1	Mécanique et ingénierie	SMS
FAVERGEON	Loïc	CR	Génie des Procédés	SPIN
FEILLET	Dominique	PR1	Génie Industriel	CMP
FOREST	Valérie	MA(MDC)	Génie des Procédés	CIS
FOURNIER	Jacques	Ingénieur chercheur CEA	Microélectronique	CMP
FRACZKIEWICZ	Anna	DR	Sciences et génie des matériaux	SMS
GARCIA	Daniel	MR(DR2)	Génie des Procédés	SPIN
GAVET	Yann	MA(MDC)	Image Vision Signal	CIS
GERINGER	Jean	MA(MDC)	Sciences et génie des matériaux	CIS
GOEURIOT	Dominique	DR	Sciences et génie des matériaux	SMS
GONDRAN	Natacha	MA(MDC)	Sciences et génie de l'environnement	FAYOL
GRAILLOT	Didier	DR	Sciences et génie de l'environnement	SPIN
GROSSEAU	Philippe	DR	Génie des Procédés	SPIN
GRUY	Frédéric	PR1	Génie des Procédés	SPIN
GUY	Bernard	DR	Sciences de la Terre	SPIN
HAN	Woo-Suck	MR	Mécanique et ingénierie	SMS
HERRI	Jean Michel	PR1	Génie des Procédés	SPIN
KERMOUCHE	Guillaume	PR2	Mécanique et Ingénierie	SMS
KLOCKER	Helmut	DR	Sciences et génie des matériaux	SMS
LAFOREST	Valérie	MR(DR2)	Sciences et génie de l'environnement	FAYOL
LERICHE	Rodolphe	CR	Mécanique et ingénierie	FAYOL
MALLIARAS	Georges	PR1	Microélectronique	CMP
MOLIMARD	Jérôme	PR2	Mécanique et ingénierie	CIS
MOUTTE	Jacques	CR	Génie des Procédés	SPIN
NIKOLOVSKI	Jean-Pierre	Ingénieur de recherche	Mécanique et ingénierie	CMP
NORTIER	Patrice	PR1		SPIN
OWENS	Rosin	MA(MDC)	Microélectronique	CMP
PERES	Véronique	MR	Génie des Procédés	SPIN
PICARD	Gauthier	MA(MDC)	Informatique	FAYOL
PIJOLAT	Christophe	PR0	Génie des Procédés	SPIN
PIJOLAT	Michèle	PR1	Génie des Procédés	SPIN
PINOLI	Jean Charles	PR0	Image Vision Signal	CIS
POURCHEZ	Jérémy	MR	Génie des Procédés	CIS
ROBISSON	Bruno	Ingénieur de recherche	Microélectronique	CMP
ROUSSY	Agnès	MA(MDC)	Génie industriel	CMP
ROUSTANT	Olivier	MA(MDC)	Mathématiques appliquées	FAYOL
STOLARZ	Jacques	CR	Sciences et génie des matériaux	SMS
TRIA	Assia	Ingénieur de recherche	Microélectronique	CMP
VALDIVIESO	François	PR2	Sciences et génie des matériaux	SMS
VIRICELLE	Jean Paul	DR	Génie des Procédés	SPIN
WOLSKI	Krzystof	DR	Sciences et génie des matériaux	SMS
XIE	Xiaolan	PR1	Génie industriel	CIS
YUGMA	Gallian	CR	Génie industriel	CMP

Mise à jour : 01/02/2016

« L'homme est la mesure de toutes choses : de celles qui sont, du fait qu'elles sont ; de celles qui ne sont pas, du fait qu'elles ne sont pas. »

Protagoras

Remerciements

Mes remerciements sont tout d'abord adressés à mes encadrants Valérie Laforest et Jonathan Villot à l'École des Mines ; Rodolphe Gaucher à l'INERIS ; Natacha Darmon, Sofia Bouhrizi, Lucien Briand et Sébastien Bor pour EDF. En particulier, merci pour ces discussions qui ont permis de redonner de l'énergie dans les moments difficiles tout au long de ces trois années.

Merci également à Catherine Azzaro-Pantel, Anne Cikankowitz, Mathias Glaus et Rémy Gourdon d'avoir évalué ce travail de thèse.

Je remercie les personnes impliquées dans le Processus de Séville et dont les discussions m'ont permis d'avancer les membres de l'EIPPCB rencontrés à Séville et en particulier Serge Roudier ; les membres du groupe miroir de l'industrie agroalimentaire et tout particulièrement Hélène Pérennou et Pierre Barrucand du CNIEL, Éric Mousset du Ministère de l'Environnement, Pierre Kerdonkuff de l'ADEME.

Un grand merci à Chadad Armiyaou pour son travail de stage sur l'analyse statistique.

De mon passage à l'École des Mines, je retiendrai les cafés pris avec les collègues qui, comme dans tout lieu de travail, sont les moments les plus enrichissants. Merci à vous.

Merci aux amis qui m'ont soutenu pendant ces trois années : Tiphaine, Greg, Will et Greg ; à Fabrice Mathieux pour m'avoir donné le goût de la recherche.

Et les personnes les plus importantes : Solange pour m'avoir supporté et Sophie pour sa joie de vivre communicative.

Sommaire

Remerciements iii
Liste des figures xi
Liste des tableaux xv
Acronymes xix
Introduction générale 1
Partie 1 : État de l'art 5
Introduction 7
Chapitre 1 : Contexte réglementaire 9
1. La réglementation des émissions industrielles : du cadre européen à l'application nationale 9
1.1. Principes de la directive 9
1.2. Les MTD : Qui sont-elles ? D'où viennent-elles ? Où les trouver ? 10
1.2.1. Qui sont-elles ? 10
1.2.2. D'où viennent-elles ? Où les trouver ? 12
1.3. Transposition de l'IPPC dans la réglementation française et retour d'expérience 13
2. Le processus décisionnel de détermination des MTD : le Processus de Séville 17
2.1. Objectifs 17
2.2. BATAEL 19
2.3. Les acteurs du Processus de Séville 19
2.3.1. Le Bureau Européen de l'IPPC (EIPPCB) 19
2.3.2. Les groupes de travail techniques (TWG) 20
2.3.3. Les groupes miroirs 21
2.3.4. Les acteurs locaux 21
2.4. Étapes du Processus de Séville et pratiques actuelles 21
2.4.1. Étape 1 : Réactivation du TWG et appel à desiderata 22
2.4.2. Étape 2 : Réunion de lancement 23
2.4.3. Étape 3 : Collecte de données 26
2.4.4. Étape 4 : Projets formels de BREF 28
2.4.5. Étape 5 : Réunion finale 29
2.4.6. Étape 6 : Projet formel final 30
2.5. Conclusions sur le Processus de Séville 30
3. Les MTD dans les secteurs non-IED 31
3.1. Définition d'une INB 31
3.2. Principaux textes réglementaires 31
3.3. Conclusion 32
4. Conclusions sur l'état de l'art réglementaire 32
4.1. Problématique générale 33
4.2. Place des experts 35
4.3. Échelles géographiques 35
Chapitre 2 : Méthodes existantes 37
1. Méthodes applicables au niveau d'un secteur industriel (échelles internationale et nationale) 39
1.1. Réponses de la littérature sur les besoins au niveau sectoriel 39
1.1.1. Sur la définition d'enjeux-clés 39
1.1.2. Sur la recherche d'installations de référence 40
1.1.3. Sur la détermination des MTD et BATAEPL 41

1.1.4. Aide à la décision dans l'existant et évaluation multicritère....43
2. Méthodes applicables au niveau d'une installation....43
2.1. Comparaison des performances....44
2.2. Sélection des MTD pour une installation....44
2.3. Évaluation des performances en fonction des conditions locales....45
3. Synthèse pour une approche intégrée des MTD de l'échelle européenne à l'échelle locale. 46
3.1. Étape 1 : Définition du périmètre et des variables....47
3.2. Étape 2 : Traitement des données....47
3.3. Étape 3 : Classification des installations....48
3.4. Étape 4 : Analyse des classifications....49
3.5. Étape 5 : Interprétation des sélections....49
3.5.1. Définition de valeurs de référence....49
3.5.2. Recherche de MTD candidates....50
4. Application de la démarche à l'échelle locale vs européenne....50
Conclusions de la partie 1....51
Partie 2 : Méthodologie proposée....53
Introduction....55
Étapes de la méthodologie et outils employés....57
1. Étape 1 : Définition du périmètre et des variables d'étude....58
1.1. Périmètre de l'étude....59
1.2. Variables d'étude....59
1.2.1. Variables d'analyse....60
1.2.2. Variables de normalisation....63
1.3. Livrable de l'étape 1....63
2. Étape 2 : Traitement des données....64
2.1. Outils statistiques explorés....65
2.1.1. Méthode d'imputation simple : K plus proches voisins....65
2.1.2. Méthode d'imputation simple : Imputation par la régression....65
2.1.3. Méthode d'imputation multiple : MissForest....65
2.2. Utilisation de Miss Forest dans la méthodologie....66
2.3. Implication des acteurs....69
2.4. Livrable de cette étape....69
2.5. Remarques sur les résultats obtenus à l'étape 2....70
3. Étape 3 : Classification des installations....70
3.1. Approche représentative....71
3.1.1. Centres mobiles....71
3.1.2. Nuées dynamiques....71
3.1.3. Classification ascendante hiérarchique....71
3.2. Approche performante....73
3.3. Complémentarité des deux approches....74
3.4. Caractérisation des groupes identifiés....74
3.5. Livrable de l'étape....76
4. Étape 4 : Analyse de la classification....76
4.1. Cas des sites représentatifs....76
4.2. Cas des sites performants....77
4.3. Livrable de l'étape 4....79
5. Étape 5 : Interprétation de la sélection....80
5.1. Détermination des valeurs de référence et du « taux d'effort » des installations....80
5.2. Aide à l'identification de techniques candidates au statut de MTD....82
5.3. Implication des acteurs....83

5.4. Livrables de l'étape 5 84
Conclusions de la partie 2 85
Partie 3 : Cas d'application 89
Chapitre 1 : Application à l'industrie laitière française 91
1. Application de la méthodologie 91
1.1. Étape 1 : Définition du périmètre et des variables d'étude 91
1.2. Étape 2 : Traitement et consolidation des données 94
1.3. Étape 3 : Classification des installations 97
1.3.1. Approche représentative et caractérisation des classes 97
1.3.1.1. Caractérisation des classes 98
1.3.1.2. Livrable de l'étape 99
1.3.2. Classification performante et caractérisation des classes 99
1.4. Étape 4 : Analyse de la classification 99
1.4.1. Liste de sites représentatifs et niveaux de référence 100
1.4.2. Liste de sites performants et niveaux de référence 101
1.5. Étape 5 : Interprétation des sélections 101
1.5.1. Taux d'effort dans les sites représentatifs 101
1.5.2. Taux d'effort dans les sites performants 102
2. Interprétation et conclusions 104
2.1. Rejets directs vs rejets indirects 104
2.2. Impact du multi-produit 104
2.3. Comparaison de la liste des sites de référence à la sélection des sites retenus à dire d'experts 104
2.4. Conclusion 106
Chapitre 2 : Application à l'industrie laitière européenne 107
1. Application de la méthodologie 108
1.1. Étape 1 : Définition du périmètre et des variables d'étude 108
1.1.1. Variable de normalisation 110
1.1.2. Variables d'analyse 111
1.1.3. Variables descriptives 112
1.1.4. Livrable de l'étape 113
1.2. Étape 2 : Traitement et consolidation des données 114
1.3. Étape 3 : Classification des installations 117
1.3.1. Classification représentative et caractérisation des classes 117
1.3.1.1. Classification des installations 117
1.3.1.2. Caractérisation des classes 118
1.3.1.3. Livrable de l'étape 119
1.3.2. Classification performante et caractérisation des classes 119
1.4. Étape 4 : Analyse de la classification 120
1.4.1. Liste de sites représentatifs et niveaux de référence 120
1.4.2. Liste de sites performants et niveaux de référence 121
1.5. Étape 5 : Interprétation des sélections 121
1.5.1. Taux d'effort dans les sites représentatifs 122
1.5.2. Taux d'effort dans les sites performants 123
1.5.3. Aide à l'identification de techniques candidates MTD 125
2. Interprétation et conclusions 125
Chapitre 3 : Application aux INB 129
1. Application à l'échelle nationale 129
1.1. Étape 1 : Définition du périmètre et des variables d'étude 130

1.1.1. Jugement d'experts dans le choix des variables....130
1.1.2. Sélection des variables d'analyse par les tests statistiques....131
1.1.3. Conclusion et livrable de l'étape 1....133
1.2. Étape 2 : Traitement et consolidation des données....134
1.3. Étape 3 : Classification des installations....134
1.3.1. Approche représentative et caractérisation des classes....135
1.3.2. Approche performante et caractérisation des classes....136
1.3.3. Livrables de l'étape 3....136
1.4. Étape 4 : Analyse de la classification....137
1.4.1. Liste de sites représentatifs....137
1.4.2. Liste de sites performants....138
1.4.3. Livrable de l'étape 4....139
1.5. Étape 5 : Interprétation des sélections....140
1.5.1. Détermination des valeurs de référence potentielles et du taux d'effort....140
1.5.1.1. Sites représentatifs....140
1.5.1.2. Sites performants....142
1.5.2. Aide à l'identification des corrélations entre caractéristiques des installations et niveaux de performance....142
1.5.2.1. Avis des experts sur les émissions de chlorures....142
1.5.2.2. Interprétation statistique....143
2. Application à l'échelle internationale....144
2.1. Étape 1 : Définition du périmètre et des variables d'étude....144
2.2. Étape 2 : Traitement des données....146
2.2.1. Livrables....147
2.3. Étape 3 : Classification des installations....148
2.3.1. Approche représentative et caractérisation des classes....148
2.3.2. Approche performante et caractérisation des classes....153
2.4. Étape 4 : Analyse des classifications....153
2.4.1. Sélection représentative....154
2.4.2. Sélection performante....155
2.5. Étape 5 : Interprétation des résultats....156
2.5.1. Interprétation de la sélection représentative....156
2.5.2. Interprétation de la sélection performante....157
2.5.3. Livrables de l'étape....157
3. Conclusions de l'application aux INB....158

Partie 4 : Discussions et perspectives....161
1. Confrontation des hypothèses avec les résultats....165
1.1. Intégration des trois échelles géographiques....165
1.2. Approche structurée, statistique et multicritère....165
1.3. Caractéristiques de la méthodologie....166
2. Discussions autour de la méthodologie proposée....167
2.1. Quantité et qualité des données....167
2.2. Outils d'aide à la décision....168
2.2.1. Outil d'aide à la sélection des installations représentatives....168
2.2.2. Outil d'aide à la sélection des installations performantes....168
2.3. Détermination de MTD candidates....169
2.4. Intégration de différents niveaux d'analyse....169
2.5. Impacts (socio-économiques) de l'approche pour les acteurs....170
3. Perspectives du travail de recherche....171
3.1. Mise en œuvre dans le Processus de Séville....171
3.2. Applications à d'autres secteurs industriels....171
3.3. Utilisation d'autres outils statistiques....171

3.4. Application à l'échelle locale 172
Conclusion 172
Conclusion générale 173
Bibliographie 175
Annexes 181
Annexe 1 : Description de MissForest 183
1. Les forêts aléatoires 183
2. MissForest 184
3. Estimation de l'erreur de la reconstruction 184
Annexe 2 : Recherche des sources d'erreur pour la reconstruction 187
Annexe 3 : ACP et CAH 189
1. Analyse en Composantes Principales (ACP) 189
2. Classification ascendante hiérarchique 190
Annexe 4 : Front de Pareto 193
Annexe 5 : Boîtes à moustaches des variables d'analyse pour les classes représentatives de l'industrie laitière française 195
Annexe 6 : Valeurs de référence pour l'industrie laitière française 197
Annexe 7 : Boîtes à moustaches pour les classes représentatives de l'industrie laitière européenne 199
1. Variable de normalisation 199
2. Variables d'analyse 199
3. Variables descriptives 201
3.1. Proportions de produits fabriqués 201
3.2. Proportion de produits entrants 202
3.3. Déchets 203
Annexe 8 : Valeurs de référence pour les installations laitières européennes 205
Annexe 9 : Publications scientifiques 207

Liste des figures

Figure 1 : Chronologie des principaux textes relatifs à l'IPPC.....15
Figure 2 : Échanges entre les acteurs nationaux et européens pendant le Processus de Séville.....18
Figure 3 : Étapes du Processus de Séville lors de la révision d'un BREF, adapté de Commission Européenne (2012)..... 22
Figure 4 : Positionnement des interprétations du concept d'installation de référence.....25
Figure 5 : Positionnement de la méthode par rapport au Processus de Séville et au groupe miroir, et positionnement de l'expert..... 34
Figure 6 : Positionnement de la littérature scientifique par rapport aux étapes de la méthodologie proposée..... 47
Figure 7 : Étapes de la méthodologie proposée..... 58
Figure 8 : Processus décisionnel sur l'aide au choix des variables d'analyse.....61
Figure 9 : Outil d'aide à la décision pour la sélection des variables proposé pour l'étape 2.....67
Figure 10 : Outil d'aide à la décision pour la sélection de la population proposé pour l'étape 2.....68
Figure 11 : Outil d'aide à la décision pour la validation d'une nouvelle population pour l'étape 2. 69
Figure 12 : Exemple de dendrogramme obtenu par la classification ascendante hiérarchique avec découpage à deux classes..... 72
Figure 13 : Exemple de boîtes à moustaches de Tukey représentant les valeurs prises par les sites de différentes classes pour une variable d'analyse..... 75
Figure 14 : Outil d'aide à la sélection des installations représentatives.....77
Figure 15 : Aperçu de l'outil d'aide à la décision pour le taux d'effort (exemple de la concentration en DCO)..... 81
Figure 16 : Livrable 5.2 – Aide au choix des techniques candidates MTD.....83
Figure 17 : Étapes de la méthodologie et outils mathématiques employés.....86
Figure 18 : Aperçu des différents apports de la méthodologie en fonction du contexte d'application..... 87
Figure 19 : Pourcentage de données manquantes dans les installations laitières françaises et taux d'erreur global de la reconstruction avec MissForest pour les 3 variables d'analyse incomplètes.....94
Figure 20 : Taux d'erreur et données manquantes pour la population d'installations laitières françaises proposée par rapport à la population initiale et par variable pour la population suggérée. 95
Figure 21 : Taux d'erreur et de données manquantes par variable pour la population d'installations laitières françaises suggérée..... 96
Figure 22 : Dendrogramme de la classification représentative des installations laitières françaises..98
Figure 23 : Boîtes à moustaches représentatives de variables d'analyse pour les deux classes d'installations laitières françaises dans l'approche représentative.....98
Figure 24 : Outils d'aide à la sélection des sites de référence pour les installations laitières françaises..... 100

Figure 25 : Taux d'effort pour la classe représentative n°1 des installations laitières françaises (flux de DCO)..... 102
Figure 26 : Taux d'effort par rapport aux installations laitières françaises performantes pour les émissions de DCO..... 103
Figure 27 : Taux d'erreur global et par variable de la reconstruction avec MissForest pour les variables d'analyse initiales de l'industrie laitière européenne..... 115
Figure 28 : Taux d'erreur global et par variable de la reconstruction avec MissForest pour les variables d'analyse modifiés et les sites laitiers européens pour lesquels la variable de normalisation est renseignée..... 116
Figure 29 : Erreurs pour la population initiale et une population suggérée de l'industrie laitière européenne..... 116
Figure 30 : Erreur par variable de la reconstruction de la population laitière européenne suggérée. 117
Figure 31 : Dendrogramme de la seconde classification représentative des installations laitières européennes..... 118
Figure 32 : Exemple de comparaison des classes pour deux variables d'analyse (DBO5 et MES) pour l'industrie laitière européenne..... 118
Figure 33 : Outils d'aide à la sélection des sites de référence pour les installations laitières européennes..... 121
Figure 34 : Taux d'effort pour les deux classes représentatives de l'industrie laitière européenne (concentration de DCO)..... 123
Figure 35 : Taux d'effort par rapport aux installations laitières européennes performantes pour les émissions de DCO..... 124
Figure 36 : Livrable 5.2 – Aide au choix du recyclage d'eau comme MTD pour l'industrie laitière européenne..... 125
Figure 37 : Dendrogramme de la classification représentative des INB françaises.....135
Figure 38 : Comparaison des classes représentatives des INB françaises pour 2 des 14 variables d'analyse..... 136
Figure 39 : Outils d'aide à la sélection des INB françaises de référence par couplage des distances intra-classes et inter-classes..... 138
Figure 40 : Positionnement des INB françaises par rapport aux valeurs de référence pour les deux classes représentatives..... 141
Figure 41 : Livrable 5.2 – Aide au choix des techniques de traitement biocide par rapport aux émissions de chlorures pour les INB françaises.....144
Figure 42 : Erreur de reconstruction des données sur les INB internationales initialement sélectionnées..... 147
Figure 43 : Dendrogramme de la première classification représentative des INB internationales.149
Figure 44 : Boîtes à moustaches pour les émissions liquides bêta-gamma et de tritium entre les 3 classes représentatives obtenues pour les INB internationales.....150
Figure 45 : Dendrogramme de la seconde classification sur les INB internationales de la classe 2..151

Figure 46 : Boîtes à moustaches pour les émissions de tritium liquide, de gaz rares et de rayonnements bêta-gamma entre les 3 classes de la seconde classification des INB internationales. 152
Figure 47 : Outil d'aide à la sélection des INB internationales de référence par couplage des distances intra-classes et inter-classes (classe 1)..... 154
Figure 48 : Représentation du taux d'effort pour la classe représentative n°1 des INB internationales..... 157
Figure 49 : Taux d'effort par rapport aux valeurs pertinentes pour l'approche performante des INB internationales..... 157
Figure 50 : Synthèse du fonctionnement de la méthodologie..... 163
Figure 51 : Exemple d'arbre de décision pour déterminer si une personne est malade (source Wikipédia)..... 183
Figure 52 : Exemple simple de reconstruction avec les forets aléatoires (d'après Pierre Gaillard).. 184
Figure 53 : Principe de l'Analyse en Composantes Principales..... 189
Figure 54 : Front de Pareto (d'après Nojhan pour Wikimedia)..... 194
Figure 55 : Dendrogramme obtenu avec une Classification Ascendante Hiérarchique..... 194

Liste des tableaux

Tableau 1 : Les 12 critères pour la sélection des Meilleures Techniques Disponibles tels qu'énoncés dans l'annexe III de l'IED..... 12

Tableau 2 : Principaux aspects des articles trouvés..... 38

Tableau 3 : Comparaison des définitions d'installations de référence en fonction des acteurs.....48

Tableau 4 : Questions, hypothèses, postulats et caractéristiques de la méthodologie dans le cadre de cette thèse..... 52

Tableau 5 : Liste des tests statistiques non paramétriques disponibles.....62

Tableau 6 : Exemple de tableau montrant les dépendances et corrélations entre variables.....63

Tableau 7 : Livrable 1 – Représentation de la base de données des variables d'analyse, descriptives et de normalisation à l'issue de l'étape 1 (la lettre X représentant une valeur quantitative, la lettre M une modalité d'une variable descriptive et le point d'interrogation une donnée manquante).....64

Tableau 8 : Comparatif des outils statistiques possibles pour la reconstruction des données de l'étape 2..... 66

Tableau 9 : Livrable 2 – Représentation de la base de données des variables d'analyse reconstruites, descriptives et de normalisation à l'issue de l'étape 2 (la lettre X représentant une valeur quantitative, R une valeur reconstruite, M une modalité d'une variable descriptive et le point d'interrogation une donnée manquante).....70

Tableau 10 : Comparatif des outils statistiques pour la classification des sites représentatifs.....73

Tableau 11 : Livrable 3.1 – Dépendances entre le classement représentatif et les variables descriptives..... 75

Tableau 12 : Livrable 3.2 – Liste des installations représentatives ou performantes.....76

Tableau 13 : Livrables 4.1 – Listes ordonnées des sites pour les approches représentatives et performantes..... 79

Tableau 14 : Livrable 5.1 – Valeurs de référence par variable et par classe et taux d'efforts associés.. 82

Tableau 15: Liste des variables d'analyse de l'industrie laitière française.....92

Tableau 16 : Pourcentage d'installations laitières françaises pour chaque nombre de variables d'analyse manquante sur les 9 variables d'analyse choisies.....93

Tableau 17 : Livrable 1 – Base de données initiale pour les installations laitières françaises.....93

Tableau 18 : Livrable 2 – Matrice de données reconstruites d'installations laitières françaises.....97

Tableau 19 : Livrable 3.2 – Code des installations laitières françaises classées par groupe représentatif..... 99

Tableau 20 : Livrable 3.2 – Liste des installations laitières françaises performantes.....99

Tableau 21 : Livrable 3.2a – Liste des installations représentatives de référence hiérarchisées par classe..... 101

Tableau 22 : Livrable 5 – Valeurs de référence par variable et par classe et taux d'efforts associés dans le cas des installations laitières françaises.....103

Tableau 23 : Résultat des tests d'hypothèse du lien entre la classification représentative et les produits fabriqués de l'industrie laitière française.....104

Tableau 24 : Position des sites laitiers français sélectionnés par le CNIEL dans la classe représentative n°1..... 105

Tableau 25 : Comparaison des valeurs de référence pour la sélection d'installations laitières françaises de la profession et les sélections issues de la méthodologie, exemple de la DCO (kg/T de matière laitière entrante)..... 105

Tableau 26 : Effectifs des installations laitières européennes par État membre, d'après EIPPCB (2015)..... 109

Tableau 27 : Importance des données manquantes ou de valeur nulle par variable dans les installations laitières européennes (valeurs annuelles)..... 110

Tableau 28 : Importance des données manquantes pour les variables de normalisation possibles de l'industrie laitière européenne..... 111

Tableau 29 : Liste des variables d'analyse de l'industrie laitière européenne (valeurs annuelles).....112

Tableau 30 : Liste des variables descriptives de l'industrie laitière européenne.....113

Tableau 31 : Livrable 1 – Extrait de la base de données initiale pour les installations laitières européennes..... 114

Tableau 32 : Livrable 3.2 – Liste des installations laitières européennes classées par groupes représentatifs..... 119

Tableau 33 : Livrable 3.2 – Liste des installations laitières européennes performantes.....120

Tableau 34 : Livrable 3.2a – Liste des installations laitières européennes de référence classées par groupe représentatif..... 121

Tableau 35 : Valeurs de référence pour l'exemple de la DCO pour l'approche représentative des sites laitiers européens..... 122

Tableau 36 : Taux d'effort pour l'exemple de la DCO pour l'approche représentative des sites laitiers européens..... 122

Tableau 37 : Valeurs de référence pour l'exemple de la DCO de la sélection performante de l'industrie laitière européenne..... 124

Tableau 38 : Livrable 5 – Valeurs de référence par variable et par classe et taux d'efforts associés de l'industrie laitière européenne exemple de la DCO.....124

Tableau 39 : Étapes de la méthodologie et tiroirs méthodologiques en fonction du niveau d'intégration dans le Processus de Séville.....126

Tableau 40 : Liste des 14 variables d'analyse choisie sur avis d'experts pour les INB françaises.....131

Tableau 41 : Extrait de la table des dépendances entre variables pour les INB françaises.....132

Tableau 42 : Liste des 14 variables suggérées suite à l'étude des dépendances entre variables d'analyse initiales des INB françaises..... 133

Tableau 43 : Comparaison des variables identifiées par les experts et par l'analyse statistique pour les INB françaises..... 133

Tableau 44 : Livrable 1 – Extrait de la base de données initiale pour les INB françaises.....134

Tableau 45 : Livrable 3.1 – Dépendances entre classes et variables descriptives pour les INB françaises..... 137

Tableau 46 : Livrable 3.2 – Liste des INB françaises classées selon les approches représentatives et performantes..... 137

Tableau 47 : Installations nucléaires françaises performantes classées selon l'INPE.....139
Tableau 48 : Livrable 4 – Listes ordonnées des INB françaises pour les approches représentatives et performantes..... 140
Tableau 49 : Taux d'effort pour les 2 classes représentatives des INB françaises.....142
Tableau 50 : Taux d'effort des INB françaises sur les émissions de chlorures par rapport aux installations performantes..... 142
Tableau 51 : Effectifs des INB internationales par pays et codage employé.....145
Tableau 52 : Livrable 1 – Base de données initiales pour les INB internationales.....146
Tableau 53 : Livrable 2 – Matrice de données reconstruites et normalisées des INB internationales. 148
Tableau 54 : Livrable 3.1 – Dépendances entre le classement représentatif et les variables descriptives des INB internationales..... 149
Tableau 55 : Livrable 3.2 – Extrait de la liste des INB internationales représentatives dans chaque classe..... 152
Tableau 56 : Livrable 3.1 – Dépendances entre classification performante et variables descriptives pour les INB internationales..... 153
Tableau 57 : Livrable 3.2 – Liste des installations performantes et non optimales des INB internationales..... 153
Tableau 58 : Livrable 4 – Extrait de la liste des INB internationales représentatives présentation des sites de la classe 1..... 155
Tableau 59 : Liste des INB internationales performantes retenues et valeur de leur INPE.....156
Tableau 60 : Livrable 5 – Valeurs de référence et taux d'effort pour les émissions de tritium liquide pour les INB internationales..... 158
Tableau 61 : Question, hypothèses, postulats et caractéristiques de la méthodologie dans le cadre de cette thèse..... 164
Tableau 62 : État de validation, limite et perspectives de la réponse de la méthodologie aux hypothèses et postulats..... 172
Tableau 63: Limites et perspectives des étapes de la méthodologie.....174

Acronymes

ACP	Analyse en Composantes Principales.
ACV	Analyse du Cycle de Vie
Al	Aluminium
ANP	*En anglais, Analytic Network Process.*
AOX	Halogène Organique Adsorbable
BATAEL	En anglais, Best Available Technique Associated Emission Levels, i.e. les niveaux d'émissions associés aux MTD ; ils font partie des BATAEPL.
BATAEPL	En anglais, Best Available Technique Associated Environmental Performance Level, i.e. tout indicateur de performance environnemental associé aux MTD en conditions normales et anormales de fonctionnement.
BREF	En anglais, Best available technique REFerence document, i.e. le document de référence sur les meilleures techniques disponibles.
CAH	Classification Ascendante Hiérarchique.
CNIEL	Centre National Interprofessionnel de l'Économie Laitière.
CO	Monoxyde de carbone
COT	Carbone Organique Total
Cr	Crome
CRT	Chlore Résiduel Total
Cu	Cuivre
CWW	En anglais, Common Waste Water and waste gas treatment / Management systems in the chemical sector document de référence sur les techniques de traitement des gaz et eaux usées communes à l'industrie chimique.
DBO_5	Demande Biologique en Oxygène à 5 jours.
DCO	Demande Chimique en Oxygène
DREAL	Direction Régionale de l'Environnement, de l'Aménagement et du Logement.
ECM	En anglais, Economics and Cross-Media Effects ; document de référence sur la prise en compte des critères économiques et des effets croisés des MTD.
EDF – DPNT – DIPDE	Électricité De France – Direction Production Nucléaire et Thermique – Division de l'Ingénierie du Parc, du Démantèlement et de l'Environnement.
. EIPPCB	En anglais, European IPPC Bureau, i.e. l'organise en charge de la mise en œuvre du Processus de Séville.
FDM	En anglais, Food – Drink & Milk ; nom du BREF agroalimentaire.
ICPE	Installation Classée pour la Protection de l'Environnement.
IED	Directive européenne 2010/75/CE sur les émissions industrielles ; en anglais, Industrial Emission Directive. Elle remplace et abroge la directive IPPC.
INERIS	Institut National de l'Environnement industriel et des RISques.
INPE	Indicateur de Niveau de Performance Environnementale.
IPPC	Directive européenne 96/62/CE relative à la prévention et à la réduction intégrée de la pollution.
KEI	En anglais, Key Environmental Issue, i.e. les enjeux environnementaux-clés pris en compte dans le Processus de Séville.
L-BAT	Local-BAT
MES	Matières En Suspension.
MTD	Meilleures Techniques Disponibles.
Na	Sodium
NH_4-N	Azote ammoniacal

NOx	Oxydes d’azote
O_2	Dioxygène
Pb	Plomb
PF-PA	Produits de Fission – Produit d’Activation
ROM	En anglais, Monitoring of emissions from IED installations ; document de référence sur les aspects généraux de mesure des émissions.
SO_2	Dioxyde de Soufre
STEP	Station d’épuration des eaux usées
TWG	En anglais, Technical Working Group, i.e. le groupe de travail technique qui participe à la rédaction des BREF.
UE	Union Européenne
VITO	Institut flamand de recherche technologique.
VLE	Valeur Limite d’Émission
Zn	Zinc

Introduction générale

La majeure partie de la réglementation française relative à l'environnement découle de règlements ou directives européennes. L'évolution des stratégies de gestion de l'environnement est essentiellement guidée par l'évolution des textes réglementaires. Ainsi, les 200 actes législatifs définis dans les années 70 sont basés sur une approche essentiellement curative permettant de limiter les pollutions en fin de chaîne de production. Cette approche curative s'intéressant de manière indépendante aux différents domaines environnementaux (déchets, eau, air, sols) par l'introduction de valeurs de rejets maximales. Or, cette approche n'était pas suffisante pour garantir une protection à long terme des ressources naturelles. Même si les activités industrielles contribuent de façon positive au développement économique, elles participent également à la pollution de l'environnement dans l'eau, l'air, le sol, à la production des déchets ainsi qu'à la consommation énergétique. Par conséquent, grâce à la prise de conscience générale des risques liés aux problèmes globaux mais aussi locaux de l'environnement, la nécessité d'adopter une approche dynamique et concertée à l'échelle européenne et internationale est devenue incontournable. Ainsi, des stratégies plus proactives et dynamiques reposant sur des approches intégrées et préventives ont vu le jour dans les années 1990. C'est ainsi que la directive européenne n°96-61-CE dite IPPC a été adoptée par la Commission Européenne, le 24 septembre 1996 pour les émissions chroniques liées aux activités « normales » des industries. Les émissions accidentelles étant gérées de façon indépendante comme avec la réglementation Seveso.

La directive IPPC, abrogée par l'IED (2010/75/CE) en 2010, définit le cadre réglementaire imposé aux États membres et le concept de Meilleures Techniques Disponibles (MTD). Ce concept de MTD et les niveaux de performance environnementale qui lui sont associés permettent de fixer des objectifs communs à l'ensemble de l'Europe tout en intégrant une prise en compte des spécificités locales. En parallèle, l'application du concept de MTD à d'autres secteurs non soumis à l'IED (par exemple : Industries extractives, Installations Nucléaires de Base) en fait un enjeu majeur pour l'ensemble de la communauté industrielle européenne. Afin de permettre l'application des MTD et de proposer une équité d'accès à l'information des différents États membres, la Commission Européenne doit, via l'article 13 de l'IED, mettre en œuvre un échange d'information pour produire les principaux éléments des documents de référence relatifs aux MTD.

Ces documents de référence sur les MTD (dénommés BREF) sont devenus incontournables pour les installations industrielles soumises à l'IED, aussi bien comme guides d'aide au choix de techniques qu'en tant que références pour l'évaluation de leurs performances environnementales. Comme le mentionne l'article 13 de la directive IED : « [...] il convient que la Commission adopte,

par la procédure de comité, des lignes directrices sur la collecte de données, sur l'élaboration des documents de référence MTD et sur leur assurance qualité. Les conclusions sur les MTD devraient servir de référence pour fixer les conditions d'autorisation. » Afin d'assurer l'amélioration continue des processus de gestion de l'environnement, un processus de révision des documents de référence existe. Ainsi, tous les dix ans environ, le processus de révision basé sur la recherche d'un consensus entre parties prenantes est déployé pour une mise à jour des BREF : le Processus de Séville.

Toutefois, les travaux de révision des BREF deviennent de plus en plus conséquents et stratégiques compte tenu des enjeux économiques et environnementaux liés à la mise en œuvre des MTD pour l'industrie. Ils font donc l'objet d'arbitrages entre l'amélioration des performances environnementales et la faisabilité technico-économique. En effet, le processus de révision comprend une partie de collecte de données relatives aux performances de sites actuellement en activité en Europe qui conduit à devoir analyser une masse d'informations croissante. Par conséquent, pour aboutir à un consensus, un traitement et une analyse de données de plus en plus complexe sont nécessaires.

Dans ce contexte réglementaire, le projet de recherche dont est issue cette thèse s'inscrit dans le cadre d'un partenariat avec l'Institut National de l'Environnement et des Risques Industriels (INERIS) et Électricité De France (EDF). Le retour d'expérience de l'INERIS comme support technique du Ministère de l'Environnement français dans le Processus de Séville a mis en exergue la nécessité d'une méthodologie formelle utilisant les données disponibles pour aider à développer un argumentaire sur le choix des installations de référence proposées à Séville visant à représenter la position nationale d'une part et pour l'aide à la sélection des techniques candidates aux MTD d'autre part. Quant à EDF, le besoin de justifier de la mise en œuvre des MTD lors du réexamen des impacts des installations nucléaires de base (INB) se fait sentir et ce, sans disposer d'un équivalent du Processus de Séville pour ce secteur.

Par conséquent, ces deux besoins consistent en un passage d'une étape d'analyse d'information à la définition de MTD en réponse à la problématique de définition de niveaux de performances permettant de positionner les MTD par rapport à un échantillon d'installations industrielles. Ainsi, ce travail de thèse s'est intéressé à la prise de décision dans le cadre de la définition des MTD et a notamment visé à répondre à la question suivante :

Quelle méthodologie déployer afin de répondre aux manques constatés dans le cadre du processus de sélection des MTD et de détermination des niveaux de performances associés ?

Une méthodologie structurée en cinq étapes et reposant sur des outils statistiques multicritères a été développée. Elle permet, à partir de données provenant d'installations industrielles, en particulier des valeurs de consommations et d'émissions, de classifier les installations selon une approche représentative ou performante. La première approche consiste à rechercher les classes d'installations représentatives de niveaux de consommations et d'émissions caractéristiques alors que la seconde repose sur l'identification d'installations optimales via l'application du front de Pareto. Chaque classification permet alors de déterminer des installations de référence qui servent à définir des valeurs de consommations et d'émissions pour la détermination des MTD et BATAEPL. La méthodologie proposée se positionne donc comme un support pour l'aide à la prise de décision des experts impliqués dans la détermination des MTD.

Ainsi, la première partie de ce mémoire s'intéresse à l'état de l'art qui a permis de mettre en évidence les limites des processus actuels de décision et de définir les caractéristiques de la méthodologie proposée. Dans la seconde partie, la méthodologie développée est présentée ainsi que les outils mathématiques utilisés à chacune de ses étapes. Dans une troisième partie, une application aux échelles géographiques nationales et internationales pour un secteur IED (l'industrie laitière) et un secteur non-IED (les installations nucléaires de base) est réalisée. Enfin, la quatrième partie fait un retour sur les différents postulats et hypothèses et présente une discussion ainsi que les perspectives de développements futurs et les limites de ce travail.

Partie 1 :
État de l'art

Introduction

Le concept de MTD est l'élément central autour duquel porte cette thèse. Il a été introduit initialement par la directive européenne IPPC (*Integrated Pollution Prevention and Control*) (Commission Européenne, 1996) puis a été renforcé et étendu dans la directive européenne sur les émissions industrielles dite IED (Commission Européenne, 2010). Cette partie du mémoire a pour objectif de présenter l'état de l'art relatif aux MTD sur des aspects tant réglementaires que scientifiques. Le premier chapitre décrit l'historique du concept de MTD et son application en France afin de mettre en exergue les besoins au regard du processus décisionnel existant (Processus de Séville) et conclure sur le positionnement de cette thèse par rapport à ce dernier. Dans une seconde partie, une revue de littérature scientifique permet d'identifier les méthodes de détermination et d'application des MTD existantes et de les positionner face aux besoins réglementaires identifiés. En dernier lieu, la question de recherche, les hypothèses directrices et les postulats présentant le cadre du travail de thèse seront énoncés.

Chapitre 1 : Contexte réglementaire

Ce premier chapitre présente l'historique réglementaire relatif aux MTD ainsi que les incidences de son application aux échelles européenne, nationale et locale. Dans un second temps, une description du Processus de Séville mis en œuvre pour la détermination de ces MTD et des niveaux de performance environnementale associés (BATAEPL – *Best Available Technique Environmental Performance Levels*) est présentée. Enfin, les implications pour l'industrie et les États membres sont déduites et servent à éclairer les besoins méthodologiques pour la détermination et l'application des MTD.

1. La réglementation des émissions industrielles : du cadre européen à l'application nationale

En 1996, la directive 96/61/CE dite IPPC (*Integrated Pollution Prevention and Control*) donne le cadre européen de la réglementation sur les émissions industrielles et introduit le concept de MTD. En 2010, elle est remplacée par la directive relative aux émissions industrielles 2010/75/UE dite IED (Commission Européenne, 2010). Son périmètre englobe celui de l'IPPC et couvre quatre autres secteurs (grandes installations de combustion, installations d'incinération de déchets, installations et activités des solvants organiques et installations produisant du dioxyde de titane), soit environ 50 000 installations en Europe. L'objectif final de l'IED est de « prévenir, réduire et, dans la mesure du possible, éliminer la pollution due aux activités industrielles, conformément au principe du "pollueur payeur" et au principe de prévention de la pollution ». Elle se concentre donc sur la réduction des émissions de substances polluantes tout en considérant des aspects environnementaux plus larges pour les six catégories d'activités industrielles énoncées dans son annexe I : les industries d'activités énergétiques ; la production et transformation des métaux ; les industries minérales ; les industries chimiques ; la gestion des déchets ; et plusieurs autres activités (installations papetières, teintureries, abattoirs, installations de transformation de matières animales, équarrissage, élevages intensifs de volailles et de porcs, traitement de surface et fabrication de carbone).

1.1. Principes de la directive

Dans la même lignée que la directive IPPC, l'IED a pour finalité de prévenir les émissions industrielles et les consommations ou, si la prévention est impossible, à les réduire. De plus, elle insiste sur l'approche intégrée qui repose notamment sur la prise en considération de

l'environnement dans son ensemble (émissions dans l'air et l'eau, déchets, bruit, odeurs, risques, etc.). En outre, elle repose sur l'obligation de l'obtention de permis d'autorisation à exploiter pour les installations concernées. Pour y parvenir, quatre grands principes (Cikankowitz et Laforest, 2013) sont énoncés : (1) l'approche intégrée pour la délivrance de l'autorisation d'exploiter et la prise en considération des effets croisés (synergies entre les différents impacts) ; (2) l'application des MTD décrites dans des documents de référence (BREF – *Best Available Technique REFerence document*) dont le processus de création repose sur un échange d'informations entre parties prenantes lors du Processus de Séville. La directive IED exige également des améliorations continues en termes de prévention ou de réduction des émissions des installations sur l'environnement. Pour cela, elle impose la mise en œuvre de techniques ayant des performances équivalentes aux MTD. Pour soutenir cette approche, les VLE doivent être fondées sur les niveaux d'émissions associés aux MTD dites BATAEL proposées dans les « conclusions sur les meilleures techniques disponibles » ; (3) la flexibilité pour la considération des conditions locales ou de la configuration de l'installation (techniques, économiques et géographiques) dans l'applicabilité des MTD. De plus, la sensibilité du milieu sur lequel l'entreprise est implantée doit être prise en compte dans la détermination des conditions d'autorisation ; (4) la participation du public :« La directive garantit que le public doit être informé des conséquences d'une autorisation à exploiter en ayant accès aux demandes d'autorisation pour donner son opinion, aux autorisations et aux résultats de surveillances des rejets dont notamment le registre européen des rejets et transferts de polluants (E-PRTR). » (Laforest et Gaucher, 2015).

1.2. Les MTD : Qui sont-elles ? D'où viennent-elles ? Où les trouver ?

1.2.1. Qui sont-elles ?

Le concept de MTD a initialement été défini dans la directive IPPC comme « le stade de développement le plus efficace et avancé des activités et de leurs modes d'exploitation, démontrant l'aptitude pratique de techniques particulières à constituer, en principe, la base des valeurs limites d'émission visant à éviter et, lorsque cela s'avère impossible, à réduire de manière générale les émissions et l'impact sur l'environnement dans son ensemble » (Communauté Européenne, 1996).

La définition du concept de MTD défini dans l'article 3 de l'IED diffère seulement de l'IPPC par l'ajout « d'autres conditions d'autorisation » afin d'« éviter et, lorsque cela s'avère impossible, de réduire les émissions et l'impact sur l'environnement dans son ensemble ».

En particulier, les trois termes composant l'expression MTD sont précisés ainsi :

- « **Techniques** » : englobe « aussi bien les techniques employées que la manière dont l'installation est conçue, construite, entretenue, exploitée et mise à l'arrêt » ; Ainsi, la notion de technique couvre les procédés de production, des installations de traitement des rejets, la substitution de produits chimiques ou encore des dispositions organisationnelles ;

- « **Disponibles** » : les techniques « mises au point sur une échelle permettant de les appliquer dans le contexte du secteur industriel concerné, dans des conditions économiquement et techniquement viables, en prenant en considération les coûts et les avantages, que ces techniques soient utilisées ou produites ou non sur le territoire de l'État membre intéressé, pour autant que l'exploitant concerné puisse y avoir accès dans des conditions raisonnables » ;

- « **Meilleures** » : c'est-à-dire les techniques « les plus efficaces pour atteindre un niveau général élevé de protection de l'environnement dans son ensemble ». La détermination des meilleures techniques disponibles repose sur douze considérations énumérées à l'annexe IV de l'IPPC.

Ces douze considérations, (Tableau 1) permettaient de qualifier une technique comme « meilleure ». Néanmoins, la portée de ces critères et leur transposition difficile en paramètres et indicateurs ne permettaient pas de déterminer les MTD de manière systématique (Cikankowitz, 2008 ; De Chefdebien, 2006 ; Laforest, 2008). Il est à noter que ces critères sont à prendre en considération en particulier lorsque les documents de références sur les MTD dits BREF (cf. ci-après) ne couvrent pas ou partiellement les installations d'un exploitant ayant l'obligation de se positionner par rapport aux MTD (Laforest et Gaucher, 2015). Ces douze considérations de l'IPPC sont reprises sous la forme de douze critères donnés dans l'annexe III de l'IED (Tableau 1) conservant les mêmes difficultés d'usage pour sélectionner ou évaluer des MTD (Laforest, 2014).

Tableau 1 : Les 12 critères pour la sélection des Meilleures Techniques Disponibles tels qu'énoncés dans l'annexe III de l'IED.

1	Utilisation de techniques produisant peu de déchets.
2	Utilisation de substances moins dangereuses.
3	Développement des techniques de récupération et de recyclage des substances émises et utilisées dans le procédé et des déchets, le cas échéant.
4	Procédés, équipements ou modes d'exploitation comparables qui ont été expérimentés avec succès à une échelle industrielle.
5	Progrès techniques et évolution des connaissances scientifiques.
6	Nature, effets et volume des émissions concernées.
7	Dates de mise en service des installations nouvelles ou existantes.
8	Délai nécessaire à la mise en place de la meilleure technique disponible.
9	Consommation et nature des matières premières (y compris l'eau) utilisées dans le procédé et efficacité énergétique.
10	Nécessité de prévenir ou de réduire à un minimum l'impact global des émissions sur l'environnement et des risques qui en résultent pour ce dernier.
11	Nécessité de prévenir les accidents et d'en réduire les conséquences sur l'environnement.
12	Informations publiées par des organisations internationales publiques.

En outre, pour être considérées comme MTD, les techniques candidates doivent être mises en œuvre à l'échelle industrielle pour le secteur concerné et accessible sur le marché à un coût économiquement acceptable.

1.2.2. D'où viennent-elles ? Où les trouver ?

Partant de cette définition, la Commission Européenne a mis en place un processus d'échange d'information autour des MTD (en réponse à l'article 13 de l'IPPC) qui s'est traduit par le Processus de Séville. Ce processus vise à fournir des documents de référence (BREF) décrivant les MTD applicables dans chaque secteur industriel ainsi que les BATAEPL. Leur contenu a évolué au cours de la période d'application de l'IPPC jusqu'à la version actuellement appliquée.

Depuis 2012, le Processus de Séville est formellement décrit dans une décision d'exécution (Commission Européenne, 2012). En particulier, les « BREF » ou « Best available techniques REFerence documents » sont le résultat de l'échange d'informations entre les parties prenantes afin d'élaborer, de réviser et, le cas échéant, de mettre à jour les MTD. Cet échange est assuré par le Bureau Européen de l'IPPC (EIPPCB) basé à Séville. Ces documents ont pris une importance croissante au fil des ans et sont devenus réglementairement les documents de référence dont découlent les conclusions sur les MTD que les autorités compétentes doivent prendre en compte

pour définir les conditions d'autorisation. Les BREF sont généralement structurés en six chapitres (1) Informations générales sur le secteur concerné ; (2) Techniques et procédés appliqués ; (3) Niveaux de consommations et d'émissions du moment ; (4) Techniques à prendre en considération pour la détermination des MTD ; (5) Conclusions sur les MTD ; (6) Techniques émergentes.

Le chapitre-clé pour les exploitants et les autorités environnementales est le cinquième chapitre (Conclusions sur les MTD) qui est le résultat d'un consensus pour la définition d'une liste de MTD. Les MTD qui y sont décrites ne sont ni exhaustives, ni prescriptives mais fournissent un support pour évaluer les performances des installations.

1.3. Transposition de l'IPPC dans la réglementation française et retour d'expérience

Préalablement à l'IPPC de 1996, la France disposait d'une législation et réglementation environnementales reposant sur une approche intégrée et préventive. Toutefois, les transpositions successives des directives IPPC et IED dans le droit national ont impliqué plusieurs changements majeurs. Ces derniers sont brièvement décrits ci-après.

Le premier texte majeur du droit de l'Environnement français est la loi n°76-663 du 19 juillet 1976 (Assemblée Nationale, 1976) sur les installations classées pour la protection de l'environnement (ICPE). Elle constitue la base juridique de l'environnement industriel en France. Dans la réglementation ICPE, l'exploitant a la charge de démontrer que les risques inhérents à son entreprise sont maîtrisés et de proposer à l'autorité des moyens de les réduire. Les activités et substances mises en jeu dans les installations sont classées selon une nomenclature qui n'a cessé d'évoluer afin de prendre en compte différents aspects environnementaux (rejets et prélèvements d'eau, émissions dans l'air, bruit), tout en tendant progressivement vers une approche de plus en plus préventive. Cette réglementation propose plusieurs régimes dont le régime des installations à autorisation qui doivent obtenir un permis d'autorisation à exploiter.

Le second texte présentant une importance dans notre périmètre de recherche est l'arrêté ministériel du 2 février 1998 relatif aux prélèvements et consommations d'eau (Assemblée Nationale, 1998) dont le champ d'application couvre toutes les activités sauf celles qui en sont explicitement exclues telles que, notamment, les carrières, installations de traitement de surface ou encore les établissements d'élevage. Cet arrêté définit, de plus, des valeurs limites d'émission en concentration ou en fonction du flux horaire de polluants.

La réglementation française préexistante portait donc déjà les notions d'approche intégrée de l'IPPC, la démarche d'autorisation à exploiter, la considération du niveau installation, la preuve à la charge de l'exploitant, les valeurs limites d'émission (VLE). La principale nouveauté pour les entreprises françaises a été la réalisation d'un bilan de fonctionnement décennal dont le contenu était décrit dans l'arrêté du 29 juin 2004 (Assemblée Nationale, 2004) et incluant une comparaison des techniques mises en œuvre avec les MTD. Il contient :

- une analyse du fonctionnement de l'installation au cours de la période décennale passée (conformité de l'installation vis-à-vis de la réglementation en vigueur, et par rapport à des valeurs limites d'émission) ;
- une synthèse de la surveillance des émissions, du fonctionnement de l'installation et de ses effets sur l'environnement ;
- l'évolution des flux des principaux polluants et l'évolution de la gestion des déchets ; un résumé des accidents et incidents qui ont pu porter atteinte aux intérêts mentionnés à l'article L. 511-1 du code de l'environnement (Code de l'Environnement, 2011), les investissements en matière de surveillance, de prévention et de réduction des pollutions ;
- une analyse des effets de l'installation sur l'environnement et la santé ;
- une analyse des performances des moyens de prévention et de réduction des pollutions par rapport aux performances des MTD ;
- les mesures envisagées par l'exploitant sur la base des MTD pour supprimer, limiter et compenser les inconvénients de l'installation ainsi que l'estimation des dépenses correspondantes ;
- les mesures envisagées pour placer le site dans un état tel qu'il ne puisse porter atteinte aux intérêts mentionnés à l'article L. 511-1 du code de l'environnement en cas de cessation définitive des activités.

Ainsi, la directive IPPC a fait émerger une série de textes législatifs pour la transposer dans la réglementation nationale (Figure 1).

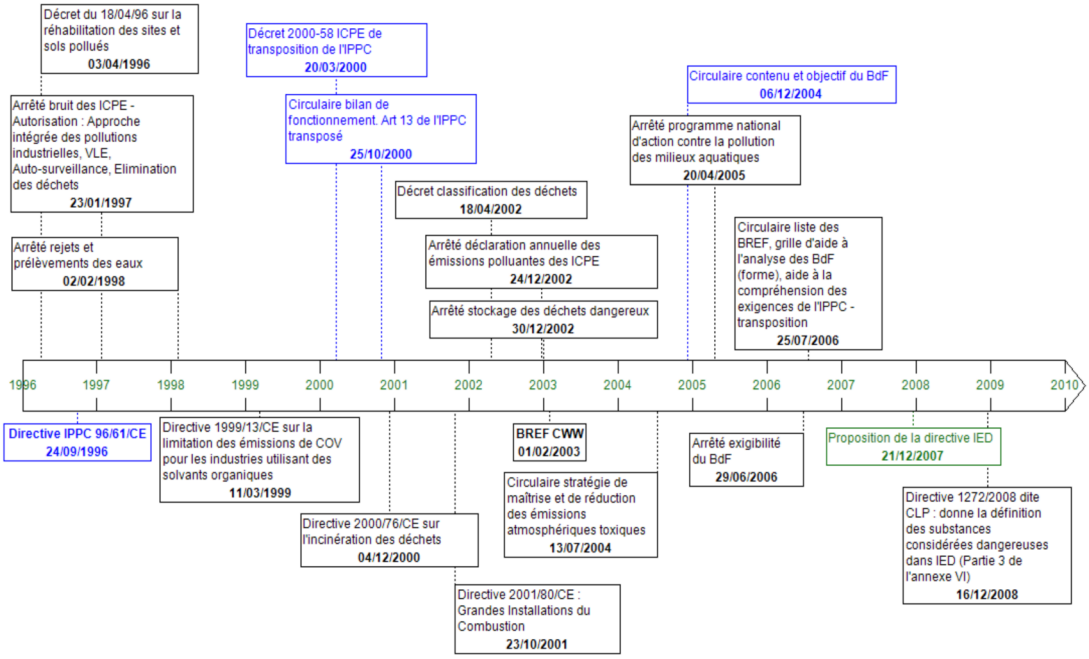


Figure 1 : Chronologie des principaux textes relatifs à l'IPPC.

Cependant, suite aux audits réalisés en 2005, la Commission Européenne (2005) a constaté que, dans de nombreux États membres, y compris en France : la mise en œuvre des MTD était insuffisante et que le respect des MTD et BATAEL variait. De plus, peu de moyens de contrôle avaient été mis en œuvre pour un champ d'application de la directive jugé trop restreint, peu clair et un corpus réglementaire trop fragmenté, ce qui rendait la charge administrative excessivement lourde.

La volonté de résolution de ces problèmes a conduit à la proposition de la directive sur les émissions industrielles (IED) dès décembre 2007.

Du fait de l'historique sur la directive IPPC et de sa transposition en droit national dans tous les États-membres dont la France, la transposition de l'IED n'a abordé que les points de différenciation et de renforcement apportés par cette dernière. En France, les premiers textes de transposition de la directive IED dans la réglementation française ont été publiés en 2012. D'abord, l'ordonnance du 5 janvier 2012 (Assemblée Nationale, 2012a) qui précise les installations visées par l'annexe I de la directive ; le recours aux conclusions sur les MTD ; le principe du réexamen périodique ; la consultation du public lors de ce réexamen ; et les conditions de remise en état du site.

Le 2 mai 2013, une série de décrets et d'arrêtés a été publiée afin d'achever cette transposition (Assemblée Nationale, 2013a ; 2013b ; 2013c ; 2013d) : transposition du chapitre II de la directive dans le Code de l'Environnement (exigences applicables aux activités définies dans son annexe I), abrogation de l'arrêté existant sur le bilan de fonctionnement, définition des rubriques 3000 de la nomenclature des ICPE transcrivant l'annexe 1 de la directive et adaptation ou abrogation d'autres dispositions existantes.

Bien qu'elles soient apparentées, les directives IPPC et IED ont impliqués des changements importants dans les entreprises. Notamment, un réexamen périodique est coordonné avec la révision des conclusions sur les MTD de leur activité principale. Ainsi, les États membres disposent d'un délai de quatre années après la publication des conclusions sur les MTD pour mettre leurs installations en conformité.

Le contenu du dossier de réexamen est identique à celui du bilan de fonctionnement qui existait avec la directive IPPC à la différence près que l'analyse de la performance des installations au regard des MTD doit être faite en lien avec les conclusions sur les MTD qui ont désormais une portée juridique. En cela, la directive IED met en place un processus d'amélioration continue des performances environnementales basé sur des cycles d'une décennie.

Néanmoins, l'impact des MTD sur la pollution est encore difficile à observer. Il semble toutefois, dans le cas de l'IPPC, que les effets sur l'environnement soient positifs grâce à l'augmentation des investissements (Testa *et al.*, 2014). En revanche, ce constat est à modérer dans la mesure où l'impact environnemental et économique d'une technique particulière dépend de son contexte d'application (Bachmann et van der Kamp, 2014) et que les bénéfices environnementaux de l'IPPC semblent également varier en fonction des substances considérées (Conti *et al.*, 2015). En effet, Conti *et al.* (2015) montrent que la crise de 2008 a eu plus d'effets sur la réduction des émissions d'oxydes de soufre que la mise en œuvre des MTD. En ce qui concerne l'IED, il est encore trop tôt pour avoir un réel retour d'expérience sur ses effets.

2. Le processus décisionnel de détermination des MTD : le Processus de Séville

Cette section présente d'abord les objectifs du Processus de Séville et ses livrables, ensuite les acteurs impliqués dans sa phase technique, puis les étapes de création ou de révision d'un BREF sont décrits. Enfin, des observations réalisées au cours du Processus de Séville de deux BREF serviront à tirer des conclusions sur les limites et besoins auxquels cette thèse vise à répondre.

2.1. Objectifs

L'objectif du Processus de Séville est de produire des documents de référence au niveau européen décrivant des MTD et leurs niveaux de performances environnementales associées qui serviront aux exploitants d'installations et aux autorités environnementales au niveau local. Ce processus décisionnel repose sur un échange d'informations entre des représentants des États membres, des industriels et des ONG environnementales. Une contrainte temporelle existe puisque la révision des BREF doit avoir lieu, en théorie, tous les huit ou dix ans. À l'heure actuelle, le processus complet dure entre deux et trois ans (Laforest et Gaucher, 2015) avec des ressources disponibles rares chez les acteurs et un besoin d'acquérir de meilleures informations plus rapidement. En cela, les acteurs du Processus de Séville sont demandeurs d'outils leur permettant de satisfaire ces échéances.

Lors du Processus de Séville, l'échange d'information au sein d'un groupe de travail technique (TWG–*Technical Working Group*) est coordonné par le Bureau Européen de l'IPPC (EIPPCB). Les membres du TWG s'assurent de transmettre les informations pertinentes à l'EIPPCB pour déterminer les MTD. Cette collecte d'informations repose sur les parties prenantes nationales (exploitants et autorités) et peut se faire par l'intermédiaire de groupes de travail mis en place optionnellement par les États membres pour discuter de la position nationale et de la manière de contribuer au TWG : les groupes miroirs. En France, ils sont généralement composés de représentants du ministère de l'Environnement, d'industriels, ainsi que d'experts techniques (Figure 2).

Le Processus de Séville peut donc être vu comme un cycle d'amélioration continue d'une décennie impliquant trois échelles géographiques européenne (TWG et EIPPCB), nationale (groupe miroir) et locale (exploitants d'installations et autorités environnementales) dont la finalité est de définir des MTD et BATAEPL décrites dans des BREF.

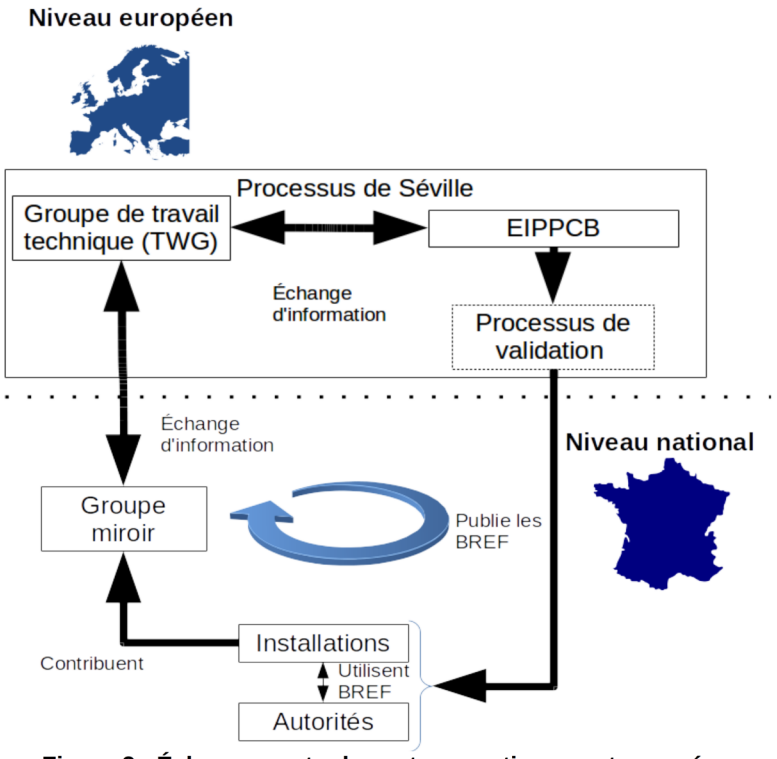


Figure 2 : Échanges entre les acteurs nationaux et européens pendant le Processus de Séville.

Les BREF sont actuellement au nombre de 33 (EIPPCB, 2016) dont 29 BREF dits « verticaux » qui couvrent autant d'activités principales et 4 BREF dits « horizontaux » ou « transversaux ». À ceux-ci sont ajoutés deux autres documents qui apportent des éléments sur le « suivi des émissions » (ROM – Commission Européenne, 2013), et sur « l'estimation des coûts et les effets multi-milieux » (ECM – Commission Européenne, 2006).

Ainsi, avec l'IED, la détermination des MTD est devenue critique pour l'industrie et les États membres dans la mesure où le montant des dépenses pour atteindre les performances des MTD peut être substantiel. Pour assurer le caractère « disponible » et « meilleure » des MTD, un processus d'échange d'information basé sur des données (Processus de Séville) les définit par une recherche de consensus au sein du TWG. La décision d'exécution du 10 février 2012 (Commission Européenne, 2012) décrit ce processus d'échange.

2.2. BATAEL

En plus des MTD, les BREF contiennent des BATAEL. Ces niveaux d'émission associés aux MTD sont des points essentiels des discussions menées à Séville et définissent quantitativement les objectifs que doivent respecter les industriels. Le dépassement de cette valeur est soumis à une demande de dérogation et peuvent s'inscrire dans le principe de flexibilité.

En réalité, deux bornes de BATAEL sont définies : une **borne supérieure** qui représente l'émission haute d'une installation mettant en œuvre les MTD en fonctionnement normal ; et une **borne inférieure** qui correspond à la performance atteignable par les MTD.

Les BATAEL sont inclus dans le concept plus large de niveau de performance environnementale associé aux MTD (BATAEPL) qui englobe également les consommations, émissions pendant « les conditions d'opération anormales » et tout autre indicateur pertinent tel que l'efficacité d'abattement ou le contrôle des procédés (Commission Européenne, 2012, sec. 3.3).

2.3. Les acteurs du Processus de Séville

Le processus de Séville repose sur l'échange d'information entre ses différents acteurs (Roudier, 2013). Le flux d'information commence au niveau de groupes de travail techniques (TWG) et des groupes miroirs nationaux pour arriver à la Commission Européenne. Cette dernière a la tâche de faciliter les échanges d'informations et d'entériner les BREF, en passant par l'EIPPCB qui coordonne les travaux des TWG, le Forum (qui est un premier validateur des projets) et le Comité (Commission Européenne, 2012). Cette section se concentre sur les acteurs de la partie technique du Processus ; les aspects de validation légale du Forum et du Comité sont considérés hors périmètre.

2.3.1. Le Bureau Européen de l'IPPC (EIPPCB)

Le Bureau Européen de l'IPPC est basé à l'IPTS (*Institute for Prospective Technological Studies*) au sein de l'unité Production et Consommation Durable (SPC – *Sustainable Production and Consumption*). Il suit les besoins de révision des BREF pour supprimer les incohérences par rapport à d'autres documents, supprime les informations obsolètes, met à jour et complète les anciennes informations à l'aide de données plus récentes.

L'EIPPCB est également chargé de l'inscription et de la gestion des membres des TWG, de l'organisation et de la présidence des réunions des TWG, ainsi que de la coordination de l'échange

d'informations. Ainsi, il vérifie la complétude et l'exactitude des informations transmises par les membres du TWG et peut demander un complément d'information. La qualité des informations transmises est assurée par :

- la production de questionnaires sectoriels que les États membres transmettent à leur industrie ;
- la vérification d'un niveau de détails et de qualité (âge des données, mesurages, unités, etc.) suffisant pour pouvoir comparer les données entre elles.

L'EIPPCB rassemble les données relatives aux émissions, aux caractéristiques techniques et au contexte de fonctionnement. Les données proviennent des membres des TWG et de la littérature.

2.3.2. Les groupes de travail techniques (TWG)

Un TWG est composé, pour chaque BREF, d'un nombre variable d'experts techniques représentant les États membres, l'Industrie, les ONG œuvrant pour la protection de l'environnement et la Commission. Ses membres sont nommés sur la base de leurs compétences techniques, économiques et environnementales, ou encore de leurs connaissances de la réglementation et du secteur visé.

Leur premier travail consiste à émettre des desiderata justifiés sur le futur BREF en préparation de la réunion de lancement de la révision ou de la création d'un nouveau BREF. Ceux-ci portent en particulier (Commission Européenne, 2012) sur :

- Le champ d'application et la structure du BREF ;
- Les MTD et les niveaux de performance environnementale associés à une MTD qui font défaut, sont dépassés, incomplets ou confus ;
- Le type et format des données qu'il convient de collecter pour alimenter la révision ;
- Les nouvelles « techniques à prendre en considération pour la détermination des MTD » ou « techniques émergentes ».

Si nécessaire, des sous-groupes peuvent être constitués pour examiner des thèmes spécifiques dans la suite du processus. Leur travail consiste ensuite à apporter et argumenter des informations mises en commun et du contenu du BREF.

2.3.3. Les groupes miroirs

Le rôle des groupes miroirs est de transférer des informations du TWG vers les associations professionnelles nationales et inversement. Tous les acteurs concernés par un BREF donné peuvent y participer (entreprises, syndicats professionnels, etc.). Leur mise en place est optionnelle et le nombre de participants à un groupe fluctue selon l'implication et la structuration du secteur. En France, ils sont pilotés par le Ministère de l'Environnement (INERIS, 2014).

2.3.4. Les acteurs locaux

Deux acteurs travaillent à l'échelle locale : les autorités environnementales et les exploitants d'installations. En effet, les autorités environnementales ont la charge de vérifier que les informations fournies par les exploitants d'installation justifient qu'ils mettent en œuvre les MTD et, si ce n'est pas le cas, les arguments technico-économiques avancés lors du réexamen de chaque installation d'un secteur.

2.4. Étapes du Processus de Séville et pratiques actuelles

Les étapes principales sont présentées dans l'annexe 2 de la décision d'exécution de 2012 (Commission Européenne, 2012). Le cas d'une révision de BREF, est illustré à la Figure 3 et chacune des étapes est décrite dans cette section. Dans la mesure où le Processus de Séville est un processus itératif, le TWG peut répéter chaque étape si nécessaire.

De plus, au cours de la thèse, plusieurs occasions d'observer l'application du Processus de Séville ont permis de constater des limites des pratiques actuelles. En ce qui concerne les 4quatre premières étapes du Processus de Séville, la révision du BREF agroalimentaire qui a débuté en 2014, en particulier le secteur laitier, a permis de suivre le travail du groupe miroir français et du TWG jusqu'à la compilation des données européennes. Pour les étapes 5 et 6, fin 2013, la réunion finale du BREF CWW a permis d'assister aux dernières discussions avant rédaction du projet final de BREF.

Tout d'abord, dans le cas du secteur agroalimentaire, la réactivation du TWG a eu lieu en mars 2014, la réunion de lancement en octobre de la même année et la collecte des données européennes au début de l'année 2015. Les enseignements tirés des pratiques observées sont indiqués pour chaque étape. Cette section regroupe les observations réalisées lors de la réunion de lancement, une réunion de travail des industriels laitiers et des réunions du groupe miroir réunissant les industriels par l'intermédiaire de leurs associations professionnelles, du Ministère de l'Environnement, de l'INERIS, des DREAL, Agences de l'eau et Ministère de l'Agriculture.

Deuxièmement, la réunion finale, dans le cadre du BREF sur les techniques de traitement des gaz et eaux usées communes à l'industrie chimique (CWW) dont le processus de révision du BREF a duré 9 ans, de 2007 à la publication des conclusions sur les MTD dans le Journal Officiel en juin 2016 a été observée.

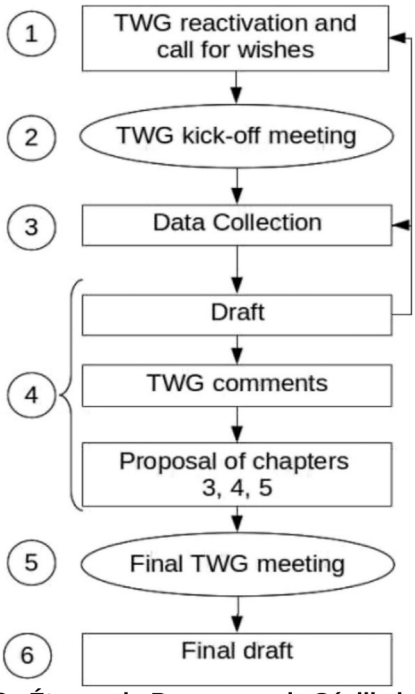


Figure 3 : Étapes du Processus de Séville lors de la révision d'un BREF, adapté de Commission Européenne (2012).

2.4.1. Étape 1 : Réactivation du TWG et appel à desiderata

Cette première étape vise à collecter les souhaits des membres du TWG pour le futur document en ce qui concerne son périmètre ; les indicateurs environnementaux (consommations ou émissions) à cibler aussi nommés KEI pour « *key environmental issues* » ; la structure du BREF, c'est-à-dire les modifications nécessaires dans le chapitre décrivant les MTD incluant les BATAEL ; et la nature des données à collecter.

Dans le cas de l'industrie agroalimentaire, après analyse du BREF existant et réalisation d'un état des lieux de l'industrie française, des données relatives aux différentes activités de l'industrie agroalimentaire ont été collectées pour définir une typologie dans chacune des filières (viande, produits laitiers, etc.) et identifier des enjeux et spécificités nationales (EIPPCB, 2015). Le recensement des installations soumises à l'IED a également été effectué et des échanges existants

avec les associations industrielles européennes menée sont été effectués pour éviter les informations redondantes.

L'objectif final de cette étape était de transmettre à Séville une liste de vœux sur la structure et le contenu de la future version du BREF. Par exemple, pour l'industrie laitière française, les produits fabriqués constituant le facteur qui, selon l'avis des experts, influent sur la performance environnementale, une décomposition du BREF par produit devait être demandée. De plus, étant donné que la plupart des sites français sont multi-produits, à la différence d'autres États membres, cette spécificité a été proposée.

Un autre objectif était d'anticiper les enjeux-clés (KEI) définis au cours de la réunion de lancement de la révision du BREF et les typologies d'installations. Par exemple, les effluents peuvent être rejetés après traitement sur site dans l'environnement (rejet direct), ou dans le réseau pour traitement en station d'épuration (rejet indirect). La prise en compte de cette distinction implique des niveaux d'émission dans l'eau en sortie de sites différents (généralement plus élevés en rejets indirects) et des techniques différentes (traitement sur site plus développé pour les rejets directs). Par conséquent, une réflexion sur les caractéristiques des sites de référence dont la liste a été transmise à Séville après la réunion de lancement a été engagée sur la base des expertises présentes. De plus, des visites de sites industriels furent proposées aux représentants nationaux pour les aider à comprendre les particularités énoncées. Cela a aidé à justifier de la pertinence des installations de référence proposées.

À ce niveau, plusieurs remarques ont été émises et démontrent la subjectivité des pratiques actuelles. Elles ont porté sur :

1. la typologie métier qui a été choisie bien que probablement peu adaptée à l'objectif final ;
2. les différences de traitement entre l'Industrie et l'État membre sur les niveaux de performance des sites de référence.

2.4.2. Étape 2 : Réunion de lancement

Une première réunion vise à définir l'agenda pour la révision sur la base des souhaits formulés à la fin de l'étape précédente. Les points à étudier et les données à collecter sur des installations existantes sont aussi définis (Commission Européenne, 2012, sections 4.6.2.2 et 5.5). Ces éléments vont constituer un questionnaire et sont susceptibles de devenir des critères d'évaluation des techniques. Ils font donc l'objet de choix stratégiques, en particulier sur les enjeux environnementaux-clés (KEI) qui vont servir à évaluer les performances environnementales.

Suite à cette réunion, les installations où vont être collectées ces données, et qui vont servir de référence, sont sélectionnées par l'industrie puis validées ou non par les États membres. La liste de ces installations de référence est partagée avec l'ensemble du TWG avant la collecte des données à proprement parler.

Cela a été réalisé en 2014 pour l'industrie agroalimentaire. Le premier sujet abordé a été le périmètre du BREF qui a fait l'objet de problèmes d'interprétation en raison de la définition des activités par matières entrantes dans le cas où de multiples produits entrent dans les installations. Par exemple, les installations produisant des yaourts aux fruits doivent-elles être considérées comme des installations laitières ou de transformation de fruits ? La réponse de l'EIPPCB a été que ce problème d'interprétation légale n'était pas un sujet technique à traiter dans le TWG. Par conséquent, le périmètre du BREF est supposé être clairement identifié au préalable. La structure du BREF a également été débattue entre la recherche de sections communes à toutes les activités agroalimentaires ou uniquement une structure par activité. Par conséquent, malgré une volonté de généricité et d'intégration, aucune approche globale ne permet de vérifier si les émissions sont semblables malgré des métiers différents.

Dans un second temps, les KEI ont été identifiés comme étant les aspects environnementaux sur lesquels doivent agir les MTD. Un consensus a émergé sur les émissions dans l'air et la plupart des émissions dans l'eau, ainsi que sur les consommations. Sur d'autres aspects qui ne sont pas considérés comme des KEI, par exemple les déchets, la conclusion de l'EIPPCB a été que le TWG collectera néanmoins des données à leur sujet. Ainsi le questionnaire couvrait les consommations (eau, énergie, techniques de réduction des consommations) et les émissions (dans l'eau, l'air et les déchets). La manière d'exprimer les valeurs a été également débattue avec les partisans d'une approche réglementaire basée sur une expression des rejets dans l'eau en concentration et les partisans de charges spécifiques exprimées en flux (généralement une masse) de produit fabriqué ou consommé par unité de temps) de produit fabriqué ou consommé. Dans le secteur laitier, les charges spécifiques ont été sélectionnées et rapportées à une unité de lait transformé (« *if possible, loads in mass / unit of processed milk* ») (Commission Européenne, 2015).

Les KEI sont donc restreints dès le début sur avis d'experts et la quantité de matière entrante ou sortante caractérise la taille des installations pour comparaison de l'efficacité. Néanmoins, la finalité étant réglementaire, l'expression de valeurs sur le même format est préférée par certains membres du TWG. De plus, la priorité est donnée aux émissions susceptibles de faire l'objet de BATAEL et couramment mesurées par les industriels. La définition du périmètre des installations peut poser problème et est due à des interprétations divergentes de l'IED.

Enfin, le panel d'installations de référence a été défini comme devant porter sur les deux types de rejets (directs et indirects, incluant l'épandage) ; des tailles variées ; des sites fabriquant un ou plusieurs produits ; d'âge différents. Le nombre d'installations faisant l'objet d'une collecte a été discuté a posteriori. Par conséquent, il est supposé que les experts techniques savent ce qui est représentatif. Des interrogations quant à la signification de ce qu'est une installation représentative ont été émises. Au cours de cette thèse plusieurs interprétations de ces « good performers » ont été rencontrées (Figure 4) :

- pour l'Union Européenne, les good-performers sont les meilleurs sites en termes de bas niveaux de consommations et d'émissions atteignables, pris individuellement ;
- pour les représentants de la France dans le groupe de travail technique du Processus de Séville, ils correspondent aux sites IED conformes à la réglementation française et approuvés par la DREAL ;
- pour les représentants de l'Industrie, tel qu'observé lors de la réunion de sélection des sites de l'industrie laitière, ce sont les sites exerçant une activité, appliquant des techniques ou présentant une particularité ;
- des sites moyens ont également été évoqués.

La multiplicité des interprétations pose problème pour harmoniser le processus entre ses acteurs et nécessite de se baser sur une définition commune à tous les acteurs du Processus de Séville.

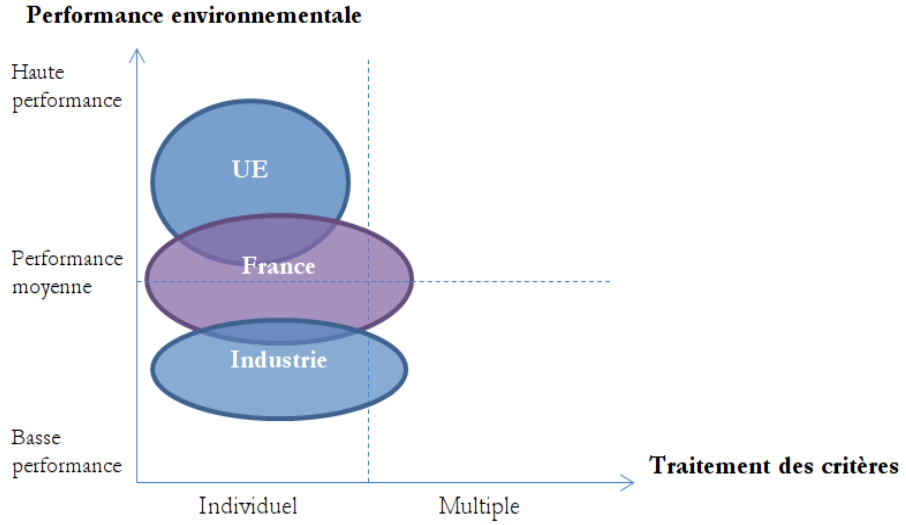


Figure 4 : Positionnement des interprétations du concept d'installation de référence.

De plus, dans les questionnaires, certaines données comme les consommations d'énergie ou les volumes produits étaient confidentielles. Par conséquent, en comptant les données non renseignées pour d'autres raisons, les données effectivement disponibles étaient parcellaires et les méthodes de mesurage variaient. Pourtant, ce sont ces informations qui vont servir à définir les BATAEPL sans qu'il n'y ait de moyen formel de fixer la quantité de données nécessaire pour atteindre une représentativité suffisante autre que le jugement d'experts. Par conséquent, ces derniers ont besoin de justifier tous les sites proposés pour montrer qu'ils ne voient pas de sites supplémentaires à ajouter ou de sites inutiles.

Il y a donc un manque de transparence, tant au niveau du traitement multicritère des données collectées qu'au niveau du traitement des données incomplètes. La nature des installations de référence était également variable en fonction des acteurs.

2.4.3. Étape 3 : Collecte de données

Cette collecte inclut les consommations et émissions des installations de référence, des indications de performance plus larges, les techniques utilisées, des éléments sur la production, des informations sur le contexte local et économique dans lequel l'installation ou la technique opère. Cette étape est essentielle pour définir des BATAEL applicables et donner de la consistance au choix des MTD déduites des performances d'installations réelles.

Pour des secteurs avec un grand nombre d'installations (par exemple, plus de 2 000 dans l'agroalimentaire), l'exhaustivité n'est pas recherchée et une sélection de quelques représentantes est nécessaire. Le choix de ces installations de référence est donc crucial pour la réussite du Processus de Séville. Pourtant, les modes de sélection diffèrent entre les États membres. En particulier, le problème de l'échantillonnage et de la vérification des données collectées à l'aide d'outils statistiques est discuté par Carretero *et al.* (2016). Leur conclusion est qu'une approche statistique permet de réduire la subjectivité des jugements d'experts dans le Processus de Séville. Toutefois, ils considèrent les différentes consommations et émissions séparément, sans recourir à une approche multicritère. Ces différences de modes de sélection des installations de référence sont détaillées dans le §2.5.1.2.

Dans le cadre de la révision du BREF agroalimentaire, la collecte des données s'est axé sur des installations existantes sélectionnées et constitue l'une des étapes cruciales du Processus de Séville. En France, le remplissage des questionnaires par les industriels se fait en coopération avec l'État membre qui les valide ensuite.

Dans le cas de l'activité laitière, la sélection par l'industrie a été effectuée suite à une collecte de données sur les installations soumises à l'IED. Ensuite, un « *mapping* » a permis de synthétiser, par famille de produit (les sites multi-produits étant une catégorie à part entière), les consommations et émissions des sites par rapport à la moyenne du secteur (+ ou – un écart-type). Chaque variable était présentée brute ou normalisée par le nombre de litres de lait équivalent déclaré pour chacun des sites. La question de la normalisation a été posée dans la mesure où la réglementation française propose une pondération de chaque produit pour calculer un équivalent litre de lait (AIDA, 2016). Pour autant, cette approche reste nationale et non reconnue par les autres États membres. Dans un second temps, une analyse plus fine en groupant les sites selon leur produit majoritairement fabriqué (famille) a été réalisée.

Au regard de ce *mapping* et classement, l'expertise des membres a servi à identifier les enjeux et spécificités techniques à retenir. L'âge des sites a été considéré comme une variable pertinente pour montrer la difficulté de modification des techniques sur des sites anciens. À partir de ces éléments, des profils de sites à sélectionner ont été déterminés. Ainsi, 12 catégories basées sur le principal produit fabriqué ont été définies. Le nombre de sites de référence a été prédéfini à 15, soit un nombre équivalent à l'Allemagne et correspondant à 10 % des sites français soumis à l'IED. Pour chacun des profils identifiés, une situation de base a été choisie puis les autres spécificités ont été discutées pour choisir précisément celles pour lesquelles il fallait trouver un représentant. Toutefois, la question de ce qu'est un site représentatif a été posée sans trouver de réponse.

Après validation des sites par le Ministère de l'Environnement et l'INERIS en trois étapes (1–Identification des enjeux que les sites choisis permettent de défendre ; 2–Vérification auprès des DREAL que les sites sont au niveau des MTD ; 3–Exclusion de sites quand le même enjeu est présent plusieurs fois), 13 sites laitiers ont été retenus pour la collecte des données sur les installations de référence et ont constitué les installations de référence françaises déterminées sur jugement d'experts.

Au niveau européen, la collecte des données d'installations agroalimentaires a concerné 353 installations dont 106 installations laitières. Le retour d'un webinaire ayant eu lieu le 2 juin 2016 a mis en exergue que les données étaient parfois inconsistantes ou incomplètes. Elles constituent pourtant les fondations sur lesquelles reposent les étapes suivantes du Processus de Séville. Ce traitement a été réalisé sur tableur, sans logiciel spécifique, au regard des KEI et selon les techniques identifiées comme possibles MTD, ainsi que sur la base des données technico-économiques disponibles.

L'observation des pratiques actuelles pour le BREF FDM a montré que les données de consommations et d'émissions et de consommation étaient essentielles au Processus de Séville pour caractériser la performance environnementale. De plus, la manière de prendre en compte la taille des installations est problématique avec des approches par produit entrant ou sortant, avec pondération ou non. Dans la mesure où des techniques assurant un haut niveau de protection de l'environnement sont cherchées, il semble qu'une certaine efficacité des installations devrait être considérée en ramenant les données sur un flux de produit utilisé ou consommé, ce que nous appellerons dans la suite de ce travail « normalisation ».

Ainsi, dans les pratiques observées, le jugement d'experts est central et présente l'inconvénient qu'il y a souvent un manque de recul pour ne sélectionner que l'essentiel et ne pas se focaliser sur des points d'exceptions. En revanche, certains de ces points d'exception sont peut-être significatifs et pourraient être justifiés par une analyse objective des données disponibles. Par ailleurs, l'approche choisie est monocritère dans la mesure où chaque émission ou consommation est étudiée séparément. Ce n'est que dans un second temps qu'il y a recherche des installations regroupant plusieurs enjeux et également la contrainte que toutes les activités doivent être représentées. Autrement dit, il y a un cycle entre analyse statistique et discussion avec les experts qui se nourriront mutuellement mais sans traiter formellement les informations dans leur ensemble.

Postulat 1 : Tous les sites échantillonnés sont viables économiquement et techniquement, ainsi que conformes à l'IED.

2.4.4. Étape 4 : Projets formels de BREF

L'EIPPCB utilise les informations collectées pour rédiger une première version de BREF. La discussion pour l'approbation de ce document par les membres du TWG concerne principalement l'applicabilité d'une technique, sa viabilité et son efficacité, en se basant sur des informations techniques et économiques. Si un consensus entre les experts est atteint sur le bénéfice environnemental d'une technique et de son niveau de performance globale, elle est considérée comme MTD.

Cette étape n'a pas été observée pour le secteur agroalimentaire car prévue pour fin 2016. En revanche, dans le cas du BREF CWW, la version projet a servi de base de discussion pour la réunion finale. À la diffusion du *draft*, les membres du TWG ont émis des commentaires sur son contenu. En revanche, les MTD et BATAEL de son chapitre de conclusion sont plus spécifiquement traités lors de la réunion finale qui suit cette étape.

Postulat 2 : La caractérisation et la sélection des sites doivent être guidées par le niveau de performance environnementale dans son ensemble.

2.4.5. Étape 5 : Réunion finale

Cette discussion finale se concentre sur le chapitre « Conclusions sur les MTD », c'est-à-dire sur la description des MTD et des BATAEL. Pendant cette réunion, chacune des MTD proposées est passée en revue jusqu'à ce qu'un consensus soit obtenu en se basant sur les consommations et les émissions des sites couverts par les questionnaires. Chaque BATAEL est déterminée en accord avec les performances d'installations réelles.

La réunion finale du BREF CWW a duré quatre jours en décembre 2013 et s'est déroulée à Séville et a rassemblé 42 participants représentants 15 institutions (1 association professionnelle, 13 États membres et la Commission Européenne). L'objectif de la réunion a été de définir les BATAEL et de finaliser les MTD pour les stations de traitement des effluents partagées de l'industrie chimique. Ainsi, la mise en place d'un système de management environnemental, d'une politique de suivi des émissions ou de stratégies de traitement des eaux usées sont des MTD utilisables par l'Industrie.

La réunion a commencé par une synthèse des données collectées et une indication de la représentativité des questionnaires reçus. La représentativité a été exprimée comme étant la part du chiffre d'affaires de l'industrie chimique par rapport à la part de questionnaires de chaque État membre, ce qui a permis de visualiser les éventuelles sur – ou sous-représentations.

Par la suite, les sujets abordés ont concerné ceux qui avaient été proposés par les membres du TWG et synthétisés dans un document de travail appelé *background paper*. À ce stade, les exceptions dans les BATAEL ont été décidées sur avis d'experts à partir des données disponibles et sans possibilité d'ajouts importants. Par ailleurs, il est à noter qu'à ce stade, un KEI peut encore être exclu s'il est démontré qu'il y a un manque de données pour le prendre en compte.

La réunion a continué par un tour de table passant en revue les conclusions sur les MTD et au cours duquel les discussions ont porté essentiellement sur des problèmes de formulation et la volonté des membres du TWG de fixer des BATAEL ou non sur chaque KEI. Des inquiétudes quant à la représentativité des données ont été émises mais aussitôt écartées du fait de l'acceptation de ces données au préalable.

Par conséquent, des données acceptées par le TWG sont considérées comme représentatives. La question de la représentativité est donc essentielle et doit être traitée le plus tôt possible sans toutefois disposer de définition formelle de la représentativité, ce qui entraîne des sélections variables et peu transparentes. De même, il n'existe aucun critère de sélection pour le nombre d'installations de référence nécessaires si ce n'est une variété d'activités, de tailles et de complexités. Par ailleurs, chaque KEI est analysé individuellement et, comme l'a fait remarquer un membre du TWG, l'interdépendance des paramètres n'est pas prise en compte. Cela pose d'autant plus problème que le nombre d'installations capables de respecter l'ensemble des BATAEL est incertain.

Il semble donc qu'il serait utile d'expliciter la manière dont les choix sont effectués à chaque étape.

2.4.6. Étape 6 : Projet formel final

Cette version finale du BREF inclut les informations et remarques pertinentes issues de la réunion finale. Elle est envoyée aux membres du TWG pour validation via une succession de phases d'analyse de données et de prises de décision. Le travail technique s'achève ici et la validation par les États membres et la Commission Européenne commence.

Le projet final de BREF CWW a été publié en juillet 2014 (EIPPCB, 2016) pour une publication dans le Journal Officiel en juin 2016. La durée entre ces deux événements a constitué la phase de validation politique et juridique du BREF.

2.5. Conclusions sur le Processus de Séville

Le Processus de Séville vise à réduire les émissions de polluants en Europe et englobe donc trois échelles géographiques d'application des MTD. Celles-ci sont interconnectées dans la mesure où la détermination et l'application des MTD les mettent toutes les trois en jeu : les niveaux européens et nationaux concernent un secteur industriel et le niveau local une installation. Par conséquent, une grande quantité d'information doit être traitée. Néanmoins, le processus actuel les considère séparément en analysant des données principalement issues d'installations. Dans la pratique, des sites de référence sont utilisés pour réduire le nombre d'installations étudiées mais la manière dont ils sont sélectionnés est variable et peu transparente.

Le Processus de Séville, bien que reposant sur la recherche de consensus a montré qu'il y avait des différences d'interprétation et d'application des principes de l'IED. Ces divergences entraînent

des difficultés à justifier les décisions prises à chaque étape et à garantir l'effet des MTD sur la pollution en Europe.

3. Les MTD dans les secteurs non-IED

Les MTD concernent également des industries hors du périmètre de l'IED qui ne disposent et ne dépendent pas du cadre du Processus de Séville. À titre d'exemple, le secteur des installations nucléaires de base (INB) peut être cité (OSPAR, 2015). L'application des MTD pour les INB concerne d'une part la conception des réacteurs nucléaires et, d'autre part, les contraintes réglementaires liées à leur exploitation.

3.1. Définition d'une INB

Les INB sont définies en France par l'article 3 de l'ordonnance n°2012-6 du 5 janvier 2012 (Assemblée Nationale, 2012a). Ce sont les réacteurs nucléaires ; les installations de préparation, d'enrichissement, de fabrication, de traitement ou d'entreposage de combustibles nucléaires ; celles de traitement, d'entreposage ou de stockage de déchets radioactifs ; celles contenant des substances radioactives ou fissiles et répondant à des caractéristiques définies par décret en Conseil d'État ; ainsi que les accélérateurs de particules. Elles font l'objet d'une réglementation spécifique sur leurs éléments nucléaires.

3.2. Principaux textes réglementaires

Depuis 2006, l'autorisation d'exploiter se base sur un dossier instruit par des instances spécifiques au secteur du nucléaire (Assemblée Nationale, 2006). Cette loi donne le cadre légal dans lequel doivent fonctionner les INB civiles. Elle insiste notamment sur la transparence et l'amélioration continue des installations sous la forme d'un réexamen de sûreté de l'exploitation devant se baser sur les meilleures pratiques internationales

Le décret 2007-157 (Assemblée Nationale, 2007) décrit, quant à lui, en particulier la manière de monter un dossier de demande d'autorisation de création d'une INB et le contenu de l'étude d'impact qui doit permettre de justifier du choix d'un procédé par rapport à plusieurs autres. Ce choix s'opère selon la hiérarchie « éviter – réduire – compenser » (ERC). Des dossiers doivent également être fournis aux autorités dans le cas de modifications des installations.

Depuis 2012, l'environnement est considéré au même niveau que les aspects de sûreté par un arrêté abordant tout le cycle de vie d'une INB (Assemblée Nationale, 2012a). Il constitue le

premier texte où les MTD sont explicitement citées au travers des 12 critères de l'IED. De plus, les valeurs limites d'émissions sont désormais fixées par rapport aux MTD et aux conditions locales.

Enfin, en 2013, une décision de l'Autorité de sûreté nucléaire (ASN, 2013) précise que les différences de performances par rapport aux MTD doivent être comblées selon le résultat d'une étude technico-économique (principe de proportionnalité). Le réexamen porte désormais sur l'installation complète alors qu'il se concentrait auparavant uniquement sur le réacteur.

3.3. Conclusion

En l'absence de documents de référence dédiés, seule la recherche de similarité avec les BREF existants ou l'utilisation de données et de moyens internes sont possibles. Par conséquent, en l'absence de BREF applicable aux INB et de méthodologie commune issue du Processus de Séville, les secteurs non-IED ont des difficultés à identifier et à objectiver le caractère MTD de leurs installations.

4. Conclusions sur l'état de l'art réglementaire

L'état de l'art a montré que le seul cadre formel de détermination et d'application des MTD est le Processus de Séville. Au regard des limites identifiées et en réponse au besoin exprimé par l'INERIS dans le cadre de sa fonction d'appui technique au Ministère de l'Environnement français dans le processus de révision des BREF, le but de cette thèse est donc de proposer des outils et méthodes fournissant un support aux experts techniques lors des phases d'analyse de données en amont des négociations pour les échelles internationale et nationale. Cette demande inclut en particulier la sélection d'un échantillon de sites de référence capable d'illustrer les caractéristiques de l'industrie afin de diminuer la quantité d'information nécessaire à l'argumentation des représentants nationaux face aux autres membres du TWG. Compte-tenu du calendrier des révisions des documents de référence sur les MTD (BREF), il a été possible de suivre la révision du BREF du secteur agroalimentaire qui a ainsi servi de terrain d'observation et d'expérimentation sur lequel a pu se développer la méthodologie proposée.

En parallèle, le besoin d'EDF porte sur des outils et méthodes d'aide au développement d'un argumentaire sur l'application des MTD. En effet, bien que le secteur des Installations Nucléaires de Base, dont les centrales de production électriques font partie, soient hors du champ de l'IED, elles doivent, en France, « *tirer parti des meilleures techniques disponibles* » (Assemblée Nationale, 2012b). Le recours aux MTD doit également être démontré dans le cas de nouveaux projets de centrale à l'international. Toutefois, il n'existe aucun mécanisme semblable au Processus de Séville qui permettrait de définir des MTD internationales applicables aux INB. Le besoin d'EDF porte

donc sur la définition de ses propres MTD au regard des données internationales disponibles dans le but d'une argumentation auprès de l'Autorité de Sûreté Nucléaire (ASN) lors des réexamens périodiques de ces installations.

Par conséquent, la méthodologie proposée vise à être applicable à tout secteur industriel, et a minima, à ces deux contextes différents mais reposant tous les deux sur la même notion de MTD et pour un objectif global commun d'être une aide à la décision pour l'identification des MTD et de leurs performances environnementales associées.

4.1. Problématique générale

Au regard de l'état de l'art et du contexte dans lequel s'inscrit cette thèse, la problématique porte sur la recherche d'une méthodologie de détermination des MTD au sens de l'IED et vise à être applicable à tout secteur industriel en considérant le Processus de Séville comme point de départ.

Ainsi, la problématique principale identifiée au regard des contraintes réglementaires et des processus décisionnels associés peut donc être formulée comme la question suivante :

Quelle méthodologie permettrait d'améliorer le processus de détermination des MTD à partir de la collecte et de l'analyse de données relatives à des installations industrielles et applicable à différentes échelles géographiques ?

En particulier, répondre à cette question revient à :

1. identifier ou confirmer les KEI ;
2. trouver des installations de référence ;
3. déterminer des MTD et BATAEPL.

Ainsi, la méthodologie doit être applicable à différentes échelles géographiques (ex : internationale, nationale, locale) afin d'obtenir une certaine homogénéité entre les niveaux, tel qu'illustré dans la Figure 5 dans le contexte de l'IED. En particulier, au niveau national, la méthodologie doit aider à effectuer un état des lieux et à identifier des spécificités (par exemple, des KEI) qui serviront ensuite à sélectionner des sites de référence à transmettre à Séville suite à la réunion de lancement d'une création ou révision de BREF.

Deuxièmement, au niveau international, à partir de l'analyse des données d'installations opérant dans différents pays, il faut identifier les techniques susceptibles d'être des MTD et leurs niveaux de performance environnementale associés.

Enfin, en disposant de données sur un parc d'installations, il s'agit de positionner un site au regard de ce parc, pour aider à justifier de son niveau de performance par rapport aux MTD. La Figure 5 illustre ces trois objectifs présentés par deux questions et sont positionnés par rapport au Processus de Séville aux échelles nationales et internationale.

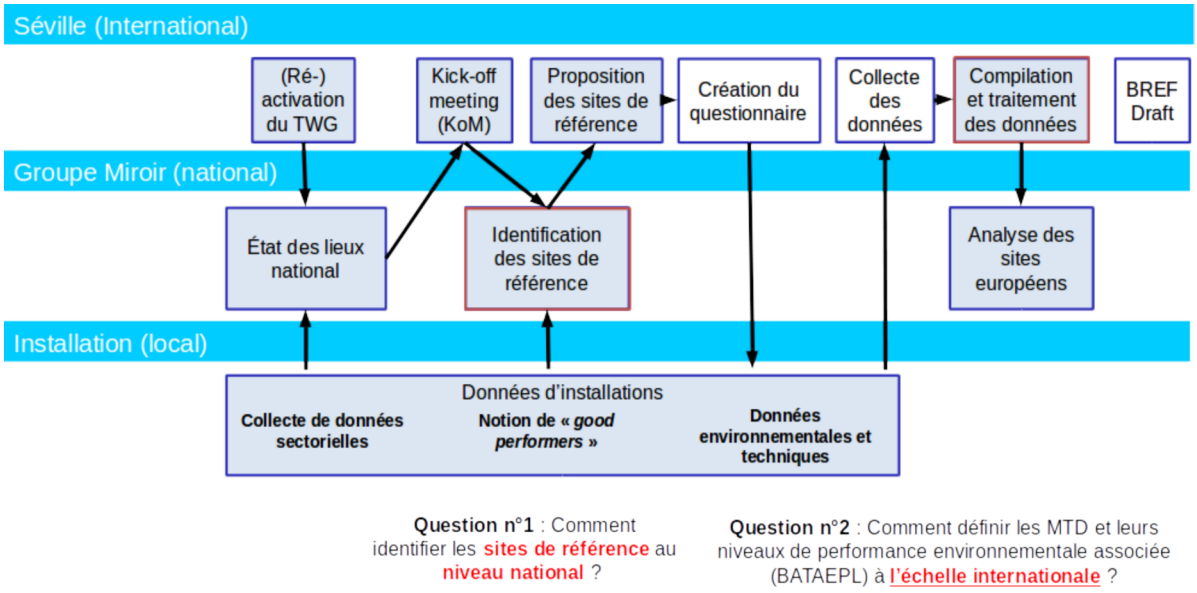


Figure 5 : Positionnement de la méthode par rapport au Processus de Séville et au groupe miroir, et positionnement de l'expert.

Concernant les installations non IED telles que les INB, la première tâche consiste à cibler les données environnementales les plus pertinentes à partir de la collecte la plus large possible sans nuire à la qualité de l'analyse. Étant donné que l'objectif est d'intégrer les résultats issus de l'application de la méthodologie dans le rapport décennal des centrales nucléaires sur la partie MTD, la justification des décisions est primordiale. Par conséquent, en l'absence de processus formel et de consensus parmi les parties prenantes, aucun enjeu-clé n'est défini au préalable. Il est donc nécessaire de hiérarchiser les caractéristiques environnementales au regard des performances des installations et de trouver des critères de décision qui permettront d'étayer les choix effectués à partir de toutes les données disponibles et du jugement d'experts. Comme dans le cas du Processus de Séville, cette tâche requiert une méthodologie aidant à traiter une grande quantité de données afin d'aider l'utilisateur à étayer et à argumenter ses choix. Par conséquent, l'approche nécessaire

varie essentiellement du Processus de Séville par la définition des enjeux-clés en préalable à la classification des installations et à la recherche des MTD.

4.2. Place des experts

La description de l'intégration de la méthodologie dans les processus existants montre que les experts y jouent un rôle central. À chaque étape, les choix humains conditionnent les paramètres qui seront utilisés dans l'analyse. La méthodologie doit donc se positionner comme une aide à l'analyse de données afin de guider les décisions prises. Les experts doivent ainsi pouvoir intervenir à chaque étape de la méthodologie en définissant chaque paramètre et en validant et en interprétant les éléments d'aide à la décision et les livrables fournis.

Dans le contexte du partenariat avec l'INERIS, les experts visés sont les représentants de la France dans le Processus de Séville qui sont susceptibles d'appliquer la méthodologie lors des réunions du groupe miroir et en préparation des échanges avec le TWG. Dans le cadre du partenariat avec EDF, ce sont les ingénieurs des centres d'ingénierie en charge de rédiger les rapports de réexamen des centrales nucléaires.

Afin d'identifier si les besoins méthodologiques sont couverts dans la littérature, une recherche bibliographique a été menée et est présentée dans le chapitre 2.

4.3. Échelles géographiques

Le Processus de Séville, qui sert de processus décisionnel de référence tout au long de cette thèse, implique des échanges à différentes échelles géographiques locale, nationale et internationale (Figure 5). De même, dans le contexte des installations nucléaires, une analyse aux échelles nationale et internationale est justifiée. Toutefois, la revue de littérature réglementaire a montré qu'aucune approche existante n'intègre une prise en compte multi-échelle qui permettrait d'assurer une cohérence entre celles-ci.

Par conséquent, nous émettons à ce niveau une première hypothèse :

Hypothèse 1 : Une approche intégrée et englobant différentes échelles géographiques permettrait d'assurer la cohérence entre celles-ci.

Chapitre 2 : Méthodes existantes

La revue de littérature scientifique effectuée au cours de cette thèse (Evrard *et al.*, 2016) sur les méthodes existantes relatives aux MTD a confirmé les enjeux d'application sur deux périmètres (secteur et installation) et différentes échelles géographiques (ex : internationale, nationale ou locale). Le Tableau 2 synthétise les principales caractéristiques de ces références et met en avant ces deux aspects. Ce chapitre explore les méthodes relatives aux MTD compte tenu de ces deux éléments et présentent leurs limites et apports en vue de l'évaluation de leur adéquation avec les objectifs de l'INERIS et d'EDF.

Tableau 2 : Principaux aspects des articles trouvés.

Périmètre	Objectifs	Évaluation	Échelle d'application	Utilisateur final	Structure	Références
Secteur industriel	Détermination de MTD et BATAEL sectorielles	Qualitative sur jugement d'experts	National	Décideurs	4 étapes	Dijkmans (2000) ; Derden *et al.* (2002) ; Derden et Huybrechts (2013)
					5 étapes	Polders *et al.* (2012)
		Quantitative			4 étapes	Geldermann et Rentz (2004)
			Régional		10 étapes	National Observatory of Athens (2006) ; Mavrotas *et al.* (2007) ; Georgopoulou *et al.* (2008)
	Comparaison de MTD candidates		Sectoriel		4 étapes	Nicholas *et al.* (2000)
					Non applicable	Halog *et al.* (2001)
	Adaptation de l'IPPC hors UE		Europe	Fournisseurs de technologies	Non applicable	Schollenberger *et al.* (2008)
		Qualitative	National	Autorités	6 étapes	Daddi *et al.* (2012)
Installation	Comparaison des performances entre une installation et des MTD	Quantitative	Installation	Exploitants et décideurs	5 étapes	Barros *et al.* (2009)
		Qualitative			4 étapes	Cikankowitz et Laforest (2013)
	Sélection de techniques pour une installation	Qualitative		Exploitants	3 étapes	Giner-Santonja *et al.* (2012)
		Quantitative			7 étapes	Bréchet et Tulkens (2009)
					4 étapes	Rodríguez *et al.* (2011)
				Non précisé	Non applicable	Bortolini *et al.* (2013)
				Exploitants	4 étapes	Ibáñez-Forés *et al.* (2013)
		Qualitative et quantitative		Non précisé	8 étapes	Aloini *et al.* (2014)
					3 étapes	Nguyen *et al.* (2014)
		Quantitative			2 étapes	Van Caneghem *et al.* (2010)
	Évaluation des performances par rapport aux conditions locales	Qualitative et quantitative			3 étapes	Kabak *et al.* (2014)
		Qualitative and quantitative		Exploitants et décideurs	4 étapes	Cikankowitz et Laforest (2014)
Les deux	Prise de décision multicritère	Non applicable	Non applicable	Non applicable	Non applicable	Ibáñez-Forés *et al.* (2014)
	Mesurages et traitement statistique des données collectés dans le cadre du Processus de Séville	Quantitative	Installation	Opérateurs, TWG	Non applicable	Carretero *et al.* (2016)

1. Méthodes applicables au niveau d'un secteur industriel (échelles internationale et nationale)

La principale différence entre les 12 articles (Tableau 2) applicables au périmètre sectoriel est qu'ils visent soit à identifier une liste de MTD (parfois avec des BATAEL) à partir d'installations dans leur périmètre d'étude, soit à comparer des MTD déjà identifiées. Ils concernent tous des décideurs sauf pour Schollenberger *et al.* (2008) qui axent leurs travaux à destination des fournisseurs et Daddi *et al.* (2013) à destination des autorités. En outre, il est à noter que Daddi *et al.* (2013) se concentrent sur le transfert de technologie et le National Observatory of Athens (2006) vise à créer un cadre d'aide à la décision pour l'évaluation des impacts engendrés par l'introduction de MTD dans l'industrie grâce à une approche économique. Toutes ces approches reposent sur une définition préalable des critères d'évaluation utilisés, sans pour autant décrire la manière de procéder. En revanche, plus d'informations sont disponibles sur la recherche d'installations de référence puis sur la détermination des MTD pour le niveau sectoriel.

1.1. Réponses de la littérature sur les besoins au niveau sectoriel

1.1.1. Sur la définition d'enjeux-clés

Toutes les méthodes se concentrent d'abord sur des variables environnementales avant les considérations économiques ; les valeurs de consommations et d'émissions étant les données quantitatives les plus disponibles. En ce qui concerne les critères techniques et économiques, le jugement d'experts ou alors une valeur économique sont employés. En revanche, aucune des méthodes trouvées ne propose de moyen de sélectionner les critères les plus pertinents dans le contexte de l'étude, si ce n'est en se concentrant sur un cadre réglementaire et une approche par substance ou par impact.

Dijkmans (2000) propose une démarche en plusieurs étapes dont la deuxième repose sur une succession d'évaluations et exclusions des techniques non satisfaisantes. Elles sont évaluées par jugement d'experts en fonction de leur faisabilité technique, performance environnementale puis de leur faisabilité économique. Les techniques restantes sont comparées et une sélection finale est réalisée par un groupe de travail sectoriel composé de représentants de l'Industrie, des agences publiques et d'experts indépendants. Le livrable de cette méthode regroupe une liste de MTD appliquées à la Flandre. Cependant, aucun enjeu clé (KEI) n'est proposé et identifié clairement dans la méthode de Dijkmans (2000). Les experts ont un rôle prépondérant. Geldermann et Rentz (2004) proposent une aide à l'identification des MTD permettant à tout secteur industriel visé de respecter les limites d'émission du Protocole de Göteborg en termes d'oxydes d'azote et de composants organiques volatiles. Il y est fait référence à la directive IPPC en ce qui concerne la

prise en compte des effets croisés. Dans cette méthode, une classification en sous-secteurs est essentielle pour l'identification des paramètres pertinents et pour s'assurer que le petit nombre d'installations de référence considéré couvre le secteur entier.

Dans la littérature, peu d'auteurs traitent de l'identification des KEI. Le choix des critères est fait essentiellement selon le contexte réglementaire et sur jugement d'experts, sur la base des données disponibles.

1.1.2. Sur la recherche d'installations de référence

À l'exception des méthodes appliquées sur un périmètre restreint dans lequel une analyse exhaustive est possible (Dijkmans, 2000), les méthodes trouvées dans la littérature scientifique se concentrent sur l'analyse d'installations de référence pouvant être appelées « installations représentatives » (Geldermann et Rentz, 2004) ou « installations typiques » (National Observatory of Athens, 2006).

Dans Geldermann et Rentz (2004), dans la mesure où toutes les installations ne peuvent pas être analysées individuellement, elles sont groupées en catégories et chacune est représentée par une « installation de référence » dont les caractéristiques peuvent être considérées comme valeurs représentatives de toutes les autres installations concernées (Geldermann et Rentz, 2004). Par conséquent, toutes celles qui sont représentées par une même installation de référence sont supposées être capables d'appliquer les mêmes options d'abattement avec une applicabilité, une efficacité et des coûts similaires pour une même option de réduction des émissions. Aucune information dans l'article n'indique si une installation de référence est une installation réelle ou fictive, si elle représente une installation moyenne, performante ou autre. De plus, il n'y a pas d'indication de la représentativité des installations de référence afin de renforcer le raisonnement en faveur d'une combinaison de MTD. Cette approche repose sur le jugement d'experts pour évaluer les techniques et combinaisons de techniques possibles, ainsi que sur l'Analyse du Cycle de Vie (ACV) pour évaluer la signification écologique des critères

Enfin, afin de considérer autant d'alternatives que possible en fonction de différents critères environnementaux, techniques et économiques, des méthodes d'évaluation mathématiques multicritères ont été développées (Georgopoulou *et al.*, 2008 ; Mavrotas *et al.*, 2007 ; National Observatory of Athens, 2006). Ces approches impliquent que toutes les installations d'un sous-secteur soient connues et relativement homogènes et qu'elles puissent toutes être couvertes par quelques installations « typiques ». Le nombre d'installations « typiques » est choisi par l'utilisateur de la méthode et leur représentativité n'est pas évaluée. De plus, une recherche de

l'optimum de Pareto est utilisée afin d'identifier les techniques qui permettent d'améliorer la performance avec une approche multicritère.

Les installations de référence, réelles ou fictives, peuvent donc être de deux sortes :

1. des installations représentatives au sens où elles contiennent des caractéristiques communes avec un groupe d'installations ;

2. des installations performantes au sens où les émissions et consommations sont basses.

1.1.3. Sur la détermination des MTD et BATAEPL

En Flandres, des études sectorielles sur les MTD couvrent des secteurs IED comme l'agroalimentaire (Derden *et al.*, 2002) ou le textile (Derden et Huybrechts, 2013) et des secteurs non-IED en relation avec la réglementation locale. C'est le BAT-Centre de l'Institut Flamand de Recherche Technologique (VITO) qui, en tant que représentant de la Belgique dans le Processus de Séville propose et développe des outils méthodologiques pour la détermination et l'application des MTD et BATAEPL. Dijkmans (2000) a été la première méthode visant à déterminer des listes de MTD pour la Flandre. Celle-ci est divisée en quatre étapes principales : (1) collecte d'information en provenance d'installations existantes du secteur visé et études pour obtenir des MTD candidates ; (2) les MTD candidates sont évaluées qualitativement ; (3) les MTD sont traduites en suggestions pour les permis ; (4) les conséquences économiques et écologiques des conditions de permis tirées des MTD sont alors estimées.

En complément à ces travaux et afin de déterminer les limites d'émission associées à ces MTD, la méthode développée par (Polders *et al.*, 2012) est utilisée. Les BATAEL sont déterminées en analysant les performances des installations concernées. À la différence du Processus de Séville, où la sélection des MTD est simultanée avec la définition des BATAEL, cette étape a été considérée individuellement par VITO. Dans ces deux méthodes, l'évaluation qualitative multicritère repose principalement sur le jugement d'experts supporté par des données locales. Le principal inconvénient de la définition des BATAEL est qu'elle présuppose qu'une liste de MTD a déjà été proposée, alors que ces deux phases sont simultanées dans le Processus de Séville. La considération d'autres niveaux de performance tels que l'efficacité énergétique ou la sûreté est laissée à la détermination des MTD dans l'approche de Dijkmans. Enfin, il n'y a pas d'analyse de l'applicabilité des BATAEL, telle qu'une indication du pourcentage de sites qui ne pourraient pas respecter les niveaux d'émissions et les raisons à cela.

De son côté, le cadre décisionnel grec a été créé plus spécifiquement pour « l'évaluation des possibles coûts et bénéfices économiques et environnementaux impliqués par l'application des MTD pour une région donnée » (National Observatory of Athens, 2006). Il est basé sur un nombre réduit d'installations « typiques » par secteur et sous-secteur industriel, citant sur cet aspect le travail de Geldermann et Rentz (2004). Le cadre est divisé en trois parties principales. La première fait référence à la détermination de MTD potentielles identifiées par des experts avec la détermination des installations « typiques » et de possibles MTD. En revanche, aucune BATAEPL n'est définie dans cette approche.

L'identification des techniques se fait donc essentiellement en considérant individuellement ou successivement les critères, à l'image du Processus de Séville. En accord avec le besoin d'une approche éprouvée pour la détermination des MTD, toutes les méthodes se basent en grande partie sur des informations provenant d'installations en activité et donc l'applicabilité ou la représentativité de terrain est conservée.

Dans les cas où une liste de MTD a déjà été déterminée, trois articles décrivent des méthodes permettant de comparer les alternatives. Ainsi, Nicholas *et al.* (2000) utilisent l'ACV (ISO, 2006) pour modéliser les procédés et évaluer les émissions et réductions d'impacts dues à l'application des MTD. Par ailleurs, cette méthode considère les dimensions techniques et sociales pour prendre une décision. La principale limite est la nécessité de modéliser le procédé dans sa totalité, ce qui, dans le cas d'une étude sectorielle présuppose qu'une certaine généricité existe dans le secteur étudié.

La seconde méthode trouvée (Halog *et al.*, 2001) utilise également l'ACV mais avec aussi une adaptation des principes de *Quality Function Deployment* (Développement des Fonctions Qualités) à la réduction des émissions. Ainsi, les demandes des parties prenantes sont définies comme catégories d'impacts, et les contraintes de conception d'une technique particulière comme substances que le processus émet et doit être réduite. Des préférences d'experts apparaissent en termes de pondération des catégories d'impacts. Par ailleurs, des objectifs d'émissions sont exprimés au travers de valeurs limites d'émissions existantes. Enfin, une hiérarchisation des substances émises, basée à la fois sur le potentiel d'impact économique et environnemental, est effectuée.

Schollenberger *et al.* (2008) abordent le transfert de technologie des MTD hors de l'UE en développant un système de mesure basé sur une approche de distance à la cible pour la sélection des MTD. Cette approche présente l'avantage de fusionner les objectifs technologiques, économiques et environnementaux au travers d'un outil d'aide à la décision multicritère. De plus,

des objectifs nationaux sont considérés pour chaque critère afin d'inclure des spécificités nationales, ce qui placerait cette méthode à l'interface entre le niveau sectoriel et installation si les spécificités locales étaient explorées plus avant.

Dans ce cas également, la littérature traite de manières différentes la comparaison de techniques avec une vision détaillée des procédés, ce qui requiert une modélisation complète des installations. Néanmoins, le système de métrique développé par Schollenberger *et al.* (2008) permet une comparaison véritablement multicritère en considérant des critères exprimés dans des unités différentes.

1.1.4. Aide à la décision dans l'existant et évaluation multicritère

Comme dans le Processus de Séville, la décision d'experts a une place importante dans les méthodes trouvées. Dans sa revue de littérature, Ibáñez-Forés *et al.* (2014) identifient plusieurs approches de prise en compte de jugements de multiples experts. Elles sont principalement utilisées pour comparer des alternatives techniques en contexte MTD ou non.

En outre, comme constaté avec l'état de l'art réglementaire, les installations de référence peuvent être soit des installations représentatives de groupes de caractéristiques déterminées par les experts ou par évaluation multicritère, soit performantes sur des critères individuels ou multiples comme avec l'optimum de Pareto recherché par Mavrotas *et al.* (2007). La recherche d'autres outils d'analyse multicritère a montré que les analyses fonctionnelles (Saporta, 2011) permettent de traiter des données quantitatives et qualitatives simultanément.

Il existe donc différentes approches d'évaluation en fonction des objectifs, soit de caractérisation de l'existant, soit de recherche d'un optimum qui ont tous pour objectifs d'apporter un soutien aux experts dans l'aide à la décision.

2. Méthodes applicables au niveau d'une installation

À l'échelle locale, c'est-à-dire au niveau d'une installation, trois objectifs principaux ont été identifiés dans la littérature (Tableau 2) : (1) comparaison de la performance entre une installation et des MTD, (2) sélection de techniques pour une installation, (3) évaluation de performance par rapport aux conditions locales.

2.1. Comparaison des performances

Seules deux méthodes ont été trouvées sur la question de la comparaison d'une installation existante avec les performances des MTD. La première, de Barros *et al.* (2009), consiste en un inventaire des procédés, des flux de matière et d'énergie ; une caractérisation des impacts environnementaux liés et un passage en revue d'une liste de MTD applicables par des experts d'une entreprise pour identifier celles qui seraient déjà en place. La dernière partie de la méthode est une revue des techniques pour vérifier la faisabilité dans l'installation étudiée. Cette méthode applique une évaluation qualitative des aspects techniques et économiques à une connaissance quantitative des consommations et émissions d'une installation.

La deuxième d'entre elles, appelée L-BAT (Local-BAT) développée par Cikankowitz et Laforest (2013, 2014) prend en compte les 12 considérations de la directive IPPC, analogues aux 12 critères de l'IED, pour déterminer les MTD au travers d'une grille d'analyse basée sur le BREF existant. Elle est la plus proche de la notion de MTD en évaluant le niveau de contrôle des techniques. Ainsi L-BAT vise à effectuer un diagnostic environnemental d'une installation comparée aux MTD décrites dans le BREF, les réglementations locales (valeurs limites d'émission) et les performances des MTD. Par ailleurs, elle considère le contexte local pour l'amélioration des performances techniques de l'installation.

Ces deux méthodes illustrent comment comparer la performance sur site avec les niveaux de performance associés aux MTD, vérifiant ainsi l'application des MTD. Elles sont toutes les deux basées sur un passage en revue du BREF pour comparer les niveaux de performance, si les MTD ne sont pas directement appliquées. La considération du contexte local est aussi traitée via la vulnérabilité de l'environnement proche et les capacités techniques et économiques des exploitants. Barros *et al.* (2009) décomposent également le processus de production selon ses principales opérations et flux (matière et énergie) comme base de comparaison.

2.2. Sélection des MTD pour une installation

Quand une installation n'atteint pas la performance des MTD, son exploitant doit trouver des techniques qui pourront l'aider à améliorer sa performance environnementale à un coût acceptable. Sur cette question, huit articles fournissent un éclairage sur la manière de sélectionner les techniques appropriées à une installation donnée.

Plutôt que de considérer toutes les émissions, Rodríguez *et al.* (2011) ciblent uniquement celles qui sont identifiées comme améliorables suite à une analyse détaillée des flux de matière et d'énergie. De plus, les impacts environnementaux causés par des substances peuvent être évaluées

grâce à une ACV (Ibáñez-Forés *et al.*, 2013). Des critères techniques et économiques sont également pris en compte dans plusieurs autres articles, soit par jugement d'experts (Giner-Santonja *et al.*, 2012) soit sur la base du coût social (Bréchet et Tulkens, 2009). L'évaluation économique proposée par (Bortolini *et al.*, 2013) considère le contexte local pour la sélection de panneaux solaires. Cette approche pourrait être appliquée pour étudier les spécificités locales d'une installation mais manque d'une approche intégrée.

Le résultat de la comparaison de techniques est soit un classement, soit une note unique des alternatives basés sur l'affichage de différents critères (Bréchet et Tulkens, 2009), ou une approche d'aide à la décision multicritère (Aloini *et al.*, 2014 ; Giner-Santonja *et al.*, 2012 ; Nguyen *et al.*, 2014) qui nécessite l'implication de décideurs multiples et une évaluation de leurs décisions et de leurs valeurs. De plus, comme Laforest (2014), la méthode de Giner-Santonja considère les 12 considérations de la directive comme préliminaire pour la détermination de ses critères d'évaluation. Le traitement des critères ou indicateurs d'évaluation peut être réalisé via l'utilisation de l'approche Analytic *Network Process* (ANP) apparaît comme dans d'autres méthodes de sélection de techniques qui la couple avec la logique floue (Kabak *et al.*, 2014 ; Nguyen *et al.*, 2014) afin d'estimer le jugement d'experts. Enfin, des techniques ou combinaisons de techniques peuvent être évaluées grâce à des indicateurs d'éco-efficience (Van Caneghem *et al.*, 2010), ce qui présente l'avantage de considérer plusieurs critères simultanément parmi des considérations environnementales, techniques et économiques.

2.3. Évaluation des performances en fonction des conditions locales

Deux articles relèvent de la considération des conditions locales pour une évaluation quantitative. Kabak *et al.* (2014) pour évaluer la performance énergétique d'un bâtiment de sorte qu'une méthode d'aide à la décision multicritère floue est utilisée pour pondérer le jugement d'experts sur la performance de divers critères locaux tels que sa localisation, son système d'éclairage ou sa consommation énergétique. Toutes les alternatives sont comparées deux à deux au travers d'un Analytic *Network Process* (Saaty, 1996) pour obtenir une note unique.

Laforest (2014) prend en compte les conditions locales pour la quantification des indicateurs d'évaluation pour différentes applications dans le traitement des eaux usées, la remédiation des sols et la valorisation des boues. Cikankowitz et Laforest (2013) considèrent qualitativement le contexte local pour définir les problèmes locaux à prendre en compte dans un planning de mise en œuvre des MTD. La prise en compte des conditions locales est donc indispensable pour établir l'applicabilité ou l'efficacité d'une technique sur un site.

Les critères techniques et économiques sont souvent peu détaillés avec une référence au BREF existant mais, en l'absence de ce document, la seule solution semble être de positionner une installation par rapport à un parc existant.

3. Synthèse pour une approche intégrée des MTD de l'échelle européenne à l'échelle locale

Les méthodes existantes portent soit sur l'échelle nationale, soit sur l'échelle d'une installation. Aucune d'entre elles ne les intègre toutes tel que dans le Processus de Séville (Figure 6) et qui permettrait une détermination et une application complètes des MTD. Le seul cadre formel existant sur les MTD étant le Processus de Séville, le développement d'une méthodologie d'aide à la détermination des MTD et BATAEL doit donc s'en inspirer.

De ce qui précède, les hypothèses directrices suivantes peuvent être émises.

Hypothèse 1 : Une approche intégrée et englobant différentes échelles géographiques permettrait d'assurer la cohérence entre celles-ci.

Hypothèse 2 : Une méthodologie structurée, statistique et multicritère permet de contribuer à l'atteinte des objectifs de l'IED en termes de performance globale.

Cependant, les apports méthodologiques montrent 5 étapes essentielles pour déterminer les MTD et BATAEPL à partir de données collectées sur des installations existantes et de leur traitement statistique, qui semble l'approche la plus appropriée pour traiter un grand nombre d'informations.

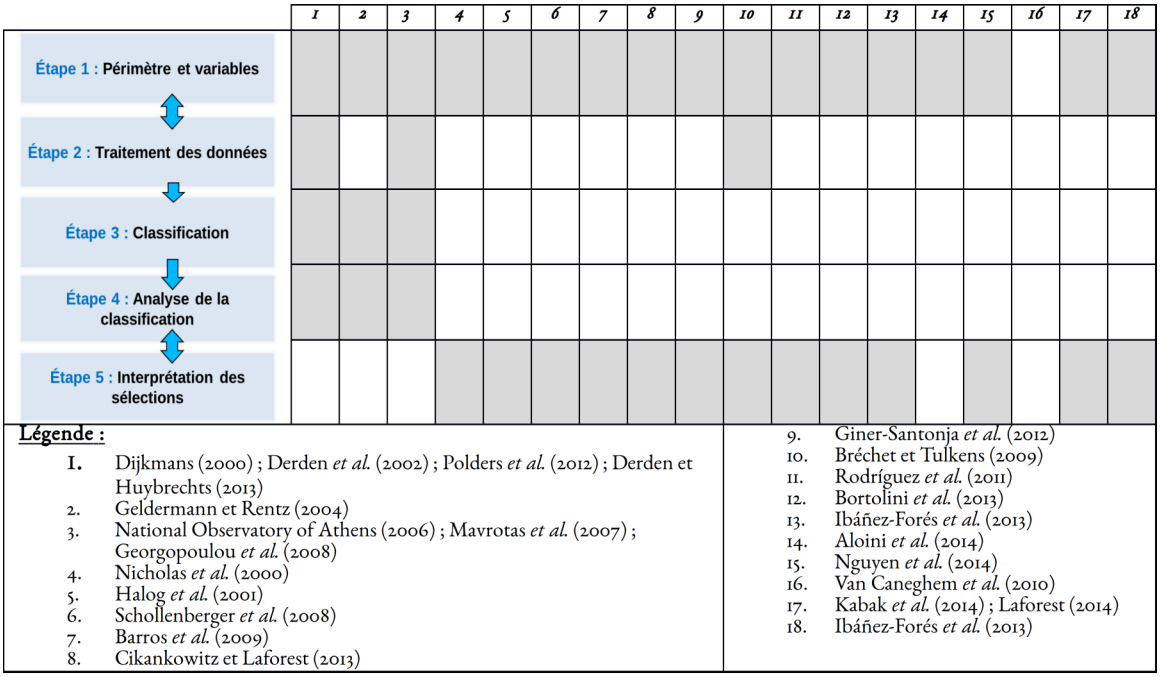


Figure 6 : Positionnement de la littérature scientifique par rapport aux étapes de la méthodologie proposée.

3.1. Étape 1 : Définition du périmètre et des variables

Le périmètre de l'étude est défini par le contexte d'application. En revanche, le choix des variables qui seront utilisées dans l'analyse est essentiel. L'intégration dans les processus décisionnels existants étant cruciale, il est nécessaire que la méthodologie proposée utilise les données disponibles. Par conséquent, l'analyse qui sera faite doit reposer sur des variables d'émissions et de consommation ; données permettant de caractériser le niveau de performance environnementale d'une installation.

3.2. Étape 2 : Traitement des données

Dans le Processus de Séville, les données collectées sont vérifiées pour assurer leur cohérence et sont analysées variable par variable avec des panels d'installations différents en fonction des données disponibles. Les valeurs moyennes de consommations et d'émissions des KEI sont considérées avec des différences de traitement selon les États membres. Comme Carretero *et al.* (2016) le suggèrent, des tests d'hypothèse pourraient être effectués pour comprendre les relations entre variables. Cependant, il semble que la prise en compte des valeurs manquantes dans l'analyse ne soit considérée qu'au travers d'une analyse multicritère.

Ainsi, il est possible de considérer l'ensemble des critères simultanément dans l'analyse en vue de déterminer les MTD et BATAEPL. Dans un tel cas, les données manquantes risquent d'avoir une influence sur les résultats et cette influence devrait être au-moins connue, sinon atténuée.

3.3. Étape 3 : Classification des installations

Deux approches ont été trouvées. D'abord, une connaissance exhaustive de toutes les installations concernées peut être cherchée (Dijkmans, 2000). Cela est réalisable dans des secteurs industriels dotés d'un petit nombre d'installations qui utilisent des techniques relativement homogènes. Dans ce cas, la similarité entre installations peut compenser des données manquantes. Deuxièmement, lorsque des informations détaillées ne sont pas disponibles pour l'ensemble des installations du secteur étudié ou qu'il y a un trop grand nombre d'installations pour un traitement exhaustif, une sélection d'un échantillon de celles-ci, par rapport à des enjeux-clés identifiés préalablement pourrait être utilisée. Néanmoins, les états de l'art réglementaire et scientifique ont montré que différentes interprétations sur la nature des installations de référence à utiliser existent (Tableau 3).

Tableau 3 : Comparaison des définitions d'installations de référence en fonction des acteurs.

	Définition	Auteurs	Installations considérées	Analyse des critères
Installation représentative	Installation présentant les caractéristiques communes à un groupe d'installations	Geldermann et Rentz (2004), Mavrotas *et al.* (2007)	Fictives	Individuelle
	Sites moyens	TWG	Réelles	Individuelle
Good-performer	Meilleurs sites en termes de bas niveaux de consommations et d'émissions	EIPPCB	Réelles	Individuelle
	Sites conformes à la réglementation et validés par l'autorité environnementale	État membre France	Réelles	Individuelle
	Sites réalisant une activité donnée, appliquant des techniques spécifiques ou présentant une particularité.	CNIEL	Réelles	Individuelle

L'IED définit les MTD comme les techniques permettant d'atteindre un haut niveau de protection de l'environnement dans son ensemble. Par conséquent, il semble qu'il soit nécessaire d'analyser un ensemble de critères simultanément et non individuellement pour répondre à cette définition. De plus, la directive, comme son nom l'indique, vise à réduire les émissions de substances dans l'environnement. Cela implique de tendre vers des valeurs d'émission plus basses que ce qui était dans la situation présente. Étant donné, la double interprétation représentative et performante rencontrée, il paraît nécessaire de considérer deux cas. D'une part, les installations de référence nécessaires à la recherche des MTD doivent être représentatives du parc actuel qui est déjà conforme aux réglementations en vigueur et susceptibles de mettre en œuvre des typologies d'installations différentes et d'avoir des émissions et consommations plus faibles. D'autre part, de rechercher les meilleures installations en considérant l'ensemble des variables utilisées dans l'analyse. Ainsi, les installations de référence peuvent être définies comme étant :

- soit des **installations représentatives**, c'est-à-dire qui possèdent les mêmes caractéristiques principales que le groupe de sites dont ils relèvent (Larousse, 2016) ;
- soit des **installations performantes** qui permettent un haut niveau de protection de l'environnement dans son ensemble tel que défini dans l'IED (Commission Européenne, 2010).

3.4. Étape 4 : Analyse des classifications

À partir du classement des installations réalisé à l'étape précédente, la recherche du petit nombre d'installations nécessaire doit être effectuée. Dans certaines approches, le nombre d'installations de référence est fixé à l'avance (Geldermann et Rentz, 2004 ; National Observatory of Athens, 2006). Néanmoins, il semble qu'une métrique permettant de caractériser le degré de représentativité inspiré par Schollenberger *et al.* (2008) serait envisageable sur une approche multicritère pour hiérarchiser les installations telle qu'une analyse factorielle (Saporta, 2011).

3.5. Étape 5 : Interprétation des sélections

Les installations de référence maintenant identifiées vont servir à déterminer les MTD et leurs niveaux de performance associés (BATAEPL).

3.5.1. Définition de valeurs de référence

Il s'agit ici d'identifier et de sélectionner les niveaux de consommations et d'émissions des installations de référence pour définir les valeurs de référence (BATAEPL) pour le secteur industriel concerné.

3.5.2. Recherche de MTD candidates

À partir des valeurs de référence identifiées précédemment, il est supposé que les installations mettant en œuvre des MTD présentent des valeurs de consommations et d'émissions égales ou inférieures aux valeurs de référence.

4. Application de la démarche à l'échelle locale vs européenne

La démarche en cinq étapes proposée ici est issue de la synthèse bibliographique au regard des besoins des acteurs (TWG notamment). Elle est essentiellement issue des approches développées à l'échelle européenne ou sectorielle, mais elle ne semble pas incompatible avec une application à l'échelle locale. Dans ce cas, l'objectif est de comparer les performances d'une installation spécifique par rapport aux performances de celles qui sont reconnues comme mettant en œuvre les MTD. Cela pourrait donc être réalisable en positionnant l'installation par rapport aux installations de référence.

Conclusions de la partie 1

À l'heure actuelle, la vision globale et cyclique de la démarche décrite dans le chapitre 1 est prise en compte partiellement dans la littérature. Une approche intégrée et englobant les trois échelles géographiques identifiées permettrait d'éviter les incompatibilités entre ces dernières. De plus, l'utilisation d'une même démarche pourrait aider à réduire les différences de traitement constatées tout en étant utilisable à différentes étapes des processus décisionnels existants. Comme il s'agit de traiter de grandes quantités de données, le recours à des outils statistiques semble pertinent. Pour répondre à ces constats et besoins, cinq étapes ont été déduites de la littérature scientifique et des principes généraux quant aux outils applicables dans chacune ont été proposés.

À chacune de ces étapes, un ou plusieurs outils d'aide à la décision ressortent de l'analyse bibliographique. En outre cette démarche globale et intégrée permet de considérer l'environnement dans son ensemble, c'est-à-dire, comme postulé précédemment, au travers d'un traitement simultané de plusieurs critères environnementaux.

En outre, cette démarche doit rester applicable aux différentes échelles et prendre en considération le contexte dans lequel elle doit être appliquée (Processus de Séville pour les installations IED ou autre processus tel que le cadre des INB, pour les installations non-IED). Cette analyse bibliographique a ainsi pu mettre en exergue les besoins des acteurs au regard de l'application du concept des MTD et du périmètre d'étude. Une problématique générale a ainsi pu être émise qui s'est traduite par une question de recherche globale. En outre, des postulats et hypothèses ont été émis.

Le Tableau 4 reprend l'ensemble de ces éléments. La méthodologie proposée permettant d'y répondre est présentée dans la partie 2 et les applications dans la partie 3.

Tableau 4 : Questions, hypothèses, postulats et caractéristiques de la méthodologie dans le cadre de cette thèse.

Problématique générale
Le seul cadre formel existant pour la détermination des MTD, le Processus de Séville, est sujet à des différences d'interprétations qui créent des problèmes de justification de décisions prises à chacune de ses étapes.
Question générale
Quelle méthodologie permettrait d'améliorer le processus de la détermination des MTD à partir de la collecte et de l'analyse de données relatives à des installations industrielles et applicable à différentes échelles géographiques ?
Hypothèse 1
Une approche intégrée et englobant différentes échelles géographiques identifiées permettrait d'assurer la cohérence entre les différentes échelles.
Hypothèse 2
Une méthodologie structurée, statistique et multicritère permet de contribuer à l'atteinte des objectifs de l'IED en termes de performance globale.
Postulats
Si un site est remonté pour son analyse lors du Processus de Séville, c'est qu'il est viable économiquement, techniquement et qu'il a un faible impact environnemental dans son ensemble. Il est donc conforme à l'IED.
La caractérisation et la sélection des sites doivent être guidées par le niveau de performance environnementale dans son ensemble.
Principales caractéristiques de la méthodologie
1. Partagée afin d'uniformiser le processus de sélection, 2. Transparente et permettant de tracer la décision (traçabilité), 3. Adaptable à des processus décisionnels variés, 4. Robuste, c'est-à-dire permettant de limiter l'incertitude, 5. Réplicable à tout secteur industriel, 6. Objective au sens où la méthode ne fait pas intervenir d'*a priori,* 7. Intégrée à l'ensemble des étapes du processus de décision.

Partie 2 :
Méthodologie proposée

Introduction

La revue de littérature a permis d'identifier cinq étapes permettant de prendre en considération les besoins relatifs à l'identification des MTD et des BATAEPL que ce soit dans un contexte IED ou non. Certaines d'entre elles mettent en œuvre des outils mathématiques spécifiques afin de fournir des éléments d'aide à la décision pour la détermination des MTD et BATAEPL. Ces outils et méthodes se positionnent éventuellement dans les phases de traitement, d'analyse et d'interprétation de données et préparent les sessions de prise de décision entre parties prenantes.

Du fait de la volonté de déploiement d'une méthodologie transverse et applicable à l'ensemble des processus décisionnels évoqués (Séville, EDF), des secteurs industriels (IED, non-IED) ainsi qu'aux différentes échelles « géographiques » (locale, nationale et internationale), la démarche méthodologique proposée dans ce chapitre combine des éléments mathématiques génériques de manière originale.

Par conséquent, la méthode met en œuvre dans un premier temps les données provenant d'installations d'un secteur industriel afin de le caractériser (Étapes 1, 2 et 3), puis identifie les installations et valeurs de référence et des taux d'effort qui serviront finalement à proposer des BATAEPL et des techniques susceptibles de devenir des MTD (Étapes 4 et 5). Ce passage par des valeurs de référence et des taux d'effort est un apport qui ne se retrouve pas ailleurs dans la littérature.

Étapes de la méthodologie et outils employés

La méthodologie proposée se compose des cinq étapes mises en évidence par l'état de l'art. Leur intégration et leur application n'est pas spécifique à un processus décisionnel particulier et permet donc des possibilités d'adaptation adéquates à l'ensemble des secteurs (IED et non-IED).

Le point commun à l'ensemble des processus décisionnels ciblés se trouve être la nécessité de traiter des quantités importantes de données techniques et environnementales provenant d'installations existantes. Ce préalable justifie le recours à des outils statistiques pour le traitement, l'analyse et l'interprétation de ces données en vue de définir les sites de référence puis d'établir des MTD et BATAEPL.

D'après les observations réalisées sur le Processus de Séville, il s'est avéré que les informations les plus facilement disponibles en termes de quantité et de qualité regroupent majoritairement des données numériques de consommations et d'émissions des sites industriels. Les BATAEPL étant issues de ces données, la méthodologie vise donc principalement, au travers des cinq étapes, à guider les experts et à leur fournir des éléments d'aide à la décision et des livrables issus d'outils statistiques intégrables dans les processus de décision grâce à une analyse multivariée et discriminante.

Dans la suite, le terme « experts » désignera toute personne utilisatrice de la méthodologie. Dans le cas des applications présentés dans la partie suivante, ce sont les experts techniques supports du Ministère de l'Environnement dans le cadre du Processus de Séville. Ils sont susceptibles d'employer la méthodologie lors des réunions du groupe miroir national ou des échanges avec le TWG au niveau européen. Dans le cas de l'application aux INB, ce sont les centres d'ingénierie chargés de la mise en œuvre des MTD pour le parc de centrales nucléaires d'EDF.

Le déroulé de la méthodologie en cinq étapes est présenté dans la Figure 7. Dans les parties suivantes, l'objectif de chacune d'entre elles est présenté, puis, lorsque différentes alternatives méthodologiques ont été trouvées, une description de celles-ci et la justification du choix de l'une d'elle dans la méthodologie proposée sont présentées. Enfin, les outils d'aide à la décision et livrables de chaque étape de la méthodologie sont présentés.

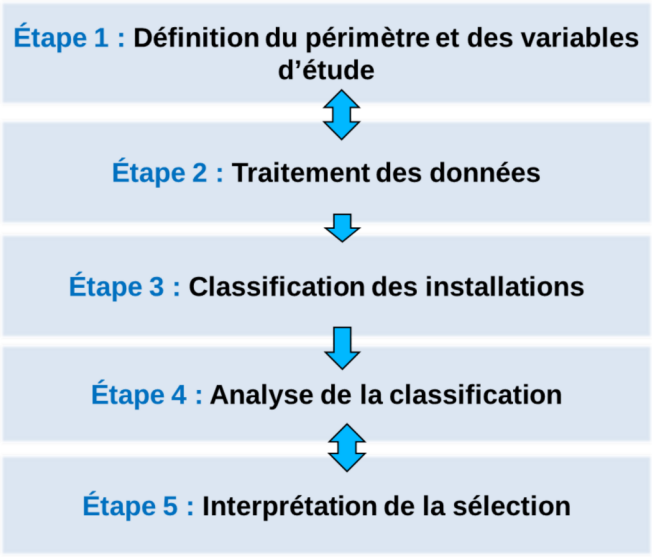


Figure 7 : Étapes de la méthodologie proposée.

1. Étape 1 : Définition du périmètre et des variables d'étude

Au regard des données préalablement collectées, la première étape consiste à définir le périmètre et les variables d'étude qui seront employés tout au long de l'analyse. Cette étape essentielle pour la transparence des résultats obtenus et la justification de la qualité des analyses proposées est effectuée par les experts appartenant au processus décisionnel. Comme indiqué dans l'état de l'art, la définition du périmètre correspond au secteur industriel et à la population d'installations étudiée. La définition des variables, quant à elle, englobe un ensemble de données collectées et classées au sein de trois catégories : les variables d'analyse, les variables de normalisation et les variables descriptives.

La classification des variables au travers de ces trois catégories repose sur les choix et l'expertise de l'utilisateur. Pour autant, et bien que les experts impliqués dans les processus décisionnels soient les plus à même d'identifier les variables les plus pertinentes en lien avec leur contexte ainsi que leur catégorisation, une aide à la sélection est proposée. Cette aide peut ainsi être employée *a priori*, c'est-à-dire, dans le but de fournir une classification par défaut des variables par catégorie ou bien *a posteriori*, comme validation de la pertinence du choix d'experts. De plus, il est possible de détecter les variables dépendantes (au sens mathématique) ou pouvant impacter de manière significative les étapes 3 et 4 de classification et d'analyse. Les conséquences que peuvent avoir la qualité et la quantité des données disponibles seront prises en compte pour les outils mathématiques retenus aux étapes suivantes.

1.1. Périmètre de l'étude

Définir le périmètre de l'étude revient à choisir l'échantillon de sites industriels parmi la population totale qui seront étudiées dans le cadre du déroulement de la méthodologie. Par conséquent, le panel sélectionné peut regrouper un ensemble de *n* sites. La valeur *n* pouvant au maximum être égale au nombre total de sites dans la population et, au minimum, être supérieure au nombre de variables d'analyse.

1.2. Variables d'étude

Les informations et données collectées sur les installations existantes sont classées, comme mentionné précédemment, dans trois catégories :

- catégorie 1 : **Les variables d'analyse**
  Ces variables regroupent les données quantitatives de consommations et d'émissions des sites industriels étudiés ;

- catégorie 2 : **Les variables de normalisation**
  Les variables de normalisation (variables quantitatives) permettent d'ajuster les séries de valeurs (les variables d'analyse) suivant une fonction de transformation afin de les rendre comparables entre elles et par conséquent de limiter les effets d'échelle et de distorsion entre les sites échantillonnés. Elles peuvent correspondre à une quantité de matière ou de produits entrants ou sortants, pondérés ou non ; par exemple, les concentrations de substances dans les effluents qui correspondent à des flux sortants divisés par le volume d'effluents ;

- catégories 3 : **Les variables descriptives**
  Ces variables (quantitatives et/ou qualitatives), englobent l'ensemble des données disponibles ne faisant pas partie des deux premières catégories. Ces variables sont employées afin de décrire et de caractériser les éventuels groupements d'installations qui auront été mis en évidence, notamment à l'étape 3. Elles peuvent, par exemple, correspondre à leur âge, leur chiffre d'affaires, ou encore leur volume de production.

Le nombre de variables par catégorie n'est pas restreint. Pour autant, il convient de respecter des règles empiriques et quelques principes d'applicabilité courants dans le cadre d'analyses statistiques. Ainsi, il conviendra de limiter le nombre de variables d'analyse afin d'obtenir au maximum, un nombre de variables d'analyse inférieur ou égal au nombre d'installations de l'échantillon étudié.

Les deux sous-parties qui suivent vont décrire plus en détail les deux premières catégories de variables d'étude.

1.2.1. Variables d'analyse

Dans la méthodologie, seules des variables quantitatives de consommations et d'émissions sont considérées et, du fait de l'objectif et du cadre d'application de la méthodologie (détermination des MTD et BATAEPL notamment), représentent des problématiques environnementales à enjeux (par exemple, consommation de ressources et émission de polluants).

Comme mentionné précédemment le choix des variables d'analyse peut être entièrement dicté par l'expertise de l'utilisateur. Néanmoins, une aide méthodologique peut être apportée au travers d'un outil mathématique à disposition de l'utilisateur pour guider son choix des variables les plus pertinentes pour l'étape 3. En particulier, dans le cas où le nombre de variables d'analyse serait supérieure au nombre de sites échantillonnés, une analyse des dépendances entre variables pourra être effectuée. Cette approche présente l'intérêt d'identifier des variables au comportement proche d'après les données disponibles, c'est-à-dire suivant une fonction monotone similaire. Par conséquent, une variable peut être remplacée par une autre dans la classification, même si les experts ne voient pas de lien, car elles y auraient un comportement semblable. Ainsi, un bouclage avec les experts est effectué pour déterminer les variables d'analyse les plus pertinentes du point de vue de la classification et du jugement d'experts (Figure 8).

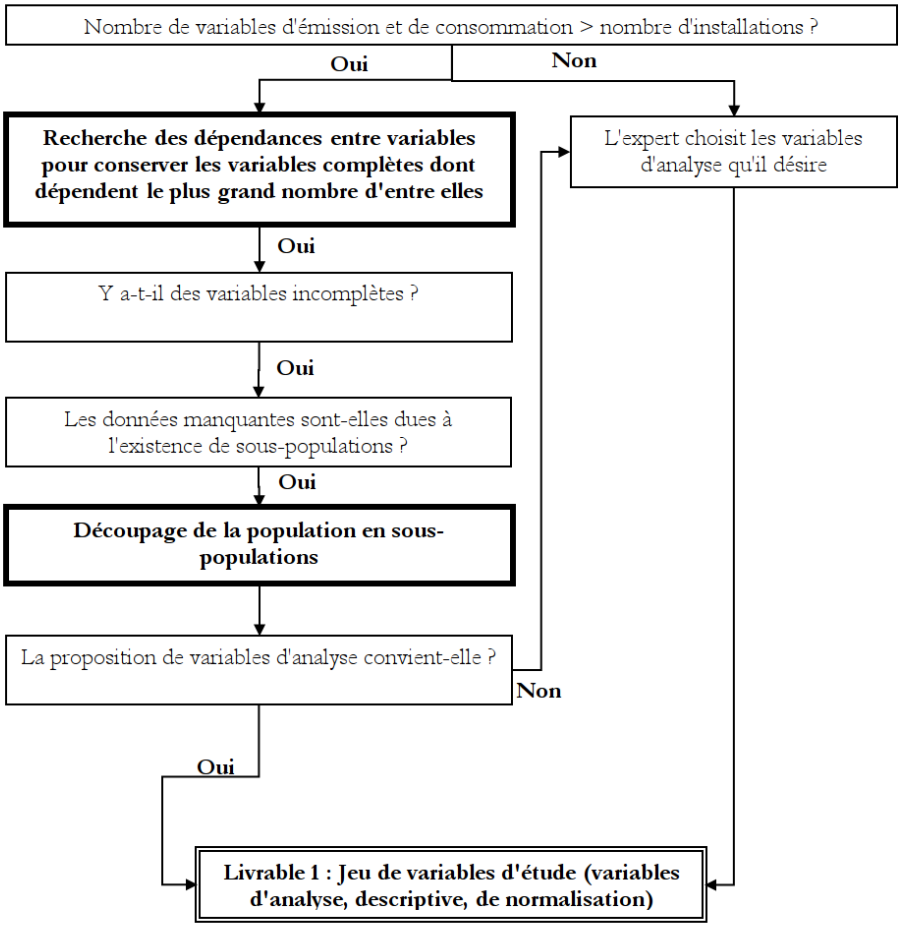


Figure 8 : Processus décisionnel sur l'aide au choix des variables d'analyse.

Parmi les tests statistiques disponibles (Ramousse *et al.*, 1996), les tests d'indépendance permettent de répondre de la manière la plus adéquate à l'identification d'un lien entre plusieurs variables. Il existe de nombreux tests selon la nature des données étudiées : variables quantitatives ou qualitatives, continues ou discrètes, ou encore issues d'une distribution déterminée (par exemple, loi normale) ou non. Afin de conserver une généricité dans notre approche, nous avons choisi, à ce stade, des tests non-paramétriques, c'est-à-dire qui ne présupposent pas que les données suivent une distribution particulière ou aient une relation connue (par exemple, corrélation linéaire). Ainsi, les tests statistiques employés font ici partie des tests non-paramétriques. Quatre approches répondent à la notion de tests non paramétriques. Le test de Spearman, de Pearson, de Wilcoxon et de Kruskal-Wallis (Tableau 5). Ces deux derniers tests sont applicables aux variables qualitatives, en fonction du nombre de modalités, c'est-à-dire les différentes valeurs qu'elles prennent.

Tableau 5 : Liste des tests statistiques non paramétriques disponibles.

Test statistique	Première variable	Seconde variable
Spearman	Quantitative	Quantitative
Pearson	Quantitative	Qualitative
Wilcoxon	Qualitative (2 modalités)	Qualitative (2 modalités)
Kruskal-Wallis	Qualitative (3 modalités et plus)	Qualitative (2 modalités et plus)

Du fait du caractère quantitatif obligatoire pour la catégorisation et la comparaison deux à deux des variables d'analyse, les tests de Wilcoxon, Kruskal-Wallis et Pearson ne peuvent pas être employés sur les variables d'analyse mais seront utilisés pour caractériser les classes représentatives obtenues à l'étape 3.

Le test de Spearman (Saporta, 2011) consiste à définir un coefficient de corrélation, en estimant l'importance de la relation entre deux variables. Cette relation peut être décrite par une fonction monotone affine ou non. Le test permet de calculer un coefficient de corrélation en fonction du classement des *n* individus d'une population en fonction d'une première variable X et de leur classement en fonction d'une seconde variable Y. Ainsi, la différence de rang d_i de chaque individu dans chacun de ces classements est utilisé pour calculer un coefficient de corrélation r_s (Équation 1). Si r_s vaut 0, les deux classements sont indépendants.

$$r_s = 1 - \frac{6\sum_i d_i^2}{n(n^2-1)} \qquad \text{Équation 1}$$

Ainsi, ce test est effectué pour chaque couple de variables d'analyse et le résultat est synthétisé au sein d'un tableau disponible et accessible aux experts (Tableau 6). Dans ce tableau, les niveaux de dépendance et indépendance sont simplifiés par l'intermédiaire de symboles « + » et « - ». Ainsi la symbolique (++) indique un fort niveau de dépendance (risque d'erreur inférieur à 1 %) ; (+) une dépendance significative à moins de 5 % de risque d'erreur ; (-) les variables indépendantes (risque d'erreur inférieur à 5 %) ; (--) pour les variables indépendantes présentant un risque d'erreur de moins de 1 %. Dans certains cas, le test de Spearman n'étant pas significatif, une cellule exempte de symbole est présentée.

Tableau 6 : Exemple de tableau montrant les dépendances et corrélations entre variables.

Variables	V1	V2	…	Vn
V1				
V2	++			
…	…	--		
Vn		+	-	

1.2.2. Variables de normalisation

Afin de pouvoir comparer des installations de tailles différentes, il est nécessaire de rapporter les variables d'analyse à une unité de référence. Cette opération de normalisation peut être effectuée en divisant les variables d'analyse par une autre variable commune à tous les sites. Dans le cas des sites industriels, les variables de normalisation peuvent indiquer une notion de « taille ». La taille des sites peut être de plusieurs ordres ; elle peut être basée sur des données socio-économiques comme le nombre de salariés, le chiffre d'affaires (INSEE, 2016), les quantités de matières utilisées pour la production ou encore le tonnage de produit sortant. De manière générale, la variable de normalisation doit correspondre à un flux de matière, d'énergie ou une donnée socio-économique qui sera considérée comme caractéristique de la taille des installations d'un secteur.

1.3. Livrable de l'étape 1

Le livrable obtenu à l'issue de cette première étape consiste en une base de données contenant les valeurs des variables d'analyse, descriptives et de normalisation (Tableau 7). Cette base est complétée par une indication du nombre et du pourcentage de données manquantes pour chacune des variables appartenant aux trois catégories évoquées. Les valeurs des variables quantitatives sont données, de même que les différentes valeurs des variables qualitatives (modalités).

Tableau 7 : Livrable 1 – Représentation de la base de données des variables d'analyse, descriptives et de normalisation à l'issue de l'étape 1 (la lettre X représentant une valeur quantitative, la lettre M une modalité d'une variable descriptive et le point d'interrogation une donnée manquante).

Population	Variable d'analyse 1	…	Variable d'analyse n	Variable descriptive 1	…	Variable descriptive m	Variable de normalisation
Installation 1	X	…	X	M	…	M	X
Installation 2	X	…	?	?	…	M	X
Installation 3	?	…	X	M	…	M	?
Installation 4	X	…	X	M	…	M	X
…	…	…	…	…	…	…	…
Installation n	X	…	X	M	…	M	X
Nombre de données disponibles	X	…	X	X	…	X	X
Nombre de données manquantes	X	…	X	X	…	X	X
% de données manquantes	X	…	X	X	…	X	X

2. Étape 2 : Traitement des données

Cette étape a pour finalité de consolider la base de données issue de l'étape 1, notamment par la reconstruction des données manquantes. La qualité des données reconstruites devant permettre une analyse précise compte tenu des outils statistiques employés, du nombre d'individus de la sélection ainsi que des variables d'analyse sélectionnées.

D'un point de vue strictement mathématique, il n'existe pas de règle permettant de définir quel niveau de données manquantes serait acceptable. Ce niveau étant apprécié au cas par cas selon le système étudié et les objectifs de l'étude. Bien entendu, la solution idéale consiste à disposer de données complètes pour l'ensemble des variables. Or, l'expérience montre qu'il est improbable de disposer de données exhaustives notamment dans le milieu industriel soumis à concurrence et détenant des données stratégiques voire confidentielles.

De ce fait la présence de données manquantes est inévitable et peut impacter de manière significative les étapes ultérieures de l'analyse. En effet, lors de la classification des installations (Étape 3), les données manquantes créent un biais en regroupant majoritairement les sites selon leur quantité de variables manquantes. La classe créée n'ayant alors aucune « réalité » ou similitude entre les sites, excepté le fait qu'ils soient pauvres en données. L'objectif de cette étape consiste donc à mettre en œuvre une méthode d'imputation statistique qui permette de « combler les vides » et donc à reconstruire les données manquantes avec la plus faible erreur possible. Trois outils statistiques ont été trouvés dans la littérature sur ce sujet. Dans la méthodologie proposée, un seul d'entre eux, qui est présenté plus en détails a été conservé : MissForest.

2.1. Outils statistiques explorés

Considérant que les données employées dans les variables d'analyse sont des valeurs quantitatives continues, il est nécessaire de trouver une méthode de reconstruction adaptée à ce type de données. Dans ce cadre, deux méthodes d'imputation simple : les K plus proches voisins (Altman, 1992) et l'imputation par la régression ; et une troisième méthode d'imputation multiple MissForest (Stekhoven et Buhlmann, 2012) basée sur les forêts aléatoires ont été explorées.

2.1.1. Méthode d'imputation simple : K plus proches voisins

Cet outil consiste à remplacer les données manquantes par des valeurs provenant d'individus similaires (les plus proches voisins de la valeur manquante) pour lesquels toute l'information est disponible. Par conséquent, appliquer cette méthode à des installations industrielles supposerait de disposer de données complètes pour plusieurs d'entre elles.

2.1.2. Méthode d'imputation simple : Imputation par la régression

Cette imputation consiste à remplacer les valeurs manquantes par des valeurs prédites selon un modèle de régression (Hastie *et al.*, 1999). Dans ce cas, il est nécessaire de pouvoir trouver un modèle de régression efficace et permettant de décrire les données.

2.1.3. Méthode d'imputation multiple : MissForest

Cet outil (Stekhoven et Buhlmann, 2012) détermine les relations entre les variables quantitatives ou qualitatives, quelles qu'elles soient, par un algorithme d'apprentissage basé sur les forêts aléatoires. Il est précisé par ailleurs (Glasson-Cicognani et Berchtold, 2010) que cet outil est particulièrement intéressant pour l'estimation des valeurs manquantes en conservant l'écart type. Ses principes théoriques sont présentés en annexe 1.

Compte-tenu de la contrainte initiale visant à disposer de plusieurs individus complets pour appliquer les K plus proches voisins et de la nécessité de présenter des relations entre variables linéaires pour appliquer l'imputation par régression, ce qui n'est *a priori* pas certain avec les données environnementales issues d'installations industrielles, ces méthodes n'apparaissent pas adéquates. MissForest, quant à elle, s'avère être la solution permettant une applicabilité la plus large possible (Tableau 8). L'intérêt pour MissForest est confirmé par la comparaison des erreurs de reconstruction de trois méthodes d'imputation effectuée par Stekhoven et Buhlmann (2012) et qui souligne l'erreur plus faible obtenue avec MissForest sur dix jeux de données différents composés de 18 à 12 626 variables et de 40 à 595 individus avec des données quantitatives, qualitatives ou mixtes. Miss Forest apparaît donc être l'outil disposant de la plus large applicabilité

pour la reconstruction des données tout en conservant la variance entre individus (ce qui sera utilisé dans la classification représentative de l'étape 3) avec la plus faible erreur.

De plus, MissForest permet de calculer l'erreur de la reconstruction en même temps que l'imputation a lieu (*out of bag error*) en utilisant une partie des arbres de la forêt aléatoire. Cette seconde approche permet d'obtenir une estimation de l'erreur très proche de l'erreur vraie. Cela a été démontré par Stekhoven et Buhlmann (2012) en comparant avec celle issue d'une analyse croisée. Ce calcul d'erreur de la reconstruction s'avère utile dans l'intégration de MissForest à la méthodologie proposée en tant qu'outil d'aide au choix des variables d'analyse et de la population d'analyse. En effet, l'utilisateur expert pourra choisir le niveau d'erreur acceptable pour la reconstruction, ce qui lui permettra d'arbitrer entre quantité de données exclues et précision de reconstruction.

Tableau 8 : Comparatif des outils statistiques possibles pour la reconstruction des données de l'étape 2.

Outil statistique	Avantages	Inconvénients	Applicabilité
K plus proches voisins	Se base sur l'ensemble des données disponibles.	Requiert quelques individus complets.	Variables quantitatives.
Imputation par la régression	Rapide et visuel.	Requiert une relation linéaire entre variables	Variables quantitatives.
MissForest	Pas de nécessité d'une relation linéaire entre variables. Faible erreur. Conserve l'écart-type.	Temps de calcul important.	Utilisable sur des variables qualitatives et quantitatives.

2.2. Utilisation de Miss Forest dans la méthodologie

L'utilisation de cet outil statistique s'opère en deux temps. Une première sous-étape consiste à estimer l'erreur globale de la reconstruction et de chaque variable prise individuellement en fonction du paramétrage qui a été défini à l'étape 1. Par la suite, les erreurs de reconstruction de chaque variable (croix dans la Figure 9) sont couplées au pourcentage de données manquantes (histogrammes) afin d'être présentées graphiquement à l'utilisateur. Pour chaque variable et pour l'ensemble de la reconstruction, les erreurs quadratiques moyennes normalisées sont calculées (Stekhoven et Bühlmann, 2012) avec la formule indiquée dans l'équation 2, avec X^{vrai} la valeur réelle prise par la donnée et X^{imp} la valeur imputée. Dans le cas présent, la valeur réelle est remplacée par les valeurs générées par les premiers arbres de la forêt.

$$NRMSE = \sqrt{\frac{moyenne\left(\left(X^{vrai} - X^{imp}\right)^2\right)}{var\left(X^{vrai}\right)}}$$ Équation 2

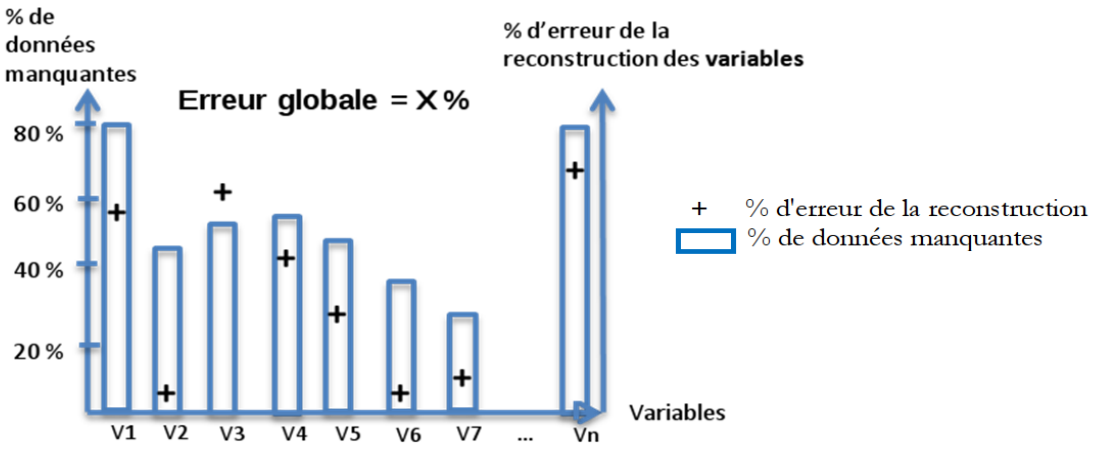


Figure 9 : Outil d'aide à la décision pour la sélection des variables proposé pour l'étape 2.

L'utilisateur a donc ici une nouvelle fois la possibilité d'exclure certaines variables. Son choix d'exclusion devant être dicté simultanément par l'erreur de la reconstruction, l'impact potentiel sur l'analyse et l'enjeu environnemental porté par la variable. Pour autant, avant d'exclure définitivement une variable complète de l'analyse, une alternative proposée par notre approche vise à tenter de réduire l'erreur de la reconstruction par la mise de côté de sites ou de variables. Cette approche purement mathématique ne s'affranchit cependant pas du jugement d'experts quant à l'impact qu'une telle décision pourrait produire.

Cette procédure de tri des installations afin de minimiser l'erreur globale de la reconstruction est décrite en annexe 2. Les résultats sont, quant à eux, transmis au travers d'une représentation graphique à l'expert afin de l'aider dans sa décision (Figure 10). Ainsi, la suppression de sites (population suggérée 1) comparé à la population initiale peut permettre une réduction globale de l'erreur comme constaté dans la Figure 10. Ce processus itératif pouvant être renouvelé autant de fois que nécessaire.

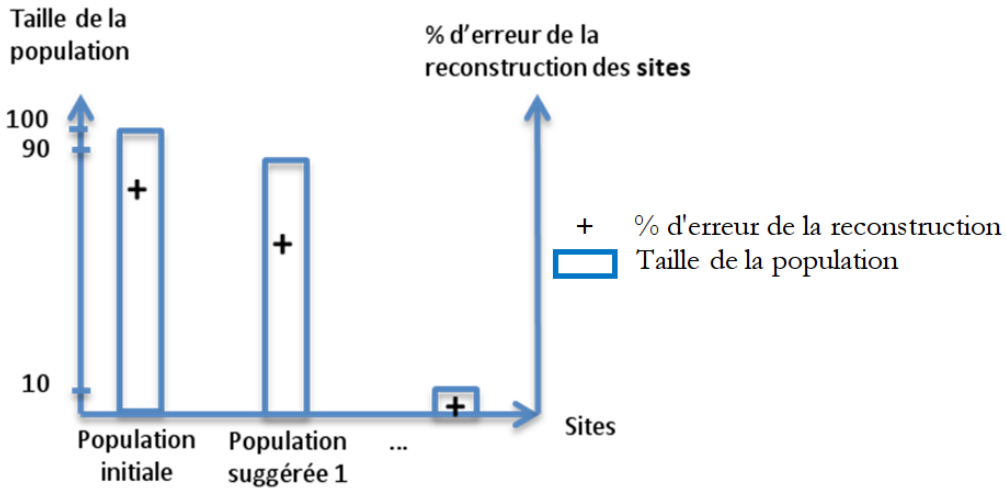


Figure 10 : Outil d'aide à la décision pour la sélection de la population proposé pour l'étape 2.

Finalement, une comparaison des erreurs par variables pour les deux populations permet de constater de leur évolution (Figure 11) et de décider de la pertinence du changement de population. Dans les cas où l'erreur de la reconstruction de certaines variables demeurerait élevée, il faudrait que l'utilisateur envisage leur transfert dans les variables descriptives, voire une exclusion définitive.

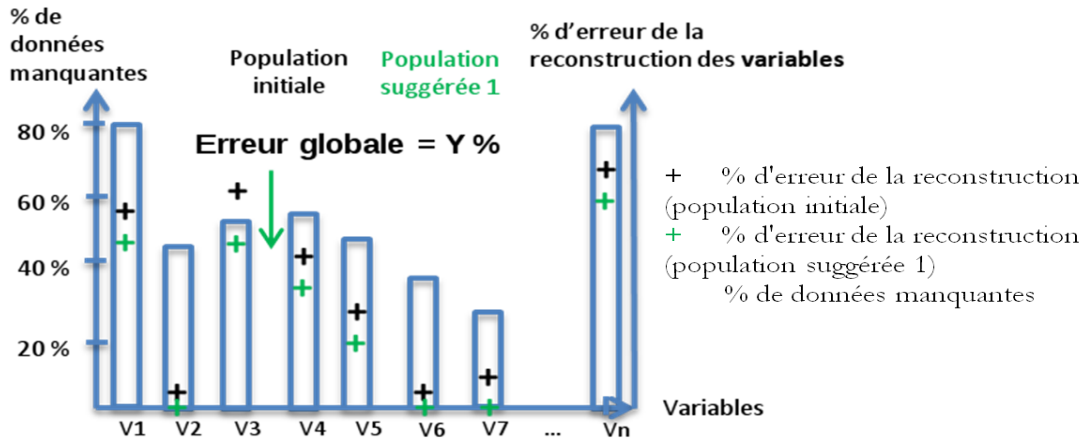


Figure 11 : Outil d'aide à la décision pour la validation d'une nouvelle population pour l'étape 2.

2.3. Implication des acteurs

Le rôle des acteurs dans l'étape 2 est important. Sur la base des informations proposées par la méthodologie (erreur de la reconstruction, pourcentage de données manquantes et taille de la population), l'expert valide la population et les variables finales d'analyse qui seront utilisées dans la suite de l'étude au regard du niveau de précision qu'il juge acceptable. Il peut également effectuer un retour à l'étape 1 pour affiner son choix de variables d'analyse au regard des erreurs de la reconstruction et chercher ainsi à la minimiser tout en optimisant la population d'étude et les variables d'analyse qui serviront à la classification.

2.4. Livrable de cette étape

Le livrable de cette étape consiste en une version 2 de la base de données issue de l'étape 1 (livrable 1). Cette base étant consolidée, c'est-à-dire que les valeurs manquantes des variables d'analyse ont été reconstruites, ainsi que par l'apposition de l'erreur associée par variable d'analyse (Tableau 9).

Tableau 9 : Livrable 2 – Représentation de la base de données des variables d'analyse reconstruites, descriptives et de normalisation à l'issue de l'étape 2 (la lettre X représentant une valeur quantitative, R une valeur reconstruite, M une modalité d'une variable descriptive et le point d'interrogation une donnée manquante).

Population	Variable d'analyse 1	…	Variable d'analyse n	Variable descriptive 1	…	Variable descriptive m	Variable de normalisation
Installation 1	X	…	X	M	…	M	X
Installation 2	X	…	Xr	?	…	M	X
Installation 3	Xr	…	X	M	…	M	?
Installation 4	X	…	X	M	…	M	X
…	…	…	…	…	…	…	…
Installation n	X	…	X	M	…	M	X
Nombre de données disponibles	X	…	X	X	…	X	X
Nombre de données manquantes	X	…	X	X	…	X	X
% de données manquantes	X	…	X	X	…	X	X

2.5. Remarques sur les résultats obtenus à l'étape 2

Cette étape permet de mettre en évidence d'éventuelles données critiques pour lesquelles des campagnes de collecte supplémentaires seraient nécessaires. D'une part, dans le cas des variables d'analyse, le choix des experts pourrait être remis en question compte tenu de l'erreur introduite par des variables incomplètes et indépendantes des autres. En effet, l'utilisation de variables qui seraient parfaitement indépendantes rendrait l'imputation inutilisable et les données manquantes deviendraient critiques pour l'application de la méthodologie. D'autre part, dans le cas de la population, des installations induisant une erreur importante pourraient être indiquées comme étant des sites susceptibles de faire l'objet d'un traitement particulier ou devant faire l'objet d'un complément d'information pour être utilisables dans l'étape suivante (Étape 3).

3. Étape 3 : Classification des installations

Cette étape débute par le choix de la définition de la notion d'installation de référence qui sera utilisée par la suite. En effet, deux approches sont proposées dans ce travail de recherche. D'une part, les installations représentatives pour lesquelles la « représentativité » est définie comme le fait de « *[posséder] les caractéristiques principales de l'ensemble dont il relève, et peut, à ce titre, le représenter* » (Larousse, 2016). Et d'autre part, les installations performantes basées sur la définition de l'article 3 de l'IED qui définit le terme « meilleure » de MTD et qui, transposé aux installations, sont celles qui « *permettent un haut niveau de protection de l'environnement dans son ensemble* ».

Les deux définitions proposées pour les sites de référence impliquent des conséquences majeures pour l'étape 3, notamment concernant l'application d'outils mathématiques différents.

L'utilisateur doit donc, à ce stade, choisir la définition la plus adéquate selon ses objectifs faisant de cette étape un élément fondamental de la méthodologie présentée.

3.1. Approche représentative

Les sites de référence, au travers de l'approche représentative, sont les sites dont les caractéristiques permettent de représenter l'ensemble de la population. De ce fait, les installations sont classées en groupes partageant des caractéristiques communes. Dans cette optique, deux types de méthodes de classification existent (Gonzalez, 2008) et peuvent être utilisées pour la classification représentative : celles dites hiérarchiques et celles dites non hiérarchiques. Deux méthodes de classification non-hiérarchiques ont été identifiées : les centres mobiles et les nuées dynamiques, ainsi qu'une méthode hiérarchique : la classification ascendante hiérarchique.

3.1.1. Centres mobiles

Cette approche consiste à placer aléatoirement des centres de classe dans un nuage de points constitué des installations en utilisant les variables d'analyse comme axes descriptifs. Par la suite les centres sont déplacés afin de minimiser la distance entre les sites d'une classe. Dans cette approche, il faut donc fixer à l'avance le nombre de classes souhaité et le centre représente alors un individu fictif.

3.1.2. Nuées dynamiques

Les nuées dynamiques sont une généralisation des centres mobiles et suivent le même principe sauf, qu'au lieu de fixer un centre fictif, l'idée est d'associer un individu comme représentant de sa classe qui soit le plus proche de son centre de gravité. Cet individu serait donc ainsi le plus représentatif de sa classe.

3.1.3. Classification ascendante hiérarchique

Dans le cas de la classification ascendante hiérarchique, les individus sont groupés 2 à 2 en fonction d'un critère d'agrégation, ce qui permet d'obtenir une représentation en arbre ou dendrogramme (Figure 12). Cet arbre est alors coupé au niveau du « saut » le plus significatif (ligne pointillée dans la Figure 12) permettant alors de créer les classes (chacune des branches coupées par la ligne en pointillée ; 2 dans cet exemple). Une approche recommandée est d'utiliser le critère de Ward qui consiste à minimiser la variance intra-classe et à maximiser la variance inter-classe. De plus, le passage par une analyse factorielle comme l'Analyse en Composantes Principales (Saporta, 2011) permet un changement de référentiel optimisant la classification à partir de ce critère (une description en est donnée en annexe 3). Ainsi, les individus sont positionnés en un

nuage de points dont les axes sont les composantes principales, ce qui permet de les analyser en fonction de leurs positions relatives. Le nombre de classes n'est, dans ce cas, pas fixé à l'avance mais proposé à l'utilisateur afin de maximiser la dissemblance entre classes, c'est-à-dire l'inertie inter-classe. Une classe est considérée comme suffisante lorsqu'elle se compose d'un nombre minimum d'installations défini par l'utilisateur. En effet, si par exemple, deux classes sont obtenues et que l'une d'entre elles ne contient qu'un ou deux individus, ceux-ci sont susceptibles d'être considéré comme des cas particuliers plutôt que comme les représentants d'une classe bien distincte. Dans un tel cas, ces installations peuvent être retirées de l'analyse en cours et une nouvelle classification réalisée. La condition d'arrêt pour la classification est définie par l'obtention d'au-moins deux classes peuplées *a minima* d'un nombre de sites fixé par les acteurs en fonction de ce qu'il considère être une classe suffisamment peuplée. Au regard des observations des pratiques de sélection effectuées auprès de l'industrie laitière française dans la partie 1 de ce mémoire, une règle par défaut fixée arbitrairement pour l'arrêt de la classification a été définie à 10 % de la population totale échantillonnée. Néanmoins, cette valeur peut être modifiée par l'utilisateur selon ses objectifs et son interprétation des classes.

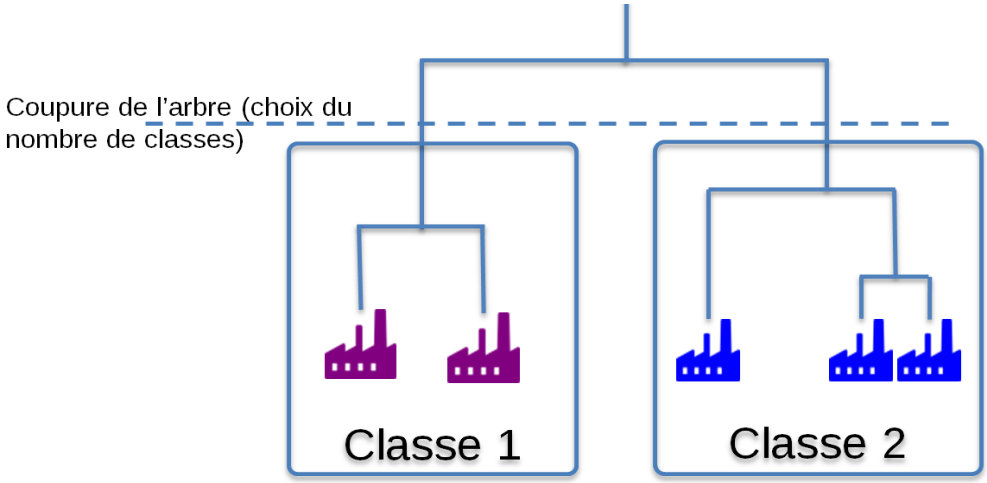


Figure 12 : Exemple de dendrogramme obtenu par la classification ascendante hiérarchique avec découpage à deux classes.

Les avantages et inconvénients de ces trois outils sont résumés dans le Tableau 10.

Tableau 10 : Comparatif des outils statistiques pour la classification des sites représentatifs.

Outil statistique	Avantages	Inconvénients	Applicabilité
Centres mobiles	Nécessite peu de calculs.	Installation fictive. Nombre de classes fixé *a priori*.	Variables quantitatives.
Nuées dynamiques	Un site représentatif des classes est choisi.	Nombre de classes fixé *a priori*.	
Classification Ascendante Hiérarchique	Nombre de classes optimal suggéré. Installations réelles représentent les classes.	Nombre de calculs importants.	

Dans le cas de la recherche des installations représentatives, il est nécessaire d'obtenir une installation réelle qui corresponde à la classification optimale compte tenu des données disponibles. Par conséquent, la classification ascendante hiérarchique est le choix qui permet le regroupement des installations. De plus, son association avec une analyse en composantes principales fournit une représentation des différences entre installations par la variance des différentes variables d'analyse. Ainsi, les sites sont classés en fonction des similitudes entre les valeurs de leurs caractéristiques environnementales. La représentativité est alors définie comme la distance au centre de la classe dans l'espace factoriel constitué des composantes principales. Une contrainte sur les données existe pour l'application de cet outil il est nécessaire de disposer de plus d'individus que de variables d'analyse ou plus généralement de maximiser le nombre d'installations tout en minimisant celui des variables d'analyse. Cela peut donc impliquer de devoir réduire le nombre de variables pour pouvoir appliquer cette approche et donc de réitérer les étapes 1 et 2.

3.2. Approche performante

Les sites de référence, au travers de l'approche performante, sont définis comme les sites présentant la meilleure performance en considérant simultanément l'ensemble des variables environnementales quantitatives analysées. Par conséquent, dans cette approche les *optima* (c'est-à-dire les sites minimisant l'ensemble des valeurs sur chacune des variables d'analyse) sont recherchés dans la population. Cette approche correspond à un problème d'optimisation multicritère et de ce fait nous le résoudrons par l'utilisation des optimums (front) de Pareto (Annexe 4).

Le front de Pareto consiste à rechercher des installations optimales comme, par exemple, les installations ayant les valeurs les plus faibles possibles pour les variables d'analyse. Cette approche est décrite comme étant la meilleure option pour trouver des compromis et est utilisée comme

critère de classement dans des méthodes d'aide à la décision multicritère (Perez Gallardo, 2013). Le nombre d'individus optimaux dépend du nombre de variables d'analyse considéré et peut, dans certains cas, être élevé. Les individus se trouvant sur le front sont des optima et ne sont donc, dans l'absolu, pas différentiables les uns des autres. Il est à noter à ce stade, que les sites optimaux présentent le meilleur compromis de performance sur l'ensemble des variables et ne sont donc par conséquent pas les meilleurs sites sur chaque variable pris indépendamment.

L'optimisation multi-objectif, comme le front de Pareto répond à une problématique qui vise à satisfaire des besoins contradictoires. Ainsi, il n'est pas toujours possible de trouver un site présentant des valeurs faibles sur l'ensemble des variables d'analyse ; il est même possible qu'un site présente une valeur élevée par rapport aux autres installations alors qu'il a des valeurs relativement faibles sur les autres variables. Dans les cas où il faut chercher un compromis dans la minimisation des variables d'analyse en sachant qu'un cas idéal est très peu probable, une telle approche fournit un moyen de distinguer les sites qui pourraient être qualifiés de « globalement performants » d'un point de vue environnemental et donc répond à notre définition d'approche performante « *permettent un haut niveau de protection de l'environnement dans son ensemble* ». À l'issue de cette classification, une liste de sites optimaux (performants) est obtenue.

3.3. Complémentarité des deux approches

Bien que les deux approches représentative et performante aient été montrées comme distinctes, il est possible de les coupler. D'une part, une première classification représentative peut permettre d'identifier des classes distinctes d'installations et, au sein de celles-ci, une classification performante peut aboutir à identifier les plus performantes compte tenu de caractéristiques de consommations et d'émissions différenciées. D'autre part, il est possible de rechercher des classes représentatives distinctes parmi un groupe d'installations performantes afin de voir quelle différenciation serait nécessaire. Le premier cas énoncé ici a été appliqué aux installations nucléaires de base internationales, dans la partie 3.

3.4. Caractérisation des groupes identifiés

Les classes obtenues (classes représentatives ou installations performantes) sont ensuite analysées au regard des valeurs des variables d'analyse pour trouver les plages de valeurs qui les caractérisent et éventuellement aboutir à un traitement différencié des installations en fonction de celles-ci. La représentation proposée ici est la boîte à moustaches de Tukey qui regroupe des informations de quartiles et de dispersion par variable (Figure 13). Ainsi, la ligne se trouvant à l'intérieur des boîtes représente la médiane et renseigne sur le milieu de la série ; l'épaisseur des deux parties de la boîte rendent compte de la dispersion des valeurs situées au centre de la série (3

premiers quartiles) ; la longueur des moustaches renseigne sur la dispersion des valeurs situées au début de la série ordonnée (les valeurs les plus petites correspondant à 25 % des observations) ou à la fin de celle-ci (les valeurs les plus grandes correspondant aussi à 25 % des observations) ; l'information fournie par les moustaches est paramétrable et peut également consister à indiquer les valeurs minimales et maximales. Généralement, la boîte et les moustaches seront d'autant plus étendues que la dispersion de la série statistique est grande.

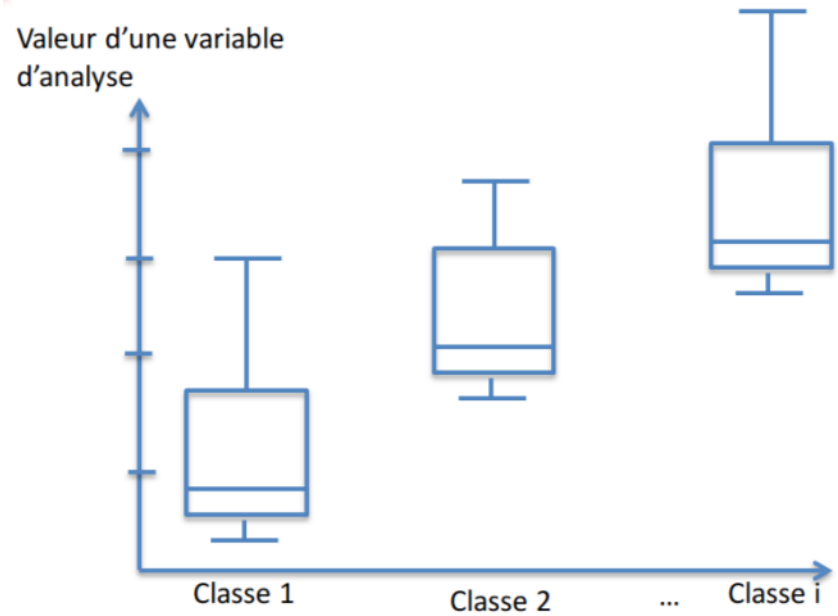


Figure 13 : Exemple de boîtes à moustaches de Tukey représentant les valeurs prises par les sites de différentes classes pour une variable d'analyse.

À partir de ce tri des installations dans différentes classes représentant différents niveaux de consommations et d'émissions, afin de mieux comprendre les variables qui caractérisent les classes et afin de vérifier que le choix de variables d'analyse est pertinent, une caractérisation sur les variables descriptives est réalisée. Ainsi, pour chaque groupe, des tests d'hypothèses Kruskal-Wallis (Tableau 5), le test le plus adapté pour comparer deux variables qualitatives (1–le n° de la classe ; 2–une des variables descriptives) sont réalisés sur les variables. On obtient ainsi une matrice des dépendances entre variables descriptives et classes (Tableau 11). Il est à noter que l'objectif est ici de mettre en avant les dépendances, à l'inverse du Tableau 6.

Tableau 11 : Livrable 3.1 – Dépendances entre le classement représentatif et les variables descriptives.

Variables	Vdesc1	Vdesc2	…	Vdescn
Classement représentatif	++		--	-

Les dépendances ou indépendances trouvées doivent ensuite être validées par jugement d'experts afin de ne garder que les plus intéressantes et d'éviter les corrélations sans réalité physique.

3.5. Livrable de l'étape

Quelle que soit l'approche choisie (représentative ou performante), le résultat de l'étape est une liste d'installations (Tableau 12). Dans l'approche représentative, les classes obtenues sont représentatives de niveaux de consommations et d'émissions spécifiques à la classe qui devraient donc faire l'objet d'un traitement, voire de valeurs de consommations et d'émissions différenciées par typologie de sites (une typologie étant reliée à une classe). En revanche, les installations performantes sont à comparer à l'ensemble de la population, sans traitement différencié possible, si aucune classification représentative n'a été effectuée.

Tableau 12 : Livrable 3.2 – Liste des installations représentatives ou performantes.

Classe 1 ou sur le front de Pareto	Classe 2 ou hors du front de Pareto	…	Classe m
Installation 1	Installation 13	…	Installation 10
Installation 2	Installation 20	…	Installation 11
Installation 3	Installation 7	…	Installation19
…	…	…	…

4. Étape 4 : Analyse de la classification

À partir du livrable précédent, l'objectif de cette étape est de prioriser par ordre de représentativité ou de performance les sites pour objectiver la sélection des installations de référence au sein de chacune des classes, voire au sein des installations présentes sur le front de Pareto (si leur nombre est estimé comme trop conséquent). Une règle de sélection de 10 % des effectifs de chaque classe avec un minimum de 3 installations de référence est proposée par défaut. Cette étape est un développement original de cette thèse et est innovant par rapport à la littérature existante.

4.1. Cas des sites représentatifs

Dans le cas des sites représentatifs, la classification ascendante hiérarchique permet d'affecter à chaque installation une distance au centre de sa classe (parangon) et une distance au centre de gravité des autres classes (typicité) qui, via l'ACP réalisée, renferment toutes les spécificités des variables d'analyse. Dans chaque classe, l'installation la plus représentative correspond donc à celle qui présente la plus faible distance intra-classe et la distance inter-classe la plus élevée. Il est donc

nécessaire dans une telle approche d'arbitrer entre minimisation de la distance intra-classe ou maximisation de la distance inter-classe.

Toutes les installations d'une même classe constituent un nuage de points dans un plan ayant pour axes les distances intra-classes et inter-classes (Figure 14). Afin de guider la décision, ce plan est divisé en secteurs correspondant aux trois premiers quartiles de chaque axe. Afin de guider le choix de l'expert, chaque secteur est numéroté dans un ordre de préférence pour la sélection correspondant aux installations les plus représentatives (minimisation de la distance intra-classe) et les plus typiques (maximisation de la distance inter-classe). Ainsi, pour l'exemple fictif de la Figure 14, les deux sites les plus représentatifs seraient ceux qui se trouvent dans le secteur n°1. Si ce nombre de sites dans cette première zone est considéré insuffisant par l'utilisateur, il est alors recommandé d'ajouter à sa sélection les 3 sites de la zone n°2, le site du secteur n°3 et ainsi de suite jusqu'à obtention du nombre d'installations représentatives souhaité.

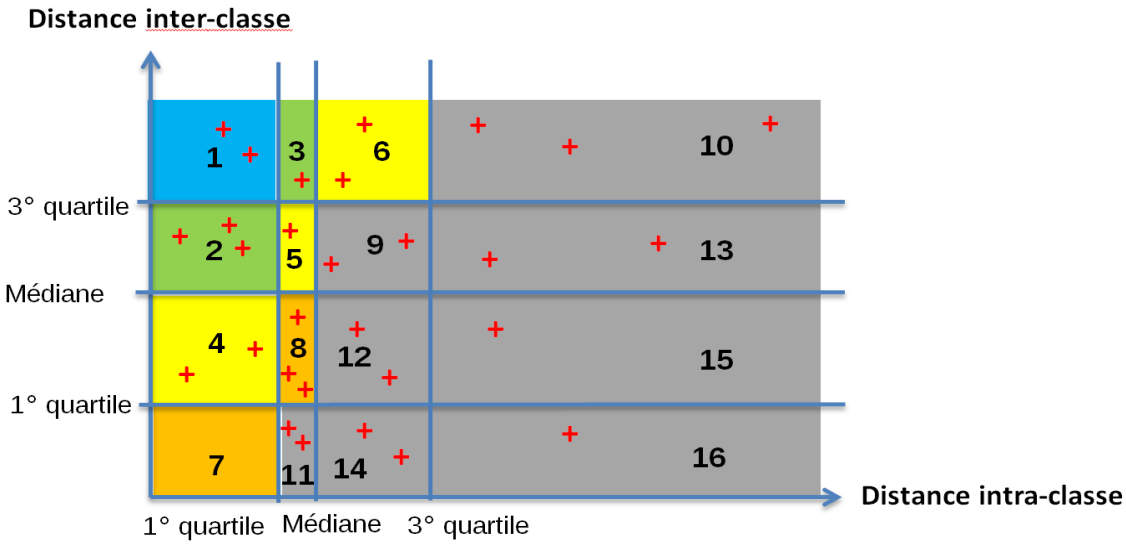


Figure 14 : Outil d'aide à la sélection des installations représentatives.

4.2. Cas des sites performants

Dans le cas de l'approche performante et comme mentionné précédemment, les sites situés sur le front de Pareto sont considérés comme « équivalents » en termes de performance globale. Il n'est donc *a priori* pas possible de les différencier au travers de cette méthode. Pour autant dans le cas où l'utilisateur souhaiterait réduire le nombre de sites parce qu'il trouve la lecture du résultat difficile, une sélection doit être envisagée. La solution proposée dans le cadre de ce travail consiste à

donner une note à chaque installation sur la base de l'ensemble des niveaux de consommations et d'émissions des variables d'analyses.

Pour pallier le problème d'addition d'unités différentes, nous proposons d'utiliser les coordonnées factorielles issues d'une ACP, ainsi que l'inertie de chaque composante principale comme coefficient de pondération afin d'obtenir un « Indicateur de Niveau de Performance Environnementale » (INPE) qui permette de hiérarchiser les installations présentes sur le front de Pareto au regard de l'ensemble des variables d'analyse.

Ainsi, l'INPE est proposé comme une aide à la différenciation des sites se trouvant sur le front de Pareto. La méthode affecte une note globale à chaque site sur la base de leurs émissions et consommations rapportée à leurs coordonnées factorielles issues d'une analyse en composantes principales avec une éventuelle réduction du nombre de composantes principales pour limiter le risque de compensation.

Quatre étapes sont nécessaires :

1. une analyse factorielle permet de positionner les sites les uns par rapport aux autres dans un espace factoriel, ce qui a l'avantage de remplacer les valeurs des variables par des coordonnées sans unité ;
2. une réduction du nombre de dimensions aux axes apportant au-moins 80 % de l'information est effectuée tout en vérifiant que le risque de compensation soit faible ;
3. l'INPE est calculé en fonction des coordonnées de chaque site et selon la part d'information portée par chaque axe ;
4. le tri croissant des sites en fonction de l'INPE permet de classer les sites du plus performant globalement à celui qui l'est le moins et de sélectionner un nombre défini de sites.

L'INPE est calculé sur la base des coordonnées factorielles d'une ACP afin d'utiliser des données sans dimension. L'approche proposée est une combinaison linéaire de coordonnées de plusieurs axes pondérée par la part d'information qu'apporte chaque axe factoriel (Équation 3).

$$K_i = \sum_{j=1}^{m} w_j \cdot c_i^j \qquad \textbf{Équation 3}$$

Où K_i est l'INPE pour l'individu i, c_i^j est la coordonnée factorielle sur l'axe *j* de l'individu *i* et w_i correspond à la part d'inertie apportée par chaque axe factoriel.

Cet indicateur fournit une mesure globale de performance environnementale en prenant en compte toutes les variables. Cependant, il y a un risque de compensation des valeurs des différents indicateurs si certaines variables croissent ensemble alors que d'autres décroissent. Pour limiter ce risque, une analyse des corrélations entre les variables d'analyse est nécessaire afin, d'une part de vérifier qu'il y a peu ou pas de corrélations négatives qui entraîneraient un phénomène de compensation, et d'autre part de réduire le nombre de composantes principales en arbitrant entre quantité d'information utilisée et impact d'éventuelles compensations.

La liste classée des installations dans l'ordre croissant de performance est fournie à l'utilisateur à qui appartient la décision finale de restreindre la sélection ou non des sites performants. Il est à noter qu'un bouclage avec les calculs des taux d'effort de l'étape 5 peut être nécessaire pour s'assurer de la représentativité des sites proposés dans le cas où la sélection serait très contraignante sur certaines variables et donc pour la définition de possibles valeurs d'émissions.

4.3. Livrable de l'étape 4

Le livrable de cette étape est semblable pour les deux approches (représentative et performante) et est constitué d'une liste ordonnée d'installations (Tableau 13). Néanmoins, l'ordre des installations des deux approches n'est pas le même étant donné que ces listes sont issues de deux approches aux objectifs et définitions différents. L'utilisateur sélectionne ensuite le nombre de sites qui lui convient avant de passer à l'étape d'interprétation de la sélection (étape 5).

Tableau 13 : Livrables 4.1 – Listes ordonnées des sites pour les approches représentatives et performantes.

Classe 1 ou sur le front de Pareto			Classe 2 ou hors du front de Pareto			…	Classe m		
Installation	**Distance intra-classe**	**Distance inter-classe**	**Installation**	**Distance intra-classe**	**Distance inter-classe**	**…**	**Installation**	**Distance intra-classe**	**Distance inter-classe**
Installation 1	Distance intra 1	Distance inter 1	Installation 13	Distance intra 13	Distance inter 13	…	Installation 10	Distance intra 10	Distance inter 10
Installation 2	Distance intra 2	Distance inter 2	Installation 20	Distance intra 20	Distance inter 20	…	Installation 11	Distance intra 11	Distance inter 11
Installation 3	Distance intra 3	Distance inter 3	Installation 7	Distance intra 7	Distance inter 7	…	Installation19	Distance intra 19	Distance inter 19
…	…	…	…	…	…	…	…	…	…

5. Étape 5 : Interprétation de la sélection

Cette étape vise à fournir des informations synthétiques à l'utilisateur pour l'aider à comprendre la ou les sélections issues des étapes précédentes et lui permettre de décider de leur intérêt et de leur pertinence au regard de ses besoins ou éventuellement de modifier le paramétrage dans une nouvelle étude.

Deux outils sont proposés pour interpréter les sélections effectuées à l'étape précédente et apporter des éléments d'aide à la décision pour les experts. D'abord une aide à la détermination des valeurs d'émissions de référence au travers de l'évaluation d'un taux d'effort permettant d'estimer l'impact qu'auraient ces valeurs sur le parc d'installations actuel selon l'échelle étudiée (nationale ou internationale). Ensuite, une aide à l'identification des MTD via une étude des liens existants entre les classes trouvées, les techniques mises en œuvre par les installations pour chaque variable d'analyse et la variable de normalisation, ainsi que la valeur de référence choisie.

5.1. Détermination des valeurs de référence et du « taux d'effort » des installations

Les installations de référence qui ont été sélectionnées, dans le cas de Processus de Séville, vont contribuer à définir des valeurs de référence pour les variables d'analyse. Dans ce cas, les autres installations dont les données ont été collectées devraient alors respecter ces valeurs. Par conséquent, plusieurs valeurs de référence correspondant aux différentes classes sont estimées (minimum, maximum, médiane, moyenne). Ces informations sont utilisées pour simuler l'impact d'éventuelles valeurs réglementaires sur la population d'installations et ainsi vérifier la représentativité des sites sélectionnés. Cette approche pourrait également être appliquée à des valeurs de référence définies au niveau européen sur le parc industriel français afin d'estimer le nombre d'installations qui auraient un effort (technique et économique) à fournir pour atteindre celles-ci. Afin de les choisir en connaissance de cause, une estimation des conséquences qu'aurait le choix d'une valeur plutôt qu'une autre est proposée à l'utilisateur et aux acteurs du processus décisionnel. Cette estimation basée sur la notion de taux d'effort, exprimé en pourcentage, présente la proportion de sites dont les valeurs actuelles de consommations et d'émissions seraient supérieures à celle proposée, c'est-à-dire celles ne respectant pas les valeurs définies (Équation 4).

$$\forall V_i, \forall Classe\ j, T_{eff}(i,j) = \frac{N_{sites}[x_i > Vref_i]}{N_{sites}[Classe\ j]} \qquad \text{Équation 4}$$

avec V_i la variable d'analyse i, Classe j, la classe considérée, $T_{eff}(i, j)$ le taux d'effort pour la variable i et la classe j, $N_{sites}[x_i > V_{ref}(i, j)]$ le nombre de sites dont la valeur x_i pour la variable d'analyse V_i est

supérieure à la valeur de référence pour la variable i et la classe j et N_{sites}[Classe j] le nombre de sites de la classe j.

La Figure 15 représente l'outil d'aide à la décision proposé pour identifier les éléments relatifs à la détermination des valeurs de référence (minimum, maximum, médiane, moyenne pour les sites de référence sélectionnés à l'étape précédente), les taux d'effort associés et la position de chaque installation d'une classe ou de la population ainsi qu'une indication des installations de référence et des installations reconstruites. Cette figure permet de choisir la valeur de référence en fonction du nombre de sites en deçà ou au-delà de chacune.

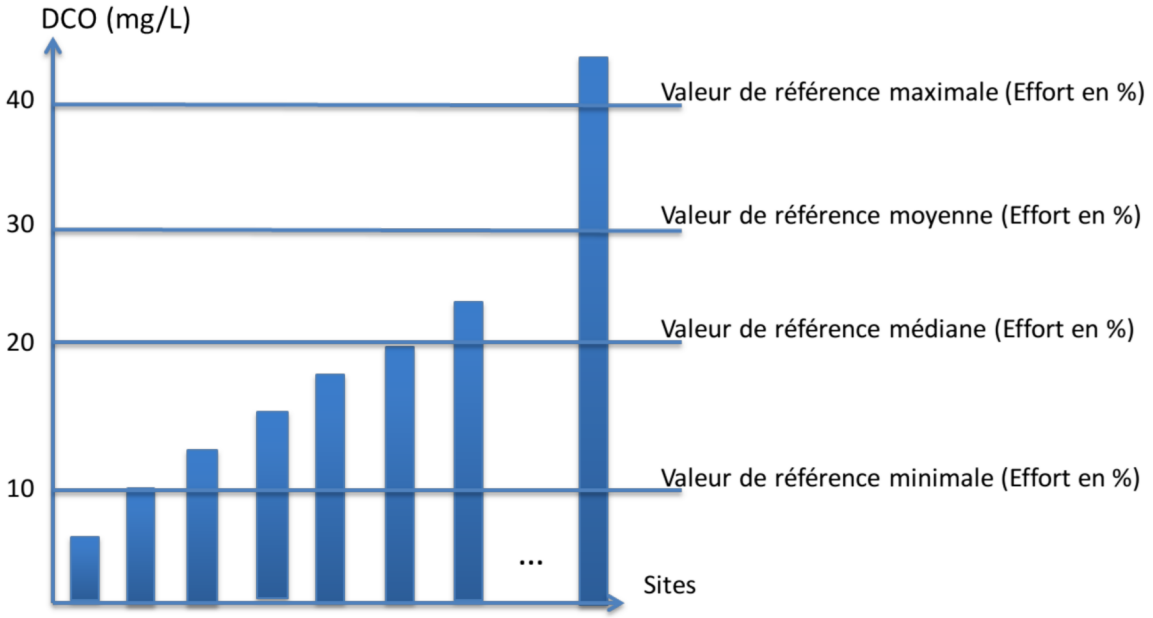


Figure 15 : Aperçu de l'outil d'aide à la décision pour le taux d'effort (exemple de la concentration en DCO).

Finalement, le livrable de cette étape se présente sous la forme du Tableau 14 afin d'obtenir une vision globale des décisions sur le secteur étudié. Pour chacune des classes identifiées à l'étape 3, les valeurs de référence issues des sites de référence trouvés à l'étape 4 sont précisées, ainsi que les taux d'effort pour chacune.

Tableau 14 : Livrable 5.1 – Valeurs de référence par variable et par classe et taux d'efforts associés.

Classes	Variable 1				Variable 2				...	Variable n			
	Min	Max	Médiane	Moyenne	Min	Max	Médiane	Moyenne	...	Min	Max	Médiane	Moyenne
Classe 1	X	X	X	X	X	X	X	X	...	X	X	X	X
Classe 2	X	X	X	X	X	X	X	X	...	X	X	X	X
Classe 3	X	X	X	X	X	X	X	X	...	X	X	X	X
...	...	...	...	...	...	...	...	...	...	...	...	...	...
Classe m	X	X	X	X	X	X	X	X	...	X	X	X	X

Classes	Taux d'effort Variable 1				Taux d'effort Variable 2				...	Taux d'effort Variable n			
	Min	Max	Médiane	Moyenne	Min	Max	Médiane	Moyenne	...	Min	Max	Médiane	Moyenne
Classe 1	X	X	X	X	X	X	X	X	...	X	X	X	X
Classe 2	X	X	X	X	X	X	X	X	...	X	X	X	X
Classe 3	X	X	X	X	X	X	X	X	...	X	X	X	X
...	...	...	...	...	...	...	...	...	...	...	...	...	...
Classe m	X	X	X	X	X	X	X	X	...	X	X	X	X

5.2. Aide à l'identification de techniques candidates au statut de MTD

Cette étape correspond au travail effectué par l'EIPPCB dans le cadre du Processus de Séville en utilisant l'ensemble des installations collectées. Une partie des techniques pouvant être considérées comme MTD devrait, par définition, se trouver dans les installations dont les performances sont au-moins égales ou inférieures au niveau des valeurs de référence définies par des installations de référence. Par conséquent, il s'agit d'établir le lien entre les installations dont les performances sont au-moins équivalentes aux niveaux de référence établis précédemment et les techniques qu'elles mettent en œuvre.

L'identification des techniques influençant les variables d'analyse est effectuée par jugements d'experts qui peuvent être complétés par les tests statistiques décrits dans l'étape 1. Pour chaque variable d'analyse, les techniques l'influençant sont positionnées sur le niveau d'émissions ou de consommations des installations pour être comparées aux valeurs de référence. Pour chaque site, les techniques mises en œuvre sont indiquées afin d'identifier les techniques présentes dans les installations dont les émissions et consommations sont inférieures aux valeurs de référence (Figure 16). Dans l'exemple donné dans la Figure 16, deux techniques liées aux émissions de DCO sont représentées ; la taille des installations est exprimée au travers de la normalisation de la variable. Des points de formes différentes représentent des sites ayant des techniques ou combinaisons de techniques différentes. Cette représentation graphique permet d'identifier rapidement si des sites emploient des techniques particulières permettant l'obtention d'une valeur d'émission ou de consommation inférieure à la valeur de référence. Compte tenu de la définition de MTD, il est

attendu que les techniques se trouvant sous la valeur de référence soient celles qui utilisent des MTD. Ainsi, dans la Figure 16, les sites utilisant la technique 1 présentent des émissions de DCO inférieures à la valeur de référence quelle que soit la taille des installations. Il est donc vraisemblable que cette technique soit une candidate MTD. Dans le cas de la technique 2, les installations de petite taille dépassent la valeur de référence, mais il semble qu'à partir d'une certaine quantité de produits fabriqués, cette même technique permette d'atteindre un niveau d'émission plus faible. Par conséquent, il apparaît que, lorsque les installations produisent plus d'une certaine quantité de produits, la technique 2 puisse être considérée comme une MTD.

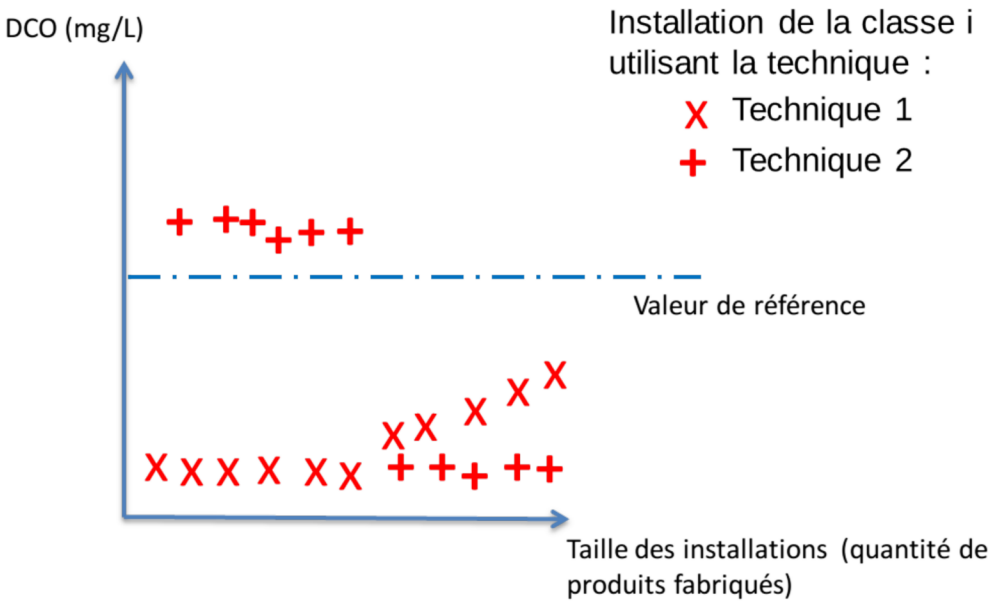


Figure 16 : Livrable 5.2 – Aide au choix des techniques candidates MTD.

Finalement, à l'issue de cette étape, les experts disposent de valeurs de référence pour l'établissement de potentielles valeurs réglementaires dont l'impact sur le parc existant aura été estimé au travers du taux d'effort ainsi que d'une liste de MTD candidates qui serviront à alimenter les discussions d'experts pour la décision finale.

5.3. Implication des acteurs

L'utilisateur peut ajuster les sélections itérativement en fonction des résultats s'il s'avère que le taux d'effort soit trop élevé pour être représentatif ou selon les discussions entre experts. Il choisit également la valeur de référence pour chaque classe dans le cas de l'approche représentative (qui correspondrait à une BATAEL différenciée dans le Processus de Séville) qui servira ensuite à

proposer des MTD candidates. En revanche, avec l'approche performante, en l'absence de classification représentative complémentaire, cette différenciation est impossible et l'ensemble de la population est jugée au regard des valeurs de référence issues de la sélection issue des optima de Pareto et éventuellement de l'INPE.

5.4. Livrables de l'étape 5

Les livrables de cette étape d'interprétation des données concernent deux éléments principaux que sont :

- livrable 5.1 : un tableau de valeurs de référence par variable et par classe associant les taux d'effort (exemple du Tableau 14) ;
- livrable 5.2 : un graphe présentant le positionnement des techniques utilisées par les installations afin d'identifier les techniques candidates MTD (exemple de la Figure 16).

Conclusions de la partie 2

Cette partie a montré l'intégration de la méthodologie indépendamment des processus de décision existants. Celle-ci se compose de 5 étapes visant à servir d'aide à la prise de décision des experts dans le cadre de processus de décision établis et ce à chacune de ces étapes. Ainsi, nous proposons d'utiliser des résultats issus de l'analyse de données provenant d'installations industrielles à l'aide d'outils mathématiques comme supports à la détermination des MTD (Figure 17).

Dans la première étape, des tests d'hypothèse sont employés comme une aide au choix des variables d'analyse pour aider à la réduction de leur nombre. Ensuite, afin de réduire l'impact des données manquantes, la seconde étape permet de les reconstruire grâce à MissForest afin de limiter l'impact de celles-ci sur l'analyse. De plus, les erreurs de l'imputation sont estimées afin de chercher le meilleur compromis entre nombre de variables d'analyse et nombre d'installations étudiées au regard de l'exploitation qui sera faite et des données disponibles. À partir, de ce jeu de données complet, la troisième étape permet de classifier les installations en fonction de deux approches pouvant être complémentaires (représentative via la classification ascendante hiérarchique ou performante via le front de Pareto) et de l'ensemble des variables d'analyse. Ainsi, les groupes d'installations identifiés représentent les différentes plages de consommations et d'émissions qu'il est possible de rencontrer dans le cas de la classification représentative, et les installations offrant les plus basses consommations et émissions sur l'ensemble des variables d'analyse pour l'approche performante. Ces groupes servent ensuite à la sélection des installations de référence dans l'étape 4 dans laquelle une aide à la hiérarchisation des sites de chaque groupe est proposée (distances intra – et inter-classes pour les installations représentatives, et Indicateur de Niveau de Performance Environnementale pour les installations performantes). Enfin, les installations de référence choisies par l'utilisateur permettent, dans la cinquième étape, de calculer des valeurs de référence utilisées pour estimer l'importance des adaptations à accomplir par les installations existantes pour une « mise en conformité » (taux d'effort) et, d'autre part, pour fixer un seuil d'émission ou de consommation qui permette de trouver les MTD dans les installations existantes.

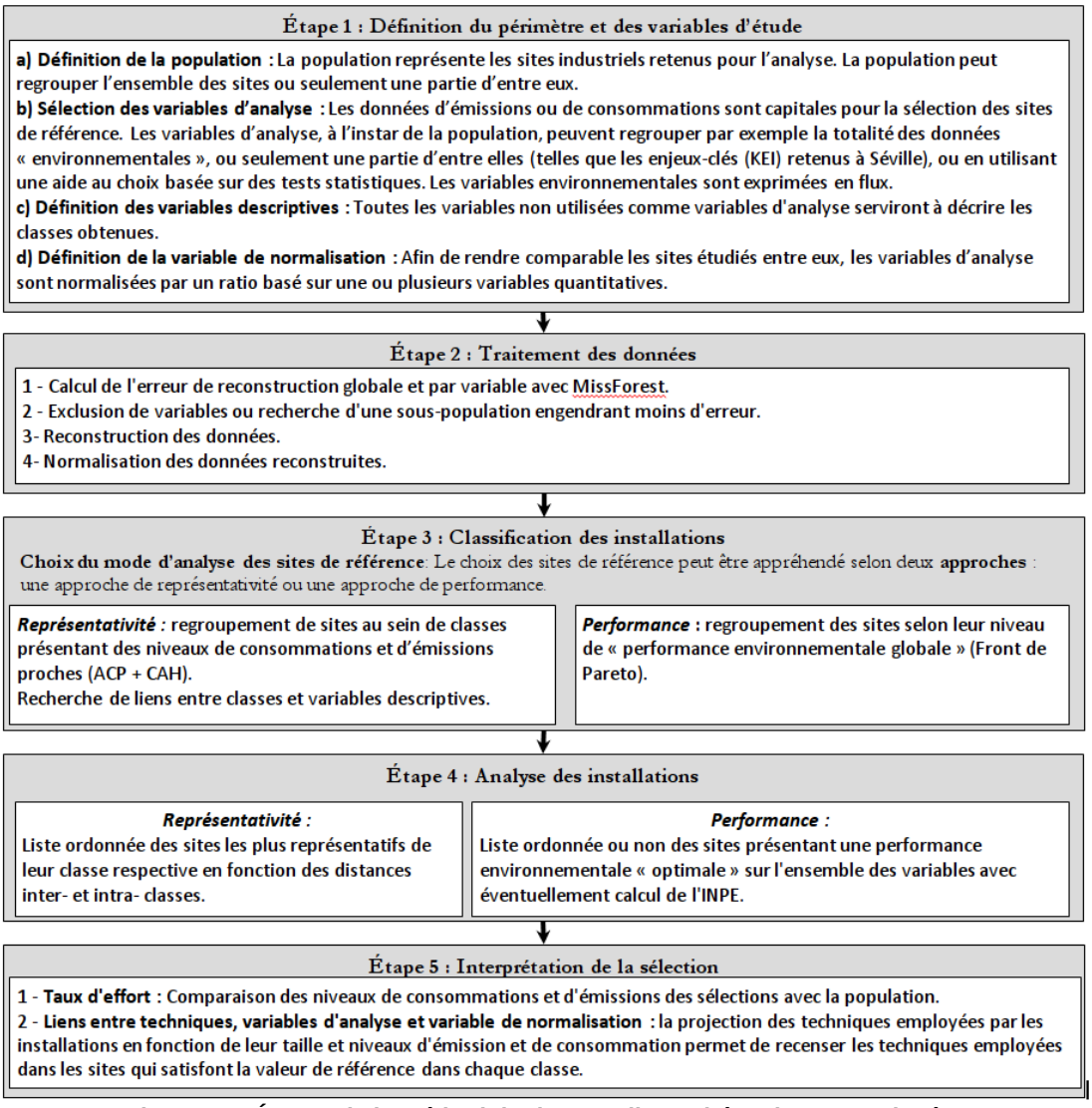
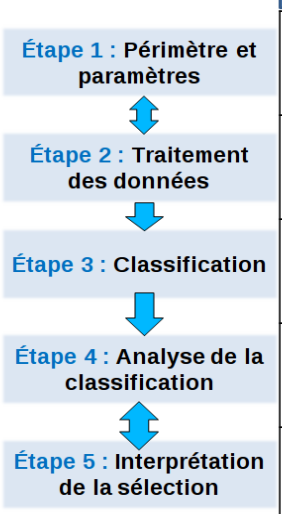


Figure 17 : Étapes de la méthodologie et outils mathématiques employés.

De plus, cette approche méthodologique est suffisamment générique pour permettre son intégration dans différents contextes et objectifs. Ainsi, comme le montre la Figure 18, les étapes peuvent servir indifféremment que l'on soit en début de projet sans avoir d'enjeux-clés (KEI) déjà établis ou si les variables d'analyse sont déjà fixées, et selon le souhait de caractériser un secteur industriel ou d'aller jusqu'à rechercher de possibles MTD. Ces différents niveaux d'intégration et d'usage de la méthodologie, sont illustrés dans plusieurs études de cas présentées dans la troisième partie de ce mémoire.



Aide à la décision apportée		
En début de projet	Avec des variables d'analyse fixées	Lors de la recherche des MTD
Éclairage sur les variables d'analyse les plus importantes pour caractériser un secteur industriel en soulignant des liens entre variables.	Vérification de la pertinence des variables d'analyse choisies par rapport à la quantité d'information unique qu'elles apportent.	Réduction du nombre de variables d'analyse.
Éclairage sur la quantité de données nécessaires à la collecte et les variables d'analyse les plus problématiques en cas de manque.	Vérification que les données sur les variables d'analyse soient suffisantes pour la précision d'analyse voulue.	Validation que la collecte de données sur les installations de référence soit suffisante pour en déduire des MTD et valeurs de référence.
Typologie des installations par rapport aux données de consommation et d'émission disponibles.		Classification des installations en vue de définir des valeurs de référence différenciées ou non.
Proposition de sites de référence ou nécessitant d'être étudiés en détail.		Proposition de valeurs de référence pour trouver des MTD.
Éclairage sur les disparités de performance entre installations d'une population donnée.	Vérification de la pertinence des installations de référence par rapport au taux d'effort, valeurs de référence et MTD candidates nationales.	Proposition de valeurs de référence et MTD candidates.

Figure 18 : Aperçu des différents apports de la méthodologie en fonction du contexte d'application.

Partie 3 :
Cas d'application

Chapitre 1 : Application à l'industrie laitière française

L'application de la méthodologie aux installations laitières françaises vise à démontrer l'applicabilité de la méthodologie proposée dans un contexte IED pour la définition d'installations de référence à l'échelle nationale. Elle s'appuie sur la démarche mise en œuvre dans le cadre du processus de Séville notamment en termes d'enjeux environnementaux-clés (KEI) et de variable de normalisation. Ainsi, suite à la réunion de lancement de la révision du BREF FDM (agroalimentaire) à Séville, des données ont été collectées par le CNIEL (Centre National Interprofessionnel de l'Économie Laitière). Ces dernières ont été utilisées en partie par le groupe miroir français pour identifier les installations de référence à transmettre à Séville. Afin de démontrer l'applicabilité de la méthodologie, ces mêmes données ont été utilisées comme base de travail.

L'étude du secteur de la transformation laitière française est d'autant plus intéressante que parmi le 762 installations du secteur, environ 150 entre dans le champ de l'IED. De plus, les données de 115 d'entre elles ont été collectées par le CNIEL. En outre, il apparaît que le secteur se démarque par une diversité d'activités avec la production de nombreux produits différents : fromages, poudres, lactosérum, etc. Il semble également que la France se démarque d'autres pays européens par la multiplicité des produits fabriqués en un même site. Par rapport au contexte européen du Processus de Séville, l'étude nationale se positionne dans le travail du groupe miroir en amont des réunions européennes.

1. Application de la méthodologie

Pour le premier objectif relatif à la détermination d'une liste d'installations de référence à l'échelle nationale, les cinq étapes de la méthodologie ont été déroulées. Les résultats de cette application sont présentés dans cette section.

1.1. Étape 1 : Définition du périmètre et des variables d'étude

Les données disponibles se composent de données techniques, d'émission, de consommations et de production. Les informations descriptives du contexte national sont les suivantes :

- **population :** 115 installations laitières françaises ayant participé à la collecte de données (sur un total estimé par la profession de 150) ;

- **variables d'analyse :** 9 variables de consommations et d'émissions dans les eaux correspondant aux KEI définis lors de la réunion de lancement de la révision du BREF (Tableau 15). Toutes les variables ont été exprimées en flux afin d'assurer une homogénéité ;

Tableau 15: Liste des variables d'analyse de l'industrie laitière française.

Variable d'analyse	Unité
Azote Ammoniacal	kg/an
Azote Total	kg/an
DBO_5	kg/an
DCO	kg/an
MES	kg/an
Phosphore Total	kg/an
Volume d'effluents	kg/an
Consommation d'énergie	MWh/an
Consommation d'eau	m^3/an

- **variables de normalisation :** 3 variables de normalisation sont possibles : tonnage de matière laitière entrante, tonnage de produits sortants et volume d'équivalents litres de lait entrant calculé à partir des produits finis (AIDA, 2016). Cette dernière variable correspond aux tonnages de produits fabriqués, pondérés par un coefficient permettant de revenir à la quantité de lait qui a été nécessaire à leur fabrication. Ici, la quantité de matière laitière a été choisie afin d'être en cohérence avec le Processus de Séville ;
- **variables descriptives :** 26 variables quantitatives et qualitatives.

Le Tableau 16 présente le pourcentage d'installations pour lesquelles certaines variables d'analyse sont manquantes. Ainsi, 28 installations sur les 115 sont complètes et autant d'installations ont une seule variable incomplète. Le reste des installations, soit plus de la moitié de la population ont entre 2 et 8 variables d'analyse manquantes.

Tableau 16 : Pourcentage d'installations laitières françaises pour chaque nombre de variables d'analyse manquante sur les 9 variables d'analyse choisies.

Nombre de variables d'analyse manquantes (sur 9)	Nombre d'installations	Pourcentage
8	1/115	0,9 %
7	18/115	16 %
6	2/115	1,7 %
5	19/115	17 %
4	1/115	0,9 %
3	9/115	8 %
2	9/115	8 %
1	28/115	24 %
0	28/115	24 %

Étant donné qu'une majorité d'installations est incomplète, il est nécessaire de reconstruire les données manquantes sur les variables d'analyse et de normalisation.

Le livrable de cette étape (Tableau 17) est composé de la base de données initiale regroupant les variables dans les catégories « variables d'analyse », « variables descriptives » et la « variable de normalisation » choisie.

Tableau 17 : Livrable 1 – Base de données initiale pour les installations laitières françaises.

	Variable de normalisation	Variables d'analyse			Variables descriptives		
Site	**Tonnage de matière laitière entrante (T/an)**	**Azote ammoniacal (kg/an)**	**…**	**Énergie (kWh/an)**	**Famille**	**…**	**Nombre de produits**
838	460		…	21 915 454	Autres		1
405	200	18 000	…		Poudres	…	1
60	237	1 839	…	28 491 589	Fromages	…	2
…	…	…	…	…	…	…	…
Nombre de données disponibles	91	41	…	107	115	…	115
Nombre de données manquantes	24	74	…	8	0	…	0
% de données manquantes	21 %	64 %	…	7 %	0 %	…	0 %

1.2. Étape 2 : Traitement et consolidation des données

Le traitement des données consiste en la reconstruction des données manquantes avec MissForest. Le taux d'erreur de la reconstruction est calculé à partir des 115 sites et 9 variables d'analyse retenues. Les histogrammes de la Figure 19 représentent le pourcentage de données manquantes pour les variables d'analyse proposées à l'étape précédente (axe de gauche). Les croix illustrent les taux d'erreur de la reconstruction de chacune des variables (axe de droite).

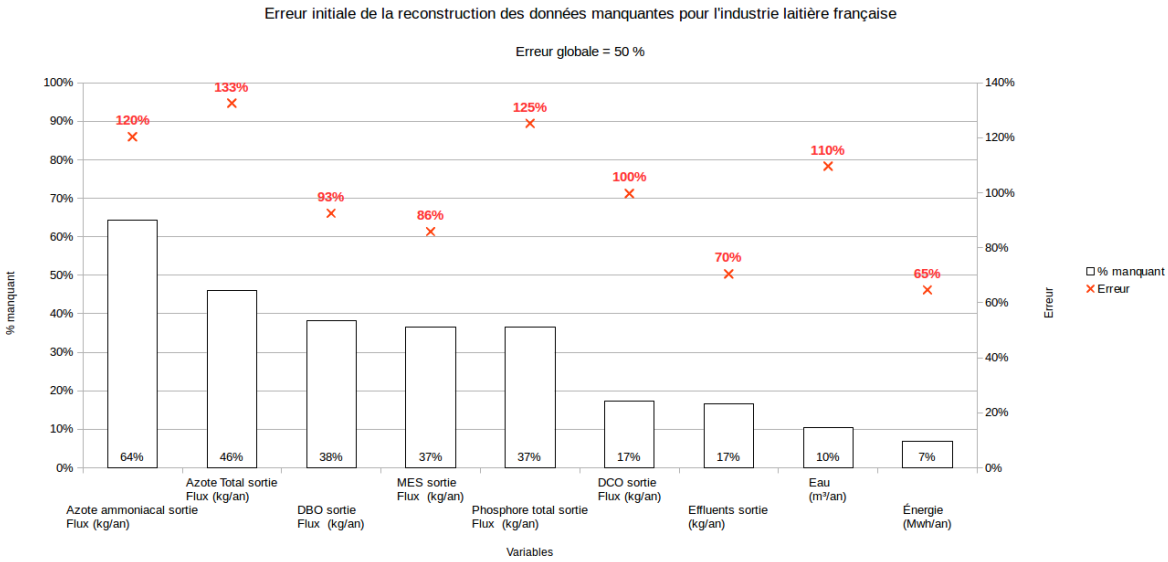


Figure 19 : Pourcentage de données manquantes dans les installations laitières françaises et taux d'erreur global de la reconstruction avec MissForest pour les 3 variables d'analyse incomplètes.

L'erreur de reconstruction observée est systématiquement supérieure à 65 % pour ces neuf variables. L'erreur quadratique moyenne normalisée de la reconstruction, au global, est évaluée à 50 % malgré un pourcentage de données manquantes inférieur à 50 % (sauf pour l'azote ammoniacal qui est de 64 %).

Dans cette application, le choix des variables étant fixé, les installations qui causent le plus d'erreur sont recherchées. Une règle générale fixée pour cette application et les suivantes est que l'erreur globale doit être inférieure à 50 % et les erreurs par variable inférieures à 100 %. L'exclusion de huit installations pour lesquelles manque la consommation d'énergie permet de passer d'une erreur globale de 50 % à 7 % (Figure 20). De plus, l'erreur par variable (Figure 21) diminue fortement pour toutes les variables d'analyse sauf l'azote ammoniacal qui apparaît difficile à reconstruire avec précision.

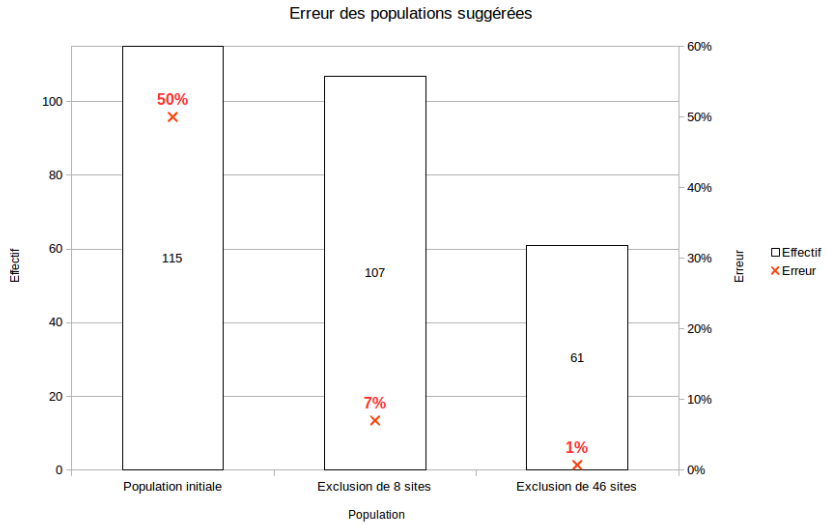


Figure 20 : Taux d'erreur et données manquantes pour la population d'installations laitières françaises proposée par rapport à la population initiale et par variable pour la population suggérée.

En continuant à chercher quels sites exclure pour réduire l'erreur globale de la reconstruction, un second palier permet d'atteindre 1 % d'erreur avec l'exclusion de 46 installations de la population initiale. Néanmoins, cette nouvelle population réduit la taille de la population à 60 % de la population initiale. Par conséquent, le choix d'experts porte ici sur la perte d'information acceptable par rapport à la qualité de l'analyse. Dans ce cas, il a été convenu qu'une erreur globale inférieure à 10 % pour moins de 10 % de réduction de la taille de l'échantillon était satisfaisant.

Ainsi, vu la réduction importante de l'erreur, la sous-population de 107 installations (115-8) a été conservée. Par jugement d'experts, bien que l'azote ammoniacal conserve une erreur élevée, il est gardé. L'ensemble des variables d'analyse est inchangé. L'approche multicritère utilisée dans l'étape suivante de classification permet d'atténuer l'impact d'une erreur élevée d'une des variables d'analyse. Après normalisation de ces variables, cette nouvelle base de données est utilisée pour la classification des installations.

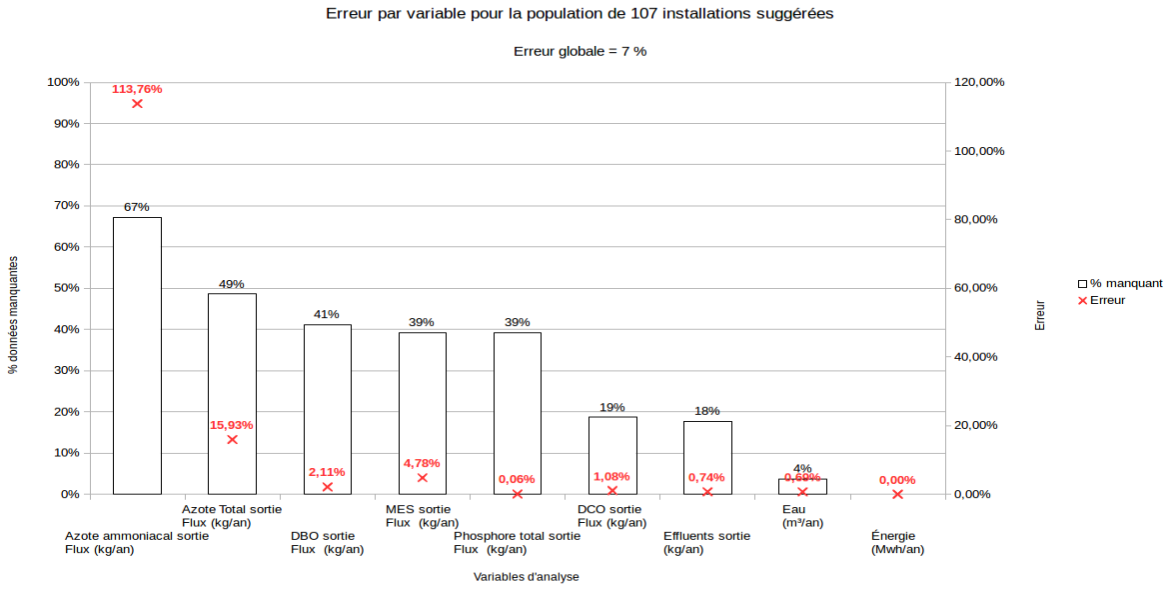


Figure 21 : Taux d'erreur et de données manquantes par variable pour la population d'installations laitières françaises suggérée.

L'objectif de cette étape est d'arbitrer entre erreur de la reconstruction et nombre de sites. Or, pour cette application, il a été décidé de conserver un maximum d'installations donc le passage à 61 installations, soit moins de la moitié de la population d'installations laitières IED, a été jugé inacceptable en comparaison d'une population de 107 sites ayant une erreur plus élevée.

Le livrable de cette étape (Tableau 18) est la matrice des données reconstruites qui serait utilisée, après normalisation par le tonnage de matière laitière entrante, à l'étape suivante.

Tableau 18 : Livrable 2 – Matrice de données reconstruites d'installations laitières françaises.

	Variable de normalisation	Variables d'analyse normalisées			Variables descriptives		
Site	**Tonnage de matière laitière entrante (T/an)**	**Azote ammoniacal (kg/T entrante)**	**…**	**Énergie (kWh/T entrante)**	**Famille**	**…**	**Nombre de produits**
838	460	18,88	…	47 642	Autres		1
60	237	7,76	…	120 218	Fromages	…	2
…	…	…	…	…	…	…	…
Nombre de données disponibles	85	72	…	107	107	…	107
Nombre de données manquantes	22	35	…	8	0	…	0
% de données manquantes	21 %	67 %	…	0 %	0 %	…	0 %
Erreur de reconstruction	75 %	113,76 %		0 %	0 %	…	0 %

1.3. Étape 3 : Classification des installations

Cette étape consiste à classifier les installations au regard de deux objectifs distincts, un objectif de représentativité et un objectif de performance.

1.3.1. Approche représentative et caractérisation des classes

Une première classification a été effectuée et a abouti à deux classes : la première de 103 installations et la seconde de 4 installations. Dans le cas où une classe possède moins de 10 % du nombre de sites total, conformément à la règle définie dans la partie méthodologie, cette classe de sites « atypiques » n'est pas retenue. Par conséquent, une seconde classification a été effectuée sur les 103 installations. Deux classes ressortent de cette nouvelle classification : la première de 90 installations et la seconde de 13 installations. Le dendrogramme issu de cette classification est visualisable sur la Figure 22.

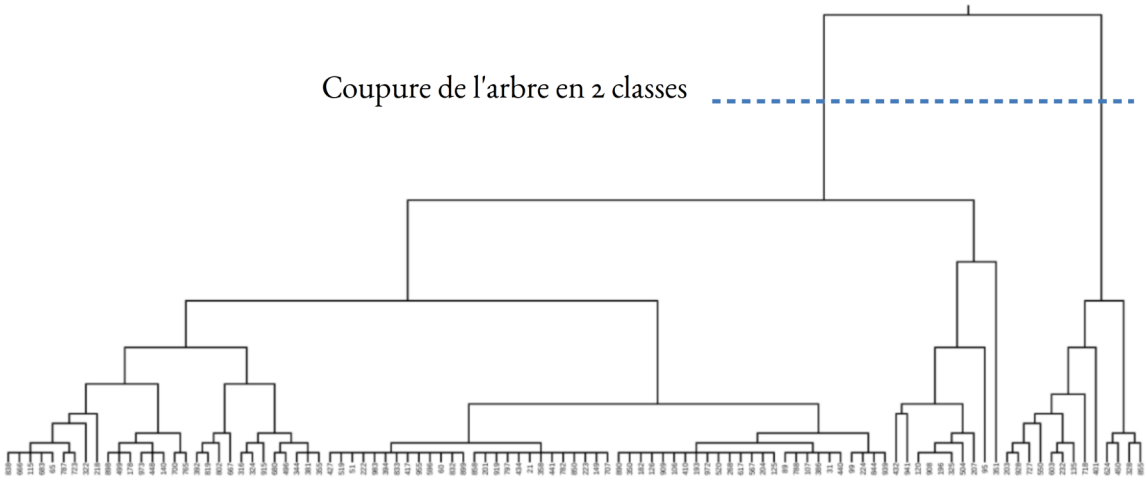


Figure 22 : Dendrogramme de la classification représentative des installations laitières françaises.

1.3.1.1. Caractérisation des classes

La comparaison des classes en fonction de leurs variables d'analyse et descriptives quantitatives est présentée sous forme de boîtes à moustaches (Figure 23) permettant de comparer les classes en fonction de leurs variables d'analyse et de leurs variables descriptives quantitatives. De manière générale, la classe 1 contient des sites moins consommateurs et émetteurs que la classe 2 par tonne de matière laitière entrante et une dispersion moindre des valeurs. La comparaison des classes avec les variables descriptives quantitatives présente des résultats similaires à ceux obtenus pour les variables d'analyse (Annexe 5).

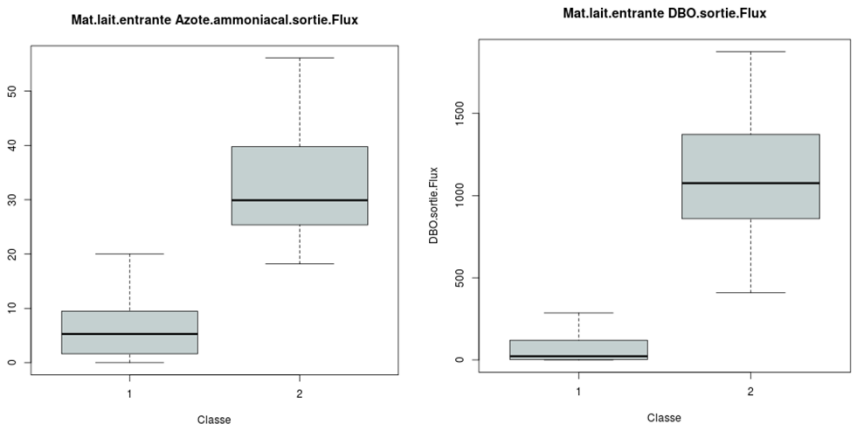


Figure 23 : Boîtes à moustaches représentatives de variables d'analyse pour les deux classes d'installations laitières françaises dans l'approche représentative.

Les tests statistiques appliqués entre le numéro de classe et les variables descriptives qualitatives ne permettent pas d'identifier des relations entre les classes et ces variables.

1.3.1.2. Livrable de l'étape

Le livrable 3.1 de l'étape, synthétisant les dépendances entre la classification et les variables d'analyse, n'est pas reproduit dans la mesure où aucune dépendance avec une variable descriptive n'a été trouvée. Seul un extrait de la table contenant les installations dans les différentes classes est présenté ici au titre du second livrable (Tableau 19).

Tableau 19 : Livrable 3.2 – Code des installations laitières françaises classées par groupe représentatif.

Installations représentatives		
Classe 1 (90 installations)	**Classe 2 (13 installations)**	**Sites « atypiques » (4 installations)**
838	855	64
149	232	629
178	328	532
860	928	725
973	203	
…	…	

1.3.2. Classification performante et caractérisation des classes

La classification performante s'opère via le front de Pareto. Sept sites optimaux pour le jeu de variables utilisé se situent sur le Front. Le Tableau 20 présente les installations qui s'y trouvent.

Tableau 20 : Livrable 3.2 – Liste des installations laitières françaises performantes.

Installations performantes	
Sites performants	**Sites non optimaux**
31	Tous les autres sites en rejets directs
65	
99	
193	
224	
440	
707	

1.4. Étape 4 : Analyse de la classification

Les installations retenues sont ensuite positionnées les unes par rapport aux autres.

1.4.1. Liste de sites représentatifs et niveaux de référence

Pour l'approche représentative, le nuage de points pour chacune des classes a été tracé en fonction des distances inter – et intra – classes. Pour la classe 1 (Figure 24), nous choisissons par défaut 10 % des sites doivent être sélectionnés comme sites de référence. La sélection s'effectue en suivant les numéros de zone établis dans la partie méthodologique. Ainsi, 8 sites de la zone 2 et 9 sites de la zone 3 sont retenus.

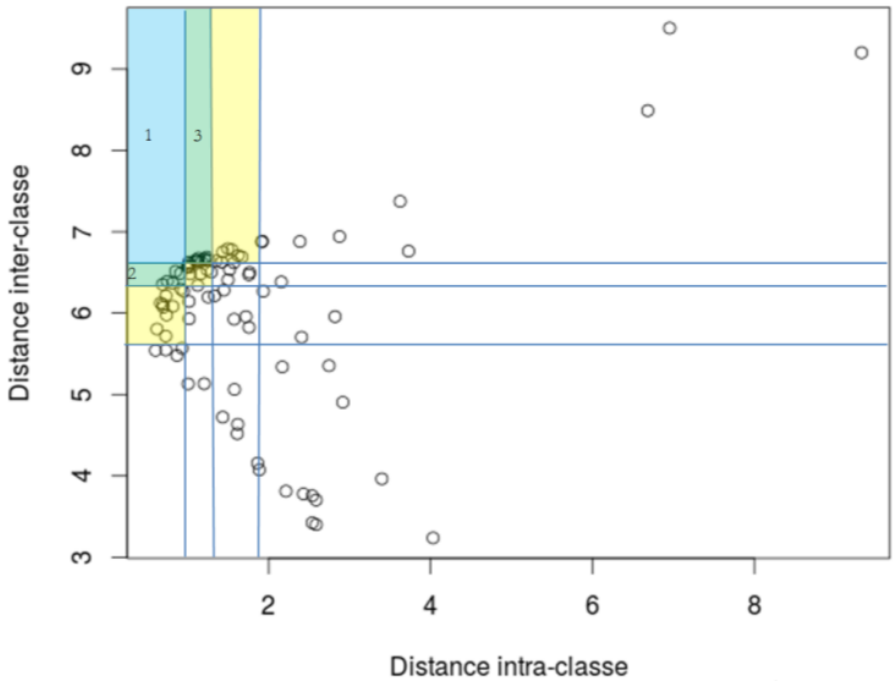


Figure 24 : Outils d'aide à la sélection des sites de référence pour les installations laitières françaises.

Une analyse similaire a été menée sur la classe 2 et deux sites de référence ont été retenus dans les zones 2 et 4.

Le Tableau 20 présente les installations retenues comme sites de référence pour l'approche représentative pour les deux classes identifiées ainsi que les 4 sites considérés comme « atypiques » lors de la première classification. Ces derniers devant être étudiés plus précisément afin d'en comprendre les particularités.

Tableau 21 : Livrable 3.2a – Liste des installations représentatives de référence hiérarchisées par classe.

Sélection représentative								
Classe 1				Classe 2				Sites « atypiques »
Sites	Distance intra-classe	Distance inter-classe	Zone	Sites	Distance intra-classe	Distance inter-classe	Zone	
417	1,10	6,66	2	450	1,85	6,68	2	64
204	1,12	6,64	2	550	1,71	6,23	2	532
933	1,13	6,75	2					629
107	1,17	6,33	2					725
125	1,20	6,66	2					
386	1,23	6,69	2					
106	1,24	6,65	2					
410	1,26	6,65	2					
880	1,35	6,64	2					
51	0,68	6,35	3					
222	0,75	6,39	3					
427	0,81	6,37	3					
983	0,85	6,52	3					
519	0,91	6,49	3					
596	0,99	6,57	3					
617	1,00	6,58	3					
394	1,01	6,63	3					

1.4.2. Liste de sites performants et niveaux de référence

Étant donné que moins de 10 % de la population (7 installations sur 107) sont sur le front de Pareto (Tableau 20), ces dernières sont toutes conservées dans la suite de l'analyse.

1.5. Étape 5 : Interprétation des sélections

Cette étape est appliquée uniquement à la classe 1 pour les installations représentatives et aux installations performantes dans le but de comparer les deux approches.

1.5.1. Taux d'effort dans les sites représentatifs

Pour rappel, le taux d'effort permet de définir les effets que peuvent avoir le choix des seuils de référence sur les entreprises du secteur d'activité concerné. Ce taux d'effort est calculé pour l'ensemble de la population d'une classe au regard des valeurs de référence (moyenne, médiane, minimale et maximale) de la variable d'analyse calculées à partir des sites de référence.

Cet exercice a été effectué pour l'ensemble des variables et les deux classes pour l'approche représentative. Les résultats ne sont pas tous représentés dans cette partie du mémoire afin de ne pas alourdir le texte.

La mise en application est présentée pour la variable DCO (Figure 25). Ainsi, les taux d'effort moyen et médian sont respectivement 77 % et 81 % pour la classe 1. Par ailleurs, les taux d'effort déduits du maximum et du minimum sont respectivement de 63 % et 96 %.

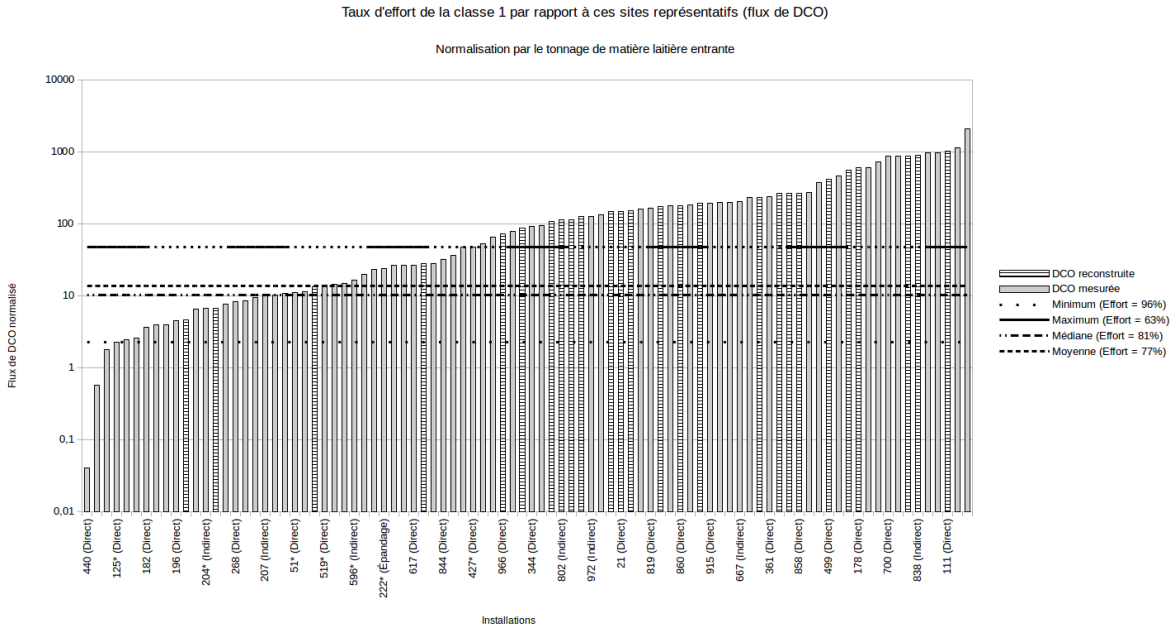


Figure 25 : Taux d'effort pour la classe représentative n°1 des installations laitières françaises (flux de DCO).

1.5.2. Taux d'effort dans les sites performants

De même que pour l'approche représentative, seul un exemple sera présenté ici afin d'illustrer les résultats. Dans l'exemple des émissions de DCO, les valeurs de référence obtenues à partir des 7 installations de référence sont comparées à l'ensemble de la population (Figure 26). Le taux d'effort médian vaut 48 % et le taux d'effort moyen est de 64 %. Dans cet exemple, les taux d'effort, si les valeurs minimale et maximale étaient retenues, seraient respectivement de 100 % et 29 % pour cet indicateur de DCO.

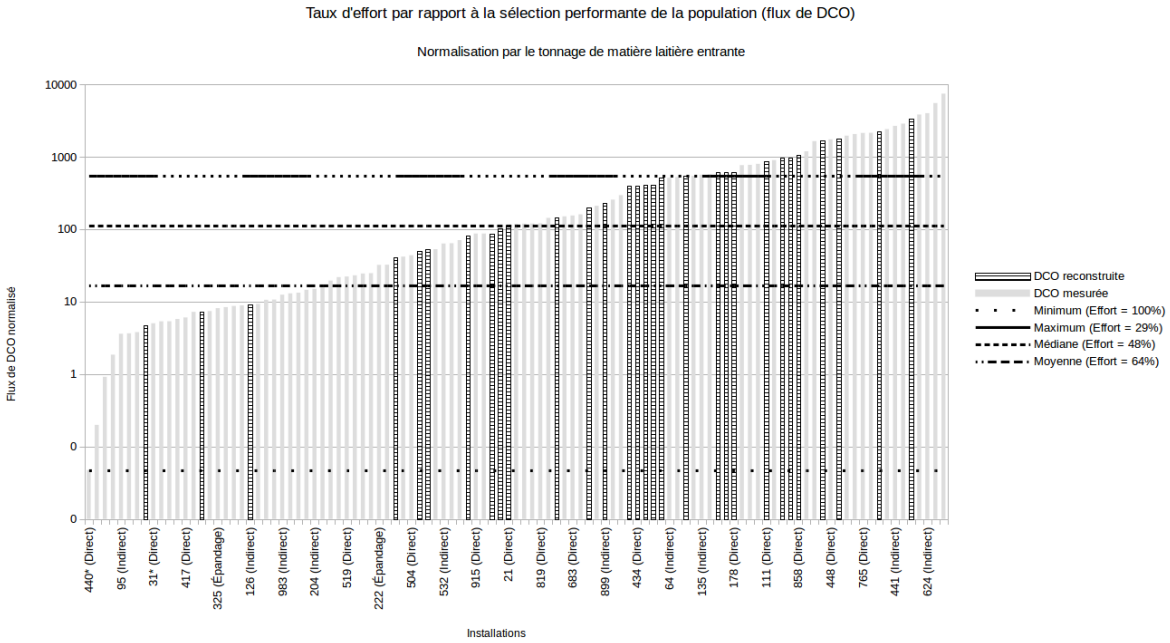


Figure 26 : Taux d'effort par rapport aux installations laitières françaises performantes pour les émissions de DCO.

Le livrable de cette étape correspond à un tableau de synthèse des valeurs de référence possibles pour chaque variable (Tableau 22). Le Tableau 22 présente un extrait des résultats obtenus pour les différentes approches (représentatives et performantes) pour la variable DCO.

Sur l'ensemble des variables d'analyse, des différences d'émission importantes sont observées (Annexe 6) entre les valeurs de référence de la classe 1 et celles de la classe 2. Une différence moindre existe entre les consommations d'eau et d'énergie. Dans le cas d'une application dans un processus de décision réel tel que le Processus de Séville, il serait utile de différencier ces deux cas.

Tableau 22 : Livrable 5 – Valeurs de référence par variable et par classe et taux d'efforts associés dans le cas des installations laitières françaises.

Classes	DCO								…
	Minimum		Maximum		Médiane		Moyenne		…
	Valeur de référence (kg/T matière laitière entrante)	Taux d'effort	Valeur de référence (kg/T matière laitière entrante)	Taux d'effort	Valeur de référence (kg/T matière laitière entrante)	Taux d'effort	Valeur de référence (kg/T matière laitière entrante)	Taux d'effort	…
Classe 1	2,24	96 %	46,89	63 %	10,17	81 %	13,59	77 %	…
Classe 2	2023,95	38 %	2641,27	23 %	2332,61	38 %	2332,61	38 %	…
Performants	0,05	100 %	540	29 %	111	48 %	17	64 %	…

2. Interprétation et conclusions

Les résultats obtenus ont soulevé plusieurs questions suite à leur comparaison avec le jugement d'experts. Tout d'abord, la distinction des rejets directs et indirects n'est pas apparue lors de l'application de la méthodologie. De même, la fabrication de multiples produits sur le même site n'a pas été trouvée comme significative. Par ailleurs, la comparaison avec les installations de référence choisies par les professionnels du lait est réalisée.

2.1. Rejets directs vs rejets indirects

En France, les contraintes réglementaires liées aux rejets directs et indirects dans le milieu récepteur sont différentes (Assemblée Nationale, 1998). Ainsi, comme les variables d'analyse sont essentiellement basées sur les paramètres de rejets dans l'eau, on aurait pu s'attendre à des classifications mettant en évidence ces deux populations dans le cadre de l'approche représentative.

Pour autant, le test de Kruskal-Wallis ne permet pas de mettre en évidence cette distinction entre rejets directs et indirects. Ceci peut être dû à la variable de normalisation qui est utilisée et au fait que les résultats sont traités en valeurs normalisées.

2.2. Impact du multi-produit

Les *a priori* d'experts tendaient à faire ressortir une différenciation des niveaux de performance selon les types de produits en sortie (par exemple : beurre, lactosérum, etc.). Cependant, les tests d'hypothèse ne permettent pas de mettre cette dépendance en exergue. Ainsi, le Tableau 23 montre que seule la production de petites séries est corrélée avec la classification.

Tableau 23 : Résultat des tests d'hypothèse du lien entre la classification représentative et les produits fabriqués de l'industrie laitière française.

Variable descriptive	Produit principal	Autres produits fabriqués	Nombre de produits fabriqués	Type de rejets	Petites séries
Classification représentative					+

2.3. Comparaison de la liste des sites de référence à la sélection des sites retenus à dire d'experts

Toutes les installations qui avaient été retenues par les professionnels de l'industrie laitière (CNIEL) se trouvent dans la classe représentative n°1. Le Tableau 24 indique les distances intra – et inter-classes pour ces 13 installations. Ainsi, seul un site (le n°427) se trouve dans la sélection qui a

été effectuée avec la méthodologie. Un autre site (358) est proche de celui-ci mais n'a pas été retenu, car il se trouve dans la zone n°4. Les autres sites sont trop éloignés du centre de la classe ou trop proches de la classe n°2 pour avoir été retenu. Cela signifie que leurs caractéristiques de consommations ou d'émissions sont éloignées de la classe identifiée. En revanche, ils peuvent être plus ou moins performants que les sites retenus avec la méthodologie.

Tableau 24 : Position des sites laitiers français sélectionnés par le CNIEL dans la classe représentative n°1.

Sites de référence du CNIEL	Distance intra-classe	Distance inter-classe	Zone de l'outil d'aide à la sélection
427	0,81	6,37	3
358	0,73	6,21	4
325	1,62	6,71	6
196	1,67	6,69	6
99	1,91	6,88	6
797	0,73	5,55	7
899	1,33	6,21	8
120	1,52	6,54	9
207	3,62	7,37	9
941	3,73	6,76	9
95	6,95	9,50	9
496	1,57	5,93	12
700	1,62	4,63	14

La comparaison des valeurs de référence issues de la méthodologie et celles calculées à partir de la sélection du CNIEL (Tableau 25) montre que l'application de la méthodologie a abouti à des valeurs plus basses. Cela est illustré pour les émissions de DCO pour lesquelles les quatre valeurs de référence (minimale, maximale, moyenne et médiane) sont présentées.

Tableau 25 : Comparaison des valeurs de référence pour la sélection d'installations laitières françaises de la profession et les sélections issues de la méthodologie, exemple de la DCO (kg/T de matière laitière entrante).

Émissions de DCO (kg / T de matière laitière entrante)			
Sélection	**Profession**	**Classe représentative n°1**	**Classe performante**
Valeur de référence minimale	3,93	2,24	0,05
Valeur de référence maximale	857,74	46,89	540
Valeur de référence moyenne	150,26	10,17	111
Valeur de référence médiane	46,89	13,59	17

2.4. Conclusion

L'application aux installations laitières françaises illustre l'applicabilité de la méthodologie à un secteur IED à partir des données collectées au cours du Processus de Séville.

La sélection des variables d'analyse est ici supposée fixée lors de la réunion de lancement de la révision du BREF et la reconstruction a visé à réduire l'impact des données manquantes sur la classification. Ainsi, l'exclusion d'installations provoquant des erreurs importantes a permis d'identifier, dès l'étape 2, des sites possédant des valeurs de consommations et d'émissions atypiques. Cette étape a également soulevé la question de l'intérêt d'intégrer l'énergie alors qu'elle n'est pas forcément liée aux processus dont les émissions sont en variables d'analyse. Dans un tel cas, la reconstruction de cette variable induit une erreur importante qui ne peut être réduite qu'en excluant les installations dont la donnée manque.

L'analyse des résultats obtenus pour la classe 1 de l'approche représentative ainsi que la sélection d'installations performantes, a montré que les valeurs de référence peuvent être plus ou moins contraignantes en fonction des variables d'analyse considérées. Les valeurs de référence issues de l'approche performante s'appliquent à l'ensemble de la population alors que l'approche représentative s'applique aux classes, donc à des sous-échantillons de la population initiale. En revanche, seule l'approche représentative permet d'identifier des classes d'installations différenciées ou des installations aux consommations et émissions atypiques et donc de pouvoir appliquer des valeurs de référence adaptées.

Chapitre 2 : Application à l'industrie laitière européenne

Cette application vise à illustrer l'utilisation de la méthodologie dans le cadre du Processus de Séville à l'échelle européenne. Dans ce contexte, il existe plusieurs différences par rapport à l'application nationale (Chapitre 1).

Tout d'abord, les installations sont ici considérées comme des sites de référence du secteur laitier européen dans la mesure où elles ont été validées par les représentants de l'Industrie et les États membres et vont servir à la détermination des BATAEPL, ainsi qu'à la définition de MTD. Le Processus de Séville demande donc aux États membres de s'engager sur le choix des sites en tenant compte notamment du fait que les sites retenus mettent normalement déjà en œuvre des MTD. De plus, les différentes données collectées constituent les K*ey Environmental Issues* (KEI) qui avaient été définis lors de la réunion de lancement du processus de révision du BREF agroalimentaire (FDM) ainsi que toute information utile pour déterminer les MTD. Par conséquent, toutes les variables disponibles devraient servir à la détermination des sites de référence et des MTD sauf celles qui seront estimées trop incomplètes pour être exploitables. Enfin, certaines variables concernent uniquement une sous-population clairement identifiée (par exemple, les émissions atmosphériques pour les producteurs de produits pulvérulents) et un traitement différencié est nécessaire dans un tel cas. Pour autant, cette étude de cas se concentre sur l'analyse du panel le plus large possible compte tenu des données disponibles.

Ainsi, les données utilisées sont issues de la collecte effectuée en 2015 dans le cadre de la révision du BREF du secteur agroalimentaire (FDM) et ne concernent que les installations laitières. Ce cas d'application vise à illustrer l'utilisation de la méthodologie lors de cette phase de traitement des données européennes pour :

1. trier les variables d'analyse afin de garder les plus pertinentes pour l'analyse en recherchant d'éventuelles thématiques ou sous-populations permettant de limiter l'impact des données manquantes ;
2. classifier la population d'étude par rapport aux variables d'analyse et proposer des BATAEPL dans le cas de l'approche représentative, et globale pour l'approche performante ;
3. d'identifier les techniques candidates aux MTD.

Les cinq étapes de la méthodologie vont donc être appliquées successivement, ainsi que les deux approches de classification.

1. Application de la méthodologie

Au regard des objectifs énoncés en introduction, les cinq étapes de la méthodologie sont déroulées et les ajustements nécessaires réalisés. Ainsi, la première étape va permettre à l'utilisateur de préciser le choix de variables d'analyse puis, l'étape 2 approfondira ce choix au regard de l'impact de la reconstruction des données manquantes sur l'analyse. Ensuite, l'étape 3 va être employée pour classifier les installations selon les deux approches représentative et performante et des valeurs de référence seront étudiées à partir d'une sélection de sites de référence dans l'étape 4 pour illustrer la recherche de potentielles MTD (étape 5).

1.1. Étape 1 : Définition du périmètre et des variables d'étude

Les données collectées concernent 105 installations laitières qui ont fait l'objet d'une collecte de données dans le cadre de la révision du BREF FDM (agroalimentaire) auprès de 14 États membres (Tableau 26). Le choix du nombre d'installations collectées ne fait l'objet d'aucune règle précise ; le recensement des installations soumises à l'IED dans chaque pays étant très incomplet (EIPPCB, 2015) comme illustré dans la table. Les variables disponibles sont donc celles qui ont été jugées significatives (KEI) pour la recherche des MTD lors de la réunion de lancement, ainsi que les déchets. Ainsi, 26 variables de consommations et d'émissions sont disponibles pour servir de variables d'analyse, et 49 variables sont potentiellement utilisables comme variables descriptives ou de normalisation (pays, proportions de produits entrants et sortants, etc.). De plus, les émissions aqueuses sont exprimées en grande majorité en concentration, ce qui correspond à des valeurs de flux normalisées par le volume d'effluents. Comme les valeurs de ces derniers sont inconnues pour de nombreux sites, la conversion des concentrations en flux induit une erreur de reconstruction. Cet élément est important car, afin de conserver de la cohérence avec les variables de déchets et de consommations qui sont des flux, toutes les émissions aqueuses ont besoin d'être converties en flux.

Tableau 26 : Effectifs des installations laitières européennes par État membre, d'après EIPPCB (2015).

Pays	Abréviation	Nombre d'installations collectées	Nombre d'installations estimé par pays
Espagne	ES	18	indéterminé
France	FR	13	150
Allemagne	DE	11	indéterminé
Danemark	DK	11	indéterminé
Pays-Bas	NL	9	indéterminé
Finlande	FI	8	indéterminé
Belgique	BE	6	indéterminé
Italie	IT	6	indéterminé
Royaume-Uni	UK	6	indéterminé
Irlande	IE	5	indéterminé
Suède	SE	5	indéterminé
Autriche	AT	4	10
Pologne	PL	3	indéterminé
Portugal	PT	1	indéterminé
Total	**14**	**105**	**indéterminé**

Le traitement primaire des données montre l'absence de données pour certaines variables (valeurs manquantes). En particulier, les consommations d'énergie, qui sont des données confidentielles sont absentes de la plupart des installations. De plus, certaines variables ne concernent que des installations spécifiques comme celles relatives aux émissions atmosphériques qui concernent uniquement les installations produisant des poudres. Par conséquent, les installations non concernées auront des valeurs nulles pour ces variables. Le Tableau 27 présente le taux des valeurs manquantes et nulles pour chaque variable.

Tableau 27 : Importance des données manquantes ou de valeur nulle par variable dans les installations laitières européennes (valeurs annuelles).

#	Variables	Valeurs manquantes		Variables	Valeurs nulles	
1	Carbone Organique Total (COT) (mg/L)	100	94 %	SO_2 (mg/Nm3)	86	81 %
2	Conso. Énergie (MWh)	99	93 %	CO (mg/Nm^3)	86	81 %
3	Chlorures (mg/L)	81	76 %	NOx (mg/Nm^3)	85	80 %
4	Azote ammoniacal (NH_4-N) (mg/L)	77	73 %	Verre (T)	65	61 %
5	Matières En Suspension (MES) (mg/L)	57	54 %	Contenu O_2 (vol-%)	62	58 %
6	Demande Biologique en Oxygène (DBO_5) (mg/L)	52	49 %	Poussières (T)	62	58 %
7	Azote total (mg/L)	36	34 %	Quantité de produits secs (T)	62	58 %
8	Phosphore total (mg/L)	31	29 %	Bois (T)	42	40 %
9	Plastiques (T)	31	29 %	Métal (T)	30	28 %
10	Contenu O_2 (vol-%)	29	27 %	Déchets dangereux (T)	30	28 %
11	Métal (T)	29	27 %	Boues (T)	26	25 %
12	Verre (T)	26	25 %	Plastiques (T)	21	20 %
13	Bois (T)	26	25 %	Déchets biologiques (T)	18	17 %
14	Demande Chimique en Oxygène (DCO) (mg/L)	26	25 %	Carton (T)	15	14 %
15	Déchets dangereux (T)	25	24 %	Déchets résiduels (T)	5	5 %
16	Carton (T)	25	24 %	COT (mg/L)	0	0 %
17	Déchets résiduels (T)	24	23 %	Conso. Énergie (MWh)	0	0 %
18	Quantité de produits secs (T)	22	21 %	Chlorures (mg/L)	0	0 %
19	Oxydes d'azote (NOx) (mg/Nm^3)	19	18 %	NH_4-N (mg/L)	0	0 %
20	Dioxyde de soufre (SO_2) (mg/Nm^3)	18	17 %	MES (mg/L)	0	0 %
21	Monoxyde de carbone (CO) (mg/Nm^3)	18	17 %	DBO_5 (mg/L)	0	0 %
22	Boues (T)	16	15 %	Azote total (mg/L)	0	0 %
23	Volume d'effluents (m^3)	15	14 %	Phosphore total (mg/L)	0	0 %
24	Déchets biologiques (T)	11	10 %	DCO (mg/L)	0	0 %
25	Poussières (mg/Nm3)	9	8 %	Volume d'effluents (m^3)	0	0 %
26	Conso. Eau (m^3)	3	3 %	Conso. Eau (m^3)	0	0 %

Cette étape vise à définir la population d'étude, la variable de normalisation, les variables d'analyse et descriptives en classant les données disponibles. Les choix sont effectués de manière à permettre une analyse transparente pour l'utilisateur. La première étape consiste à choisir la variable de normalisation. Dans un second temps la liste des variables d'analyse est précisée. Enfin, les variables qui en sont exclues sont ajoutées aux variables descriptives.

1.1.1. Variable de normalisation

Les données relatives aux quantités de matière entrantes ou sortantes ne sont pas disponibles pour des raisons de confidentialité et n'ont donc pas pu servir de variables de normalisation. Les seules variables de normalisation possibles, c'est-à-dire susceptibles de fournir une indication de la taille des installations, sont la capacité de production quotidienne et le volume d'effluents.

Normaliser par le volume d'effluents revient à exprimer les variables d'émissions dans l'eau en concentration. En revanche, cette approche n'a pas de sens par rapport aux émissions atmosphériques ou aux déchets. Par ailleurs, la capacité de production est une valeur théorique qui ne reflète pas nécessairement l'activité réelle d'une installation.

Par conséquent, le choix de la normalisation va conditionner les variables d'analyse qui peuvent être employées dans ce cas d'application :

- avec la capacité de production, ce qui correspond également au critère de taille choisi à Séville, une évaluation globale est possible mais alors toutes les variables doivent être exprimées comme des flux ;
- avec le volume d'effluents, seule l'analyse des variables relatives aux consommations et émissions d'eau est possible. Étant donné que cette variable est incomplète (Tableau 28), l'étape 2 de la méthode va permettre d'orienter ce choix en fonction des erreurs de reconstruction des données manquantes.

Le choix de la variable de normalisation peut aussi être réalisé en prenant en compte le nombre de données manquantes. Ainsi, dans notre cas, le volume d'effluents est plus complet que la capacité de production (Tableau 28) ; il a donc été retenu. Le choix de la variable de normalisation sera vérifié dans l'étape 2, en fonction des variables qui pourraient être exclues de l'analyse.

Tableau 28 : Importance des données manquantes pour les variables de normalisation possibles de l'industrie laitière européenne.

Variables	Nombre de données manquantes	Pourcentage sur 105 installations
Volume d'effluents (m^3/an)	15	14 %
Capacité production quotidienne (T/jour)	18	17 %

1.1.2. Variables d'analyse

Le Tableau 27 montre que certaines variables ont une valeur nulle pour une grande proportion d'installations. Cela indique que ces variables s'appliquent pour une sous-population particulière. Les trois variables pour lesquelles il y a 80 % ou plus de valeurs nulles sont SO_2, CO et NOx, soit des variables relatives aux émissions atmosphériques qui ne sont mesurées que sur les sites producteurs de poudres. Par conséquent, ces trois variables et également le contenu en O_2 des poussières et la poussière totale devraient être traités à part et faire l'objet d'une thématique spécifique pour une sous-population particulière. De même, l'émission de chlorures est due majoritairement au salage des fromages et est donc spécifique à certaines activités. Par conséquent,

la reconstruction de cette variable pour tous les autres sites où elle est absente ne pourra pas être effectuée car fausse. Elle ne peut donc pas être utilisée dans l'analyse en l'absence d'information sur les installations qu'elle concerne ou non. Finalement, le nombre de variables d'analyse après ce premier tri pour une analyse globale, passe à 18 (Tableau 29).

Tableau 29 : Liste des variables d'analyse de l'industrie laitière européenne (valeurs annuelles).

#	Variables	Données manquantes sur 105	
		Effectif	Pourcentage
1	COT (mg/L)	100	94 %
2	Conso. Énergie (MWh)	99	93 %
3	NH_4-N (mg/L)	77	73 %
4	MES (mg/L)	57	54 %
5	DBO_5 (mg/L)	52	49 %
6	Azote total (mg/L)	36	34 %
7	Phosphore total (mg/L)	31	29 %
8	Plastiques (mg/L)	31	29 %
9	Bois (T)	30	28 %
10	Métal (T)	30	28 %
11	DCO (T)	26	25 %
12	Verre (T)	26	25 %
13	Carton (T)	25	24 %
14	Déchets dangereux (T)	25	24 %
15	Déchets résiduels (T)	24	23 %
16	Boues (T)	16	15 %
17	Volume d'effluents (m^3)	15	14 %
18	Conso. eau (m^3)	3	3 %

Deux sous-populations peuvent être distinguées, les effluents aqueux en rejet direct et ceux en rejet indirect. Dans cette étude, seule l'analyse des installations en rejets directs sera présentée.

1.1.3. Variables descriptives

Les variables qui ont été exclues sont devenues des variables descriptives et serviront à une classification thématique sur les émissions atmosphériques. Elles complètent ainsi les variables quantitatives et qualitatives déjà présentes pour un total de 36 variables descriptives (Tableau 30).

Tableau 30 : Liste des variables descriptives de l'industrie laitière européenne.

Variables descriptives				
Qualitatives	Pays	Rejet direct, indirect ou épandage		
Quantitatives	% matière entrante : autre (%)	% matière entrante : crème (%)	% matière entrante : fromage (%)	% matière entrante : lactose (%)
	% matière entrante : lait (%)	% matière entrante : lait concentré (%)	% matière entrante : poudres (%)	% produits fabriqués : autres (%)
	% produits fabriqués : autres desserts (%)	% produits fabriqués : beurre et graisse à tartiner (%)	% produits fabriqués : crème glacée (%)	% produits fabriqués : fromages (%)
	% produits fabriqués : lactose (%)	% produits fabriqués : lait (%)	% produits fabriqués : lait fermenté (%)	% produits fabriqués : poudres (%)
	% produits fabriqués : produits liquides concentrés (%)	Chlorures (mg/L)	CO (mg/NM^3)	Capacité journalière (T/jour)
	Flux d'air (Nm^3/h)	Poussières totales (mg/NM^3)	SO_2(mg/NM^3)	Contenu en 0_2 des émissions atmosphériques (vol-%)
	NOx (mg/NM^3)	Déchets métalliques (T)	Déchets Boues (T)	Déchets biologiques (T)
	Déchets Carton et Papier (T)	Déchets Verre (T)	Tonnage de matières premières (T)	Tonnage de produits fabriqués (T)
	Tonnage de produits secs fabriqués (T)	Volume d'eau recyclé (m^3)		

1.1.4. Livrable de l'étape

La base de données initiale se compose donc de 18 variables d'analyse, 36 variables descriptives et 2 variables de normalisation. Un extrait est présenté dans le Tableau 31.

Tableau 31 : Livrable 1 − Extrait de la base de données initiale pour les installations laitières européennes.

	Variable de normalisation		Variables d'analyse			Variables descriptives		
Code du site	**Volume d'effluents (m³/an)**	**Capacité de production (T/jour)**	**COT (mg/L)**	**…**	**DCO (mg/L)**	**Pays**	**…**	**Déchets résiduels (T)**
3		500		…	35	Autriche	…	
4		500	6	…	13	Autriche	…	104
17	528 003	200		…	10	Belgique	…	
18	354 665	685		…	37	Belgique	…	125
…	…	…	…	…	…	…	…	…
Nombre de données disponibles	90	87	5	…	79	105	…	24
Nombre de données manquantes	15	18	100	…	26	0	…	81
% de données manquantes	14 %	17 %	92 %	…	25 %	0 %	…	23 %

Dans la mesure où il y a plus d'installations que de variables d'analyse, il n'y a pas besoin *a priori* de chercher à réduire le nombre de ces variables. L'étape suivante permettra de confirmer ou d'infirmer cette décision.

1.2. Étape 2 : Traitement et consolidation des données

À partir de ce premier jeu de variables le taux d'erreur de la reconstruction avec MissForest est calculé (Figure 27). En premier lieu, l'erreur globale de la reconstruction est ici de 42 %, soit une valeur inférieur au seuil de 50 % spécifiée dans l'application nationale du chapitre 1. Elle servira également de référence pour comparer les erreurs de différents jeux de variables et aider les experts dans leur choix. Les histogrammes (Figure 27) représentent le pourcentage de données manquantes pour les variables incomplètes proposées à l'étape précédente (axe de gauche). Les croix illustrent les taux d'erreur de la reconstruction de chacune des variables (axe de droite). Il apparaît que les erreurs de reconstruction de 11 des 18 variables d'analyse sont supérieures à 100 % notamment pour les variables relatives aux déchets. Une seule des deux conditions énoncées au chapitre précédent (erreur globale < 50 % et erreurs par variable < 100 %), est donc remplie. Il est donc recommandé dans un tel cas d'exclure des installations ou des variables afin de parvenir à faire baisser les erreurs globales et propres à chaque variable.

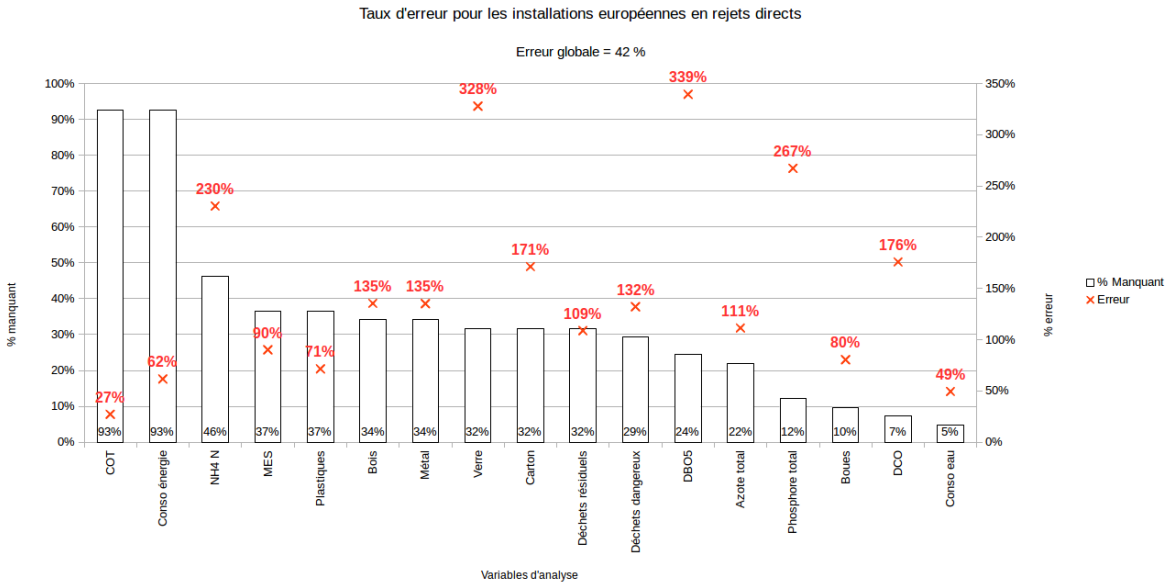


Figure 27 : Taux d'erreur global et par variable de la reconstruction avec MissForest pour les variables d'analyse initiales de l'industrie laitière européenne.

L'avis de l'utilisateur est indispensable pour chercher à faire diminuer l'erreur de la reconstruction. Au regard de la Figure 27, il apparaît que deux variables sont reconstruites alors que moins de 10 % des installations disposent de valeurs. Il est raisonnable de penser que la reconstruction à partir d'une information aussi incomplète peut conduire à des résultats aberrants. De plus, dans la pratique, il semble difficile que le TWG définisse une BATAEL à partir de si peu de données. Par conséquent, les variables COT et consommation d'énergie ont été exclues de l'analyse. Compte tenu de l'erreur élevée de la reconstruction des variables de déchets (plus de 100 %), celles-ci sont exclues également. Ainsi, les variables d'analyse sont désormais au nombre de 7 et toutes relatives à l'eau : NH4-N, MES, DBO5, Azote total, Phosphore total, DCO et consommation d'eau. La variable de normalisation la plus pertinente est par conséquent le volume d'effluents.

Les erreurs de reconstruction sont de nouveau calculées. Il apparaît que ces modifications font augmenter l'erreur globale, car de l'information utile à la reconstruction semble avoir été supprimée et les erreurs pour chacune des variables ont soit légèrement augmentées ou diminuées (Figure 28). Il est possible qu'un petit nombre d'installations soit responsable de la plus grosse partie de l'erreur. Dans un tel cas, il suffirait de les exclure pour faire fortement diminuer les erreurs de reconstruction.

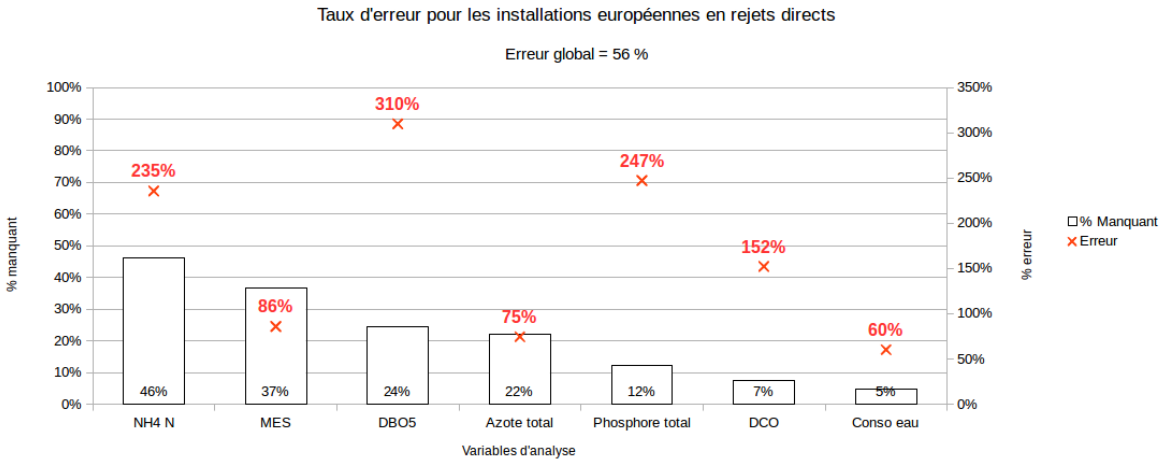


Figure 28 : Taux d'erreur global et par variable de la reconstruction avec MissForest pour les variables d'analyse modifiés et les sites laitiers européens pour lesquels la variable de normalisation est renseignée.

Dans le cas présent, il s'avère que trois installations sont les principales sources d'erreur et qu'il suffit de les exclure pour faire passer l'erreur globale de la reconstruction de 56 % à moins de 1 % (Figure 29). Ceux-ci correspondent aux trois installations pour lesquels la variable de normalisation est inconnue.

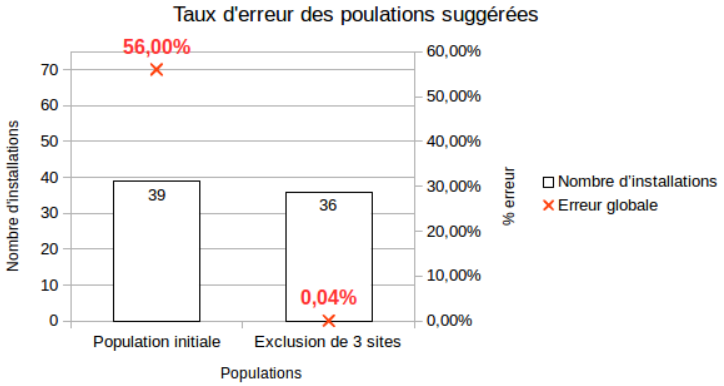


Figure 29 : Erreurs pour la population initiale et une population suggérée de l'industrie laitière européenne.

Les erreurs de reconstruction pour chaque variable ne sont pas améliorées par la nouvelle population (Figure 30), à l'exception des variables MES et DCO et de l'erreur globale. Néanmoins, ces erreurs élevées sont préférables aux données manquantes dans la suite de l'analyse ; il s'agira ensuite de vérifier que les valeurs de référence ne reposent pas sur des valeurs aberrantes. Trois

choix s'offrent à l'utilisateur dans ce cas : compléter les données manquantes, réduire encore le nombre de variables d'analyse ou la population, ou continuer l'analyse avec ces données et juger au regard des résultats de leur pertinence. La première option étant impossible ici et la seconde obligeant à exclure plus de la moitié des installations de référence européenne, l'analyse continue dans cette application avec les 37 sites en rejets directs et 7 variables d'analyse.

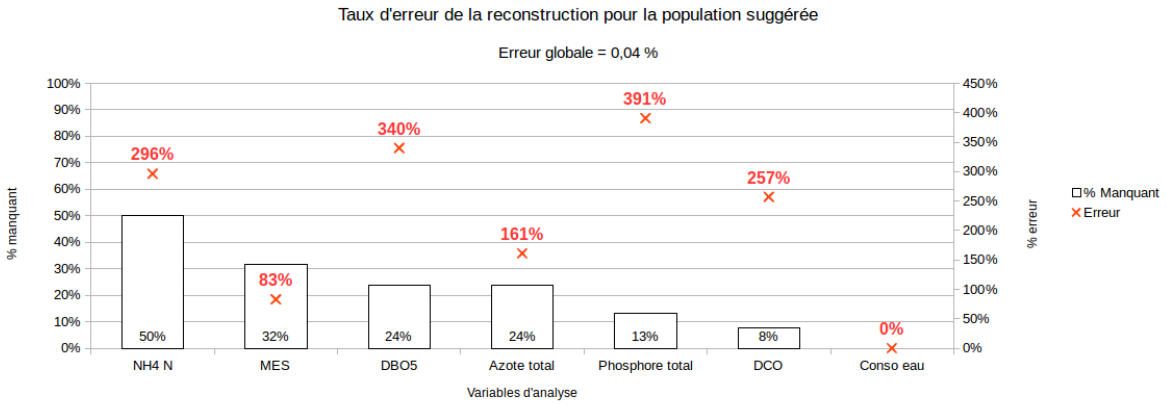


Figure 30 : Erreur par variable de la reconstruction de la population laitière européenne suggérée.

Finalement, toutes les variables précédemment exclues deviennent des variables descriptives, ce qui permettra de juger la classification en fonction de celles-ci également. Enfin, toutes les variables sont normalisées.

1.3. Étape 3 : Classification des installations

1.3.1. Classification représentative et caractérisation des classes

1.3.1.1. Classification des installations

La classification aboutit à la constitution de quatre classes : la première est constituée de 34 installations, les trois autres, d'une seule installation (respectivement le n°126, 255 et 405). Ces trois installations sont donc conservées à part et nécessitent plus d'investigations sur les raisons de leurs différences.

Une seconde classification est effectuée sur les 34 installations de la première classe et aboutit à trois classes (Figure 31) : une première classe de 18 installations, une seconde de 15 installations et une troisième d'une seule installation (Tableau 32). Cette dernière rejoignant les autres installations « atypiques ».

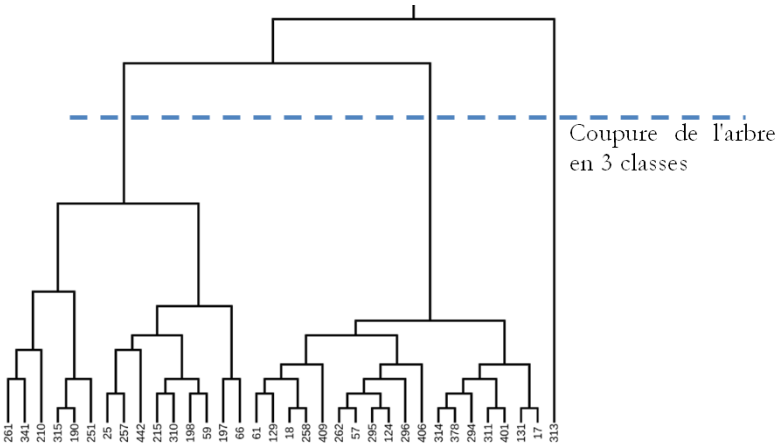


Figure 31 : Dendrogramme de la seconde classification représentative des installations laitières européennes.

1.3.1.2. Caractérisation des classes

La comparaison des classes en fonction de leurs variables d'analyse et descriptives quantitatives est présentée dans l'annexe 7 sous la forme de boîtes à moustaches. La Figure 32 présente deux exemples pour la DBO5 et les MES à titre d'illustration.

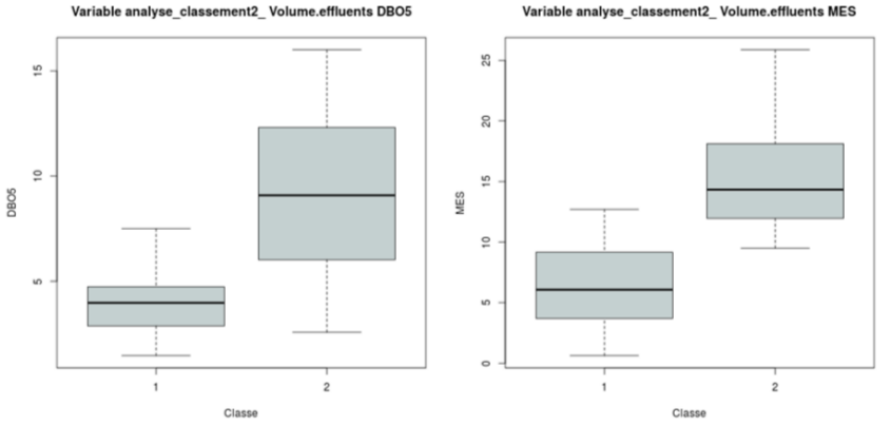


Figure 32 : Exemple de comparaison des classes pour deux variables d'analyse (DBO5 et MES) pour l'industrie laitière européenne.

De manière générale, la classe 1 contient des sites moins consommateurs et émetteurs que la classe 2 par volume d'effluents émis et une dispersion moindre des valeurs. La comparaison des classes avec les variables descriptives quantitatives présente des résultats similaires avec les conclusions établies pour les variables d'analyse (Annexe 7). Aucune corrélation n'a été trouvée entre ces deux classes et les variables descriptives qualitatives. En revanche, toutes les installations

de certains pays se trouvent dans l'une ou l'autre classe. Cela est susceptible de révéler des spécificités nationales qui ne pourront être vérifiées que par l'analyse d'experts.

1.3.1.3. Livrable de l'étape

Le livrable 3.1 de l'étape synthétisant les dépendances entre la classification et les variables d'analyse n'est pas reproduit ici dans la mesure où aucune dépendance avec une variable descriptive n'a été trouvée. Seul le second livrable contenant la répartition des installations dans les différentes classes est présenté dans le Tableau 32.

Tableau 32 : Livrable 3.2 – Liste des installations laitières européennes classées par groupes représentatifs.

Installations représentatives		
Classe 1	**Classe 2**	**Sites atypiques**
17	25	126
18	59	313
57	66	134
61	190	255
124	197	405
129	198	
131	210	
258	215	
262	251	
294	257	
295	261	
296	310	
314	315	
378	341	
401	442	
406		
409		

1.3.2. Classification performante et caractérisation des classes

La classification des installations performantes avec les mêmes variables d'analyse que l'approche représentative se fait via la représentation du front de Pareto. La recherche du front de Pareto fait apparaître 16 sites optimaux pour le jeu de variables utilisé (Tableau 33).

Tableau 33 : Livrable 3.2 – Liste des installations laitières européennes performantes.

Installations performantes	
Sites performants	**Sites non optimaux**
17	Tous les autres sites en rejets directs
57	
124	
126	
129	
131	
294	
295	
296	
314	
341	
378	
401	
406	
409	
442	

1.4. Étape 4 : Analyse de la classification

Les installations contenues dans les listes obtenues précédemment ont été hiérarchisées en fonction de l'approche suivie.

1.4.1. Liste de sites représentatifs et niveaux de référence

Pour l'approche représentative, un nuage de points pour chacune des classes a été tracé en fonction des distances inter – et intra – classes (Figure 33) :

- pour la classe 1, la première zone contenant des installations est la zone n°2. Deux sites s'y trouvent et seront donc les installations représentatives sélectionnées, auxquels s'ajoute un site de la zone 3. Il est décidé ici de s'arrêter à ces trois installations conformément à la règle définie dans la partie méthodologie ;
- pour la classe 2, les trois installations des zones 3 et 4 ont été sélectionnées sur dire d'experts.

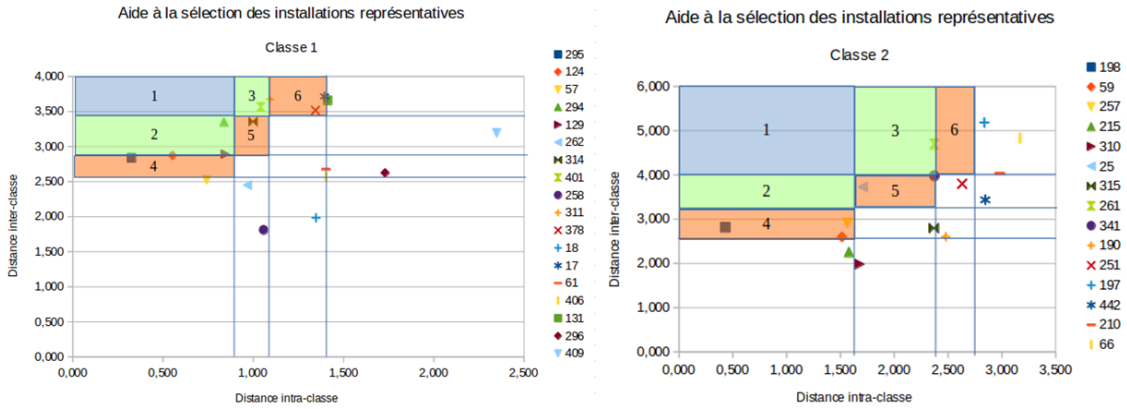


Figure 33 : Outils d'aide à la sélection des sites de référence pour les installations laitières européennes.

Par conséquent, le livrable de cette étape pour les installations représentatives comprend le Tableau 34 reprenant les informations de hiérarchisation des sites par classe ainsi que des sites « atypiques ».

Tableau 34 : Livrable 3.2a – Liste des installations laitières européennes de référence classées par groupe représentatif.

Sélection représentative								
Classe 1				Classe 2				Sites atypiques
Sites	Distance intra-classe	Distance inter-classe	Zone	Sites	Distance intra-classe	Distance inter-classe	Zone	
129	0,85	2,89	2	198	0,43	2,81	3	126
294	0,84	3,35	2	257	1,57	2,91	4	313
401	1,04	3,56	3	261	2,37	4,70	4	255
								405

1.4.2. *Liste de sites performants et niveaux de référence*

Étant donné que les installations étudiées sont déjà considérées comme étant des installations de référence pour la détermination des MTD par le TWG, l'ensemble des 16 installations se trouvant sur le front de Pareto est ici conservé. En l'absence de retour d'experts sur ce nombre, une réduction des sites de référence aurait pu être réalisée via l'utilisation de l'INPE.

1.5. Étape 5 : Interprétation des sélections

Afin de ne pas alourdir le texte, seules les applications aux émissions de DCO pour les installations représentatives et performantes sont présentées dans le but de comparer les deux approches. Les résultats sur les autres variables sont présentés à l'annexe 8.

1.5.1. Taux d'effort dans les sites représentatifs

Afin de calculer le taux d'effort, il est nécessaire de définir des valeurs seuils de référence qui pourront être utilisées comme base pour les BATAEPL. Ces valeurs de référence sont définies sur la base des valeurs des variables des installations de référence issues du livrable 3.2a (Tableau 34). Ainsi, les valeurs de référence sont présentées dans le Tableau 35.

Tableau 35 : Valeurs de référence pour l'exemple de la DCO pour l'approche représentative des sites laitiers européens.

DCO (mg/L)	Concentration min	Concentration moyenne	Concentration médiane	Concentration max
Classe 1	15,32	26,99	27,30	39,60
Classe 2	44,50	47,22	47,94	52,11

Le taux d'effort calculé à partir des valeurs de référence correspond à l'effet que peut avoir l'application des valeurs de référence sur l'ensemble de la population de chaque classe proposée. Le Tableau 36 présente l'exemple des taux d'effort pour la DCO. La Figure 34 présente par des histogrammes le positionnement des installations de chaque classe au regard de ces valeurs de référence.

Tableau 36 : Taux d'effort pour l'exemple de la DCO pour l'approche représentative des sites laitiers européens.

DCO (mg/L)	Taux d'effort minimal	Taux d'effort moyen	Taux d'effort médian	Taux d'effort maximal
Classe 1	6 %	50 %	50 %	44 %
Classe 2	19 %	38 %	38 %	88 %

La Figure 34 montre que les effets sont plus importants pour la classe 1 que pour la classe 2 sauf pour le taux d'effort max. Cela est dû au fait qu'un plus petit nombre d'installations avec une plus grande variabilité est considéré. Comme nous l'avions montré lors de la classification, les valeurs de référence proposées pour la classe 1 sont inférieures aux valeurs de référence pour la classe 2 d'une amplitude d'environ 20 mg/L.

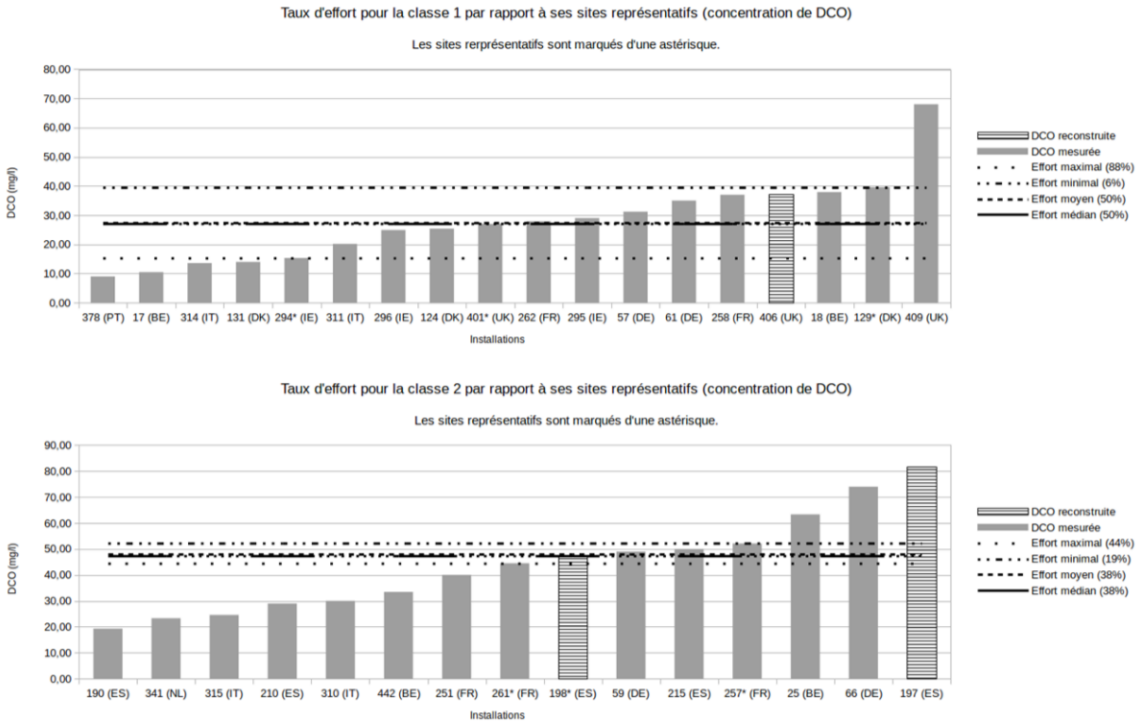


Figure 34 : Taux d'effort pour les deux classes représentatives de l'industrie laitière européenne (concentration de DCO).

1.5.2. Taux d'effort dans les sites performants

En considérant que les sites sélectionnés à l'étape précédente sont ceux qui serviraient à définir ces valeurs de référence, leur impact sur le reste de la population est une information pertinente pour les acteurs du Processus de Séville et pourrait faire le lien avec une étude de faisabilité technico-économique sur leur atteinte.

Dans l'exemple des émissions de DCO, les valeurs de référence obtenues à partir des 17 installations de référence (Tableau 37) sont comparées à l'ensemble de la population (Figure 35). Les sites dits « atypiques » (255, 405, 126) identifiés lors de la classification par l'approche représentative apparaissent ici comme les installations les plus émettrices de DCO, à l'exception du n°313. Le taux d'effort médian vaut 68 % pour cette variable et, à cause de la valeur maximale bien plus élevée que les autres, le taux d'effort moyen est de 10 %. Les valeurs maximale et minimale appartiennent à la sélection et les taux d'effort qui leur sont associés seraient donc respectivement de 0 % et de 100 % si elles servaient de valeurs de référence. Une attention particulière des experts devrait être portée sur l'installation n°126 dont l'émission de DCO est très élevée. Notamment, il

serait opportun de vérifier auprès de l'industriel que ce site rejette bien directement dans le milieu naturel (rejet direct). D'une manière générale, les éléments atypiques constatés au niveau de la base de données ou bien lors de la définition des valeurs de référence devraient être validés et questionnés par les experts du processus de Séville.

Tableau 37 : Valeurs de référence pour l'exemple de la DCO de la sélection performante de l'industrie laitière européenne.

DCO en mg/L	Valeur de référence minimale	Valeur de référence moyenne	Valeur de référence médiane	Valeur de référence maximale
Approche performante	9	205,05	25,40	3064

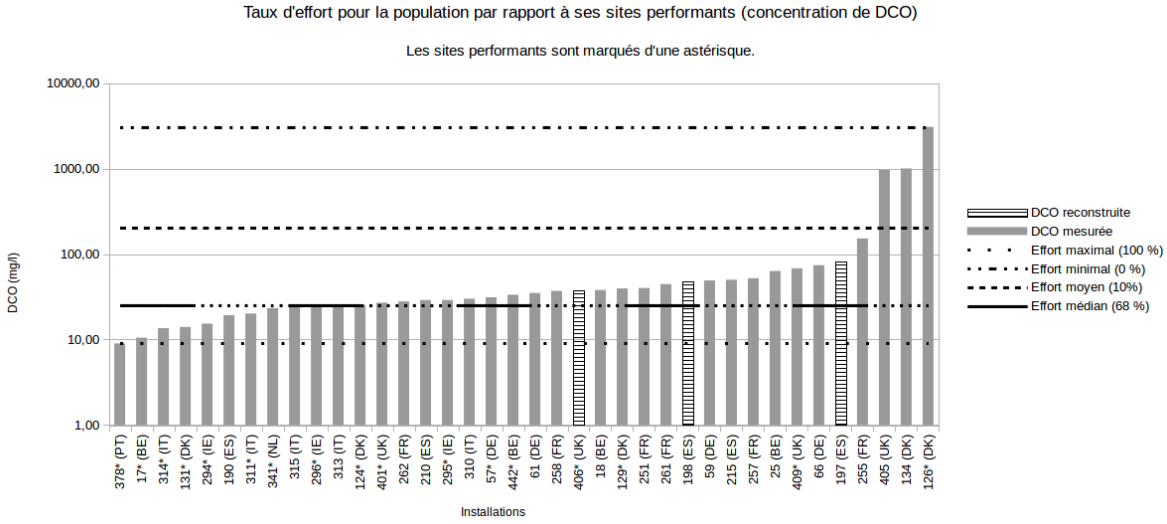


Figure 35 : Taux d'effort par rapport aux installations laitières européennes performantes pour les émissions de DCO.

En synthèse de cette cinquième étape, un tableau regroupant les valeurs de référence possibles pour chaque variable est proposé. Ce tableau (Tableau 38) constitue le livrable 5 de l'étude. Le résultat pour la DCO illustre les différences retrouvées sur les autres variables.

Tableau 38 : Livrable 5 – Valeurs de référence par variable et par classe et taux d'efforts associés de l'industrie laitière européenne exemple de la DCO.

Classes	DCO							
	Minimum		Maximum		Médiane		Moyenne	
	Valeur de référence (mg/L)	Taux d'effort	Valeur de référence (mg/L)	Taux d'effort	Valeur de référence (mg/L)	Taux d'effort	Valeur de référence (mg/L)	Taux d'effort
Classe 1	15,32	88 %	39,60	6 %	26,99	50 %	27,30	50 %
Classe 2	44,50	44 %	52,11	19 %	47,22	38 %	47,94	38 %
Performants	9,00	97 %	3064,00	0 %	25,40	68 %	205,05	10 %

1.5.3. *Aide à l'identification de techniques candidates MTD*

Cette étape est fortement dépendante de la disponibilité de données techniques, *a minima*, d'une indication binaire quant à la présence ou non d'une technique. L'exemple présenté ici consiste à chercher si un système de recyclage de l'eau pourrait être une MTD par rapport à une des valeurs de référence issues des installations performantes. Les consommations d'eau sont comparées aux quatre valeurs de référence qui se trouvent dans le livrable 5 (Annexe 8) : 0,26 (minimum) ; 1,60 (maximum) ; 0,85 (moyenne) ; 0,88 (médiane). Dans la Figure 36, deux groupes d'installations appliquant le recyclage sont visibles. Le premier comporte 7 installations dont les consommations d'eau sont inférieures à la valeur de référence minimale et 9 installations pour lesquelles elles sont supérieures aux valeurs minimales, moyennes et médianes mais inférieures à la valeur de référence maximale. Indépendamment du taux de recyclage effectif dans les installations, il semble donc que le recyclage d'eau présente deux niveaux d'applicabilité sans toutefois que les données disponibles ne permettent d'en comprendre les raisons. De plus, cette conclusion est à modérer avec le jugement d'experts sur la valeur de référence choisie. En effet, si on se basait sur la valeur de référence maximale, alors cette distinction n'aurait plus de raison d'être.

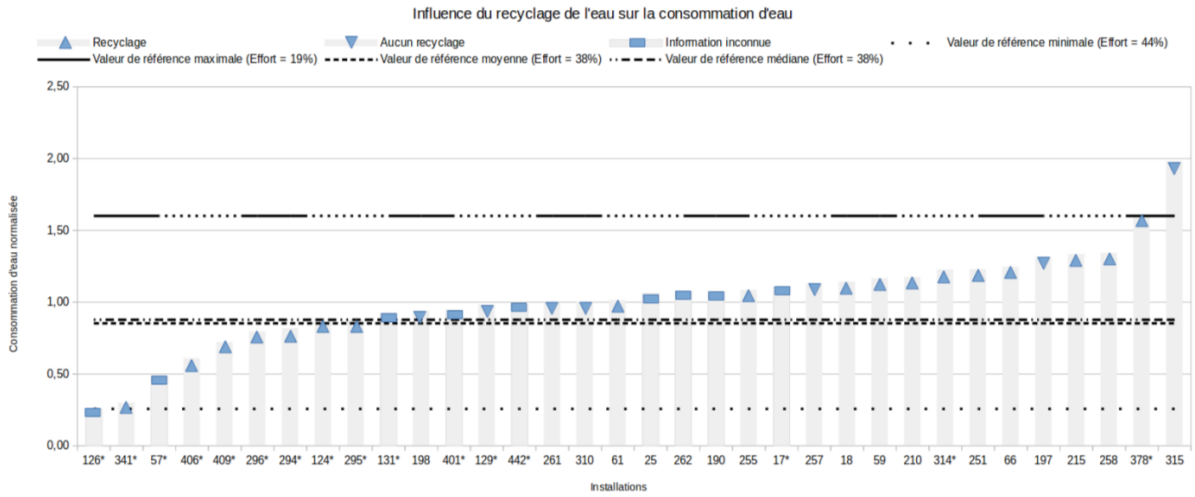


Figure 36 : Livrable 5.2 – Aide au choix du recyclage d'eau comme MTD pour l'industrie laitière européenne.

2. Interprétation et conclusions

Ce cas d'application aux installations laitières européennes illustre l'applicabilité de la totalité de la méthodologie à un secteur IED à partir des données collectées au cours du Processus de Séville. La méthodologie a permis de réduire le nombre de variables d'analyse via la définition de

plusieurs thématiques analysées séparément, car soit elles concernent des sous-populations précises, soit la normalisation choisie ne s'applique pas à l'ensemble des variables de consommations et d'émissions disponibles. Ainsi, des installations de référence ont été identifiées pour les deux approches (représentative et performante). Des valeurs de référence ont ensuite été retenues comme base de détermination des BATAEPL. Puis, les taux d'effort permettant d'évaluer l'effet de l'application des valeurs de référence comme valeur limite d'émission (VLE) ont été émis. La méthodologie a donc été déroulée dans sa globalité. Comme le montre le Tableau 39, l'utilisation de cette méthodologie dans le Processus de Séville permettrait d'apporter un éclairage au regard des différentes étapes. En particulier, suite à l'activation du TWG, elle permettrait de définir la quantité nécessaire de données à collecter et de la concentrer sur les variables d'analyse les plus problématiques en cas de manque. Après la réunion de lancement, elle permettrait de vérifier que les données sur les KEI soient suffisantes pour effectuer une analyse de qualité.

Tableau 39 : Étapes de la méthodologie et tiroirs méthodologiques en fonction du niveau d'intégration dans le Processus de Séville.

	Étapes du Processus de Séville		
	Activation TWG	Réunion de lancement	Analyse données européennes
Étape 1 : Périmètre et paramètres	Éclairage sur les variables d'analyse les plus importantes pour caractériser un secteur industriel en soulignant des liens entre variables et sur les KEI.	Vérification de la pertinence des KEI par rapport à la quantité d'information unique qu'ils apportent.	Réduction du nombre de variables d'analyse.
Étape 2 : Traitement des données	Éclairage sur la quantité de données nécessaires à la collecte et les variables d'analyse les plus problématiques en cas de manque.	Vérification que les données sur les KEI sont suffisantes pour permettre une analyse de qualité suffisante.	Validation que la collecte de données sur les installations de référence soit suffisante pour en déduire des MTD et BATAEPL.
Étape 3 : Classification	Typologie des installations nationales par rapport aux données de consommation et d'émission disponibles.	Classification des installations nationales par rapport aux KEI en préalable à la sélection d'installations de référence.	Classification des installations de référence européennes pour définir des BATAEL différenciées ou non.
Étape 4 : Analyse de la classification	Proposition de sites de référence ou intéressants pour les visiter et comprendre des enjeux et spécificités.	Aide pour dresser la liste des installations de référence qui seront transmises à Séville et estimation de BATAEPL nationales.	Proposition de BATAEPL pour trouver des MTD
Étape 5 : Interprétation de la sélection	Éclairage sur les disparités de performance entre installations d'un même pays.	Vérification de la pertinence des installations de référence par rapport au taux d'effort, valeurs de référence et MTD candidates nationales	Proposition de BATAEPL et MTD candidates.

De plus, les deux cas d'applications au secteur laitier ont montré que les données collectées à l'heure actuelle ne sont pas suffisantes pour parvenir à tirer des conclusions sur des techniques candidates aux MTD. Néanmoins, l'application de la méthodologie dès les premières phases du Processus de Séville semble pouvoir permettre de révéler les données nécessaires à une compréhension des techniques mises en œuvre et leurs liens avec les consommations et émissions.

Une boucle de rétroaction semble donc nécessaire pour assurer que les données permettent de tirer des conclusions sur les techniques.

Pour des raisons de confidentialité, l'étude du positionnement des sites français entre l'application française et européenne n'est pas publiable dans ce mémoire ; les codes des installations anonymisées différemment ne permettant pas de faire de recoupement. La seule comparaison imparfaite qui puisse être effectuée porte sur les valeurs de référence pour les classes performantes (annexes 6 et 8). Ainsi, les valeurs de référence trouvées étaient globalement plus faibles pour les installations européennes, sauf pour les émissions d'azote total pour lesquelles les sites français de référence présentaient des valeurs plus faibles qu'au niveau européen. En l'absence d'un positionnement plus fin, l'interprétation de ces résultats ne peut être plus détaillée.

Chapitre 3 : Application aux INB

L'application aux installations nucléaires de base (INB) de la méthodologie proposée a pour finalité de valider l'applicabilité de celle-ci en contexte non-IED. Pour cela, la méthodologie doit permettre de positionner les installations au regard de leurs performances environnementales, en lien avec les techniques existantes :

- définir les variables d'analyse les plus pertinentes au regard des émissions et de l'impact sur l'environnement dans son ensemble, à partir d'une confrontation du jugement d'experts à la sélection purement statistique ;
- classifier les installations en fonction de leurs performances environnementales globales, en lien avec les techniques mises en œuvre et avec les caractéristiques (notamment environnementales) des installations ;
- établir des liens entre performances environnementales globales, techniques mises en œuvre et caractéristiques (notamment environnementales) des installations.

Pour répondre à ces objectifs, la méthodologie a été appliquée dans un premier temps à l'échelle nationale puis à l'échelle internationale. Les cinq étapes ont ainsi été déroulées et les résultats sont présentés dans les deux sections suivantes. Plus précisément, les deux premières étapes de la méthodologie ont servi à répondre au premier objectif (sélection des variables d'analyse) et les étapes 3, 4 et 5 aux deux derniers objectifs. Il est à noter que seules les données nationales contiennent des informations relatives aux techniques mises en œuvre et couplées aux caractéristiques (notamment environnementales) des installations.

Les données utilisées pour l'application de la méthodologie au secteur nucléaire sont des données fictives mais cohérentes et réalistes au regard de la situation réelle du secteur.

1. Application à l'échelle nationale

Le territoire français compte 19 centrales nucléaires en activité. Les données quantitatives de performances environnementales incluent principalement des données de prélèvement d'eau et d'émissions, liquides et gazeuses, thermiques, chimiques et radioactives.

1.1. Étape 1 : Définition du périmètre et des variables d'étude

Cette étape consiste à définir la population d'installations qui sera étudiée (périmètre), ainsi qu'à répartir les variables dans les trois catégories (variables d'analyse, de normalisation et descriptives). Les informations utilisées pour l'application à l'échelle nationale sont les suivantes :

- **population :** 19 centrales nucléaires françaises ;
- **données quantitatives de performance environnementale :** 55 variables de consommations et d'émissions annuelles ;
- **variable de normalisation :** puissance électrique installée ramenée au nombre de jours annuels de fonctionnement du site ;
- **variables descriptives :** 11 variables.

Les variables d'analyse pour cette application sont les données quantitatives de performance environnementale. Or, le nombre de données est ici supérieur au nombre d'installations de la population, ce qui rend les méthodes de classification de l'étape 3 inapplicables (cf. partie 2). Par conséquent, il est nécessaire de sélectionner les données les plus pertinentes pour constituer un jeu de variables d'analyse qui soit à la fois compatible avec l'objectif de classification au regard de l'environnement dans son ensemble et avec l'applicabilité des méthodes de classification.

Afin de constituer un jeu de variables d'analyse le plus pertinent d'un point de vue industriel et statistique, la sélection des données doit être faite en prenant en compte les avis d'experts et les résultats des tests de dépendance statistique entre variables.

1.1.1. Jugement d'experts dans le choix des variables

Ce choix de variables d'analyse reflète le besoin d'avoir un classement au regard de l'ensemble des enjeux environnementaux des centrales nucléaires françaises que sont les prélèvements et consommations d'eau, les rejets thermiques, chimiques et radioactifs, le risque de dispersion des micro-organismes pathogènes, les déchets et le bruit. Compte-tenu de ce besoin, 14 variables ont été sélectionnées par deux ingénieurs d'EDF (Tableau 40). Toutes ces variables étant complètes, la reconstruction n'est pas nécessaire ici.

Tableau 40 : Liste des 14 variables d'analyse choisie sur avis d'experts pour les INB françaises.

#	Catégorie	Variable
1	Prélèvement et consommation d'eau	Volume entrant
2		Volume d'eau évaporée
3	Rejets chimiques	CRT
4		AOX
5		Chlorures
6		Acide borique
7		Hydrazine
8		Morpholine/Éthanolamine
9		Phosphates
10	Rejets radioactifs	Iodes gazeux
11		Gaz rares
12		Autres PF/PA liquides
13	Rejets thermiques	Rejets thermiques
14	Déchets	Résines

En complément de cette sélection « experts », et compte tenu du nombre important de variables de consommations et d'émissions disponibles dans la base de données, des tests statistiques ont été également effectués afin de réduire le nombre de variables d'analyse aux plus pertinentes pour la classification.

1.1.2. Sélection des variables d'analyse par les tests statistiques

Le principe est de proposer une aide au choix des variables d'analyse ; soit par inclusion des variables indépendantes, soit par exclusion des variables dépendantes, jusqu'à l'obtention d'un nombre de variables d'analyse aussi réduit que possible. Ainsi, des tests statistiques (tests de Spearman) ont été effectués sur chaque couple possible de variables d'analyse afin d'évaluer leur degré de dépendance. Pour ce faire, deux éléments sont pris en compte : la part de données manquantes et leur degré de dépendances (++ ou +) ou d'indépendances (- ou --).

En cas de dépendances avérées, l'une des deux variables d'analyse exclues peut passer dans le groupe des variables descriptives et ne pas servir à la classification, ce qui permet de diminuer le nombre de variables d'analyse tout en conservant un maximum d'information. Les variables incomplètes sont remplacées en priorité. Dans l'exemple illustré dans le Tableau 40, les données de rejet d'acide borique apparaissent comme dépendantes avec les données de rejets de Pb, Al et Zn.

Tableau 41 : Extrait de la table des dépendances entre variables pour les INB françaises.

Variables		Pb (kg/an)	Al (kg/an)	Cr (kg/an)	Na (kg/an)	Cu (kg/an)	Zn (kg/an)	Volume entrant (Mm^3/an)	CRT (kg/an)	Acide borique (kg/an)	Autres PF-PA liquides (GBq/an)	
	Données absentes	6	4	4	4	1	1	0	0	0	0	
Pb(kg/an)	6											
Al(kg/an)	4											
Cr(kg/an)	4	+										
Na(kg/an)	4											
Cu(kg/an)	1											
Zn(kg/an)	1		+			++						
Volume entrant(Mm^3/an)	0	+	++		-							
CRT(kg/an)	0	-			++			-				
Acide borique(kg/an)	0	+	+				++		-			
Autres PF-PA liquides(GBq/an)	0		++			+		++		+		
Rejets thermiques (Mth/an)	0		+					++	-		++	...
...	...											...

Ainsi à partir de ces résultats (Tableau 41), le choix peut s'opérer de deux façons :

1. préférentiellement, à partir d'échanges avec les experts ;
2. à défaut, les variables dont dépendent le plus grand nombre d'autres variables sont choisies en priorité dans un objectif de réduire au maximum le nombre de variables d'analyse pour rendre les méthodes utilisées aux étapes 3 ; 4 et 5 les plus efficaces possibles. La validité des choix effectués sera ensuite vérifiée par la classification elle-même qui montrera le caractère discriminant des variables sélectionnées avec les boîtes à moustaches (Figure 38).

En appliquant la deuxième façon de sélectionner les variables d'analyse, une sélection de 14 variables d'analyse est possible ; elle est présentée dans le Tableau 42. Seules les données relatives aux rejets de chrome (Cr) sont incomplètes (21 % de données manquantes) et seront susceptibles d'être reconstruites. La qualité de sa reconstruction pouvant éventuellement entraîner son exclusion. C'est ici le rôle central de l'expert de juger de l'acceptabilité ou non de la reconstruction en fonction du degré d'incertitude jugé recevable. Par ailleurs, les 30 variables exclues des variables d'analyse deviennent des variables descriptives qui pourront être utilisées pour caractériser la classification réalisée lors de la troisième étape.

Tableau 42 : Liste des 14 variables suggérées suite à l'étude des dépendances entre variables d'analyse initiales des INB françaises.

#	Variable	% manquant
1	Chrome (kg/an)	21 %
2	Azote total (kg/an)	0 %
3	Phosphates (kg/an)	0 %
4	Volume IN (Mm^3/an)	0 %
5	Acide borique (kg/an)	0 %
6	Tritium liquide (TBq/an)	0 %
7	Iodes gazeux (GBq/an)	0 %
8	Autres PF/PA liquides (GBq/an)	0 %
9	Gaz rares (TBq/an)	0 %
10	Hydrazine (kg/an)	0 %
11	CRT (kg/an)	0 %
12	AOX (kg/an)	0 %
13	Thermique (Mth/an)	0 %
14	Résines (m^3/an)	0 %

1.1.3. Conclusion et livrable de l'étape 1

La comparaison des variables suggérées par l'analyse statistique et par jugement d'experts (Tableau 43) montre que trois variables seulement sur 14 diffèrent. Ces trois variables sont les suivantes : rejets de Chrome, d'Azote Total et de Tritium liquide pour la sélection statistique ; Volume évaporé, rejets de Morpholine / Éthanolamine et rejets de Chlorures pour la sélection d'experts. Le choix final porte donc sur les trois variables spécifiques à la sélection statistique qui ont été jugées moins intéressantes par les experts. Par conséquent, les 14 variables choisies par les experts sont celles qui seront utilisées dans les étapes suivantes.

Tableau 43 : Comparaison des variables identifiées par les experts et par l'analyse statistique pour les INB françaises.

Experts	Analyse statistique
Acide borique (kg/an)	Acide borique (kg/an)
AOX (kg/an)	AOX (kg/an)
Autres PF/PA liquides (GBq/an)	Autres PF/PA liquides (GBq/an)
Chlorures (kg/an)	**Azote total (kg/an)**
CRT (kg/an)	CRT (kg/an)
Gaz rares (TBq/an)	Gaz rares (TBq/an)
Hydrazine (kg/an)	Hydrazine (kg/an)
Iodes gazeux (GBq/an)	Iodes gazeuses (GBq/an)
Phosphates (kg/an)	Phosphates (kg/an)
Rejets thermiques (Mth/an)	Rejets thermiques (Mth/an)
Résines (m^3/an)	Résines (m^3/an)
Volume d'eau évaporé (Mm^3/an)	**Tritium liquide (TBq/an)**
Volume entrant (Mm^3/an)	Volume entrant (Mm^3/an)
Morpholine/Éthanolamine (kg/an)	**Chrome (kg/an)**

Le livrable de cette étape (Tableau 44) est la base de données initiale regroupant l'ensemble des variables classées dans les trois catégories « variables d'analyse », « variables descriptives » et « variables de normalisation ».

Tableau 44 : Livrable 1 – Extrait de la base de données initiale pour les INB françaises.

Site	Variables d'analyse			Variables descriptives		Variable de normalisation
	AOX (kg/an)	…	Résines (m³/an)	Type	…	Production (MWe / jour)
FR01	0	…	12,6	CP0	…	7,20
FR02	0	…	2,1	P'4	…	16,77
FR03	1110	…	15,2	CP2	…	8,71
FR04	117	…	2,5	CP2	…	5,86
FR05	0	…	14,5	CP1	…	10,81
FR06	4634	…	8,4	P4	…	25,24
FR07	211	…	6,3	P'4	…	50,98
FR08	27 156	…	8,4	CP1	…	10,95
FR09	467	…	18,9	P'4	…	43,33
FR10	8425	…	19,7	P4	…	19,69
FR11	186	…	12,3	CP0	…	8,71
FR12	0	…	25,2	P'4	…	12,77
FR13	593	…	10,5	N4	…	22,48
FR14	0	…	8,4	N4	…	15,51
FR15	0	…	14	CP2	…	16,14
FR16	154	…	6	CP1	…	11,28
FR17	8004	…	16,8	P'4	…	44,82
FR18	0	…	14,7	P4	…	24,52
FR19	0	…	5,5	CP1	…	14,63
Nombre de données disponibles	19	…	19	19	…	19
Nombre de données manquantes	0	…	0	0	…	0
% de données manquantes	0 %	…	0 %	0 %	…	0 %

1.2. Étape 2 : Traitement et consolidation des données

Dans la mesure où les données sont complètes pour les variables d'analyse, cette étape de reconstruction n'est pas appliquée.

1.3. Étape 3 : Classification des installations

Comme indiqué dans le chapitre 2, cette étape consiste à classifier les installations au regard de deux approches, une approche dite représentative et une approche dite performante.

1.3.1. *Approche représentative et caractérisation des classes*

La classification par l'approche représentative a abouti à la constitution de deux classes (Figure 37). Le découpage des classes correspond à la plus haute branche de l'arbre : une classe de 8 sites et une classe en contenant 10.

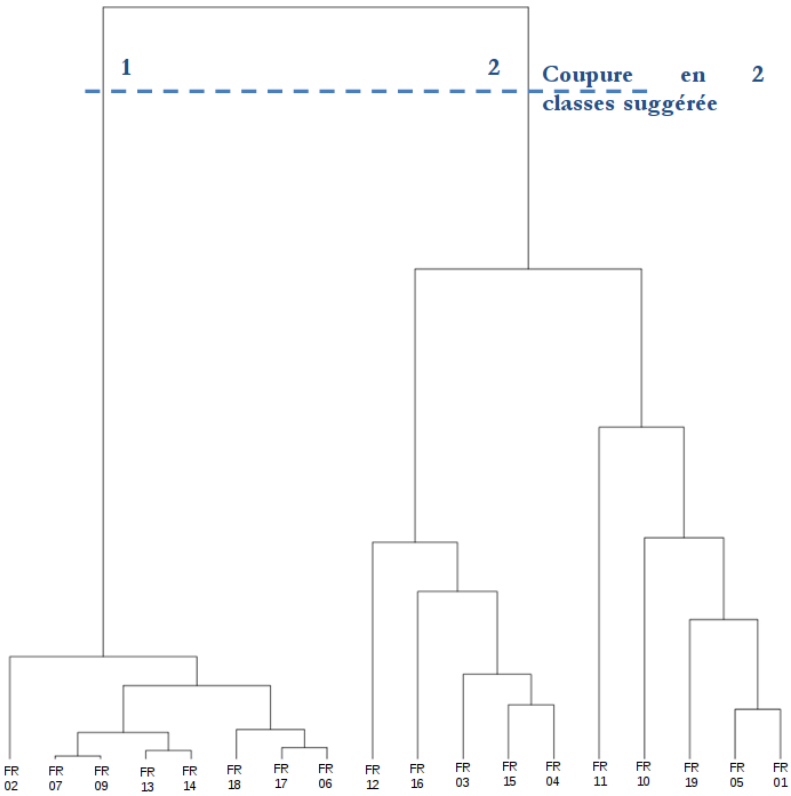
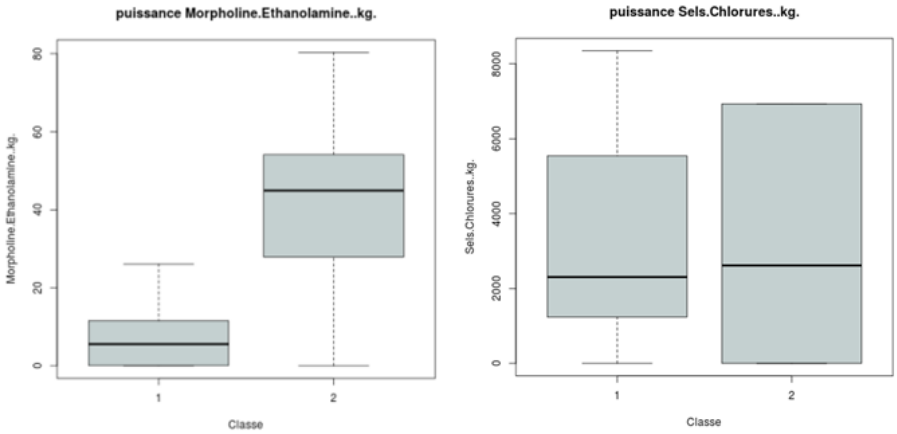


Figure 37 : Dendrogramme de la classification représentative des INB françaises.

Pour certaines variables, la classe 1 contient des sites moins consommateurs et émetteurs que la classe 2 et une dispersion moindre des valeurs comme dans l'exemple des émissions de Morphine-Éthanolamine (Figure 38a). Pour d'autres variables, comme dans le cas des chlorures par exemple, la distinction entre les classes n'est pas évidente dans la mesure où les valeurs se confondent (Figure 38b).



(a) (b)

Figure 38 : Comparaison des classes représentatives des INB françaises pour 2 des 14 variables d'analyse.

La comparaison des classes avec les variables descriptives quantitatives présente des résultats similaires avec les conclusions établies pour les variables d'analyse. La seule corrélation mise en exergue entre ces deux classes et les variables descriptives qualitatives concerne le nombre de tranches (Tableau 45). Cela révèle également qu'avec un aussi petit nombre d'installations, la limite d'applicabilité des outils statistiques est atteinte.

1.3.2. Approche performante et caractérisation des classes

Sur les 19 sites français étudiés, 13 apparaissent sur le front de Pareto au travers de l'approche performante. Dans ce cas d'application, il a été choisi d'utiliser l'Indicateur de Niveau de Performance Environnementale (INPE) pour sélectionner les installations les plus performantes au regard de l'environnement dans son ensemble (§1.4.2).

1.3.3. Livrables de l'étape 3

Les deux livrables de l'étape sont d'une part un tableau récapitulant les résultats du test d'hypothèse entre le classement et les variables descriptives (Tableau 45), et d'autre part le classement des installations (Tableau 46) pour les deux approches appliquées ici. Le premier livrable permet de constater si certaines caractéristiques se retrouvent majoritairement dans une classe ou l'autre. Le Tableau 45 montre des résultats différents selon l'approche utilisée. Dans le cas de l'approche représentative, il semble que les classes définies soient différenciées par le nombre de tranches. Dans l'approche performante, il semble que la localisation du site influe sur son classement.

Tableau 45 : Livrable 3.1 – Dépendances entre classes et variables descriptives pour les INB françaises.

Variables	Localisation	Nombre de tranches	…	Conditionnement
Classement représentatif		+	…	
Classement performant	++		…	

Tableau 46 : Livrable 3.2 – Liste des INB françaises classées selon les approches représentatives et performantes.

Installations représentatives		Installations performantes
Classe 1	Classe 2	
FR02	FR01	FR01
FR06	FR03	FR02
FR07	FR04	FR05
FR09	FR05	FR06
FR13	FR10	FR07
FR14	FR11	FR09
FR17	FR12	FR10
FR18	FR15	FR14
	FR16	FR15
	FR19	FR16
		FR17
		FR18

1.4. Étape 4 : Analyse de la classification

Cette étape consiste à analyser les résultats en vue d'établir une liste hiérarchisée d'installations représentatives ou performantes. Pour l'approche représentative, cela correspond à ordonner les sites en fonction de leur proximité au centre de leur classe (du plus proche au plus éloigné) et au centre des autres classes (du plus éloigné au plus proche). Pour l'approche performante, la hiérarchisation ne concerne que le cas où un nombre trop important de sites se trouverait sur le front de Pareto par rapport à la population totale. Dans ce cas, l'INPE permet d'ordonner les sites se trouvant sur le front et de sélectionner le nombre de sites souhaité par l'expert.

1.4.1. Liste de sites représentatifs

Pour rappel, dans le cadre de l'approche représentative, les sites les plus pertinents d'une classe sont ceux pour lesquels les distances inter-classes et intra-classes sont respectivement les plus élevées et les plus faibles. La sélection des sites se fait suivant la représentation de ces distances inter-classes et intra-classes et les décisions d'experts. La Figure 39 présente les résultats pour les deux classes obtenues à l'étape 3.

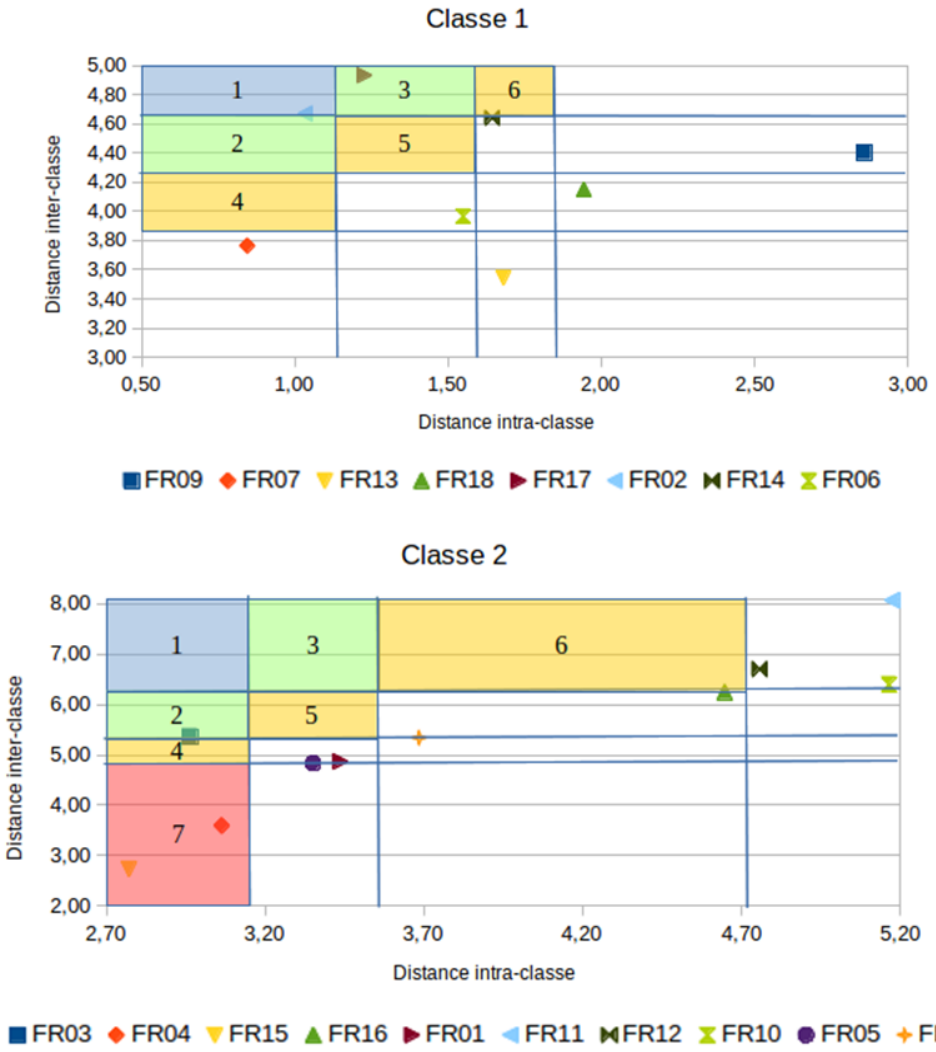


Figure 39 : Outils d'aide à la sélection des INB françaises de référence par couplage des distances intra-classes et inter-classes.

La méthodologie fournit une liste ordonnée des installations dans l'ordre croissant des zones présentées sur la Figure 39 et synthétisées dans le Tableau 48. Ensuite, les experts sélectionnent le nombre d'installations qui leur convient en suivant cet ordre et en incluant éventuellement d'autres considérations (ici, 3 sites dans chacune).

1.4.2. Liste de sites performants

Pour rappel, dans le cadre de l'approche performante, tous les sites se trouvant sur le front de Pareto sont pertinents. Néanmoins, si leur nombre est trop important, l'aide proposée par la méthodologie consiste en l'attribution d'une note en fonction des coordonnées factorielles issues

d'une analyse en composantes principales pondérées par la quantité d'information portée par chacun des axes sont les plus faibles (cf. INPE – Partie 2, 4.2). Le Tableau 47 présente les sites performants ordonnés grâce à cet outil. La sélection repose ensuite sur le jugement d'experts afin de mettre cette liste ordonnée en perspective avec d'autres facteurs que les données ne permettent pas de prendre en compte (contraintes environnementales locales, réglementaires, etc.). Ainsi, les trois sites en gras dans le Tableau 47 sont les sites de référence qui vont être utilisés dans la suite de l'étude.

Tableau 47 : Installations nucléaires françaises performantes classées selon l'INPE.

#	Installations performantes	
	Code	INPE
1	**FR07**	**637**
2	FR09	604
3	**FR17**	**558**
4	FR02	474
5	FR06	451
6	FR18	423
7	FR14	414
8	FR15	333
9	FR19	292
10	FR16	210
11	**FR05**	**176**
12	FR10	160
13	FR01	0

1.4.3. Livrable de l'étape 4

Le livrable de cette étape est un tableau listant les installations par ordre de pertinence en fonction des deux approches de classifications (Tableau 48). Pour les installations représentatives, les distances intra – et inter-classes qui ont servi à la construction du nuage de points de la Figure 39 sont indiqués ; pour les installations performantes, les valeurs de l'INPE pour les installations se trouvant sur le front de Pareto sont indiquées.

Tableau 48 : Livrable 4 – Listes ordonnées des INB françaises pour les approches représentatives et performantes.

#	Sites représentatifs								Sites performants	
	Classe 1				Classe 2					
	Code	Distance intra-classe	Distance inter-classe	Zone	Installation	Distance intra-classe	Distance inter-classe	Zone	Code	INPE
1	**FR09**	**1,03**	**4,67**	**1**	**FR03**	**2,96**	**5,36**	**2**	**FR07**	**637**
2	**FR07**	**1,23**	**4,93**	**3**	**FR04**	**3,06**	**3,59**	**7**	FR09	604
3	FR13	0,84	3,76	7	FR15	2,77	2,73	7	**FR17**	**558**
4	FR18	1,55	3,97	8	FR16	4,65	6,25	8	FR02	474
5	FR17	1,64	4,64	9	FR01	3,44	4,87	9	FR06	451
6	**FR02**	**2,86**	**4,40**	**13**	FR11	5,18	8,07	10	FR18	423
7	FR14	1,68	3,55	14	**FR12**	**4,76**	**6,70**	**10**	FR14	414
8	FR06	1,94	4,15	15	FR10	5,17	6,40	10	FR15	333
9					FR05	3,35	4,83	11	FR19	292
10					FR19	3,68	5,34	12	FR16	210
11									**FR05**	**176**
12									FR10	160
13									FR01	0

Le choix des installations pertinentes proposé dans la méthodologie consiste à suivre l'ordre indiqué dans le livrable 4 (Tableau 48). La sélection finale se fait avec l'appui des experts qui apportent des informations complémentaires relativement aux caractéristiques globales (environnementales, locales, réglementaires, techniques, etc.) des installations proposées. Ainsi, une sélection de trois installations dans chacune des classes a été décidée ; les deux premiers sites correspondant aux recommandations de la méthodologie et le dernier servant à illustrer des spécificités connues des experts (sites en gras dans le Tableau 48).

1.5. Étape 5 : Interprétation des sélections

Deux niveaux d'interprétation sont proposés ici en vue, d'une part, de proposer des valeurs pertinentes en termes d'indicateurs de performance et, d'autre part, d'identifier l'influence des caractéristiques globales (environnementales, locales, réglementaires, techniques, etc.) de chaque installation sur ces valeurs.

1.5.1. Détermination des valeurs de référence potentielles et du taux d'effort

Des valeurs de référence potentielles différentes entre l'approche représentative et performante sont définies à cette étape. Celles-ci impliquent donc des taux d'effort différents.

1.5.1.1. Sites représentatifs

Dans un premier temps, les valeurs de référence potentielles (moyenne, médiane, maximale, minimale) sont calculées par classe à partir de la sélection de l'étape 4. Dans un deuxième temps, le taux d'effort est calculé ; il représente le pourcentage de l'ensemble des installations de la classe

devant « faire un effort » pour atteindre la valeur de référence relative indiquée (et donc des sites ayant des valeurs par variable au-dessus de la valeur de référence). La Figure 40 illustre la situation des sites de chaque classe par rapport aux valeurs de référence et le Tableau 49 présente un extrait des taux d'effort associés aux valeurs de référence pour chacune des deux classes identifiées.

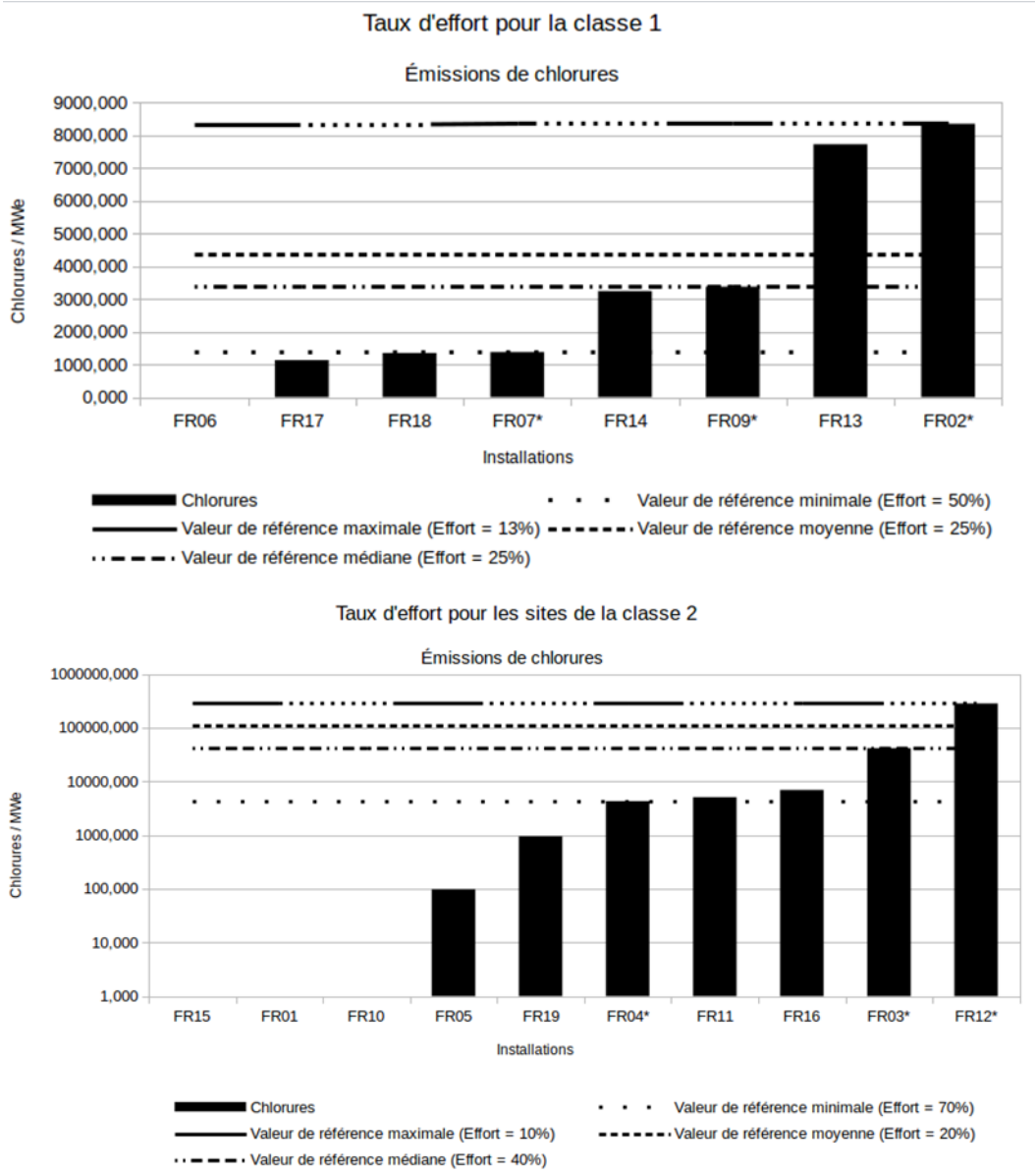


Figure 40 : Positionnement des INB françaises par rapport aux valeurs de référence pour les deux classes représentatives.

Tableau 49 : Taux d'effort pour les 2 classes représentatives des INB françaises.

Classes	Chlorures			
	Minimum	Maximum	Médiane	Moyenne
Classe 1	50 %	0 %	25 %	25 %
Classe 2	40 %	0 %	10 %	10 %

1.5.1.2. Sites performants

Les valeurs de référence potentielles des sites performants sont calculées par rapport à l'ensemble de la population. Le Tableau 50 présente les taux d'effort par rapport aux valeurs maximales des sites se trouvant sur le front de Pareto. On constate notamment que l'approche par compromis du front de Pareto peut faire apparaître des taux d'effort plus faibles en comparaison de l'approche représentative (par exemple, 25 % pour la classe 1 contre 11 % par rapport aux sites performants).

Tableau 50 : Taux d'effort des INB françaises sur les émissions de chlorures par rapport aux installations performantes.

Classes	Chlorures			
	Minimum	Maximum	Médiane	Moyenne
Performants	53 %	5 %	11 %	42 %

1.5.2. Aide à l'identification des corrélations entre caractéristiques des installations et niveaux de performance

Cette étape consiste à étudier les caractéristiques notamment techniques qui pourraient être potentiellement corrélées à l'atteinte des valeurs de performance données. L'illustration choisie pour exemple concerne les émissions de chlorures des installations. L'objectif est d'essayer d'identifier les corrélations statistiques, à partir des connaissances d'experts, entre les valeurs d'émissions et les caractéristiques des installations et de les confronter aux avis d'experts. La représentation en histogrammes est ici utilisée pour la recherche de techniques (Figure 41).

1.5.2.1. Avis des experts sur les émissions de chlorures

D'après les experts, plusieurs facteurs influencent les émissions de chlorures. Ils correspondent aux variables descriptives du traitement anti-tartre, de la régénération des résines, du type de circuit de refroidissement, du type de condenseur et du traitement biocide. Cependant, les tests statistiques (dépendances) effectués sur ces variables ne permettent pas de retrouver ces informations à partir des données.

Deux explications sont possibles pour l'expliquer :

1. les émissions de chlorures ayant des causes multiples, il n'est pas possible de trouver de corrélation simple et des outils plus avancés seraient nécessaires ;
2. le nombre de sites est insuffisant pour que les tests statistiques fournissent un résultat.

En l'absence de résultat statistique concluant, seul le jugement d'experts est en mesure de fournir une conclusion à cette étape.

1.5.2.2. Interprétation statistique

À titre d'illustration de l'outil d'aide au choix des techniques, les trois principales sources d'émissions de chlorures que sont le traitement biocide, le traitement anti-tartre et la régénération des résines sont positionnées dans la Figure 41. Il apparaît que l'utilisation de l'acide chlorhydrique pour la régénération des résines ou le traitement anti-tartre et l'utilisation de monochloramine pour le traitement biocide sont bien des sources de chlorures.

Les sites qui ont les émissions de chlorures les plus basses utilisent l'acide sulfurique. Il serait pertinent d'approfondir l'étude pour vérifier s'il n'y a pas transfert de pollution des émissions de chlorures aux sulfates en cohérence avec la démarche de prise en compte de l'environnement dans son ensemble. De plus, il serait nécessaire d'étudier d'autres conditions d'utilisation pour le choix d'une technique plutôt que l'autre. En outre, compte tenu du petit nombre de données disponibles, ces conclusions sont à prendre avec prudence et la validation d'un expert est indispensable pour vérifier d'éventuels effets croisés. Cette conclusion illustre néanmoins le type de résultats relatifs à l'analyse des techniques qui peuvent être obtenus avec la méthodologie.

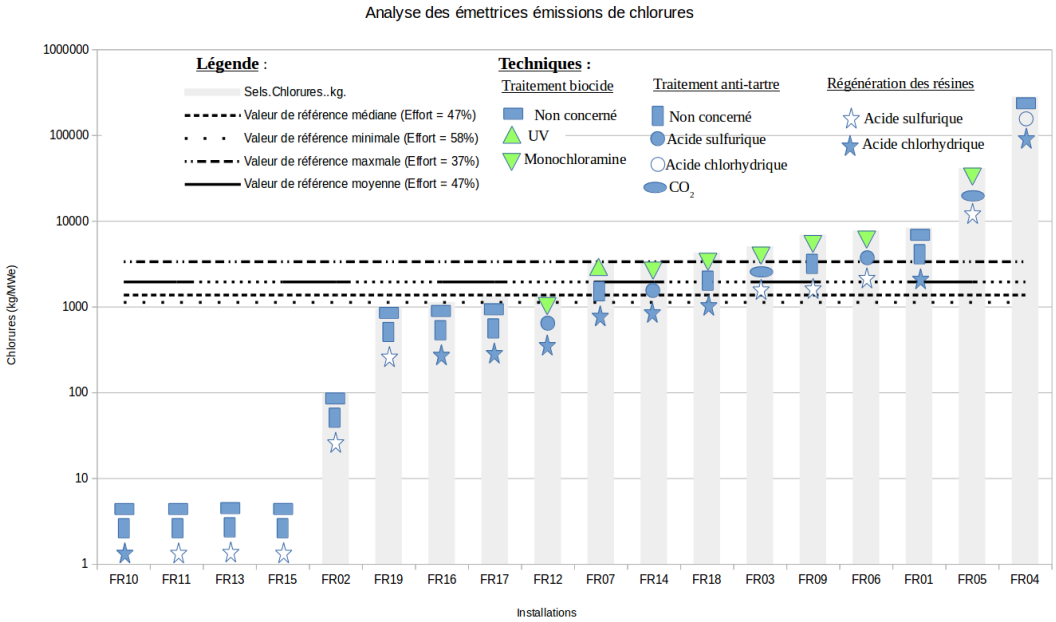


Figure 41 : Livrable 5.2 – Aide au choix des techniques de traitement biocide par rapport aux émissions de chlorures pour les INB françaises.

De même qu'aux deux chapitres précédents, l'application de l'outil d'aide au choix des techniques met en exergue le besoin de cohérence entre les données techniques disponibles et leurs relations avec les différentes variables d'analyse. Au vu de cette multiplicité d'influences, une approche multidimensionnelle et des informations plus détaillées sur les contributions des techniques semblent nécessaires à cette étape.

2. Application à l'échelle internationale

Les données internationales concernent 76 centrales nucléaires incluant les 19 sites français.

2.1. Étape 1 : Définition du périmètre et des variables d'étude

Les spécificités du périmètre d'étude sont les suivantes :

- **population :** 76 installations situées dans 9 pays. Les effectifs par pays sont indiqués dans le Tableau 51, ainsi que le codage permettant d'identifier les sites (code pays – numéro de site) ;

Tableau 51 : Effectifs des INB internationales par pays et codage employé.

Pays	Code pays	Nombre d'installations	Code pays
États-Unis	US	37	US1 à US37
France	FR	19	FR1 à FR19
Allemagne	AL	9	AL1 à AL9
Espagne	ES	4	ES1 à ES4
Belgique	BE	2	BE1 et BE2
Finlande	FI	1	FI1
Pays-Bas	NL	1	NL1
Suède	SW	1	SW1
Royaume-Uni	UK	1	UK1

- **variables d'analyse :** 8 variables d'émission annuelle de radionucléides (rejets gazeux de carbone 14, rejets gazeux d'Iodes, rejets liquides d'Iode 131, rejets gazeux bêta-gamma, rejets liquides bêta-gamma, rejets de tritium liquide, rejets de tritium gazeux et rejets de gaz rares) ;
- **variable de normalisation :** Étant donné la disponibilité des données, la variable de normalisation choisie est la puissance électrique installée (non ramenée au nombre de jours annuel de fonctionnement du site comme fait pour l'application à l'échelle nationale) ;
- **Variables descriptives :** 3 variables (Pays, année de mise en service, nombre de tranches).

Compte tenu des différences significatives de conception et de réglementation (points de contrôle et fréquences d'échantillonnage, méthode de mesure et de comptabilisation, spectre de référence, etc.), la comparaison directe des données de rejets d'effluents radioactifs ne peut être utilisée qu'à titre indicatif. En particulier, dans la plupart des pays, les activités volumiques inférieures à la limite de détection (LD) ne sont pas prises en compte pour l'établissement des bilans de rejets d'effluents radioactifs. A contrario, en France, la règle est plus conservative : les activités volumiques des radionucléides du spectre de référence inférieures ou égales à la LD sont par convention remplacées par une valeur égale à la moitié de cette LD (LD/2) ; pour les autres radionucléides, les résultats sont pris en compte dès lors que l'activité volumique est supérieure à la LD. Ceci peut conduire, pour un même rejet réel, à des différences notables de rejet déclaré. Ainsi, il est nécessaire de rester prudent quant à l'interprétation des données internationales relatives aux rejets d'effluents radioactifs des sites nucléaires.

Par ailleurs, il est à noter que la variable de normalisation utilisée est la puissance nominale installée. Ce choix par défaut (faute d'autre donnée disponible pour les installations internationales) induit également des sous-estimations plus ou moins importantes des valeurs de rejets normalisées. En effet, pour une même valeur de rejet brut, plus l'installation aura fonctionné

dans l'année, plus la valeur de rejet normalisée sera proche de la valeur vraie. De même, plus le nombre de jours d'arrêt de l'installation sera important plus la valeur rejet normalisée sera sous-estimée.

Les variables d'analyse pour cette application sont les données quantitatives de performance environnementale. Le nombre de données étant bien inférieur au nombre d'installations de la population, il n'est pas nécessaire de faire une sélection de variable. Cependant, les données relatives aux rejets gazeux de carbone 14 sont manquantes pour tous les sites étasuniens. Par conséquent, cette variable a été exclue du jeu de variables d'analyse.

La base de données initiale est produite, elle reprend l'ensemble des données correspondant aux variables descriptives, d'analyse et de normalisation pour chaque installation. Le Tableau 52 synthétise l'ensemble des informations et constitue le livrable 1.

Tableau 52 : Livrable 1 – Base de données initiales pour les INB internationales.

	Variables d'analyse			Variables descriptives			Variable de normalisation
Site	**I131 Liquide (GBq/an)**	**…**	**Tritium Liquide (TBq/an)**	**Pays**	**…**	**Nombre de tranches**	**Puissance (MWe)**
BE01		…	41,7	Belgique	…	4	2 775
BE02	9,79E-3	…	33,7	Belgique	…	3	2 037
AL01	1,9E-4	…	18,6	Allemagne	…	2	2 407
AL02		…	22	Allemagne	…	1	1 370
…	…	…	…	…	…	…	…
Nombre de données disponibles	63	…	75	76	…	76	76
Nombre de données manquantes	13	…	1	0	…	0	0
% de données manquantes	17 %	…	1,3 %	0 %	…	0 %	0 %

2.2. Étape 2 : Traitement des données

Cette étape consiste à reconstruire les données. Au regard du jeu de variables d'analyse défini à l'étape précédente, il est nécessaire d'estimer l'influence des données manquantes après reconstruction et en calculant leurs taux d'erreur.

Suite à la première reconstruction des données avec MissForest, des premiers niveaux d'erreur peuvent être estimés. Malgré une erreur globale de 95 %, la Figure 42 présente des erreurs de plus de 150 % pour 5 variables et de 59 % pour les émissions liquides de tritium. Avant d'envisager

d'exclure ces variables, une recherche des sites qui seraient les principales causes d'erreur a été effectuée.

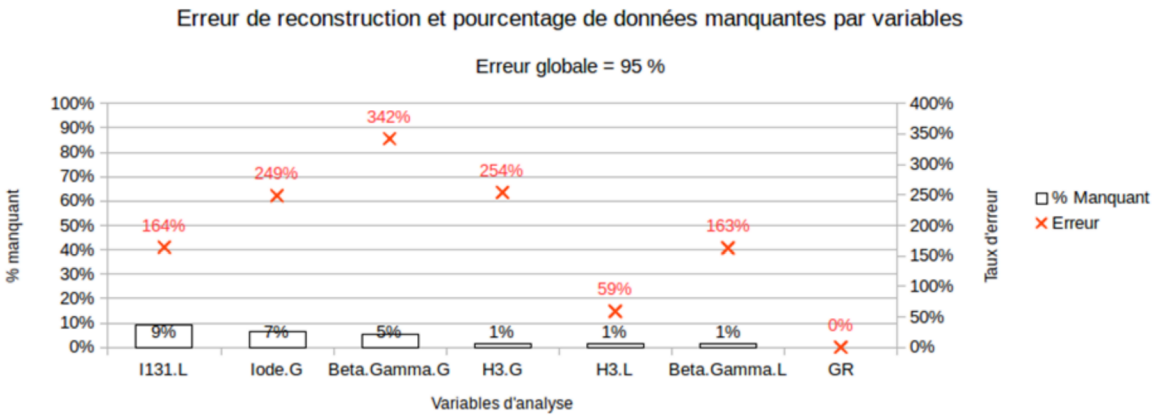


Figure 42 : Erreur de reconstruction des données sur les INB internationales initialement sélectionnées.

Ainsi, il est apparu que l'exclusion de l'installation US21 permet de réduire fortement l'erreur globale (qui passe à 3 %) en rendant la variable émission gazeuses de tritium complète. Néanmoins, l'erreur associée change peu pour les autres variables d'analyse. Les variables d'analyse ayant été choisies par les experts, il a été décidé ici de ne pas en exclure, ni d'exclure d'autres installations compte tenu de la faible réduction de l'erreur globale engendrée par de nouvelles exclusions de sites.

2.2.1. Livrables

La base de données reconstruite et normalisée est produite à l'issue de cette étape pour les 75 installations nucléaires de base restantes. Elle constitue le livrable 2 et est présentée dans le Tableau 53.

Tableau 53 : Livrable 2 – Matrice de données reconstruites et normalisées des INB internationales.

Site	Variables d'analyse			Variables descriptives			Variable de normalisation
	Gaz rares (TBq/MWe)	…	Tritium Liquide (TBq/MWe)	Pays	…	Nbre. de tranches	Puissance (MWe)
BE01	6,13E-6	…	1,5E-2	Belgique	…	4	2 775
BE02	1,41E-2	…	1,6E-2	Belgique	…	3	2 037
AL01	1,45E-4	…	7,7E-3	Allemagne	…	2	2 407
AL02	3,79E-4	…	1,6E-2	Allemagne	…	1	1 370
…	…	…	…	…	…	…	…
Nombre de données disponibles	63	…	75	76	…	76	76
Nombre de données manquantes	13	…	1	0	…	0	0
% de données manquantes	17 %	…	1,3 %	0 %	…	0 %	0 %
Erreur de la reconstruction	0 %	…	51 %	0 %	…	0 %	0 %

2.3. Étape 3 : Classification des installations

La classification se fera suivant les deux approches proposées : approche représentative d'une part et approche performante d'autre part.

2.3.1. Approche représentative et caractérisation des classes

La classification par l'approche représentative a réparti la population dans six classes présentées sur la Figure 43. La première contient 16 installations, la classe n°2 contient 44 sites, dont les 19 installations françaises, la classe n°6 contient 8 installations et les trois autres classes contiennent de 1 à 3 sites. Par défaut, dans la méthodologie, une classe contenant moins de 10 % de la population est considérée comme regroupant des installations ayant des émissions spécifiques et atypiques par rapport aux autres sites sur au-moins une des variables d'analyse. Il est donc proposé aux experts des investigations supplémentaires afin d'identifier si un traitement particulier doit être opéré pour ces sites. Ainsi, dans notre étude, seules les classes n°1, 2 et 6 seront considérées dans la suite de l'analyse des classes.

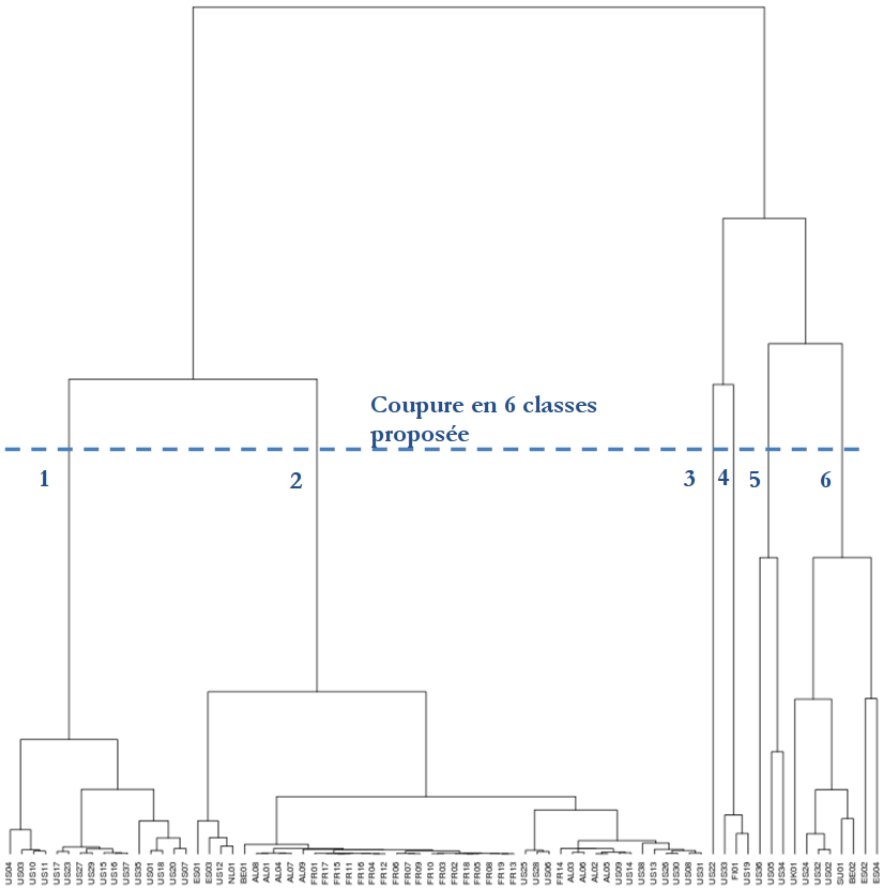


Figure 43 : Dendrogramme de la première classification représentative des INB internationales.

L'analyse des classes, c'est-à-dire les tests d'hypothèse entre une variable constituée du numéro de la classe et des trois variables descriptives (pays, année de mise en service, nombre de tranches), montre que le pays influence le plus la classification (Tableau 54).

Tableau 54 : Livrable 3.1 – Dépendances entre le classement représentatif et les variables descriptives des INB internationales.

Variables	Pays	Année de mise en service	Nombre de tranches
Classification représentative	**++**		

Les boîtes à moustaches permettant de comparer les classes 1, 2 et 6 par rapport aux variables d'analyse (Figure 44) et indiquent que la classe 2 est composée d'installations dont les valeurs des variables d'analyse sont les plus faibles (comparativement aux autres classes) et dont la dispersion est relativement réduite malgré un plus grand nombre d'installations. Néanmoins, la distinction

avec la classe 2 n'est franche que pour deux variables : émissions gazeuses bêta-gamma et émissions liquides de tritium.

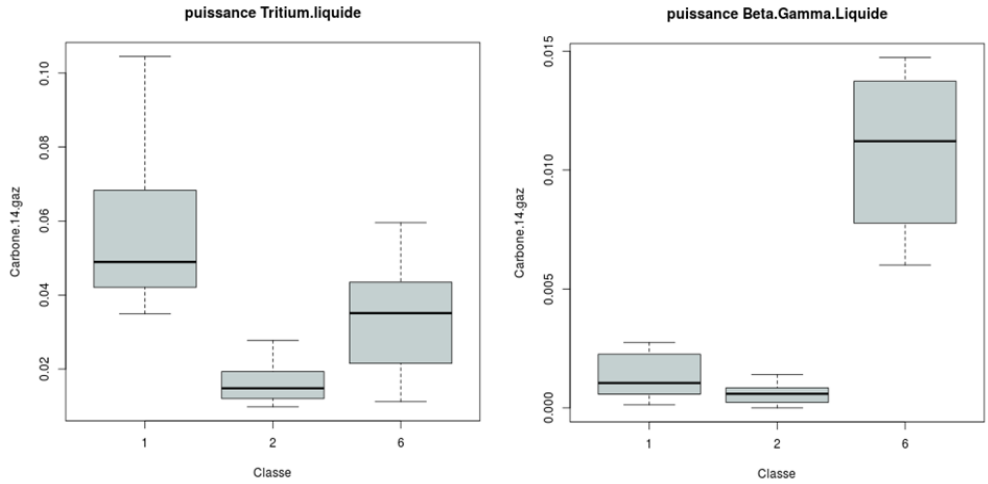


Figure 44 : Boîtes à moustaches pour les émissions liquides bêta-gamma et de tritium entre les 3 classes représentatives obtenues pour les INB internationales.

Étant donné que la classe 2 est peuplée de plus de la moitié de la population, dont notamment l'ensemble des sites français, une seconde classification a été effectuée uniquement sur la population de cette classe.

La classification sur la classe n°2, contenant les installations françaises, a permis de décomposer les 44 installations la peuplant en 3 classes (Figure 45). Toutes les installations françaises (19 sites) se trouvent dans la classe n°1 contenant 30 sites, à l'exception de FR01 qui est dans la classe 2 en raison d'émission gazeuses de tritium doubles des autres installations françaises. Cette seconde classe regroupe 8 sites.

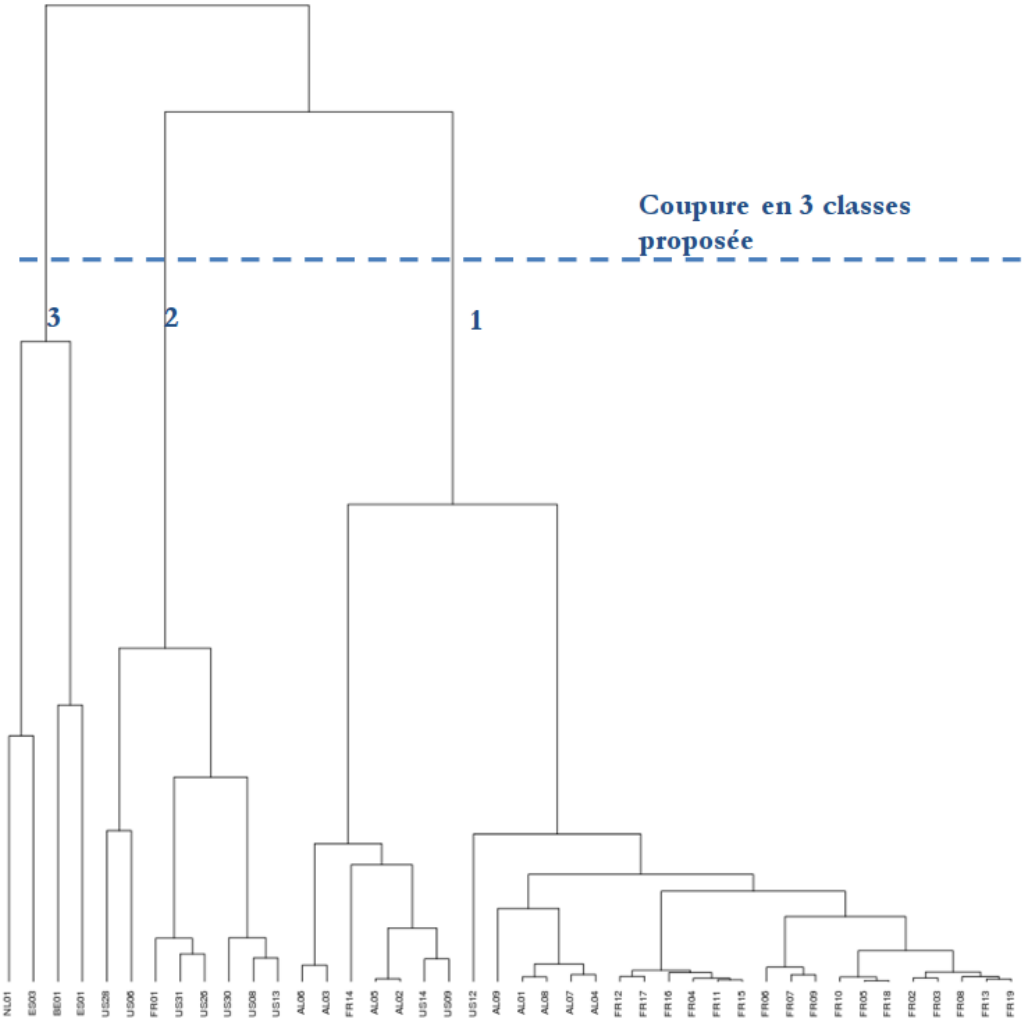


Figure 45 : Dendrogramme de la seconde classification sur les INB internationales de la classe 2.

L'analyse des classes, visualisable sur la Figure 46, montre qu'elles sont moins affirmées qu'avec la précédente, avec des dispersions et des niveaux d'émissions différents d'une variable à l'autre. Par ailleurs, les tests d'hypothèse sur cette seconde classification ne permettent pas d'établir de lien statistique avec les variables descriptives.

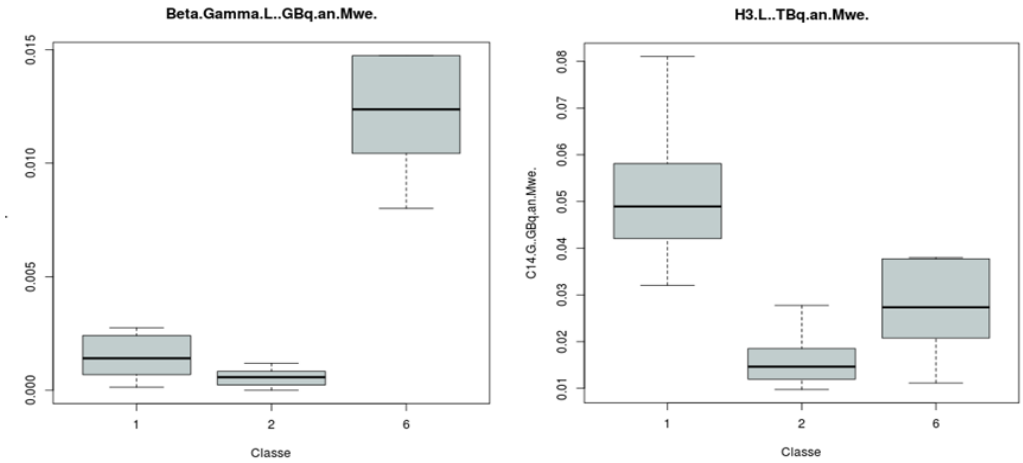


Figure 46 : Boîtes à moustaches pour les émissions de tritium liquide, de gaz rares et de rayonnements bêta-gamma entre les 3 classes de la seconde classification des INB internationales.

Finalement, la liste classée des installations pour cette seconde classification est résumée dans le Tableau 55 et constitue le livrable 3.2. Les installations françaises se trouvent toutes dans les classes 1 et 2.

Tableau 55 : Livrable 3.2 – Extrait de la liste des INB internationales représentatives dans chaque classe.

Installations représentatives			
Classe 1		Classe 2	Classe 3
AL01	FR08	FR01	BE01
AL02	FR09	US06	ES01
AL03	FR10	US08	ES03
AL04	FR11	US26	NL01
AL05	FR12	US28	
AL06	FR13	US30	
AL07	FR14	US31	
AL08	FR15	US13	
AL09	FR16		
FR02	FR17		
FR03	FR18		
FR04	FR19		
FR05	US09		
FR06	US12		
FR07	US14		

2.3.2. *Approche performante et caractérisation des classes*

La classification par l'approche performante effectuée sur les sept variables d'analyse conservées à l'issue de l'étape 2 a fait apparaître 50 installations sur le front de Pareto sur les 76 (Tableau 57), avec toutes les installations françaises sur le front de Pareto à l'exception de FR09, qui, malgré des émissions proches des autres installations françaises n'est dominant au sens de Pareto sur aucune des variables d'analyse. La recherche des dépendances entre les variables descriptives et la présence ou non des installations sur le front de Pareto apparaît comme étant liée au pays et au nombre de tranches (Tableau 56).

Tableau 56 : Livrable 3.1 – Dépendances entre classification performante et variables descriptives pour les INB internationales.

Variables	Pays	Année de mise en service	Nombre de tranches
Classification performante	+		+

Le livrable de cette étape est constitué de la liste des installations performantes (Tableau 57). Étant donné que 50 installations sont sur le front de Pareto, la liste est présentée en ordre inversé, en présentant les sites non optimaux, c'est-à-dire qui ne sont pas sur le front de Pareto.

Tableau 57 : Livrable 3.2 – Liste des installations performantes et non optimales des INB internationales.

Installations performantes	Installations non optimales	
Tous les autres sites	BE01	US11
	BE02	US15
	ES02	US18
	ES04	US19
	FR14	US20
	NL01	US22
	SU01	US25
	UK01	US27
	US01	US29
	US03	US34
	US04	US35
	US07	US36
	US09	US37

2.4. Étape 4 : Analyse des classifications

Cette étape consiste à analyser les résultats en vue d'établir une liste hiérarchisée d'installations représentatives ou performantes. Seule la classe n°1 contenant la plupart des installations françaises est reportée ici.

2.4.1. *Sélection représentative*

La sélection des sites représentatifs est issue du positionnement graphique des installations classées par l'approche représentative dans un couple d'axes distance inter-classe et distance intra-classe du positionnement des sites par rapport aux centres de chaque classe. Le nuage de points obtenu constitue un outil d'aide à la décision permettant de sélectionner les installations représentatives pour la classe 1 (Figure 47). Dans cet exemple, un objectif d'approcher un nombre d'installations de référence de 10 % de la population de chaque classe a été fixé (valeur définie par l'utilisateur et subjective). Par conséquent, la classe contenant 44 installations, l'objectif est de sélectionner les 5 sites les plus pertinents. Le livrable de cette étape correspond au Tableau 58 ; pour chaque classe, les codes des installations, leurs distances intra – et inter-classes, ainsi que la zone correspondante à la Figure 47 sont indiqués.

Le livrable de l'étape, dont un extrait est présenté dans le Tableau 58, est une liste ordonnée des installations par ordre de représentativité avec indication des distances intra – et inter-classes et de la zone de sélection des sites.

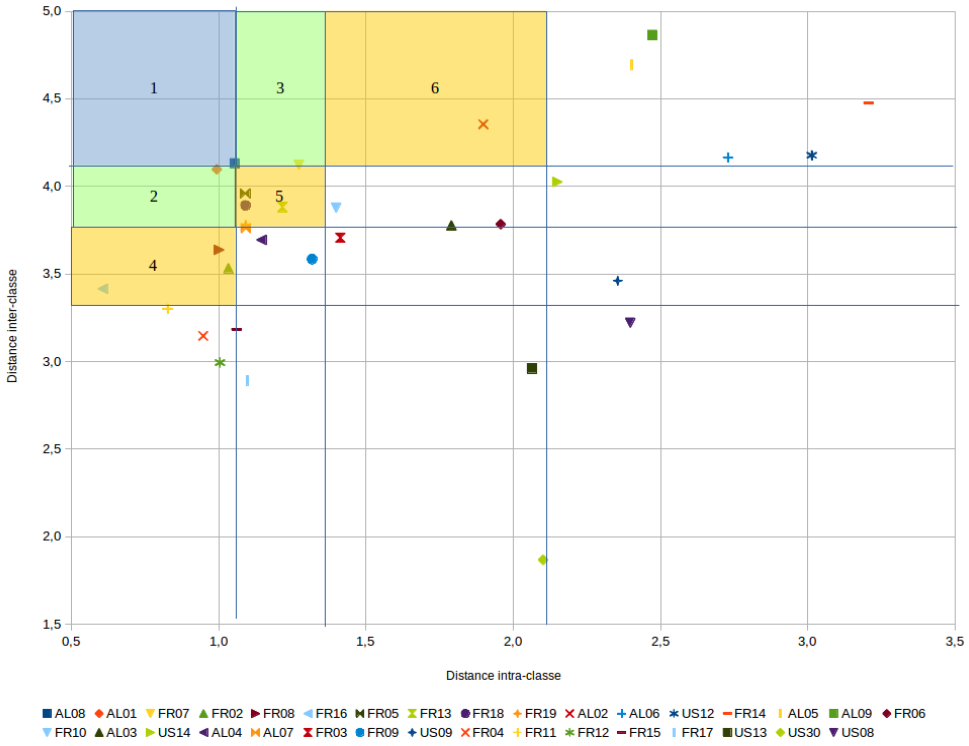


Figure 47 : Outil d'aide à la sélection des INB internationales de référence par couplage des distances intra-classes et inter-classes (classe 1).

Tableau 58 : Livrable 4 – Extrait de la liste des INB internationales représentatives présentation des sites de la classe 1.

#	Installations représentatives			
	Classe 1			
	Installation	Distance intra-classe	Distance inter-classe	Zone de sélection
1	**AL08**	**1,05**	**4,13**	**1**
2	**AL01**	**0,99**	**4,09**	**2**
3	**FR07**	**1,27**	**4,12**	**3**
4	**FR02**	**1,03**	**3,53**	**4**
5	FR08	**1,00**	**3,63**	4
6	FR16	**0,60**	**3,41**	4
7	**FR05**	1,08	3,95	**5**
8	FR13	1,21	3,88	5
9	FR18	1,09	3,89	5
10	FR19	1,09	3,77	5
11	**AL02**	1,89	4,35	**6**
…	…	…	…	…

Les experts interviennent ensuite pour la sélection finale des installations les plus pertinentes selon l'approche représentative avec, en complément, des installations dont les caractéristiques sont jugées indispensables. La sélection finale est constituée des quatre premières installations du Tableau 58 ; soit les sites AL08, AL01, FR07 et FR02. Elle est complétée de deux sites par les experts : FR05 et AL02.

2.4.2. Sélection performante

Pour être en cohérence avec l'approche représentative, un nombre d'installations de l'ordre de 10 % de la population totale, soit 8 sites ont été sélectionnés via le calcul de l'INPE (Tableau 59). En effet, étant donné le nombre d'installation sur le Front de Pareto (50) il n'est pas envisageable de les garder tous. Ainsi l'INPE a été calculé et la sélection se fera en fonction de l'évolution décroissante de l'indicateur (plus l'indicateur est fort, meilleur est le site).

Tableau 59 : Liste des INB internationales performantes retenues et valeur de leur INPE.

#	Installations performantes	
	Installations	INPE
1	US12	725
2	AL09	699
3	FR05	695
4	FR03	684
5	FR19	675
6	FR13	669
7	FR10	663
8	FR18	663

2.5. Étape 5 : Interprétation des résultats.

Pour rappel, cette étape a pour objectif de proposer des valeurs de référence et d'identifier des techniques permettant d'atteindre ces valeurs et pouvant constituer les MTD du secteur. Dans ce mémoire, seule la classe n°1, contenant la plupart des installations françaises, et l'approche performantes sont présentées. Étant donné qu'aucune information n'était disponible sur les techniques pour les installations autres que françaises, seule l'identification des valeurs de performances pertinentes a pu être effectuée dans cette étape.

2.5.1. Interprétation de la sélection représentative

Dans les figures qui suivent, le nom de sites de référence contient une « * » et les sites français sont hachurés.

Pour la classe 1, les taux d'effort sont tous supérieur à 50 % pour la variable donnée en exemple (émissions liquides de tritium). Les installations françaises sont réparties parmi les installations internationales sans constituer de groupe apparent qui permettrait de les distinguer autrement que comme membre à part entière de cette classe (Figure 48).

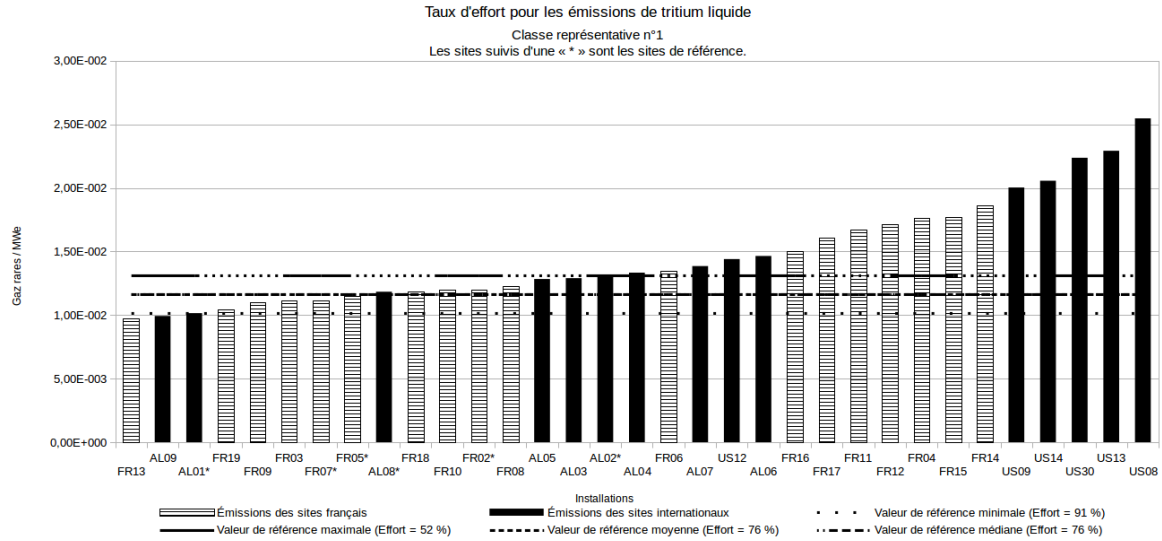


Figure 48 : Représentation du taux d'effort pour la classe représentative n°1 des INB internationales.

2.5.2. *Interprétation de la sélection performante*

Dans le cas de l'approche performante, les taux d'effort sont présentés dans la Figure 49.

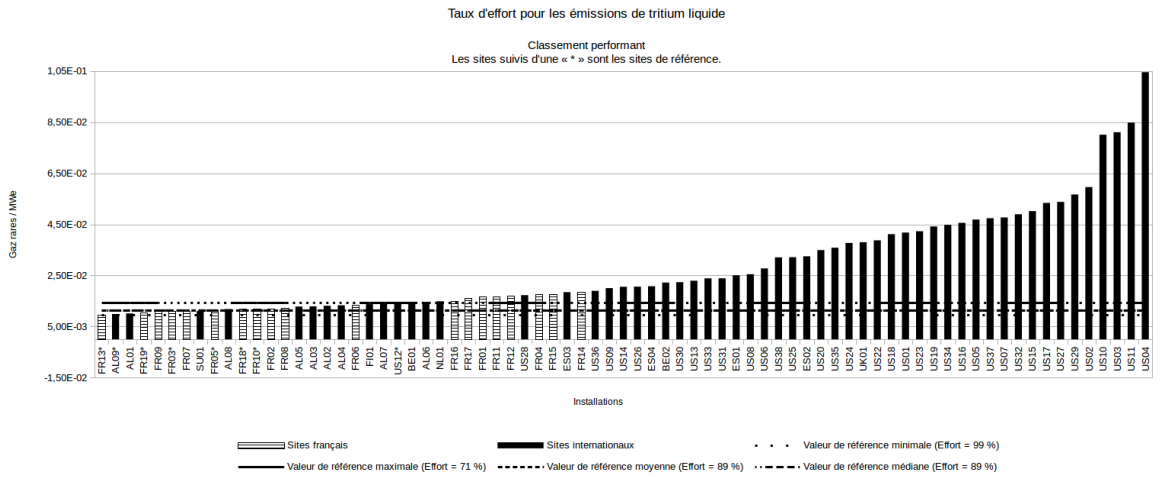


Figure 49 : Taux d'effort par rapport aux valeurs pertinentes pour l'approche performante des INB internationales.

2.5.3. *Livrables de l'étape*

Le livrable de l'étape résume les valeurs pertinentes pour chaque approche et les taux d'effort associés pour chaque variable d'analyse. Seule la classe 1 contenant la plupart des installations

françaises et les installations situées sur le front de Pareto ont été étudiées. Dans le Tableau 60, seules les émissions de tritium liquide sont présentées avec les taux d'effort internationaux et appliqués aux seuls sites français. Pour chacune des valeurs de référence, les taux d'effort sont supérieurs dans l'approche performante à celles de l'approche représentative. De plus, les taux d'effort pour les sites français sont plus faibles que pour l'ensemble des sites internationaux, ce qui montre qu'ils sont positionnés parmi les sites les moins émetteurs.

Tableau 60 : Livrable 5 – Valeurs de référence et taux d'effort pour les émissions de tritium liquide pour les INB internationales.

Taux d'effort sur les émissions de tritium liquide												
Classement	Minimum			Maximum			Moyenne			Médiane		
	International		National	International		National	International		National	International		National
	Valeur (TBq/MWe)	Effort	Effort	Valeur (TBq/MWe)	Effort	Effort	Valeur (TBq/MWe)	Effort	Effort	Valeur (TBq/MWe)	Effort	Effort
Classe 1	1,02E-02	91 %	89 %	1,50E-02	33 %	32 %	1,21E-02	64 %	47 %	1,19E-02	70 %	58 %
Performants	9,75E-03	99 %	95 %	1,44E-02	71 %	42 %	1,14E-02	89 %	74 %	1,13E-02	89 %	74 %

3. Conclusions de l'application aux INB

Des applications à deux échelles pour les Installations Nucléaires de Base ont été effectuées : la première concernant essentiellement les installations françaises et la deuxième concernant un panel d'installations internationales. Ils ont de plus permis de positionner les décisions des experts et de valider des suggestions faites par la méthodologie à chaque étape.

La première application à l'échelle nationale française a montré que la méthodologie est applicable dans sa complétude, sous réserve de limiter le nombre de variables d'analyse. La méthodologie peut aider à la réduction du nombre de variables d'analyse dans le but de permettre la classification. Au cours de cette étape, les échanges avec les experts ont montré que l'approche multicritère employée était suffisamment robuste pour éviter que les choix des experts n'influent complètement sur la classification d'après des tendances monocritères. Malgré le faible nombre d'installations, il a été possible d'identifier des corrélations entre les performances environnementales et les caractéristiques des installations, notamment la localisation et le nombre de tranches.

La seconde application à l'échelle internationale a présenté un plus grand nombre d'installations disponibles. Cette étude a permis de mettre en exergue que, pour l'analyse représentative, les installations françaises se positionnent dans une même classe dont la dispersion et les niveaux d'émission de radionucléides sont parmi les plus faibles. De plus, les installations françaises sont apparues sur le front de Pareto, ce qui tend à montrer que le contexte français soit

spécifique et performant. Afin de conclure plus précisément sur ces points, des données supplémentaires seraient nécessaires.

La méthodologie a ainsi fourni des éléments d'aide à la décision qui ont pu guider les choix des experts. Cependant, le regard de l'expert est important pour que les choix soient en accord avec la réalité du terrain.

Partie 4 : Discussions et perspectives

Grâce à plusieurs outils statistiques et des avis d'experts, la méthodologie proposée permet de traiter des données issues d'installations industrielles en activité (Figure 50). Ce traitement, décomposé en cinq étapes fournit à ces mêmes experts des outils d'aide à la décision et des livrables les aidant à prendre des décisions dans l'identification d'installations de référence, la définition de valeurs de référence en considérant des taux d'effort et l'identification des MTD. De plus, les résultats obtenus vont permettre de soulever des questions sur les données utilisées et affiner les besoins en la matière.

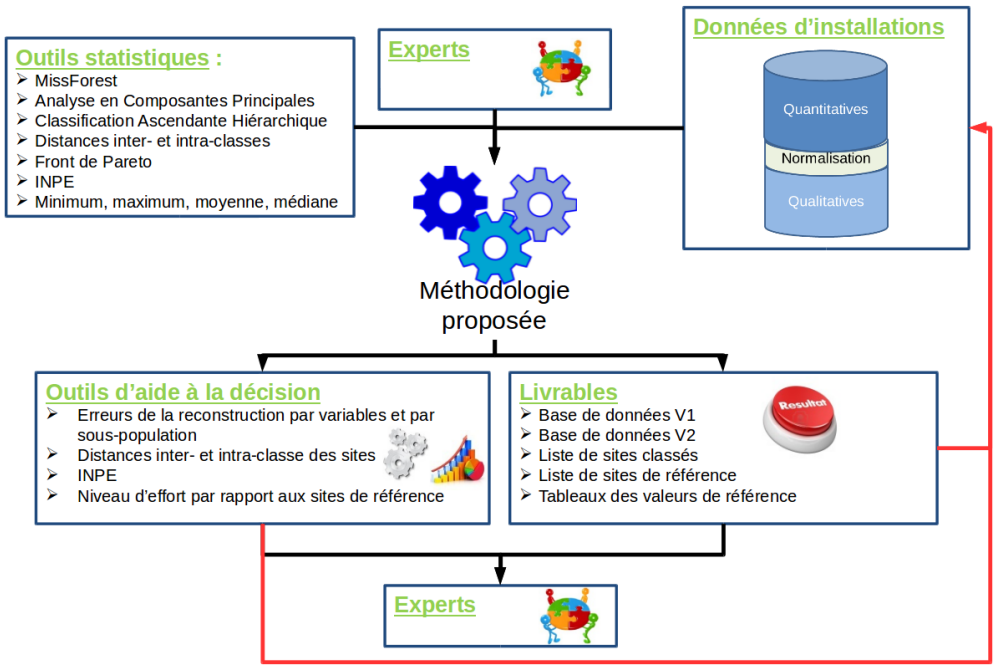


Figure 50 : Synthèse du fonctionnement de la méthodologie.

Cette partie vise à confronter la méthodologie proposée aux hypothèses et postulats sur lesquels a reposé ce travail de thèse (Tableau 61). En outre, les apports et limites de la méthodologie sont discutés dans ce chapitre et les réponses possibles à ces limites sont mises en perspective.

Tableau 61 : Question, hypothèses, postulats et caractéristiques de la méthodologie dans le cadre de cette thèse.

Problématique générale
Le seul cadre formel existant pour la détermination des MTD, le Processus de Séville, est sujet à des différences d'interprétations qui créent des problèmes de justification de décisions prises à chacune de ses étapes.
Question générale
Quelle méthodologie permettrait d'améliorer le processus de la détermination des MTD à partir de la collecte et de l'analyse de données relatives à des installations industrielles et applicable à différentes échelles géographiques ?
Hypothèse 1
Une approche intégrée et englobant les trois échelles géographiques identifiées permettrait d'éviter les conflits entre les différentes échelles.
Hypothèse 2
Une méthodologie structurée, statistique et multicritère permet de contribuer à l'atteinte des objectifs de l'IED en termes de performance globale.
Postulats
Si un site est remonté pour son analyse lors du Processus de Séville, c'est qu'il est viable économiquement, techniquement et qu'il a un faible impact environnemental dans son ensemble. Il est donc conforme à l'IED.
La caractérisation et la sélection des sites doivent être guidées par le niveau de performance environnementale dans son ensemble.
Principales caractéristiques de la méthodologie
1. Partagée afin d'uniformiser le processus de sélection, 2. Transparente et permettant de tracer la décision (traçabilité), 3. Adaptable à des processus décisionnels variés, 4. Robuste, c'est-à-dire permettant de limiter l'incertitude, 5. Réplicable à tout secteur industriel, 6. Objective au sens où la méthode ne fait pas intervenir d'*a priori*, 7. Intégrée à l'ensemble des étapes du processus de décision.

1. Confrontation des hypothèses avec les résultats

L'état de l'art (Partie 1) a permis de formuler deux hypothèses et des caractéristiques de la méthodologie auxquelles les cas d'application (Partie 3) ont visé à répondre.

1.1. Intégration des trois échelles géographiques

Les cas d'application ont démontré que la méthodologie était applicable au niveau international pour classifier et comparer des installations provenant de différents pays en fonction de leurs consommations et émissions. De même, l'application nationale a permis de démontrer la faisabilité de la démarche en se basant uniquement sur le parc industriel français (laitier ou nucléaire). Par conséquent, la méthodologie semble bien applicable à ces deux échelles ; l'origine des données étant la seule différence.

Cependant, au niveau local, son applicabilité n'a pas pu être démontrée. Toutefois, il semble que le positionnement d'une installation particulière par rapport aux résultats de l'application de la méthodologie aux échelles nationale ou internationale soit possible. En effet, la comparaison de la performance des installations aux valeurs de référence et MTD est possible dans la mesure où le site appartient à une des classes proposées si l'approche représentative est retenue ou aux résultats obtenus par l'approche performante. Au cas où le site en question ne s'intégrerait avec aucune des classes identifiées, alors son positionnement en tant qu'exception doit être discuté. Cela peut éventuellement se faire en exécutant de nouveau une classification incluant cette installation pour confirmer ou infirmer son caractère unique ou modifier les classes identifiées. La vérification de la cohérence multi-échelle nécessiterait d'analyser chacune d'elles selon les mêmes critères.

Validation partielle de l'hypothèse 1 : « Une approche intégrée et englobant différentes échelles géographiques identifiées permettrait d'assurer la cohérence entre les différentes échelles. »

1.2. Approche structurée, statistique et multicritère

La méthodologie proposée est structurée en cinq étapes qui mettent en jeu des outils d'analyse statistique. Elle permet de proposer des niveaux de consommations et d'émissions exprimés par des valeurs de référence issues des installations globalement performantes ou représentatives, en fonction d'un niveau d'effort acceptable pas les experts impliqués dans le processus décisionnel existant. À la différence des pratiques actuelles (voir partie bibliographique), ces résultats ont été établis en considérant un ensemble de variables simultanément. Ainsi, même en utilisant les valeurs de référence maximales, le niveau d'effort n'est pas nul pour toutes les variables d'analyse et

permettrait donc d'aboutir à une amélioration continue des émissions et consommations. Ce mode d'analyse entraîne néanmoins des difficultés de perception par les experts qui ne retrouvent pas forcément les caractéristiques qu'ils ont l'habitude d'appréhender. De plus, la complémentarité des approches représentatives et performantes permet d'identifier des performances différenciées au sein d'une population donnée. Ainsi, la méthodologie permet d'établir des valeurs de référence susceptibles de contribuer au principe d'amélioration continue de l'IED et d'intégrer l'environnement dans son ensemble. Un outil d'aide au choix des techniques fournit un éclairage sur les celles qui sont appliquées dans les sites atteignant ces valeurs de référence et susceptibles d'être des candidates aux MTD étayées par les données.

Validation de l'hypothèse 2 : « Une méthodologie structurée, statistique et multicritère permet de contribuer à l'atteinte des objectifs de l'IED en termes de performance globale. »

1.3. Caractéristiques de la méthodologie

En reprenant les caractéristiques formulées dans le Tableau 61, la méthodologie proposée est disponible et conçue pour être **partagée** entre les acteurs dans la mesure où sa structure en étapes et ses outils mathématiques accessibles permettent son application par n'importe quel acteur disposant des données d'entrée.

Elle est également **transparente** dans la mesure où chaque livrable et décision est explicite et reproductible et où toutes les hypothèses et périmètres sont présentés et connus.

De plus, elle est **robuste** dans la mesure où les mêmes entrées et choix d'experts aboutissent aux mêmes résultats. L'influence du choix du périmètre et des variables d'étude a été étudiée sur les installations laitières sur 32 combinaisons de populations, de variables d'analyse et de normalisation. Les résultats montrent qu'en modifiant un paramètre à la fois, l'approche représentative est la plus sensible aux variations, alors que l'approche performante est très peu impactée par la variation des paramètres. La conclusion de cette analyse de sensibilité est que les facteurs les plus « influents » sont, dans l'ordre, la variable de normalisation, les variables d'analyse et enfin la population. Par ailleurs, le seul élément stochastique introduit dans la méthodologie, qui peut entraîner des différences de classification, est la reconstruction des données manquantes.

Bien que les experts puissent agir de façon ponctuelle sur des orientations (par exemple, le choix des variables d'analyse, du nombre de sites de référence, ou des valeurs de référence, etc.), la méthode est **objective**, car elle reste fondée sur l'analyse des données qui ne met pas en jeu d'*a priori* quant aux facteurs causant les consommations et émissions.

Les cas d'application ont montré qu'elle était **adaptable** à deux secteurs industriels différents et qu'elle ne dépend pas du secteur d'application. En cela, elle semble donc réplicable à tout secteur industriel, en contexte IED ou non, et pour une population d'une vingtaine à une centaine d'installations.

Enfin, elle est **intégrable** à l'ensemble des étapes du processus de décision. Dans le cas du Processus de Séville, elle a été appliquée *a priori* aux échelles nationales et internationales. Elle devrait également pouvoir l'être *a posteriori* pour vérifier des décisions prises par les experts ; en particulier sur la pertinence des BATAEL.

2. Discussions autour de la méthodologie proposée

2.1. Quantité et qualité des données

La principale limite à une approche multicritère est la disponibilité des données. La méthodologie proposée utilise donc les données de consommations et d'émissions qui sont des informations industrielles disponibles. L'ajout d'autres données quantitatives économiques, techniques ou environnementales serait possible en l'état. Néanmoins, si l'on cherchait à affecter une pondération entre critères ou à inclure des variables qualitatives pour la classification, il serait nécessaire de modifier les outils mathématiques utilisés (par exemple : pour l'approche représentative, l'analyse factorielle de données mixtes). En revanche, ne prendre que les données complètes réduit la représentativité de l'étude et la prise en compte de l'environnement au global. La reconstruction pourrait être améliorée avec d'autres algorithmes pour choisir le plus adapté aux données. De même, la procédure de décision pourrait inclure une phase de comparaison de différentes solutions de reconstruction des données. Il est de plus à signaler qu'en l'absence d'informations pour une variable et un site donné, un doute subsiste pour la reconstruction. Notamment, lorsqu'une donnée est vide, cela peut signifier soit que le site n'est pas concerné (dans ce cas la mention NA devrait apparaître) soit que le site n'a pas renseigné la valeur (o ou non connue). Dans le premier cas, la reconstruction ne doit pas s'appliquer, contrairement au deuxième cas. Une pré-lecture de la base de données initiale par un expert est donc indispensable pour orienter les « données/cases vides » dans l'une ou l'autre des alternatives (NA ou non renseigné) pour voir si la reconstruction est pertinente.

Les émissions et consommations étant souvent liées entre elles, une analyse de ces corrélations multiples pourrait faire l'objet de futures recherches. À l'heure actuelle, seul un arbitrage humain peut décider des paramètres utilisables ou pouvant être exclus de la méthodologie. La subjectivité est alors modérée par la transparence des processus décisionnels et leur fonctionnement collégial.

Par ailleurs, il semble qu'il y ait un lien entre le choix des variables d'analyse et de la variable de normalisation. En effet, dans le cas où une des variables d'analyse n'est pas corrélée avec la variable de normalisation (par exemple, la consommation d'énergie et le volume d'effluents), sa pertinence dans l'étude devrait être analysée plus en détail pour déterminer de possibles limites dans l'application de la méthodologie et estimer l'impact de différentes variables de normalisation et d'analyse. Dans tous les cas, des applications de la méthodologie à différentes échelles ne sont comparables, en termes d'installations de référence et de valeurs de référence, qu'avec une même normalisation. Par exemple, dans le cas de l'industrie laitière, la normalisation par le tonnage de matière laitière entrante va « pénaliser » les installations utilisant principalement des produits secs, alors que la normalisation par tonnage de produits sortants classera comme moins performants les installations productrices de produits secs. Par conséquent, la comparaison, à différentes échelles, d'une classification basée sur les variables normalisées différemment est impossible.

2.2. Outils d'aide à la décision

2.2.1. Outil d'aide à la sélection des installations représentatives

La sélection des installations de référence par l'approche représentative s'effectue en fonction des distances intra – et inter – classes. Le découpage en zones par les différents quartiles vise à suggérer un ordre de sélection qui ne correspond pas à des sous-classes. En effet, un site se trouvant dans les zones à la numérotation la plus élevée peut tout aussi bien avoir des valeurs d'émissions élevées ou basses par rapport au site le plus représentatif sur une ou plusieurs des variables d'analyse.

Par conséquent, sélectionner des sites les plus éloignés n'est pas un gage de choix des moins contraignants et montre donc que la représentativité est ici objective. L'ajout de sites plus éloignés dans les zones par les experts peut donc être plus contraignant en termes de valeurs de consommations et d'émissions mais permettre de mettre en avant une caractéristique particulière que seuls les experts peuvent détecter.

2.2.2. Outil d'aide à la sélection des installations performantes

Le front de Pareto est une approche par compromis qui permet d'identifier les installations optimales. Cependant, comme avec l'ACP, elle requiert de disposer d'un plus grand nombre d'installations que de variables d'analyse. En effet, dans le cas où il y a autant ou plus de variables que d'installations, le front est alors constitué des sites dont une des variables est le minimum de l'échantillon, faisant perdre le bénéfice d'une approche multicritère. L'ajout de critères de sélection supplémentaires tels que l'INPE est une solution possible. Cependant, bien que l'INPE permette

de discriminer des installations optimales qui ne peuvent *a priori* pas l'être, l'agrégation de plusieurs indicateurs en un score unique présente des risques de pertes d'information et de compensation commune à toutes les approches de ce type.

Une solution pour répondre aux limites de l'INPE (et du front de Pareto) pourrait être de suivre une approche type programmation linéaire en incluant aux objectifs de minimisation de consommations et d'émissions des contraintes de valeurs maximales à ne pas dépasser. Ainsi, une installation se trouvant sur le front de Pareto et avec une bonne valeur d'INPE pourrait être écartée, si elle avait une valeur plus élevée qu'une valeur limite réglementaire existante sur une des variables.

2.3. Détermination de MTD candidates

Établir le lien entre techniques et émissions est l'élément qui demande le plus d'informations et devient complexe dans le cas de combinaisons de techniques influençant plusieurs variables. Dans ce cas, les experts sont encore plus importants que dans les autres étapes. De plus, la méthodologie semble plus pertinente pour justifier qu'une technique donnée se trouve bien dans les sites les plus performants que pour déduire les techniques candidates MTD à partir des installations. En fait, il est difficile d'associer une performance d'une installation au regard d'une valeur de référence sur une seule technologie sachant que cette dernière peut être couplée à d'autres ou mise en œuvre de façon particulière. La démarche proposée dans cette thèse peut seulement faire ressortir la fréquence d'apparition d'une technique dans les installations par rapport à une valeur de référence et donc guider le décideur vers de potentielles techniques candidates aux MTD. Pour aller plus loin dans l'identification des techniques, des données sur les contributions de chacune d'elles sur les différentes variables d'analyse et un outil multidimensionnel seraient nécessaires.

2.4. Intégration de différents niveaux d'analyse

Les objectifs d'application pouvant différer en fonction du contexte, il serait intéressant de proposer différents niveaux d'analyse avec, dans un premier temps, une intégration *a priori* dans le cadre de la recherche de MTD et une intégration *a posteriori* permettant de vérifier des décisions prises précédemment uniquement sur jugement d'experts. Dans un second temps, il est envisageable de proposer à chaque étape de la méthodologie une pluralité d'outils adaptés à différentes applications en fonction de la taille de la population, de la nature des données ou du niveau de détail recherché.

Ainsi, pour les étapes 1 et 2 de la méthodologie, le traitement des données pourrait se faire en fonction d'un intervalle de confiance pour les différentes variables qui s'ajouterait au choix statistique des variables selon plusieurs approches possibles : tests de dépendances et indépendances, calcul des erreurs de reconstruction des données manquantes, recherche des variables contribuant le plus à la construction des axes factoriels d'une analyse en composantes principales.

Pour la troisième étape de classification des installations, des variables qualitatives (technologie, contexte local, pratiques managériales, présence ou non de certains éléments, typologie de déchets, etc.) pourraient être incluses en utilisant des outils d'analyse factorielle de données mixtes. De plus, les variables d'analyse pourraient être pondérées en fonction, par exemple, de leur présence ou non dans les KEI (exemple des substances dangereuses prioritaires de la directive cadre sur l'eau).

2.5. Impacts (socio-économiques) de l'approche pour les acteurs

Une application au niveau national du Processus de Séville, via les groupes miroirs permettrait de structurer et harmoniser les approches entre différentes activités concernées par un même BREF. En outre, cela permettrait de faciliter la collecte de données en se concentrant directement sur les données permettant d'aboutir aux objectifs fixés en termes d'installations de référence. Les différentes hypothèses des industriels pourraient également être vérifiées au regard des données disponibles, fournissant ainsi les arguments nécessaires à la défense des spécificités nationales.

L'application de la méthodologie au niveau européen est plus hypothétique et se trouve dans un horizon temporel plus lointain en raison du nombre de parties à convaincre pour une utilisation commune. Néanmoins, de même qu'au niveau des groupes miroirs, son utilisation dans ce cadre formel permettrait d'améliorer la cohérence et la qualité des données collectées et traitées et de disposer d'une analyse multicritère permettant de classifier les installations en fonction de différences de performances environnementales et non en fonction d'avis d'experts qui ne sont pas toujours justifiables à partir des données. La perspective d'inclure dans les variables d'analyse des critères économiques et techniques permettraient de rattacher explicitement la définition des MTD données dans l'IED avec les techniques apparaissant comme possibles MTD.

Enfin, hors du cadre de l'IED, l'application de la méthodologie à un parc d'installations constituerait le cadre qui manque à l'heure actuelle pour justifier qu'une installation ait un niveau de performance environnemental équivalent aux MTD. De même, l'ajout de critères économiques fournirait des éléments utiles à la justification technico-économique de l'application ou la non-application d'une technique donnée.

3. Perspectives du travail de recherche

3.1. Mise en œuvre dans le Processus de Séville

Bien que le développement de la méthodologie ait été en parallèle de la révision du BREF agroalimentaire, son application ne s'est pas déroulée en condition réelle. Il serait donc pertinent de mettre en œuvre la méthodologie avec une application en temps réel lors de la révision d'un nouveau BREF au travers d'une démarche *a priori* correspondant aux applications de la partie 3.

D'autres applications, sur d'autres secteurs industriels, tel que le BREF sur les panneaux de bois (*Wood-Based Panels*) a été révisé récemment, seraient envisageables dans le but de dérouler la méthode *a posteriori*, c'est-à-dire afin de vérifier les décisions par l'application de la méthodologie. En particulier, il serait intéressant de confronter les typologies d'installations et les BATAEL obtenues au niveau de Séville avec les classifications et valeurs de référence issues de la méthodologie et donc en collaboration avec l'ensemble des acteurs du processus de Séville.

3.2. Applications à d'autres secteurs industriels

Afin de mieux identifier les limites d'applicabilité de la méthodologie, tel qu'une mise en œuvre dans d'autres secteurs industriels, qu'ils soient IED ou non, serait à envisager. En particulier, une application sur un secteur comprenant plusieurs centaines d'installations permettrait d'éprouver sa validité et d'approfondir la phase de recherche de liens entre techniques et performances environnementales.

3.3. Utilisation d'autres outils statistiques

Les outils statistiques présentés dans ce mémoire sont les plus pertinents compte tenu des applications qui ont été réalisées. D'autres outils pourraient être utilisés dans des contextes différents. Ainsi, dans le cas où un grand nombre d'installations seraient disponibles, les techniques de krigeage pour la reconstruction des données pourraient être plus efficaces que MissForest. En ce qui concerne les classifications des installations, des approches telles que le *Data Envelopment Analysis* semblent pertinentes pour la classification représentative et d'autres méthodes d'optimisation multi-objectif que le front de Pareto existent et pourraient être étudiées.

Ces approches supplémentaires permettraient d'étoffer la méthodologie afin de pouvoir proposer aux experts la ou les approches les plus pertinentes compte tenu des données dont ils disposent.

3.4. Application à l'échelle locale

Comme dit précédemment, l'application de la méthodologie à l'échelle locale n'a pas pu être testée. Y parvenir nécessiterait de lui ajouter les moyens de considérer des éléments de contexte local afin de croiser les données sur les MTD et BATAEL avec ses techniques et émissions. Les recherches sur ce sujet semblent requérir de disposer de données sectorielles afin d'appliquer la méthodologie puis de positionner une installation donnée par rapport aux classes identifiées par la méthodologie.

Conclusion

La méthodologie proposée a permis de répondre à la question de recherche et aux hypothèses formulées bien que certains éléments restent à éclaircir. Ainsi la première hypothèse a été validée aux échelles nationale et internationale (Tableau 62). Une ou plusieurs études de cas supplémentaires permettraient de démontrer l'applicabilité de la méthodologie à l'échelle locale. Bien qu'elle semble possible, les adaptations nécessaires n'ont pas été envisagées dans le cadre de cette thèse, de même que la manière de prendre en considération certaines caractéristiques locales autrement que par jugement d'experts. Une de ces applications supplémentaires pourrait inclure un déroulé complet de la méthodologie au cours du Processus de Séville afin de comparer ses résultats avec les décisions prises par les experts et de valider complètement l'apport de la méthodologie aux objectifs de l'IED. Cela constituerait une double application *a priori* et *a posteriori*.

Tableau 62 : État de validation, limite et perspectives de la réponse de la méthodologie aux hypothèses et postulats.

	État de la validation	Limite	Perspectives
Hypothèse 1 : Une approche intégrée et englobant les trois échelles géographiques identifiées permettrait d'éviter les conflits entre les différentes échelles.	Partielle	L'application locale n'a pas été effectuée.	Extension de la méthodologie à une installation particulière.
Hypothèse 2 : Une méthodologie structurée, statistique et multicritère permet de contribuer à l'atteinte des objectifs de l'IED en termes de performance globale.	Validée	L'application complète sur des données du Processus de Séville n'a pas été réalisée.	Application au cours d'une révision de BREF ou sur les données d'un BREF révisé.

Enfin, la méthodologie et en particulier ses seconde et cinquième étapes apportent un éclairage sur la quantité et la qualité des données suffisantes pour remplir les objectifs de la directive et atteindre un niveau de détail suffisant.

Conclusion générale

Bien que l'objectif de l'IED soit la réduction des pollutions en Europe grâce à l'utilisation des MTD et demeure le fil conducteur des discussions lors du Processus de Séville, les effets de cette réglementation n'ont pas été mesurés à l'heure actuelle. L'équilibre des considérations environnementales vis-à-vis des critères techniques et économiques portés par les différents membres des TWG repose sur ce processus décisionnel collégial. Malgré, les critiques qui peuvent être émises sur le Processus de Séville, il permet de fournir aux industriels et aux autorités environnementales des documents de référence servant de base à la comparaison des techniques mises en œuvre dans les installations avec les MTD. Dans le cas des secteurs devant mettre en œuvre les MTD hors du périmètre de l'IED, ces documents de référence sont inexistants et les difficultés encore plus grandes.

Cette thèse a visé à proposer une méthodologie capable de s'intégrer dans les processus décisionnels existants et permettant d'expliciter et de guider les décisions au travers d'une analyse statistique multicritère de données industrielles disponibles. En contexte IED ou non, la méthodologie permet, aux échelles nationales et internationales, d'identifier des installations représentatives et performantes puis de calculer des valeurs de référence pour la détermination de MTD et de BATAEPL. Elle permet en outre, grâce au taux d'effort, de visualiser l'impact des valeurs de référence sur le parc existant.

Cette méthodologie en cinq étapes a été appliquée avec succès dans deux secteurs différents (IED et non-IED), à deux échelles géographiques différentes (nationale et internationale). Ces applications ont permis de valider les hypothèses quant à la pertinence d'une approche commune à ces deux échelles en vue de contribuer aux objectifs de l'IED et d'identifier des limites à plusieurs des étapes de la méthodologie. Les principales d'entre elles (Tableau 63) ouvrent plusieurs perspectives pour l'utilisation d'outils mathématiques supplémentaires. Ainsi, le manque de connaissances des liens qui peuvent exister entre les différentes variables pourrait être approfondi par une analyse plus approfondie des corrélations multiples entre variables. En ce qui concerne la deuxième étape de reconstruction des données, d'autres approches comme le krigeage pourraient être étudiées. Il en est de même pour l'étape de classification avec la *Data Envelopment Analysis* dont l'adoption modifierait également le mode de sélection des installations représentatives de l'étape 4. Enfin, l'outil d'aide à la sélection des techniques de l'étape 5 pourrait intégrer une approche multicritère de manière à prendre en compte plus précisément les liens complexes qui existent entre consommations, émissions et techniques.

Tableau 63: Limites et perspectives des étapes de la méthodologie.

	Limites	Perspectives
Étape 1 : Définition du périmètre et des variables d'étude	Connaissances des liens entre variables	Analyser les corrélations multiples
Étape 2 : Traitement des données	Reconstruction des données	Chercher d'autres approches et leurs applicabilités
Étape 3 : Classification des installations	Classification sur les données quantitatives	Analyses factorielles de données mixtes.
Étape 4 : Analyse de la classification		
Étape 5 : Interprétation de la sélection	Difficulté d'établir le lien entre techniques et émissions ou consommations	Recherches supplémentaires pour une approche multicritère.

Par ailleurs, la troisième échelle d'application (locale) n'a pas pu être traitée et devra faire l'objet d'autres recherches. Elle constitue l'une des perspectives de recherche qui, en plus de nouvelles applications et de l'intégration d'outils supplémentaires permettant des analyses de niveaux de détail différents, *a priori* et *a posteriori*. En particulier, une application à un secteur comportant plusieurs centaines d'installations est envisagée.

Enfin, l'apport de la méthodologie dans le cas d'une application à l'échelle européenne reste à démontrer sur un cas d'application.

Bibliographie

AIDA, 2016. 2230. Réception, stockage, traitement, transformation, etc. du lait [En ligne]. URL http://www.ineris.fr/aida/consultation_document/10565 (accédé le 07/06/2016).

Aloini, D., Dulmin, R., Mininno, V., 2014. A peer IF-TOPSIS based decision support system for packaging machine selection. Expert Syst. Appl. 41, 2157–2165.

Altman, N.S., 1992. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. Am. Stat. 46, 175–185.

ASN, 2013. Décision n°2013-DC-0360 de l'ASN du 16 juillet 2013.

Assemblée Nationale, 2013a. Décret n° 2013-374 du 2 mai 2013 portant transposition des dispositions générales et du chapitre II de la directive 2010/75/UE du Parlement européen et du Conseil du 24 novembre 2010 relative aux émissions industrielles (prévention et réduction intégrées de la pollution).

Assemblée Nationale, 2013b. Décret n° 2013-375 du 2 mai 2013 modifiant la nomenclature des installations classées.

Assemblée Nationale, 2013c. Ordonnance n° 2012-7 du 5 janvier 2012 portant transposition du chapitre II de la directive 2010/75/UE du Parlement européen et du Conseil du 24 novembre 2010 relative aux émissions industrielles (prévention et réduction intégrées de la pollution).

Assemblée Nationale, 2013d. Arrêté du 2 mai 2013 modifiant l'arrêté du 29 juin 2004 relatif au bilan de fonctionnement prévu à l'article R. 512-45 du code de l'environnement.

Assemblée Nationale, 2012a. Ordonnance n°2012-6 du 5 janvier 2012 modifiant les livres Ier et V du code de l'environnement.

Assemblée Nationale, 2012b. Arrêté du 7 février 2012 fixant les règles générales relatives aux installations nucléaires de base.

Assemblée Nationale, 2007. Décret n°2007-157 du 5 février 2007 relatif aux substances vénéneuses et modifiant le code de la santé publique (dispositions réglementaires).

Assemblée Nationale, 2006. Loi n°2006-686 du 13 juin 2006 relative à la transparence et à la sécurité en matière nucléaire.

Assemblée Nationale, 2004. Arrêté du 29 juin 2004 relatif au bilan de fonctionnement prévu par le décret n° 77-1133 du 21 septembre 1977 modifié.

Assemblée Nationale, 1998. Arrêté du 2 février 1998 relatif aux prélèvements et à la consommation d'eau ainsi qu'aux émissions de toute nature des installations classées pour la protection de l'environnement soumises à autorisation.

Assemblée Nationale, 1976. Loi n°76-663 du 19 juillet 1976 relative aux Installations Classées pour la Protection de l'environnement.

Bachmann, T.M., van der Kamp, J., 2014. Environmental cost-benefit analysis and the EU (European Union) Industrial Emissions Directive : Exploring the societal efficiency of a DeNOx retrofit at a coal-fired power plant. Energy 68, 125–139.

Barros, M.C., Magán, A., Valiño, S., Bello, P.M., Casares, J.J., Blanco, J.M., 2009. Identification of best available techniques in the seafood industry : a case study. J. Clean. Prod. 17, 391–399.

Bortolini, M., Gamberi, M., Graziani, A., Mora, C., Regattieri, A., 2013. Multi-parameter analysis for the technical and economic assessment of photovoltaic systems in the main European Union countries. Energy Convers. Manag. 74, 117–128.

Bréchet, T., Tulkens, H., 2009. Beyond BAT : Selecting optimal combinations of available techniques, with an example from the limestone industry. J. Environ. Manage. 90, 1790–1801.

Carretero, A.L., de la Rosa, J., Sanchez-Rodas, D., 2016. Applying statistical tools systematically to determine industrial emission levels associated with the best available techniques. J. Clean. Prod. 112, 4226–4236.

Cikankowitz, A., 2008. Méthodologie d'évaluation des performances environnementales de techniques en vue de les comparer puis de les valider "meilleures techniques disponibles" (Thèse de doctorat). École Nationale Supérieure des Mines, Saint-Etienne, France.

Cikankowitz, A., Laforest, V., 2013. Using BAT performance as an evaluation method of techniques. J. Clean. Prod. 42, 141–158.

Code de l'Environnement, 2011. Code de l'environnement – Article L511-1, Code de l'environnement.

Commission Européenne, 2015. Kick-off meeting for the review of the reference document on best available techniques for food, drink and milk industries. Séville.

Commission Européenne, 2013. JRC Reference Report on Monitoring of emissions from IED installations.

Commission Européenne, 2012. Décision d'exécution de la Commission du 10 février 2012 établissant les lignes directrices sur la collecte de données, sur l'élaboration de références MTD et sur leur assurance qualité, visées par la directive 2010/75/UE du Parlement européeen et du Conseil relative aux émissions industrielles.

Commission Européenne, 2010. Directive 2010/75/EU of the European Parliament and of the Council of 24 November 2010 on industrial emissions (integrated pollution prevention and control).

Commission Européenne, 2006. Reference document on Economics and Cross-Media Effects.

Commission Européenne, 2005. Rapport de la Commission au Conseil et au Parlement Européen – rapport de la Commission concernant la mise en œuvre de la directive 96/61/CE relative à la prévention et à la réduction intégrées de la pollution.

Commission Européenne, 1996. Directive 96/61/CE du Conseil du 24 septembre 1996 relative à la prévention et à la réduction intégrées de la pollution.

Conti, M.E., Ciasullo, R., Tudino, M.B., Matta, E.J., 2015. The industrial emissions trend and the problem of the implementation of the Industrial Emissions Directive (IED). Air Qual. Atmosphere Health 8, 151–161.

Daddi, T., De Giacomo, M.R., Dils, E., Polders, C., Lepe, G.R., Calvo, V.V., 2013. Transferring the Integrated Pollution Prevention and Control (IPPC) Approach and Best Available Techniques (BAT) Concepts to Egypt, Tunisia and Morocco. Sustainability 5, 2944–2959.

De Chefdebien, H., 2006. BAT & BREF : WI-BREF ? BATs pour l'incinération.

Derden, A., Huybrechts, D., 2013. Brominated flame retardants in textile wastewater : reducing Deca-BDE using best available techniques. J. Clean. Prod. 53, 167–175.

Derden, A., Vercaemst, P., Dijkmans, R., 2002. Best available techniques (BAT) for the fruit and vegetable processing industry. Resour. Conserv. Recycl. 34, 261–271.

Dijkmans, R., 2000. Methodology for selection of best available techniques (BAT) at the sector level. J. Clean. Prod. 8, 11–21.

EIPPCB, 2016. Reference documents under the IPPC Directive and the IED [En ligne]. Eur. IPPC Bur. URL http://eippcb.jrc.ec.europa.eu/reference/ (accédé le 06/06/2016).

EIPPCB, 2015. Kick-off Meeting Report for the review of the Food, Drink and Milk Industries (FDM) BREF [En ligne]. URL http://eippcb.jrc.ec.europa.eu/reference/BREF/FDM_KoM_Report_102014.pdf (consulté le 4.15.15).

Evrard, D., Laforest, V., Villot, J., Gaucher, R., 2016. Best Available Technique assessment methods : a literature review from sector to installation level. J. Clean. Prod. 121, 72–83.

Geldermann, J., Rentz, O., 2004. The reference installation approach for the techno-economic assessment of emission abatement options and the determination of BAT according to the IPPC-directive. J. Clean. Prod. 12, 389–402.

Georgopoulou, E., Hontou, V., Gakis, N., Sarafidis, Y., Mirasgedis, S., Lalas, D.P., Loukatos, A., Gargoulas, N., Mentzis, A., Economidis, D., Triantafilopoulos, T., Korizi, K., 2008. BEAsT : a decision-support tool for assessing the environmental benefits and the economic attractiveness of best available techniques in industry. J. Clean. Prod. 16, 359–373.

Giner-Santonja, G., Aragonés-Beltrán, P., Niclós-Ferragut, J., 2012. The application of the analytic network process to the assessment of best available techniques. J. Clean. Prod. 25, 86–95.

Glasson-Cicognani, M., Berchtold, A., 2010. Imputation des données manquantes : Comparaison de différentes approches. Presented at the 42èmes Journées de Statistique.

Gonzalez, P.-L., 2008. Méthodes de classification.

Halog, A., Schultmann, F., Rentz, O., 2001. Using quality function deployment for technique selection for optimum environmental performance improvement. J. Clean. Prod. 9, 387–394.

Hastie, T., Tibshirani, R., Sherlock, G., Eisen, M., Brown, P., Botstein, D., 1999. Imputing missing data for gene expression arrays. Division of Biostatistics, Stanford university.

Ibáñez-Forés, V., Bovea, M.D., Azapagic, A., 2013. Assessing the sustainability of Best Available Techniques (BAT) : methodology and application in the ceramic tiles industry. J. Clean. Prod. 51, 162–176.

Ibáñez-Forés, V., Bovea, M.D., Pérez-Belis, V., 2014. A holistic review of applied methodologies for assessing and selecting the optimal technological alternative from a sustainability perspective. J. Clean. Prod. 70, 259–281.

INERIS, 2014, La Directive sur les émissions industrielles, [rapport en ligne]. URL http://www.ineris.fr/centredoc/ficheied-2014-1416904340.pdf.

INSEE, 2016. Insee – Définitions, méthodes et qualité – Entreprise de taille intermédiaire / ETI [En ligne]. URL http://www.insee.fr/fr/methodes/default.asp ? page=definitions/entreprise-taille-intermedi.htm (accédé le 30/06/2016).

ISO, 2006. ISO 14040 :2006 environmental management – life cycle assessment – principles and framework.

Kabak, M., Köse, E., Kırılmaz, O., Burmaoğlu, S., 2014. A fuzzy multi-criteria decision making approach to assess building energy performance. Energy Build. 72, 382–389.

Laforest, V., 2014. Assessment of emerging and innovative techniques considering best available technique performances. Resour. Conserv. Recycl. 92, 11–24.

Laforest, V., 2008. Applying Best Available Technologies in environmental management accounting : from the definition to an assessment method, in : Environmental Management Accounting for Cleaner Production, EMAN. Springer, New York, pp. 29–48.

Laforest, V., Gaucher, R., 2015. Industrial Emisson Directive (IED) and Best Available Techniques. Directive sur les émissions industrielles (IED) et meilleures techniques disponibles. Tech. Ing.

Larousse, 2016. Représentatif. Petit Larousse Illus.

Mavrotas, G., Georgopoulou, E., Mirasgedis, S., Sarafidis, Y., Lalas, D., Hontou, V., Gakis, N., 2007. An integrated approach for the selection of Best Available Techniques (BAT) for the industries in the greater Athens area using multi-objective combinatorial optimization. Energy Econ. 29, 953–973.

National Observatory of Athens, 2006. Decision Aid Framework for the assessment of impacts from the introduction of Best Available Techniques in industry (DAF-BAT) (Final Project Report Executive Summary).

Nguyen, H.-T., Dawal, S.Z.M., Nukman, Y., Aoyama, H., 2014. A hybrid approach for fuzzy multi-attribute decision making in machine tool selection with consideration of the interactions of attributes. Expert Syst. Appl. 41, 3078–3090.

Nicholas, M.J., Clift, R., Azapagic, A., Walker, F.C., Porter, D.E., 2000. Determination of "Best Available Techniques" for Integrated Pollution Prevention and Control : A Life Cycle Approach. Process Saf. Environ. Prot. 78, 193–203.

OSPAR, 2015. Best Available Techniques (BAT) and Best Environmental Practice (BEP) [En ligne]. URL http://www.ospar.org/content/content.asp ? menu=00120000000063_000000_000000 (accédé le 16/04/2015).

Perez Gallardo, J.R., 2013. Ecodesign of large-scale photovoltaic (PV) systems with multi-objective optimization and Life-Cycle Assessment (LCA). Université de Toulouse, Toulouse.

Polders, C., Van den Abeele, L., Derden, A., Huybrechts, D., 2012. Methodology for determining emission levels associated with the best available techniques for industrial waste water. J. Clean. Prod. 29–30, 113–121.

Ramousse, R., Le Berre, M., Le Guelte, L., 1996. Statistique non paramétrique [En ligne]. Introd. Aux Stat. URL http://www.cons-dev.org/elearning/stat/ (accédé le 30/06/2016).

Rodríguez, M.T.T., Andrade, L.C., Bugallo, P.M.B., Long, J.J.C., 2011. Combining LCT tools for the optimization of an industrial process : Material and energy flow analysis and best available techniques. J. Hazard. Mater. 192, 1705–1719. Roudier, S., 2013. De l'IPPC à l'IED – Évolution des documents BREF et modalités. Colloque AFITE. La directive IED. Quels enjeux ? Quelles conséquences ?, Paris, 14 novembre 2013. 25 p.

Saaty, T.L., 1996. The Analytic Network Process, PA : RWS Publications. ed. PA : RWS Publications, Pittsburgh.

Saporta, G., 2011. Probabilités, analyse des données et statistique. Editions TECHNIP. 622 p.

Schollenberger, H., Treitz, M., Geldermann, J., 2008. Adapting the European approach of Best Available Techniques : case studies from Chile and China. J. Clean. Prod. 16, 1856–1864.

Stekhoven, D.J., Buhlmann, P., 2012. MissForest--non-parametric missing value imputation for mixed-type data. Bioinformatics 28, 112–118.

Testa, F., Daddi, T., De Giacomo, M.R., Iraldo, F., Frey, M., 2014. The effect of Integrated Pollution Prevention and Control regulation on facility performance. J. Clean. Prod. 64, 91–97.

Van Caneghem, J., Block, C., Van Hooste, H., Vandecasteele, C., 2010. Eco-efficiency trends of the Flemish industry : decoupling of environmental impact from economic growth. J. Clean. Prod. 18, 1349–1357.

Annexes

Annexe 1 : Description de MissForest

1. Les forêts aléatoires

Les forêts aléatoires, aussi appelées forêts d'arbres décisionnels, ont été proposées en 2001 par Leo Breiman et Adèle Cutler. Elles font partie des techniques d'apprentissage automatique, c'est-à-dire que cet algorithme effectue un apprentissage sur de multiples arbres de décision entraînés sur des sous-ensembles de données légèrement différents.

L'algorithme consiste à créer de nouveaux ensembles d'apprentissage par un double processus d'échantillonnage, d'une part en utilisant un tirage avec remise d'un nombre fixé d'observations identique à celui des données d'origine (bootstrap), et d'autre part, sur les prédicteurs, en n'en retenant qu'un échantillon. Sur chaque échantillon, on entraîne un arbre de décision (Figure 51), en limitant sa croissance par validation croisée. Les prédictions de la variable d'intérêt pour chaque observation d'origine sont stockées. La prédiction de la forêt aléatoire est alors un simple vote majoritaire. Dans l'exemple donné dans la Figure 51, une succession de questions vise à déterminer si une personne est malade ou non. La première d'entre elles consiste à demander si elle a de la fièvre (Température < 37,5 °C ?) et, selon la réponse donnée, le diagnostic est clair ou soulève une autre question sur l'irritation éventuelle de sa gorge. Cet exemple simple avec des variables qualitatives peut être transposé à des valeurs quantitatives, comme dans le cas des variables de consommations et d'émissions d'une installation industrielle.

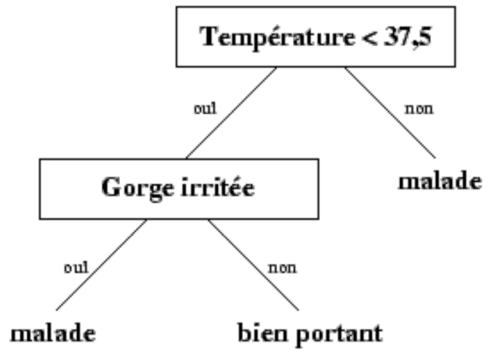


Figure 51 : Exemple d'arbre de décision pour déterminer si une personne est malade (source Wikipédia).

Cet outil permet ainsi d'extrapoler des valeurs manquantes à partir des valeurs existantes d'autres variables du même individu et des autres individus de la population en recourant à plusieurs arbres de décisions ; la forêt étant constituée de ces arbres.

2. MissForest

Dans ce travail de thèse, l'algorithme qui a été employé est celui créé par Stekhoven et Buhlmann (2012). Ces auteurs ont proposé une méthode de complétion basée sur les forêts aléatoires appelée MissForest. Cette méthode nécessite une première imputation « naïve », par défaut une complétion par la moyenne, afin d'obtenir un échantillon d'apprentissage complet. Puis une série de forêts aléatoires est ajustée jusqu'à la première dégradation du modèle. Autrement dit, un sous-échantillonnage est effectué afin d'extrapoler les valeurs manquantes en utilisant un arbre de décision permettant de déduire les manques.

Dans un exemple simple avec 14 individus pour lesquels les valeurs de deux variables contextuelles x_1 et x_2 sont connues (Figure 52). Une troisième variable décrit ces individus mais est connue seulement pour 13 individus ; elle manque donc pour l'individu « ? ». L'objectif est de trouver la valeur manquante d'après les valeurs des autres individus et à partir des valeurs connues de t que sont x_{1t} et x_{2t} (Figure 52a). La première étape réalise un premier sous-échantillonnage aléatoire des individus qui servira à établir les proximités moyennes entre individus. Ensuite, dans chaque sous-échantillonnage (Figure 52b), les valeurs probables de « ? » sont prédites à partir des informations apportées par des arbres de décision constitués d'autres individus. Enfin, l'algorithme stoppe dès que la valeur prédite est dégradée par rapport à l'échantillonnage initial. La valeur prédite est alors la moyenne de toutes les prédictions qui ont été réalisées (Figure 52c).

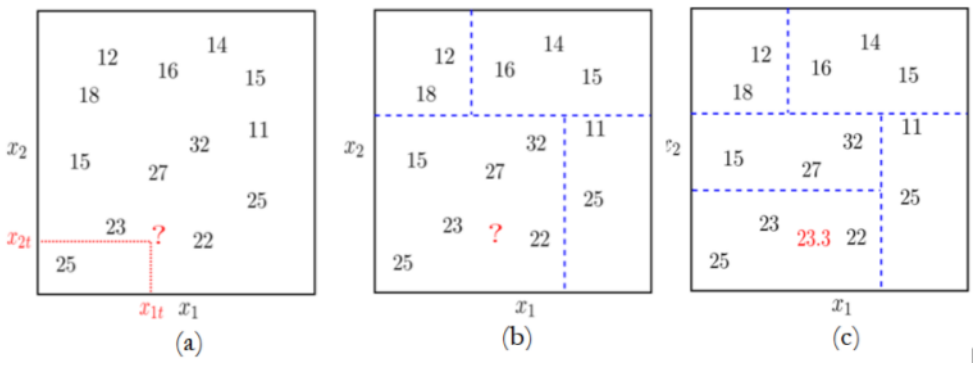


Figure 52 : Exemple simple de reconstruction avec les forets aléatoires (d'après Pierre Gaillard).

3. Estimation de l'erreur de la reconstruction

L'estimation proposée ici est l'erreur quadratique moyenne qui est la moyenne des carrés de la distance entre la valeur vraie et la valeur estimée. Ainsi, dans le cas de n valeurs estimées, la valeur Y et son estimation $\hat{Y}$, l'erreur quadratique moyenne MSE vaut :

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i) \qquad \text{Équation 5}$$

Dans le cas des forets aléatoires, l'erreur quadratique moyenne est égale à la somme de la variance et du carré du biais. De plus, lorsque l'arbre est suffisamment profond, le biais est faible et l'erreur équivaut à la variance.

Annexe 2 : Recherche des sources d'erreur pour la reconstruction

Dans MissForest, l'erreur n'est pas forcément liée au nombre de données manquantes. En effet, la reconstruction se basant sur les informations existantes, la difficulté de « déduire » les valeurs manquantes est la plus problématique. Néanmoins, il est certain que les sites complets n'engendrent pas d'erreur. Par conséquent, la recherche des installations sources d'erreur s'effectue de la manière suivante :

1. Pour les variables d'analyse et de normalisation choisies, les installations sont classées par nombre de données manquantes décroissant.

2. L'erreur globale de la reconstruction est calculée pour la population complète (p installations) puis en excluant 1 à 1 les installations les plus incomplètes.

3. Lorsqu'un saut dans la valeur de l'erreur est détecté par l'utilisateur après la nième installation, les n premières installations sont séparées des autres.

4. Chacune des n installations est ensuite ajoutée aux p-n installations se situant après le saut d'erreur détecté et l'erreur globale de la reconstruction est calculée à chaque fois.

5. Les n installations sont ensuite replacées dans la population par ordre d'erreur décroissante.

En répétant ces étapes pour chaque saut d'erreur détecté, la base de données est alors triée avec les sites qui sont les principales sources d'erreur placés en premier. Les populations suggérées correspondent donc à celles pour lesquelles un saut dans le taux d'erreur apparaît.

Annexe 3 : ACP et CAH

1. Analyse en Composantes Principales (ACP)

L'Analyse en Composantes Principales est une méthode d'analyse factorielle multivariée permettant d'explorer les liaisons entre variables et les ressemblances entre individus. Dans la méthode proposée, elle sert de préliminaire à une classification ascendante hiérarchique afin de regrouper les installations en fonction de leurs ressemblances exprimées par les valeurs que prennent plusieurs variables de consommations et d'émissions choisies.

Elle consiste d'abord à placer les installations sur les axes formés par les variables d'analyse centrées et réduites, donc dans un espace de dimensions le nombre de variables (Figure 53a). Ensuite, elle permet de constituer une représentation dans un sous-espace qui fera perdre le moins d'information possible (Figure 53b). Pour cela, des axes orthogonaux sont recherchés pour placer les installations dans un nuage de points ; ce sont les composantes principales. Minimiser la perte d'information consiste à minimiser la somme des carrés des distances entre les axes « vrais » et les composantes principales. Autrement dit, on recherche la projection de variance (inertie) maximale orthogonale au premier axe puis aux axes suivants.

Les composantes principales sont donc une relation linéaire des variables initiales. L'ACP est en général utilisée pour réduire le nombre de dimensions considérées et accélérer des temps de calcul. Toutefois, dans cette thèse, cette réduction du nombre de composantes principales a seulement été utilisée dans le calcul de l'INPE.

Figure 53 : Principe de l'Analyse en Composantes Principales.

Ainsi, le premier axe apporte le plus d'information possible. Le second axe apporte de l'inertie non prise en compte par le premier axe, et ainsi de suite jusqu'à ce que toute l'information soit considérée par la somme des inerties de chaque axe. L'inertie est donc la somme pondérée des

carrés des distances des individus au centre de gravité du nuage de points. Elle mesure donc la dispersion du nuage et est égale à la somme des variances des variables étudiées.

Dans des cas simples, la force de l'ACP est de pouvoir visualiser graphiquement les positions des différents individus selon des axes factoriels calculés en fonction de l'inertie ou de la variance des différentes variables en projetant sur les deux ou trois premières composantes principales. Néanmoins, dans les cas où plus de trois axes sont utilisés, une classification visuelle est difficile et une méthode de classification est nécessaire : la classification ascendante hiérarchique. Cela a en outre permis de ne pas avoir à réduire le nombre d'axes et de conserver ainsi l'intégralité de l'information.

2. Classification ascendante hiérarchique

La classification ascendante hiérarchique consiste à regrouper 2 à 2 des individus puis des groupes d'individus en fonction d'un critère de distance. Sa première étape consiste à partir d'une partition où chaque individu constitue une classe. Ensuite, les individus les plus proches suivant une distance d définie sur les individus sont agrégés et l'on remplace les individus agrégés par leur centre de gravité. Les classes obtenues sont ensuite fusionnées avec les plus proches à partir d'une distance D appelée indice d'agrégation des classes jusqu'à obtenir une classe formée de tous les individus.

Il existe plusieurs méthodes pour définir l'indice d'agrégation D. Dans la méthode proposée, la méthode de Ward a été retenue. Le critère de Ward consiste à agréger les individus qui font le moins varier l'inertie intra-classe. L'indice de dissimilarité entre deux classes est alors égal à la perte d'inertie interclasse résultant de leur regroupement (l'inertie correspond à la distance entre le centre de gravité des classes et celui du nuage).

Dans la méthode de Ward, l'indice d'agrégation utilisé est défini par l'Équation 6 :

$$D(A, B) = \frac{\sum (P_A * P_B)}{P_A + P_B} * d^2(g_A, g_B) \qquad \textbf{Équation 6}$$

où P_A, P_B représentent les masses respectives de A et B et g_A, g_B leur centre de gravité.

On obtient ainsi un classement des installations dans des classes caractérisées par leur centre de gravité. Il est possible d'obtenir pour chaque site sa distance au centre de sa classe ; le plus proche étant le plus représentatif (parangon) et également, la distance au centre des autres classes ; le plus éloigné étant le plus typique de sa classe.

De plus, la classification ascendante hiérarchique permet de déterminer le nombre optimal de classes sans intervention humaine, évitant par conséquent d'introduire de la subjectivité. Ainsi, dans une représentation en dendrogramme telle que présentée dans la Figure 55, la hauteur des sauts entre les différentes branches indique la qualité du découpage en classes. Dans cet exemple, le découpage en deux classes est le meilleur (ligne en pointillées dans la Figure 55).

Annexe 4 : Front de Pareto

La recherche du front de Pareto consiste, si l'on utilise une métaphore, à désigner les meilleurs élèves de la classe, même s'ils ne sont pas les meilleurs dans toutes les matières prises indépendamment. On cherche donc ceux qui sont globalement les meilleurs.

Dans un problème d'optimisation multi-objectif avec :

1. *m* variables de décision qui sont les valeurs à optimiser (minimiser dans le cas de cette thèse). Ces valeurs sont notées x_i , $i \in \{1..., m\}$. Le vecteur X de *m* variables de décision est représenté par $X=(x_1, x_2, \ldots, x_m)$. Dans cette thèse, les variables de décision sont les variables de consommations et d'émissions.

2. *n* fonctions objectifs à optimiser notées f_i , $i \in \{1,..., n\}$ et le vecteur f de *m* objectifs pour $x \in X$ est représenté par :

$$f(x) = (f_1(x), f_2(x), \ldots, f_n(x)) .$$

Ici, *f* est la fonction identité ($f(x)=x$) car on souhaite minimiser les variables de décision elles-mêmes.

Les objectifs peuvent prendre des contraintes. Les contraintes limitent les valeurs que peuvent prendre les variables de décision. Elles sont notées $g_i(x), i \in (1,\ldots,p)$. Elles pourraient correspondre à des valeurs limites d'émission existantes.

Un problème d'optimisation multi-objectif est donc un triplet (X, f, g) qui consiste à minimiser $f(x)$ pour $x \in X$ sachant que $g(x) \leq 0$.

L'algorithme consiste ensuite à comparer les installations deux à deux. Soit $U=(u_1, u_2, \ldots, u_m)$. et $V=(v_1, v_2, \ldots, v_m)$ deux vecteurs de décision correspondant aux valeurs prises par deux installations pour les variables d'analyse choisies. Il est dit que V domine U $(V < U)$ pour un problème de minimisation si et seulement si

$$\begin{cases} \forall i \in 1,\ldots,n \; f_i(V) \leq f_i(U) \\ \exists i \in 1,\ldots,n \; f_i(V) < f_i(U) \end{cases}$$

Le front de Pareto FP^* est l'ensemble des vecteurs de décision qui ne sont pas dominés.

$$FP^{*} = \left\{ X \in X \mid \nexists X' \in X \quad X < X' \right\}$$

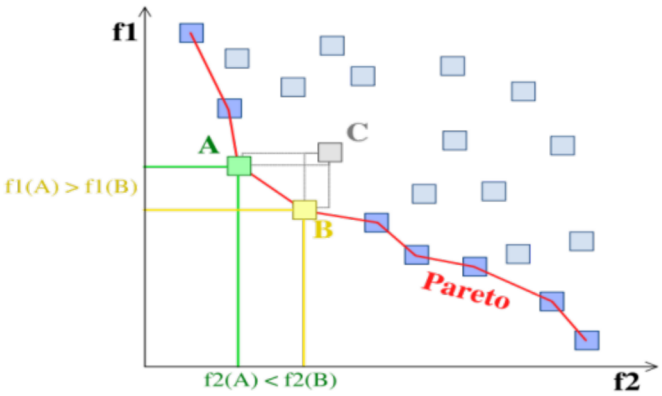


Figure 54 : Front de Pareto (d'après Nojhan pour Wikimedia).

Le front de Pareto est l'ensemble des solutions de compromis. Sur la Figure 55, les points A et B sont deux points du front de Pareto : A ne domine pas B, B ne domine pas A, mais tous les deux dominent le point C. Le but de l'optimisation multi-objectif est de déterminer le front de Pareto pour un problème donné.

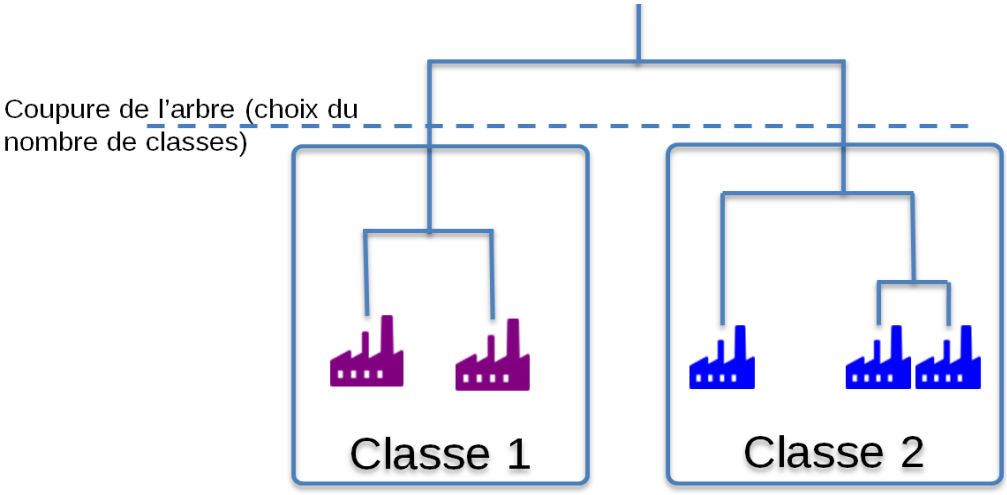
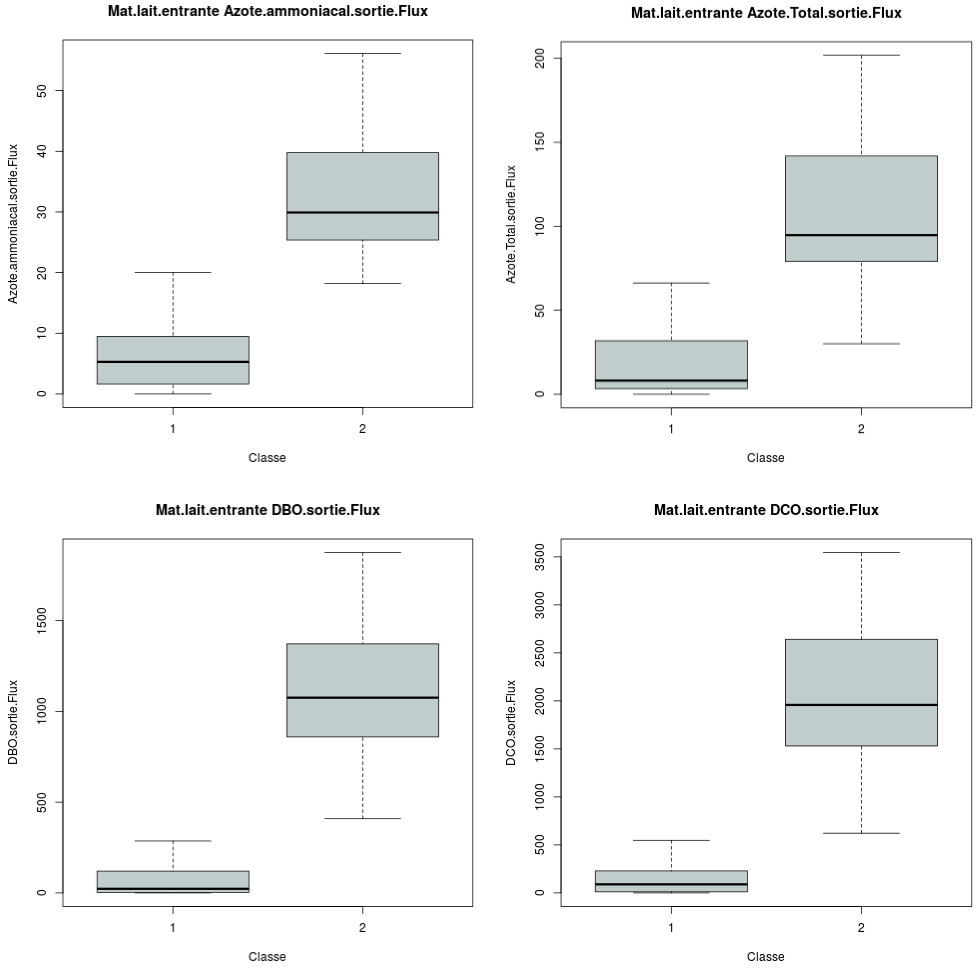
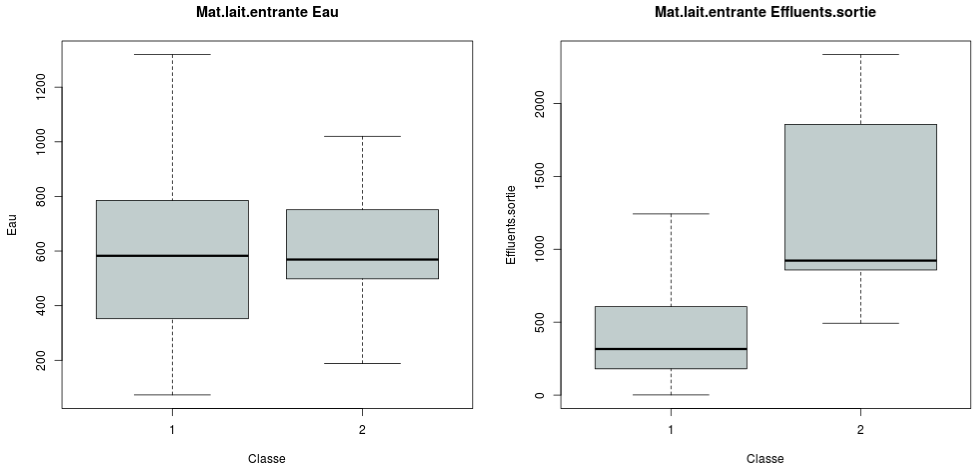


Figure 55 : Dendrogramme obtenu avec une Classification Ascendante Hiérarchique.

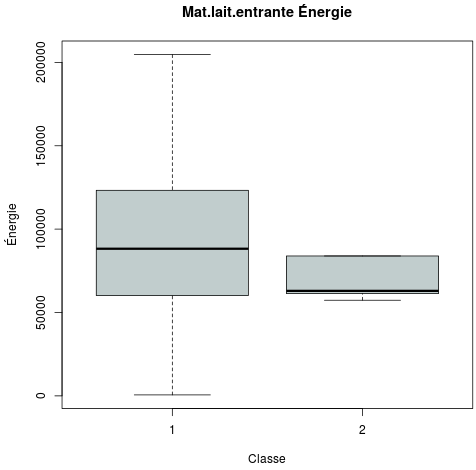
Les installations industrielles performantes sont donc les sites non dominés sur les variables d'analyse. Le nombre de sites performants obtenu avec cette méthode peut parfois être important et il n'est alors pas *a priori* possible de décider quels sites parmi ceux sur le front de Pareto est le meilleur au regard des valeurs de consommations et d'émissions. Cette étape se fait donc sans intervention humaine.

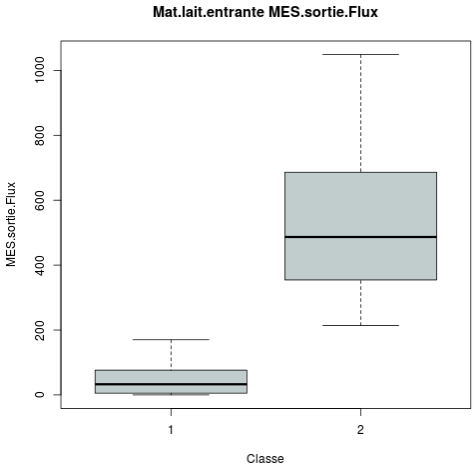
Annexe 5 : Boîtes à moustaches des variables d'analyse pour les classes représentatives de l'industrie laitière française

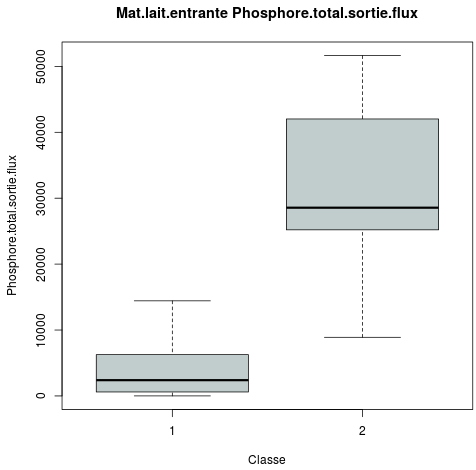



Mat.lait.entrante Eau
Eau
200
400
600
800
1000
1200
1
2
Classe

Mat.lait.entrante Effluents.sortie
Effluents.sortie
0
500
1000
1500
2000
1
2
Classe


Mat.lait.entrante Énergie
Énergie
0
50000
100000
150000
200000
1
2
Classe


Mat.lait.entrante MES.sortie.Flux
MES.sortie.Flux
0
200
400
600
800
1000
1
2
Classe


Mat.lait.entrante Phosphore.total.sortie.flux
Phosphore.total.sortie.flux
0
10000
20000
30000
40000
50000
1
2
Classe

Annexe 6 : Valeurs de référence pour l'industrie laitière française

Classes	Azote ammoniacal (kg/T matière laitière entrante)							
	Minimum		Maximum		Moyenne		Médiane	
	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort
Classe 1	0,29	93 %	5,50	47 %	1,51	79 %	1,08	83 %
Classe 2	29,91	46 %	39,78	23 %	34,84	38 %	34,84	38 %
Performants	0,00	99 %	19,61	20 %	3,98	63 %	0,27	95 %

Classes	Azote total (kg/T matière laitière entrante)							
	Minimum		Maximum		Moyenne		Médiane	
	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort
Classe 1	0,45	96 %	8,02	50 %	3,22	76 %	2,54	82 %
Classe 2	110,91	38 %	135,96	31 %	123,44	38 %	123,44	38 %
Performants	0,01	99 %	46,66	21 %	7,64	59 %	0,56	95 %

Classes	DBO_5 (kg/T matière laitière entrante)							
	Minimum		Maximum		Moyenne		Médiane	
	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort
Classe 1	0,73	92 %	11,08	57 %	2,77	74 %	1,48	82 %
Classe 2	1030,98	54 %	1371,94	23 %	1201,46	38 %	1201,46	38 %
Performants	0,01	99 %	481,68	21 %	83,60	40 %	0,38	95 %

Classes	MES (kg/T matière laitière entrante)							
	Minimum		Maximum		Moyenne		Médiane	
	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort
Classe 1	0,96	96 %	18,01	59 %	6,14	74 %	3,29	83 %
Classe 2	477,90	62 %	686,13	23 %	582,01	38 %	582,01	38 %
Performants	0,01	99 %	229,35	21 %	37,82	51 %	1,47	94 %

Classes	Phosphore total (kg/T matière laitière entrante)							
	Minimum		Maximum		Moyenne		Médiane	
	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort
Classe 1	205,77	94 %	2460,49	48 %	761,04	73 %	477,02	81 %
Classe 2	27068,59	62 %	42022,46	23 %	34545,52	38 %	34545,52	38 %
Performants	3,31	99 %	18300,29	20 %	2824,68	50 %	168,50	96 %

Classes	DCO (kg/T matière laitière entrante)							
	Minimum		Maximum		Médiane		Moyenne	
	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort
Classe 1	2,24	96 %	46,89	63 %	10,17	81 %	13,59	77 %
Classe 2	2023,95	38 %	2641,27	23 %	2332,61	38 %	2332,61	38 %
Performants	0,05	100 %	540	29 %	111	48 %	17	64 %

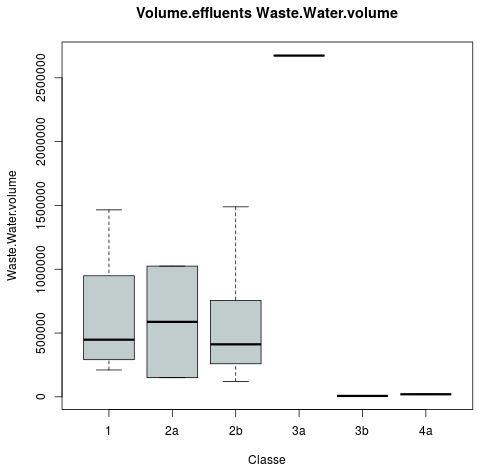
Classes	Volume d'effluents (kg/T matière laitière entrante)							
	Minimum		Maximum		Moyenne		Médiane	
	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort
Classe 1	121,86	87 %	1242,67	17 %	436,25	39 %	346,10	47 %
Classe 2	858,11	69 %	1856,83	23 %	1357,47	31 %	1357,47	31 %
Performants	1,10	100 %	632,25	35 %	159,86	82 %	68,30	93 %

Classes	Consommation d'énergie (MWh/T matière laitière entrante)							
	Minimum		Maximum		Moyenne		Médiane	
	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort
Classe 1	51959,17	83 %	147568,09	13 %	86336,26	53 %	86334,31	53 %
Classe 2	62871,38	54 %	203870,47	8 %	133370,92	23 %	133370,92	23 %
Performants	568,16	99 %	118737,49	30 %	55836,65	79 %	53501,31	79 %

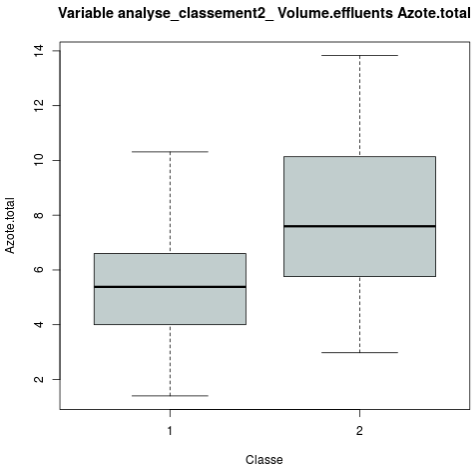
Classes	Consommation d'eau (m^3/T matière laitière entrante)							
	Minimum		Maximum		Moyenne		Médiane	
	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort
Classe 1	142,34	91 %	811,62	20 %	528,34	56 %	569,00	51 %
Classe 2	188,17	92 %	1020,09	8 %	604,13	38 %	604,13	38 %
Performants	73,24	99 %	600,32	48 %	213,97	86 %	106,59	96 %

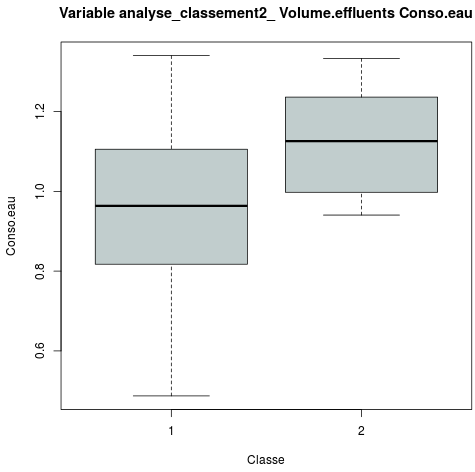
Annexe 7 : Boîtes à moustaches pour les classes représentatives de l'industrie laitière européenne

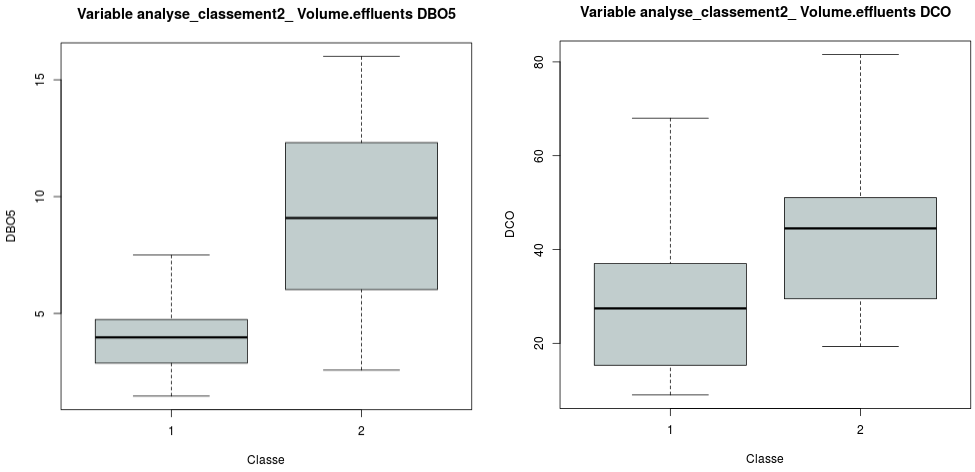
1. Variable de normalisation



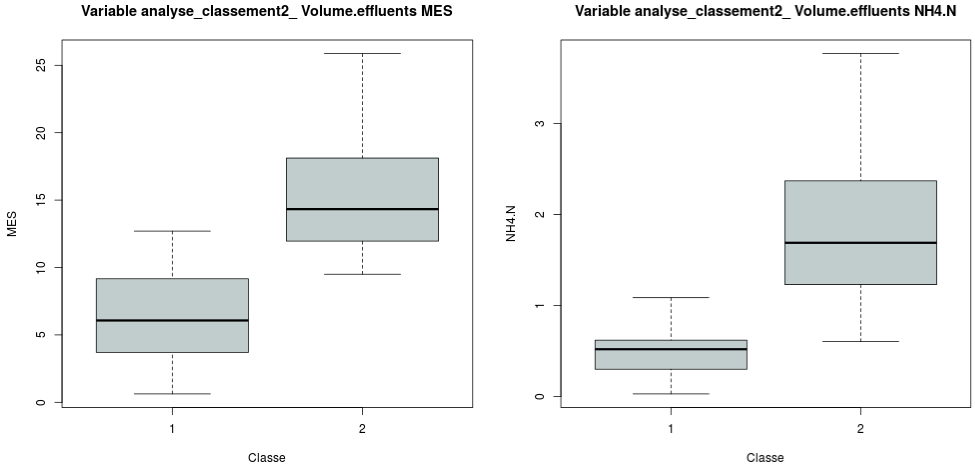
2. Variables d'analyse



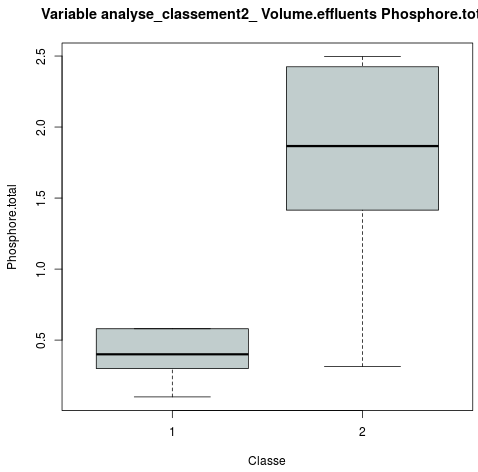



Variable analyse_classement2_ Volume.effluents DBO5
DBO5
Classe

Variable analyse_classement2_ Volume.effluents DCO
DCO
Classe

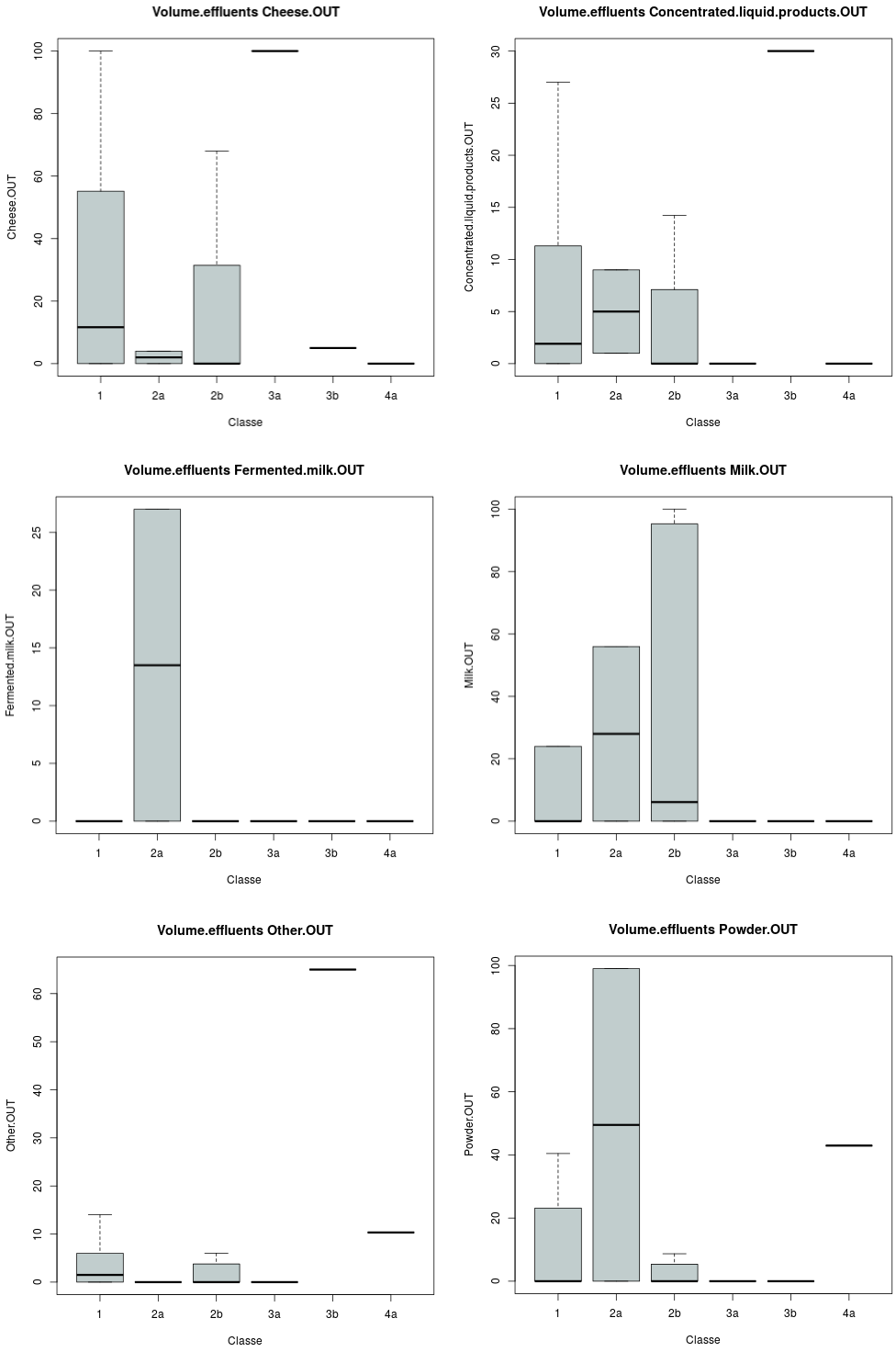

Variable analyse_classement2_ Volume.effluents MES
MES
Classe

Variable analyse_classement2_ Volume.effluents NH4.N
NH4.N
Classe

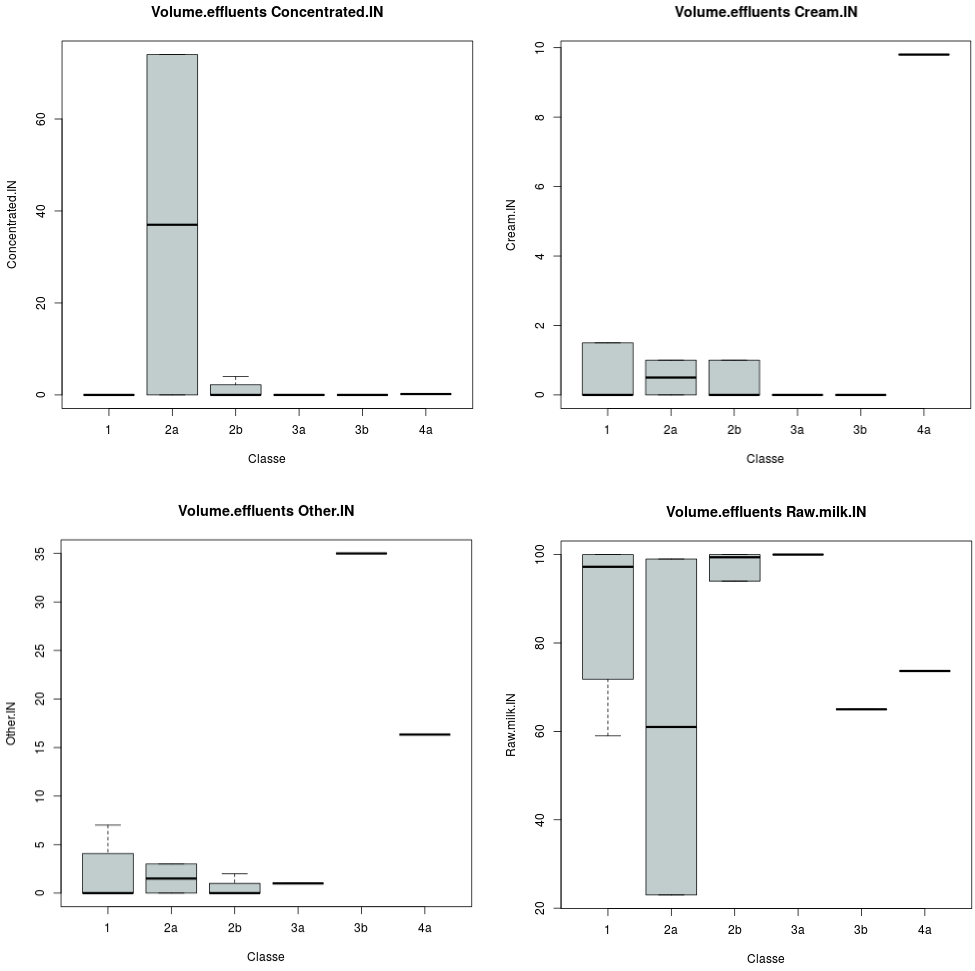

Variable analyse_classement2_ Volume.effluents Phosphore.total
Phosphore.total
Classe

3. Variables descriptives

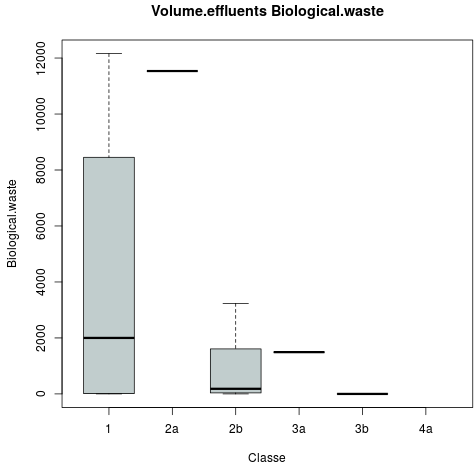
3.1. Proportions de produits fabriqués

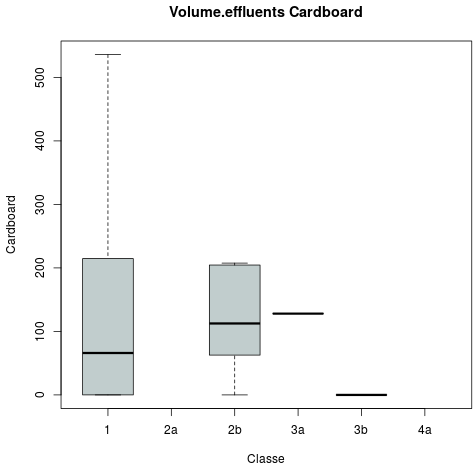


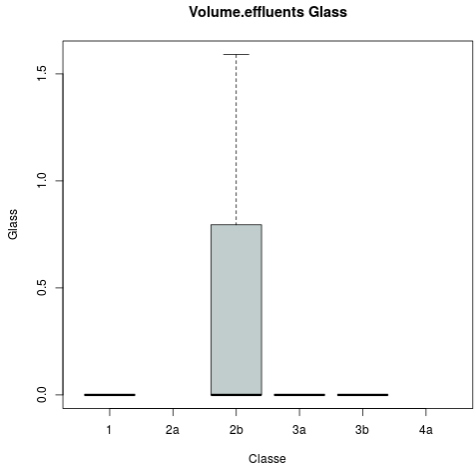
3.2. Proportion de produits entrants

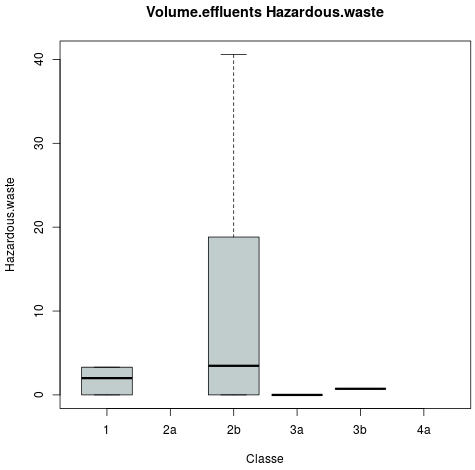


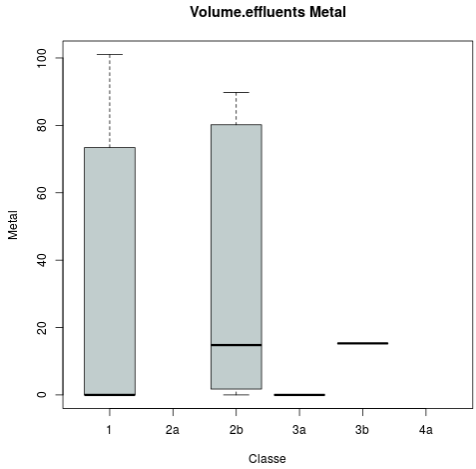
3.3. Déchets

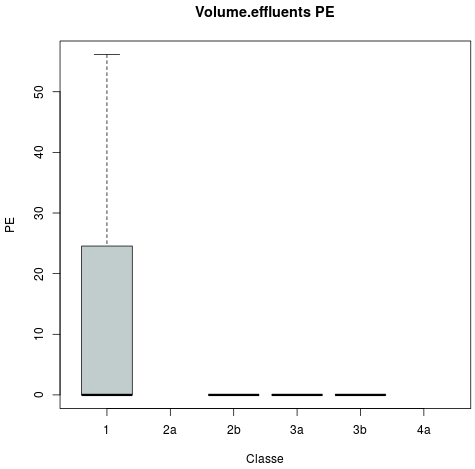

Volume.effluents Biological.waste
Biological.waste
0
2000
4000
6000
8000
10000
12000
1
2a
2b
3a
3b
4a
Classe

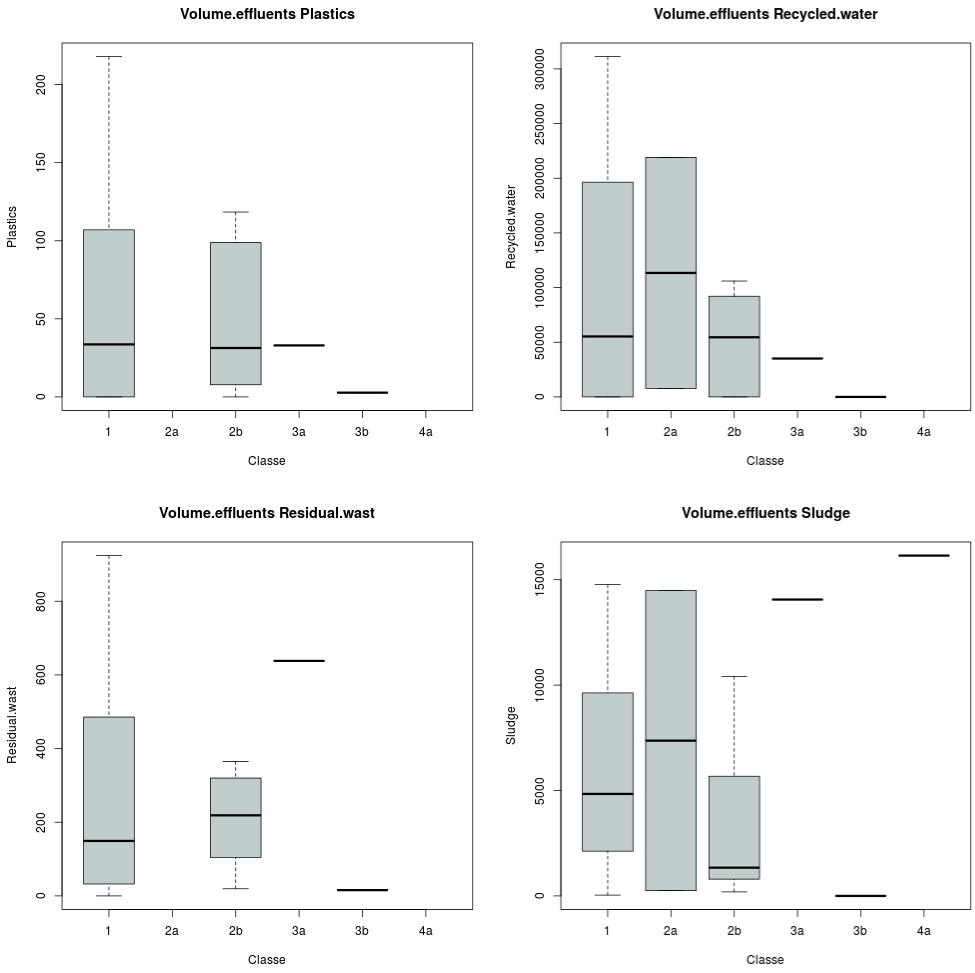

Volume.effluents Cardboard
Cardboard
0
100
200
300
400
500
1
2a
2b
3a
3b
4a
Classe


Volume.effluents Glass
Glass
0.0
0.5
1.0
1.5
1
2a
2b
3a
3b
4a
Classe


Volume.effluents Hazardous.waste
Hazardous.waste
0
10
20
30
40
1
2a
2b
3a
3b
4a
Classe


Volume.effluents Metal
Metal
0
20
40
60
80
100
1
2a
2b
3a
3b
4a
Classe


Volume.effluents PE
PE
0
10
20
30
40
50
1
2a
2b
3a
3b
4a
Classe


Volume.effluents Plastics
Plastics
0
50
100
150
200
1
2a
2b
3a
3b
4a
Classe
Volume.effluents Recycled.water
Recycled.water
0
50000
100000
150000
200000
250000
300000
1
2a
2b
3a
3b
4a
Classe
Volume.effluents Residual.wast
Residual.wast
0
200
400
600
800
1
2a
2b
3a
3b
4a
Classe
Volume.effluents Sludge
Sludge
0
5000
10000
15000
1
2a
2b
3a
3b
4a
Classe

Annexe 8 : Valeurs de référence pour les installations laitières européennes

Classes	Azote ammoniacal (kg/T matière laitière entrante)							
	Minimum		Maximum		Moyenne		Médiane	
	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort
Classe 1	0,03	94 %	0,53	44 %	0,21	78 %	0,06	89 %
Classe 2	1,20	73 %	2,72	20 %	1,89	40 %	1,73	40 %
Performants	0,03	97 %	3,11	13 %	0,69	58 %	0,54	74 %
Classes	Azote total (kg/T matière laitière entrante)							
	Minimum		Maximum		Moyenne		Médiane	
	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort
Classe 1	3,60	78 %	5,59	44 %	4,79	56 %	5,17	50 %
Classe 2	7,50	53 %	9,71	33 %	8,27	47 %	7,60	47 %
Performants	1,40	97 %	204,00	0 %	17,71	11 %	5,77	61 %
Classes	DBO_5 (kg/T matière laitière entrante)							
	Minimum		Maximum		Moyenne		Médiane	
	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort
Classe 1	2,00	83 %	4,74	22 %	3,68	67 %	4,29	33 %
Classe 2	6,42	67 %	10,49	33 %	7,97	53 %	7,00	53 %
Performants	1,46	97 %	363,34	3 %	26,48	11 %	3,87	76 %
Classes	MES (kg/T matière laitière entrante)							
	Minimum		Maximum		Moyenne		Médiane	
	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort
Classe 1	0,62	94 %	6,09	44 %	4,17	72 %	5,80	56 %
Classe 2	16,72	33 %	20,00	13 %	18,60	27 %	19,09	20 %
Performants	0,62	97 %	45,77	8 %	9,82	55 %	6,07	76 %
Classes	Phosphore total (kg/T matière laitière entrante)							
	Minimum		Maximum		Moyenne		Médiane	
	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort
Classe 1	0,30	67 %	0,55	28 %	0,43	50 %	0,43	44 %
Classe 2	0,74	80 %	5,00	7 %	2,70	20 %	2,35	27 %
Performants	0,10	97 %	91,00	0 %	6,3	5 %	0,40	74 %
Classes	DCO (kg/T matière laitière entrante)							
	Minimum		Maximum		Médiane		Moyenne	
	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort
Classe 1	15,32	72 %	39,60	6 %	27,30	50 %	26,99	50 %
Classe 2	44,50	47 %	52,11	20 %	47,94	40 %	47,22	40 %
Performants	9,00	97 %	3064,00	0 %	205,05	8 %	25,40	68 %
Classes	Consommation d'eau (m^3/T matière laitière entrante)							
	Minimum		Maximum		Moyenne		Médiane	
	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort	Valeur de référence	Taux d'effort
Classe 1	0,82	72 %	0,98	44 %	0,91	61 %	0,95	50 %
Classe 2	0,94	87 %	1,13	47 %	1,02	67 %	1,00	73 %
Performants	0,26	97 %	1,60	8 %	0,84	82 %	0,87	79 %

Annexe 9 : Publications scientifiques

Athens Institute for Education and Research
ATINER

ATINER's Conference Paper Series
ENV2015-1750

Identification of Representative Installations in the Best Available Techniques Determination Process: A Methodological Proposal

Damien Evrard
PhD Student
Mines Saint-Etienne
France

Valerie Laforest
Professor
Mines Saint-Etienne
France

Jonathan Villot
Lecturer
Mines Saint-Etienne
France

Rodolphe Gaucher
Head of Unit "Clean Technologies"
INERIS
France

Natacha Darmon
Head of Group Regulation & Installations
EDF-CIDEN
France

Sofia Bouhrizi
Engineer
EDF-CIDEN
France

ATINER CONFERENCE PAPER SERIES No: ENV2015-1750

An Introduction to ATINER's Conference Paper Series

ATINER started to publish this conference papers series in 2012. It includes only the papers submitted for publication after they were presented at one of the conferences organized by our Institute every year. This paper has been peer reviewed by at least two academic members of ATINER.

Dr. Gregory T. Papanikos
President
Athens Institute for Education and Research

This paper should be cited as follows:

Evrard, D., Laforest, V., Villot, J., Gaucher, R., Darmon, N. and Bouhrizi, S. (2015). **"Identification of Representative Installations in the Best Available Techniques Determination Process: A Methodological Proposal",** Athens: ATINER'S Conference Paper Series, No: **ENV2015-1750.**

Athens Institute for Education and Research
8 Valaoritou Street, Kolonaki, 10671 Athens, Greece
Tel: + 30 210 3634210 Fax: + 30 210 3634209 Email: info@atiner.gr URL: www.atiner.gr
URL Conference Papers Series: www.atiner.gr/papers.htm
Printed in Athens, Greece by the Athens Institute for Education and Research. All rights reserved. Reproduction is allowed for non-commercial purposes if the source is fully acknowledged.
ISSN: 2241-2891
14/12/2015

Identification of Representative Installations in the Best Available Techniques Determination Process: A Methodological Proposal

Damien Evrard

Valerie Laforest

Jonathan Villot

Rodolphe Gaucher

Natacha Darmon

Sofia Bouhrizi

Abstract

Best Available Techniques (BATs) in Europe have become crucial for the industrial sectors concerned with the Industrial Emission Directive (IED) and for other sectors like the nuclear industry. In the IED, the creation and revision process to define a list of BATs which will be used as a reference by operators and environmental authorities is known as the "Sevilla Process". The decision process is based upon exchange of information between stakeholders, data analysis and expert judgement. Several methods to assess technique performances compared to BATs from reference documents or to support the determination and the application of BATs have been developed. However, existing tools integrate either the European or the local level but not both levels. This situation makes the application of BATs highly dependent on expert judgements. In order to contribute to strengthen the robustness of BAT determination and application, the authors argue that there is a need for an integrated method to help define a list of BATs and to support the decision for their application. This paper presents the first results of an on-going research aiming to answer these questions. It first introduces the context in which the BAT concept is used. In a second section, a method is proposed to select installations considered as representative of these issues in a view of illustrating the current situation in the sector and fuelling the discussions between experts. Finally, the proposed method is discussed and perspectives for selecting BATs are given.

Keywords: Best Available Technique, Industrial Emission Directive, Statistical analysis

Acknowledgments: Our thanks to CNIEL for allowing us to use the data of the French dairy industry they gathered in our study.

Introduction

A growing number of industrial sectors must apply Best Available Techniques (BATs). Since 1996, BATs have been applied in the industrial sectors within the scope of the Integrated Pollution Prevention and Control (IPPC) directive (European Commission, 1996). The role of these BATs has then been reinforced and strengthened in the Industrial Emission Directive (IED) (European Commission, 2010)

BATs have implied obligations at two levels: (1) the European Commission has had to set the appropriate process in order to provide sector-specific reference documents describing these techniques (called BREF); (2) local environmental authorities and operators have had to compare local installation performances with the information contained into these reference documents to define emission limit values.

In the IED, "Best Available Technique" is defined in article 3 as "*the most effective and advanced stage in the development of activities and their methods of operation which indicates the practical suitability of particular techniques for providing the basis for emission limit values and other permit conditions designed to prevent and, where that is not practicable, to reduce emissions and the impact on the environment as a whole*". Considered individually, the three words have the following meaning:

- "Technique" includes both the production technology used and the way in which the installation is designed, built, maintained, operated and decommissioned. In other words, a technique is not limited to a pollution abatement device but also encompasses management approaches such as environmental management systems.
- "Availability" means the technique considered is developed on a scale which allows an implementation in the industrial sector, under economically and technically viable conditions. These conditions take into consideration the costs and advantages, whether or not the technique is used or produced inside the Member State in question, and if it is reasonably accessible to the operator (Laforest, 2008).
- "Best" means that the technique considered is the most effective for achieving a high general level of protection of the Environment as a whole.

Some of the sectors which must apply BATs are outside the scope of the IED and, therefore, they have to refer to twelve criteria given in annex III of the IED to propose techniques corresponding to these definitions. This is the case, for example, in France for the nuclear industry (Arrêté, 2012). Nevertheless, these criteria are a general guidance for the Industry and are difficult to use on-site (Laforest, 2014).

In the IED, the process of BAT selection is formalised in an implementing decision (European Commission, 2012). This process, called the Sevilla Process, is an iterative process in which the selection of techniques in an industrial sector is carried out through expert judgement, and based on the collection of data from actual European sites. It is therefore the best basis for the determination of BATs, even in sectors outside the scope of the directive.

Furthermore, considering the high amount of information to consider in this process, the authors argue that a statistical approach may help experts in their decision-making process: (1) to identify key issues specific to a given industrial sector; (2) to classify installations according to their level of performance and peculiarities; (3) to select in each group, a sample of installations which can illustrate the whole.

This paper presents the first developments of a method for the selection of representative sites within an industrial sector based on a statistical analysis. These sites are meant to illustrate the sectoral issues and performances in order to select those which are likely to apply to BATs. Firstly, the context of this research work is drawn considering the Sevilla Process and the previous works on BAT selection. Secondly, the steps of the method are presented. Then, the limits and perspectives of the results are discussed. The paper ends with a conclusion about further studies to answer the limits identified.

Context of Application of Best Available Techniques

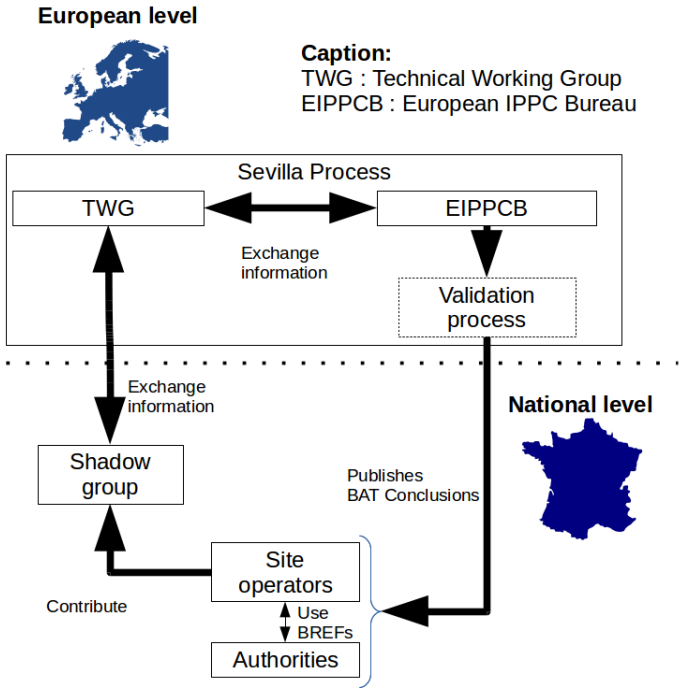
The Sevilla Process

The Sevilla Process is the official process to create or revise BREFs. It is based on information exchanges between Member States, the European Commission, the Industry, and environmental NGOs (European Commission, 2012). With the legal requirements associated with BAT-associated emission levels (BATAELs), the outcome of their revision has become strategic.

Figure 1 illustrates the interactions between the groups involved in the technical work within the Sevilla Process in order to publish a BREF at European and national levels, with the example of France. Thus, a Technical Working Group (TWG) composed of representatives from the organisations listed previously is formed in order to define the scope and plan the work. An important step is the collection of data about the performances of sites currently operating in Europe. The extensive analysis of information is requiring an increasingly complex data processing phase. The transition from this step of information analysis to the definition of BATs highly relies on the expertise of its actors with a risk of a biased assessment.

The objective of this on-going research is to look for a way to help, with a scientific method, the national experts in the shadow group (Figure 1) during their collection of the most relevant information to provide to the Sevilla Process. Shadow groups are meant to assist TWG members in their preparatory work for the Sevilla Process. They can be composed of the national TWG members, representatives from national authorities and industrials. In order to accomplish this, data about national installations which may illustrate the performances reachable thanks to a good application of BATs are sought. The Sevilla Process is used as a reference framework but the developments of this research are aimed to be applicable to other industrial sectors outside the scope of the IED, in particular to the nuclear industry.

Figure 1. *Exchange between National and European Actors in the IED*



The present on-going research follows the development of the revision of the reference document for the Food, Drink and Milk industry (EIPPCB, 2015). It especially focuses on the dairy sub-sector. All the following developments were experimented on data provided by the French dairy industry. For confidentiality reasons, details on the data used cannot be published in this article.

Previous Works on Bat Selection

Several existing methods to determine sectoral BATs were found in the scientific literature. The first one is a local BAT determination method which mainly relies on expert judgement supported by mostly qualitative data (Dijkmans, 2000; Polders et al., 2012) and was developed in Flanders. Then, two other methods aiming to reduce subjective elements are described. The first one by Geldermann and Rentz (2004) relies on more factual parameters to assess techniques, and the second one by the National Observatory of Athens (2006) uses mathematical approaches to guide the decision.

To sum up, these approaches to determine BATs are greatly influenced by their area of application (from Europe-wide or a given region) and their context (actors, end-user and data sources). Thus, either a method starts from a given sector to consider all relevant pollutants, or it begins from one particular environmental issue to select the relevant sectors which could contribute to an improvement thanks to the use of BATs. Moreover, they differ in the number of installations considered. When this number is too high to consider every

single installation, a sampling is used (European Commission, 2012; Geldermann and Rentz, 2004).

The Sevilla Process and Dijkmans' method seem to focus on emissions, though also consider larger performance considerations, whereas Geldermann and Rentz focus on environmental impacts, and the National Observatory of Athens on economical valuation of impacts. Other mathematical approaches can be found in the literature, although in articles about the selection of techniques on-site.

Two other methods for the determination of sectoral BATs were found (Nicholas et al., 2000; Halog et al., 2001) with a focus on one very specific sector. Halog et al. (2001) have a concern for the assessment and comparison of techniques by expert judgement, whereas Nicholas et al. (2000) use Life Cycle Assessment. Moreover, in Halog et al. (2001), a preliminary step of determination of the key impacts is first carried out and then, the issues for BAT determination are considered as stakeholders' requirements. Therefore, these two methods only illustrate the issue of the "BAT candidate" assessment and the comparison of technique performance without detailing any representativeness assessment of the candidate techniques. In this aspect, they might be more appropriate if considered as tools at the installation level.

Overall, all the methods found focus on the selection of techniques and performance levels associated to techniques without explaining the upstream decision process based on the selection of installations. Such a preliminary process would be essential firstly to identify sectoral key environmental issues, then to select the installations which are bound to use techniques that might contribute to these issues, whether the level of performance is high or not. Another limit of these methods is that they are aimed to be applied in very specific contexts and do not consider both European and local applications illustrated in Figure 1.

Applied Methodological Approach

Data and Tools Used

Data Used

In parallel to the work undertaken by the French representatives in the Sevilla Process and the French Industry, a data collection was carried out by the dairy national federation in order to have an overview of the sector's performances and specific issues. Thus, the data on production, consumptions and emissions have been collected, in addition to a few elements of the applied techniques for 117 IED installations located in France. Therefore the data are either qualitative or quantitative, with levels of completeness varying from one variable to another or for one site to another. To sum up, the gathered data concern:

- **Generalities:** activities, production capacity, classification in the French regulation, date of construction, production of small series, seasonal activities;

- **Inputs:** nature and mass of milk and other dairy products, ingredients, water, and energies.
- **Outputs:** nature and mass of products (e.g. cheese, milk powder, whey, etc.).
- **Production and mitigation techniques:** curing, drying, filtration system, etc.
- **Emissions:** volume of wastewater, characteristics of wastewater, dust emissions, and waste.

Identified Specificities of the Sector

The main peculiarity of the sector is that most installations produce several products. Therefore, a site which is identified as a cheese manufacturer may also produce other products like whey. However, another one may also make whey but without cheese. This possibility of having various products made in the same installation makes the comparison between installations difficult.

Tools Used

Spreadsheet

The first tool to be used in this study is spreadsheet software in order to calculate basic statistics on the available quantitative data such as the averages, standard deviations, medians, minima, and maxima.

Statistical Analysis

In order to consider qualitative data such as the application of certain techniques, a statistical analysis software was used ("Coheris Analytics SPAD - Data Scientist - Logiciel Data Mining - Coheris," 2015). The method used in this study is the factorial analysis method of the Multiple Correspondence Analysis, which enables to group individuals according to their similarities and according to various characteristics.

Proposed Method

Overview

This study was decided to focus on the *key environmental issues* which had been defined during the kick-off meeting of the revision process for the FDM BREF (EIPPCB, 2015). Several of them were discarded due to a lack of available data: Total Organic Carbon and ammonium nitrogen. Thus, nine variables were considered to classify the installations: energy consumption, water consumption, chemical oxygen demand (BOC), biological oxygen demand (BOD), total suspended solids (TSS), chlorides (Cl^-), total nitrogen (TN), total phosphorus (TP), and dust emissions.

Due to the large amount of datasets to consider (9 variables for 117 installations, i.e. a maximum of 1,053 datasets), a statistical approach had seemed relevant to handle simultaneously so much information.

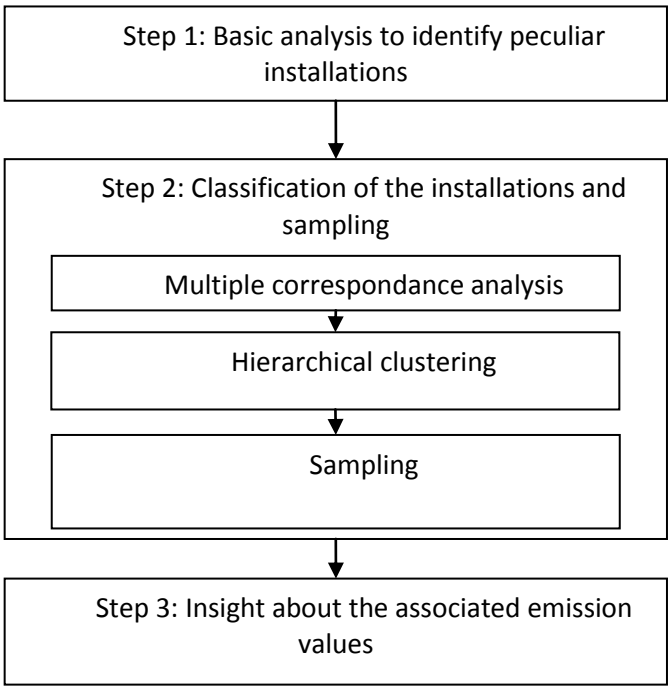
Besides, the installations differed in sizes and activities. Therefore, a common denominator appeared to be necessary to be able to compare all the

sites in spite of their differences. Because of the lack of an official and clear rule to normalise the installations, three common denominators have been considered in this study:

- **Input of processed milk in tons:** this is the preferred approach in Europe (EIPPCB, 2015); all the products are considered as undistinguishable (e.g. 1 T of cheese = 1 T of milk powders). This approach is relevant for installations which make only one kind of product, but seem to be a source of over- or under- estimation when considering several products at once.
- **Input of dairy matter in tons, assessed from milk-equivalent:** in this approach, the various products are converted into milk-equivalent weight according to conversion factors. The factors used are the one stated in the French regulation (INERIS, 2015).
- **Output of dairy products in tons:** it focuses on the product but does not consider the differences between products.

Without prejudice about which denominator is the most relevant, the three of them have been tested in order to find out the differences in classification that they may imply. Considering these three denominators to calculate ratios for each of the nine variables, the statistical analysis itself was carried out according to the three steps presented in Figure 2.

Figure 2. ***Steps of the Proposed Method***



Step 1: Basic Analysis to Identify Peculiar Installations

The objective of this step was to identify the installations whose characteristics were peculiar, i.e. to highlight those whose performances were different from the rest of the French dairy sector. The study first applied to the raw data in order to compare the values of each installation to target-values,

e.g. existing regulations. The aim here was to alert in case of deviation whether a site was out of bound. Then, a second analysis was carried out on the normalised data to search for installations deviating from the average of the sector. The aim here was also to alert about "extreme" installations. Therefore, experts' attention could be drawn on some of the sites to help them decide if they present peculiar issues or if they should be excluded from the analysis.

Further work should include the definition of cut-off rules in order to exclude odd installations from the beginning of the study. Nevertheless, exclusion would require ensuring that there are good reasons for it to prevent excluding singular but interesting installations. Therefore, this decision would be critical in a full-scale application and the decision may only be taken by expert judgment.

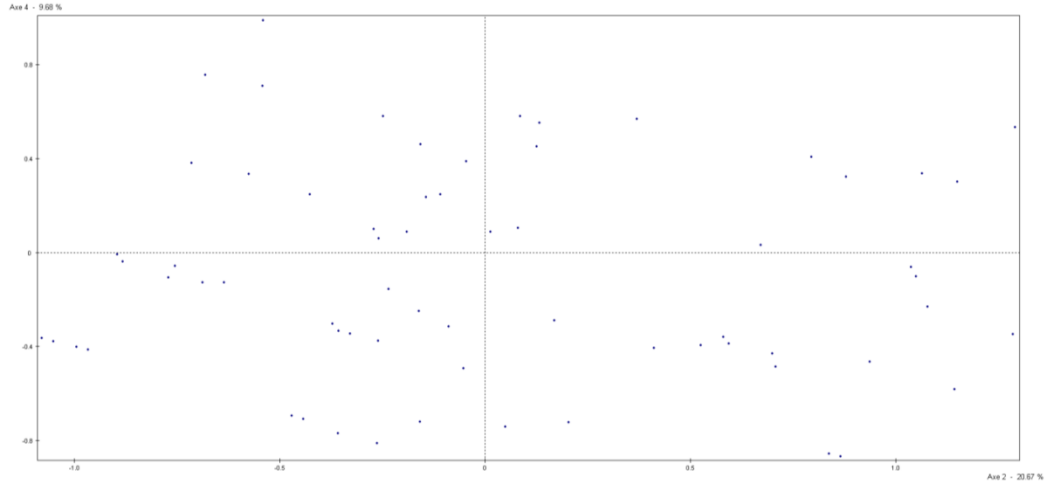
Step 2: Classification of the Installations and Sampling

The objective of this step was to classify the installations according to their performances described through the nine variables previously considered. Then, to deduce from the sub-groups identified one or several installations which illustrated their group as best as possible. In the study presented in this article, the approach was to focus on key environmental issues in order to suggest a short-list of installations to experts who would make the final decision. The final goal was to provide a decision-support tool to guide the experts' choice by considering simultaneously a large number of parameters and installations. Several operations were carried out on the normalised data (Figure 2).

Multiple Correspondance Analysis (MCA)

This factor analysis method enables us to study the relation between several nominal variables in a view to establish a classification based on their coordinates in a Euclidian space (Saporta, 2011). Its outcome is a graphical display which can be used to visualise the proximity between nominal categorical variables. However, considering the number of variables taken into account in this study, and the dispersion of the data, the graphical display was not enough to clearly identify groups of installations (dots in Figure 3). Therefore a formal classification method had to be used.

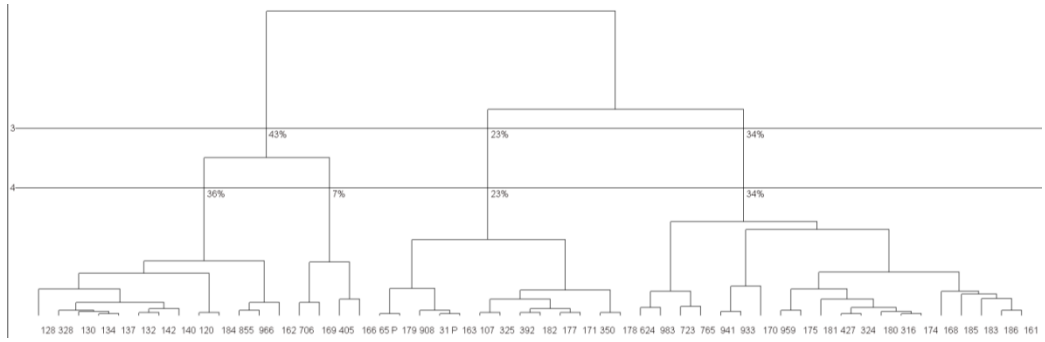
Figure 3. *Example of MCA Carried out on French Dairy Installations*



Hierarchical Clustering

Hierarchical clustering is an iterative method which calculates the dissimilarity between individuals, then pairs them according to an aggregation criterion to minimise cluster inertia. This operation is repeated until all individuals are in a cluster. The result is presented as a dendrogram as illustrated in Figure 4.

Figure 4. *Example of Dendrogram of Clustering of the French Dairy Installations*



This dendrogram enables us to select the most relevant number of classes. In Figure 4, two partitions would be the most relevant, statistically speaking. However, in order to better discriminate the installations, 4 groups were considered in order to take into account the specificities of the small second group. Thus, the first cluster would encompass 36% of the installations, the second 7%, the third 23%, and the fourth 34%.

Sampling of the Classes

The objective is to select a small number of installations which could illustrate their class in order to have a short-list of installations embodying the whole sector and its specificities. In this study, the closest individuals to the centre of each class were chosen, while introducing a few selection rules based on expert judgement. Thus:

- The list of the installations is ordered in ascending order by class and distance to its centre.
- Installations are considered according to their main activity.
- Once a main activity has been selected in a class, it cannot be selected a second time.
- If an installation is the only representative of its main activity, it will always be selected if it is the closest to the centre of its class.
- The selection within a class ends when the number of installations of a given main activity reaches a certain limit. For example, if 30% of the installations in a class are cheese-making installations and the following site to select is a milk powder installation whose activity is done by 20% of the installations in the class, then 50 % of the installations in the class should be covered and the selection in this class ends.
- If installations with the same main activity and distance to the centre of the class should be selected, all of them are selected.

In the end, a total of 19 installations were selected. In a real application, they would be submitted to experts for validation and possibly a new selection according to different selection criteria could be undergone.

Step 3: Insight about the Associated Emission Values

Thanks to the selected sites, it was possible to consider the emission values associated with them to assess the number of installations which would be excluded if they were used to create national emission values.

Thus, the percentage of installations whose values for the key environmental issues were below each selected site were assessed. Another use for this percentage could be to state a level of conformity for the selected sites. For example by deciding that 80% of the existing installation should be within the chosen limit value.

Limits and Perspectives

Common Denominator

The same method applied to the three common denominators showed that there were important differences in the classification of installations and sampling. Therefore, the impact on the clustering step should be further studied.

This could be answered by translating each qualitative variable (k discrete values) into k binary variables in each denominator. This should show invariant installations, i.e. installations which remain grouped together whatever the common denominator used.

Relevance of the Statistical Approach

General Considerations

This first exploration of the use of a statistical approach for installation classification and selection raised several questions about the depth of its relevance. Indeed, how far can it go to highlight installation specificities and to find out what technical characteristics influence their level of performance?

Another general issue is the management of missing data from the information collected on actual installations. Indeed, for some key environmental issues, 50% of the installations did not provide any values. How can such a situation be considered? What is its impact on the classification? What are the limits of applicability of the MCA?

Answers may be found thanks to sensitivity analyses, with or without their exclusion, or data reconstruction.

For Installation Classification

The first step of the proposed method implies to compare values according to their distance to a sectoral average but in cases where the values are very different from one installation to the next, the standard deviation may be too large to discriminate installations in a relevant manner. Therefore, are there any other approaches which may be more interesting such as the median and deviation to the median? How can expert judgment be used to target relevant sub-population without creating too important a bias?

These issues also concern the choice of the number of classes to consider. Is there a method which could help select the most appropriate number of classes to discriminate installations effectively? The meaning of the classes and the significance of a sub-division may bear part of the answer.

For Installation Selection

In the same class, several installations may appear to be at the same distance from its centre. So far, it is impossible to choose one rather than another and the decision seems to be dependant from experts. Is there a method to guide this decision or highlight the right information to help the decision-making?

A more fundamental issue is: what is a “representative” installation? In this article, it is presented as the closest to the centre of its class. In the Sevilla Process, “good-performers” are sought. In any cases, the criteria to select representative installations are still unclear. Therefore, a clear definition and indicators should be sought to assess the representativeness of installations.

Conclusions

To conclude, BAT determination is a complex issue even where a framework exists such as the Sevilla Process. Considering the large amount and wide variety of information to provide to decision makers, a statistical approach seems relevant although the availability of data makes it difficult to

relate to every variable. Nevertheless, a performance-oriented approach seems to be relevant since the purpose of BATs, as stated in the IED directive is to *"achiev[e] a high general level of protection of the Environment as a whole"*.

Further study in this project shall include a more critical analysis of the contribution of the statistical approach to the whole issue of representative and BAT identification. In addition, application to other sectors shall help identify where a generic approach might be possible.

References

Arrêté du 7 février 2012 fixant les règles générales relatives aux installations nucléaires de base, 2012.

Coheris Analytics SPAD - Data Scientist - Logiciel Data Mining - Coheris, 2015.

Dijkmans, R., 2000. Methodology for selection of best available techniques (BAT) at the sector level. Journal of Cleaner Production 8, 11–21. doi:10.1016/S0959-6526(99)00308-X.

EIPPCB, 2015. Kick-off Meeting Report for the review of the Food, Drink and Milk Industries (FDM) BREF [WWW Document]. URL http://bit.ly/1P1LWAS [accessed 4.15.15], 45p.

European Commission, 1996. Council directive 96/61/EC of 24 September 1996 concerning integrated pollution prevention and control, 22p.

European Commission, 2010. Directive 2010/75/EU of the European Parliament and of the Council of 24 November 2010 on industrial emissions (integrated pollution prevention and control). Off J Eur Union 2010, L334/17-118.

European Commission, 2012. Commission implementing decision of 10 February 2012 laying down rules concerning guidance on the collection of data and on the drawing up of BAT reference documents and on their quality assuranc referred to in Directive 2010/75/EU of the European Parliament and of the Council on industrial emissions. Off J Eur Union 2012, L63/1-39.

Geldermann, J., Rentz, O., 2004. The reference installation approach for the techno-economic assessment of emission abatement options and the determination of BAT according to the IPPC-directive. Journal of Cleaner Production 12, 389-402. doi:10.1016/S0959-6526(03)00032-5.

Halog, A., Schultmann, F., Rentz, O., 2001. Using quality function deployment for technique selection for optimum environmental performance improvement. Journal of Cleaner Production 9, 387–394. doi:10.1016/S0959-6526(00)00080-9.

INERIS, 2015. Rubrique ICPE 2230. Réception, stockage, traitement, transformation etc. du lait [Reception, storage, treatment, processing, etc. of milk] [WWW Document]. AIDA. URL http://bit.ly/1m0uWiN [accessed 4.15.15].

Laforest, V., 2008. Applying best available technologies in environmental management accounting: from the definition to an assessment method. In: Environmental Management Accounting for Cleaner Production. 5th EMAN book, 29–47.

Laforest, V., 2014. Assessment of emerging and innovative techniques considering best available technique performances. Resources, Conservation and Recycling 92, 11–24. doi:10.1016/j.resconrec.2014.08.009.

National Observatory of Athens, 2006. Decision Aid Framework for the assessment of impacts from the introduction of Best Available Techniques in industry (DAF-BAT) (Final Project Report Executive Summary), 39p.

Nicholas, M. J., Clift, R., Azapagic, A., Walker, F. C., Porter, D. E., 2000. Determination of "Best Available Techniques" for Integrated Pollution Prevention and Control: A Life Cycle Approach. Process Safety and Environmental Protection 78, 193-203. doi:10.1205/095758200530637.

Polders, C., Van den Abeele, L., Derden, A., Huybrechts, D., 2012. Methodology for determining emission levels associated with the best available techniques for industrial waste water. Journal of Cleaner Production 29–30, 113–121. doi:10.1016/j.jclepro.2012.02.008.

Saporta, G., 2011. Probabilités, analyse des données et statistique [Probability, data analysis and statistics]. Editions TECHNIP, 656p.

Journal of Cleaner Production 121 (2016) 72–83

Contents lists available at ScienceDirect

Journal of Cleaner Production

journal homepage: www.elsevier.com/locate/jclepro

Review

Best Available Technique assessment methods: a literature review from sector to installation level

Damien Evrard [a, *], Valérie Laforest [a], Jonathan Villot [a], Rodolphe Gaucher [b]

[a] UMR CNRS 5600 EVS, Mines Saint-Étienne, 158 cours Fauriel, 42023 Saint-Etienne cedex 02, France
[b] INERIS, Parc Technologique ALATA BP 2, 60550 Verneuil-en-Halatte, France

ARTICLE INFO

Article history:
Received 12 May 2015
Received in revised form
27 January 2016
Accepted 30 January 2016
Available online 12 February 2016

Keywords:
Best Available Technique
BAT selection
BAT assessment
Performance
IED
Industrial emission directive

ABSTRACT

The Industrial Emission Directive (IED) has strengthened the requirement to apply Best Available Techniques (BATs) previously introduced by the IPPC (Integrated Pollution Protection and Control) Directive. BATs in the European Union are defined and revised through an official framework: the Sevilla Process. It describes a step-by-step procedure of information exchange among stakeholders supported by data analysis and expert judgements. However, it lacks formal tools, applicable at each step, to support data analysis and the decision process leading to the determination of BATs and BAT-Associated Environmental Performance Levels (BATAPELs). This article is aimed at reviewing the methodological approaches available in the scientific literature since the late 90s for regional or local applications.

This literature review begins with a presentation of the legal context in which the BATs are used, especially under the IED. Then, the methodology used to establish the literature review is detailed; followed by an analysis of the different methods. The outcome is an identification of their objectives, methodological approaches and limits in the case of BAT determination for an industrial sector and the comparison of site performance with identified BATs with a view to selecting techniques adapted to a given installation. From the literature review, seven steps to determine and apply BATs are deduced: (1) Scope and objective definitions; (2) Knowledge of the sector studied to identify key performance indicators; (3) Identification of representative installations; (4) Data collection; (5) Selection of a list of BATs and determination of BATAELs; (6) Site performance analysis and comparison with BATs; (7) Selection of the most appropriate BATs. Finally, this proposition is discussed and perspectives are identified.
© 2016 Elsevier Ltd. All rights reserved.

Contents

1. Introduction 73
2. Legal context 74
 2.1. Actors involved in the Sevilla Process 75
 2.2. Steps of the Sevilla Process 75
  2.2.1. Step 1: reactivation of the TWG and call for wishes 75
  2.2.2. Step 2: kick-off meeting 75
  2.2.3. Step 3: data collection 75
  2.2.4. Step 4: drafts 75
  2.2.5. Step 5: final meeting 75
  2.2.6. Step 6: final draft 75
 2.3. Conclusions on the legal context 75

Acronyms: AHP, Analytical Hierarchy Process; ANP, Analytical Network Process; BAT, Best Available Technique; BATAEL, Best Available Technique Associated Emission Level; BATAEPL, Best Available Technique Associated Environmental Performance Level; BREF, BAT Reference document; EIPPCB, European Integrated Pollution Prevention and Control Bureau; EU, European Union; IED, Industrial Emission Directive; IPPC, Integrated Pollution Prevention and Control; L-BAT, Local-BAT; LCA, Life Cycle Assessment; NGO, Non-Governmental Organisation; TWG, Technical Working Group; VITO, Flemish Institute for Technological Research.

* Corresponding author. Tel.: +33 (0)477426671.
E-mail address: damien.evrard@mines-stetienne.fr (D. Evrard).

http://dx.doi.org/10.1016/j.jclepro.2016.01.096
0959-6526/© 2016 Elsevier Ltd. All rights reserved.

3. Search method for the literature review 76
4. Results of the literature review 77
4.1. General aspects 77
4.2. BAT-related methods at industrial sector level 77
4.2.1. Methods for BAT determination 77
4.2.2. Methods for BAT comparison 78
4.3. BAT-related methods at installation level 79
4.3.1. Performance comparison between installation and BATs 79
4.3.2. Technique selection for an installation 79
4.3.3. Performance evaluation according to local conditions 79
4.4. From sectorial to installation level: a step-by-step approach 79
4.4.1. Step 1: scope and objective definition 79
4.4.2. Step 2: knowledge of the sector studied 80
4.4.3. Step 3: identification of representative installations 80
4.4.4. Step 4: data collection 80
4.4.5. Step 5: selection of BATs and determination of BATAELs 80
4.4.6. Step 6: site performance analysis and comparison with BATs 80
4.4.7. Step 7: selection of the most appropriate BATs 80
4.5. Conclusion on the methods 81
5. Discussion: towards an adaptive approach to help decision-makers 81
5.1. Issues to consider for an integrated approach: criteria selection 82
5.2. Issues to consider for an integrated approach: selection of representative installations 82
6. Conclusion 82
Acknowledgements 83
References 83

1. Introduction

In 1996, the Integrated Pollution Prevention and Control (IPPC) directive (Council Directive, 1996) introduced the concept of Best Available Technique (BAT) to the European environmental regulation. Since then, applying BATs has become of great importance for the industries concerned and has implied obligations at two levels. On the one hand, the European Commission has had to set out the appropriate process in order to provide sector-specific reference documents describing the BATs. On the other hand, local environmental authorities and operators have had to compare local installation performances with the information contained in these reference documents.

In 2010, the IPPC directive was replaced by the Industrial Emission Directive (IED) (European Commission, 2010) which has extended and reinforced the role of BATs. The industrial sectors concerned with the implementation of BATs are listed in its annex I. This represents about 50,000 installations within in Europe from various industrial sectors (e.g. food, drink and milk; wood-based panels; large combustion plants; or sanitary landfills). The overall goal of the IED is "*to prevent, reduce and as far as possible eliminate pollution arising from industrial activities in compliance with the 'polluter pays' principle and the principle of pollution prevention*". Furthermore, BATs also concern non-IED industries which do not even have the Sevilla Process as a framework; e.g. in the nuclear sector (OSPAR, 2015).

Article 3 of the IED defines "Best Available Technique" as "*the most effective and advanced stage in the development of activities and their methods of operation which indicates the practical suitability of particular techniques for providing the basis for emission limit values and other permit conditions designed to prevent and, where that is not practicable, to reduce emissions and the impact on the environment as a whole*" (European Commission, 2010).

The meaning of each of these three words is important to understand the concept of BAT:

- "Technique" includes both the production "*technology used and the way in which the installation is designed, built, maintained, operated and decommissioned*" (European Commission, 2010). In other words, a technique is not limited to a pollution abatement device but also encompasses management approaches such as environmental management systems.
- "Availability" means the technique considered is "*developed on a scale which allows an implementation in the industrial sector, under economically and technically viable conditions*" (European Commission, 2010). These conditions take into consideration

Table 1
The 12 criteria for selecting Best Available Techniques as stated in annex III of the Industrial Emission Directive.

1	The use of low-waste technology.
2	The use of less hazardous substances.
3	The furthering of recovery and recycling of substances generated and used in the process and of waste, where appropriate.
4	Comparable processes, facilities or methods of operation which have been tried with success on an industrial scale.
5	Technological advances and changes in scientific knowledge and understanding.
6	The nature, effects and volume of the emission concerned.
7	The commissioning dates for new or existing installations.
8	The length of time needed to introduce the Best Available Technique.
9	The consumption and nature of raw materials (including water) used in the process and energy efficiency.
10	The need to prevent or reduce to a minimum the overall impact of the emissions on the environment and the risks to it.
11	The need to prevent accidents and to minimise the consequences for the environment.
12	Information published by public international organisations.

the costs and advantages, whether or not the technique is used or produced inside a given Member State, and if it is reasonably accessible to the operator.
- "Best" means that the technique considered is the "*most effective for achieving a high general level of protection of the environment as a whole*" (European Commission, 2010).

Furthermore, BATs are characterised by twelve criteria given in annex III of the IED (Table 1). However, using these criteria to select or to assess BATs is not easy. According to Laforest (2008, 2014), the phrasing used mixes objectives, criteria, indicators and parameters and makes the qualities required by a candidate technique unclear.

Companies within the scope of the IED can use reference documents called BREFs (EIPPCB, 2015) to identify BATs and their associated emission levels. The creation and revision process of these BREFs, the Sevilla Process, is detailed in an implementing decision (European Commission, 2012). Since its creation, it has been evolving so as to reach a three-year revision process and to improve the accuracy of the BREF. This process is a response to the requirement stated in article 13 of the IED to set up an information exchange while creating or revising the BREFs. It is coordinated by the European IPPC Bureau (EIPPCB) based at the Joint Research Centre in Seville (Spain). Although it defines how the information is collected and shared among stakeholders (European Commission, 2012), the only formal tools to help in the determination of BATs concern the issue of cross-media effects and economic performance (European Commission, 2006). Thus, to increase the robustness of the technique performance assessment to determine BATs, several academic methods have been developed since the late '90s (e.g. Dijkmans, 2000; Nicholas et al., 2000; Halog et al., 2001). The existing methods address three issues: (1) the local application of the IPPC directive or the IED for operational permits; (2) the determination of BATs at the industrial sector scale; (3) the determination of emission levels associated with BATs (BATAELs). Nowadays, most of the industrial sectors concerned which have to comply with the IED possess a draft or a final version of a BREF (31 published BREFs and 2 on-going) (EIPPCB, 2015). In other sectors like in the nuclear industry (OSPAR, 2015), operators must prove that they apply BATs but have no reference documents and therefore must find their own references to assess the performances of their installations. The Sevilla Process is therefore the framework for the definition of BATs at European level in the context of the IED. The process, described in the following section, is based on information exchange and discussion among several stakeholders. In addition to this framework, tools are needed at each of its steps to guide and strengthen a decision-making process to determine BATs. Furthermore, since the end users of BATs are industrial operators, the application of BATs would also need tools.

This article is aimed at reviewing in the scientific literature, the existing methods which can contribute to determine BATs for a given industrial sector or to apply BATs at installation level, including the comparison of site and BAT performance and selection of the most appropriate techniques for a given installation. After a presentation of the legal context and the Sevilla Process, the search method for the literature review is presented. Thanks to this review, general conclusions on seven essential steps towards a possible integrated approach considering both the sectoral and installation levels are drawn. Finally, perspectives for further research are proposed.

2. Legal context

In the IED, sectoral reference documents describe BATs and their associated performances. These BREFs are divided into six main chapters (European Commission, 2012): (1) General information about the sector concerned; (2) Applied processes and techniques; (3) Current emission and consumption levels; (4) Techniques to consider in the determination of BATs; (5) BAT conclusions; (6) Emerging techniques. The key chapter for operators and environmental authorities is the "BAT Conclusions" which is published in the Official Journal of the European Union and contains a list of BATs and BATAELs. The BATs described are neither exhaustive nor prescriptive but provide support to assess the performance of

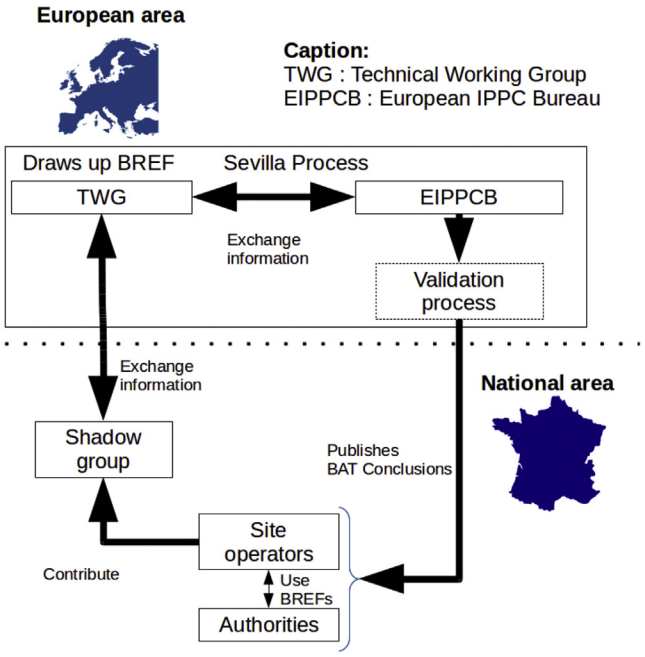


Fig. 1. Exchange between national and European actors in the Industrial Emission Directive.

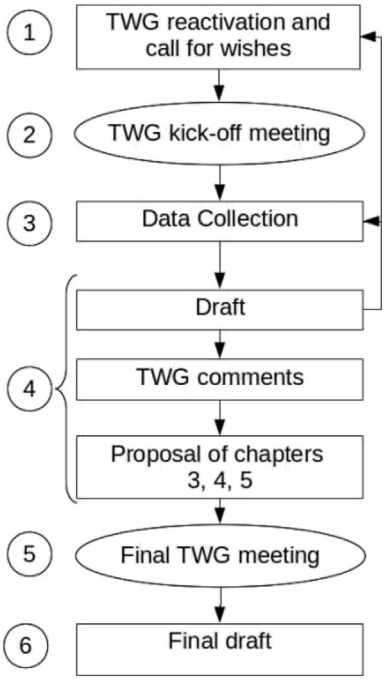


Fig. 2. Steps of the Sevilla Process during the revision of a BREF, adapted from the European Commission (2012).

installations. BATAELs are included in the wider concept of BAT-Associated Environmental Performance Level (BATAEPL) which also encompasses consumption, emissions during "other than normal operating conditions" and any other relevant performance indicators such as abatement efficiency or process controls (European Commission, 2012, section 3.3).

The creation process of these BREFs is defined in an official framework named the Sevilla Process. Decisions made during the process are based on a consensus among Member States, the European Commission, the Industry, and environmental NGOs gathered in a Technical Working Group (TWG) (European Commission, 2012, section 4.4.1). Information exchange among actors is essential to reach a consensus and to produce a document which will be applicable and will contribute to reducing industrial pollution in the EU. Therefore, BATs encompass two areas: a European area with the Sevilla Process and a national area with the contribution of national stakeholders within a shadow group and local application of BATs by operators and environmental authorities.

2.1. Actors involved in the Sevilla Process

The information exchange in the TWG is coordinated by the European IPPC Bureau which also manages the whole revision of the BREFs, collects and verifies data and performs an independent analysis to define BATs. The experts in the TWG ensure transmission of relevant information to the EIPPCB, comments of the proposals by the EIPPCB and participation in the TWG meetings in order to determine BATs. They are nominated by their respective organisation on the basis of their technical, economic and environmental expertise or their knowledge of the regulation (European Commission, 2012).

The link between the TWG and national stakeholders (site operators and authorities) can be established by shadow groups, i.e. national groups optionally set up by Member States to discuss the national position, how to support it and to prepare the TWG meetings. In France, they are usually composed of representatives of the Ministry of Environment and the Industry, as well as technical experts (Fig. 1).

A validation process is set out by article 13 of the IED (European Commission, 2010, article 13). Since this validation does not involve technical discussions but legal aspects, it is therefore not within the scope of this research project.

2.2. Steps of the Sevilla Process

The TWG works along the six steps defined in the Sevilla Process with an objective to finish within two years (Laforest and Gaucher, 2015). Since the Sevilla Process is an iterative process, the TWG can repeat each step if necessary. The main steps are pictured in appendix 2 of the implementing decision (European Commission, 2012). In the case of a BREF revision, the first step is the reactivation of the TWG and a call for wishes for the coming revision (Fig. 2).

2.2.1. Step 1: reactivation of the TWG and call for wishes

This first step aims at collecting the TWG members' wishes for the future documents: its scope; environmental indicators (consumption or emission) to be targeted as "key environmental issues"; structure of the BREF, changes required in the BAT chapter including BATAELs, and nature of data to collect (type and format).

2.2.2. Step 2: kick-off meeting

A first meeting aims at defining the agenda for the revision on the basis of the wishes sent at the end of the previous step. The issues to be investigated and the data to be collected by questionnaires are also defined (European Commission, 2012, sections 4.6.2.2 and 5.5). The outcome is a questionnaire template which will be used to collect data within existing installations. The items of the questionnaire are bound to become criteria for the assessment of techniques.

2.2.3. Step 3: data collection

The questionnaire is then sent to installations selected by the TWG members and approved by their Member State. This collection includes their consumption and emissions, wider performance indications, techniques used, production and economic elements, and information about the local context in which a technique is operated. Besides, technical and economic data are gathered to ponder the feasibility of techniques. This step is essential to define applicable BATAELs and to give consistency to the choice of BATs derived from real installation performances.

For sectors with a large number of installations (e.g. over 2000 in the Food, Drink and Milk sector), a selection of a few "representative" installations is made (European Commission, 2012). The choice of installations reported in a questionnaire and the validation of the data sent to the EIPPCB differ among Member States.

2.2.4. Step 4: drafts

The EIPPCB uses the information collected to write a draft BREF. Discussion for approval of this document by the TWG members mainly concerns the applicability of a technique, its viability and its efficiency, based on technical and economic information. If a consensus among experts is reached on the matter of the environmental benefit of a technique and its overall performance level, it is considered as a BAT.

2.2.5. Step 5: final meeting

This final discussion focuses on the "BAT conclusions", i.e. the description of the BATs and BATAELs. During the meeting, every proposed BAT is reviewed until a consensus is obtained based on the emissions or consumption of the sites covered by questionnaires. Each BATAEL is determined in accordance with the performances of real installations. Members of the TWG disagreeing with the consensus may express a "split-view". If accepted by the EIPPCB, it is recorded in the concluding remarks in the BREF.

2.2.6. Step 6: final draft

This final version of the BREF includes the relevant information from the final meeting. It is sent to the TWG members for validation. The technical work ends here and the validation by the Member States and the European Commission begins.

2.3. Conclusions on the legal context

The Sevilla Process is the official framework for the determination of BATs at EU level. It is highly dependent on the expert judgement of the TWG members and their ability to reach a consensus. Besides, discrepancies in the transmission of information may occur due to the lack of tools to strengthen expert judgement and provide explicit arguments to their decisions at every step of the process. This problem is also due to differences in objectives and methods among members of the TWG. The extensive amount of information to be analysed implies an increasingly complex data processing phase. The transition from this step of information analysis to the definition of BATs relies highly on the quality of the information provided by its actors with the risk of a biased assessment due to differences in interpretation of the characteristics of installations to be selected.

Table 2
Number of search results with the keywords used in the different databases in May 2015.

Keyword associated with IPPC or IED and "Best Available Technique" or BAT and the keywords	Science Direct	Web of Science	Google scholar	Research Gate	Total
Cross media effect	0	4	4	0	8
Decision support	10	6	51	0	67
Economic assessment	15	3	28	0	46
Eco-efficiency	9	4	48	0	61
Environmental assessment	17	17	120	0	154
Analysis	8	19	18	12	57
Determination	9	9	34	4	56
Assessment	15	19	58	6	98
Performance	8	17	52	56	133
Total	**91**	**98**	**413**	**78**	**680**

At a more local level, the implementation of the directive in the different Member States depends on the transposition into national regulations. Thus, at national level, environmental authorities can use BAT conclusions to determine emission limit values in their regulation. They must also provide general administrative provisions for the installations and are liable towards the European Commission for the proper implementation of the IED (European Commission, 2010). For each installation concerned, operators must periodically demonstrate that the level of performance of the techniques they use is at least as good as the BATAELs. The launch of this reconsideration depends on the publication of new or updated relevant "BAT conclusions" for their main activity. Thus, after their publication or update, the installations concerned must comply with them within four years (European Commission, 2010). In cases where a technique is not considered in a BREF and no BATAEL has been defined on the environmental issues related to this technique, the operator would have to justify the installation performances.

Therefore, the legal context shows that the European and national areas of application of the BATs are closely interwoven and that a large amount of information has to be studied to cover all aspects of the BATs. The issue of BATs, whether their determination or application, has existed since the publication of the IPPC directive, therefore research to develop methods to identify them within a sector or a geographical area already exists but does not seem to be used in the Sevilla Process. Besides, on-site comparison of performance with BATs and the identification of techniques or combination of techniques applicable to reach BAT performances may be found in the literature.

3. Search method for the literature review

This literature review was carried out by searching methods related to BAT in international journals indexed in recognised databases and accessible from Mines Saint-Étienne. Moreover, searches were made in English and in French and additional information for the understanding of the articles was asked to their authors when necessary. The four databases used were: ScienceDirect, Web of Science, Google scholar and Research Gate. Therefore, a search using the terms "Best Available Technique" (or "BAT") associated with "IPPC" or "IED" was carried out in association with one of the following keywords: "analysis"; "determination"; "assessment" and "performance". In addition, other keywords related to more specific aspects of BATs were included: "cross media effect"; "economic assessment"; "eco-efficiency"; "environmental assessment" and "decision-support method".

The initial search by association of keywords in the databases provided a large number of results (680 references) (Table 2).

In order to narrow down the number of results and to ensure that only relevant results were found, the results of each database were cross-checked to exclude duplicates. At this stage, 320 articles remained. Then, a look at all their keywords and their title or abstract enabled us to exclude 285 articles which did not actually concern the directives or BATs. Table 3 summarises the criteria of exclusion and the number of articles concerned. Besides off-topic articles, the reasons for excluding an article were the focus on policy application and legal study; the study of a single technique applied in a given installation; a survey of situation in a sector; environmental, technical or economic study; the description of a new process; articles related to the IPPC or IED directives but not about BAT determination or application; the modelling and simulation of an installation using BATs.

At this stage, 35 articles remained. This number was small enough to read the remaining articles in detail and select only those which described a method, either aimed at determining BATs or assessing on-site technical performances. Two of them were about improvable flows and four about Life Cycle Assessment applications. Only one of each which best described an original method for technique assessment and selection, was kept (4 articles excluded).

Table 3
Number of off-topic articles excluded.

Topic	Articles concerned
Sports	47
About policy and legal study	43
Other topics (medicine, electronics, animal, building, nanomaterial, etc.)	70
Study of a single technique	34
Sectoral survey	22
Environmental assessment or environmental impact assessment of a site or technique	17
Monitoring or data collection	5
Waste management	16
Technical or economic assessment only	8
New processes	5
IPPC or IED-related but not about BAT assessment	15
Modelling and simulation of an installation using BATs	3
Total excluded	**285**

Table 4
Main aspects of the articles found.

Scope	Objectives	Assessment	Area of application	End user	Structure	References
Industrial sector level	Determination of sectoral BATs and BATAELs	Qualitative by experts	National	Policy-makers and decision-makers	4 steps	Dijkmans (2000); Derden et al. (2002); Derden and Huybrechts (2013)
					5 steps	Polders et al. (2012)
		Quantitative			4 steps	Geldermann and Rentz (2004)
			Regional		10 steps	National Observatory of Athens (2006); Mavrotas et al. (2007); Georgopoulou et al. (2008)
	Comparison of candidate BATs		Sectoral		4 steps	Nicholas et al. (2000)
					Not applicable	Halog et al. (2001)
	Adaptation of the IPPC to non-EU countries		European	Technology suppliers	Not applicable	Schollenberger et al. (2008)
Installation level	Performance comparison between installation and BATs	Quantitative	Installation	Operators and decision-makers	5 steps	Barros et al. (2009)
		Qualitative			4 steps	Cikankowitz and Laforest (2013)
	Technique selection for an installation	Qualitative		Operators	3 steps	Giner-Santonja et al. (2012)
		Quantitative			7 steps	Bréchet and Tulkens (2009)
					4 steps	Rodríguez et al. (2011)
				Operators	4 steps	Ibáñez-Forés et al. (2013)
		Qualitative and quantitative		Not specified	8 steps	Aloini et al. (2014)
					3 steps	Nguyen et al. (2014)
		Quantitative			2 steps	Van Caneghem et al. (2010)
	Performance evaluation according to local conditions	Qualitative and quantitative			3 steps	Kabak et al. (2014)
		Qualitative and quantitative		Operators and decision-makers	4 steps	Laforest (2014)
Both	Multi-criteria decision-making	Not applicable	Not applicable	Not applicable	Not applicable	Ibáñez-Forés et al. (2014)

Several articles may be about the same method but described a specific aspect or case study. In these cases, the description of the overall method was kept. Thus, seven more articles were excluded.

Finally, the number of results was reduced to 24. Specifically, three methods were found to answer the question of the determination of sectoral BATs (Dijkmans, 2000; Geldermann and Rentz, 2004; National Observatory of Athens, 2006). Two methods could be applied at installation level to compare installation and BAT performances (Barros et al., 2009; Cikankowitz and Laforest, 2013). The other methods concerned a specific aspect which could be used in a part of the process, mainly for assessment or decision support. Among these articles, a literature review (Ibáñez-Forés et al., 2014) analyses some of the articles analysed here but its goal is to present an overview of the assessment and selection methods for optimal technological alternatives from a sustainable perspective. This article was kept as a synthesis of possible approaches in a broader scope than IPPC or IED-related articles.

4. Results of the literature review

4.1. General aspects

The main characteristics of the methods found in the literature review are presented in Table 4. They differ in scope (industrial sector or installation level), area of application (European, national, regional, sectoral or installation) or objectives. Thus, the main distinction among the articles is whether they are applicable at industrial sector level with a qualitative or quantitative approach, or at installation level in order to assess and compare performances or to find the most appropriate technique in a given situation.

Considering the industrial sector level, most of these articles deal with a quantitative method except for Dijkmans (2000), Derden et al. (2002), Polders et al. (2012) and Derden and Huybrechts (2013), all from the same Flemish institute (Table 4). Moreover, two articles (Nicholas et al., 2000; Halog et al., 2001) concern the comparison of candidate BATs for an industrial sector without geographical considerations. Finally, Schollenberger et al. (2008) focus on commenting the Sevilla Process with a view to adapting BATs outside the EU.

At installation level, performance assessment (Kabak et al., 2014; Laforest, 2014), comparison (Barros et al., 2009; Cikankowitz and Laforest, 2013) or selection of BATs were considered (Bréchet and Tulkens, 2009; Rodríguez et al., 2011; Giner-Santonja et al., 2012; Bortolini et al., 2013; Ibáñez-Forés et al., 2013; Aloini et al., 2014; Nguyen et al., 2014). A regional method was also considered for its insight on the relations between environmental and economic performances (Van Caneghem et al., 2010). Most of these articles deal with quantitative assessment except for Cikankowitz and Laforest (2013).

4.2. BAT-related methods at industrial sector level

Eleven methods aim to determine BATs at industrial sector level (Table 4). The main difference among the articles is that they either aim to identify a list of BATs (possibly with BATAELs) from installations in their area of concern, or to compare BATs already identified (possibly to find the most important substances). They all concern decision or policy-makers except for Schollenberger et al. (2008) focussing on technology transfer.

4.2.1. Methods for BAT determination

Eight articles illustrate the determination of BATs. One of them presents a qualitative approach for BAT determination (Dijkmans, 2000) with two case studies (Derden et al., 2002; Derden and Huybrechts, 2013) and an extension to determine BATAELs (Polders et al., 2012). The other four are quantitative approaches: the first one by the National Observatory of Athens (2006) with methodological focuses in two other articles (Mavrotas et al., 2007; Georgopoulou et al., 2008); the second one by Geldermann and Rentz (2004).

4.2.1.1. Qualitative approach for BAT and BATAEL determination. The method Dijkmans (2000) developed aims to determine lists of BATs for Flanders. It is also being used by the BAT-Centre of the

Flemish Institute for Technological Research (VITO) in its work as a representative of Belgium in the Sevilla Process. In Flanders, BAT sectoral studies cover IED sectors like the food processing (Derden et al., 2002) or textile (Derden and Huybrechts, 2013) industries and non-IED industrial sectors in relation to the local regulation.

The method is divided into four main steps: (1) information collection from existing installations and studies to obtain candidate BATs; (2) candidate BATs are qualitatively evaluated; (3) BATs are translated into suggestions for permit conditions; (4) economic and ecological consequences of BAT-based permit conditions are pondered. The second step of evaluation consists of a succession of assessments and exclusion of unsatisfactory techniques. They are assessed according to their technical feasibility, environmental performance, and economic feasibility. The remaining techniques are compared and a final selection is made by a Sector Working Group composed of representatives of the Industry, public agencies and independent experts consulted at every step. The result of this method is a list of BATs applied in Flanders.

Then, to determine the emission limits associated with these BATs, the method developed by Polders et al. (2012) is used. BATAELs are determined by analysing the performances of the installations concerned. Unlike the Sevilla Process, where the selection of BATs is interwoven with the determination of BATAELs, this step was considered individually by VITO. For this second method, five steps are applied to emission data of installations from a given industrial sector. In both methods, the qualitative multi-criteria assessment mostly relies on expert judgement supported by local data. The main drawback of the BATAEL definition is that it presupposes that a list of BATs has already been drafted, whereas these two phases are simultaneous in the Sevilla Process. The consideration of other performance levels such as energy efficiency or safety is left to the determination of BATs with Dijkmans' approach. Finally, there is no analysis of the applicability of BATAELs, such as an indication of the percentage of sites which could not comply with the emission levels and the reason why.

In later approaches like in Geldermann and Rentz (2004) and the National Observatory of Athens (2006), quantitative approaches were used.

4.2.1.2. Quantitative approach for BAT determination. Several quantitative methods have been identified for BAT determination at sectoral level. The two methods presented as examples in this section have the characteristics to be based on pollutant approaches, for all sectors and their users are meant to be decision and policy-makers.

4.2.1.2.1. Geldermann and Rentz's example. This approach was developed in 2001 by Geldermann and Rentz (2004) as a support to identify BATs enabling all emission-relevant industrial sectors to satisfy the obligations of the Gothenburg Protocol concerning the emissions of nitrogen oxides and Volatile Organic Components, therefore a few substances are considered. The reference to the IPPC directive concerns cross-media considerations. Two case studies in the sector of paint application and glass manufacturing are partially presented.

The four steps of this method are (1) first screening, (2) mass and energy balance, (3) calculation of characteristic figures, (4) decision support. The reference installation approach is used within the first three steps of this method. Since all the installations cannot be assessed individually, they are grouped into categories and each category is represented by a reference installation. Thus, a reference installation is an installation whose characteristics "*can be considered as representative values for all concerned installations*" (Geldermann and Rentz, 2004). Therefore, all the installations represented by the same reference installation are supposed to be able to apply the same abatement options and to have similar applicability, efficiency and costs for a given emission reduction option. No information in the article indicates whether a reference installation is a real or fictional installation, whether it stands for an average, good-performing or the most challenging installation.

Although resorting to reference installations may reduce the number of installation categories to study and allow for the definition of cut-off rules, existing Life Cycle Assessment (LCA) studies are necessary to collect information about the ecological significance of the criteria. Besides, there is no indication of the representativeness of the reference installations in order to strengthen a rationale in favour of a combination of BATs. This approach relies on expert judgement to assess techniques and combinations of techniques. In order to consider as many alternatives as possible according to various environmental, technical and economic criteria, mathematical multi-criteria assessment methods were developed.

4.2.1.2.2. National Observatory of Athens' example. This Greek project aimed to create a Decision Aid Framework for the assessment of impacts from the introduction of BATs in industry. It is an economy-centred approach to determine BATs. This framework is said to be more specifically aimed at the "*evaluation of the possible economic and environmental costs and benefits implied by the application of BATs in a given region*" (National Observatory of Athens, 2006). It is based on a small number of "typical" installations per industrial sector or sub-sector, quoting in this aspect the work by Geldermann and Rentz (2004). The framework is divided into three main parts. The first part relates to the determination of potential BATs identified by experts with determination of "typical" installations and possible BATs. In the second part, these potential BATs are environmentally and economically assessed in order to determine costs and benefits. Finally, judging from this assessment, a decision is made to find out which ones can actually be considered as BATs and to define overall strategies for policy-makers.

In this method, a classification into sub-sectors is essential for identification of relevant parameters and to ensure that the small number of "typical" installations considered covers the entire sector. Two decision-support tools were developed: COMBAT (Mavrotas et al., 2007) to help decision-making through goal programming or interactive filtering and BEAsT for "*assessing the environmental benefits and the economic attractiveness of BATs*" (Georgopoulou et al., 2008). This approach implies that all the installations in a sub-sector are known and relatively homogeneous and that they can all be covered with a few "typical" installations. However, representativeness of the chosen reference installations may be an issue if significant heterogeneity is found in the sector studied. The number of "typical" installations is chosen by the user of the method and their representativeness is not assessed.

4.2.2. Methods for BAT comparison

In cases where a list of BATs has already been determined, three articles describe methods to compare alternatives. Thus, Nicholas et al. (2000) use LCA (ISO, 2006) to model processes and assess emission and impact reduction due to the application of BATs. Moreover, it considers technical and social dimensions to make a decision. The main drawback is the necessity to model a whole process, which, in the case of a sectoral study presupposes that some genericness can be found within the sector studied.

The second method found (Halog et al., 2001) also uses LCA with an adaptation of Quality Function Deployment principles to emission reduction. Thus, stakeholders' requirements are defined as impact categories and design requirements of a particular technique as substances that the process emitted and which need to be reduced. The contribution of the substances to impact categories expresses the link between them. Expert preferences appear in

terms of the weighting of impact categories. Moreover, emission targets were expressed through existing legal emission limit values. Finally, a ranking of the substances emitted, based on both cost and environmental impact potential, is carried out.

The last article concerns technology transfer of BATs outside the EU (Schollenberger et al., 2008). The authors have developed a metric based on a distance-to-target approach for the selection of BATs. This approach presents the advantage of merging the technological, economic and environmental goals through a multi-criteria decision-support tool. Besides, national targets are considered for each criterion to consider national specificities, which places this method at the interface between sectoral and installation level if local specificities were further explored.

4.3. BAT-related methods at installation level

At installation level, three main objectives have been identified in the literature (Table 4): (1) performance comparison between installation and BATs, (2) technique selection for an installation, (3) Performance evaluation according to local conditions.

4.3.1. Performance comparison between installation and BATs

Only two methods have been found on the issue of comparing an existing installation with BAT performances. The first one, by Barros et al. (2009) is an application on a specific sector. It consists of an inventory of the processes, material and energy flows; a characterisation of the related environmental impacts and a review of a list of applicable BATs by corporate experts to identify the BATs already in place. The last part of the method is a review of the techniques which could be added or cannot be applied to the installation studied. Therefore, this method also partially concerns the following section.

The second one, called L-BAT method (Local-BAT) developed by Cikankowitz and Laforest (2013) takes into account the 12 considerations of the IPPC directive, analogous to the 12 criteria of the IED for determining BATs, through an analysis grid based on the BREF. It is closer to the notion of BAT by assessing the level of control of the techniques, thus L-BAT aims to perform an environmental diagnosis of an installation compared with the BATs described in the BREF, local regulations (emission limit values) and BAT performances. Moreover, it considers the local context for the improvement of the technical performances of the installation.

These two methods illustrate how to compare on-site performance with BAT-associated environmental performance level, therefore verifying the application of BATs. They are both based on a review of the BREF to compare the level of performance, if BATs are not directly applied. The consideration of the local context is also treated considering the vulnerability of the close environment and the technical and economic capabilities of the operators. Barros et al. (2009) also decompose the production process into its main operations and flows (materials and energy) as a base for comparison.

4.3.2. Technique selection for an installation

When an installation does not reach the performance of the BATs, its operator must find techniques which would help to improve its environmental performance at an acceptable cost. On this issue, eight articles provide insight into how to select techniques appropriate for a given installation.

Rather than considering all emissions, Rodríguez et al. (2011) only target the ones which can be improved using a detailed material and energy flow analysis to identify them. In addition, environmental impacts caused by substances can be assessed thanks to an LCA (Ibáñez-Forés et al., 2013). Technical and economic criteria are also considered in several articles, either through expert judgement (Giner-Santonja et al., 2012) or social cost (Bréchet and Tulkens, 2009). Bortolini et al. (2013) propose an economic assessment considering the local context for the selection of solar panels. This approach could be applied to consider the local specificities of an installation but lacks an integrated approach.

The outcome of technique comparison is either a ranking or a score of alternatives based on a display of various criteria (Bréchet and Tulkens, 2009), or a multi-criteria decision approach (Aloini et al., 2014; Nguyen et al., 2014; Giner-Santonja et al., 2012) which requires the implication of multiple decision-makers and an assessment of their decisions and value.

Furthermore, like Laforest (2014), Giner-Santonja's method considers the 12 considerations of the directive as a preliminary for the determination of its assessment criteria. Then, the use of the Analytical Network Process (ANP) approach appears in other technique selection methods coupled with fuzzy logic (Kabak et al., 2014; Nguyen et al., 2014) in order to estimate expert judgement. Finally, techniques or combinations of techniques can be assessed thanks to eco-efficiency indicators (Van Caneghem et al., 2010), which presents the advantage of considering several criteria among environmental, technical and economic considerations simultaneously.

4.3.3. Performance evaluation according to local conditions

Two articles relate to the consideration of local conditions for a quantitative assessment. Kabak et al. (2014) to assess the building energy performance whereby a fuzzy multi-criteria decision-making method is used to ponder expert judgement on the performance of various local criteria such as location or lighting system and energy consumption. All the alternatives are compared pairwise through an ANP to obtain a single score.

Laforest (2014) takes into account local conditions for the quantification of the assessment indicators for different applications in wastewater treatment, soil remediation and sludge valorisation. Cikankowitz and Laforest (2013) qualitatively consider the local context to define local issues to be taken into account in the implementation planning of BATs.

4.4. From sectorial to installation level: a step-by-step approach

The legal context has showed that BATs are applied in two areas which are closely interrelated and that a large amount of information has to be studied to cover all aspects of BATs. The literature review has showed that methods do exist for sectoral BAT determination and application at installation level in the context of the IED. Moreover, BAT determination and alternative technique selection are issues for regulations other than the IPPC directive or the IED. Analysis of the methods showed that several steps are common or essential to determine and apply BATs: (1) scope and objective definitions, (2) knowledge of the sector studied, (3) identification of representative installations, (4) data collection, (5) selection of BATs and determination of BATAELs at sectoral level, (6) site performance analysis and comparison to BATs and (7) selection of the most appropriate BATs at installation level.

4.4.1. Step 1: scope and objective definition

BAT characteristics depend on the criteria selected. Therefore, the scope and objectives, including the environmental indicators to be used, need to be clear from the start. In the IED, they are “key environmental issues” identified in existing regulations or studies. Thus, a clear statement of the sector concerned and its installations, and the desired level of detail ensure that the study is relevant. This can be based on the LCA goal and scope definition like in Geldermann and Rentz (2004), or a decision from an expert group including the relevant stakeholders (European Commission, 2012).

Besides, the available data may differ whether a reference document on the possible BATs for the sector exists or not. This aspect is easier in the framework of the IED where an official process exists; it may be more difficult in cases where there is no reference and an installation operator has to define the relevant BATAEPLs, e.g. in the nuclear sector (OSPAR, 2015).

4.4.2. Step 2: knowledge of the sector studied

A common starting point in all the methods found is the identification of specific issues and characteristics of the industrial sector studied. In the case of the IED, the Sevilla Process requires for sectoral key environmental issues to be determined as early as this stage in order to target the data to be collected in step 3. According to the literature review, various national and international information sources can be used. This step should be closely related to step 1 in so far as new information might modify the scope and knowledge of the sector is a condition to the availability of a BAT.

4.4.3. Step 3: identification of representative installations

For this step, two approaches have been found. Firstly, extensive knowledge of all the installations concerned can be sought (Dijkmans, 2000). This might be feasible for sectors with few installations using relatively homogeneous techniques. In this case, similarity among installations may compensate for the missing data. Secondly, where detailed information may not be available for every installation concerned or there may be too much to obtain an exhaustive list, a selection of a few installations, according to sectoral key issues identified in step 2, could be used (Geldermann and Rentz, 2004; National Observatory of Athens, 2006; European Commission, 2012). Thus, in order to justify their relevance, these "representative" installations have to share common characteristics with the total population of the sector.

It seems that several approaches could be explored by either using a fictitious reference installation (Geldermann and Rentz, 2004) or a real representative installation (European Commission, 2012). In both cases, an explanation of their representativeness is necessary since experience shows that a common understanding of representativeness is not a simple issue.

This step raises the more general issue of the meaning of a "representative installation". In the Sevilla Process, they are "good-performers", i.e. installations which can claim to be considered as having the highest level of environmental performance (European Commission, 2010). The number of installations to analyse so as to be able to have a representative overview of the sector would also be crucial.

4.4.4. Step 4: data collection

The methods found in the literature present the sources and the nature of information they use but they do not detail how data is collected and processed. In the Sevilla Process, questionnaires are created by the TWG members in order to collect the necessary information. The transmission of these questionnaires is performed by the TWG members and a quality check is carried out by the Member States based on the regulatory data reported by operators. However, in cases outside the framework of the IED where operators have to define BATs on their own, in-house data and expert judgement might be the most detailed sources of information available (European Commission, 2012; Laforest, 2014).

4.4.5. Step 5: selection of BATs and determination of BATAELs

This review has showed that different approaches are used from expert judgement (Dijkmans, 2000) to site performance analysis or cost/benefit evaluation (National Observatory of Athens, 2006). The choice of the relevant approach would depend on the goal and scope. The data collected in the previous step would help to identify sites likely to use techniques which may be considered as BATs. Then, these "candidate BATs" are assessed according to their environmental and technical performances, and cost to ponder whether they can be considered as BATs or not. Numerous assessment approaches exist; they can be based on expert judgement (Dijkmans, 2000; European Commission, 2012), multi-criteria assessments such as LCA (Nicholas et al., 2000), or more economics-oriented approaches (National Observatory of Athens, 2006). Finally, the qualification of "candidate BATs" takes place. In the Sevilla Process the determination of BATs and BATAELs is based on discussions among the members of the TWG. Multi-criteria decision-making methods could be used to provide arguments at this step. Here again, the appropriate tool to be used may depend on the scope and objective definitions carried out in step 1.

Thus, the outcome of a BAT determination method could be a list of BATs with a description of each technique and its BATAEPLs. After obtaining this information, operators would be able to compare the performance of their own installation with BATs and, if necessary, identify the best combination of BATs they could set up to reach the BAT performance. This stage is described in the last two steps.

4.4.6. Step 6: site performance analysis and comparison with BATs

An operator needs to analyse the situation of his installation to have a view of the material and energy flows circulating within it, in order to describe the preventive and abatement measures used, and the level of compliance of the installation with the applicable regulations. Then, a comparison of this situation and the techniques in place, based on the definition of BATs and the 12 criteria, would help him assess the need for modification.

Besides, although BREFs are meant to be applicable for a whole sector, this step requires analysing BATs in the particular situation of the site. Therefore, the local context, i.e. any specific conditions (economic, technical or environmental), which may make a generic BAT irrelevant, need to be considered (Giner-Santonja et al., 2012; Cikankowitz and Laforest, 2013). Approaches based on Material and Energy Flow Analysis could be used (Rodríguez et al., 2011) with the drawback of not being simple and quick enough to be accepted by Industry.

This diagnosis of an installation then provides indications for the key points which need to be remedied. This helps to choose the BATs for the installation, whether they are described in an official reference document or in-house studies.

4.4.7. Step 7: selection of the most appropriate BATs

This step is necessary in two situations: (1) if an installation does not comply with BATAELs, and (2) when an operator is unable to prove that he uses techniques with performances similar to BATs. In these cases, an operator needs a way to assess the technical alternatives which could help him reach BAT performances.

In this case, it is necessary to simulate the outcome of alternative techniques or combinations of techniques to find out whether they may help to reach BAT performances. Here, the problem is slightly different to the one answered in step 5 since the objective is to prove that current performances are similar to the BATs. Considering that a list of generic BATs was determined by the regulators or the operator at step 5, the issue is to test different scenarios under specific conditions and to select the most relevant one. This step seems to be feasible through expert judgement in most cases. Nevertheless, in cases where the choice of alternative is complex or very specific local issues are identified, multi-criteria decision-support methods, like in Giner-Santonja et al. (2012), may be necessary to assist operators both in their choice of alternatives and in the determination of justification to be provided to environmental authorities.

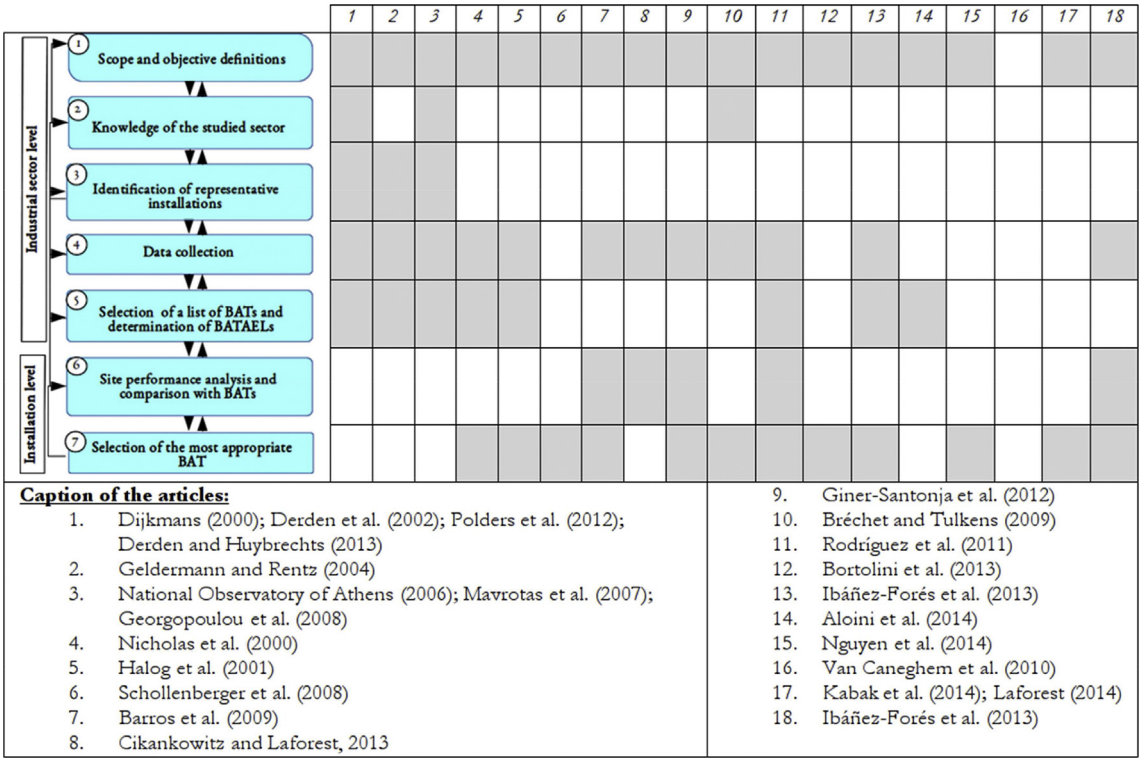


Fig. 3. Aggregated steps for the determination and application of Best Available Techniques.

4.5. Conclusion on the methods

Considering the sectoral level and in the framework of the Sevilla Process, Dijkmans (2000) mainly focuses on emissions with consideration of environmental, technical and economic criteria, whereas Geldermann and Rentz (2004) target environmental impacts, and the National Observatory of Athens (2006) aims at the economic valuation of impacts. Beyond these differences, the three methods described herein present interesting elements that should be kept in future work and limitations that may be subject to further research. Other approaches can be used to assess candidate techniques. Thus, Halog et al. (2001) are interested in the assessment and comparison of techniques by a mathematical model, whereas Nicholas et al. (2000) and Halog et al. (2001) use a Life Cycle Assessment approach. Moreover, in Halog et al. (2001), a preliminary step of determination of key impacts is first carried out and then the issues for BAT determination are considered as stakeholder requirements. Therefore, these two methods illustrate the issue of the choice of criteria to select BATs.

The determination of BATAELs is considered by Polders et al. (2012) after BATs have been determined considering individual emissions. The approach developed by Schollenberger et al. (2008) classifies emissions according to multiple criteria which could be interesting to search for combinations of applicable BATAELs and not only on the best level for every single criterion.

At installation level, three aspects were found. Firstly, performance comparison with BATs is required to determine whether the following steps are necessary or not. There is a method (Cikankowitz and Laforest, 2013) to verify the application of BATs provided by the BREF and BAT performance considering local context and its sensitivity. In order to quantify the performances of BAT, LCA could be used like in Barros et al. (2009) to model energy, material flows and environmental impacts. Secondly, the selection of techniques through multi-criteria decision-making has been studied more extensively and several methods do exist which have been applied. However, they are all dependent on the choice of criteria and their weighting made by a group of experts. Therefore, the definition of the role of experts seems to be essential. Finally, the consideration of local conditions for application at installation level is important in considering ways to apply or discard BATs determined at a larger scale (for the whole European Union in the case of the IED).

After crosschecking the seven steps deduced from the review and the steps presented in the articles, the first observation is that none of the articles found covers all the steps deduced (Fig. 3). Besides, because they were not developed for the same purpose and context, they cannot be used complementary to each other in a whole case study. Therefore, an integrated approach, considering both levels described in Fig. 1 remains to be developed.

The variety of tools found in the literature tends to show that not one but a set of tools could be applied at every step depending on the complexity of the issues considered. In the case where sectoral BATs are sought, the first five deduced steps would be sufficient. Where a justification of performance compared to BATs without any reference document is sought, a simplified but somewhat adapted version of these five steps could probably be used to provide a list of BATs to be used in steps 6 and 7. Such an integrated approach provides perspectives for further research presented in section 5.

5. Discussion: towards an adaptive approach to help decision-makers

As stated in the beginning of this article, the actors concerned with the determination of BATs or their application could either be the members of a shadow group in the Sevilla Process, operators of IED or non-IED installations, or local environmental authorities. For IED installations, the actors' goal is to provide information to the

Sevilla Process on national installations enabling identification of BATs and the associated performance level. For non-IED installations, operators would have to carry out the same task without the guidance of a formal process. In both cases, operators would have to justify to environmental authorities that their installation complies with the BATAEPLs. Because of this variety of situations, different methodological approaches would have to be considered to select or apply BATs at different levels and for different purposes. Therefore, several tools adapted to the level of detail and applicability conditions of each study would have to be identified and their conditions of applicability defined.

The main issue would be to identify the appropriate means for every possible situation and to take into account all aspects borne within the concept of BAT. Judging from this literature review, such a progressive approach would concern the choice of installations considered to determine BATs. Moreover, the multi-criteria assessment methods used would range from expert judgement (Dijkmans, 2000; Polders et al., 2012; Giner-Santonja et al., 2012; European Commission, 2012) to detailed material and energy flow analysis with or without impact assessment (Rodríguez et al., 2011; Ibáñez-Forés et al., 2013). This evaluation would have to consider a combination of environmental, economic and technical performances in order to help stakeholders make a decision. Two main issues need to be addressed first: the selection of the criteria used in multi-dimensional assessments and the selection of representative installations thanks to these criteria.

5.1. Issues to consider for an integrated approach: criteria selection

The selection of relevant criteria concerns both the determination of BATs and the assessment of performance on site. Indeed, BATs encompass environmental, economic and technical aspects which would have to be considered as accurately as possible. The choice of criteria raises the question of how to define the performance of an installation or a technique. Indeed, BATs need to be both “best” and “available” and, beyond the issue of collecting and processing of consumption and emission data, the question of the determination of the relevant information which contributes to characterising the performance still needs to be answered. Therefore, the question here is how to identify all the criteria which would describe the BATAEPL.

5.2. Issues to consider for an integrated approach: selection of representative installations

When the number of installations to be considered prevents from taking into account all of them or when a sector is too heterogeneous to judge which installations are representative, a guided selection of a small number of them is necessary. Two approaches were found in this case: (1) the use of fictitious reference installations which bear the characteristics of the installations of a sub-group (Geldermann and Rentz, 2004; Mavrotas et al., 2007); (2) the use of real representative installations of the sector (European Commission, 2012). The best approach depends on the objectives of the study. Indeed, in the case where the goal is to help a Member State prepare its contribution to the Sevilla Process, i.e. to identify installations with which it can argue on a specific issue, resorting to fictitious installations is not relevant. Nevertheless, fictitious reference installations could be used when quick screening of potential BATs is required. Therefore, a distinction is needed between fictitious and real installations used to represent a large number of installations. However, both approaches might be relevant considering the two levels of concern in this study and both should be strengthened by work on the assessment of the representativeness according to multiple criteria.

6. Conclusion

This literature review about Best Available Technique assessment methods has shown that methods exist at two levels: (1) to determine BATs at sectoral level, (2) to apply BATs at installation level. These methods were developed in different legal frameworks, specific contexts, different decision processes and have various objectives. Nevertheless, none of them offers a method covering both levels of assessment and is designed to be integrated or to be a support to the existing overall decision process.

An important aspect is the resort to reference installations whose definition may vary according to the objectives of the stakeholders. Therefore, its possible meanings should be specified before any methodological development. Specifically, in the framework of the Industrial Emissions European Directive, which aims at improving the environmental performance of Industry with the definition and the implementation of BATs, one might want to select the “best” installations to define BATs, whereas others might consider that installations complying with current regulations (including current BATs) is a necessary and a sufficient condition to be selected as “good candidate” for the BAT assessment process. In this second case, the reduction of the number of installations required to define BATs by grouping them according to common characteristics and then selecting the best representatives would be a challenge to have a more robust, transparent and efficient process. Therefore, clarification of the concepts of representativeness and consequently of BAT performances should be made.

From this analysis of existing methods, it seems to be possible to define seven steps for an integrated approach taking both sector and installation levels into account. At sector level, BATs and their associated environmental performances should be analysed independently of their legal framework and the targeted industrial sector. Under the IED, the Sevilla process proposes a framework to collect relevant information and could be used as a basis for further methodological developments. For non-IED sectors, data availability may be a problem due to a lack of an existing supranational framework and the absence of reference documents at European level. Nevertheless, a formal method similar to the one used in the context of the IED could be applicable within the limits of data availability. At installation level, the issue is to compare the performances of a given industrial installation with identified sectoral BATs and BATAELs.

However, several limitations of the existing approaches were identified in terms of criteria selection and choice of reference installations containing the necessary information. Thus, their objectives are economics or a limited number of substances as environmental focus. Therefore, the focus of future research should be on the formal selection of relevant indicators, in accordance with the definition of BAT in the legal framework studied (e.g. the environment as a whole in addition to technical and economic feasibility, like stated in the IED). The decision process may also be different in that it may involve public or private actors and concern different areas of application. This implies a need to consider the integration of the method into existing decision processes. Furthermore, several scientific databases could not be accessed by the authors (e.g. Scopus) and existing methods might have been ignored. Nevertheless, judging from the references used in the available articles and the direct contact with authors, the risk of having missed existing researches should be low.

Finally, future developments should concern the definition of a comprehensive method to support and follow expert decision at each level and step and focus on environmental performances while being able to be integrated to existing decision processes. Considering these facts, a formal objective method could be developed. It should be, applicable to any industrial sectors, and

should help to select representative installations through a multivariate approach, thus strengthening existing selection processes by reducing uncertainty of results, variability of decisions, and difficulty to compare performances. In such a multicriteria approach, selection and assessment of sites and techniques would imply consideration of various environmental indicators with techno-economic criteria. Moreover, it should be possible to find an integrated approach taking into account various parameters related to the concept of BAT as defined in the IED and the twelve criteria from annex III of the IED. This approach could also be adapted to non-IED sectors. After the definition and selection of relevant assessment criteria, multicriteria assessment associated with a decision support tool would enable analysing an industrial sector from a BAT perspective. The determination of reference installations from these criteria and the way to use them for the selection of a small number of installations could be a preliminary to have a reduced and comprehensive set of data for the determination of BATs. Data collected from operating installations and an explicit methodology to process them through statistical analysis could be used at this stage for site and technique selection.

Acknowledgements

Our thanks to the authors cited in this literature review who answered our questions, in particular VITO and EIPPCB.

References

Aloini, D., Dulmin, R., Mininno, V., 2014. A peer IF-TOPSIS based decision support system for packaging machine selection. Expert Syst. Appl. 41 (5), 2157–2165.

Barros, M., Magán, A., Valiño, S., Bello, P., Casares, J., Blanco, J., 2009. Identification of best available techniques in the seafood industry: a case study. J. Clean. Prod. 17 (3), 391–399.

Bréchet, T., Tulkens, H., 2009. Beyond BAT: selecting optimal combinations of available techniques, with an example from the limestone industry. J. Environ. Manag. 90 (5), 1790–1801.

Cikankowitz, A., Laforest, V., 2013. Using BAT performance as an evaluation method of techniques. J. Clean. Prod. 42, 141–158.

Council directive 96/61/EC of 24 September 1996 Concerning integrated pollution prevention and control, European Commission Publication Paper, L257/26.

Derden, A., Huybrechts, D., 2013. Brominated flame retardants in textile wastewater: reducing deca-BDE using best available techniques. J. Clean. Prod. 53, 167–175.

Derden, A., Vercaemst, P., Dijkmans, R., 2002. Best available techniques (BAT) for the fruit and vegetable processing industry. Resour. Conserv. Recycl. 34 (4), 261–271.

Dijkmans, R., 2000. Methodology for selection of best available techniques (BAT) at the sector level. J. Clean. Prod. 8 (1), 11–21.

EIPPCB, 2015. The European IPPC Bureau. Available from: http://eippcb.jrc.ec.europa.eu [19 June 2015].

European Commission, 2006. Reference Document on Economics and Cross-media Effects. European Commission, DG JRC, Sevilla, 159 p.

European Commission, 2010. Directive 2010/75/EU of the European parliament and of the council of 24 November 2010 on industrial emissions (integrated pollution prevention and control). Off. J. Eur. Union 2010, L334 17–118.

European Commission, 2012. Commission implementing decision of 10 February 2012 laying down rules concerning guidance on the collection of data and on the drawing up of BAT reference documents and on their quality assurance referred to in directive 2010/75/EU of the European parliament and of the council on industrial emissions. Off. J. Eur. Union 2012, L63 1–39.

Geldermann, J., Rentz, O., 2004. The reference installation approach for the techno-economic assessment of emission abatement options and the determination of BAT according to the IPPC-directive. J. Clean. Prod. 12 (4), 389–402.

Georgopoulou, E., Hontou, V., Gakis, N., Sarafidis, Y., Mirasgedis, S., Lalas, D., Loukatos, A., Gargoulas, N., Mentzis, A., Economidis, D., Triantafilopoulos, T., Korizi, K., 2008. BEAsT: a decision-support tool for assessing the environmental benefits and the economic attractiveness of best available techniques in industry. J. Clean. Prod. 16 (3), 359–373.

Giner-Santonja, G., Aragonés-Beltráan, P., Niclós-Ferragut, J., 2012. The application of the analytic network process to the assessment of best available techniques. J. Clean. Prod. 25, 86–95.

Halog, A., Schultmann, F., Rentz, O., 2001. Using quality function deployment for technique selection for optimum environmental performance improvement. J. Clean. Prod. 9 (5), 387–394.

Ibáñez-Forés, V., Bovea, M., Azapagic, A., 2013. Assessing the sustainability of best available techniques (BAT): methodology and application in the ceramic tiles industry. J. Clean. Prod. 51, 162–176.

Ibáñez-Forés, V., Bovea, M.D., Pérez-Belis, V., 2014. A holistic review of applied methodologies for assessing and selecting the optimal technological alternative from a sustainability perspective. J. Clean. Prod. 70, 259–281.

ISO, 2006. ISO 14040:2006 Environmental Management Life Cycle Assessment Principles and Framework. ISO, Geneva, 23 p.

Kabak, M., Köse, E., Kirilmaz, O., Burmaoğlu, S., 2014. A fuzzy multicriteria decision making approach to assess building energy performance. Energy Build. 72, 382–389.

Laforest, V., 2008. Applying best available technologies in environmental management accounting: from the definition to an assessment method. In: Environmental Management Accounting for Cleaner Production, pp. 29–47, 5th EMAN book.

Laforest, V., 2014. Assessment of emerging and innovative techniques considering best available technique performances. Resour. Conserv. Recycl. 92, 11–24.

Laforest, V., Gaucher, R., 2015. Industrial Emisson Directive (IED) and Best Available Techniques. Directive sur les émissions industrielles (IED) et meilleures techniques disponibles. Techniques de l'Ingénieur. Tech. l'ingénieur, 2015, 15p. Available from: http://www.techniques-ingenieur.fr/base-documentaire/genie-industriel-th6/ingenierie-et-responsabilites-42598210/directive-sur-les-emissions-industrielles-ied-et-meilleures-techniques-disponibles-g4097/ [1 October 2005]. [only available in French].

Mavrotas, G., Georgopoulou, E., Mirasgedis, S., Sarafidis, Y., Lalas, D., Hontou, V., Gakis, N., 2007. An integrated approach for the selection of best available techniques (BAT) for the industries in the greater Athens area using multi-objective combinatorial optimization. Energy Econ. 29 (4), 953–973.

National Observatory of Athens, 2006. Decision Aid Framework for the Assessment of Impacts from the Introduction of Best Available Techniques in Industry (Final project report executive summary, National Observatory of Athens), 39p.

Nguyen, H.-T., Dawal, S.Z.M., Nukman, Y., Aoyama, H., 2014. A hybrid approach for fuzzy multi-attribute decision making in machine tool selection with consideration of the interactions of attributes. Expert Syst. Appl. 41 (6), 3078–3090.

Nicholas, M.J., Clift, R., Azapagic, A., Walker, F.C., Porter, D.E., 2000. Determination of best available techniques for integrated pollution prevention and control: a life cycle approach. Process Saf. Environ. Prot. 78 (3), 193–203.

OSPAR, 2015. Best Available Techniques (BAT) and Best Environmental Practice (BEP). Available from: http://www.ospar.org/about/principles/bat-bep [16 June 2015].

Polders, C., Van den Abeele, L., Derden, A., Huybrechts, D., 2012. Methodology for determining emission levels associated with the best available techniques for industrial waste water. J. Clean. Prod. 2930, 113–121.

Rodríguez, M.T., Andrade, L.C., Bugallo, P.B., Long, J.C., 2011. Combining LCT tools for the optimization of an industrial process: material and energy flow analysis and best available techniques. J. Hazard. Mater. 192 (3), 1705–1719.

Schollenberger, H., Treitz, M., Geldermann, J., 2008. Adapting the European approach of best available techniques: case studies from Chile and China. J. Clean. Prod. 16 (17), 1856–1864.

Van Caneghem, J., Block, C., Van Hooste, H., Vandecasteele, C., 2010. Eco-efficiency trends of the Flemish industry: decoupling of environmental impact from economic growth. J. Clean. Prod. 18 (14), 1349–1357.

Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 48 (2016) 520 – 525

www.elsevier.com/locate/procedia

23rd CIRP Conference on Life Cycle Engineering

Best Available Techniques as a sustainability tool in manufacturing: case study in the dairy sector

Damien Evrard [a*], Valérie Laforest [a], Jonathan Villot [a], Rodolphe Gaucher [b]

[a]*UMR CNRS 5600 EVS, Mines Saint-Étienne, 158 cours Fauriel, 42023 Saint-Etienne cedex 02, France*
[b]*INERIS, Parc Technologique ALATA BP 2, 60550 Verneuil-en-Halatte, France*

* Corresponding author. Tel.: +33 (0)477426671; fax: +33 (0)4 77 42 66 33. *E-mail address:* damien.evrard@mines-stetienne.fr

Abstract

Best Available Techniques encompass preventive and end-of-pipe solutions aimed to contribute to the sustainability of the European industry. They are determined by the official Sevilla Process based on extensive data collection and analysis, supporting formal negotiation steps. This article presents a statistical multicriteria method applied to the dairy sector to help determine reference sites likely to use BATs. This 5-step methodology is based on two classifications: representative or performant sites. Performant sites selected by the Pareto front analysis are better than representative sites. In the representative analysis, the size of installations seems to be inversely proportional to their environmental impacts.

© 2016 The Authors. Published by Elsevier B.V This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Peer-review under responsibility of the scientific committee of the 23rd CIRP Conference on Life Cycle Engineering

Keywords: Best Available Technique; statistical analysis; reference installations; dairy sector

1. Introduction

Best Available Techniques (BAT) were first introduced in 1996 by the Integrated Pollution Prevention and Control Directive [1]. Their role was then extended and strengthened by the Industrial Emission Directive in 2010 [2]. Moreover, they have become an essential tool of the European regulation for regulating industrial emissions. The industrial sectors within the scope of the directive encompass about 50,000 installations (e.g. food, drink and milk; wood-based panels; large combustion plants; or sanitary landfills).

The overall goal of the IED is "to prevent, reduce and as far as possible eliminate pollution arising from industrial activities in compliance with the 'polluter pays' principle and the principle of pollution prevention" [2]. Furthermore, the concept of "Best Available Technique" is defined in the directive as "the most effective and advanced stage in the development of activities and their methods of operation which indicates the practical suitability of particular techniques for providing the basis for emission limit values and other permit conditions designed to prevent and, where that is not practicable, to reduce emissions and the impact on the environment as a whole" [2].

Thus, in the concept of BAT:

- **"Technique"** encompasses both "the technology used and the way in which the installation is designed, built, maintained, operated and decommissioned" [2]. Therefore, it is not limited to a pollution abatement device but can also be a management approach such as an environmental management system.
- **"Available"** means the technique considered is "developed on a scale which allows an implementation in the industrial sector, under economically and technically viable conditions" [2]. These conditions take into consideration its costs and advantages, whether it is used or produced inside a given Member State or not, and if it is reasonably accessible to the operator.
- **"Best"** means that the technique considered is the "most effective for achieving a high general level of protection of the environment as a whole" [2].

2212-8271 © 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Peer-review under responsibility of the scientific committee of the 23rd CIRP Conference on Life Cycle Engineering
doi:10.1016/j.procir.2016.03.098

This concept has involved obligations at two levels. First, at European level, sector-specific reference document need to be drawn up. An official framework named the "Sevilla Process" has been established for information exchange on BATs. This process leads to the creation of Best Available Technique Reference documents (BREFs). It is thus based on a consensual step to gather the European "good-performing" industries. Because a large number of installations may be targeted, only a few reference installations can usually be studied and consequently need to be identified.

Secondly, at local level, operators have to compare environmental performances of a given installation with the information contained in their reference documents, in particular with BAT-Associated Environmental Performance Levels (BATAEPL). If they do not reach these BATAEPLs, they will have to provide a plan to improve their environmental performances or justify this impossibility with technical and economic arguments.

After a brief presentation of its context of application, this article describes a statistical method, applied to the dairy sector, for the determination of these sectoral reference installations. Then, the use of the concept of BAT as a sustainability tool in manufacturing, beyond its legal context of application is explored.

2. Context

2.1. Legal background

BATs are defined during an exchange information process named the Sevilla Process. This framework is described in an implementing decision [3,4]. Its first steps are aimed to define the environmental, economic and technical information about installations and techniques to be collected and shared among stakeholders. Fig. 1 illustrates the interactions among the various groups involved in this technical work in coordination with the European IPPC Bureau (EIPPCB). Thus, at European level, a Technical Working Group (TWG) composed of representatives of Member States, the European Commission, the Industry, and environmental NGOs is created. Its first task is to define the scope and the “key environmental issues” which will be considered. In order to coordinate the national contributions to the Sevilla Process, discussions among stakeholders are optionally undergone by a "shadow group". This shadow group can include representatives of the industry, national authorities and environmental non-governmental organisations, depending on the choices of the Member State.

After publication of a BREF, site operators and environmental authorities are concerned with the application of the BAT conclusions since they use the BREF in order to verify that an installation has a level of environmental performance comparable to BATs'.

Then, an extensive data collection is carried out, targeting performances and characteristics of sites currently operating in Europe. The transition from this step of information analysis to the definition of BATs highly relies on the expertise of its actors with a risk of biased assessment due to differences of interpretation.

The outcome of this process is a reference document (BREF) whose most important aspect is a description of sectoral BATs and emission levels associated with these techniques (BATAEL). The “BAT conclusions” extracted from these BREFs are published as European Commission Decisions and thus, bear a legal value which makes them essential to the directive.

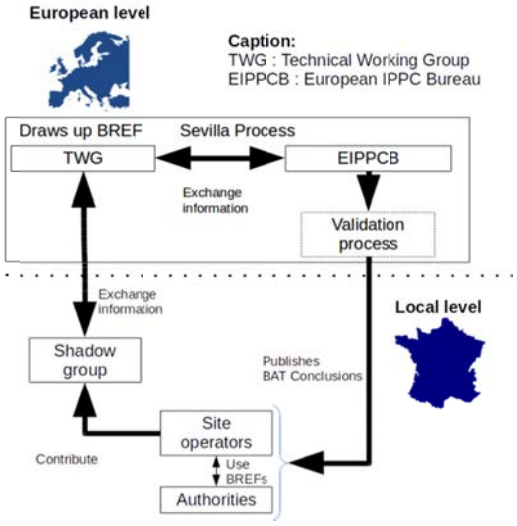


Fig. 1. Interactions among stakeholders involved in the IED.

Furthermore, other sectors, outside the scope of the IED must also apply BATs (e.g. nuclear installations in France [5]) whereas they do not possess any framework similar to the Sevilla Process.

2.2. Previous works on BAT selection

Several methods have been developed since the late 90s to help decision-makers to determine BATs at sector or installation levels [6–10]. They have been analyzed for this project in a previous literature review [11]. The main teachings of the study of existing researches was that they address three main issues: (1) local application of the IPPC directive or the IED for operational permits [9,10]; (2) selection of BATs at industrial sector scale with tools ranging from expert judgment [6] to Life Cycle Assessment [7] or potential impact assessment [8]; (3) determination of emission levels associated with BATs (BATAELs) [12]. In sectors outside the scope of the IED, like in the nuclear industry [2], operators must prove that they apply BATs but have no reference documents and therefore have to find their own references to assess their installations. They can do so using existing BREFs to find applicable techniques although they were not made for their sector or resort to their own resources to look for any helpful data.

Previous works on the topic of BAT identification differ according to their goals and scope. A local BAT determination method mainly relying on expert judgment was found [6,12], while two other methods were aimed to reduce subjective elements [13,14]. Thus, Geldermann and Rentz

[13] relied on more factual parameters to assess techniques, and the National Observatory of Athens [14] used mathematical approaches to guide the decision.

Moreover, the number of installations considered differed. When this number was too high to consider every single installation, a sampling was used like in the Sevilla Process [3,13,14].

The Sevilla Process and Dijkmans' method [6] focused on emissions, while also considering larger performance criteria. However, Geldermann and Rentz [13] targeted environmental impacts, and the National Observatory of Athens [14] on economical valuation of impacts.

Basically, all the existing methods were about on the selection of techniques and their associated performance levels without explaining the upstream decision process based on the selection of installations, except in Polders et al. [12]. The only reference for this preliminary step, which is essential to identify environmental performances associated with BATs, is the Sevilla Process [3,4].

3. Methodological approach applied to the dairy industry

In the framework of the IED, reference installations are used to define BATs and to determine associated environmental performance levels. The method, presented here in an application to the French dairy industry, is a statistical approach which considers several environmental indicators in a classification and selection process. Its outcome is a list of a small number of installations chosen according to their shared environmental characteristics or their low values of consumption and emissions. Economic and technical aspects are considered through expert judgment in a later phase which is not presented here.

3.1. General methodology

The developed method consists of five steps (Fig. 2). First, the scope and objectives of the study are defined. Three main items are required: the population to be studied, the variables considered in the statistical analysis, and the normalization used to consider the size of the installations.

Secondly, data previously collected are processed in order to correct possible errors and to ensure consistency of data types (numerical or textual). Missing data may dramatically influence the analysis; therefore an imputation method based on random forests [15] was used in order to mitigate this effect.

Thirdly, a statistical analysis is carried out to classify the installations. The language for statistical computing R [16] was used to identify groups of installations based on their consumption and emissions. Two approaches were used. The first one is a characterization of the consumption and emissions of the installations thanks to a Principal Component Analysis with hierarchical clustering [17] which enables to identify clusters of installations with similar characteristics (representative approach). The most representative installations are here defined as the closest to the center of their class. The second approach is aimed to look for optimal installations minimizing the variables considered (performant approach) thanks to the calculation of a Pareto Front in order to select the installations with the lowest values considering all the variables and therefore, the environment as a whole [18].

Fourthly, a few installations are selected, according to their proximity to the center of their class in the representative approach and their presence on the Pareto front or not for the performant approach.

Finally, performance levels of these selections for each variable are analyzed and compared to the population studied.

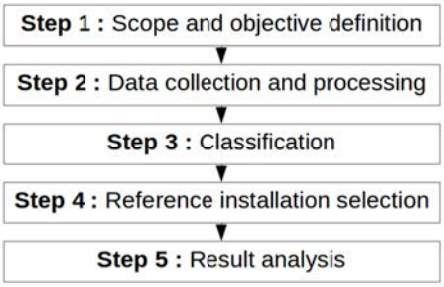


Fig. 2. Steps of the method.

3.2. Data used

The French dairy federation (CNIEL) provided data collected to prepare the revision of the "Food, Drink and Milk" (FDM) BREF. Thus, qualitative and quantitative information on installation characteristics, production, consumption and emission levels were made available for this study. They concerned a total of 115 IED installations located in France. The data collection was not comprehensive and some data were missing.

3.3. Application to the French dairy industry

3.3.1. Step 1: Scope and objective definition

The goal was to propose a list of reference installations for the French industry. Considering the choice of possible parameters, 16 scenarios could be carried out. The parameters chosen were the closest to the scope and key environmental issues defined during the kick-off meeting of the revision process for the Food, Drink and Milk (FDM) BREF [19].

Thus, nine variables, expressed as flows, were considered to classify the installations: energy consumption (EC), water consumption (WC), volume of waste water (WW), chemical oxygen demand (COD), biological oxygen demand (BOD), total suspended solids (TSS), total nitrogen (TN), ammonium nitrogen (AN) and total phosphorus (TP). Several variables were discarded due to a lack of data (Total Organic Carbon and chloride ion) or because of their specificity to a certain activity (dust emissions for installations manufacturing milk powder). In this example, emissions were expressed as specific loads (mass of pollutant released per mass of product manufactured or mass of raw material used [3]).

Finally, the size of installations was considered according to the input of processed milk and only 48 installations discharging directly their waste water into the environment were considered [19].

3.3.2. Step 2: Data collection and processing

The data collected by the French industry were verified and missing data were imputed to obtain a complete set of installations [15].

3.3.3. Step 3: Classification

Classifications were made according to both representative and performant approaches. Thus, 3 classes (groups) were found for the representative approach: class 1 contained 36 sites whose emissions and consumption are similar enough to form a cluster; class 2 was constituted of 11 sites whose energy consumption, water consumption and volume of waste water is larger than in the other class (Fig. 3). Besides, it had lower average levels of emission per ton of input of processed milk than class 1 except for total and ammonium nitrogen. Finally, class 3 contained only 1 site whose energy consumption was lower than the other two and emissions were much higher for each variable.

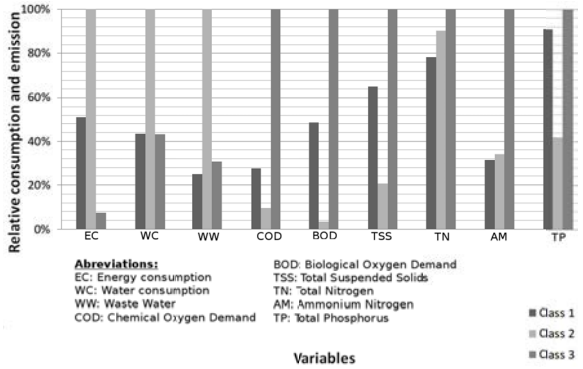


Fig. 3. Average consumption and emissions for the classes obtained with the representative classification for direct discharge sites.

Then, statistical hypothesis testing was done to look for other variables which might influence the classification. In this case, only the input of processed milk and the mass of manufactured products appeared as significant in the classification. The main product or other characteristics did not seem to influence the classification. Thus, class 1 contained the largest installations both in terms of input dairy matter and mass of products manufactured. Class 2 included the sites with the smallest productions and class 3 was a single site with lower than average input dairy matter for an intermediate production. In terms of consumption and emission levels, class 1 had lower values than class 2 for each variable whereas installations in class 2 were smaller than in class 1. Therefore, emission and consumption levels appeared to be inversely proportional to the size of the sites. It is to notice that this trend was similar in the other scenarios studied.

For the performant approach, the calculation of the Pareto front provided 6 installations whose consumption and emission levels were the lowest considering all the variables [18].

3.3.4. Step 4: Selection of reference installations

In the representative approach, installations were sorted according to their class and distance to their center. In this example the selection criterion was to obtain at most 15 installations. Thus, 3 sites in class 2 and 11 in class 1 were selected due to their closeness to the center of their respective class. The site in class 3 occupies the last spot but, considering its singular consumption and emission values, it may be a particular case whose selection would have to be approved by expert judgment or further investigations. In this example, it was considered separately from the others to prevent it from altering the comparison among average values in the following step.

With the performant approach, 6 optimal installations were identified and therefore kept. The number of installations obtained with this approach may vary according to the number of variables considered. Another set of variables may have provided a different outcome.

3.3.5. Step 5: Result analysis

This step aims at comparing the sites selected with the rest of the population in order to consider the "representativeness" of the selections. Thus, two analyses have been carried out. Since the relative levels of emission and consumption have been used in the classification, the first result considered here was a comparison of the values for the variables used in the statistical analysis (step 3). Thus, a comparison of the average consumption and emission levels showed that both selections were below the population average (Fig. 4). Emissions for the performant approach appeared to be lower than the representative selection except for COD and BOD. This seemingly odd result could be explained by the Pareto front approach. Indeed, it seeks for a compromise among variables; therefore the installations considered cannot have low levels on all the variables. A more detailed statistical analysis of these results has been carried out but is not showed in this article. It confirms that the methodological approaches lead to coherent results compared to the definition of representative and performant installations.

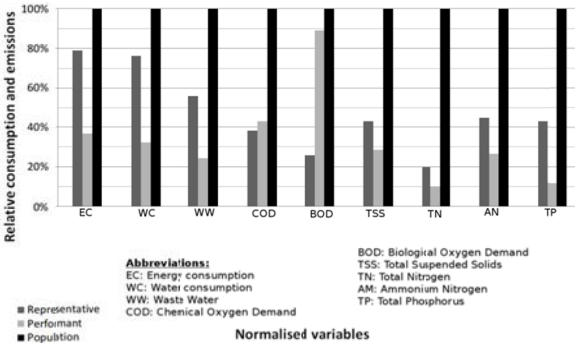


Fig. 4. Comparison of the average levels of consumption and emission of the selections and the population.

Since an important concern is the potential impact of the decisions made in the selection of reference installations, the percentages of sites from the whole population which have, for each variable, a larger value than the selection are given (Fig.5). In this simplified example, the comparison was made with the average value of the selected sites. Thus, the performant selection would be more stringent if used to determine emission limit values. Such a tool could be helpful

at the level of shadow groups to foresee the possible impact of a given selection for the national industry. This can be viewed as an estimation of the number of installations which would need to adapt their emissions to potential emission limit values deduced from these selections (Fig. 5).

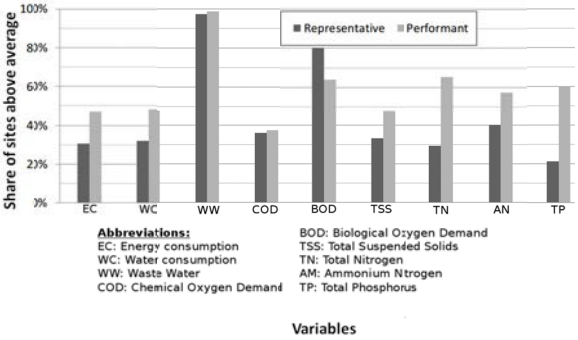


Fig.5. Share of sites which would have to make an effort to reach the average of the selections.

Thus, a larger share of the population had higher values in the performant selection compared to the representative approach. For the latter, 31 to 52 % of the total number of installations would have to adapt its emission levels, if deduced from this selection, whereas 53 to 92 % would be concerned with the performant approach, which would imply dramatic changes for the sector. This result is consistent with what is expected from a representative or performant selection of installations.

4. Discussion: BATs as a contributor to sustainability

The very definition of BAT already involves a relation with sustainability as defined in the Brundtland report [20] since it contains environmental, economic and technical considerations. Indeed, the environmental dimension is taken into account in the "high general level of protection of the environment as a whole" stated in the IED [2]. The economic aspects are included in the availability of a technique. Social considerations are included in the examination of chronic and accidental risks of industrial activities as stated in annex III of the IED [2] and a wider consideration of the impacts of BAT implementation for companies (in terms of employment, occupational safety, etc.) and for the society (e.g. impacts on health in the vicinity). Therefore, the concept of BAT contributes to the sustainability of the European industry.

Moreover, BATs, as defined under the IED, mainly focus on targeted substances and leave the assessment of their significance and impacts to other studies and expert judgment. Therefore a holistic approach could take place in the preliminary steps in order to use this statistical approach to highlight the most significant criteria. Indeed, although targeting key issues may ease the decision process, it implies a risk to miss important elements. In this aspect, a statistical approach may strengthen the decision to reduce the number of data.

Anyhow, balance between completeness and feasibility of such an analysis would have to be sought according to the objectives of the study. Moreover, social and economic aspects should be considered at the same level than the environment.

Since the BAT approach is a sectoral approach, it would be possible to assess the impact of technical alternatives on an entire industrial sector in a given geographical area by considering their availability or diffusion.

5. Conclusions

This article introduced a statistical multicriteria method to help to select reference installations in a view to identify Best Available Techniques. These reference installations can either be representative of a panel or the most performant among a studied population according to a multivariate approach considering their levels of environmental consumption and emission. This ongoing research is aimed to be applicable to industrial sectors under the IED and any other sectors concerned with BATs. Moreover, it is meant to be valuable for the existing decision processes carried out at any geographical levels, from local to international, providing that a sufficient collection of homogeneous data is provided. Developments on the method are still in progress for the determination of potential BATs from the reference installations.

Beyond the legal context of application of BATs, this statistical approach could probably be included in a sustainability tool to compare alternatives in a view to improve the sustainability of the industry by helping to identify key-issues to be investigated and improved according to the characteristics of an industrial sector. Moreover, with further research the proposed approach may be helpful to compare technical alternatives from a sustainable point of view.

Acknowledgements

Our thanks to CNIEL for allowing us to use the data of the French dairy industry they gathered in our study.

References

1. Council directive 96/61/EC of 24 September 1996 concerning integrated pollution prevention and control, European Commission Publication Paper, L257/26.
2. European Commission, 2010. Directive 2010/75/EU of the European parliament and of the council of 24 November 2010 on industrial emissions (integrated pollution prevention and control). Off J Eur Union 2010, L334/17-118.
3. European Commission, 2012. Commission implementing decision of 10 February 2012 laying down rules concerning guidance on the collection of data and on the drawing up of BAT reference documents and on their quality assurance referred to in directive 2010/75/EU of the European parliament and of the council on industrial emissions. Off J Eur Union 2012, L63/1-39.
4. Laforest V., Gaucher R., 2015. Industrial Emisson Directive (IED) and Best Available Techniques, Techniques de l'ingénieur, 2015, 15p. [only available in French].
5. OSPAR, 2015. Best Available Techniques (BAT) and Best Environmental Practice (BEP). Available from: <http://www.ospar.org/about/principles/bat-bep> [16 June 2015].

6. Dijkmans, R., 2000. Methodology for selection of best available techniques (BAT) at the sector level. Journal of Cleaner Production 8 (1), 11–21.
7. Nicholas, M. J., Clift, R., Azapagic, A., Walker, F. C., Porter, D. E., 2000. Determination of best available techniques for integrated pollution prevention and control: A life cycle approach. Process Safety and Environmental Protection 78 (3), 193–203.
8. Halog, A., Schultmann, F., Rentz, O., 2001. Using quality function deployment for technique selection for optimum environmental performance improvement. Journal of Cleaner Production 9 (5), 387–394.
9. Cikankowitz, A., Laforest, V., 2013. Using BAT performance as an evaluation method of techniques. Journal of Cleaner Production 42, 141–158.
10. Laforest, V., 2014. Assessment of emerging and innovative techniques considering best available technique performances. Resources, Conservation and Recycling 92, 11–24.
11. Evrard, D., Laforest, V., Villot, J., Gaucher, R., 2016. Best Available Technique assessment methods: a literature review from sector to installation level. Journal of Cleaner Production 121, 72–83.
12. Polders, C., Van den Abeele, L., Derden, A., Huybrechts, D., 2012. Methodology for determining emission levels associated with the best available techniques for industrial waste water. Journal of Cleaner Production 2930, 113–121.
13. Geldermann, J., Rentz, O., 2004. The reference installation approach for the techno-economic assessment of emission abatement options and the determination of BAT according to the IPPC-directive. Journal of Cleaner Production 12 (4), 389–402.
14. National Observatory of Athens, 2006. Decision aid framework for the assessment of impacts from the introduction of best available techniques in industry. (Final project report executive summary, National Observatory of Athens), 39p.
15. Stekhoven DJ, Buhlmann P. MissForest--non-parametric missing value imputation for mixed-type data. Bioinformatics. 2012 Jan 1;28(1):112–8.
16. CRAN. The Comprehensive R Archive Network [Internet]. [23 Nov 2015]. Available from: https://cran.r-project.org/
17. Saporta G. Probabilités, analyse des données et statistique. Editions TECHNIP; 2011. 622 p.
18. Mavrotas, G., Georgopoulou, E., Mirasgedis, S., Sarafidis, Y., Lalas, D., Hontou, V., Gakis, N., 2007. An integrated approach for the selection of best available techniques (BAT) for the industries in the greater Athens area using multi-objective combinatorial optimization. Energy Economics 29 (4), 953–973.
19. EIPPCB. Kick-off meeting for the review of the reference document on Best Available Techniques for Food, Drink and Milk Industries − Meeting report [Internet]. 2015, 45. p. Available from: <http://eippcb.jrc.ec.europa.eu/reference/BREF/FDM_KoM_Report_102014.pdf>.21. United Nations.
20. Our Common Future - Brundtland Report. Oxford University Press; 1987, 204 p.

École Nationale Supérieure des Mines
de Saint-Étienne

NNT : *2016LYSEM015*

Damien EVRARD

Integrated approach for multicriteria assessment of Best Available Techniques

Speciality: Environmental sciences and engineering

Keywords: Best Available Techniques, multicriteria assessment, statistical analysis, industrial emissions, environmental performance, decision-making support.

Abstract:

In order to reduce chronic emissions due to industrial activities, the Industrial Emission Directive (IED − 2010/75/CE) defines, for the Member States, a legal framework and the concept of Best Available Technique (BAT). This concept of BAT and their associated environmental performance levels (BATAEPL) enables to set common objectives in Europe while considering local specificities. Moreover, the application of the concept of BAT to non-IED sectors (e.g. Nuclear installations) makes it a major issue for the entire European industrial community. BAT reference documents (called BREFs) have become essential to help to choose techniques and are references for environmental performance assessment of installations. The process of their revision can be improved because of the differences of interpretation that exist among involved actors for industrial information analysis. Thus, the contribution of this thesis is a methodology structured in 5 steps based, on the one hand, on multicriteria statistical tools (hierarchical clustering, PCA, Pareto front) to help to determine reference installations and applicable BATs thanks to representative or performant approaches and, on the other hand, to define associated environmental performance levels, at international and national scales. This methodology has demonstrated its applicability and integration into current decision processes thanks to applications to both geographical scales, and to IED and non-IED activities.

École Nationale Supérieure des Mines
de Saint-Étienne

NNT : 2016LYSEM015

Damien EVRARD

APPROCHE INTEGREE POUR UNE EVALUATION MULTICRITERE DES MEILLEURES TECHNIQUES DISPONIBLES

Spécialité : Sciences et génie de l'environnement

Mots clefs : Meilleures Techniques Disponibles, Évaluation multicritère, analyse statistique, émissions industrielles, performance environnementale, aide à la décision.

Résumé :

Dans le but de réduire les émissions chroniques issues des activités industrielles, la directive sur les émissions industrielles dite IED (2010/75/CE) définit le cadre réglementaire imposé aux États membres et le concept de Meilleures Techniques Disponibles (MTD). Ce concept de MTD et les niveaux de performance environnementale qui lui sont associés (BATAEPL) permettent de fixer des objectifs communs à toute l'Europe tout en considérant des spécificités locales. Par ailleurs, l'application du concept de MTD à des secteurs non soumis à l'IED (ex : Installations Nucléaires de Base) en fait un enjeu majeur pour l'ensemble de la communauté industrielle européenne. Les documents de référence sur les MTD (dénommés BREF) sont devenus incontournables comme aides au choix de techniques et références pour l'évaluation des performances environnementales des installations. Le processus de leur révision est perfectible car des différences d'interprétation et d'application existent entre les différents acteurs impliqués dans l'analyse des informations industrielles. Ainsi, l'apport de ce travail de thèse est une méthodologie structurée en 5 étapes, reposant d'une part sur des outils statistiques multicritères (classification ascendante hiérarchique, ACP, front de Pareto) d'aide à la détermination de sites de référence et de MTD applicables via des approches représentative ou performante et, d'autre part, à la définition de niveaux de performance environnementale associée, à l'échelle internationale et nationale. Cette méthodologie a démontré son applicabilité et son intégration dans les processus de décision actuels par les applications effectuées aux deux échelles géographiques et aux activités IED et non-IED.