



HAL
open science

Gestion Patrimoniale Multi-Echelles des réseaux d'eau potable (GePaME) - Rapport final 2023

Selma Baati, Kevin Caillaud, Marie Collet, Vincent Couallier, Lina Essafi, Alain Husson, Yves Le Gat, Eddy Renaud, Nicolas Rodriguez, Bénédicte Rulleau

► **To cite this version:**

Selma Baati, Kevin Caillaud, Marie Collet, Vincent Couallier, Lina Essafi, et al.. Gestion Patrimoniale Multi-Echelles des réseaux d'eau potable (GePaME) - Rapport final 2023. INRAE UR ETTIS; SWDE; IMB - Institut de Mathématiques de Bordeaux. 2023. hal-04311166

HAL Id: hal-04311166

<https://hal.inrae.fr/hal-04311166v1>

Submitted on 28 Nov 2023

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

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

Projet GePaME

Gestion Patrimoniale Multi-Echelles des réseaux d'eau potable

Rapport final 2023

Selma Baati, Kévin Caillaud, Marie Collet, Vincent Couallier, Lina Essafi, Alain Husson, Yves Le Gat, Eddy Renaud, Nicolas Rodriguez, Bénédicte Rulleau.

Etabli le 01/09/2023
Révisé le 20/11/2023

Table des matières

Introduction.....	14
Contexte	14
Contenu du projet GePaME	14
Organisation du rapport.....	15
1 Tâche 01 – Prise en main et traitement des données.....	16
1.1 Retour aux conduites non agréées.....	16
1.2 Exploitation des données de 2022	16
1.3 Complétion des données des conduites hors service	16
1.4 Analyse des codes de cause de panne	17
1.5 Nouvelle organisation géographique de la SWDE.....	17
1.6 Correspondances des numéros de cellules entre le SIG et SAP pour la RF.....	18
1.7 Notions d’adduction et de distribution	19
2 Tâche 02 – La sécurisation par la performance. Analyse sociologique du mode de gouvernance interne à la SWDE.....	21
2.1 Introduction.....	21
2.2 La Société Wallonne Des Eaux : recomposition des enjeux et reconfigurations du service .	23
2.2.1 De l’hygiénisme local à la sécurisation régionale du service d’eau potable	23
2.2.2 À la recherche d’un mode d’organisation et de gestion « performant ».....	24
2.3 La mise en œuvre du paradigme de la performance	27
2.3.1 Une stratégie d’entreprise orientée vers l’agir collectif et l’anticipation des dysfonctionnements.....	27
2.3.2 L’agir collectif : un travail de définition et d’écriture compliqué.....	28
2.3.3 Une « rationalisation » inachevée.....	30
2.4 Un pilotage par les indicateurs.....	32
2.4.1 Chiffrer la performance : de la mesure du challenge à l’épuisement socio-professionnel	32
2.4.2 Automatiser les décisions : une désocialisation progressive au terrain	33
2.4.3 L’implication managériale en soutien du chiffre.....	36
2.5 Conclusion	39
3 Tâche 03 – Implémentation de Casses.....	42
3.1 Analyse descriptive des tronçons et des casses	42
3.2 Calage	42
3.2.1 Jeux de données	42
3.2.2 Choix des covariables pour chaque strate.....	43

3.2.3	Résultats des calages.....	43
3.3	Validation	45
3.3.1	Données jusqu'en 2021	45
3.3.2	Données de 2022.....	47
3.4	Prévision	48
4	Tâche 04 – Intégration des outils dans le système d'information	52
4.1	Nouveaux éléments de contexte.....	52
4.2	Positionnement des défaillances dans le SIG et propositions de corrections	52
4.3	Production des fichiers pour Casses.....	53
4.4	Importance du lien relationnel entre anomalies et tronçons.....	54
4.5	Calcul des débits.....	55
5	Tâche 05 – Exploitation des données de la sectorisation	56
5.1	Evolution des débits	56
5.2	Étude des débits consommés par les abonnés des cellules.....	58
5.2.1	Structure des données	58
5.2.2	Calcul des débits de compteurs et de cellules	59
5.2.3	Vérification des calculs	59
5.2.4	Méthode de correction des débits de consommation.....	59
5.2.5	Bilan sur les débits consommés par les abonnés	60
5.3	Lien entre débits et caractéristiques de la cellule.....	60
5.3.1	Evolution des <i>ILP</i>	60
5.3.2	Indicateurs de pertes.....	63
5.3.3	Relation entre indicateurs de pertes et caractéristiques des cellules (extrait de l'Annexe T05_02)	64
5.3.4	Analyse chronologique des $(Q_{nuit}-Q_{ref})/L$ et des fuites, crash et RF	69
5.3.5	Exploration de référence alternative	70
5.4	Complément à l'ILP de référence : Pente.....	70
6	Tâche 06 - Analyse des pratiques de recherche de fuites.....	73
7	Tâche 07 – Modélisation de la probabilité d'un tronçon d'être fuyard	74
7.1	Principe et rappel	74
7.2	Le nouveau modèle : Zero-inflated binomial (ZIB).....	75
7.2.1	Principe.....	75
7.2.2	Modèle d'inspection.....	76
7.2.3	Modèle d'état fuyard.....	76
7.2.4	Modèle ZIB.....	76

7.2.5	Estimation des paramètres	76
7.3	Test de la robustesse du modèle.....	76
7.3.1	Principe.....	76
7.3.1.1	Estimation des paramètres dans les cas imparfaits	77
7.3.1.2	Impact du choix des paramètres initiaux de l'optimisation sur l'estimation des paramètres	78
7.3.2	Résultats.....	78
7.3.2.1	Estimation des paramètres dans les cas imparfaits	78
7.3.2.2	Impact du choix des paramètres initiaux de l'optimisation sur l'estimation des paramètres	80
7.4	Analyse des données de la SWDE.....	84
7.4.1	Données d'entrée.....	84
7.4.1.1	Campagnes	84
7.4.1.2	Données finales	85
7.4.1.3	Méthodes	86
7.4.2	Résultats.....	87
7.4.2.1	Sélection de modèles	87
7.4.2.2	Meilleur modèle : modèle ZIB 17	89
7.5	Perspectives.....	93
8	Tâche 08 - Analyse de la vulnérabilité des clients.....	95
8.1	Introduction.....	95
8.2	La méthode des choix multi-attributs	96
8.3	Revue de littérature sur l'évaluation de la performance des services d'eau potable	98
8.4	Protocole d'enquête.....	99
8.4.1	Construction des séries de choix.....	99
8.4.2	Structure du questionnaire	101
8.4.3	Protocole d'enquête.....	101
8.5	Résultats.....	103
8.5.1	Analyse statistique descriptive.....	103
8.5.2	Analyse bivariée par rapport au groupe de communes.....	107
8.5.3	Analyse économétrique des choix.....	110
8.5.3.1	Rappel des attributs et de leurs niveaux.....	110
8.5.3.2	Résultats du logit conditionnel.....	111
8.5.3.3	Résultats du modèle à classes latentes.....	113
8.5.4	Caractérisation des classes et estimation des consentements-à-payer	119

8.5.4.1	Caractérisation des classes.....	119
8.5.4.2	Représentation des liens entre groupes de communes et classes	120
8.5.4.3	Estimation des consentements-à-payer.....	123
8.6	Conclusion	125
9	Tâche 09 - Modélisation statistique couplée des fuites et des crashes.....	127
9.1	Développement d'une variante de Cassettes et du modèle LEYP	127
9.2	Mise en œuvre du modèle de type LEYP avec délai.....	128
9.3	Application un modèle Multi-états Markovien pour la dégradation des tronçons	128
9.3.1	Généralités	128
9.3.2	Intégration de covariables.....	129
9.3.3	Le cas particulier des données de panel.....	130
9.3.4	Le cas particulier des données d'inspection générant une possibilité de mauvaise classification : modèle HMM.....	130
9.3.5	Synthèse du formatage des données SWDE pour le calage du modèle HMM	131
9.3.6	Ajustement et modélisation par le package msm de R.....	132
9.3.7	Résultats	133
9.3.7.1	Préparation des données	133
9.3.7.2	Choix des variables	135
9.3.7.3	Résultats du calage par msm, pour les trajectoires initiales $N_{def0}=0$: modèle F ..	136
9.3.7.4	Résultats du calage par msm sur toutes les trajectoires : modèle G	139
9.3.7.5	Tableaux de sorties des prédictions (modèle G)	142
9.3.7.6	Conclusion et perspectives.....	144
10	Tâche 10 – Outil multicritère de programmation court terme	146
10.1	Objectif de la tâche et décisions stratégiques.....	146
10.2	Objectifs, critères, et indicateurs de performance et d'opportunité.....	147
10.3	Calcul des indicateurs.....	148
10.3.1	Indicateurs continus utilisés pour le rang des tronçons.....	149
10.3.2	Indicateurs discrets à caractère informatif	151
10.4	Agrégation des indicateurs.....	152
10.5	Format des résultats de l'outil	154
11	Tâche 11 - Efficacité des actions de recherche de fuite.....	156
11.1	Objectifs et mise en œuvre	156
11.2	Résultats par secteur et par technique	158
11.2.1	Pré-localisation	158
11.2.2	Localisation.....	159

11.2.3	Réparation	160
11.2.4	Débits.....	161
11.3	Résultats par cellules (secteurs de Namur et Verviers)	163
11.3.1	Données des cellules retenues	163
11.3.2	Regroupement des cellules par type de campagne	163
11.3.3	Prise en compte de la typologie des cellules.....	164
11.4	Conclusion	165
12	Tâche 12 – Etude de scénarios long terme et simulation	168
12.1	Objectif de la tâche et déroulement	168
12.2	Principe de la simulation long terme (SLT).....	168
12.3	Hypothèses utilisées.....	169
12.4	Résultats	170
12.5	Prochaines étapes	171
13	Tâche 13 - Coordination du Projet et valorisation	173
13.1	Coordination du projet.....	173
13.2	Valorisations	173
	Conclusion générale	176
	Principaux résultats	176
	Autres enseignements.....	176
	Références.....	178

Figures

Figure 1. Vue d'ensemble de l'articulation des tâches et des volets du projet.	15
Figure 2. Les sociétés de distribution d'eau potable en Wallonie.	21
Figure 3. Structuration de la SWDE depuis Juillet 2022.	26
Figure 4. Les Programmes de la SWDE.	27
Figure 5. Résumé des 15 strates utilisées par le modèle LEYP dans Casses.	42
Figure 6. Validation 2020-2021 des casses prédites sur la distribution FG-Verviers.	46
Figure 7. Validation 2020-2021 des casses prédites sur l'adduction.	46
Figure 8. Validation 2022 des casses prédites sur la distribution.	47
Figure 9. Validation 2022 des casses prédites sur l'adduction.....	48
Figure 10. Carte des prévisions de défaillances au 1:5000 de Charleroi-Jumet (distribution).	49
Figure 11. Carte des prévisions de défaillances au 1:50 000 de Charleroi (distribution).....	50
Figure 12. Carte des prévisions de défaillances au 1:700 000 de la SWDE (distribution).....	50
Figure 13. Carte des prévisions de défaillances au 1:700 000 de la SWDE (adduction).	51
Figure 14. Défaillances SAP (points) positionnées sur les tronçons de manière uniforme.	53
Figure 15. Evolution des débits moyen jour et nuit 56	56
Figure 16. Evolution des ILP en fonction des réparations..... 57	57
Figure 17. Ecart ILP0.01 - ILPref..... 58	58
Figure 18. ILP moyen de 2000 cellules en fonction de la densité de raccordement..... 60	60
Figure 19. ILP moyen en fonction de la densité de raccordement 61	61
Figure 20. ILP moyen en fonction de la densité de logement..... 62	62
Figure 21. Evolution des ILP moyen annuel de 2017 à 2021..... 62	62
Figure 22. ILP 0.01 en fonction de la densité de logement..... 63	63
Figure 23. Illustration de la dépendance de Q_{nuit} , Q_{ref} à la taille de la cellule (N et L). 63	63
Figure 24. $(Q_{nuit}-Q_{ref})/L=f(cRF/L)$, influence du type de classe (secteur, N, Dr, L). 64	64
Figure 25. Nombre de fuites trouvées par km en fonction du nombre de campagnes réalisées par km. 69	69
Figure 26. Evolution annuelle des fuites et crashes et du $(Q_{nuit}-Q_{ref})/L$ 69	69
Figure 27. Quantile de Q_{nuit}/L réparti par classe de raccordement. 70	70
Figure 28. Pente stabilisée en 2020 et 2021. 71	71
Figure 29. Instabilité des ILP de la cellule confirmée en 2020 et 2021. 72	72
Figure 30. Schéma du modèle de données. Z est la seule variable observée..... 75	75
Figure 31 . Estimation de θ_0 79	79
Figure 32. Estimation de θ_1 79	79
Figure 33. Estimation de θ_2 79	79
Figure 34. Estimation de θ_3 79	79
Figure 35. Valeurs estimées de θ_0 en fonction de θ_1 selon différentes valeurs initiales avec un faible écart type des données ($\sigma = 1$)..... 80	80
Figure 36. Valeurs estimées de β_1 en fonction de β_2 selon différentes valeurs initiales avec un faible écart type des données ($\sigma = 1$)..... 81	81
Figure 37. Valeurs estimées de θ_0 en fonction de θ_1 selon différentes valeurs initiales avec un écart type des données augmenté ($\sigma = 2$)..... 81	81
Figure 38. Valeurs estimées de β_1 en fonction de β_2 selon différentes valeurs initiales avec un écart type des données augmenté ($\sigma = 2$)..... 82	82

Figure 39. Opposé de la log-vraisemblance après optimisation du modèle ZIB selon les 31 jeux de paramètres initiaux d'optimisation.....	84
Figure 40. Valeur des estimations du nombre de défaillances passée (Ndef) et de la pression (pression100).....	84
Figure 41. Courbe de validation du modèle complet.....	87
Figure 42. Courbe de validation du modèle 15.....	88
Figure 43. Courbe de validation du modèle 16.....	88
Figure 44. Courbe de validation du modèle 17.....	88
Figure 45. Courbe de validation du modèle 18.....	88
Figure 46. Fréquence cumulée des probabilités calculées par le modèle 17.....	93
Figure 47 : Exemple de série de choix.....	100
Figure 48 : Mail envoyé par la SWDE à un échantillon aléatoire de clients habitant dans les communes d'enquête.....	103
Figure 49 : Distribution des enquêtés par commune.....	105
Figure 50 : Répartition en classes par groupe de commune.....	121
Figure 51 : Distribution spatiale de la densité des enquêtés par groupe de commune (Classe 1).....	121
Figure 52 : Distribution spatiale de la densité des enquêtés par groupe de commune (Classe 2).....	122
Figure 53 : Distribution spatiale de la densité des enquêtés par groupe de commune (Classe 3).....	122
Figure 54 : Distribution spatiale de la densité des enquêtés par groupe de commune (Classe 4).....	123
Figure 55. Exemple de graphe de Markov, tiré de Chris Jackson (2014).	128
Figure 56. Matrice des intensités du graphe de Markov.	129
Figure 57. Graphe de Markov de succession des états.....	129
Figure 58. Exemple de trajectoire.....	130
Figure 59. Matrice de probabilité d'erreur de classification.....	131
Figure 60. Trajectoire d'un tronçon.....	131
Figure 61.Exemple d'une trajectoire vierge d'évènement.....	132
Figure 62.Exemple d'une trajectoire avec un crash (2 trajectoires créées).....	132
Figure 63. Exemple d'une trajectoire avec fuite trouvée par RF, puis réparée (2 trajectoires).....	132
Figure 64. Exemple de sortie du package msm.....	137
Figure 65. Résultats du calage par msm sur toutes les trajectoires : modèle G.....	140
Figure 66. Probabilité de transition entre les états en fonction du temps pour un tronçon en fonte ductile de caractéristiques autres connues.....	143
Figure 67.Probabilité de transition entre les états en fonction du temps pour un tronçon en acier de caractéristiques autres connues.....	144
Figure 68. Outil multicritère SWDE en place depuis 2017.....	146
Figure 69. Objectifs, critères, et indicateurs retenus pour l'outil.....	147
Figure 70. Zone de Chantier (trait épais) et Zone Périmètre (polygone) dans PoWalCo.....	152
Figure 71. Distribution normalisée cumulée en longueur des tronçons d'un indicateur I_j	153
Figure 72. Longueur inspectée des cellules en fonction du temps passé selon la technique.....	159
Figure 73 : Chaîne de combinaison des données associées aux actions et aux résultats.....	166
Figure 74. Trajectoire d'un tronçon dans la simulation long terme.....	169
Figure 75. Simulation long terme avec un budget probablement surestimé.....	170
Figure 76. Simulation long terme avec un budget probablement sous-estimé.....	171
Figure 77 : Planning final de réalisation du projet GePaME.....	173

Tableaux

Tableau 1. Tâches du projet GePaME.	14
Tableau 2 : Incohérences entre cellules SAP et SIG	18
Tableau 3 : Règles employées pour distinguer adduction/distribution pour la tâche 03.....	20
Tableau 4 : Règles employées pour distinguer adduction/distribution pour les tâches 10 et 12	20
Tableau 5. Résumé des caractéristiques des strates utilisées pour les modèles LEYP dans Casses.	43
Tableau 6. Résultats du calage du modèle LEYP sur les 15 strates.	44
Tableau 7. Résumé des résultats de validation sur les données jusqu'en 2021.	45
Tableau 8. Distribution du pourcentage de tronçons à risque élevé de défaillances (les 2% les plus à risque en linéaire cumulé) selon les Entités Fonctionnelles.	49
Tableau 9. Règles possibles de gestion des ID de tronçon en cas d'évolution du réseau.....	54
Tableau 10. Exemple de résultats d'estimation des paramètres du sous-modèle d'inspection du ZIB dans le cas d) du cadre 3	79
Tableau 11. Covariables du modèle ZIB utilisé pour tester l'influence des paramètres initiaux d'optimisation sur l'estimation des paramètres	83
Tableau 12. Récapitulatif des AIC des 19 modèles testés.	87
Tableau 13. Résultat de l'estimation des paramètres du modèle ZIB 17.	91
Tableau 14 : Liste des communes d'enquête par groupe	102
Tableau 15 : Analyse descriptive des caractéristiques socio-géo-économiques des répondants	106
Tableau 16 : Résultats de l'analyse bivariée par groupe de communes	109
Tableau 17 : Description des attributs et de leurs niveaux.....	111
Tableau 18 : Résultats du logit conditionnel	112
Tableau 19 : Comparaison entre les modèles selon le nombre de classes latentes	114
Tableau 20 : Résultats du modèle à 4 classes latentes	116
Tableau 21 : Résultats de l'analyse bivariée par classes	119
Tableau 22 : Estimation des CAP issus du modèle à classes latentes	124
Tableau 23. Effectifs des transitions observées des 3ZF après suppressions des double crashes (écarts en jours <=8 jours).....	134
Tableau 24. Résultats du modèle F : rapports des risques par variable et par transition.	138
Tableau 25. Résultats du modèle G : rapports des risques par variable et par transition.....	141
Tableau 26. Extrait de l'Annexe T09_01.....	142
Tableau 27. Répartition selon matériau et longueur des 63791 tronçons considérés dans les 3 ZF testées.	142
Tableau 28. Synthèse des indicateurs continus et des principaux facteurs les influençant.	149
Tableau 29. Nouvelles données demandées et reçues pour le développement de l'outil multicritère.	150
Tableau 30. Variables utilisées dans les indicateurs.	150
Tableau 31. Poids obtenus	154
Tableau 32. Format des résultats de l'outil multicritère.	155
Tableau 33. Répartition et caractéristiques cumulées par secteur des cellules sélectionnées.....	157
Tableau 34. Répartition et caractéristiques cumulées par secteur des cellules effectivement étudiées	157
Tableau 35. Campagnes de pré-localisation par type et par secteur.....	158
Tableau 36. Temps et longueurs inspectées par type de pré-localisation.....	158

Tableau 37. Techniques de localisation des fuites par secteur.....	159
Tableau 38. Durées de localisation selon la technique utilisée	159
Tableau 39. Réparations par secteur et par type d'organe.....	160
Tableau 40. Temps de réparations par organe puis par type de défaillance	160
Tableau 41. Débits moyens gagnés par les réparations selon l'organe réparé et le type de défaillances (m ³ /h).....	161
Tableau 42. Débits moyens gagnés par les réparations selon le type de pré-localisation (hors Mons) (m ³ /h).....	161
Tableau 43. Bilan cumulé des débits (m ³ /h) et des ILP (m ³ /j.km) hors secteur de Mons.....	162
Tableau 44. Extrait des données recueillies sur les cellules retenues	163
Tableau 45. Données et ratios des cellules regroupées par type de campagne.....	164
Tableau 46. Gain en débit des fuites réparées par journée de campagne (par technique et par classes de variables)	165
Tableau 47 : Communications des résultats de GePaME dans des conférences	174

Liste des sigles et abréviations

Abréviation	Objet	Description
AEP	A limentation en E au P otable	
BDD (BD)	B ase D e D onnées	
BI	B orne I ncendie	
CR	C ompte R endu	
CSO	C omité de S uivi O opérationnel	Réunion périodique de suivi du projet.
CLC	C orine L and C over	Occupation du sol
CT	C ourt T erme	
DDC	D ate D e C asse	Date à laquelle l'élément du réseau AEP (comme la conduite) a cassé.
DDP	D ate D e P ose	Date à laquelle la conduite a été posée.
DE	Diamètre Extérieur	diamètre extérieur (mm) d'un tuyau ou d'une conduite.
DHS	D ate de mise H ors S ervice	Date à laquelle la conduite a été mise hors service, ou abandonnée. La DHS de la canalisation est une donnée nécessaire à l'implémentation du modèle LEYP2s.
DIA	D iamètre	Selon le contexte, peut désigner DN ou DE.
DN	D iamètre N ominal	diamètre nominal (mm) d'un tuyau ou d'une conduite.
EPSPG	E uropean P etroleum S urvey G roup	
ES	E n S ervice	Utilisé pour décrire des conduites qui sont encore utilisées à l'heure actuelle pour l'AEP.
FG	F onte G rise	
GPI	G estion P atrimoniale des I nfrastructures	
HS	H ors S ervice	Synonyme de « abandonné », utilisé pour les conduites hors service.

ICAR	Inventaire Centralisé des Adresses et des Rues en Wallonie	Système utilisé par le SPW pour identifier de façon unique chaque rue ou point d'adresse par un numéro.
ID	Identifiant	Code utilisé dans un fichier pour identifier de façon unique chaque objet. L'OT SAP est un type d'ID.
INS (code)	Institut National de la Statistique	Code à 5 caractères unique pour chaque commune en Belgique.
LEYP	Linear Extended Yule Process	Modèle probabiliste à base du processus de Yule utilisé pour la prédiction des défaillances (fuites et casses) du réseau AEP (Le Gat, 2009).
LT	Long Terme	
MAT	Matériau	Matériau de l'élément du réseau AEP.
MNT (DTM)	Modèle Numérique de Terrain	Représentation numérisée de la topographie du terrain de surface.
NPS	Nominal Pipe Size	Diamètre de tuyau ou de conduite exprimé en pouces.
OT	Ordre de Travail	Numéro utilisé dans SAP pour identifier de façon unique chaque intervention.
PE	Polyéthylène	
PICC	Projet Informatique de Cartographie Continue	Cartographie des bâtiments de la Wallonie.
PoWalCo	Portail Wallon de Coordination	Portail internet permettant l'échange d'informations sur les intentions de travaux, la coordination des chantiers et la gestion des autorisations en Wallonie.
PVC	Polychlorure de Vinyle	
QGIS	Quantum Geographic Information System	Logiciel de SIG.
RF	Recherche de Fuite	
SAP	Systems Applications and Products in Data Processing	Progiciel de gestion intégré utilisé par la SWDE notamment pour archiver les interventions.
SCR	Système de Coordonnées de Référence	Informations mathématiques sur la méthode de projection de géométrie sphérique sur un plan.

SI	S ystème d' I nformation	Système (logiciel) utilisé pour stocker et traiter l'information.
SIG	S ystème d' I nformation G éographique	Système (logiciel) utilisé pour stocker et traiter l'information géographique.
SPW	S ervice P ublic de W allonie	Administration publique Wallonne.

Introduction

Contexte

Partant du constat de leur intérêt commun à rechercher des réponses scientifiques et opérationnelles aux questions soulevées par la gestion patrimoniale des réseaux d'eau potable, SWDE et INRAE ont décidé de s'associer pour conduire des travaux de recherche dans ce domaine. Ils ont ainsi conclu, le 9 juin 2020, un contrat de recherche qui établit une collaboration pour mener à bien le projet « Gestion Patrimoniale Multi-Echelles des réseaux d'eau potable » (GePaME) (Annexe T00_01). Le projet GePaME, d'une durée de trois ans, a démarré le 1^{er} septembre 2020 et s'est terminé le 31 août 2023. Pour conduire les recherches dans le domaine des statistiques appliquées, INRAE s'est associé avec l'Institut de Mathématiques de Bordeaux (IMB – UMR 5251) dans le cadre d'un contrat de collaboration en vigueur du 1^{er} novembre 2022 au 31 août 2023 (Annexe T00_02).

Le présent rapport est le rapport final du projet GePaME, il présente les résultats obtenus au cours du projet. Il fait référence, sans les reprendre dans le détail, aux résultats déjà exposés dans les rapports intermédiaires 2021 et 2022 (Husson et al. 2022; Rodriguez et al. 2021).

Contenu du projet GePaME

Le projet GePaME est organisé en trois volets :

- Volet 1 : compréhension du processus de fuite et efficacité de la recherche de fuite (RF) ;
- Volet 2 : aide à la sélection multi-objectifs et contrainte des chantiers de renouvellement ;
- Volet 3 : simulation long terme de stratégies de gestion patrimoniale.

Il se décompose en treize tâches opérationnelles (Tableau 1) qui doivent aboutir à des avancées scientifiques valorisées par l'INRAE et à des outils d'aide à la décision qui ont vocation à être intégrés dans les pratiques et dans le système d'information de la SWDE.

Tableau 1. Tâches du projet GePaME.

Tâche	Dénomination
01	Prise en main et traitement des données
02	Analyse de la gouvernance de la SWDE en matière de GPI
03	Implémentation de Casses
04	Intégration des outils dans le système d'information
05	Exploitation des données de la sectorisation
06	Analyse des pratiques de recherche de fuites
07	Modélisation de la probabilité d'un tronçon d'être fuyard
08	Analyse de la vulnérabilité des usagers
09	Modélisation statistique couplée des fuites et des crashes
10	Outil multicritère de programmation court terme
11	Efficacité des actions de recherche de fuite
12	Etude de scénarios long terme et simulation
13	Coordination du Projet et valorisation

La Figure 1 présente une vue d'ensemble de l'articulation des tâches en cohérence avec les trois volets du projet.

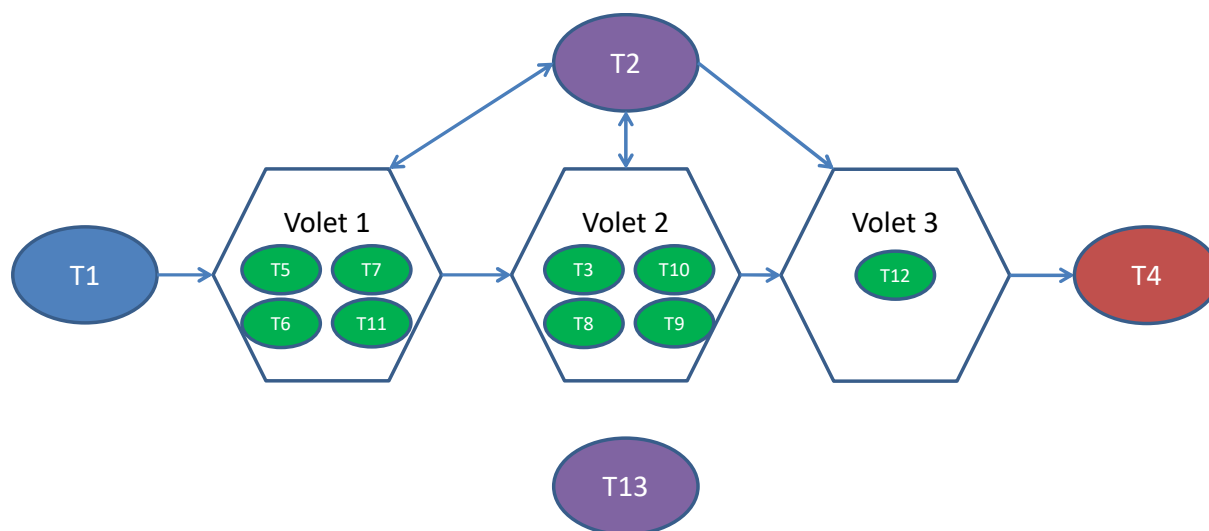


Figure 1. Vue d'ensemble de l'articulation des tâches et des volets du projet.

Organisation du rapport

Le présent rapport présente les résultats par tâche. Chacune d'entre elle fait l'objet d'un chapitre, elles apparaissent dans leur ordre de numérotation initial. Le texte renvoie à des annexes, identifiées par tâche et fournies dans des fichiers séparés.

En appui au rapport, toutes les données, codes de calcul ou autres documents pertinents, produits ou utilisés par le projet, sont fournis en archive par tâche au sein d'un ou plusieurs fichiers.

1 Tâche 01 – Prise en main et traitement des données

1.1 Retour aux conduites non agrégées

La décision la plus importante concernant la prise en main des données pour la dernière année du projet consiste en l'utilisation des tronçons d'origine de la SWDE (au contraire des tronçons agrégés décrits précédemment dans les rapports intermédiaires ; Rodriguez et al., 2021 ; Husson et al., 2022) dans toutes les analyses et tous les outils afin de travailler sur le même référentiel de conduites que dans les futures utilisations des méthodes et modèles. **Ainsi, tous les résultats présentés dans les sections suivantes concernent les tronçons « non agrégés », contrairement aux rapports précédents.**

Le transfert des défaillances des tronçons agrégés vers les tronçons d'origine s'est fait en affectant au plus long de l'agrégat l'ensemble des défaillances du tronçon agrégé (le détail du positionnement des défaillances est explicité au paragraphe 4.2). Pour l'utilisation dans les modèles (Casses et modèle d'état fuyard par exemple), cela signifie que les tronçons qui ne sont pas les plus longs de leur agrégat sont systématiquement à ignorer pour le calage car ils n'ont jamais eu de défaillances du point de vue du jeu de données. Néanmoins, parmi ces tronçons ignorés, ceux de longueur suffisante (au moins 6 mètres) sont pertinents et sont donc à inclure dans les résultats. En conséquence, il est nécessaire de propager les résultats des modèles des tronçons les plus longs vers les moins longs des agrégats (voir Annexe T01_01 pour les taux de casse et probabilités d'état fuyard).

1.2 Exploitation des données de 2022

Il a été précédemment décidé de ne pas prendre en compte les données SWDE au-delà du 31/12/2021 pour le développement et le calage des modèles dans GePaME. Néanmoins, les données de l'année 2022 ont été utilisées pour valider les modèles ou pour faire des comparaisons avec les données des années précédentes.

L'analyse des données de défaillances de 2022 a révélé que l'encodage des défaillances ne se fait en 2022 toujours pas exclusivement via des anomalies SIG, puisque près de 70 % des crashes ne sont connus qu'en tant qu'enregistrement SAP (sans géolocalisation ni tronçon identifié). Ainsi, une dernière affectation des défaillances (uniquement celles de 2022) a dû être effectuée de manière quasiment identique à la précédente (Rodriguez et al., 2021 ; Husson et al., 2022). Comme pour la précédente affectation, l'extrait SIG des conduites début 2022 a été utilisé et considéré comme représentatif de 2022. Notons que pour l'affectation 2022, le traitement des fichiers fuites s'est fait à la main par simplicité et rapidité de traitement.

Une différence remarquable concernant l'affectation des défaillances de 2022 est l'apparition d'environ 500 anomalies SIG en 2022 (extrait de début 2023) ayant des ID de tronçons inconnus du fichier de conduites de début 2022 mais connus du fichier de conduites de début 2023 alors qu'il s'agit de conduites posées avant 2022. En réalité, il s'agit de modifications d'ID de conduite qui n'ont pas été prises en compte dans les attributs des anomalies (voir section 4.4 pour des propositions de solutions). Dans ces cas, l'ID du tronçon le plus proche de l'anomalie SIG a été choisi.

1.3 Complétion des données des conduites hors service

Les conduites abandonnées (ou conduites mises hors service, i.e., HS) au même titre que les conduites en service sont une source de données indispensables pour l'analyse de la gestion du patrimoine SWDE. Les conduites HS sont indispensables pour l'affectation des défaillances aux conduites, car elles

peuvent être concernées par des fuites. En pratique, les conduites abandonnées nécessitent des traitements pour pouvoir être utilisées dans le modèle de prédiction des défaillances. Il s'agit notamment de renseigner les attributs obligatoires que sont : MATERIAU, DIAMETRE, DHS = année d'abandon, DDP = année de pose.

Les méthodes automatiques développées sous R pour compléter les données sont décrites en Annexe T01_02 et les résultats transmis à la SWDE sous forme d'une couche SIG (Annexe T01_03). Elles nécessitent d'assurer la connexité du réseau pour propager l'information par les nœuds et une géométrie des tronçons non ambiguë.

1.4 Analyse des codes de cause de panne

Les extraits SAP de réparations de défaillances utilisés dans le projet contiennent une information qui n'a pas été complètement valorisée jusqu'à présent : le code de cause de panne. Précédemment, seuls les codes PM-C310-100, PM-D999-999, et PM-FUITE-999 avaient été écartés car considérés comme n'étant pas des pannes ni des fuites ou crashes (Husson et al., 2022).

Une analyse plus détaillée fournit une liste plus complète des enregistrements SAP qui sont à écarter des jeux de données, car non pertinents. Les groupes de codes suivants ont été créés selon le libellé associé au code et selon le nombre d'utilisations du code pour des types d'activités concernant les pertes d'eau (réparations ou remplacements de conduites, branchements, hydrants, appareils...) :

- indéterminé ;
- pas de panne (notamment les 3 codes ci-dessus) ;
- hors pertes en eau ;
- pertes cause tiers ;
- pertes hors réseau SWDE ;
- pertes autres équipements ;
- pertes sur canalisation et autre ;
- pertes sur canalisation.

Les enregistrements SAP ont donc été marqués comme étant à conserver pour Casses (oui/non) et à conserver pour leur effet sur les pertes (oui/non) selon ces groupes de codes et parfois selon le type d'activité (résultats en Annexe T01_04).

1.5 Nouvelle organisation géographique de la SWDE

Le 1^{er} juillet 2022, la SWDE a adopté une nouvelle organisation. L'impact pour le projet GePaME se limite uniquement à la définition de nouvelles entités géographiques parfois plus pertinentes dans la modélisation et pour la présentation des résultats :

- Les Zones Fonctionnelles (ZF), qui englobent un ensemble de conduites dans lequel l'analyse des risques est pertinente. Les ZF sont des regroupements de ZDE. Il s'agit notamment d'une notion plus hydraulique (fonctionnelle) que les secteurs (basés sur l'organisation des moyens et des employés). La ZF a pour ambition de décloisonner les notions de production et de distribution afin de mieux comprendre le parcours de l'eau depuis son captage jusqu'au client et de répondre plus rapidement en cas de crise par exemple.
- Les entités fonctionnelles (EF), au nombre de 4 (Mons, Charleroi, Namur, Verviers), qui sont des regroupements de ZF.

Bien qu'au nombre de 17 comme les secteurs, les zones fonctionnelles ne redessinent pas les limites des secteurs, malgré une certaine ressemblance des polygones. Même remarque pour les EF, qui ressemblent aux 4 anciennes succursales (A, B, C, et D) mais ne redessinent pas les mêmes polygones.

Ainsi, pour Casses, la notion de ZF et d'EF est conservée au lieu des secteurs et succursales. Pour le modèle d'état fuyard (section 7), les 4 notions sont conservées car la compréhension de l'inspection des tronçons par la RF est plus pertinente au niveau de l'organisation humaine (secteur) que fonctionnelle (ZF), alors que cette dernière est plus pertinente pour l'analyse des facteurs influençant les fuites et casses.

1.6 Correspondances des numéros de cellules entre le SIG et SAP pour la RF

Du point de vue de SAP, la notion de cellule inspectée par la RF n'est pas identique à celle du SIG. En effet, une cellule dans SAP est un « poste technique » servant d'objet de référence pour les OT de RF (OT « systématiques »). Le numéro de cellule associé à ce poste technique n'est pas mis à jour automatiquement comme cela peut être fait pour les objets en relation avec le SIG. Par conséquent, il peut exister des incohérences entre les numéros des cellules réellement inspectées par la RF (numéros de cellule du SIG) et les numéros de cellules sur lesquels les prestations de RF ont été encodées (numéros de cellule SAP). Au cours des traitements de données pour le modèle d'état fuyard, il a été constaté que de telles incohérences se produisent réellement (exemples dans le Tableau 2) :

- Des numéros de cellule dans le SIG n'existent pas dans SAP et inversement ;
- Des numéros de cellule différents sont associés au même nom de cellule.

Tableau 2 : Incohérences entre cellules SAP et SIG

Numéro SIG	Nom	Poste technique	Numéro SAP
56102038	CRISNEE KEMEXHE+FIZE	1000001547	56202038
56104359	FERNEL HINGEON	1000001434	
56102118	OUPEYE BARONHAIE	1000001568	56202118
56102103	OUPEYE RESEAU MOY INF RE	1000001555	56202103
56102039	REMICOURT MOMALLE	1000001548	56202039
56102117	OUPEYE CE RESEAU BAS OUP	1000001567	56202117
56105075	HERON LAVOIR	1000001489	
56102115	OUPEYE SP HACCOURT GAU	1000001565	56202115

Au total, 266 cellules SAP n'existent pas dans le SIG, ce qui signifie qu'il est impossible de rattacher aux cellules du SIG les inspections par la RF sur ces cellules. Une inspection des données par la SWDE a permis de retrouver 92 correspondances par égalité de nom de cellule, et 18 correspondances par des changements connus de numéro de secteur ces dernières années (ex : 5520->5510, 5650-5610) (voir Tableau 2). Une deuxième analyse a permis de retrouver 6 correspondances supplémentaires par égalité approchée des noms de cellule (comparaison des groupes de mots ou de la ressemblance de la chaîne de caractères).

Au final, il reste 110 cellules pour lesquelles les prestations de RF n'ont pas pu être rattachées à des cellules existantes dans le SIG. Ces données sont donc ignorées.

1.7 Notions d'adduction et de distribution

Un enjeu pour plusieurs tâches du projet et de savoir si une conduite est une « distribution », en général de plus petit diamètre et avec des raccordements, ou bien une « adduction », en général de plus gros diamètre, sans abonnés raccordés, et servant uniquement au transport entre ouvrages ou d'un ouvrage à un lieu de distribution. En effet, l'approche par les risques de la SWDE requiert une distinction entre les conduites « normales » (de distribution) et les conduites plus « critiques » (les adductions) où la gestion du risque n'est pas similaire. Les conduites de la SWDE possèdent un attribut « classe » (production, eau brute, branchement industriel, distribution...), mais ce dernier n'est pas suffisant car il traduit un point de vue plus comptable que technique (d'après la SWDE). Notons que la définition des conduites « critiques » de façon durable dans le SI est un enjeu fort perçu par la SWDE.

Ainsi, pour les besoins du projet, un attribut supplémentaire (ADD=0/1 pour adduction non/oui) a été calculé pour chacune des conduites et utilisé pour scinder le jeu de données des tronçons en deux pour la modélisation dans Casses, l'outil multicritère, et l'outil de simulation long terme. Au cours des avancées des tâches concernées, respectivement les tâches 03, 10, et 12, la notion d'adduction/distribution a légèrement évolué car il a été constaté qu'une simplification pouvait avoir lieu dans sa définition.

5 variables binaires traduisant la possibilité qu'une conduite soit une adduction ont été définies :

- classe0/1 : vaut 1 si classe = « adduction mixte » ou « production » ou « eau brute », et 0 sinon ;
- typecell0/1 : vaut 0 si la cellule est de type déclaré « Cellule Distribution », et 1 sinon ;
- numsect0/1 : vaut 1 si le numéro de secteur débute par 4, et 0 sinon ;
- numcell0/1 : vaut 1 si le numéro de cellule est de la forme 4XXXXXXX, et 0 sinon ;
- nomcell0/1 : vaut 1 si le nom de cellule contient « ADD », 0 sinon

Du point de vue des tronçons inclus dans Casses, le fractionnement du jeu de données entre adductions (ADD=1) et distribution (ADD=0) a suivi les règles suivantes (Tableau 3).

Cette définition induit notamment que des conduites de classe production, eau brute, ou adduction mixte de petit diamètre sont modélisées par Casses comme de la distribution. De plus, des conduites abandonnées, des galeries, et des conduites de décharge et de liaison ont été affectées à une catégorie et incluses dans les prédictions avec Casses.

Tableau 3 : Règles employées pour distinguer adduction/distribution pour la tâche 03

classe0/1	typecell0/1	numsect0/1	numcell0/1	nomcell0/1	diamètre	ADD	Commentaire
1	1	1	1	0 ou 1		1	Concerne la plupart des « eaux brute » et « production »
1	0	0	0	0	≥ 250	1	Concerne la plupart des « adductions mixte » au sein de cellules de distribution
0	1	1	1	1	≥ 250	1	Concerne vraisemblablement les conduites à la classe imprécise.
0 ou 1	0 ou 1	0 ou 1	0 ou 1	0 ou 1	≥ 400	1	Toute conduite de très gros diamètre est très probablement une adduction
Tout le reste						0	Contient quasiment toutes les conduites de classe « distribution »

En comparaison, par souci de simplification pour l'interprétation des résultats, pour l'outil multicritère court terme et pour l'outil de simulation long terme, les règles suivantes ont été appliquées (Tableau 4).

Tableau 4 : Règles employées pour distinguer adduction/distribution pour les tâches 10 et 12

classe0/1	typecell0/1	numsect0/1	numcell0/1	nomcell0/1	diamètre	ADD	Commentaire
1	0 ou 1	0 ou 1	0 ou 1	0 ou 1		1	
0	1	1	1	1	≥ 150	1	
0 ou 1	0 ou 1	0 ou 1	0 ou 1	0 ou 1	≥ 400	1	
Branchement industriel						0	
Décharge, liaison, galerie, ou abandonnée						NA	
Tout le reste						0	

Suivant cette nouvelle définition, les conduites ayant NA pour ADD ne sont pas incluses dans les résultats de l'outil court terme ni dans les simulations long terme.

Notons que ce changement des règles entre tâches 03, 10 et 12 pour séparer le jeu de données entre adduction et distribution n'a pas une grande importance puisque dans les deux cas les jeux de données résultants sont proches, et que les prédictions faites par Casses n'en sont que peu affectées (les jeux de données sont larges et les conduites d'une même catégorie possèdent des covariables cohérentes entre elles).

Archive Tâche 01 :

- Dossier codes_et_notices : ensemble des codes utilisés pour le traitement des données et leurs notices
- Dossier conduites : fichier de données sur les conduites
- Dossier Covariables_AH : code pour l'ajout des covariables aux tronçons et aux cellules
- Dossier donnees_rf : fichier de correspondance Poste technique-Cellules
- Dossier fichier_casses : fichier des casses
- Dossier fichier_fuites : fichier des fuites concaténées
- Analyse des données SAP

2 Tâche 02 – La sécurisation par la performance. Analyse sociologique du mode de gouvernance interne à la SWDE

2.1 Introduction

La SWDE constitue le principal opérateur de service d'eau potable de la Wallonie. Elle assure la production et la distribution d'eau potable pour plus de 2,5 millions d'habitants répartis sur 190 communes (cf. Figure 2), ce qui représente 67% de la population régionale et plus de 1,1 millions d'abonnés. Son réseau technique est constitué d'environ 40 000 km de canalisations (dont 27 000 km dédiés à la distribution), et s'étend sur un territoire d'environ 13 000 km² (SWDE, 2022).

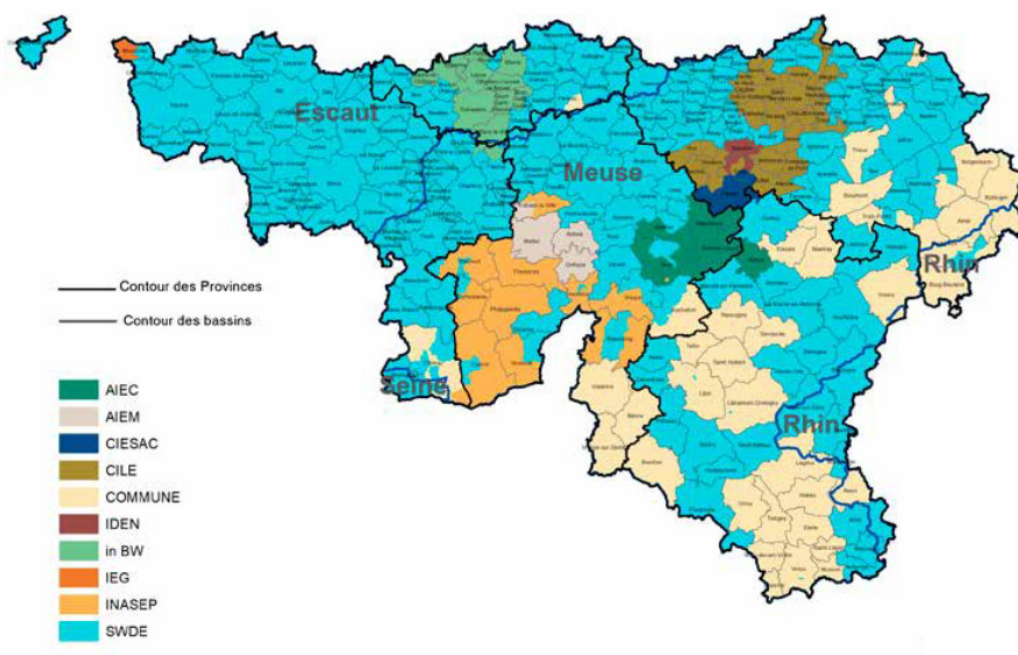


Figure 2. Les sociétés de distribution d'eau potable en Wallonie. Source : (AquaWal 2022)

En 2022, la SWDE a distribué plus de 102 millions m³ d'eau potable, pour un coût – stable dans le temps – de 2,62 €/m³ HT. Son réseau présente un niveau de performance limité lié à l'état fuyard du patrimoine enterré¹ et à un déficit de son renouvellement², malgré l'investissement constant de la Société (environ 117 millions d'euros annuels pour son patrimoine infrastructurel) (SWDE 2019). Un enjeu de performance s'impose ainsi à la Société, d'une part pour réduire ses coûts d'exploitation et maîtriser le prix et la qualité du service, et d'autre part pour répondre à de nouvelles contraintes macro-structurelles liées la sécurisation du service (directive européenne « eau potable » ; décret PoWalCo pour la coordination des travaux sur voirie ; défis socio-écologiques et sanitaires ; coûts des marchés de travaux et d'équipements...).

Dès lors, comment la SWDE s'organise-t-elle pour répondre à ces enjeux ? Quelles sont les logiques, les dynamiques et les facteurs qui y concourent ? À partir d'une approche relevant de la sociologie des

¹ L'indice linéaire de volumes non enregistrés s'élevait en 2019 à 5,42 m³/km/j, contre le seuil souhaité de 4,4 m³/km/j.

² Le taux de réhabilitation des canalisations n'a cessé de diminuer entre 2014 et 2019, atteignant 0,44 % lors de cette dernière année d'exercice.

organisations et du travail, il s'agit ici d'apporter des éléments de compréhension relatifs à la structuration de la Société et à son mode de gestion, afin d'en analyser les effets sur l'organisation du travail et la gestion patrimoniale.

Pour ce faire, nous nous appuyons sur l'hypothèse d'une restructuration de la SWDE visant à « rationaliser » et à « normaliser » son organisation, ses métiers, ses activités et ses outils, dans l'objectif de les « maîtriser »³ et d'atteindre un niveau « optimal » de gestion et de fonctionnement. Nous défendons la thèse d'une montée en puissance d'un « esprit gestionnaire » (Ogien 1995), également appelé « gestionnarisation » (Boussard 2008; Craipeau et Metzger 2011), au sein de la Société. Cette dynamique se caractérise par : un idéal de performance ; des objectifs quantifiés, déclinés en séries d'indicateurs destinés à « objectiver » la connaissance des situations et préparer la décision ; des jugements évaluatifs portant sur la performance globale de l'entreprise et l'activité des agents ; une réorganisation du travail et une évolution des métiers pour améliorer l'activité ; le déploiement d'un mode de management dit « de la qualité » destiné à faire adhérer les agents à la stratégie d'entreprise. Bien qu'elle favorise une certaine vision holistique des enjeux d'exploitation et de maintenance, cette dynamique butte sur des effets sociaux, en imposant un changement de culture du travail (passage d'une logique séquentielle vers une logique métier ; recherche de transversalité de l'action) et en altérant – ne serait-ce que momentanément – les ressources d'action (humaines, expertise, connaissance...) et le sens vécu du travail.

Notre matériel d'enquête repose sur une quinzaine d'entretiens réalisés entre 2021 et 2023 auprès de cadres de différents niveaux (de l'administrateur général au chef de secteur). Ils ont été menés par : Selma Baati, dans le cadre de son post-doctorat relatif à l'activité de recherche de fuites (Annexe T06_01 : Baati and Caillaud, 2023; rapport intermédiaire 2022 : Husson et al., 2022) ; Simon Fidelle, dans le cadre de son stage de M2 sur la gouvernance territoriale de la SWDE (Annexe T02_01 : Fidelle, 2022); Kevin Caillaud, en tant que responsable des tâches 2 et 6 du projet GePaME. Les entretiens ont porté sur : la trajectoire d'évolution de la SWDE ; l'organisation du travail ; le management ; le pilotage de l'activité et ses outils.

Notre analyse est structurée autour de trois parties. La première s'attache à retracer les principales étapes de recomposition de la SWDE, afin de mettre en évidence les logiques de structuration et l'évolution progressive des enjeux du service, conduisant à l'affirmation d'un paradigme gestionnaire centré sur la performance du service. La deuxième partie s'intéresse à la manière dont ce paradigme se traduit aujourd'hui par une approche centrée sur l'anticipation et la prévention, et ce que cela produit comme effet sur l'organisation et le travail. La troisième partie se concentre enfin sur le pilotage de l'activité, notamment en matière de gestion patrimoniale, dans l'objectif de montrer comment le paradigme de la performance structure la manière d'objectiver et d'évaluer le réel et l'activité, et comment les outils agissent sur l'organisation, le travail et les agents.

³ Le terme « maîtriser » renvoie ici à l'ambition de contrôler le système dans sa totalité, afin d'écarter l'ensemble des risques, aléas et incertitudes susceptibles d'affecter le fonctionnement « idéal » de l'activité, notamment la qualité et la continuité du service.

2.2 La Société Wallonne Des Eaux : recomposition des enjeux et reconfigurations du service

2.2.1 De l'hygiénisme local à la sécurisation régionale du service d'eau potable

À l'instar des services d'eau potable en Europe de l'Ouest (Barbier et al. 2016; Barraqué 1995), celui géré par la SWDE s'inscrit dans une histoire séculaire à l'origine liée à un enjeu d'hygiène et de salubrité publique (SWDE 2014). Tandis que les premiers services d'eau potable en réseau sont créés à la fin du 19^e siècle à l'initiative des villes, dans une logique de proximité, une loi est adoptée en 1913 instituant la Société Nationale des Distributions d'Eau. Cette dernière a pour mission de déployer un service public unique desservant l'ensemble du pays, dans l'optique de dépasser les approches « individualistes » et leurs effets de silo.

Dans un contexte de crise économique, la Régionalisation du pays au cours des années 1980 conduit à transférer la compétence Eau potable aux Régions Wallonne et Flamande, et par conséquent à dissoudre la Société Nationale. La Société Wallonne des Distributions d'Eau est ainsi créée en 1986⁴, sous le statut de Société coopérative⁵.

Rapidement, elle est positionnée pour structurer le secteur de l'eau potable à l'échelle de la Wallonie, d'abord par la fusion-absorption d'opérateurs locaux, puis par un travail d'échanges et de coordination avec ses homologues régionaux. Ainsi, à la fin des années 2000, la SWDE atteint un « effet de seuil » (Caillaud 2017), qui se matérialise certes par un attachement de certains élus locaux pour leur service de proximité, mais surtout par une échelle territoriale qui n'apporte plus de marges de progression et d'optimisation dans la gestion du service. Dès lors, l'ambition gouvernementale de faire de la SWDE « l'opérateur unique » de l'eau potable en Wallonie est abandonnée, pour lui préférer davantage une fonction de chef de file via : l'animation de groupes de travail (gestion patrimoniale, systèmes d'informations, recherche de fuites...) au sein de la fédération des opérateurs d'eau AquaWal ; le pilotage du schéma régional des ressources en eau (document stratégique de sécurisation de l'eau potable à l'échelle Régionale) ; l'appui au législateur (préfiguration de textes législatifs ; expertise/conseils).

En parallèle, l'enjeu de sécurisation du (des) service(s) se précise aux plans économique, patrimonial et socio-écologique. Ainsi, dans le contrat de gestion 2012-2017 signé entre la Région et la SWDE, on apprend que la « modernisation et [la] rationalisation des infrastructures » doit être « poursuiv[ie] et amplifi[ée] » (p7-9) à des fins de sécurité d'alimentation et de maîtrise du prix (Région Wallonne et SWDE 2012). L'enjeu économique est d'autant plus marqué au cours du temps, que les contrats successifs d'un côté soulèvent les possibilités d'emprunt auprès de la Banque Européenne d'Investissement, et de l'autre imposent de réduire les coûts de fonctionnement et de générer de nouvelles entrées financières pour mieux les réinvestir dans une politique de gestion patrimoniale (Région Wallonne et SWDE 2022).

Face à ces défis, les dirigeants de la SWDE engagent à partir de 2008 toute une série de « réformes » du modèle organisationnel, technique et professionnel de la Société.

⁴ Elle deviendra la Société Wallonne Des Eaux en 2006.

⁵ Le capital social de la SWDE est aujourd'hui composé de 207 communes, 10 intercommunales, 4 provinces, la Région Wallonne et la Société Publique de Gestion de l'Eau.

2.2.2 À la recherche d'un mode d'organisation et de gestion « performant »

Précisément, quatre principales « réformes » ont été mises en œuvre pour adapter la Société aux nouveaux enjeux socio-écologiques, économiques et réglementaires qu'elle rencontre dans son activité.

Suite à la réglementation de la structure tarifaire des services d'eau et à la mise en place du Plan Comptable de l'Eau par le gouvernement Wallon au mitant des années 2000, la recherche de performance est devenu un enjeu central pour les services d'eau, afin de garantir la stabilité des prix. La première réforme (2008) de la SWDE a alors consisté à déconcentrer son activité d'exploitation, dans le but de rapprocher le service de ses publics et d'adopter une approche globale du petit cycle de l'eau. Concrètement, huit succursales ont été créées sur la base des sous-bassins hydrographiques.

La deuxième réforme (2012), intitulée « *Plan 20 cents* », a explicitement visé la réduction du coût de fonctionnement, dans l'objectif de garantir la stabilité du prix de l'eau (appelé « *coût-vérité-distribution* »). Face à l'incompressibilité des charges d'exploitation liées à la qualité de service, le principal levier d'action identifié pour améliorer la performance globale de l'entreprise fut d'agir sur les ressources humaines, ce qui se traduit par une diminution de l'effectif du personnel (moins 400 agents environ).

La troisième réforme (2017) a quant à elle ciblé l'enjeu de maîtrise des fuites, dans une optique à la fois économique et écologique. Elle est considérée comme une réforme « *structurellement* » importante, puisqu'elle a directement agi sur l'organisation des métiers d'exploitation, de distribution et de maintenance, en créant une équipe de fontainiers (appelée « *cellule de recherche de fuite* ») au sein du Pôle Distribution, dédiée à la localisation des fuites sur le réseau (Annexe T06_01 : Baati and Caillaud, 2023). En professionnalisant ce métier, cette réforme s'est aussi accompagnée d'un déploiement de l'activité de recherche de fuite à l'échelle de la Société : implémentation des systèmes d'information pour accueillir, enregistrer et traiter les données afférentes ; développement des procédures d'intervention ; interactions nouvelles entre équipes pour gérer les fuites ; etc.

La dernière réforme est intervenue en juillet 2022. Elle a redéfini l'organisation générale de la Société, pour passer notamment d'une approche dite « *séquentielle* » fondée sur la séparation des cycles « production » et « distribution » d'eau potable en deux entités à part entière, vers une « *logique métier* ». Celle-ci se concrétise par la constitution d'équipes dont les périmètres de mission et d'activité relèvent d'un même métier, à savoir l'hydraulique opérationnelle Vs la maintenance. L'objectif de cette dernière « *réforme* » vise à améliorer la performance de l'entreprise d'une part grâce à une spécialisation accrue des métiers, et d'autre part au moyen d'une lecture davantage holistique de l'activité du service et la mise en place de nouvelles collaborations inter-équipes.

« [La production et la Distribution] C'est vraiment notre cœur de métier et c'est comme ça qu'on était organisé jusqu'en Juillet [2022]. [...] Ici, avec la nouvelle réorganisation, on a poussé le curseur [de la performance] à un niveau au-dessus, en remettant une logique Processus, une logique Métier, avec l'Exploitation et la Maintenance, alors qu'historiquement les deux métiers cohabitaient dans la plupart des équipes. Ils faisaient de l'exploitation ET de la maintenance. Là, il y a vraiment une volonté – je ne vais pas dire "prise de conscience" – qu'on sera plus efficaces avec des métiers spécialisés : le métier de la maintenance, le métier de l'exploitation. »

Désormais, la SWDE se structure autour d'un Comité de Direction pilotant huit « processus clés » (SWDE 2022), qualifiés par certains de « *processus primaires* », complétés de huit autres plus

« secondaires » (cf. Figure 3). Au plan opérationnel, l'exploitation répond à un découpage en 17 « zones fonctionnelles », qui représentent des portions de territoire « où l'ensemble des éléments du réseau sont hydrauliquement connectés » (SWDE 2020) et « *normalement autonomes ou suffisamment autonomes* ».

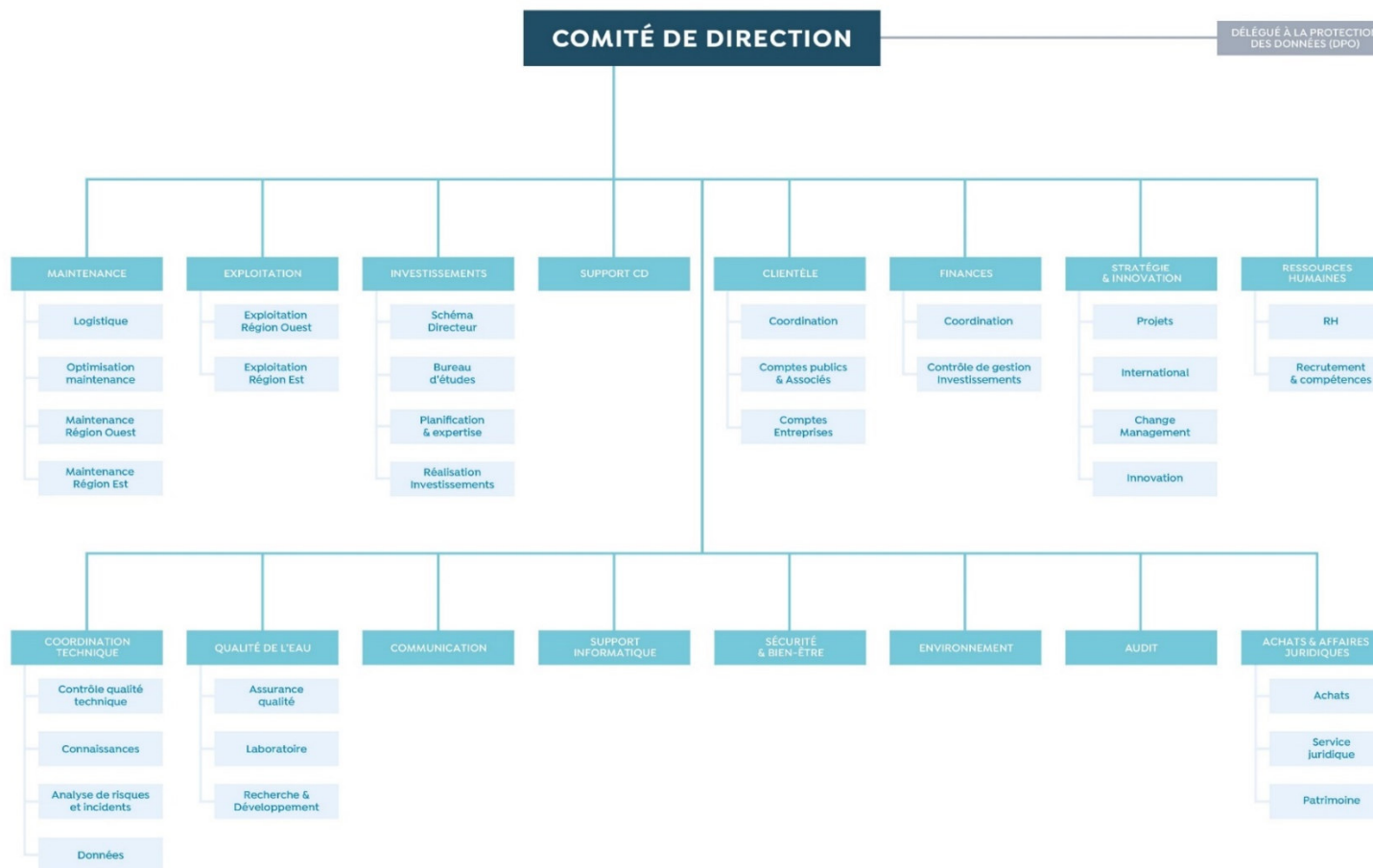


Figure 3. Structuration de la SWDE depuis Juillet 2022. Source : SWDE, <https://www.swde.be/fr/propos-de-la-swde>

2.3 La mise en œuvre du paradigme de la performance

2.3.1 Une stratégie d'entreprise orientée vers l'agir collectif et l'anticipation des dysfonctionnements

La « réforme » de 2022 s'appuie sur une révision de la stratégie d'entreprise, pour faire face aux enjeux socio-écologiques, économiques et réglementaires. Elle a été définie lors d'un séminaire de travail réalisé en 2019-2020, à la demande du Comité de Direction. L'objectif consistait à clarifier les missions des différents organes de la SWDE. Animé par des consultants extérieurs, ce séminaire a mobilisé les directeurs de Pôles (ancienne structuration). Les « missions » identifiées ont donné lieu à huit « Programmes » (cf. Figure 4), auxquels contribuent tout ou partie des Processus de la Société. Véritable « référentiel » (Muller 2005) d'action, ces programmes expriment une ambition à réaliser et impriment une nouvelle ligne directrice et un mode d'action tournés vers l'agir collectif et l'anticipation.



Figure 4. Les Programmes de la SWDE. Source : (SWDE Date inconnue).

Plusieurs éléments ressortent de ces programmes. Dans une optique managériale, le choix des termes répond à la volonté d'établir un récit collectif dans le but de donner du sens à la stratégie d'entreprise et au travail des agents. Chaque programme renvoie ainsi à un slogan qui exprime soit un objectif à atteindre (e.g. « Objectif zéro incident »), soit une valeur (e.g. « Nos données, une autre ressource »), soit une démarche (e.g. « Le client, c'est l'affaire de chacun »), voire qui incarne ces différentes dimensions, comme l'indique un cadre de la Société.

« [Ces programmes] on leur a trouvé des noms qui voulaient dire quelque chose. Par exemple, il y a un programme "Osons un horizon durable". Qu'est-ce que ça veut dire ? Ça veut dire qu'il faut prendre des risques, il faut oser. Pour les investissements, on va faire un pari sur le futur ; sur la base de ce qu'on connaît aujourd'hui, on va se dire ce qu'on doit mettre en place demain. Et "durable", c'est tout ce qui est environnemental, énergétique, etc. Les mots sont importants. Ça veut dire qu'on prend un risque, qu'on va faire de l'innovation, etc. et d'un autre côté ça doit être durable. Et le mot "horizon", ça veut dire qu'on veut voir loin. Ce qui caractérise le processus "investissons aujourd'hui pour demain", c'est un mot d'action

“investir” : on prend une action aujourd’hui, sur base de ce qu’on connaît aujourd’hui, pour que le futur soit meilleur. Donc dans ce processus, il y a une notion d’anticipation et de challenge : il faut faire de la veille, se préparer, etc. Car le cycle d’un projet, chez nous, c’est entre 4 et 7 ans. »

Parmi ces programmes, l’Objectif zéro incident revêt une place spécifique au sein de la nouvelle organisation. C’est – aux dires d’un enquêté – « *le seul programme qui est commun à deux Processus* » puisqu’il mobilise pleinement l’Exploitation et la Maintenance. Ce programme s’appuie en filigrane sur deux notions majeures au sein de la SWDE : « *la gestion du risque* » et l’approche « *qualité* ». Ensemble, ils définissent un nouveau mode opératoire fondé sur l’anticipation des dysfonctionnements, c’est-à-dire sur une action davantage préventive et coordonnée guidée par l’analyse de signaux existants, plutôt que curative et réactive.

« On a un programme zéro incident, et l’équipe de direction a décidé que si on voulait réussir ce programme, il fallait qu’on professionnalise la maintenance. [...] Donc le but de la réorganisation, c’était ça et à travers aussi un outil important qui est la gestion du risque. [...] Et il y a un concept aussi important à la SWDE, c’est la Zone fonctionnelle, [... dans laquelle] on délimite des risques et donc ces zones fonctionnelles vont guider nos investissements, notre maintenance et notre façon d’exploiter. »

Concrètement, se mettent en place des réunions tri ou quadripartites, mobilisant les responsables de l’Exploitation, de la Maintenance, de l’Investissement voire de la Coordination technique, afin d’établir des plans d’action tenant compte des enjeux de gestion patrimoniale et de continuité de service.

2.3.2 L’agir collectif : un travail de définition et d’écriture compliqué

Un préalable à la coordination des équipes consiste à créer les conditions d’un agir collectif, à savoir constituer et partager en premier lieu un socle commun de connaissances, d’objectifs et de règles permettant de tenir compte des logiques, des ressources et des contraintes d’action de chaque partie prenante, et de les faire converger vers la prise en charge d’un même enjeu opérationnel. Or ce préalable n’est pas si anodin, surtout lorsque la collaboration inter-équipes n’est pas à l’origine aussi centrale dans la culture professionnelle de l’entreprise, du fait du séquençage historique de l’activité.

« Ces logiques en processus un peu hand to hand métier, vont obliger beaucoup plus de collaboration entre l’Exploitation et la Maintenance, là où Production et Distribution devaient quand même collaborer à minima mais gérant l’ensemble de son parc d’asset et avec une logique plus géographique, ça permettait réellement moins de collaboration. Ici, le manque de collaboration conduira à des clashes quotidiens. Si les actions de maintenance nécessitent de mettre certains trucs hors service, et si l’exploitation n’est pas au courant, ça n’ira pas. Et inversement. »

Chaque équipe n’étant certes pas étrangère aux métiers des autres, c’est néanmoins une culture professionnelle qui est amenée à évoluer, et pour laquelle il est nécessaire de « *redéfinir des rôles et responsabilités, organiser des transferts de compétences* », etc. C’est donc un important travail de réécriture et de convention qui est engagé, pour faire advenir la nouvelle organisation et les interrelations en son sein. Cette réécriture porte principalement sur trois domaines socio-cognitifs et opérationnels de la Société : les notions fondamentales ; les indicateurs centraux du suivi stratégique ; les périmètres de juridiction et leurs « flux » relationnels.

Les notions fondamentales renvoient au socle conceptuel structurant l'ambition politique et l'organisation du travail autour d'un ensemble de représentations et de savoirs, comme on le voit avec le programme Objectif zéro incident. Tout d'abord, le terme même d'« incident » exige d'être précisé pour distinguer les divers événements et leurs niveaux d'impacts sur le fonctionnement du service. Les cadres de la SWDE concernés par ce programme s'accordent alors pour définir l'incident comme renvoyant à tout « *dysfonctionnement avec impact sur le client* ». C'est donc l'insatisfaction (totale ou partielle) de l'utilisateur final qui accorde le statut d'incident, contrairement aux « *dysfonctionnements d'un équipement sans impact sur le client* » qui, pour leur part, sont qualifiés de « *défaillances* »⁶. Lorsqu'il se répète, l'incident change de statut et devient un « *problème* ». Les notions de « *risque* » et de « *qualité* » font elles aussi l'objet de définitions (plus vagues), en intégrant dans le système d'évaluation les indicateurs réglementaires, le statut des consommateurs et d'autres facteurs macro-structurels (le climat, l'état des ressources en eau, etc.).

Le deuxième domaine de définition et d'écriture a pour objectif de faire exister ces notions dans la sphère gestionnaire, pour être en mesure de les enregistrer, de les comptabiliser et de les évaluer. Cela repose donc sur un travail de création d'indicateurs et de données, pour lesquels la quantification des observations sont sources d'objectivation et vecteurs d'arbitrage, tout en reposant sur des « *conventions sociales* » (Desrosières 2014). C'est d'autant plus important suite au changement de paradigme intervenu en 2022, que la gestion patrimoniale et le Zéro incident doivent désormais reposer sur une analyse de risques, qui s'accommode difficilement du seul vécu des agents ou de leur « *sens des réalités* » (Berlin 2011). Ce travail s'avère enfin primordial, du fait qu'il conditionne également – pour partie – l'infrastructure technique de travail et les dispositifs de suivi qui devront être déployés et utilisés pour faire exister et traiter les données.

« On est en phase prospective donc on essaie de voir quelle donnée a un impact sur la qualité de l'eau. Ça, c'est plutôt les données contextuelles. Et après, il y a toutes les données d'incidents ?, de problèmes ?, de défaillances ?. Et là, on est en train de mettre en place un système d'enregistrement [...]. C'est important car c'est quelque chose qui va alimenter notre analyse de risques, qui aujourd'hui, il faut le dire : quand on définit la probabilité d'un incident, on la définit un peu au doigt mouillé, avec les gens du terrain : est-ce que c'est important ou pas, est-ce que c'est arrivé beaucoup de fois, pas beaucoup de fois, plus d'une fois par semaine, moins. Et après, lorsqu'on aura une base de données de 4-5 ans, on pourra avoir quelque chose de plus solide. Mais d'abord, c'est comment on enregistre, à quel niveau de précision on enregistre. [...]. Aujourd'hui, on a construit un système sur lequel on a une évaluation par rapport au vécu des gens mais des fois ça reste assez subjectif. Donc demain, ce qu'il va falloir faire, c'est d'objectiver ça par de la donnée, bien enregistrée au bon équipement, etc. Par contre, on a une analyse objective sur les événements dangereux car on les connaît et ils sont enregistrés, donc ça, ça va. »

Le troisième domaine renvoie aux rôles et aux responsabilités des différents acteurs impliqués dans le programme Objectif Zéro incident, et par extension dans l'activité de gestion patrimoniale. Autrement dit, il s'agit de clarifier les « *périmètres de juridiction* » (Abbott 1988), afin d'attribuer à chaque organe opérationnel et à chaque agent une feuille de route précise de l'étendue de ses missions. Outre la

⁶ Cela correspond par exemple à une vidange inopinée d'un château d'eau n'entraînant pas de coupure d'eau ou de baisse de pression chez les clients.

recherche de cohérence d'ensemble des missions attribuées à chaque corps de métier, ce travail de définition est d'autant plus sensible qu'il touche directement à l'identité professionnelle des agents, au sens de leur travail et à l'allocation de moyens. Le sentiment de dépossession ou au contraire de surcharge peut alors être exprimé et peser dans la négociation des conditions d'exercice.

« Parfois, ça s'apparente à des peurs de perte de pouvoir : “oui mais je ne ferais plus ça...” ben oui, mais est-ce que c'était vraiment à toi de le faire ? [...] il y a aussi parfois des discussions tendues, ça fritte un peu : “ça c'est toi, ce n'est pas moi, ce n'est pas de la maintenance, c'est de l'exploitation”, “mais non, c'est pas de l'exploitation, ça n'a pas trait aux flux d'eau”, “moi je veux bien le faire mais il faut me rajouter quinze personnes.” »

L'établissement des périmètres conduit de manière complémentaire à identifier les interactions avec d'autres processus, et par conséquent à définir les « flux », c'est-à-dire l'ensemble des relations et des collaborations, leur sens et leur contenu.

« Par exemple, un des ateliers [de redéfinition des flux] c'était la manœuvre des vannes, donc qui manœuvre les vannes ? Dans la logique, on peut se dire : qui en a la responsabilité ? C'est l'exploitant parce qu'il est responsable du flux [d'eau]. [...] Il a la vision globale de l'eau qui circule dans les conduites. Mais certains disent “oui mais à la Maintenance, ils ont la connaissance, ils sont en nombre, ils sont plus proches du terrain donc est-ce qu'ils ne pourraient pas aller le faire ?” [...] J'aurais tendance à dire “d'accord mais il faut que cela se fasse sous la responsabilité de l'Exploitation. Donc comment est-ce que vous déléguez ? Est-ce que vous venez et on le fait sur place ? Est-ce que vous donnez un plan, etc. ? [...] Et puis est venu sur le tapis “qui manœuvre les vannes dans le cas d'un renouvellement d'une grosse conduite ?” Car là, on n'est plus dans le processus Exploitation ou Maintenance, mais dans le processus Investissement... »

Enfin, dans une optique d'appropriation sociale et de mise en œuvre par les agents, les flux sont traduits en schémas, en notices et/ou en protocoles, ce qui a pour effet de formaliser, d'ordonner et de standardiser les relations. En d'autres termes, l'écriture des périmètres de juridiction et des flux « travaille l'organisation » (Cochoy, Garel, et de Terssac 1998), ce qui impose toutefois de poser certaines limites, notamment celle d'accepter que des « vides » ou des « impensés » demeurent au risque de créer ultérieurement des situations inédites, afin d'éviter une surcharge protocolaire contraire à la fluidité de l'action.

« Depuis une petite dizaine d'années, c'est vrai qu'on a amené pas mal de choses, en passant d'abord par les certifications ISO qui ont quand même amené des contraintes. [...] il a d'ailleurs parfois fallu le rappeler parce que avec la dernière réorganisation... pensez flux, pensez simple ! [...] on a refait tous les flux : les vues sur maintenant, les vues sur demain, on en a fait des km. En fait, on dessinait les flux et on devait les redessiner pour les simplifier, éviter les pertes de temps dans certains flux, etc. [...] car certains étaient jusqu'au-boutistes, à vouloir aller jusqu'à préciser la situation exceptionnelle... »

2.3.3 Une « rationalisation » inachevée

Alors que la nouvelle organisation ambitionne une certaine « rationalisation » structurelle et opérationnelle de la Société, la démarche apparaît à ce stade inachevée.

La nouvelle organisation est ainsi ressentie par certains comme étant « beaucoup plus complexe, beaucoup plus enchevêtrée » du fait des nombreuses interactions entre équipes. Le renforcement des collaborations impose par ailleurs de créer de nouveaux espaces d'échanges pour favoriser

l'interconnaissance et la coordination des équipes et des activités. C'est ainsi qu'un « Groupe Techn'Eau », composé des directeurs des Processus Exploitation, Maintenance, Investissement et Coordination technique, s'est constitué « un peu en marge du reste » pour stabiliser et réguler de manière autonome leur nouvel univers commun d'exercice.

« Ça n'a pas été vu d'un bon œil au début que les processus techniques se réunissent entre eux. Donc tous les lundis, on se voit parce qu'on a des besoins de coordination entre l'Investissement, la Maintenance, l'Exploitation. Car on est incapable de vivre tout seul. [...] Donc on s'était dit qu'on créerait bien un petit groupe à quatre, pour se parler de choses communes qu'on a à régler. [...] Il y a quelques semaines, on a fait nos objectifs croisés car on s'est rendu compte qu'on ne pouvait pas avoir d'objectifs tout seul en maintenance, en exploitation, en investissement sans se préoccuper des autres. Par exemple, en Investissement, leurs objectifs sont hyper dépendants de nous parce que c'est nous qui leur fournissons les besoins d'investissement. Si on ne les alimente pas en besoin d'investissement, ils sont au chômage. [...] si on ne leur fournit pas nos besoins d'investissement avec un bon degré de précision, les études prennent un temps fou ou on nous étudie des choses qui ne sont pas intéressantes pour nous. Donc c'est de tout cela qu'on discute dans ce fameux Groupe Techn'Eau, et de plein d'autres choses. [...] Mais ça, ça a été très mal vu par certains [...] ça a été perçu au début comme un État dans l'État, un contre-pouvoir : les techniciens se liguent contre les autres. »

Cet espace spécifique est reproduit à l'échelle des centres opérationnels, pour programmer et coordonner les actions sur le réseau dans le cadre de réunions hebdomadaires ou bimensuelles. La communication et l'échange d'informations sont ainsi primordiales pour éviter que des interventions ne créent davantage de dysfonctionnements, voire d'incidents pour les usagers.

« Avec la réorganisation, le gros challenge c'est quand même la passation de connaissance, le fait d'avoir les mêmes logiques et que les communications puissent se faire. [...] Donc ce qu'on va discuter le plus, c'est tout ce qui va être stratégique. Par exemple, le remplacement d'un point de comptage va être important parce que derrière il y a beaucoup d'organisation puisque ça va impacter potentiellement un ouvrage, la clientèle, etc. [...] donc à la réunion, on va parler de ce qui bloque, de tout ce qui pose problème. »

Ces groupes de travail témoignent ainsi d'une organisation qui, si ce n'est cherche son fonctionnement, au moins demeure dans une phase de consolidation. Par exemple, les interventions de première ligne lors d'appels clients soulèvent encore aujourd'hui des interrogations sur l'attribution des responsabilités inter-équipes, du fait d'un manque de personnel pour internaliser durablement l'activité au sein du processus Exploitation. En complément, la connaissance de terrain doit être retransférée en partie vers les équipes du processus d'Exploitation, pour accroître l'efficacité de leurs interventions. De nombreux outils restent aussi à construire ou à parfaire, pour que l'activité globale de la Société redevienne fluide et qu'elle se routinise. Ainsi, le Processus Maintenance s'est donné pour mission de dresser un inventaire exhaustif de l'ensemble des équipements d'exploitation, pour être en mesure, à terme, de programmer de manière préventive leurs besoins de maintenance, au même titre que pour les canalisations. Il s'agit donc avant tout de faire exister officiellement et objectivement ce patrimoine, pour pouvoir ensuite lui appliquer le même mode de gestion qu'au réseau souterrain. Dans le même ordre d'idée, l'infrastructure numérique de la Société, pilotée par le processus Coordination Technique, est amenée à se perfectionner, pour qu'elle puisse pleinement jouer son rôle d'appui à l'action : le système d'information géographique doit être amélioré, les données doivent être plus précises et standardisées, les bases de données davantage toilettées, etc.

« Il faudrait que les équipes d'exploitation et les analystes commencent à s'approprier un peu plus le réseau, y compris pour les équipes de performance, que le SIG soit le plus précis et complet pour que cela aide la Maintenance à comprendre comment fonctionne le réseau, et que petit à petit ça libère un peu ces équipes de cela. »

« [Pour ce qui est de l'intégration par exemple du logiciel Casses de INRAE] on est aujourd'hui en maturité 2 par rapport à une maturité 4, qui est pour moi une maturité bien construite. On a passé le premier cap qui était des données non maîtrisées. Aujourd'hui, on maîtrise les données, on va faire des plans d'action de contrôle des données. Demain, le grand défi sera d'avoir des données qui rentrent dans le système, qui auront une valeur ultime plus grande qu'aujourd'hui. Donc il faut convaincre, persuader les gens de terrain qu'on ne peut avoir des résultats intéressants que si au départ on a une donnée super intéressante et que le service data a moins de travail à faire sur la donnée. Car pour moi, ce serait magnifique que ce service ne soit là que pour contrôler la donnée, pas pour l'enrichir ou la corriger. »

2.4 Un pilotage par les indicateurs

2.4.1 Chiffrer la performance : de la mesure du challenge à l'épuisement socio-professionnel

L'impératif de performance apparu au cours des années 2000 fait rapidement émerger un besoin d'évaluer le service et son activité, pour identifier des marges de manœuvre et des pistes d'amélioration.

Une première démarche initiée par le Comité de Direction fut de se mesurer à d'autres services, pour comparer les résultats. Dès la fin des années 2000, la Société s'inscrit donc dans un travail de benchmark auprès d'opérateurs Wallons et Européens, qui aura pour premier effet d'accompagner l'élaboration du Plan 20 cents destiné à réduire les coûts de fonctionnement de la SWDE.

« On a fait un benchmark, on participe au benchmark avec plein de sociétés qui sont différentes de la nôtre. Et puis le Comité de Direction a commandé un benchmark avec des entreprises qui avaient été identifiées comme étant assez proches de nous : réseau vieillissant, autant de fonte, qualité de l'eau ok. Et c'est à ce moment-là qu'on a défini les 20 cents de réduction de running cost. »

Ce Plan va s'avérer particulièrement structurant dans la trajectoire d'évolution de la Société, et dans la manière de piloter son activité. En effet, l'utilisation du benchmark comme outil phare de l'évaluation de la performance d'entreprise induit deux effets majeurs (Bruno et Didier 2015) : un pilotage par le chiffre et la statistique ; la montée en puissance d'une logique concurrentielle.

Pour se mesurer puis se comparer aux autres, l'objectivation du chiffre apparaît comme un rempart au jugement subjectif et par conséquent comme un vecteur de connaissance indiscutable (Salais 2010). L'activité est donc révélée par la construction et l'utilisation d'indicateurs *a priori* impartiaux, fondés sur la raison statistique : taux de renouvellement, nombre de fuites trouvées, indice linéaire de perte, volumes retrouvés, etc. Ces indicateurs sont par ailleurs imaginés et construits par les experts du domaine, que sont en l'occurrence les directeurs de processus et leurs équipes d'ingénieurs. La légitimité des indicateurs repose donc sur l'expertise technique de leurs créateurs, et les valeurs cibles sont identifiées par un double jugement, à savoir le positionnement dans le benchmark et l'estimation de marges de manœuvre empiriques pressenties et *a priori* réalistes.

« [La définition des objectifs] c'est encore très top-down [...]. Fondamentalement, c'est le top management qui décline la stratégie élaborée par le Comité de Direction, qui est très vague, et donc c'est le top

management qui décide sur quels critères il est important de travailler. [...] La valeur des indicateurs, c'est vraiment fixé globalement avec les directeurs de processus. C'est le directeur de processus, à partir de son expertise technique, qui fixe. Après, il y a un apport du benchmark mais c'est ça qui nous fait faire le choix d'un niveau de cible à atteindre. Par exemple, les fuites trouvées par jour, on s'était rendu compte qu'on était à 0,7, on savait qu'on avait un potentiel de 1, donc c'est nous qui avons dit en Comité de Direction que c'était 1. Le 4,40 en ILP [Indice Linéaire de Perte], c'était plus un truc de retour sur investissement qui a été calculé et on s'est dit qu'on pouvait avoir ça comme performance, par rapport aux bonnes pratiques. Ce n'est pas une approche qui a été tirée au sort. Ce sont les directeurs de processus avec les experts de la société qui définissent des niveaux. »

Or le verbatim précédent l'indique : la démarche de construction des indicateurs et des cibles a pour effet d'introduire de la normativité (rationalité purement technique, benchmark, « bonnes pratiques ») et de la subjectivité (estimation des marges de manœuvre et des gains possibles) dans l'établissement des outils de mesure et des objectifs. Moins neutres qu'en apparence, ces outils expriment en arrière-plan des visions spécifiques de ce que devrait être un service performant, ce qui leur confère par conséquent un pouvoir performatif (Denis 2006). De plus, le caractère en partie estimatoire des marges de manœuvre contribue à forger une culture du « *challenge* » au sein de la Société, ne serait-ce qu'à l'échelle des cadres et ingénieurs. En effet, face au défi de maîtriser le coût du service dans un contexte socio-écologique, économique et réglementaire en tension, la démarche qui consistait à améliorer la performance globale de l'entreprise tend à prendre la forme d'un cercle infernal, voire d'une fuite en avant, puisque les équipes se retrouvent à « *se challenger en permanence* ». « *On est pris au jeu* », avoue un cadre : les projets s'enchaînent, les objectifs se renforcent, les pistes de travail se multiplient, les réorganisations se succèdent.

Cette dynamique a alors pour effet (pervers) d'un côté de revaloriser l'entreprise et le travail (et l'investissement) des agents, et de l'autre de les dévitaliser par une dégradation ressentie des conditions d'exercice et l'apparition d'un épuisement professionnel. Dès lors, un dilemme nouveau se pose aux cadres dirigeants, qui vise à concilier les logiques techniques et gestionnaires liées au niveau d'exigence et de performance avec celle davantage sociale ayant trait aux risques psycho-sociaux.

« Aujourd'hui, on est dans un moment critique car on sent qu'il y a une fatigue qui est présente. [...] Par exemple, on a beaucoup de jeunes cadres qui quittent leur boulot, alors que ça n'arrivait pas avant. Mais comme les gens qui sont chez nous sont des gens intéressants sur le marché du travail et qu'on s'intéresse à pas mal de trucs intéressants, avant on était inconnu ou plutôt connu comme une administration un peu ringarde. Mais maintenant qu'on est un peu plus sexy... je pense qu'on est pris à notre propre jeu de se challenger et de faire bouger sans cesse les choses. [...] Mais la question que je me pose, c'est donc quel est le juste milieu entre inertie et changement ? Parce que c'est vrai qu'en dix ans, ça s'est dégradé [les conditions de travail et le bien-être dans l'entreprise]. Mais a contrario, si on regarde du côté technique, il n'y a jamais eu autant de chouettes projets à traiter. »

2.4.2 Automatiser les décisions : une désocialisation progressive au terrain

Outre le chiffrage de la performance, la SWDE s'équipe de données contextuelles et de fonctionnement dans le but de « rationaliser » son activité, et notamment sa gestion patrimoniale. Ces données ont pour fonction de dresser un diagnostic du patrimoine infrastructurel et des enjeux territoriaux, en vue de renforcer la pertinence et la cohérence des actions d'exploitation, de maintenance et d'investissement. Avec le programme Objectif zéro incident et l'approche par le risque, la stratégie de gestion patrimoniale

est donc moins repositionnée pour servir le maintien global de l'état du réseau que pour améliorer la situation des sites en tensions, c'est-à-dire ceux présentant d'importantes fragilités systémiques et pour lesquels l'impact clientèle est tangible. Ce faisant, il s'agit désormais de « territorialiser » (Offner 2006) davantage la gestion patrimoniale, au risque de différencier les efforts de renouvellement des infrastructures selon les zones fonctionnelles.

« Ce que l'on souhaite, c'est "est-ce qu'on améliore ou pas la situation ?" Cette question va avoir un impact sur la priorisation. Donc si on a le choix entre deux mêmes travaux mais dont l'un va améliorer la situation et répondre à des difficultés locales, alors c'est celui-là qui sera sans doute mis en priorité par rapport à l'autre. Par exemple, on a une conduite de diamètre 80 et de temps en temps le week-end, on a des pertes de pression parce que le débit que les gens consomment est beaucoup trop grand, parce qu'en fait le diamètre 80 est sous-dimensionné, alors on va choisir de remettre ici une 100 ou 125, plutôt que d'en remettre une 80 là où il y en a déjà une qui ne pose pas de problème aux usagers. »

Pour y parvenir, des critères de criticité sont mobilisés afin de hiérarchiser les priorités de renouvellement, selon leur degré d'intensité. Trois catégories d'indicateurs issus de l'activité des processus Exploitation et Maintenance, ainsi que de la réglementation, sont utilisés pour établir l'indice de criticité des tronçons et évaluer sa priorité. Les indicateurs qualitatifs et quantitatifs de fonctionnement du service (critères PGSSE, nombre d'événements, mesures d'impacts, volumes affectés, etc.) servent ainsi à rendre visible et à mesurer la fréquence et le poids des incidents provoqués par tronçon. Les indicateurs davantage contextuels (concomitance de travaux, présence de clients sensibles, etc.) ont pour objectif de révéler le degré d'« opportunité »⁷ d'un potentiel renouvellement du tronçon. Ces différents indicateurs alimentent une matrice *ad hoc* de la SWDE, qui attribue des « scores » d'intensité aux tronçons du réseau, 1 correspondant à une criticité minimale, 5 à un taux maximal appelant un remplacement.

Ce faisant, cet outil constitue une aide à la décision, en proposant de hiérarchiser les travaux à planifier, sur la base d'une analyse des enjeux et des risques. Surtout, il témoigne d'une volonté d'automatiser autant que faire se peut ce travail de priorisation, en accordant au calcul matriciel des données de terrain la valeur prévalente de l'évaluation des enjeux et de l'intérêt à agir⁸. Par extension, il exprime en filigrane une érosion de la confiance (Chiapello et al. 2015) à l'endroit de la capacité de discernement et de jugement « rationnels » des individus, sur la base de leur expérience de terrain : « Il suffit que la personne a eu deux gros crashes il n'y a pas longtemps, donc directement elle aura l'impression que ce truc-là arrive souvent. »

Il en résulte un effet de désocialisation, voire de déshumanisation de la gestion patrimoniale, en faisant prévaloir l'outil et le calcul au détriment des logiques et des perceptions sociales et professionnelles. Non seulement l'outil contraint le pouvoir discrétionnaire des agents, mais il favorise également une forme d'« administration »⁹ de l'activité et du travail, en imposant de plus en plus aux agents une tâche de

⁷ Le terme « opportunité » est à entendre dans un double sens : la circonstance rendant l'action plus favorable (e.g. la concomitance de travaux sur la voirie) ; ce qui est adapté à la situation (e.g. une vétusté du tronçon entraînant de nombreux dysfonctionnements).

⁸ L'outil multicritère développé dans le projet GePaME est également censé servir cette finalité et répondre à cette logique.

⁹ Au sens d'action visant à gérer l'activité de manière impersonnelle.

documentation normalisée (comprenant la constitution de données standardisées) dont le contenu et l'intérêt n'est pas toujours évident du point de vue strictement opérationnel. Cette administration du travail et de la donnée est vécue de manière d'autant plus normative par certains agents (notamment les chercheurs de fuites – Annexe T06_01 : Baati and Caillaud, 2023), qu'elle est pour partie centralisée et contrôlée par le processus Coordination Technique, qui entretient un rapport plus distancié au travail opérationnel de terrain que les processus Exploitation et Maintenance.

Dès lors, l'administration du travail et de la donnée contribue à éloigner les agents du terrain, en les enfermant dans des démarches « comptables » destinées avant tout à alimenter le système. C'est donc une forme de déconnexion au réel qui s'opère, comme le reconnaissent certains enquêtés. Ce phénomène n'est cependant pas le fait du « chiffre en soi [... mais d'] une certaine pratique du chiffrage, [c'est-à-dire que] ce sont les conditions socio-organisationnelles de la *valuation* » (Chiapello et al. 2015) qui déconnectent l'individu. Partant, la déconnexion au réel s'accompagne d'une dégradation du savoir empirique.

« Aujourd'hui, les agents sont moins sur le terrain, ils font beaucoup plus de papier. Ils sentent moins les choses et ils sont moins nombreux aussi. Donc ils ont plus de mal à exprimer leurs besoins d'investissement. »

Or, ce phénomène conjoint de déconnexion et de dégradation du savoir risque de s'accroître en cas de technologisation poussée du travail d'exploitation du service, ce qui pose de nouveau la question du sens de la donnée et de son usage dans l'opérationnalisation des missions de la SWDE, comme l'expriment certains cadres.

« Je suis très sceptique sur tout ce qui est connecté, connectable, IT [Intelligence Technology]. On travaille dans le sol, dans l'inconnu, sur ce qu'on ne voit pas. C'est quand on ouvre qu'on se rend compte que la conduite est là ou pas. On a des réseaux pour lesquels on n'a pas de plan parce qu'on a racheté le réseau d'une commune qui n'avait pas de plan. [...] et pour faire des plans, il faut faire du sondage et faire des trous. [...] on n'a pas à aller sur Mars. [...] On avait fait un projet il y a deux ans, pour faire des jumeaux numériques de nos réseaux. [...] on s'était dit qu'on ferait un calage automatique en temps réel pour avoir toujours un modèle auto-calé. [...] et on s'est rendu compte qu'il fallait 80 personnes simplement pour faire l'entretien des équipements qu'on allait placer. C'est énorme. Et le 2^e point, c'est les données : qu'est-ce qu'on va avoir comme données et qu'est-ce qu'on va en faire ? Le vendeur de capteur vous dit "on peut faire une mesure toutes les secondes". Mais qu'est-ce que vous allez faire avec ça ? »

« Ce qu'il faut, c'est de bien étudier la gouvernance de la donnée, qui est responsable de quoi entre l'IT, le data, le business. [...] Il ne faut pas partir dans un format boîte noire. Il faut garder un contact avec la donnée et pas que ce soit des modèles obscurs, des boîtes noires qui sortent des résultats dont personne ne sait d'où ça vient. [...] Et il faut persuader les personnes que l'info qu'elles transmettent, elles la collectent bien. Mais pour ça, on ne doit pas leur demander de collecter 36 infos. On doit vraiment être dans le minimalisme, aller à l'essentiel. »

Cette désocialisation liée à l'emprise de la donnée et à l'automatisation des décisions peut s'accroître et être vécue sous la forme d'une déshumanisation, en rendant la gestion patrimoniale sourde à l'endroit des attentes des acteurs de terrain. En effet, en écartant des espaces de concertation et de décision les bourgmestres et autres acteurs chargés de l'aménagement du territoire, et en les remplaçant par une

analyse froide des données de réseau, la gestion patrimoniale devient hermétique à toute subjectivité locale, qui pourtant constitue un ingrédient fondamental de la vie locale et de la réception du service.

« Il se pourrait qu’il y ait maintenant de fortes disparités d’investissement, beaucoup plus fortes qu’avant. Avant, même s’il y avait des différences, il y avait quand même un saupoudrage dans l’idée de ne pas choquer les bourgmestres. [...] Demain, il faudra vraiment faire attention à bien expliquer au terrain pourquoi on investit dans une zone plutôt qu’une autre. Parce qu’on peut très bien avoir une zone où le bourgmestre dit “moi, j’ai ma conduite qui pète tout le temps !” Mais si ce n’est pas une zone à risque important et qu’on fait plutôt des travaux dans la zone fonctionnelle d’à côté, il faudra faire un effort pédagogique. On devra avoir de la transparence et une vue sur un peu plus long terme, en disant “toi, on peut attendre 3-4 ans pour faire”. »

2.4.3 L’implication managériale en soutien du chiffre

Le fonctionnement de la Société et son pilotage par le chiffre reposent enfin sur un important travail de management, afin de favoriser l’appropriation sociale et la mise en œuvre de la stratégie d’entreprise, et pour lever d’éventuelles résistances.

Très tôt, une démarche participative s’est donc installée dans la Société, à des fins opérationnelles et de gestion. Cette démarche a progressivement été précisée et généralisée, au point de constituer aujourd’hui un marqueur de la culture et de l’identité managériale. L’« animation managériale » est ainsi un thème central du discours des acteurs, lorsqu’il s’agit d’évoquer les relations de travail au sein des Processus et la gestion des équipes.

« C’est vrai qu’on a beaucoup insisté sur l’animation managériale ; on en fait beaucoup mais je suis un vieux de la vieille et il y a trente ans, on ne parlait pas du tout de ça mais j’avais des réunions avec les chefs de secteur, où on examinait ensemble les données, on leur donnait des informations, on leur donnait des indicateurs. C’est vrai, on avait un peu moins d’indicateurs [...] mais on faisait des réunions. On n’appelait pas ça RMO [Revue de Management Opérationnel] mais réunion du lundi. »

À partir des années 2000, la recherche d’optimisation et de performance a par ailleurs conduit à doubler la restructuration de la Société, d’une normalisation de ses pratiques de travail et de son management. En obtenant la certification ISO 9001, la SWDE a ainsi repositionné son organisation et son fonctionnement pour épouser les principes du management dit « de qualité »¹⁰, à savoir : l’orientation client ; la responsabilité des équipes de direction ; la collaboration ; l’approche par « processus » ; la recherche d’amélioration ; l’objectivation des décisions (preuves) ; la vision systémique.

« Parfois on a l’impression qu’on a découvert la pierre philosophale mais on n’a rien découvert du tout. On a juste amélioré le système. Et l’ISO est venu renforcer ces besoins de structuration. Enfin, ça nous a obligé – parce que c’est obligatoire dans les normes – de faire régulièrement des revues [d’indicateurs] à différents niveaux. »

Guidées par les besoins de l’action, la démarche « lean », la « cascade managériale » et les « revues » sont ainsi des dispositifs devenus centraux du management de la Société, mais qui sont à l’origine conçus pour accroître la compétitivité des entreprises sur les marchés internationaux. La rationalité économique et

¹⁰ <https://www.france-certification.com/les-certifications/iso-9001/>

gestionnaire en constitue donc l'essence, ce qui par conséquent renforce le caractère normatif de ces dispositifs.

La démarche lean provient plus précisément du modèle de « l'organisation flexible », pour lequel « la compétitivité tient surtout à la capacité d'innover, de diversifier, d'introduire rapidement un changement qualitatif plus ou moins important dans la production » (Stroobants 2009), en l'occurrence ici : la production du service. Ce modèle d'organisation prône notamment les nouvelles technologies et la fluidification de l'action pour améliorer la performance. Les chaînes de production sont appelées à réduire au maximum les stocks, les flux, les délais, les pannes, les défauts, le papier¹¹. C'est ce qui s'est traduit à la SWDE par l'objectif zéro incident, la révision des « flux » relationnels, le déploiement des nouvelles technologies, le pilotage par les données chiffrées, etc.

Les marges de manœuvre sur lesquelles ce type d'organisation cherche à agir, sont alors identifiées par les principaux agents de production (les ouvriers) et leurs hiérarchies, dans le cadre d'ateliers de réflexion. En effet, non seulement ce sont les plus à mêmes d'exprimer un point de vue circonstancié sur les situations de travail et l'activité au quotidien pour proposer des modifications réalistes, mais cette approche *bottom-up* favorise en outre l'appropriation des mesures, du fait qu'elles émanent de la base plus qu'elles ne sont imposées par le haut. C'est donc une manière d' enrôler les agents dans la transformation de leurs propres conditions de travail (Bernoux 2004). À la SWDE, ces ateliers ont par exemple accompagné l'élaboration des différentes « réformes » initiées depuis 2012. Ils ont également été mis en œuvre lors de la dernière réorganisation, afin de redéfinir les « flux ». Aux dires des cadres enquêtés, ils ont été un réel succès en permettant « de rationaliser pas mal de choses ».

Tandis que les ateliers favorisent une remontée d'information, la « cascade managériale » en est la clef de transmission et le pendant, en faisant de chaque échelon hiérarchique un point de passage obligé entre deux niveaux. Autrement dit, chaque agent en position de responsabilité a le devoir d'un côté de faire remonter l'information à son supérieur, et de l'autre de transmettre les directives et de les expliciter à ses collaborateurs. Ce principe de fonctionnement favorise donc l'implication et l'engagement des agents en position d'encadrement, en leur faisant porter une part de responsabilité dans le fonctionnement et dans la performance de leurs équipes.

« Le système de management à la société, se veut fort cascadié de niveaux en niveaux. Et on essaye d'éviter de sauter des niveaux. Chaque niveau hiérarchique est responsable de l'information des gens qui sont directement en dessous. C'est une cascade managériale qu'on souhaite mettre en place pour favoriser l'appropriation à tous les niveaux, montrer que le responsable hiérarchique est bien le référent de l'équipe, que ce n'est pas quelqu'un d'autre qui vient lui souffler la place quand il y a quelque chose à dire. Ça veut dire aussi qu'il est responsable du message qu'il porte. Donc on essaye de faire ça de manière à responsabiliser, à ce que chacun devienne l'ambassadeur du message qu'il porte, en étant le plus possible convaincu. »

Enfin, les « revues » représentent les moments d'échanges et de partage autour desquels sont élaborés les bilans intermédiaires et les feuilles de route, en examinant collectivement différents indicateurs de

¹¹ La réduction du papier est obtenue par la montée en puissance de l'informatique et des outils numériques dans l'organisation.

fonctionnement. Elles constituent à la fois des formes d'évaluation et des temps de consolidation du collectif. Quatre principales revues existent à la SWDE, de la plus opérationnelle à la plus stratégique : les revues de management opérationnel ; les revues de processus ; les revues de programmes ; la revue globale des indicateurs.

Les revues de management opérationnel consistent à examiner tous les mois les indicateurs opérationnels d'activité, afin d'apprécier « *le niveau de performance du personnel* ». Elles sont menées à l'échelle locale entre le directeur régional et ses collaborateurs. Ces revues permettent d'échanger autour de l'activité quotidienne des agents (difficultés de terrain, événements particuliers, sécurité au travail, etc.) et des objectifs. Les revues de processus proposent un examen d'indicateurs plus stratégiques, afin d'apprécier la performance des processus dans leur ensemble. Elles sont réalisées de manière semestrielle ou annuelle et portent sur des indicateurs agrégés relativement standards : la réglementation, les impacts, les audits, etc. Elles sont animées par deux membres du Comité de Direction et convoquent l'ensemble des cadres des processus. Les revues de programme sont consacrées aux projets qui les composent. Il s'agit de dresser des bilans intermédiaires en présence d'au moins un membre du Comité de Direction qui encadre le programme. En termes d'apports, les deux premières revues peuvent entraîner directement des adaptations opérationnelles de l'activité afin de répondre aux difficultés de terrain, ou générer une remontée d'information pour que la question soit discutée et traitée à un échelon hiérarchique plus stratégique. Les revues de programme peuvent quant à elles aboutir à des propositions de décision stratégique à soumettre directement au Comité de Direction. Enfin, la revue globale des indicateurs est réalisée annuellement pour dresser le bilan d'activité et de gestion. Elle sert à synthétiser l'activité, à évaluer la performance globale de l'entreprise et à se comparer dans le cadre des benchmarks.

Ces différents temps et niveaux de suivi évaluatif ont ainsi pour objet d'opérer un contrôle régulier des données et de l'activité. En s'appuyant sur les données quantifiées de terrain, progressivement retraduites en de nouveaux indicateurs statistiques agrégés, l'évaluation porte finalement moins sur l'analyse du réel (hormis dans les revues de management opérationnel) que sur sa traduction statistique. Les chiffres sont ensuite remis en mots et des éléments contextuels et conjoncturels sont de nouveau convoqués pour leur redonner du sens social. C'est d'ailleurs ce dernier qui semble aujourd'hui faire le plus défaut pour nombre d'agents. Bien que le management soit destiné à conserver l'« enrôlement » (Callon 1986) des agents, le pilotage par les chiffres et la succession des réformes au sein de la Société créent malgré tout de l'incompréhension et favorisent l'enfermement des équipes sur leurs propres objectifs et moyens.

« La question des objectifs et des réformes, c'est vraiment des critiques qu'on entend. [...] On est dans une phase un peu critique pour le moment. On a réorganisé en juillet [2022] et on n'a pas été du tout dans le détail des activités, et tout ça est en train seulement de se redéfinir et se redéployer. [...] Aujourd'hui, la logique est assez peu collaborative. Elle est plutôt "j'ai ça comme troupe dans mon processus, voilà ce que je peux faire et le reste, prenez-le chez vous". [...] En parallèle, on travaille aussi à écrire "L'histoire du changement". C'est une espèce de document qui fait 6-7 pages, qui est l'histoire, comment elle commence, pourquoi, parce que. Donc c'est une histoire pour que les gens comprennent mieux qu'un tableau excel, de dire "la cible, pour notre nouveau plan stratégique, c'est 31 %" car ça, c'est moins excitant. »

Cette perte de sens se couple par ailleurs avec une dégradation ressentie des conditions de travail liée à la politique du chiffre de l'entreprise : « *on veut faire plus mais à périmètre constant* », raconte un enquêté qui observe le décalage entre les objectifs, les moyens mis en œuvre et sa réalité de terrain.

Cette question du sens au travail prend d'ailleurs de l'ampleur au sein de la Société, au point qu'elle est aujourd'hui perçue comme une réelle faiblesse pour l'entreprise. C'est pourquoi, un service dédié à l'accompagnement au changement a été créé en 2020, pour faciliter l'appropriation sociale des réformes et leur mise en œuvre, limiter la déstabilisation du (des) collectif(s) et éviter les dysfonctionnements. Ce travail d'intéressement et d'enrôlement (Callon 1986) vise ainsi à créer du sens chez les agents et à les motiver à adhérer à la stratégie d'entreprise, en les équipant aux plans sociocognitif et pratique de nouveaux récits collectifs fondés sur des valeurs et des finalités, et de nouvelles ressources pour agir.

« La méthode qu'on utilise, c'est la méthode ADKAR [...]. C'est le modèle de conduite du changement qui est le plus répandu aujourd'hui. C'est un acronyme de cinq termes en anglais. Le A, c'est awareness : avoir conscience. C'est que les gens doivent avoir conscience de l'intérêt pour la société, pourquoi on fait ce choix d'aller par exemple vers un modèle GePaME. Parce qu'on a un réseau vieillissant, des investissements limités et donc il faut cibler les plus pertinents. Il y a moyen d'expliquer, de raconter un peu une histoire sur le A. Le D, c'est desire : le désire d'y aller. C'est plus individuel : qu'est-ce que moi en tant qu'individu dans la SWDE, j'ai à gagner ou au minimum de me convaincre que c'est quand même une bonne chose ? Ça va me demander un peu plus d'énergie mais c'est vraiment top pour les collègues et la SWDE. Le K, c'est knowledge : la connaissance. Ça, c'est vraiment le volet formation. On a déployé l'outil GePaME, si on n'explique pas aux gens comment s'en servir, ça ne risque pas d'être utilisé. Le A, c'est le ability. C'est la notion entre une formation théorique où on va donner une explication en salle pendant une demie journée et il y aura sans doute pendant les trois mois qui suivent, des actions spécifiques pour s'assurer que les gens prennent bien l'outil en main et qu'il n'y a pas de problème. [...] Et puis le R, c'est reinforcement, s'assurer que tout ce qu'on a déployé est intégré au day to day des nouvelles activités et rentre même dans une logique d'amélioration continue. »

2.5 Conclusion

Depuis la création de la SWDE en 1986, un objectif de performance s'est progressivement imposé à la Société, afin de répondre aux nouveaux impératifs socio-écologiques, économiques et réglementaires qui encadrent la délivrance du service. Pour le satisfaire, les dirigeants de l'entreprise ont entrepris de restructurer l'organisation et le fonctionnement de la Société, afin d'être en mesure de mieux suivre et d'évaluer l'activité, et d'identifier les marges de manœuvre permettant d'améliorer la qualité du service et de réduire ses coûts. Ce faisant, un « esprit gestionnaire » (Ogien 1995) a gagné l'organisation et s'est profondément ancré dans le pilotage de l'activité.

Le « référentiel » (Muller 2005) d'action qui en découle, est résolument tourné vers la maîtrise des coûts et de l'activité. Il se déploie au travers d'une approche systémique et préventive des enjeux, soutenue à la fois par : une spécialisation accrue des tâches et des métiers opérationnels ; une recherche de collaborations inter-équipes ; une analyse systématique des risques ; une logique d'amélioration continue.

Pour mesurer l'atteinte des objectifs et évaluer l'activité, un besoin d'objectiver le réel s'est imposé aux dirigeants de la Société. L'objectivation passe alors par la création de données contextuelles et d'activité, dont le format quantifié favorise la production d'indicateurs statistiques, constituant ainsi un rempart contre toute interprétation subjective des situations. Dès lors, un pilotage par les chiffres s'est généralisé au sein de la SWDE, dans le but également d'automatiser autant que faire se peut les décisions d'action.

Bien que ce repositionnement progressif de la Société présente d'importants résultats en termes de maîtrise et d'amélioration du service, il s'accompagne également d'effets plus délétères, qui se font aujourd'hui ressentir.

La restructuration de la SWDE en 2022 autour d'une logique métier, conduit à rompre avec d'anciennes habitudes de travail, et à redéfinir les « périmètres de juridiction » (Abbott 1988). Outre la question de l'allocation de moyens pour chaque processus, ce travail de redécoupage déstabilise l'identité professionnelle des agents, et génère des craintes liées au sens et aux conditions de travail. L'agir collectif censé émerger de cette restructuration, notamment grâce à l'établissement de nouvelles collaborations inter-équipes, semble encore aujourd'hui balbutiant, en raison de contraintes sociales (manque d'effectif), culturelles (habitudes de travail) et d'activité (flux d'activité imprécis) qui pèsent sur l'organisation.

Le pilotage par les chiffres provoque tout d'abord, chez certains agents, une perte de sens au travail puisqu'ils ne voient pas toujours l'intérêt et l'utilité des données que la hiérarchie leur impose de créer et de faire remonter. La recherche d'amélioration des résultats constitue ensuite un risque psycho-social pour les employés placés en situation « *se challenger en permanence* ». Cela se traduit alors par de l'épuisement professionnel et des désaffections. Enfin, l'ambition d'automatiser le plus possible les décisions en alimentant toujours davantage les outils de données multiples, conduit à éloigner les individus du terrain, et progressivement à les désocialiser, voire à déshumaniser certaines décisions et situations en faisant prévaloir la rationalité du calcul numérique sur les logiques sociales.

Pour accompagner cette démarche d'objectifs et de performance, les dirigeants de la Société ont néanmoins déployé un mode de management dit « de la qualité », afin d'impliquer autant que faire se peut les agents dans une démarche d'innovation, et ainsi les faire adhérer aux transformations organisationnelles, du travail et de l'activité. Ce management atteint cependant ses limites, du fait qu'il intervient en appui de la politique du chiffre, plus qu'il ne protège les agents, leurs conditions de travail et le sens vécu.

Pour finir, cette analyse contrastée du fonctionnement de la SWDE et de son mode de gestion autorise un dernier commentaire sous forme de conjecture et de point de vigilance. La dynamique de « gestionnarisation » (Boussard 2008; Craipeau et Metzger 2011) identifiée dans notre travail présente à terme le risque d'installer un cercle vicieux, dont le scénario extrême serait un phénomène d'auto-entraînement, d'auto-renforcement dans lequel la politique du chiffre et de la performance s'appuierait sur une technologisation et une numérisation accrue de l'activité, désincarnant toujours davantage le travail et le terrain en les transformant par une production démultipliée de données. Celles-ci favoriseraient l'automatisation des décisions, l'identification de nouvelles sources de performance et d'économie, et par conséquent de nouveaux challenges. Les agents – dont les conditions de travail

auraient été dégradées à l'extrême en raison de l'enjeu de réduction des coûts – seraient alors contraints à surinvestir un travail qui aurait perdu d'autant plus de sens qu'il serait déconnecté du réel et qui aurait pour principale fonction d'alimenter en continue la machine numérique et la politique du chiffre.

Archive Tâche 02 :

- Notes méthodologiques
- Grille/ Guide d'entretien

3 Tâche 03 – Implémentation de Casses

3.1 Analyse descriptive des tronçons et des casses

La grande quantité disponible de données permet d'appliquer le modèle LEYP via le logiciel Casses à différents sous-jeux de données ayant un comportement différent vis-à-vis du phénomène de défaillances. Afin de définir ces sous jeux de données, appelés ensuite strate, une analyse descriptive des tronçons et des casses a été réalisée afin d'explorer les liens entre les différentes covariables et le phénomène de casses (Annexe T03_01).

Après cette analyse descriptive, il ressort que la division du jeu de données est optimale avec 15 strates (Figure 5). Il a ainsi été décidé de ne pas stratifier les données d'adduction, et de stratifier par matériau et secteur géographique les données de distributions.

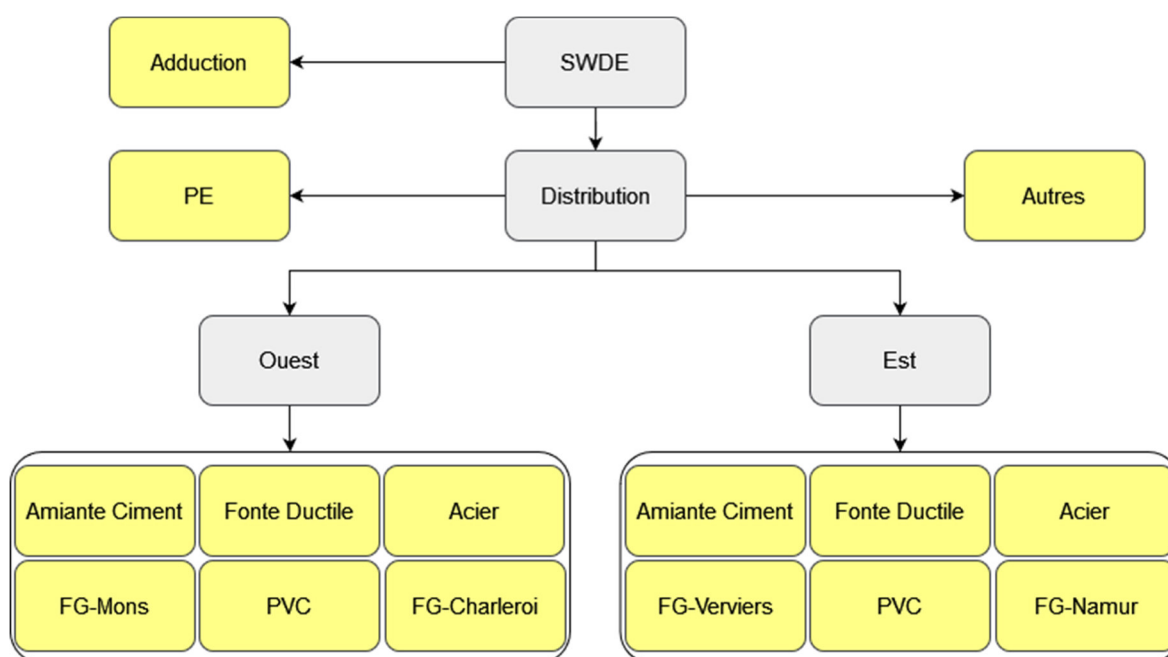


Figure 5. Résumé des 15 strates utilisées par le modèle LEYP dans Casses.FG : Fonte Grise

3.2 Calage

3.2.1 Jeux de données

Selon les strates, les périodes utilisées pour le calage du modèle LEYP varient (Tableau 5). La période standard et majoritaire est un calage sur 2011-2019, mais 4 strates ont un calage effectué sur 2016-2020 et une sur 2011-2018.

Tableau 5. Résumé des caractéristiques des strates utilisées pour les modèles LEYP dans Casses. Les périodes de calage et de validation en gras et italiques sont les périodes qui diffèrent de la période standard 2011-2019 / 2020-2021.

Strate	Période de calage	Période de validation	Nombre de casses observées	Linéaire (km)	Taux de casses observé (u/km/an)
Adductions	2011-2019	2020-2021	1823	3370	0.062
FG-Mons	2011-2019	2020-2021	3640	2715	0.149
FG-Charleroi	2011-2019	2020-2021	5237	2600	0.224
FG-Namur	2011-2019	2020-2021	3249	1816	0.199
FG-Verviers	2011-2019	2020-2021	4052	1622	0.278
Acier-Ouest	2011-2019	2020-2021	1208	614	0.219
Acier-Est	2011-2019	2020-2021	2691	1226	0.244
FD-Ouest	2011-2019	2020-2021	1103	1549	0.093
FD-Est	2016-2020	2021	526	1552	0.074
Amiante Ciment-Ouest	2016-2020	2021	1787	2180	0.164
Amiante Ciment-Est	2016-2020	2021	919	794	0.231
PVC-Ouest	2016-2020	2021	949	3006	0.064
PVC-Est	2011-2019	2020-2021	2519	3141	0.092
PE	2011-2019	2020-2021	1523	726	0.245
Autres	2011-2018	2019-2021	137	84	0.204

3.2.2 Choix des covariables pour chaque strate

A partir des données brutes, 2 covariables sont créées de manière systématique :

- Le logarithme de la longueur : $\ln LNG = \ln(LNG)$
- Le logarithme du nombre de branchement +1 : $\ln BR1 = \ln(NB_{RACC} + 1)$

Sur le jeu de données de calage, nous avons appliqué une méthode de sélection de modèles dite « backward », c'est-à-dire que nous sommes partis du modèle le plus complet avec le plus de covariables et nous avons laissé Casses supprimer au fur et à mesure les covariables non significatives (avec une hausse de la limite des itérations à 20 000). Des regroupements de modalités non significatives des variables qualitatives avec la référence ont parfois été nécessaires et différentes selon chaque strate, de même que la discrétisation en classe du diamètre. Enfin, nous avons réalisé une discrétisation en classes des dates de poses puisque la fixation de δ à 1 a été nécessaire à la convergence des modèles du fait de l'influence non monotone du vieillissement sur l'intensité de défaillances.

3.2.3 Résultats des calages

Les covariables ont été regroupées en différentes catégories. La catégorie environnement regroupe le type de voirie, les CLC, les proportions de tronçons sous chaussée/acotement/dans l'argile/dans le limon, avec différents regroupements de modalités selon la strate. La catégorie territoire regroupe les différents découpage géographiques (EF / ZF).

Un résumé des résultats est présenté en Tableau 6, et les résultats exhaustifs sont disponibles dans l'Annexe T03_02. L'effet du logarithme de la longueur et du logarithme du nombre de raccordement +1 est significatif et positif pour toutes les strates (l'intensité de défaillance augmente quand la variable augmente, $\beta_{lnlong} \in [0.3, 0.6]$, $\beta_{lnbr1} \in [0.13, 0.18]$), excepté pour l'adduction où le nombre de raccordement n'est pas pris en compte. Ces valeurs ne sont donc pas présentées dans le Tableau 6.

Il est intéressant de noter certaines différences :

- L'effet du vieillissement (δ) n'est pas monotone sauf sur l'amiante ciment de l'Est.
- L'effet des défaillances passées (α) est plus élevé sur les plastiques.
- L'effet du matériau est significatif dans le modèle d'adduction : l'intensité de défaillances est supérieure dans le PE que dans l'AMCI, elle-même supérieure aux autres (NB : Les stratifications étant faites par matériau pour la distribution, cet effet n'est pas mesurable en distribution).
- L'effet du diamètre est différent selon les matériaux : l'intensité de défaillance augmente quand le diamètre augmente pour l'Amiante-Ciment et le PE et diminue pour les autres.
- Le caractère chimique de l'eau doit être lié à un autre facteur non mesuré. En effet, on remarque par exemple des effets inverses de la corrosion sur l'AMCI entre Est et Ouest.

Tableau 6. Résultats du calage du modèle LEYP sur les 15 strates. α : effet des défaillances passées. δ : effet du vieillissement du tronçon. Dia : diamètre. Caractère de l'eau : incr, eau incrustante, corr, eau corrosive, (+)/(-), effet positif/négatif sur l'intensité de défaillances. Envt : nombre de covariables présentes dans le modèle de la catégorie environnement.

Strate	Nb covariables	α	δ	Dia ($\times 10^{-3}$)	Pression ($\times 10^{-3}$)	Caractère de l'eau	Envt	Territoire
Adductions	7	4,79	3 classes	-1,37			3	
FG-Mons	9	2,51	4 classes	-1,29			5	
FG-Charleroi	10	2,38	6 classes	-1,41	1,93	(-) incr	3	ZF
FG-Namur	9	1,96	5 classes	-1,43			4	ZF
FG-Verviers	10	1,88	5 classes	-1,27		(-) incr	5	
Acier-Ouest	9	2,99	3 classes	-1,60		(-) corr incr	4	
Acier-Est	9	3,61	3 classes	-2,00	1,75	(-) corr.	2	EF
FD-Ouest	9	4,52	2 classes	-1,68		(-) corr incr	4	
FD-Est	8	4,04	3 classes	-4,76	6,51	(-) corr.	2	
AMCI-Ouest	6	4,50	3 classes	1,55	2,88	(+) corr.		
AMCI-Est	6	2,55	1,35	2 classes	3,07	(-) corr incr	1	
PVC-Ouest	6	5,54	4 classes	4 classes	4,81		1	EF
PVC-Est	7	6,21	5 classes	-5,60	2,71	(-) corr.	1	
PE	8	5,69	3 classes	2 classes	3,56	(-) corr.	1	EF
Autres	7	3,12	2 classes	2 classes	5,17		2	

3.3 Validation

3.3.1 Données jusqu'en 2021

La validation a été effectuée tout d'abord que la période restante après le calage jusqu'au 31/12/2021 (Tableau 5). La validation des 15 strates a montré des résultats satisfaisants, avec des aires sous la courbe comprises entre 0.63 et 0.81 et des pourcentages de casses évitées en changeant 5% du linéaire entre 13 et 42% (Tableau 7, Figure 6 et Figure 7).

Tableau 7. Résumé des résultats de validation sur les données jusqu'en 2021. A_L : aire sous la courbe en longueur. C_{5L} : % des casses évitées en changeant 5 % du linéaire.

Strate	A_L	C_{5L}
Adductions	0,807	32,3
FG-Mons	0,653	15,5
FG-Charleroi	0,653	13,3
FG-Namur	0,673	14,5
FG-Verviers	0,640	14,2
Acier-Ouest	0,730	27,1
Acier-Est	0,705	22,6
FD-Ouest	0,682	23,2
FD-Est	0,678	23,9
Amiante Ciment-Ouest	0,700	25,9
Amiante Ciment-Est	0,695	21,3
PVC-Ouest	0,714	24,5
PVC-Est	0,710	25,2
PE	0,813	42,1
Autres	0,625	14,8

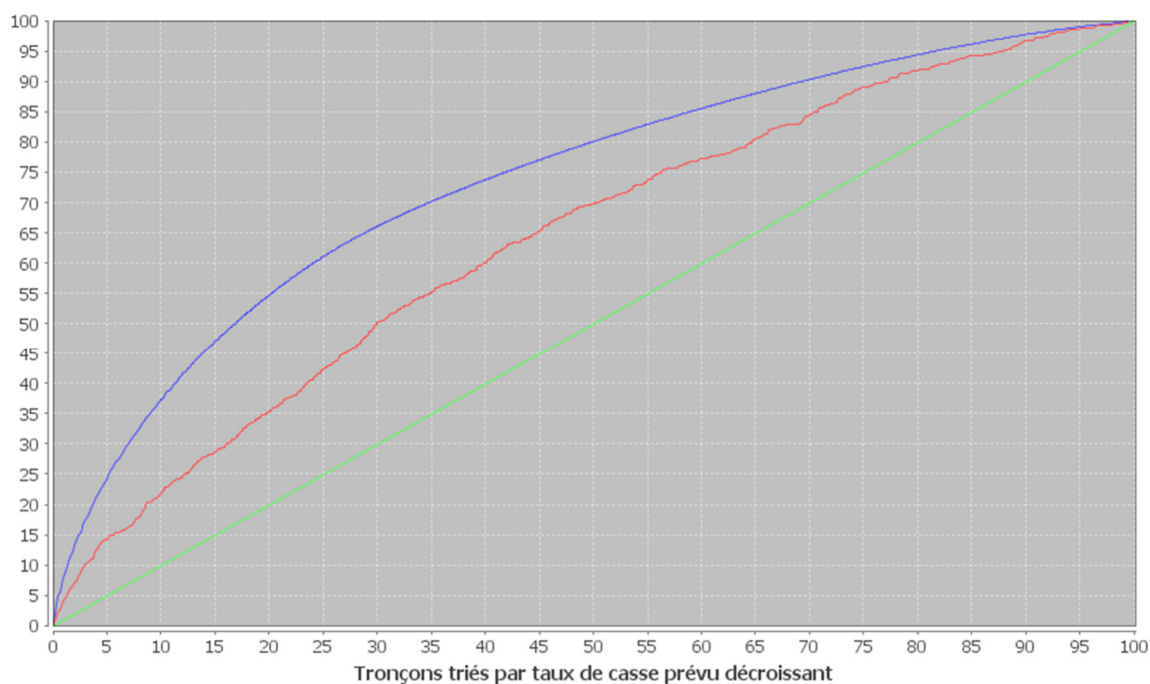


Figure 6. Validation 2020-2021 des casses prédites sur la distribution FG-Verviers. Bleu : casses prédites évitées, rouge : casses observées réellement évitées, vert : renouvellements aléatoires. Aire sous la courbe = 0.640.

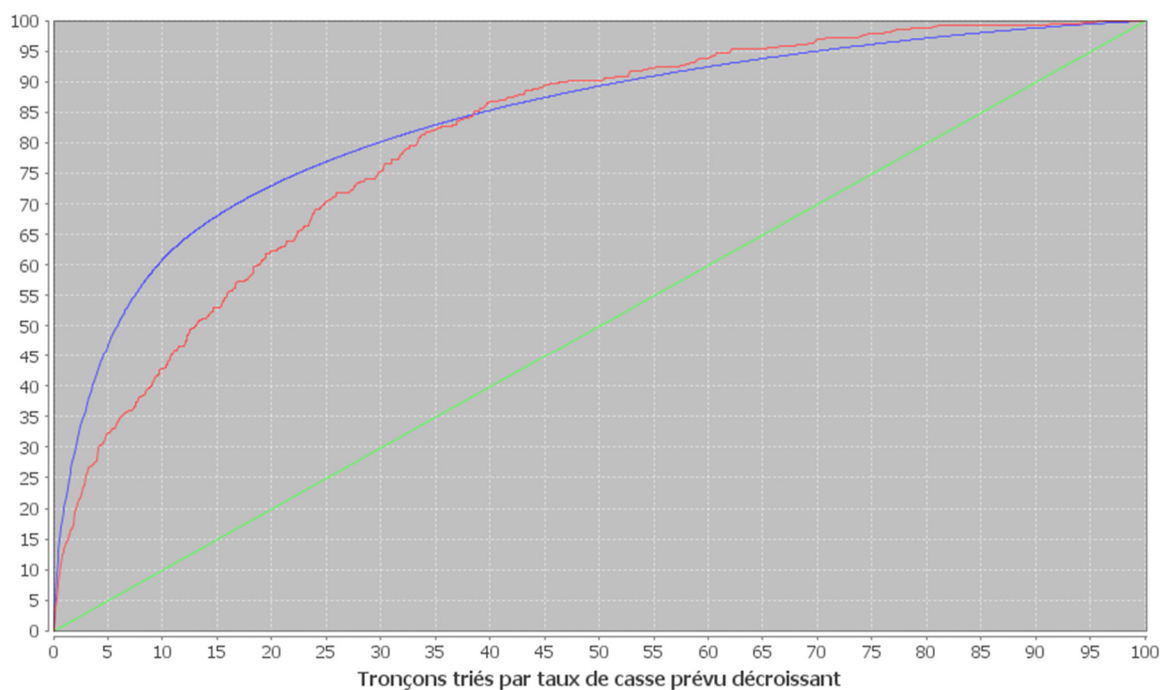


Figure 7. Validation 2020-2021 des casses prédites sur l'adduction. Bleu : casses prédites évitées, rouge : casses observées réellement évitées, vert : renouvellements aléatoires. Aire sous la courbe = 0.807.

3.3.2 Données de 2022

Le traitement des données de défaillances de 2022 (section 1.2) offre la possibilité de comparer les casses prédites en 2022 avec les casses observées et ainsi de faire une validation supplémentaire. Cette dernière montre que les casses sont prédites de manière très satisfaisante sur la distribution (aire sous la courbe = 0.735, Figure 8) et sur l'adduction (aire sous la courbe = 0.822, Figure 9).

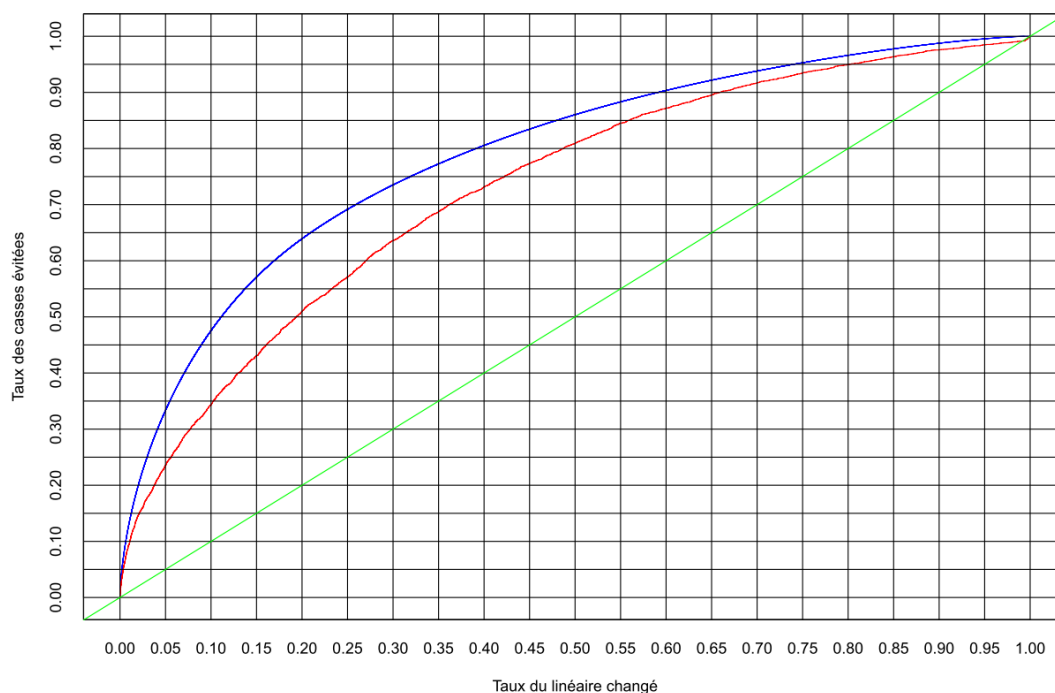


Figure 8. Validation 2022 des casses prédites sur la distribution. Bleu : casses prédites évitées, rouge : casses observées réellement évitées, vert : renouvellements aléatoires. Aire sous la courbe = 0.735.

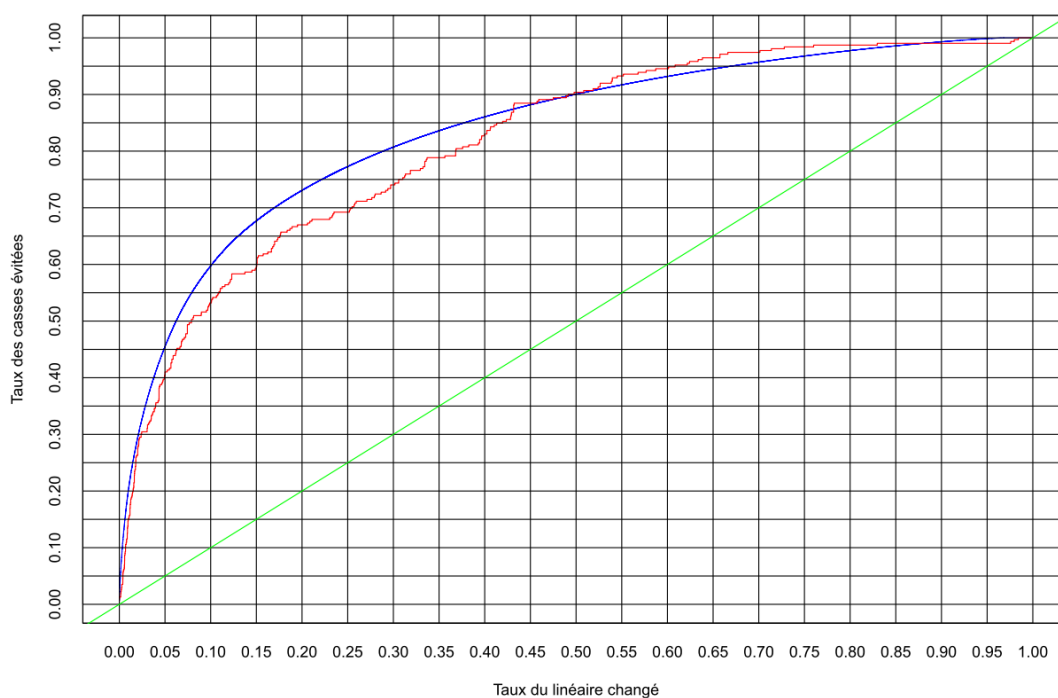


Figure 9. Validation 2022 des casses prédites sur l'adduction. Bleu : casses prédites évitées, rouge : casses observées réellement évitées, vert : renouvellements aléatoires. Aire sous la courbe = 0.822.

3.4 Prévision

Les prévisions de Casses pour la période 2023-2025 sont faites sur l'ensemble des périodes de calage et de validation pour toutes les strates séparément. Les résultats des prévisions des 15 strates ont été regroupés et propagés aux tronçons écartés des modèles (par paquet d'agrégats) afin de créer les représentations cartographiques associées (Figure 10 à Figure 13). Les seuils retenus pour distinguer les tronçons à risque élevé, modéré et faible sont respectivement 2%, 5% et supérieur à 5% du pourcentage cumulé du linéaire lorsque les tronçons sont classés par taux de casses prévu décroissant. On peut noter l'utilisation de traits pleins pour distinguer les conduites dont l'affectation des défaillances passées est plus sûre (affectation jugée « correcte » à « certaine »), et de pointillés pour les autres (Husson et al., 2022), sauf dans le cas des cartes de l'ensemble du territoire pour des raisons de visibilité.

Une analyse géographique des prévisions a de plus permis de montrer une hétérogénéité des risques sur les territoires de la SWDE (Tableau 8). Une alternative pourrait ainsi être de classer les tronçons par taux de défaillances prévus sur chaque EF séparément plutôt que sur le réseau entier.

Tableau 8. Distribution du pourcentage de tronçons à risque élevé de défaillances (les 2% les plus à risque en linéaire cumulé) selon les Entités Fonctionnelles.

EF	Distribution	Adduction
NAMUR	1.86	4.36
CHARLEROI	2.39	0.85
VERVIERS	2.93	1.53
MONS	0.87	1.35
TOTAL	2.00	2.00

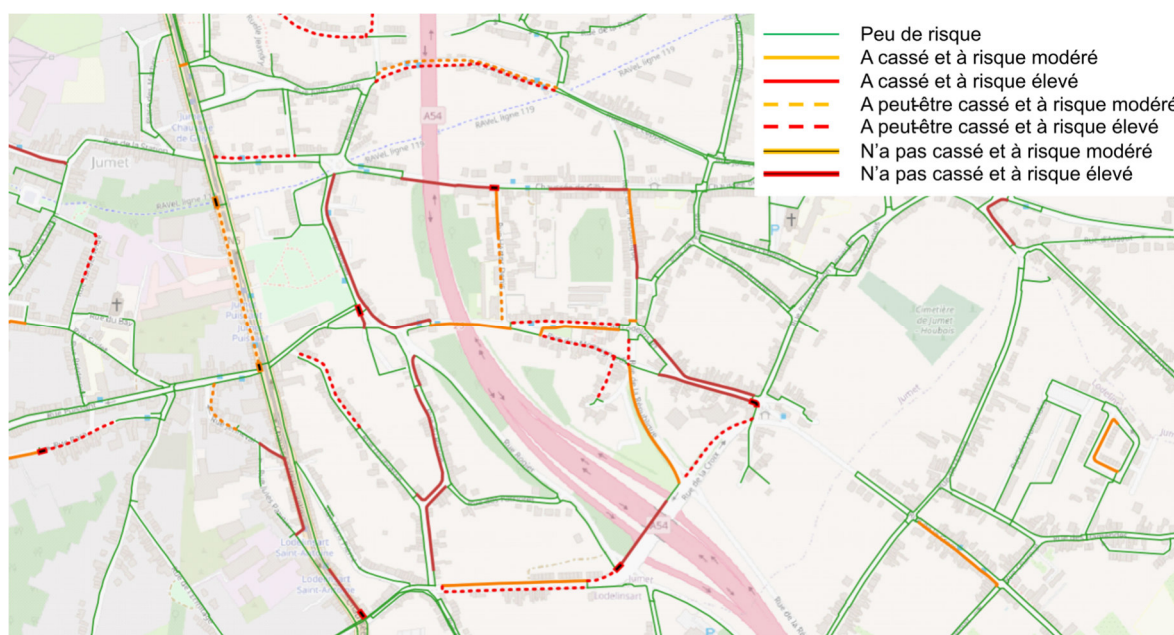


Figure 10. Carte des prévisions de défaillances au 1:5000 de Charleroi-Jumet (distribution).

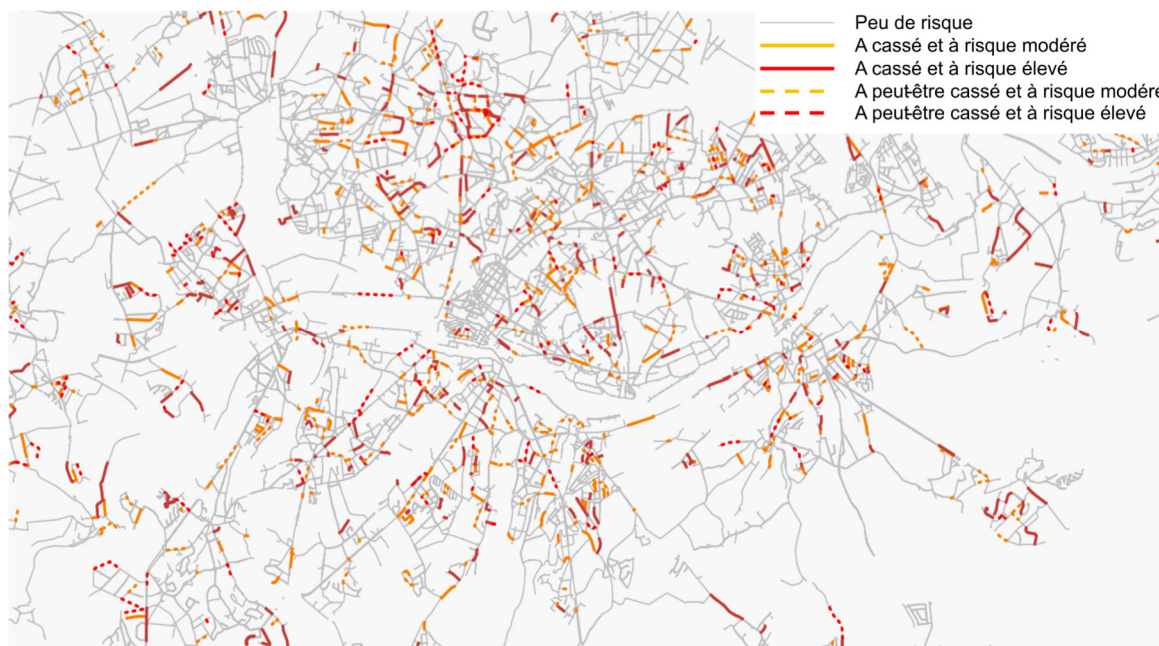


Figure 11. Carte des prévisions de défaillances au 1:50 000 de Charleroi (distribution).

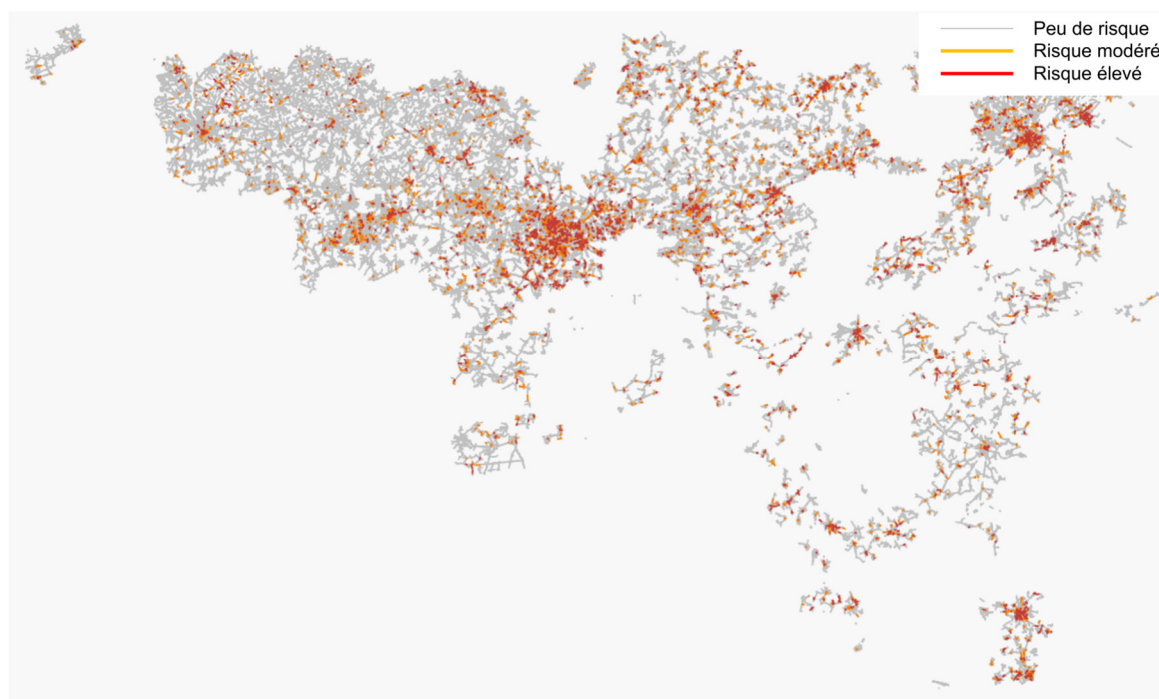


Figure 12. Carte des prévisions de défaillances au 1:700 000 de la SWDE (distribution).

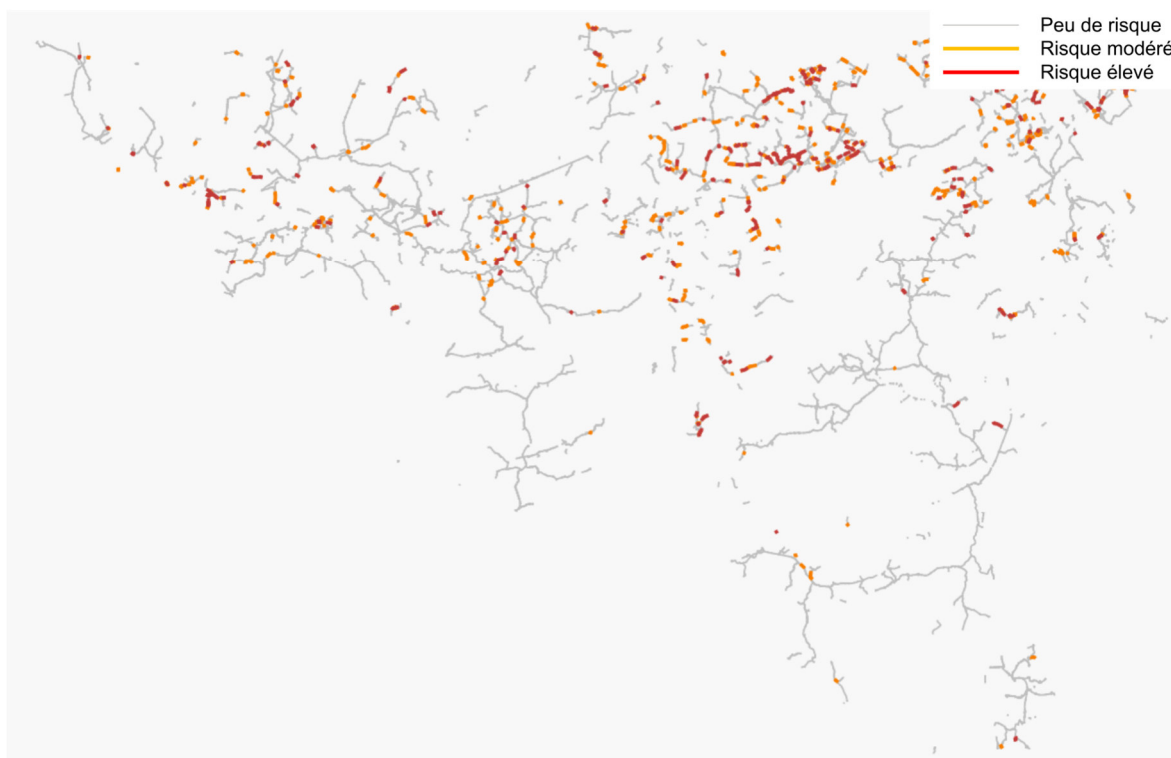


Figure 13. Carte des prévisions de défaillances au 1:700 000 de la SWDE (adduction). On peut noter le risque particulièrement élevé près de Villers-le-Bouillet (Carrière de Moha, Amiante Ciment de > 70 ans).

Archive Tâche 03 :

- Dossier LEYP2s_R :
 - Marie_Collet : Codes et résultats du LEYP2s sous R sur le réseau SWDE fait par Marie Collet
 - Nicolas_Rodriguez : Résultats LEYP/LEYP2s sous R sur le réseau Mons-Ath-Gaurain 2022 par Nicolas Rodriguez
- Dossier Casses : Données, projets Casses et prédictions du logiciel Casses sur le réseau entier SWDE fait par Nicolas Rodriguez

4 Tâche 04 – Intégration des outils dans le système d'information

4.1 Nouveaux éléments de contexte

Dans la continuité des avancées précédentes sur cette tâche (Husson et al., 2022), au cours de la dernière année du projet, les échanges INRAE-SWDE se sont concentrés sur le transfert des méthodes pour produire des fichiers pour Casses. Notons que le transfert des autres outils développés dans GePaME (modèle de probabilité d'état fuyard, outil multicritère court terme, outil de simulation long terme) a été profondément redéfini par trois éléments importants :

- Le changement du SIG de la SWDE au cours des années 2022 et 2023 de 1Spatial Elyx Aqua vers une solution développée par ESRI.
- L'externalisation à partir de 2022 du service informatique SWDE vers la société « Digit'Eaux » qui dispense ses services à l'ensemble des grands acteurs (SPGE, inBW, SWDE, IDEA...) du secteur de l'eau en Wallonie (membres du projet concernés : Frédéric Haan et Gaëtan Jacques).
- Le recrutement confirmé au cours de l'année 2022 de Nicolas Rodriguez pour le 1^{er} septembre 2023 au sein du nouveau « Processus Transformation et Données » de la SWDE à Couillet, qui prend progressivement en charge l'ensemble des données de la SWDE (ainsi leurs méthodes de traitement, d'amélioration, et d'exploitation).

4.2 Positionnement des défaillances dans le SIG et propositions de corrections

Comme décrit dans le rapport précédent (Husson et al., 2022), les défaillances historiques n'étant présentes que dans SAP jusqu'au 31/12/2021 ont été positionnées géographiquement de manière uniforme le long des tronçons après leur avoir été affectées (Figure 14) et transférées sous le format d'un fichier « anomalies SIG » (Annexe T04_01). Les défaillances concernées portent les types d'activités pertinents pour les pertes d'eau, concernant les branchements, conduites, hydrants, et appareils (vannes et autres) qui sont fuyards ou cassants (Husson et al., 2022). Le positionnement des défaillances est fait selon 3 groupes distincts :

- Les réparations de conduites ;
- Les réparations de raccordements ;
- Tout le reste.

Pour chacun de ces groupes, la position des points est choisie de façon à diviser le tronçon en $n+1$ parts de longueurs égales pour n points à positionner. Il peut donc par exemple y avoir une réparation de conduite et une réparation de raccordement affectées à un tronçon, et dans ce cas les deux points seront situés au milieu du tronçon et superposés. Des attributs supplémentaires à ceux des anomalies SIG ont été ajoutés au fichier transmis afin de mieux comprendre : l'origine des données (SAP seul ou SAP et fichiers fuites), la précision de l'affectation au tronçon (rue avec ou sans numéro d'immeuble), la nécessité de l'intervention de la RF oui/non, et l'utilité pour les modèles (Casses ou modèle d'état fuyard) selon l'objet concerné et le type de panne.

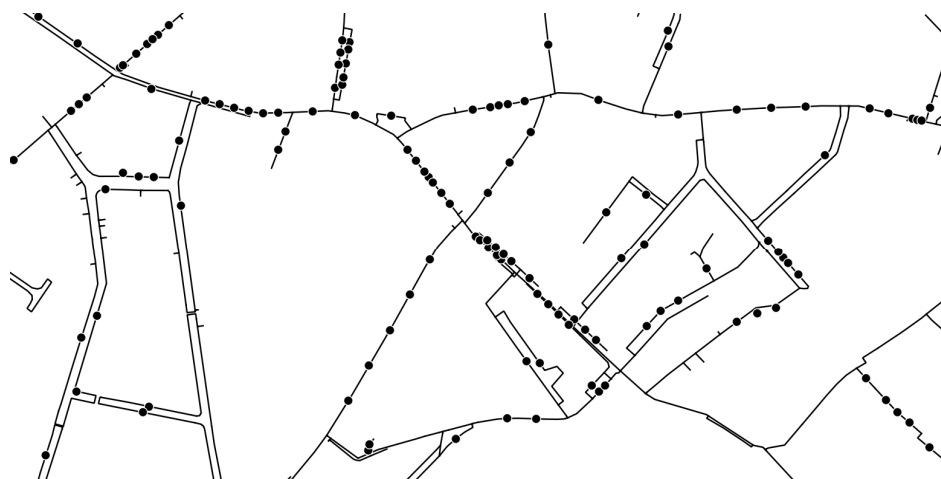


Figure 14. Défaillances SAP (points) positionnées sur les tronçons de manière uniforme.

En complément, partant du principe que les anomalies SIG et opérations SIG seront ajoutées aux défaillances historiques pour créer les fichiers pour Casses, ces défaillances SIG ont été analysées pour détecter les ajustements éventuels à leur apporter pour qu'elles soient toujours valides pour Casses. Cela consiste dans un premier temps à écarter les anomalies ou opérations (Annexe T04_02) qui :

- Sont le doublon de l'une ou l'autre (en conservant en priorité les opérations doublons d'une anomalie) ;
- Possèdent des données obligatoires manquantes et impossibles à déduire (ID de tronçon ou DDC inconnus) ;
- Possèdent des types d'activités non pertinents pour les outils GePaME.

Ce travail a notamment permis de faire des rapprochements entre défaillances SAP (avec OT connu) et points SIG sans OT renseigné (Annexe T04_03).

Dans un second temps, des propositions de modification de certains attributs des anomalies et opérations SIG ont pu être avancées sur la base d'incohérences détectées entre différentes sources de données sur les défaillances et tronçons (SIG, SAP, fichiers fuites). Pour les anomalies, cela concerne les attributs suivants : date de localisation, date de réparation, type d'anomalie, type d'activité, ID de tronçon. Pour les opérations, il s'agit seulement de la date de réparation. Enfin, comme pour les défaillances SAP, des attributs supplémentaires à ont été ajoutés aux anomalies et opérations afin de comprendre si l'intervention de la RF a eu lieu oui/non, et l'utilité de la défaillance pour les modèles les utilisant (Casses ou modèle d'état fuyard) selon l'objet concerné et le type de panne.

Notons que bien qu'il ait été décidé d'arrêter les traitements de données au 31/12/2021 pour GePaME, le même travail que ci-dessus a été effectué une nouvelle fois en 2023 pour les données de l'année 2022 (Annexe T04_04 à T04_06 reprenant les mêmes fichiers que les Annexes T04_01 à T04_03 pour les données de 2022) afin d'aider la SWDE à améliorer son jeu de données.

4.3 Production des fichiers pour Casses

Toutes les méthodes permettant le calcul des covariables pour les tronçons ont été transférées et adaptées au SI de la SWDE (principalement en tant que workbench FME, parfois faisant appel à du code

R). Il est donc normalement possible de produire sur demande, au sein du SI, un fichier des tronçons au bon format et avec les covariables nécessaires pour Casses.

Concernant le fichier des casses, l'analyse des données SIG (anomalies et opérations) de l'année 2022 entière montre que près de 100 % des fuites sont dans les anomalies SIG. En revanche, seulement environ 30 % des crashes sont encodés dans le SIG, bien que la consigne d'encoder tous les crashes comme des anomalies ait été transmise aux agents. La géolocalisation des défaillances de 2022 (voir 1.2) a permis de corriger ce problème. Il est cependant essentiel de pouvoir désormais générer directement le fichier des casses à partir des anomalies SIG, qui sont beaucoup plus fiables que les données SAP à la géolocalisation incertaine (rue et numéro d'immeuble) à très incertaine (rue).

4.4 Importance du lien relationnel entre anomalies et tronçons

L'intégration des fichiers décrits dans les sections précédentes par la SWDE a soulevé la question de l'utilité d'un lien relationnel entre les anomalies et les tronçons (le fait que l'ID de tronçon soit un attribut de l'anomalie), au lieu d'un simple lien géographique (le positionnement XY de l'anomalie proche du tronçon concerné). Il est important de rappeler que pour Casses, un individu (tronçon) doit garder le même nom et cumuler correctement toutes les défaillances le concernant tout au long de sa vie. Ainsi, en raison de l'évolution du réseau (découpage, renommage, renouvellement, modification du tracé des conduites...), un lien relationnel doit être maintenu en plus d'une géolocalisation XY quelle que soit l'origine de l'anomalie (SIG ou SAP), pour éviter les incohérences ou ambiguïtés dans le futur (exemple : où placer une anomalie lors de la mise hors service ?). Par conséquent, des règles (Tableau 9) ont été suggérées pour la gestion des ID de tronçon de façon à avoir un suivi cohérent du nombre de casses des tronçons.

Tableau 9. Règles possibles de gestion des ID de tronçon en cas d'évolution du réseau.

Évolutions possibles d'un tronçon	ID de tronçon que doivent porter les anomalies du tronçon
Modification du tracé sans renommage	ID d'origine
Modification du tracé avec renommage	Nouvel ID
Découpage (avec/sans renommage)	ID du morceau le plus proche
Mise hors service	ID du tronçon HS, idéalement le même que l'original

Notons que les deux dernières lignes du Tableau 9 peuvent être combinées, et que dans ce cas la solution dépend de la localisation de l'anomalie par rapport au découpage et à la partie mise hors service. Il est aussi important d'insister sur le besoin de conserver les ID de tronçons lors de la mise hors service au lieu de les modifier sans garder le lien entre l'ancien et le nouveau ID (en cours de validation pour le nouveau SIG de la SWDE).

4.5 Calcul des débits

Les calculs de débits ont été réalisés à partir des données brutes historiques (2017 – 2021) fournies au pas de temps du ¼ d’heure tels que décrits dans le rapport précédent (Husson et al., 2022). Les résultats sont résumés dans l’Annexe T04_07 qui reprend à la fois les données de contexte, de patrimoine et d’exploitation du réseau agrégés à l’échelle de la cellule.

Pour la tâche 11, les débits sont calculés directement à partir d’extractions des débits journaliers issus de Perf’O.

Archive Tâche 04 :

- Dossier Calcul_debit : Données, codes et résultats du calcul de débit à partir des données de compteurs.
- Dossier Donnees2022 : fichiers Anomalies et défaillances hors SIG ainsi que leur traitement pour l’année 2022
- Envoi_FH_JGB_NR_2023-01-06 : Tous les codes et fichiers ayant été transférés à Frédéric Haan le 06/01/2023

5 Tâche 05 – Exploitation des données de la sectorisation

5.1 Evolution des débits

L'étude de l'historique des débits de la cellule renseigne sur les minimums déjà atteints, les fuites potentielles de la cellule, l'efficacité de la recherche de fuites. Lorsque l'on observe les chroniques de débits d'une cellule, on constate une grande variabilité aussi bien des débits de nuit (Q_{nuit} , point vert sur la Figure 15) que des débits moyen jour (Q_{moy} en noir). Les évolutions journalières sont dues aux écarts de consommation d'un jour à l'autre (avec un éventuel effet weekend et y compris des tirages occasionnels tel que purges ou manœuvre de BI), aux fuites ou crashes et à leurs réparations. Pour s'affranchir des fluctuations dues à la consommation y compris l'effet « weekend ») et se focaliser sur l'impact des fuites, des crashes et de leurs réparations, les débits sont calculés en moyenne glissantes sur 7 jours :

- $Q_{nuit_7j} = \frac{\sum_{j-6}^j Q_{nuit_i}}{7}$ (en rouge sur la Figure 15)

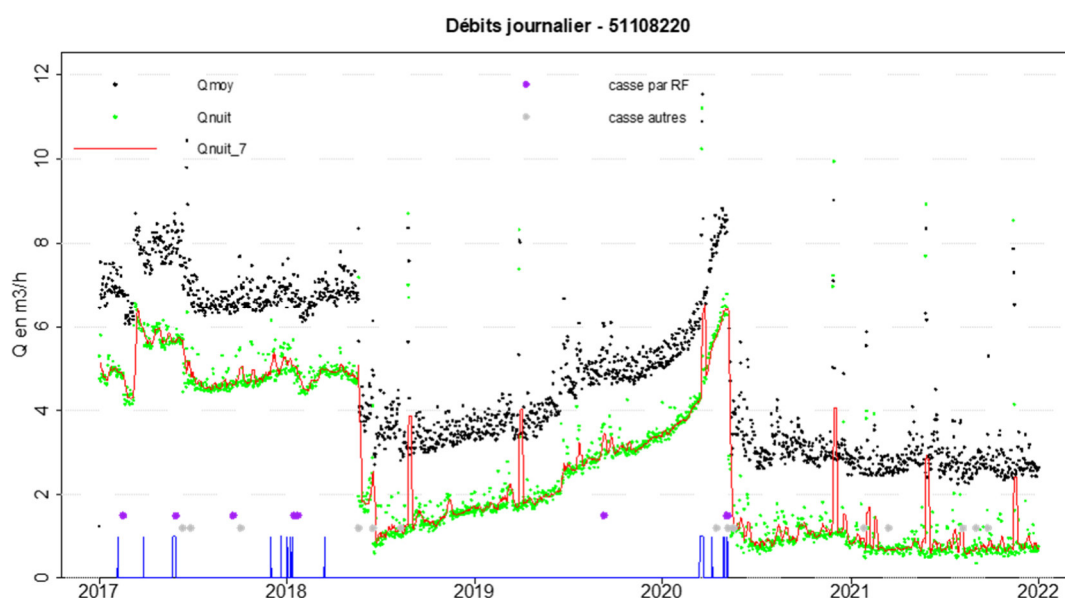


Figure 15. Evolution des débits moyen jour et nuit

Par analogie aux débits de nuit, les ILP sont aussi calculés sur 7 jours glissants pour être comparés à l' ILP de référence retenu par la SWDE :

- $ILP_{ref} = 0.06 \times \frac{\text{nombre de raccordement}}{\text{longueur de la cellule}} + 1.2$

L'écart à l' ILP_{ref} constitue un des éléments déclencheur des actions de recherche de fuite.

L'évolution des ILP de la cellule peut être étudiée en fonction des campagnes de recherche de fuite réalisées (RF en bleu sur la Figure 16), des fuites et crashes trouvés par la RF (point en rouge) ou la 1^{ère} ligne (point en vert). Sur cette cellule 51108220, 2 réparations ont permis le retour à un ILP proche de la référence (en marron), et même de descendre sous cet objectif à partir de mi 2020.

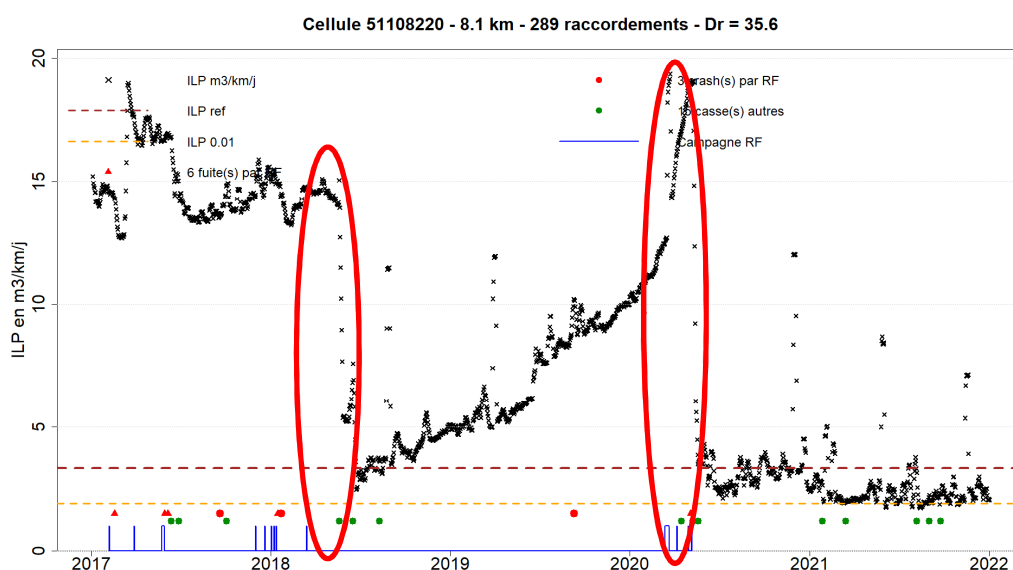


Figure 16. Evolution des ILP en fonction des réparations

Dans l'optique de cibler les recherches de fuites sur les cellules qui en ont besoin, l'*ILPref* même s'il est un peu élevé est une référence adaptée à cette cellule. Par contre, la cellule 51108819 représentée en Figure 17, bien qu'ayant fait l'objet de nombreuses campagnes de recherche de fuite et de réparation, ne voit pas ses *ILP* s'approcher de la référence.

Pour caractériser la valeur d'*ILP* minimale atteinte, l'*ILP 0.01* a été défini. Il représente le quantile 1% de l'*ILP*, soit la valeur en dessous de laquelle les débits sont descendus au moins 1% du temps sur la période étudiée, 5 ans pour cette cellule. Il convient alors de s'interroger sur l'écart entre *ILPref* et *ILP 0.01*, qui peut être lié aux caractéristiques de la cellule (niveau ou variations de la pression, type d'urbanisme, etc.) ou à celles de ses consommateurs (consommations nocturnes atypiques), et d'envisager si nécessaire un ajustement de la référence. Pour certaines cellules, à l'inverse, *ILP 0.01* est significativement inférieur à *ILPref*.

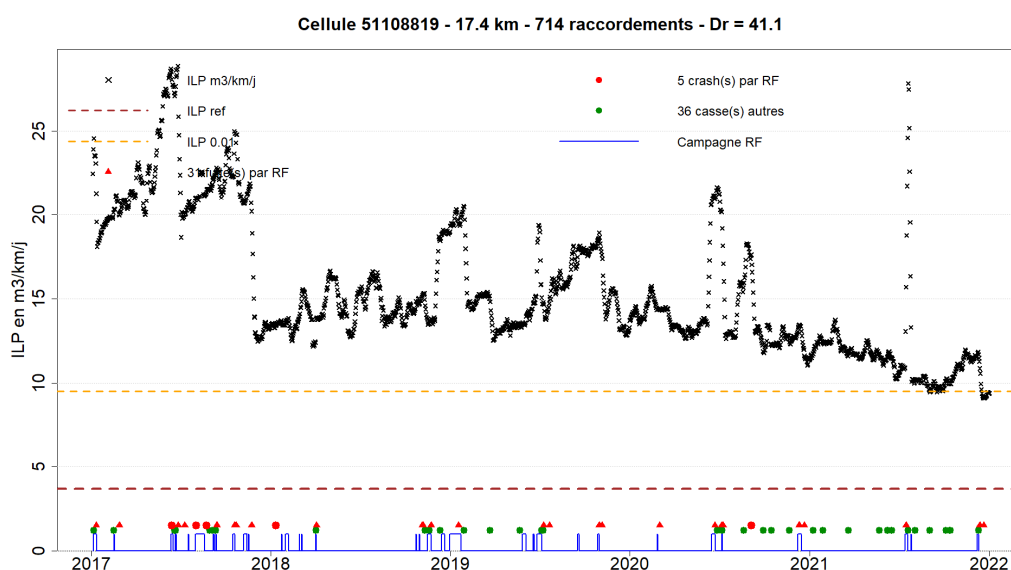


Figure 17. Ecart ILP0.01 - ILPref

5.2 Étude des débits consommés par les abonnés des cellules

Le calcul de l'ILP repose sur l'hypothèse que le débit de la cellule entre 1h et 5h du matin est uniquement constitué de pertes. Des consommations peuvent tout de même avoir lieu pendant ces heures. Cette section décrit les efforts déployés pour estimer les débits consommés par les abonnés des cellules dans l'optique de comparer ces derniers aux débits nocturnes issus de la sectorisation.

5.2.1 Structure des données

Les données brutes (12 millions de lignes) proviennent de SAP et concernent les relevés de compteurs d'abonnés SWDE entre 2016 et 2022 inclus (voir structure détaillée en Annexe T05_01). Il y a environ 1 million de branchements et un 1.1 million de points de consommation (PDC) qui sont relevés en général 7 fois (une fois par an). On s'intéresse uniquement aux installations standard (hors comptages internes SWDE), ce qui représente 99 % des compteurs présents dans les données (ceux qui ont ANLART = 900 ou 901).

Chacun des 1.5 millions de compteurs d'abonnés est identifié de façon unique par la combinaison des identifiants {cellule, branchement, PDC, installation, numéro d'équipement SAP}. Ne pas utiliser l'ensemble de ces informations peut mener à des incohérences et doublons dans l'identification des compteurs.

Il est important de noter que pour les données utilisées dans cette étude, 12 % des branchements n'ont pas de cellule connue du fait du rattachement non exhaustif des branchements entre SAP et le SIG. Leur distribution n'est pas uniforme spatialement et il ne s'agit pas uniquement de gros consommateurs. Le volume consommé par conséquent non attribué aux cellules sur 2016-2022 est 118 871 752 m³, comparable au volume mis en distribution en une année.

5.2.2 Calcul des débits de compteurs et de cellules

Après nettoyage des données, les débits des compteurs entre chacun de leurs relevés sont déduits, ré-échantillonnés à l'échelle journalière, et sommés par cellule, en tenant compte des changements de compteurs sur un même point de consommation (détails en Annexe T05_01).

On s'intéresse seulement aux moyennes des débits consommés par les abonnés des cellules sur les années 2018, 2019, et 2020, où le nombre de compteurs suivis est stable et où tous les compteurs ont des relevés. Ces valeurs sont les seules étant fiables.

5.2.3 Vérification des calculs

La cohérence des débits consommés à l'échelle des cellules Q_{conso} est testée par l'inégalité :

$$Q_{moy} - Q_{nuit} \leq Q_{conso} \leq Q_{moy}$$

où Q_{moy} et Q_{nuit} sont les débits moyens (sur 24 h) et de nuit (entre 1 h et 5 h du matin) des cellules calculés à partir des compteurs de sectorisation. $Q_{moy} - Q_{nuit}$ est une estimation pessimiste (minimale) du débit consommé sous l'hypothèse que le débit de nuit est entièrement constitué de pertes.

Cette comparaison est faite seulement pour l'année 2020, qui a les valeurs les plus fiables. Seules 949 cellules sont sélectionnées pour la comparaison, sur les critères suivants :

- Débits de sectorisation calculables au moins pour les années 2020 et 2021 ;
- Pas de fusion/division au cours de l'existence de la cellule ;
- Au moins 200 jours de débit de sectorisation valides en 2020;
- Débit de sectorisation calculé en accord 7 jours sur 7 avec un extrait hebdomadaire de Perf'O de 2021.

Seulement 481 cellules respectent l'inégalité proposée sur les 949, principalement car Q_{conso} est inférieur à $Q_{moy} - Q_{nuit}$. C'est vraisemblablement car des consommations n'ont pas pu être affectées à ces cellules suite aux 12 % de branchements non rattachés aux cellules.

5.2.4 Méthode de correction des débits de consommation

Les consommations des branchements sans cellule sont affectées à une ou plusieurs cellules en utilisant l'information de rue (ICAR) attachée au branchement dans SAP. Par exemple, pour un branchement sans cellule, situé dans une rue ayant 25 % de son linéaire de conduites dans la cellule C1 et 75 % dans la cellule C2, 25 % du débit du branchement est affecté à C1, et 75 % à C2.

Sur 14 % de compteurs sans cellule à l'origine, il n'en reste plus que 1 % après cette affectation (13 488 compteurs restants, ayant consommé environ 541 m³ en moyenne sur 2016-2022). 112 millions de m³ sur les 119 millions manquants à l'origine sont alors rattachés aux cellules. Il reste donc 7 millions de m³ toujours non affectés aux cellules.

721 cellules sur 949 respectent alors l'inégalité suite à cette correction.

Distribuer aléatoirement les 7 millions de m³ restants sur les cellules où Q_{conso} est toujours trop faible (187 cas) permet d'avoir au mieux 839 cellules respectant alors l'inégalité.

Au final, malgré toutes les corrections envisagées, 12 % de cellules ont des débits de consommation incohérents par rapport aux débits de sectorisation. De plus, cette estimation ne tient pas compte des cellules où la comparaison n'est pas effectuée.

5.2.5 Bilan sur les débits consommés par les abonnés

Les débits consommés par les abonnés des cellules ne sont pas suffisamment cohérents avec les débits issus de la sectorisation pour être pris en compte. L'estimation actuelle de l'*ILP* basée sur le débit total de nuit de la cellule (pertes + consommations éventuelles) ne peut a priori pas être améliorée par la connaissance des débits consommés par les abonnés à une échelle généralement annuelle.

5.3 Lien entre débits et caractéristiques de la cellule

5.3.1 Evolution des *ILP*

Compte tenu du nombre important de données et de leur foisonnement (Figure 18), l'analyse est effectuée par classe. Pour ce faire, les cellules sont triées par densité de raccordement ou de logement croissante, puis l'échantillon est divisé en 20 classes identiques en nombre de cellules (aux arrondis près). Pour une même classe de n individus, les indicateurs sont recalculés pour constituer une valeur par classe :

- $Dr = \frac{\sum_1^n Nb\ racc_i}{\sum_1^n L_i}$, densité de raccordement par classe
- $Dl = \frac{\sum_1^n Nb\ lgt_i}{\sum_1^n L_i}$, densité de logement par classe
- $ILP = \frac{\sum_1^n (ILP_i \times L_i)}{\sum_1^n L_i}$, *ILP* moyen par classe

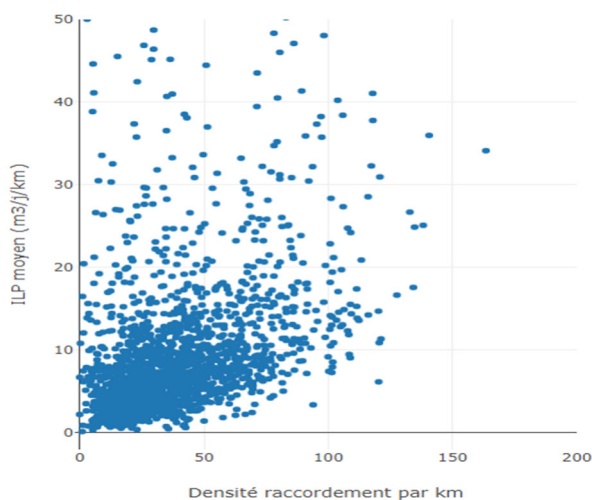
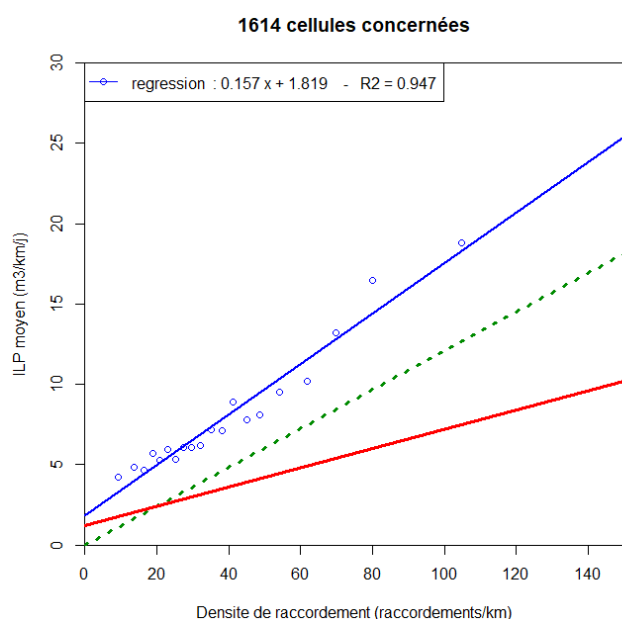


Figure 18. *ILP* moyen de 2000 cellules en fonction de la densité de raccordement

Les valeurs extrêmes qui correspondent à des cellules atypiques sont écartées parce qu'elles présentent des particularités du fait leurs caractéristiques (densité de raccordements Dr , de logements Dl très faible), leur influence sur les résultats est disproportionnée.

La Figure 19 présente la relation entre l'*ILP* et la densité de raccordement (pour 20 classes de densité de raccordement). On note que la pente de la courbe de l'*ILP* de référence en rouge est bien en dessous et

représente l'objectif à atteindre. La droite de régression ne passe pas par l'origine, l'*ILP* est ici calculé à partir des débits de nuit et comporte donc une part de consommation qui s'ajoute aux pertes.



Conditions imposées :

- 20 classes de densité de raccordement
- Densité de raccordement ≥ 5 racc/km
- $ILP_{moy} \leq 150 \text{ m}^3/\text{jour}/\text{km}$
- Nombre d'*ILP* exploitables entre 2017 et 2021 ≥ 365

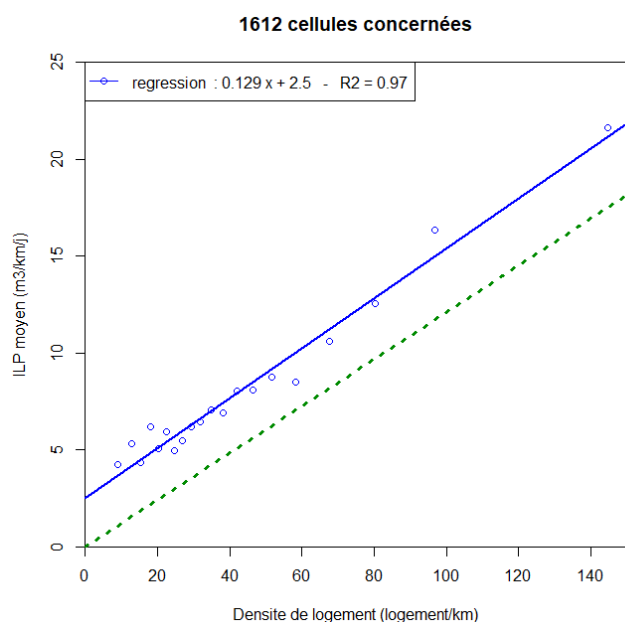
En pointillé vert (référentiel français), pente 0.121

En trait plein rouge $ILPref (0.06 * Dr + 1,2)$

Figure 19. ILP moyen en fonction de la densité de raccordement

Pour comparaison, sur la Figure 19 est tracé en pointillé vert la valeur moyenne de l'indicateur issu du référentiel français d'appréciation du niveau des pertes d'un service d'eau potable (Renaud, 2022). Il est calculé à partir du lien entre l'Indice Linéaire des Volumes Non Comptés (*ILVNC*) et la densité d'abonnés sur la base des données de l'observatoire national des services publics d'eau et d'assainissement français.

La pente est sensiblement différente (0.157 pour 0.121) car on se base sur le nombre de raccordements qui dans le cas général est plus faible que le nombre d'abonnés. La même construction en Figure 20 avec le nombre de logements (plus proche du nombre d'abonnés) donne des pentes beaucoup plus proches des valeurs moyennes françaises (0.129 pour 0.121).



Conditions imposées :

- 20 classes de densité de logement
- Densité de logement ≥ 5 lgt/km
- $ILP_{moyen} \leq 150$
- Nombre d'ILP exploitable ≥ 365

En pointillé vert (référentiel français),
 pente 0.121

Figure 20. ILP moyen en fonction de la densité de logement

Si l'on s'intéresse à l'évolution annuelle des ILP entre les années 2017 à 2021 en ne retenant que les cellules pour lesquelles on dispose d'au moins 180 valeurs d'ILP par an, la Figure 21 ne montre pas d'évolution notable de l'ILP moyen annuel pendant ces 5 années.

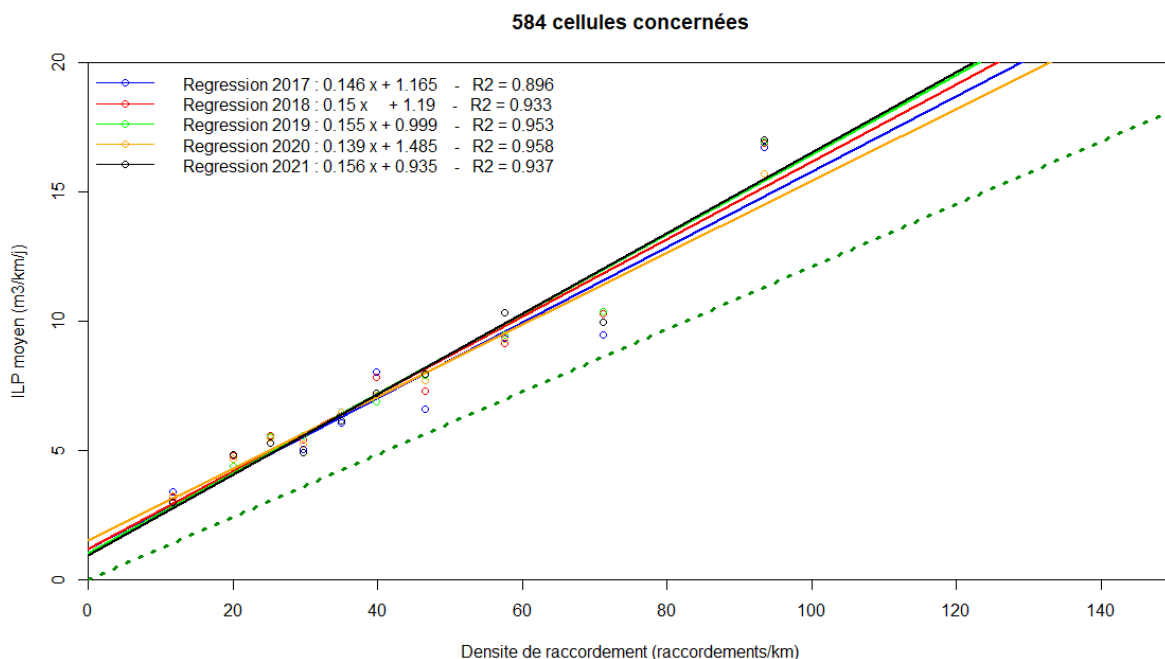
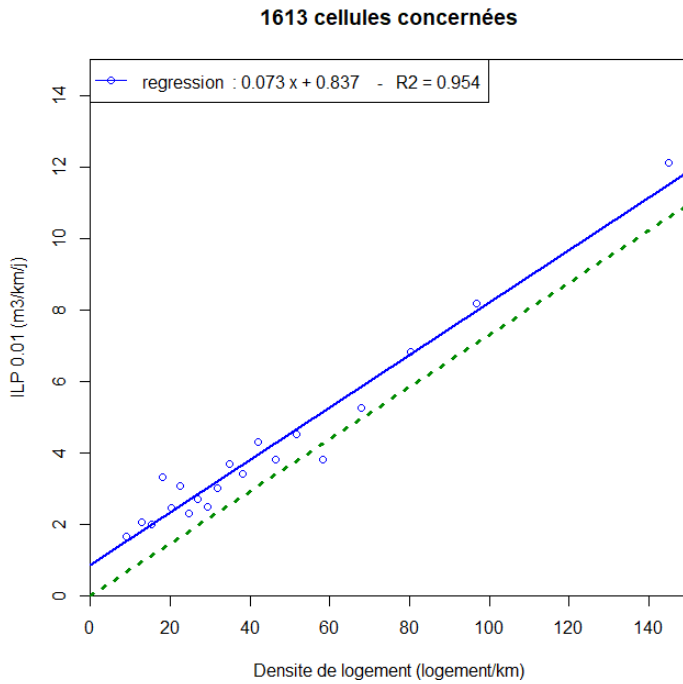


Figure 21. Evolution des ILP moyen annuel de 2017 à 2021.

Les ILP minimums atteints pendant la période d'observation des cellules que l'on peut caractériser avec les ILP 0.01 sont classés et représentés en fonction de la densité de logements en Figure 22. La pente de la droite de régression est proche de celle du seuil bas de l'indicateur issu du référentiel français d'appréciation du niveau des pertes d'un service d'eau potable (Renaud, 2022).



Conditions imposées :

- 20 classes de densité de logement
- Densité de logement ≥ 5 lgt/km
- $ILP\ 0.01 \leq 150$
- Nombre d'ILP exploitable ≥ 365 entre 2017 et 2021

En pointillé vert (référentiel français),
 pente perte seuil bas : 0.073

Figure 22. ILP 0.01 en fonction de la densité de logement.

5.3.2 Indicateurs de pertes

Les Q_{nuit} ou le Q_{ref} tiré de l' ILP_{ref} sont des indicateurs des pertes de la cellule, mais ils sont très dépendants de sa taille (Figure 23). Pour étudier les relations avec les caractéristiques de la cellule, il faut s'affranchir de cette dépendance en rapportant ces grandeurs à L (la longueur de la cellule), N (le nombre de raccordements), le Q_{jour} ou le Q_{ref} (Figure 23).

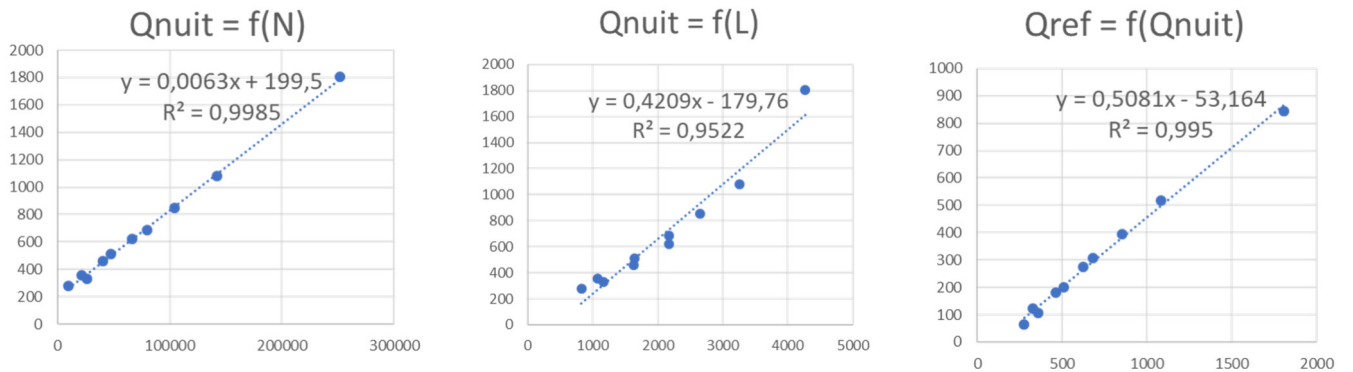
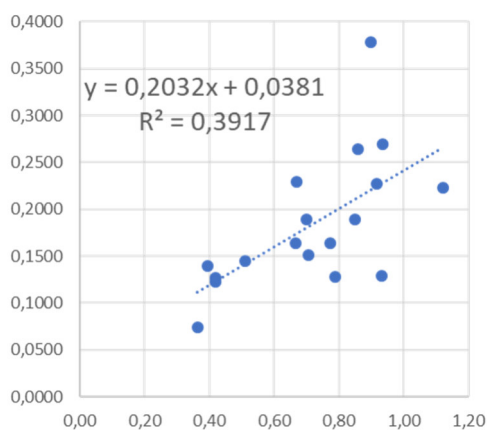
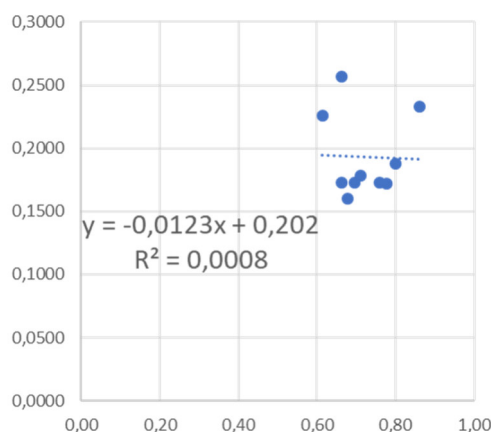


Figure 23. Illustration de la dépendance de Q_{nuit} , Q_{ref} à la taille de la cellule (N et L).

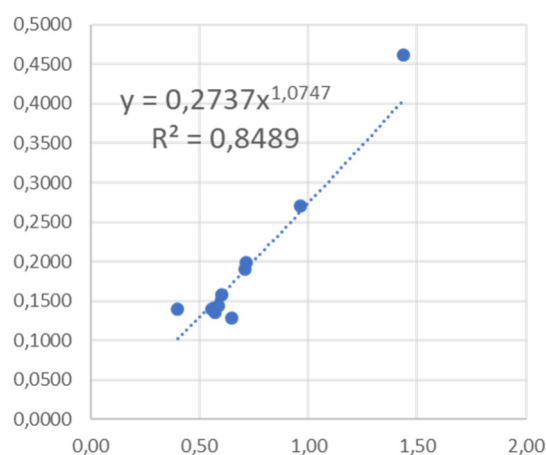
Le foisonnement des données impose également comme précédemment une étude par classe. Le choix de la variable de construction des classes a une importance, on note dans la Figure 24 par exemple que des classes construites sur les secteurs SWDE, ou N masque une éventuelle relation entre $(Q_{nuit}-Q_{ref})/L$ et cRF/L (nombre de campagne de RF sur la longueur des canalisations de la cellule). On privilégiera donc des classes de densité de raccordement (Dr) ou de longueur (L).



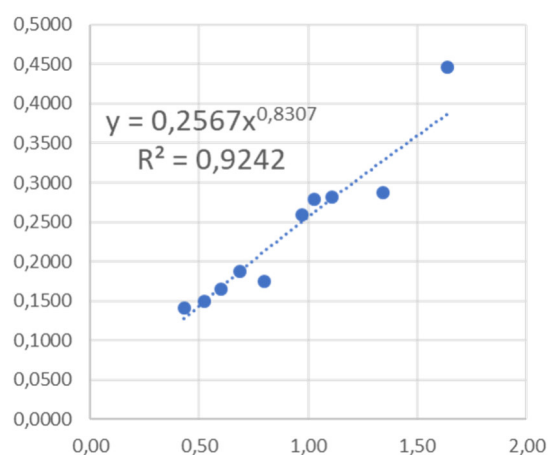
17 classes de secteur



10 classes de N



10 classes de Dr



10 classes de L

Figure 24. $(Q_{\text{ nuit}} - Q_{\text{ ref}})/L = f(cRF/L)$, influence du type de classe (secteur, N, Dr, L).

5.3.3 Relation entre indicateurs de pertes et caractéristiques des cellules (extrait de l'Annexe T05_02)

Si le débit de nuit est un bon indicateur des pertes, il est très dépendant de la taille de la cellule. Pour comparer les cellules entre elle, le ratio $Q_{\text{ nuit}}/L$ est plus approprié. On retient pour la suite :

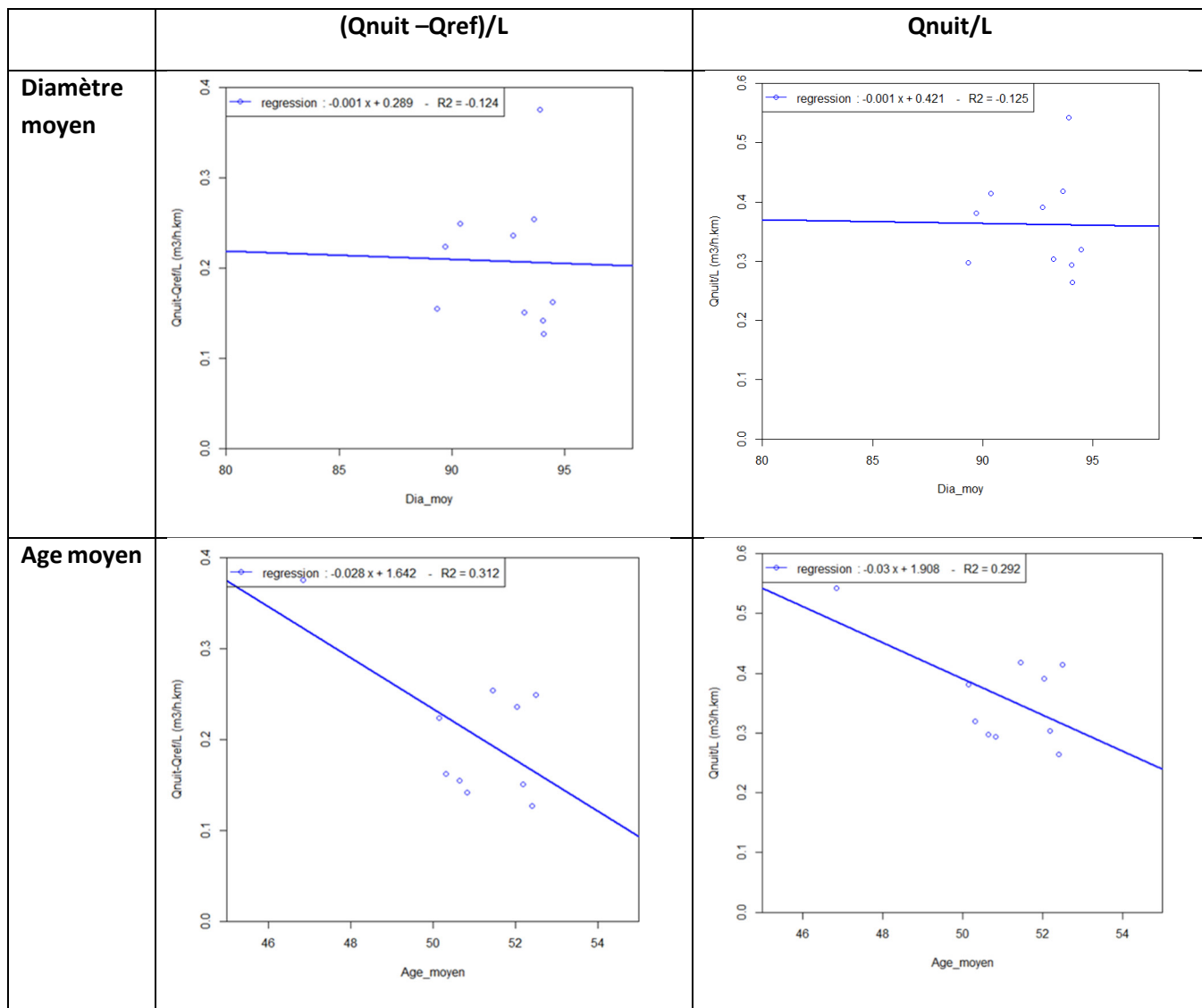
- $Q_{\text{ nuit}}/L$ qui caractérise les pertes linéaires
- $Q_{\text{ nuit}} - Q_{\text{ ref}}/L$ qui reflète l'écart par rapport à la référence SWDE

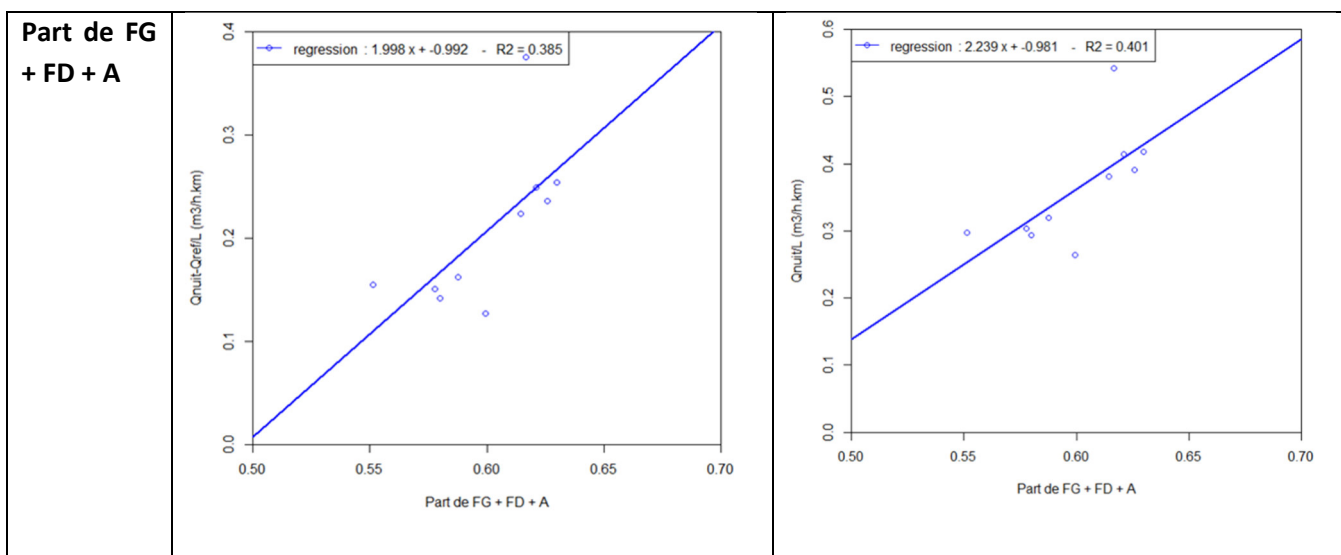
Les indicateurs de pertes $Q_{\text{ nuit}} - Q_{\text{ ref}}/L$ et $Q_{\text{ nuit}}/L$ sont exprimés en $\text{m}^3/\text{h}/\text{km}$, l'exploration des relations entre ces indicateurs et les caractéristiques des cellules s'effectue en utilisant des classes de longueur de canalisations.

Les cellules retenues doivent avoir des débits valables sur plus de 3 ans ($\geq 1\ 000\ Q_{\text{ nuit}}$ entre 2017 et 2021), une longueur de canalisation $\leq 50\ \text{km}$ et avoir des débits cohérents avec les volumes extraits de Perf'O.

Les 1 001 cellules sélectionnées sont confrontées aux variables caractéristiques des cellules (réseau, contexte, fonctionnement et défaillances, RF, débits et consommations) :

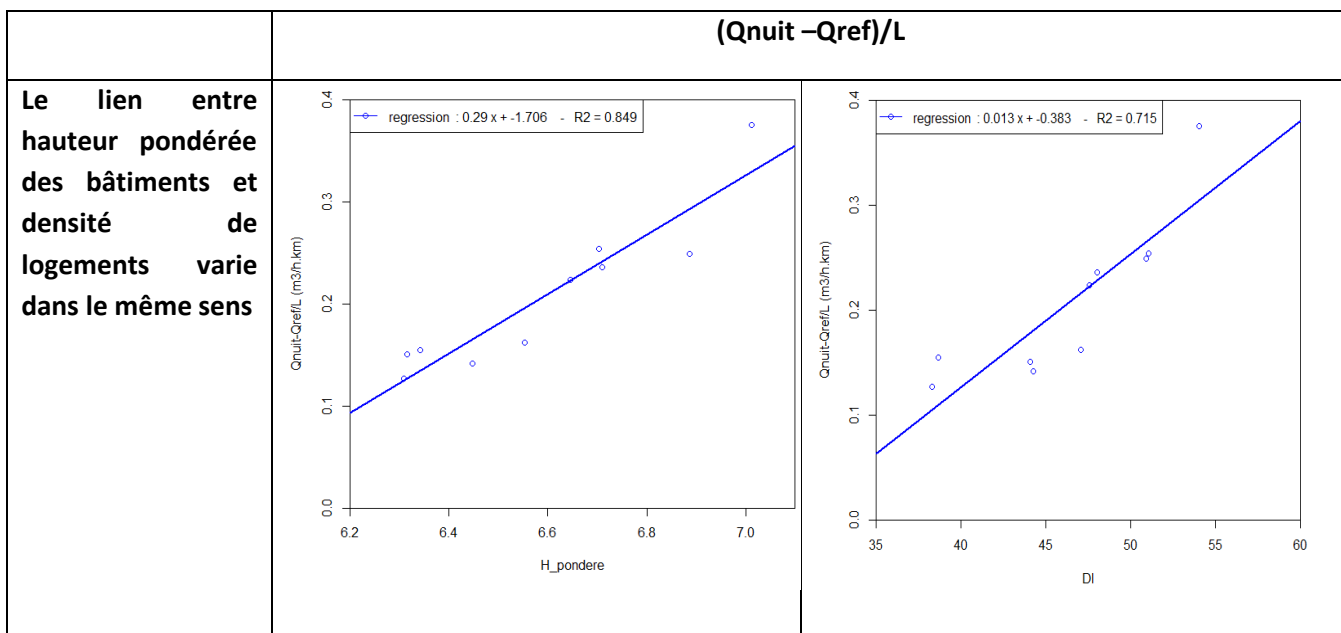
- Variables réseau : pas de lien notable

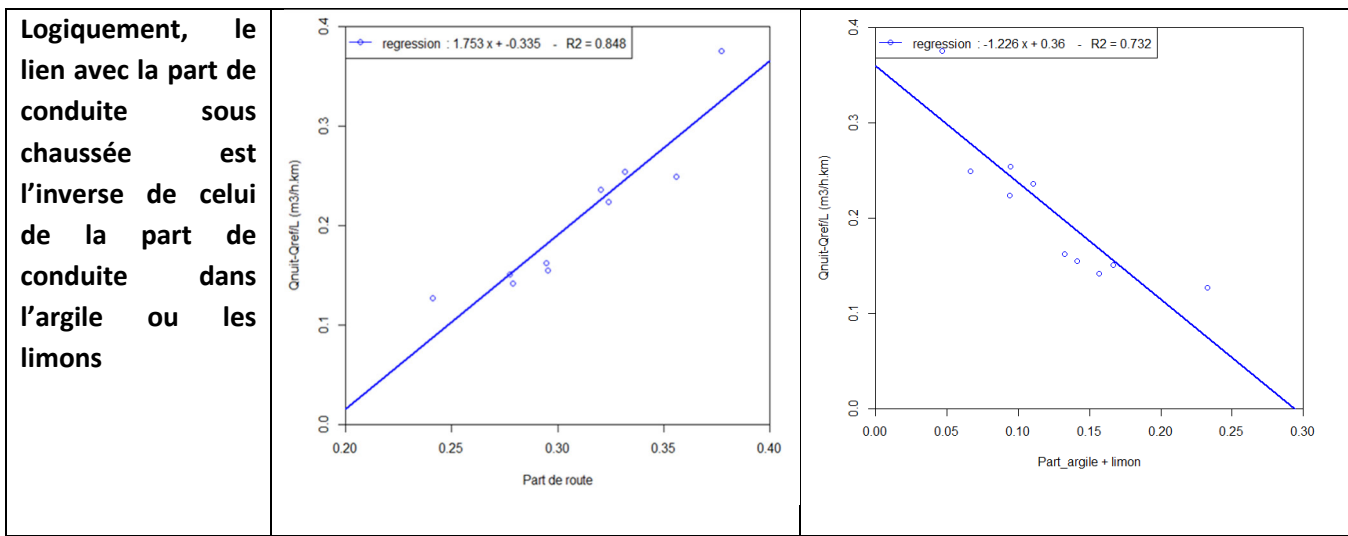




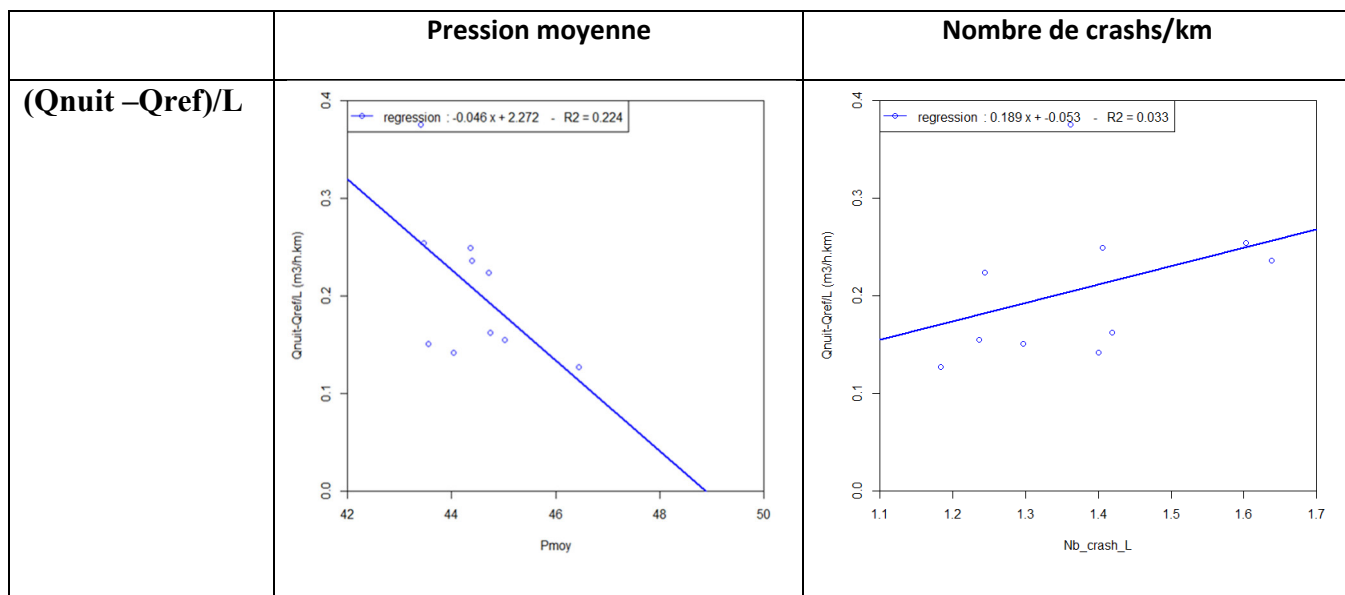
Les liens entre les variables réseau et $Q_{nuit}-Q_{ref}/L$ ou Q_{nuit}/L sont de même type, c'est également le cas pour les variables de contexte, de fonctionnement ou de RF. Dans la suite, seuls les graphiques relatifs à $Q_{nuit}-Q_{ref}/L$ sont présentés (les graphiques représentant Q_{nuit}/L sont disponibles en Annexe T05_02).

- **Variables de contexte** : Les variables de contexte sont parfois dépendantes, par exemple pour l'aspect urbain rural : hauteur des bâtiments, densité de logement, part de CLC1, ...

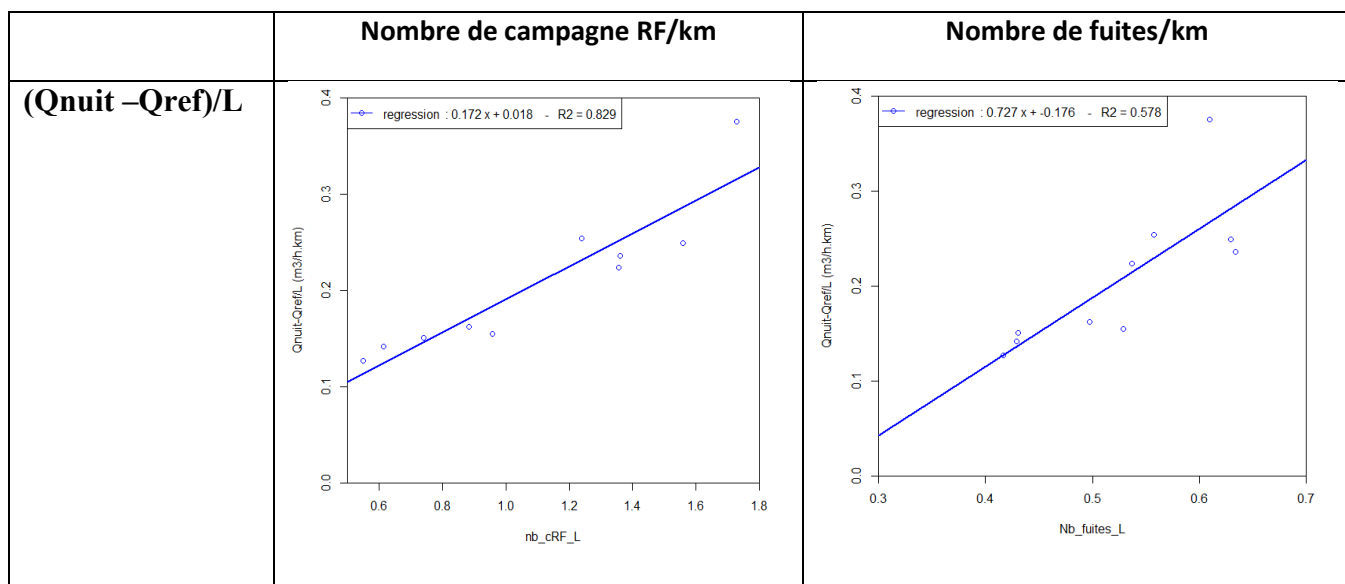




- **Variables de fonctionnement** : Pas de lien notable avec les variables de fonctionnement type pression moyenne ou nombre de crashes



- **Variables de RF**: Plus les débits de nuit par km s'écartent de la référence, plus le nombre de campagnes et le nombre de fuites par kilomètre sont élevés. C'est logique puisque l'augmentation des débits entraîne la mobilisation des équipes de RF. La Figure 25 montre également que plus les campagnes de RF sont nombreuses, plus des fuites sont trouvées.



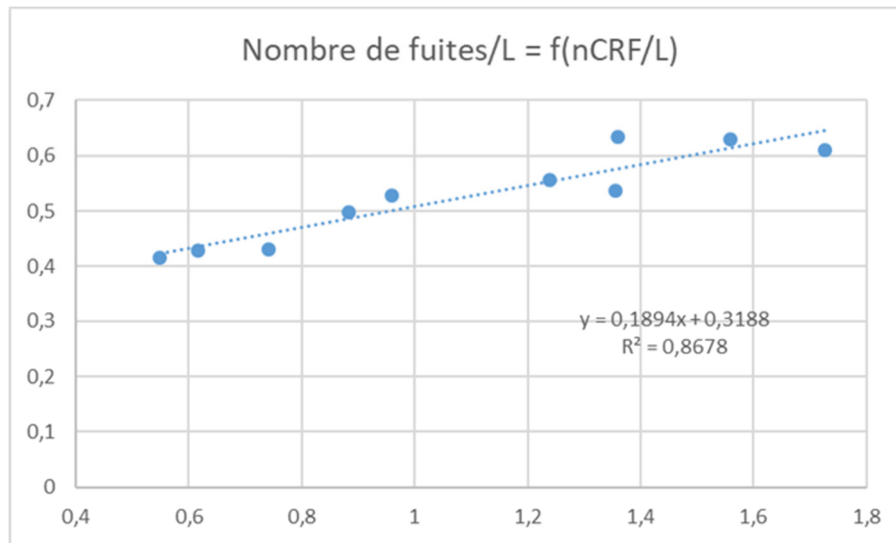


Figure 25. Nombre de fuites trouvées par km en fonction du nombre de campagnes réalisées par km.

5.3.4 Analyse chronologique des (Qnuit-Qref)/L et des fuites, crash et RF

Les cellules retenues ont un minimum de 4 ans de données (nombre de Qnuit valide ≥ 1460), soit un total de 583. Les pertes restent relativement stables et ne suivent pas les variations de l'intensité de la RF ou du nombre de défaillances (fuites et crashes) réparées par km (cf Figure 26 et Annexe T05_02).

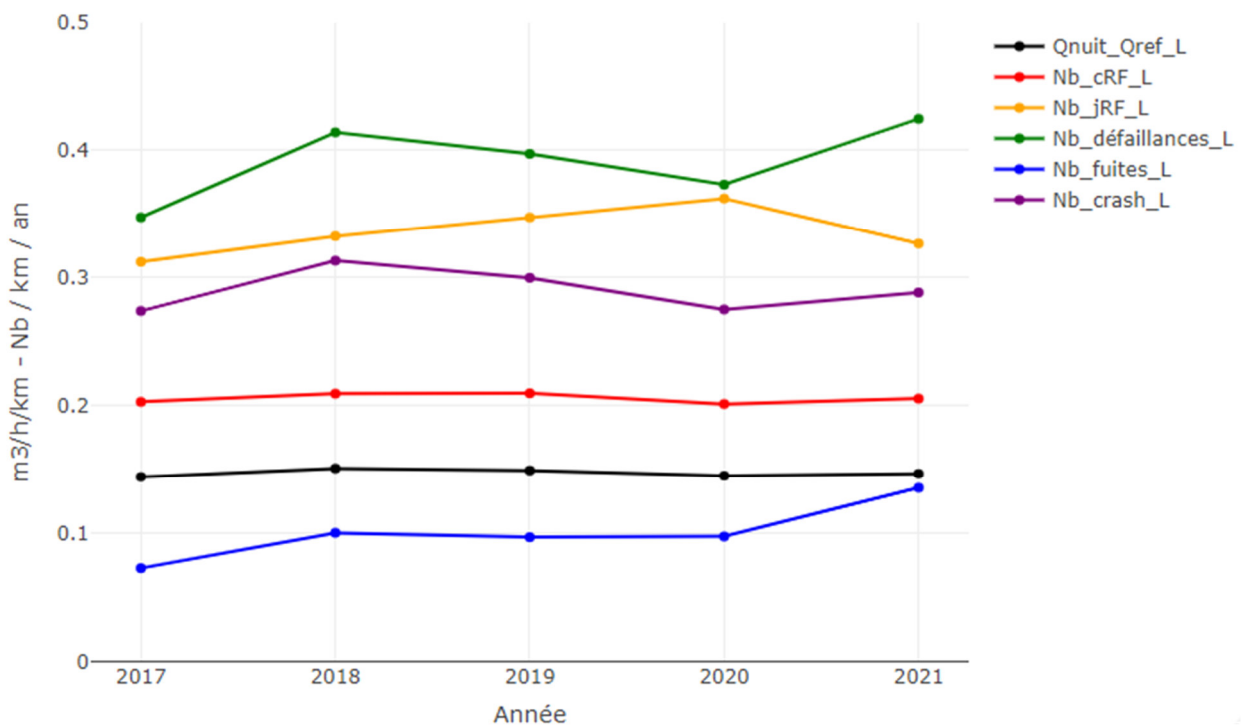


Figure 26. Evolution annuelle des fuites et crashes et du (Qnuit-Qref)/L

5.3.5 Exploration de référence alternative

Sur la sélection des 1 001 cellules (plus de 1 000 Q_{nuit} valables entre 2017 et 2021), les Q_{nuit}/L et Q_{nuit}/N sont étudiés par classes (Annexe T05_02). 10 classes de longueur, puis de raccordements et de densité de raccordement sont créées. Chacune de ces classes est divisée en quantile et comparée à la valeur du Q_{ref} de la classe. L'objectif est de trouver un quantile qui soit stable par rapport à la référence quelle que soit sa classe. En Figure 27, le Q_{nuit}/L quantile 0.30 ou 0.25 apparaît comme un candidat intéressant.

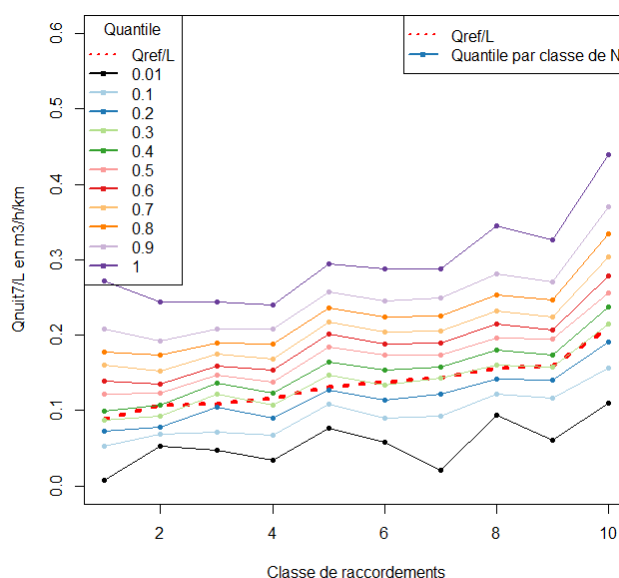


Figure 27. Quantile de Q_{nuit}/L réparti par classe de raccordement.

Des modèles multilinéaires ont été construits pour caractériser $(Q_{nuit}-Q_{ref})/L$ et Q_{nuit}/L (quantile 0,25), cependant si les prédictions par classe sont plutôt bonnes, les prédictions sur les cellules sont trop dispersés pour retenir ces références.

5.4 Complément à l'ILP de référence : Pente

Utiliser l'écart d'ILP à une référence permet de sélectionner des cellules où une réduction des pertes est possible à la fois par des renouvellements ou de la RF sans savoir quelle action serait la plus appropriée. En complément, il est nécessaire de s'intéresser à la dérive temporelle de l'ILP, témoin de l'incapacité à maîtriser les pertes par des actions de RF seules. A l'issue de travaux exploratoires, il apparaît que la pente de la distribution des ILP peut être un bon indicateur, en ignorant les valeurs hautes et basses de débits, l'objectif étant de cibler des dérives tendancielle et non ponctuelle. Cette pente est donc calculée entre les quantiles 75 et 25 de la série chronologique des ILP de l'année en cours.

On définit la pente comme suit :

$$P_n = \frac{(Q_{75} - Q_{25}) / (0.75 - 0.25)}{\min(Q_5, ILP_{ref})}$$

où Q_x est la valeur supérieure à $x\%$ des ILP journaliers de l'année n .

La pente est calculée seulement si au moins 183 valeurs de débits journaliers de la cellule sont disponibles l'année n considérée.

En Figure 28, la pente des *ILP* est faible et stable entre 2020 et 2021, le pic provoqué par un crash début 2021 n'est pas pris en compte. Par contre en Figure 29, la pente des *ILP* est forte en 2021, confirmée en 2020, les *ILP* sont instables.

La variable « pente » est utilisée comme indicateur de la tâche 10.

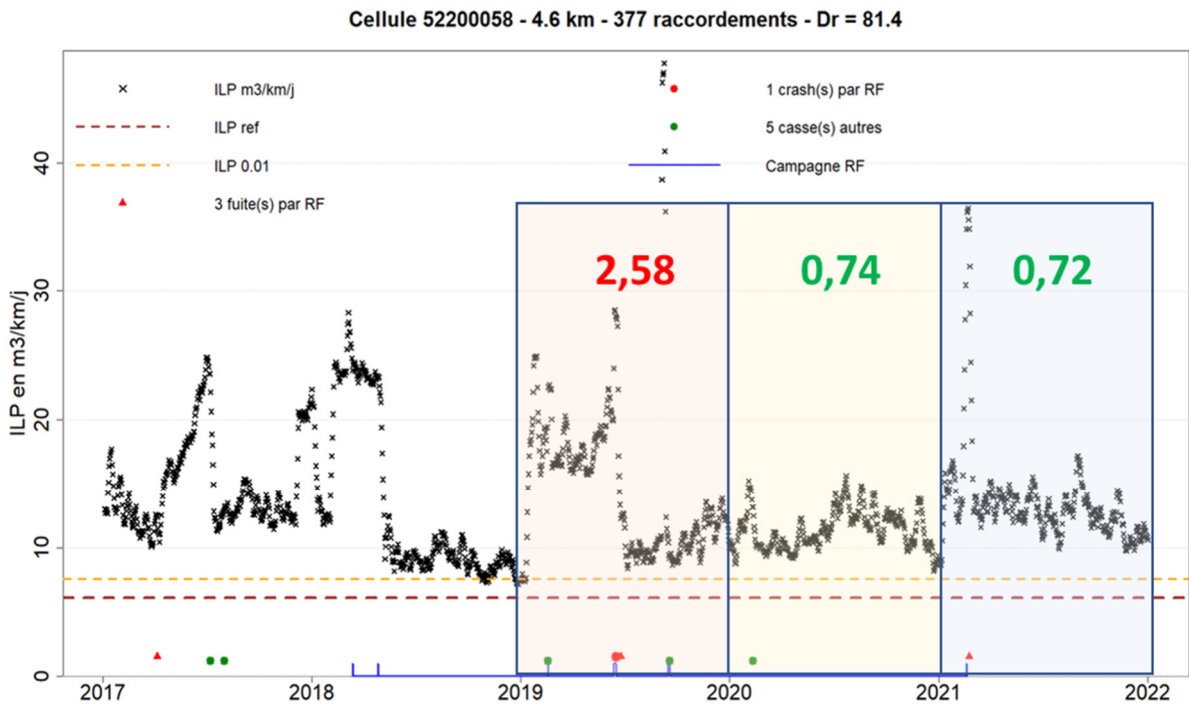


Figure 28. Pente stabilisée en 2020 et 2021.

Cellule 54400098 - 10.6 km - 418 raccords - Dr = 39.4

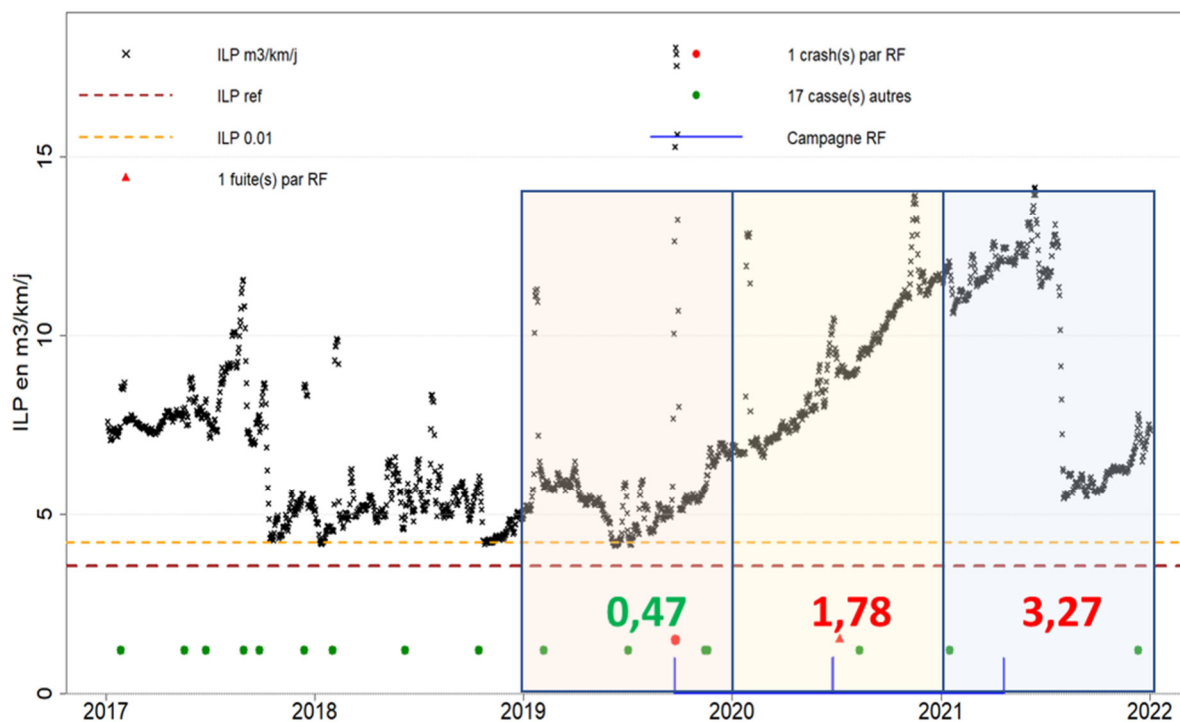


Figure 29. Instabilité des ILP de la cellule confirmée en 2020 et 2021.

Archive Tâche 05 :

- Dossier evolution_cellules : fichiers de suivi des évolutions des noms/numéros de cellule au cours du temps

6 Tâche 06 - Analyse des pratiques de recherche de fuites

La tâche 6 s'intéresse aux conditions de mise en œuvre de la démarche de recherche de fuites développée par la SWDE, afin d'en analyser les effets sur les activités de gestion patrimoniale du réseau de distribution. Pour ce faire, la tâche 6 postule des effets de réciprocité entre la démarche de recherche de fuites (et ses résultats) et son milieu sociotechnique de déploiement (dispositifs techniques, modes d'organisation, dimension sociale). L'analyse fine de ce dernier vise alors à : mieux saisir les finalités de la démarche ; mieux connaître les pratiques qui lui sont liées ; rendre compte des interactions entre les dispositifs techniques, les individus et les modes d'organisation ; évaluer l'appropriation de la recherche de fuites aux activités de gestion du réseau d'eau de la SWDE.

La tâche 6 s'appuie sur une enquête sociologique qualitative réalisée entre janvier et mai 2022, auprès des agents de recherche de fuite et de leur hiérarchie. Le matériau est constitué de plus de quarante sessions d'observation in situ, complétées d'une quinzaine d'entretiens semi-directifs.

Le rapport intégral de l'étude (Annexe_T06_01 : Baati et Caillaud 2023) dévoile l'existence d'un milieu sociotechnique reposant sur la production de données, et par extension sur des chaînes de traduction censées objectiver et localiser les fuites sur le réseau. La recherche de fuites apparaît alors comme une démarche autant – voire davantage – sociocognitive que technique, du fait du poids des connaissances (en partie diffuses) et des compétences pratiques et sensorielles que l'activité requiert sur le terrain. Enfin, la recherche de fuites renvoie à une série d'activités dont la pérennité, la maturité et la « naturalisation » sont cependant rendues compliquées par les transformations successives des dispositifs techniques et organisationnels de la SWDE.

Archive Tâche 06 :

- Notes méthodologiques

7 Tâche 07 – Modélisation de la probabilité d'un tronçon d'être fuyard

7.1 Principe et rappel

La maîtrise des pertes en eaux suppose une compréhension du phénomène de fuite des canalisations d'eau potable. Le repérage via les équipes de recherche de fuite et la réparation rapide des fuites limitent les pertes en eau du réseau. Cependant, une connaissance *a priori* de l'état fuyard du réseau permettrait des actions de prévention ou de recherche de fuite ciblée et optimiserait les plans de renouvellement. La tâche 07 de modélisation de la probabilité d'état fuyard des canalisations s'inscrit donc dans cette volonté de meilleure compréhension du phénomène de fuite et de sa durée avant qu'elle n'évolue en crash (voir l'Annexe 12 du rapport intermédiaire 2021, (Rodriguez et al. 2021).

Cette modélisation se base sur les données des équipes de recherche de fuite, à la fois sur l'action de recherche sur le terrain (une « campagne » de recherche de fuite, ou « campagne RF ») et sur le résultat de cette recherche (une fuite et/ou un crash trouvé(e) ou non à l'issue de cette campagne). La variable d'intérêt est donc le résultat d'une inspection pour chaque tronçon de la cellule inspectée : soit 0 si le tronçon ne fuyait pas, soit 1 si une fuite est détectée sur le tronçon (la valeur 1 est donc donnée uniquement aux fuites –et non aux crashes– ayant eu lieu sur la canalisation elle-même -types d'activités SAP 100, 105 ou 999- trouvée par les équipes de RF).

De ce fait, nous étudions ici la probabilité qu'un tronçon soit dans un état fuyard ($Y=1$) et qu'il a été inspecté ($R=1$), et selon ses propres covariables (X) et celles qui ont mené à l'inspection de la cellule (W) :

$$P(Z = 1) = P(Y = 1 \cap R = 1, X, W) \quad (1)$$

La nature binaire (0 ou 1) de notre variable d'intérêt Y nous empêche ainsi d'utiliser des méthodes de régression linéaire classiques qui travaillent sur des variables continues sur $] -\infty; +\infty[$. De plus, les régressions (logistiques ou non) se basent sur des données censées être indépendantes les unes des autres. Or, ce n'est pas vrai dans le cas présent :

- Un tronçon peut apparaître plusieurs fois dans le jeu de données si sa cellule a été inspectée plusieurs fois ;
- La gestion des équipes de recherche de fuites au niveau des secteurs de distribution implique que tous les tronçons de la SWDE ne sont pas inspectés aléatoirement par une équipe de RF aléatoire mais que les équipes sont réparties par zones géographiques (secteur de distribution, cellules...), ce qui peut avoir un impact sur l'efficacité de la recherche (avec en particulier un « effet secteur »).

La première tentative de modélisation de ce phénomène était une modélisation via une régression logistique simple ou mixte (voir Rapport intermédiaire 2022 : Husson et al. 2022, Partie 8), modélisation classique pour une variable à expliquer binaire. Cependant, cette modélisation n'était pas satisfaisante pour une raison fondamentale : elle supposait l'inspection **exhaustive** de la cellule concernée lors d'une campagne RF. Or cette hypothèse forte n'est pas du tout respectée sur le terrain puisque les agents s'arrêtent de chercher après avoir localisé une fuite « correspondant vraisemblablement » à celle recherchée (Voir Section 6). Cela conduisait à la prédiction d'une probabilité d'état fuyard extrêmement faible pour tous les tronçons puisque l'événement considéré inclut l'effort inspection sur la cellule et la

fuite réelle du tronçon, sans distinction possible. (voir Rapport intermédiaire 2022 : Husson et al. 2022, Partie 8.3).

Afin de pallier le problème de l'exhaustivité de l'inspection des cellules lors d'une campagne RF, nous avons développé un modèle entièrement nouveau adapté aux données disponibles, inspiré des modèles de présence/absence utilisés en écologie pour étudier les dynamiques de populations animales.

7.2 Le nouveau modèle : Zero-inflated binomial (ZIB)

7.2.1 Principe

L'observation disponible sur laquelle se base les prédictions d'état fuyard est le résultat d'une campagne de RF ciblée qui est binaire : 1 si le tronçon a été observé¹² et trouvé fuyard, 0 sinon. Pour autant, la valeur 0 couvre deux phénomènes bien distincts : le tronçon a été inspecté et il n'était pas fuyard ou le tronçon n'a pas été inspecté alors qu'il faisait partie du périmètre de la campagne (*i.e.* il appartient à la cellule sur laquelle se déroule la RF).

Afin de distinguer les deux cas de présence non détectée (faux négatifs et absence), nous avons développé un modèle basé sur le principe des modèles zéro-augmentés (Blasco-Moreno et al. 2019; Lambert 1992; Moriña, Puig, et Navarro 2021; Mullahy 1986) en considérant notre variable aléatoire d'observation Z comme étant le produit de deux variables aléatoires latentes, Y étant l'état réel de l'individu (état fuyard du tronçon) et R étant l'indicateur d'observation de l'individu lors de la campagne de recherche. On suppose alors que ces deux variables latentes ne sont pas systématiquement observées mais on observera plutôt Z définie par $Z = R \times Y$ (2)

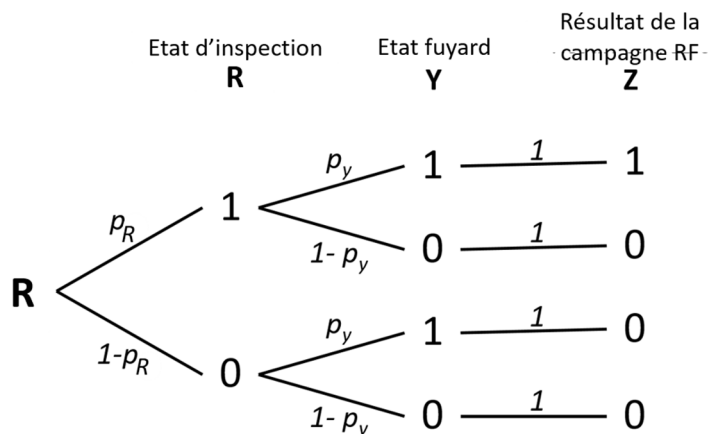


Figure 30. Schéma du modèle de données. Z est la seule variable observée. Inspiré de Moriña et al., (2021).

Sous l'hypothèse que Y et R sont indépendantes (ou conditionnellement indépendantes sachant un ensemble de covariables), on a simplement $P(Z = 1) = P(R = 1) \times P(Y = 1)$, et le problème se résume alors à estimer le modèle de probabilité de présence Y à partir des observations Z . Dans

¹² Nous utilisons ici le terme « d'observation » au sens large, c'est-à-dire que le tronçon a été inspecté de quelque manière que ce soit (écoute, tronçonnage...) par les équipes de RF.

l'hypothèse de couverture globale, $R_i = 1$ pour tout i , alors $Z_i = Y_i$ et les états exhaustivement observés de présence (ou de non-présence) permettent une régression logistique standard qui peut estimer statistiquement les paramètres du modèle :

$$P(Y = 1|X = x) = \frac{1}{1 + e^{-\beta^T x}} = \frac{e^{\beta^T x}}{1 + e^{\beta^T x}} \quad (3)$$

où β_x est la combinaison linéaire $\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$, avec $x = (1, x_1, \dots, x_p)$. Dans la suite on définit un modèle permettant d'estimer le paramètre β à partir de données (Z_i, X_i) , $i = 1 \dots n$.

7.2.2 Modèle d'inspection

La variable binaire R codant pour l'inspection de l'individu i lors de la campagne c vérifie :

$$P(R_i = 1|W = w_i) = \frac{1}{1 + e^{-\theta^T w_i}} \quad (4)$$

7.2.3 Modèle d'état fuyard

La variable binaire Y codant pour l'inspection de l'individu i lors de la campagne c vérifie :

$$P(Y_i = 1|X = x_i) = \frac{1}{1 + e^{-\beta^T x_i}} \quad (5)$$

7.2.4 Modèle ZIB

Sous l'hypothèse d'indépendance de Y et R , on a :

$$P_{(\beta, \theta)}(Z = 1|W = w, X = x) = \left(1 + e^{-\theta^T w}\right)^{-1} \left(1 + e^{-\beta^T x}\right)^{-1} \quad (6)$$

7.2.5 Estimation des paramètres

Nous proposons d'estimer l'ensemble des paramètres (θ, β) par maximisation de la log-vraisemblance calculée sur un échantillon $(z_i, x_i, w_i)_{i=1 \dots n}$. Pour plus de détails sur le calcul de la fonction de vraisemblance ainsi que du score de Fischer nécessaire à l'optimisation rapide via l'algorithme BFGS choisi, vous pouvez vous référer à l'Annexe T07_01.

Le modèle, ainsi que les différentes fonctions afférentes nécessaires à l'estimation des paramètres ont été codées sous forme de fonctions sous le logiciel R, dont tous les paramètres sont présentés dans l'Annexe T07_02.

7.3 Test de la robustesse du modèle

7.3.1 Principe

Afin de tester la robustesse et la fiabilité du modèle ZIB présenté ci-dessus, nous avons effectué plusieurs simulations afin de vérifier si :

- 1) Le modèle était capable d'estimer correctement les paramètres sur un petit jeu de données fictif simulé dans le cas parfait (toutes les covariables sont connues) et dans plusieurs types de cas imparfaits (covariable oubliée, ou mise dans le modèle d'état fuyard alors qu'elle concernait l'inspection, etc...).
- 2) Le choix des paramètres initiaux de l'optimisation influençait le résultat final d'estimation. Ce travail a été effectué à la fois sur des données simulées et sur les données réelles de la SWDE.

7.3.1.1 Estimation des paramètres dans les cas imparfaits

7.3.1.1.1 Principe de simulation

Six variables ont été simulées par tirage dans une loi normale de moyenne μ_i et d'écart-type $\sigma_i = \mu_i \times CV$, avec CV le coefficient de variation :

- $X_1 \sim \mathcal{N}(4, 4 \times CV)$, $X_2 \sim \mathcal{N}(2, 2 \times CV)$ et $X_3 \sim \mathcal{N}(1.5, 1.5 \times CV)$ pour les covariables du sous-modèle d'état fuyard.
- $W_1 \sim \mathcal{N}(4, 4 \times CV)$ et $W_2 \sim \mathcal{N}(2, 2 \times CV)$ pour les covariables du sous-modèle d'inspection.
- $E \sim \mathcal{N}(8, 8 \times CV)$ comme covariable supplémentaire simulant une « erreur », c'est-à-dire une covariable n'étant pas impliquée dans le calcul des probabilités des sous-modèles.

A partir de ces données, les probabilités $P(R)$ et $P(Y)$ ont été calculées avec des paramètres θ et β fixés, i.e. $\theta = \{\theta_0: -2; \theta_1: -0.5; \theta_2: 0.5\}$ et $\beta = \{\beta_0: -3; \beta_1: 0.3; \beta_2: 0.7; \beta_3: -1\}$ (6). Les variables latentes R et Y ont été simulées grâce à un tirage dans une loi binomiale de paramètre $p_R = P(R)$ et $p_Y = P(Y)$ respectivement. Enfin, la variable observable Z a été calculée comme le produit de R et Y (2).

Trois cadres de simulations ont été étudiés :

- 1) Petite population de tronçons ($N = 50000$) et petite variance des données simulées ($CV = 0.3$)
- 2) Petite population de tronçons ($N = 50000$) et grande variance des données simulées ($CV = 0.7$)
- 3) Grande population de tronçons ($N = 100000$) et grande variance des données simulées ($CV = 0.7$)

La comparaison des cadres 1 et 2 nous a permis de déterminer l'impact de la variance des données sur l'estimation des paramètres du modèle, tandis que la comparaison des cadres 2 et 3 nous a permis de déterminer l'impact de la taille de la population sur l'estimation.

7.3.1.1.2 Cas étudiés

Neuf cas à l'intérieur de chaque cadre ont été abordés afin d'estimer la robustesse du modèle :

- a) Cas « parfait » de référence. Toutes les covariables utilisées pour calculer Z ont été incluses dans le modèle ZIB. Données : $Y \sim X_1 + X_2$, $R \sim W_1 + W_2$ et estimation : $ZIB \sim Y = X_1 + X_2$; $R \sim W_1 + W_2$
- b) Cas croisement 1. La variable W_1 est utilisée pour simuler R mais est ajoutée à la fois dans Y et R pour l'estimation du ZIB. Données : $Y \sim X_1 + X_2$, $R \sim W_1 + W_2$ et estimation : $ZIB \sim Y = X_1 + X_2 + W_1$; $R \sim W_1 + W_2$

- c) Cas croisement 2. Cas inverse du croisement 1 avec X_1 . Données : $Y \sim X_1 + X_2$, $R \sim W_1 + W_2$ et estimation : $ZIB \sim Y = X_1 + X_2$; $R \sim W_1 + W_2 + X_1$
- d) Cas croisement 3. Mélange des deux précédents. Données : $Y \sim X_1 + X_2$, $R \sim W_1 + W_2$ et estimation : $ZIB \sim Y = X_1 + X_2 + W_1$; $R \sim W_1 + W_2 + X_1$
- e) Cas variable commune. La même covariable X_2 a été utilisée pour simuler Y et R. Données : $Y \sim X_1 + X_2$, $R \sim W_1 + X_2$ et estimation : $ZIB \sim Y = X_1 + X_2$; $R \sim W_1 + X_2$
- f) Cas variable oubliée. La variable X_3 a été utilisée pour simuler Y mais « oubliée » dans l'estimation du ZIB. Données : $Y \sim X_1 + X_2 + X_3$, $R \sim W_1 + W_2$ et estimation : $ZIB \sim Y = X_1 + X_2$; $R \sim W_1 + W_2$
- g) Cas variable ajoutée. La variable E est incluse dans l'estimation du ZIB alors qu'elle n'a pas servi à simuler les données. Données : $Y \sim X_1 + X_2$, $R \sim W_1 + W_2$ et estimation : $ZIB \sim Y = X_1 + X_2 + E$; $R \sim W_1 + W_2$
- h) Cas mixte variables oubliées et ajoutées. Mélange des deux cas précédents. Données : $Y \sim X_1 + X_2 + X_3$, $R \sim W_1 + W_2$ et estimation : $ZIB \sim Y = X_1 + X_2 + E$; $R \sim W_1 + W_2$
- i) Cas mixte double. Données : $Y \sim X_1 + X_2$, $R \sim W_1 + W_2$ et estimation : $ZIB \sim Y = X_1 + W_2$; $R \sim W_1 + X_2$

Chaque cas a été simulé 100 fois pour les 3 cadres afin d'obtenir une distribution des estimations des paramètres intra-cas et intra-cadre.

7.3.1.2 Impact du choix des paramètres initiaux de l'optimisation sur l'estimation des paramètres

Les algorithmes d'optimisation quels qu'ils soient peuvent avoir leurs performances impactées par les valeurs initiales des paramètres à estimer. En effet, la démarche globale étant de trouver un jeu de paramètres maximisant la vraisemblance du modèle (ou minimisant l'opposé de sa vraisemblance selon les algorithmes) par pas itératifs, le jeu de paramètres initiaux peut conduire l'algorithme à rester « coincé » dans un optimum local, et il n'atteindra ainsi jamais les valeurs du jeu de paramètres correspondant à l'optimum global de la vraisemblance.

De ce fait, nous avons testé l'impact de ce jeu de paramètres initiaux en tirant aléatoirement des jeux de paramètres initiaux pour quatre variables explicatives sur une population de 50000 individus : $Y \sim X_1 + X_2$ et $R \sim W_1 + W_2$. Les 4 variables explicatives suivent des lois normales, de moyennes respectives 4, 2, 4 et 2 et d'écart type 1 ou 2 selon la simulation. 50 répliques de 100 simulations ont été effectuées pour chaque valeur d'écart type, soit au total un tirage aléatoire de 10000 jeux de 6 paramètres initiaux ($\beta_0, \beta_1, \beta_2, \theta_0, \theta_1, \theta_2$).

7.3.2 Résultats

7.3.2.1 Estimation des paramètres dans les cas imparfaits

Le travail sur un jeu de données simulées simples (maximum 3 covariables par sous-modèle d'inspection ou d'état fuyard) a permis de montrer que le modèle ZIB a un comportement robuste dans la grande majorité des situations créées. Il a ainsi été montré que dans un cadre avec une population de tronçons importante ($N = 100000$) et des covariables avec une variance suffisante, le modèle estime correctement les paramètres demandés, avec de petits écart-types (l'estimation est donc précise). Ainsi,

le modèle est capable de discriminer les variables explicatives de chacun des sous-modèles séparément (Tableau 10, Figure 31 à Figure 34).

Seul le cas i) (mixte double) a posé des difficultés au modèle ZIB, avec l'incapacité pour lui de différencier les variables ajoutées des oubliées, et ce même dans le cas le plus favorable avec une grande population et une grande variance des données. Cependant, le jeu de données contenait peu de covariables (2 pour chaque sous modèles), ce qui explique vraisemblablement la difficulté du modèle à séparer leur effet. Cela devrait de fait être atténué sur les données réelles ou les deux sous-modèles auront une dizaine de covariables associées chacun. De plus, le cas i) est la simulation du cas le plus défavorable sur des données sans « sens métier », et la connaissance des processus de recherche et de fuite devraient permettre d'éviter cette situation dans les estimations réelles.

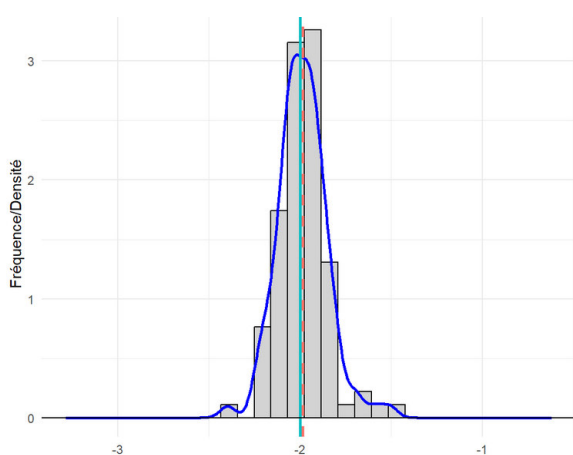


Figure 31 . Estimation de θ_0

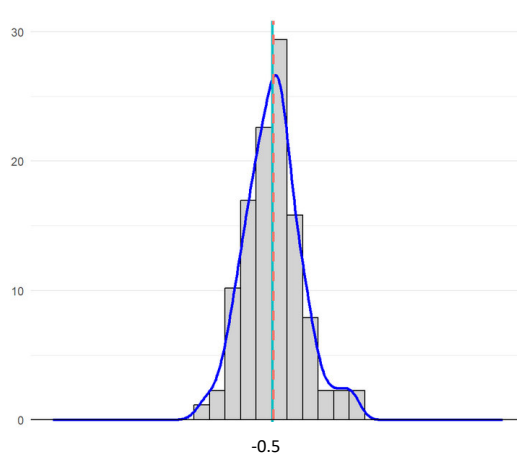


Figure 32. Estimation de θ_1

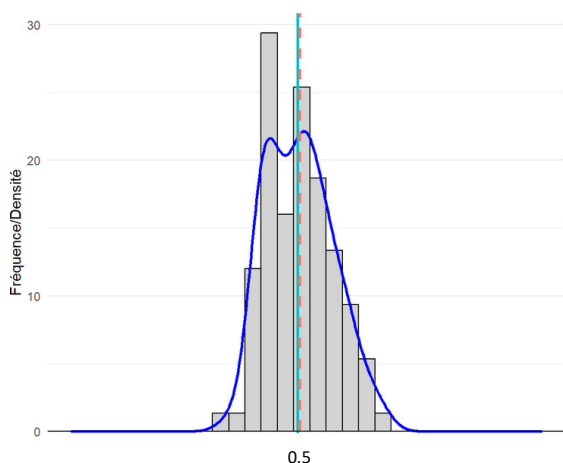


Figure 33. Estimation de θ_2

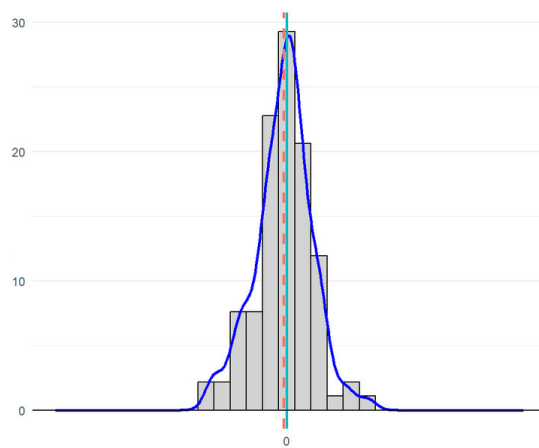


Figure 34. Estimation de θ_3

Tableau 10. Exemple de résultats d'estimation des paramètres du sous-modèle d'inspection du ZIB dans le cas d) du cadre 3 (i.e. Croisement de variable de type 3 dans une grande population avec une grande variance des données). La densité est superposée aux histogrammes de résultats (courbes bleues), la valeur attendue est représentée par le trait vertical bleu clair et la moyenne de la distribution par le trait pointillé orange. θ_3 est le coefficient associé à la variable X_1 qui n'est pas utilisé pour simuler les données d'inspection, sa valeur attendue est donc 0.

7.3.2.2 Impact du choix des paramètres initiaux de l'optimisation sur l'estimation des paramètres

7.3.2.2.1 Données simulées

Les paramètres à estimer étaient fixés à $\theta = (2, -0.5, 0.3)$ et $\beta = (-3, 0.3, -0.1)$. Lorsque l'écart type était petit ($\sigma = 1$), les estimations ne convergeaient pas toujours vers les valeurs réelles correspondantes et les cas aberrants pouvaient être nombreux (Figure 35 et Figure 36). Cependant, dès que la dispersion des données augmente ($\sigma = 2$), les estimations convergent nettement mieux vers les valeurs réelles (Figure 37 et Figure 38). L'impact des valeurs initiales est donc très net en cas de données peu dispersées mais très fortement diminué lorsque la dispersion des données augmente sur ces données simulées simples. Nous avons donc testé cet impact sur un modèle de données réelles afin de vérifier la bonne convergence des estimations.

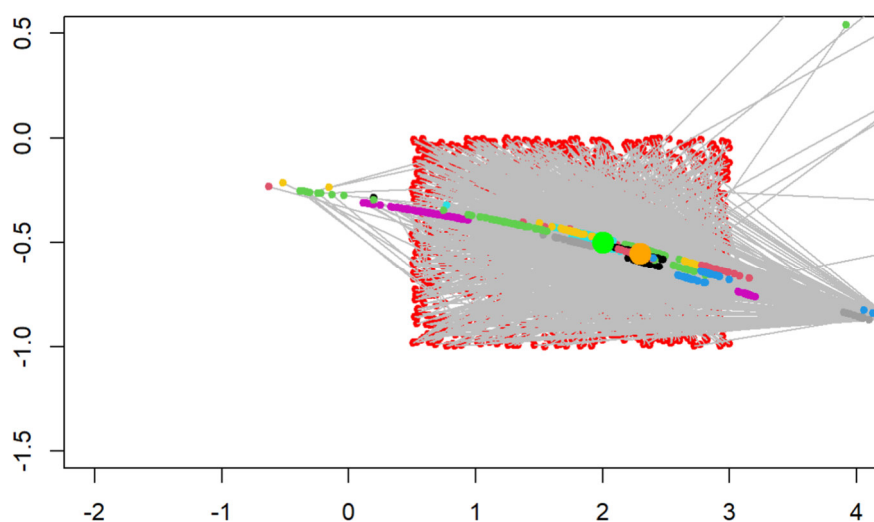


Figure 35. Valeurs estimées de θ_0 en fonction de θ_1 selon différentes valeurs initiales avec un faible écart type des données ($\sigma = 1$). Les points rouges représentent le couple de valeurs initiales qui est relié en gris à l'estimation correspondante par l'algorithme BFGS (petits points de couleur). Le gros point vert clair représente le couple de valeur réelles (θ_0, θ_1).

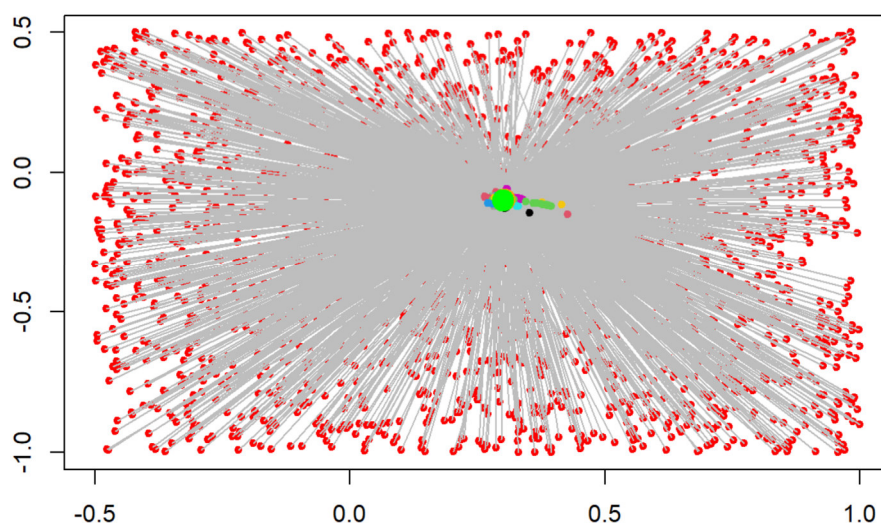


Figure 36. Valeurs estimées de β_1 en fonction de β_2 selon différentes valeurs initiales avec un faible écart type des données ($\sigma = 1$). Les points rouges représentent le couple de valeurs initiales qui est relié en gris à l'estimation correspondante par l'algorithme BFGS (petits points de couleur). Le gros point vert clair représente le couple de valeur réelles (β_1, β_2).

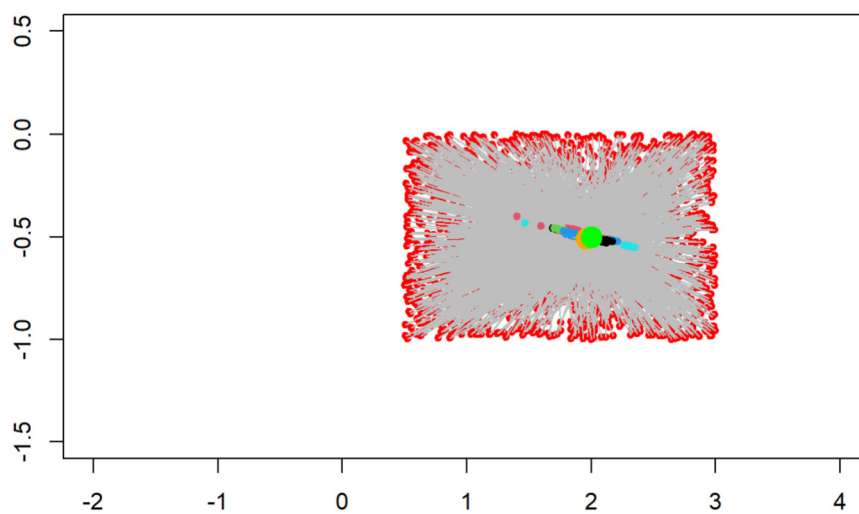


Figure 37. Valeurs estimées de θ_0 en fonction de θ_1 selon différentes valeurs initiales avec un écart type des données augmenté ($\sigma = 2$). Les points rouges représentent le couple de valeurs initiales qui est relié en gris à l'estimation correspondante par l'algorithme BFGS (petits points de couleur). Le gros point vert clair représente le couple de valeur réelles (θ_0, θ_1).

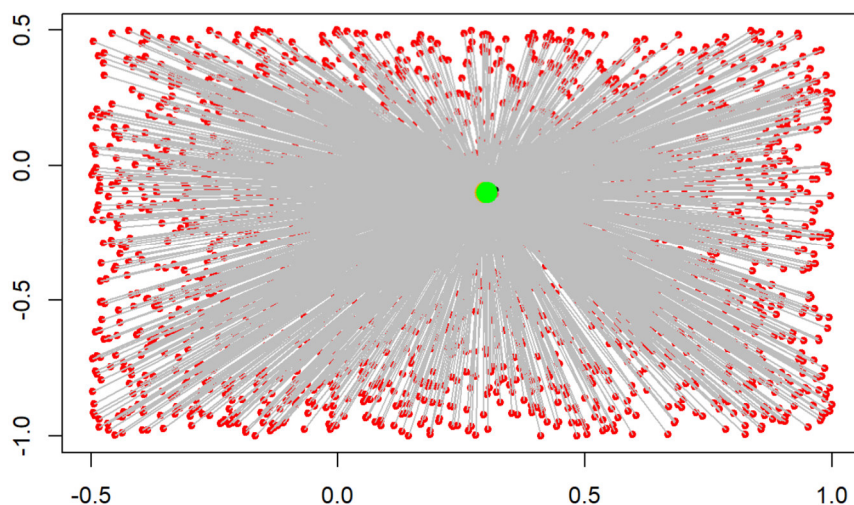


Figure 38. Valeurs estimées de β_1 en fonction de β_2 selon différentes valeurs initiales avec un écart type des données augmenté ($\sigma = 2$). Les points rouges représentent le couple de valeurs initiales qui est relié en gris à l'estimation correspondante par l'algorithme BFGS (petits points de couleur). Le gros point vert clair représente le couple de valeur réelles (β_1, β_2).

7.3.2.2.2 Données réelles

L'apparition de cas aberrants avec les données simulées nous a interrogé sur la capacité du modèle à converger vers une estimation stable avec les données réelles. Nous avons donc utilisé un jeu de données contenant les informations des campagnes de RF sur 2018-2021 sur 190 689 tronçons sur 1935 cellules lors de 19253 campagnes de recherche de fuite. Au total, 4969 fuites ont été trouvées lors de ces campagnes de RF. Les données se composaient ainsi de 2 522 445 lignes contenant 99,8% de $Z = 0$. Les données utilisées n'avaient à l'époque pas encore été complétées comme expliqué en Annexe T01_01.

Un modèle ZIB a été construit en utilisant 11 variables pour le modèle d'état fuyard et 7 pour le modèle d'inspection, conduisant à estimer 60 paramètres (voir Tableau 11). Les paramètres initiaux de référence de l'optimisation étaient les estimations d'une régression logistique standard pour le sous-modèle d'état fuyard et 0 pour le sous-modèle d'inspection (excepté $\theta_0 = 1$), ce qui correspond au comportement standard de la fonction encodée si l'utilisateur ne précise pas de paramètres initiaux. Nous avons ensuite effectué 30 tirages aléatoires du jeu de paramètres initiaux (*i.e.* $\{\theta_0, \theta_1, \dots, \theta_{28}, \beta_0, \beta_1, \dots, \beta_{32}\}$) dans l'intervalle $[0.5 \times \theta_i; 1.5 \times \theta_i]$ (respectivement β_i) et estimé les 60 paramètres à partir de ces 31 jeux de paramètres initiaux.

Tableau 11. Covariables du modèle ZIB utilisé pour tester l'influence des paramètres initiaux d'optimisation sur l'estimation des paramètres

Etat fuyard Y		
Matériau (7 modalités)	Age à l'inspection (siècle)	Proportion d'argile dans le sol
Longueur (m, échelle log)	Nombre de défaillances RF passées sur conduites	Corrosivité de l'eau
Diamètre (dm)	Pression (millibars)	Secteur (17 modalités)
Nombre de branchements (échelle log)	Proportion du tronçon sous la chaussée	
Etat d'inspection R		
Durée de la campagne (h)	Matériau (7 modalités)	Nombre de branchements (échelle log)
Longueur de la cellule (km)	Diamètre (dm)	Proportion d'argile dans le sol
Secteur (17 modalités)		

Sur les 31 estimations, 7 ont eu un comportement aberrant, conduisant à des estimations de paramètres différentes des 23 autres (Figure 39 et Figure 40). Cependant, quelque soient les covariables, les estimations issues du jeu de paramètre initial issus de la régression logistique standard convergeaient systématiquement avec les estimations majoritaires. De plus, cette optimisation conduisait à une log-vraisemblance maximale parmi les 31 (Figure 39, opposé de log-vraisemblance minimal), ce qui confirme que le choix par défaut des paramètres initiaux de l'optimisation sont judicieux. Cela démontre aussi la prudence à avoir si l'utilisateur choisi lui-même les paramètres initiaux et donc nécessite de faire des vérifications supplémentaires afin de vérifier que l'optimisation se fait correctement.

Des résultats plus exhaustifs sont disponibles dans l'Annexe T07_03.

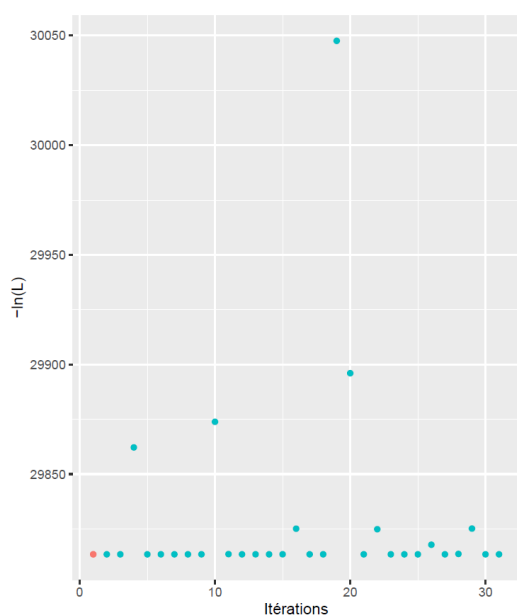


Figure 39. Opposé de la log-vraisemblance après optimisation du modèle ZIB selon les 31 jeux de paramètres initiaux d'optimisation. La meilleure optimisation est celle avec $-\ln(L)$ minimum. En rouge, le jeu de paramètres issu de la régression logistique standard.

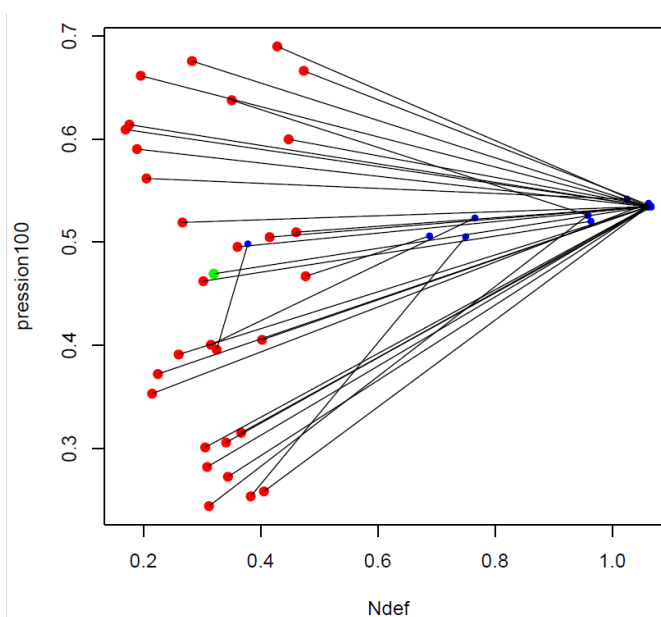


Figure 40. Valeur des estimations du nombre de défaillances passée (Ndef) et de la pression (pression100). En vert, les paramètres initiaux issus de la régression logistique, en rouges, ceux issus des 30 tirages aléatoires et en bleu les estimations des paramètres après optimisation.

En conclusion, tous les tests effectués sur le modèle ZIB, en données simulées comme réelles, ont démontré une convergence rapide (une vingtaine de minutes pour les modèles avec plus de 60 paramètres à estimer sur les données réelles), fiable et un modèle théoriquement satisfaisant robuste aux erreurs.

Nous avons donc appliqué ce modèle aux dernières données disponible de la SWDE afin de déterminer la combinaison de covariables optimales expliquant les données et de prédire les probabilités d'inspection et d'état fuyard sur le réseau entier.

7.4 Analyse des données de la SWDE

7.4.1 Données d'entrée

7.4.1.1 Campagnes

Tout comme la régression logistique utilisée précédemment, le modèle ZIB se base sur la notion de campagne de recherche de fuite (voir Rapport intermédiaire 2022 : Husson et al. 2022, Partie 2.5.3 et 2.5.4 et son Annexe 8). Pour rappel, cela consiste à regrouper ensemble des Ordres de travail de RF rapprochés dans le temps et l'espace en une seule campagne RF ayant un résultat (fuite ou crash réparé à la suite de cette campagne ou pas de réparation effectuée). Le jeu de données comporte ainsi la duplication des tronçons du nombre de fois que leur cellule a été soumise à une campagne RF. De ce fait,

le nombre de campagne est directement responsable du nombre de zéros et donc de faux zéros présents dans les données (voir Figure 30 et paragraphe 7.2.1 page 75).

Le nombre de campagnes est influencé par deux délais :

- Le délai inter-opération, qui est le délai maximum entre deux OT permettant de les regrouper en une seule et même campagne RF (en jours ouvrés, optimum à 3 jours ouvrés, Annexe 8 du Rapport intermédiaire 2022 : Husson et al. 2022). Si le délai inter-opération augmente, le nombre de campagnes RF diminue.
- Le délai inter-campagne, qui est le délai maximum entre 2 campagnes permettant de les regrouper en « méga-campagnes ». Ce délai a été introduit afin de limiter le nombre de campagnes infructueuses (*i.e.* n'ayant trouvé ni crash ni fuite sur les types d'activités concernant les pertes en eau), et on ne peut ainsi donc pas regrouper deux campagnes fructueuses entre elles.

Nous avons donc testé l'influence de ces délais sur les estimations du modèle ZIB, en faisant varier le délai inter-opération à 3,5 ou 15 jours ouvrés, et les données avec ou sans regroupement en mégacampagnes pour les 3 délais sus-cités (délai inter-campagne de 2 ans). Il faut bien noter ici que la seule variation entre les jeux de données est la proportion de $Z = 1$ et non leur nombre absolu. Les covariables choisies étaient peu nombreuses (5 pour l'état fuyard, 2 pour l'inspection, conduisant à l'estimation de 14 paramètres) afin de garder un modèle simple à valeur d'exemple.

Si le regroupement en méga-campagnes et l'augmentation du seuil inter-opération modifiaient effectivement le nombre de campagnes (d'environ 20 000 pour 3 jours ouvrés sans regroupement à 7500 pour 15 jours ouvrés avec regroupement) et donc la proportion de $Z = 1$ (de 0.2 à 0.5% pour les deux cas cités précédemment), les estimations des paramètres du ZIB ne variaient pas significativement. De plus, les variations des délais ne changeaient pas les rangs des tronçons sur leur probabilité d'état fuyard. De ce fait, nous avons décidé de maintenir les choix précédents d'un délai inter-opération de 3 jours ouvrés et de ne pas regrouper les campagnes en mégacampagnes.

Les résultats complets de cette étude sont disponibles dans l'Annexe T07_04.

7.4.1.2 Données finales

Les données finales se composent des données de campagnes et de leurs défaillances rattachées sur la période 2018-2022, après complétion des caractéristiques des tronçons comme spécifié en Annexe T01_01. Seuls les tronçons les plus longs de leur agrégat et de plus de 6 m sont utilisés. De plus, seuls les tronçons encore en service auront une prédiction de leur probabilité d'état fuyard et d'inspection.

Comme pour le modèle LEYP (Rapports intermédiaires 2021 et 2022), nous avons découpé le jeu de données complet en deux :

- Pour la calibration du modèle, seules les campagnes du 01/01/2018 au 31/12/2021 sont prises en compte. Le jeu de données comprend 2 522 445 lignes, qui représentent 19 253 campagnes RF sur 190 689 tronçons dans 1935 cellules. Les campagnes ont trouvé 4969 fuites, et donc la variable Z est composée de 99.8% de zéros.

- Pour la validation du modèle, seules les campagnes du 01/01/2022 au 31/12/2022 sont prises en compte. Le jeu de données comprend 589 688 lignes, qui représentent 4362 campagnes RF sur 147 799 tronçons dans 1374 cellules. Les campagnes ont trouvé 1244 fuites, et donc la variable Z est composée de 99.8% de zéros.

7.4.1.3 Méthodes

Sur le jeu de données de calibration, nous avons appliqué une méthode de sélection de modèles dite « backward », c'est-à-dire que nous sommes partis du modèle le plus complet avec le plus de covariables et nous avons supprimé au fur et à mesure les covariables non significatives et regroupé les modalités non significatives des variables qualitatives avec la référence, sur le même principe que la sélection de modèle du logiciel Casses.

Une variable est considérée comme significative quand la p-value de la variable (test du Z score) est inférieure à 0.05, c'est-à-dire lorsque nous avons 5% de risque au maximum de nous tromper en disant que la variable a un impact significatif sur la probabilité d'inspection ou d'état fuyard.

La comparaison des différents modèles entre eux se fait via le critère d'Akaike (AIC, Akaike, 1974), qui se calcule simplement comme

$$AIC = 2k - 2 \ln \mathcal{L} \quad (7)$$

avec k le nombre de paramètres à estimer par le modèle et $\ln \mathcal{L}$ la log-vraisemblance maximisée du modèle. Le meilleur modèle est ainsi celui ayant l'AIC minimum, sachant que l'on considère classiquement que deux modèles sont équivalents si leur différence d'AIC est inférieure à 2.

Nous avons ainsi testé 18 modèles descendants du modèle complet initial (pour la composition exacte des 19 modèles, voir l'Annexe T07_05). Parmi ces 19 modèles, nous avons effectué une validation croisée sur le modèle complet initial ainsi que les 3 meilleurs restants après sélection. La validation croisée consiste à calculer les probabilités prédites de Z sur les données 2022 grâce aux estimateurs calculés sur la période de calibration et de comparer les probabilités prédites aux défaillances observées réellement. De la même façon que le logiciel Casses, nous avons produit des courbes de validation et calculé l'aire sous la courbe comme métrique. Plus l'aire est proche de 1, plus le modèle est bon.

Les modèles sélectionnés via AIC contiennent tous la variable Year_camp, qui est l'année durant laquelle s'est déroulée la campagne. De ce fait, une validation simple ne peut pas se faire puisque le modèle a estimé β_{2019} , β_{2020} et β_{2021} (si pas de regroupement des modalités) et donc ne peut travailler sur les données de validations où Year_camp n'a qu'un seul niveau de facteur égal à 2022. Il a donc été choisi d'utiliser le β_{2021} comme estimateur de Year_camp 2022 (β_{N-1} pour généraliser). Nous avons de plus testé le meilleur modèle (le 17) en lui enlevant la variable Year_camp (modèle 18) afin de connaître l'impact du choix de l'utilisation du β_{2021} .

Les courbes de validation sont la représentation du pourcentage du nombre de tronçons (courbe bleue, tri décroissant de la probabilité $P(Z)$ ou du pourcentage du linéaire de réseau (courbe orange, tri décroissant de la probabilité $P(Z)/\text{Longueur}$) en fonction du pourcentage du nombre de $Z=1$ observés (Figure 41 à Figure 45). Le point $f(0.05)$ est aussi calculé : il représente le pourcentage de fuite évitées en

changeant les 5% les plus à risques (en nombre pour le tri sur $P(Z)$ ou en longueur pour le tri sur $P(Z)/\text{Longueur}$).

7.4.2 Résultats

7.4.2.1 Sélection de modèles

Le modèle complet comprend 23 covariables et 69 paramètres à estimer dû aux variables qualitatives où un paramètre est estimé par modalité (sauf pour la modalité de référence). Il converge en 22 min et sa courbe de validation est déjà bonne avec une aire sous la courbe de 0.84 malgré son AIC élevé (Tableau 12 et Figure 41).

Tableau 12. Récapitulatif des AIC des 19 modèles testés.

Modèle	Complet	1	2	3	4	5	6	7	8	9
AIC	62468	62470	62414	62454	62412	62470	62612	62458	62822	63220
Modèle	10	11	12	13	14	15	16	17	18	
AIC	62482	62412	62412	62412	62412	62412	62402	62400	62432	

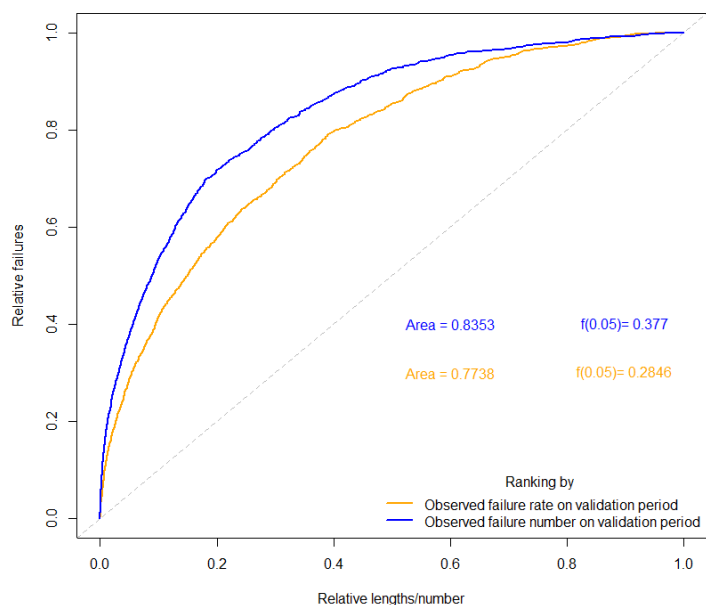


Figure 41. Courbe de validation du modèle complet.

Nous avons ainsi appliqué la validation croisée aux 3 meilleurs modèles (15, 16 et 17) ainsi qu'au modèle 18 (Figure 42 à Figure 45). Les 4 validations ont des aires sous la courbe très élevées (minimum 0.75) et des pentes à l'origine fortes ($f(0.05)_{min} = 0.24$), ce qui confirme la pertinence de notre modèle et sa validité à estimer les paramètres sur des données réelles.

Les résultats exhaustifs de calibration et de validation de ces 5 modèles sont disponibles dans l'Annexe T07_06.

Au vu de ces résultats, nous avons choisi d'utiliser le modèle 17, qui a l'AIC minimum et une très bonne validation, pour produire les probabilités prédites de $P(Y)$ utilisées par la Tâche 10 (voir page 146). Ses résultats sont détaillés dans le paragraphe suivant.

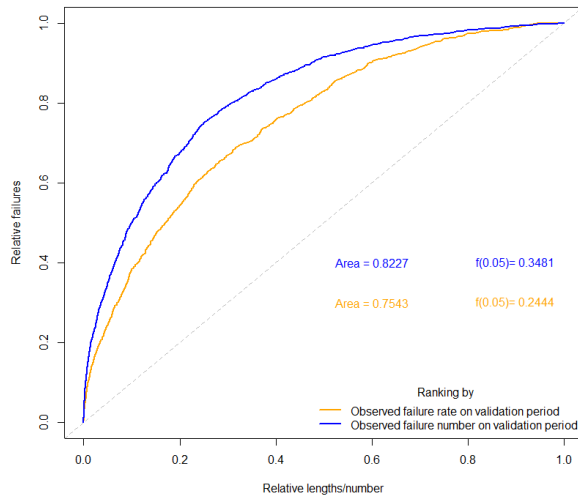


Figure 42. Courbe de validation du modèle 15.

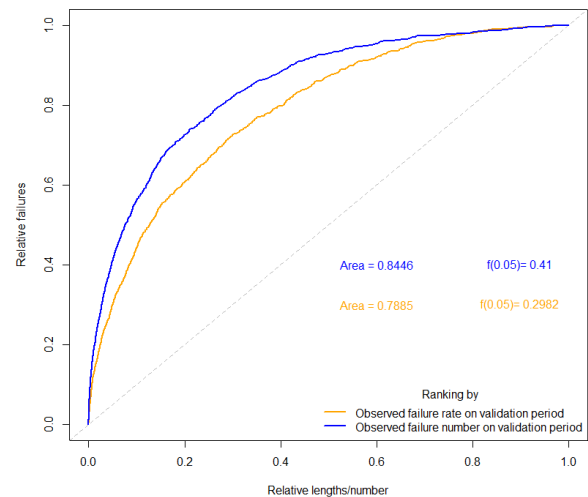


Figure 43. Courbe de validation du modèle 16.

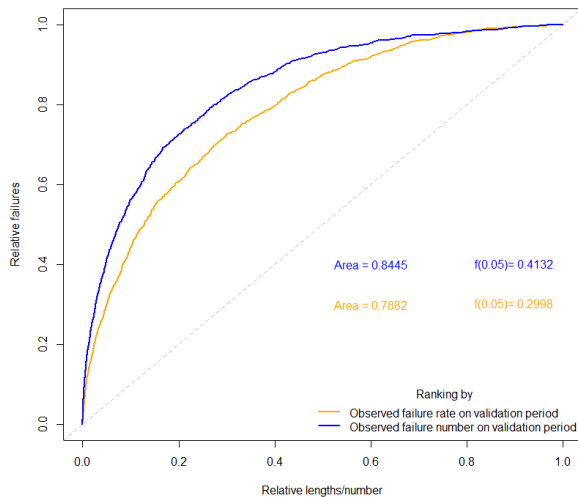


Figure 44. Courbe de validation du modèle 17.

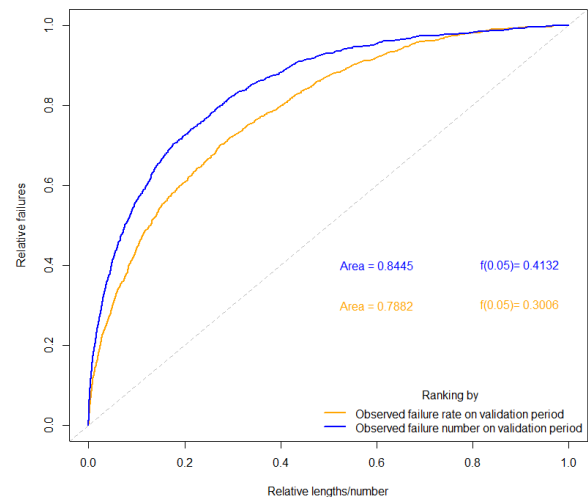


Figure 45. Courbe de validation du modèle 18.

7.4.2.2 Meilleur modèle : modèle ZIB 17

7.4.2.2.1 Calibration

Le modèle ZIB estime 47 paramètres et converge rapidement (13 minutes).

Les deux sous-modèles utilisés sont les suivants :

Etat fuyard Y		
MAT3 : Matériau (6 modalités)	Age_Inspect_S : Age à l'inspection (siècle)	Pargile : Proportion d'argile dans le sol
Lnlong : Longueur (m, échelle log)	Ndef : Nombre de défaillances sur conduites RF passées	Plimon : Proportion de limon dans le sol
Diam100 : Diamètre (dm)	Pression100 : Pression (millibars)	ID_ZF3 : Zones fonctionnelles (9 modalités)
Lnbr1 : Nombre de branchements (échelle log)	Proute : Proportion du tronçon sous la chaussée	CLC1 : proportion du tronçon dans le Corine Land Cover 1 (centre urbain)
Etat d'inspection R		
Duree : Durée de la campagne (h)	MAT_IS2 : Matériau (2 modalités)	CLC1 : proportion du tronçon dans le Corine Land Cover 1 (centre urbain)
InLcell : Longueur de la cellule (km, échelle log)	VOI3 : type de voirie (3 modalités)	Year_camp : Année de la campagne (3 modalités)
Secteur2 : Secteur (14 modalités)	Ndrf : Nombre de défaillances RF passées	

Les modalités qualitatives sont explicitées ci-dessous (modalité de référence en gras).

- MAT_IS2, 2 modalités : **Métal + Autre** et Plastique.
- VOI3, 3 modalités : **VCO**, NTL et Autre.
- Secteur2, 14 modalités (**5810 + 5120 + 5510 + 5821**).
- Year_camp, 3 modalités (**2018 + 2019**).
- MAT3, 6 modalités : **Acier + Autre**, AMCI, Fonte grise, Fonte ductile, PE et PVC.
- ID_ZF3, 9 modalités : **VE05 + CH01+ CH02 + CH04 + NA02 + NA03 + NA04 + VE02 + MO02**.

Les modalités de références géographiques correspondent au Secteur (5810) et à la ZF (VE05) ayant la moyenne maximum de tronçons en CLC4 + CLC5, regroupées avec leurs modalités respectives non-significatives.

Pour le sous-modèle d'état fuyard, il est intéressant de noter que les variables présentes dans les modèles LEYP (Tâche 03, page 42) agissent significativement dans le modèle ZIB et dans le même sens : le logarithme de la longueur et du nombre de raccordement +1, les défaillances passées et la pression augmentent les risques de défaillances et d'état fuyard en augmentant et le diamètre le diminue en

augmentant. L'effet des matériaux est aussi très significatif, avec tous les matériaux qui ont un risque moindre d'état fuyard que la référence Acier + Autre. Les variables environnementales agissent elles aussi en augmentant le risque de fuite quand la proportion sous la route, dans un centre urbain, dans de l'argile ou dans du limon augmente.

Le sous-modèle d'inspection fait ressortir des variables intéressantes à analyser en termes d'organisation du travail. En effet, on observe par exemple une plus forte probabilité d'inspection des plastiques, peut-être due à la difficulté d'écoute sur ces matériaux qui demande de multiplier les inspections par rapport au métal. De même, une forte proportion du tronçon dans un centre urbain engendre une probabilité d'inspection plus faible, vraisemblablement corrélée à la difficulté d'accès de ces tronçons pour une inspection en pleine journée. Dans la même optique, on remarque que les tronçons ont une probabilité plus forte d'être inspectés dans les secteurs de référence, qui ont été choisis au départ comme étant les secteurs avec un maximum de tronçons avec un CLC4+CLC5 élevé, soit des secteurs très ruraux. Enfin, on observe la montée en puissance de la recherche de fuite grâce à la covariable d'année de campagne, où la probabilité des tronçons d'être inspectés augmente d'année en année.

Tableau 13. Résultat de l'estimation des paramètres du modèle ZIB 17. Les 23 premières estimations correspondent au sous-modèle d'inspection R et les 24 suivantes au sous-modèle d'état fuyard Y.

Label	Estimate	Std. Dev.	Z score	Pval	lower 0.95	upper 0.95
(Intercept)	-1.77	0.12	-14.75	0.00	-2.00	-1.53
MAT_IS2PLASTIQUE	0.75	0.14	5.54	0.00	0.49	1.02
VOI3AUTRE	-0.26	0.15	-1.75	0.04	-0.56	0.03
VOI3NTL	0.16	0.04	4.31	0.00	0.09	0.24
CLC1	-1.07	0.49	-2.19	0.01	-2.03	-0.11
Ndrf	0.10	0.01	7.47	0.00	0.07	0.12
lnLcell	-0.69	0.02	-28.37	0.00	-0.73	-0.64
Secteur25110	-0.93	0.09	-10.68	0.00	-1.10	-0.76
Secteur25210	-0.73	0.14	-5.31	0.00	-0.99	-0.46
Secteur25220	-1.05	0.14	-7.30	0.00	-1.33	-0.77
Secteur25230	-1.04	0.16	-6.39	0.00	-1.36	-0.72
Secteur25310	-0.62	0.09	-7.25	0.00	-0.79	-0.45
Secteur25330	-0.75	0.11	-6.62	0.00	-0.97	-0.53
Secteur25420	-0.99	0.12	-8.34	0.00	-1.23	-0.76
Secteur25440	-0.50	0.06	-8.99	0.00	-0.61	-0.39
Secteur25530	-0.52	0.08	-6.42	0.00	-0.68	-0.36
Secteur25610	-0.74	0.06	-13.22	0.00	-0.85	-0.63
Secteur25710	-0.83	0.09	-9.01	0.00	-1.01	-0.65
Secteur25720	-0.49	0.08	-5.84	0.00	-0.65	-0.32
Secteur25820	-1.30	0.21	-6.11	0.00	-1.71	-0.88
duree	0.01	0.00	13.19	0.00	0.01	0.01
Year_camp2020	0.15	0.04	4.20	0.00	0.08	0.22
Year_camp2021	0.20	0.04	5.49	0.00	0.13	0.27
(Intercept)	-5.84	0.18	-32.14	0.00	-6.19	-5.48
lnlong	0.71	0.03	28.53	0.00	0.66	0.76
Age_Inspect_S	1.05	0.16	6.68	0.00	0.74	1.36
MAT3AMIANTE CIMENT	-0.17	0.08	-2.02	0.02	-0.33	-0.00
MAT3FONTE DUCTILE	-1.01	0.11	-9.26	0.00	-1.22	-0.79
MAT3FONTE GRISE	-0.16	0.06	-2.41	0.01	-0.28	-0.03
MAT3PE	-0.83	0.18	-4.59	0.00	-1.18	-0.48
MAT3PVC	-1.94	0.17	-11.60	0.00	-2.26	-1.61
diam100	-0.44	0.05	-9.63	0.00	-0.53	-0.35
pression100	0.77	0.10	7.37	0.00	0.56	0.97
lnbr1	0.13	0.02	6.01	0.00	0.09	0.17
Proute	0.15	0.05	3.02	0.00	0.05	0.24
CLC1	1.24	0.63	1.98	0.02	0.01	2.46
Pargile	0.64	0.25	2.58	0.00	0.15	1.13

Suite page suivante

Label	Estimate	Std. Dev.	Z score	Pval	lower 0.95	upper 0.95
Plimon	0.14	0.08	1.71	0.04	-0.02	0.30
Ndef	0.92	0.05	17.63	0.00	0.82	1.02
ID_ZF3CH03	-0.31	0.14	-2.23	0.01	-0.59	-0.04
ID_ZF3MO01	-0.34	0.15	-2.29	0.01	-0.64	-0.05
ID_ZF3MO03	-0.43	0.10	-4.54	0.00	-0.62	-0.25
ID_ZF3NA01	-0.95	0.17	-5.62	0.00	-1.28	-0.62
ID_ZF3VE01	0.82	0.33	2.46	0.01	0.17	1.47
ID_ZF3VE03	-0.24	0.11	-2.19	0.01	-0.45	-0.03
ID_ZF3VE04	-0.41	0.10	-4.28	0.00	-0.60	-0.22
ID_ZF3VE06	-0.50	0.07	-7.18	0.00	-0.64	-0.36

7.4.2.2.2 Validation

La représentation de la validation du modèle ZIB 17 sur les données de 2022 est présentée Figure 44. Elle présente des valeurs très élevées (0.84 pour le tri sur le nombre de tronçons, et $f(0.05)_{min} = 0.41$), ce qui indique un très bon pouvoir prédictif de notre modèle 17 avec 40% de fuites détectées évitées en changeant les 5% des tronçons les plus à risque. Il est cependant important de noter que, dans l'incapacité de séparer inspection d'état fuyard dans nos données, la validation se fait uniquement sur la variable $Z = R \times Y$, et ne peut se faire sur la variable d'intérêt Y seule.

7.4.2.2.3 Probabilité d'état fuyard $P(Y)$ et d'inspection $P(R)$

Le modèle ZIB nous permet ainsi de calculer séparément la probabilité d'état fuyard et d'inspection de chaque tronçon. Il faut cependant noter que le modèle ZIB calcule ces probabilités pour chaque tronçon à chaque inspection. De ce fait, nous avons choisi de conserver la probabilité d'état fuyard $P(Y)$ maximum, puisqu'elle augmente avec l'âge des tronçons et leur nombre de défaillances, et la probabilité d'inspection $P(R)$ moyenne, puisqu'elle dépend par exemple de la durée de campagne ou de l'année de la campagne (Annexe T07_07).

Nous les avons ainsi comparées aux probabilités prédites par le modèle logistique classique (GLM logistique) qui considère une couverture d'inspection des cellules de 100% à chaque inspection (Figure 46).

Notre modèle produit ainsi des prédictions de probabilités beaucoup plus cohérentes avec la réalité : le GLM classait 99.99% des tronçons avec une probabilité quasi nulle de fuite et 0.01% avec une probabilité quasi certaine, quand notre modèle produit une courbe plus infléchie avec une pente à l'origine beaucoup plus faible, ce qui correspond à un continuum de probabilité d'état fuyard et d'inspection.

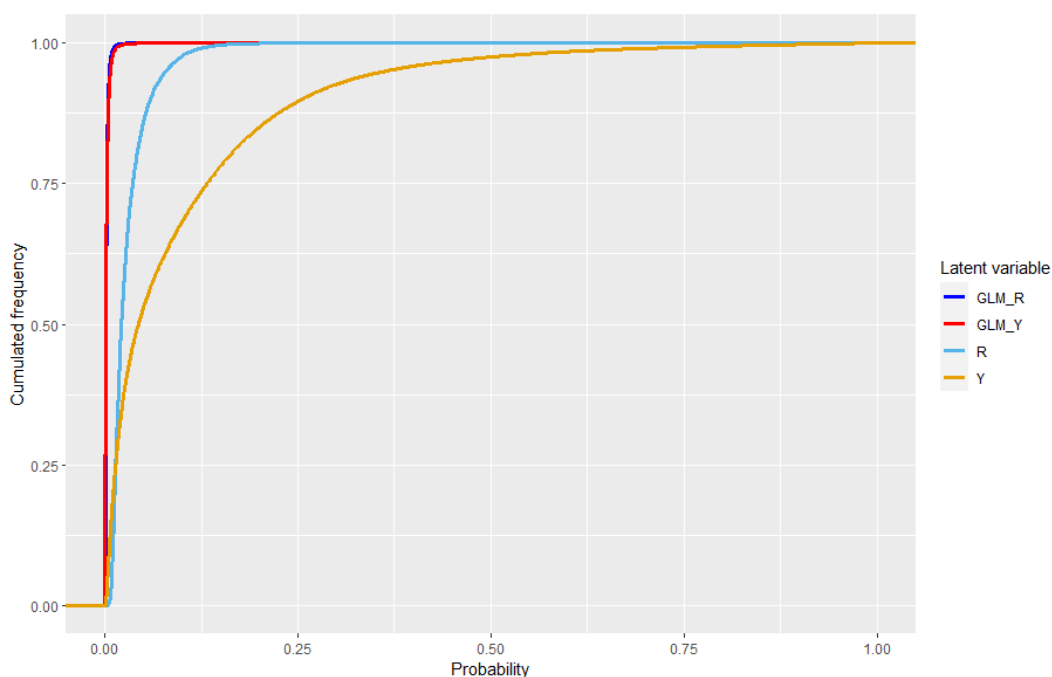


Figure 46. Fréquence cumulée des probabilités calculées par le modèle 17. Y (courbe orange) représente la probabilité d'état fuyard maximum des tronçons. R (courbe bleue claire) représente la probabilité d'inspection moyenne des tronçons. Les courbes GLM_Y (rouge) et GLM_R (bleu foncé) sont les homologues des précédentes, avec des probabilités (respectivement maximum et moyennes) calculées avec un modèle de régression logistique standard.

7.5 Perspectives

Plusieurs champs d'étude restent ouverts à la fin de cette Tâche 07- Modélisation de la probabilité d'un tronçon d'être fuyard.

Premièrement, une utilisation des résultats de la Tâche 11 (page 156) pourrait potentiellement permettre de valider séparément les probabilités d'inspection et d'état fuyard quand aujourd'hui, seule la validation sur Z est possible.

Deuxièmement, une visualisation de $P(Y)$ sous forme de carte comme celle produite lors de la Tâche 03 est envisageable afin d'exploiter ces résultats au-delà de l'inclusion de cette probabilité dans l'outil multi-critère en Tâche 10 (indicateur IP_{pertes} , Tableau 28).

Enfin, la combinaison des phénomènes des casses et des fuites reste ouverte et fait l'objet d'une tâche spécifique, la Tâche 09 (page 95).

Archive Tâche 07 :

- Dossier Données : Données et code nécessaires pour produire le fichier d'entrée du modèle ZIB
- Dossier Stage_Malick : codes et résultats du test de la robustesse du modèle ZIB (stage Malick N'Diaye M2)
- Dossier ZIB : Code du modèle ZIB et résultats du meilleur modèle choisi

8 Tâche 08 - Analyse de la vulnérabilité des clients

8.1 Introduction

La Wallonie jouit d'un climat tempéré, avec une forte variabilité des cumuls annuels de précipitations d'une année à l'autre, la moyenne annuelle s'établissant à environ 843 mm sur la période 1981-2019 (données de l'Institut Royal Météorologique de Belgique). Néanmoins, ces dernières années, un déficit répété et accentué des précipitations entraîne une multiplication des épisodes de sécheresse (Van Loon et al. 2016). Les années 2017, 2018 et 2019 notamment ont été caractérisées par des périodes de canicules longues (Thibaut et Ozer 2021) et cette tendance se poursuit depuis. Sur les huit premiers mois de l'année 2022, le déficit moyen de précipitations s'élevait à 40 %, soit un manque de 220 mm par rapport aux moyennes historiques (CRA-W 2022), les mois de juillet et août enregistrant même un déficit des précipitations de 86 %. Ces aléas climatiques ont de lourdes conséquences sur les systèmes naturels et agricoles (Schwalm et al. 2017), notamment sur les ressources en eau avec des étiages parfois sévères et une baisse des niveaux de certains barrages et des nappes souterraines, une dégradation de la qualité des cours d'eau, une augmentation de la concentration en nitrate des eaux souterraines, etc.

Les zones affectées par les sécheresses sont aussi plus susceptibles d'être confrontées à des pénuries d'eau et de nourriture, à des incendies de forêt ou encore à l'apparition de maladies d'origine hydrique et alimentaire (Sinisi et Aertgeerts 2011). En Wallonie, des pénuries d'eau potable ont ainsi déjà été observées localement (Bastin et al. 2022) et certaines communes mettent régulièrement en place des arrêtés de police visant à limiter l'usage de l'eau du robinet (par exemple interdiction de lavage des véhicules, de remplissage des piscines et citernes, de l'arrosage des jardins d'agrément). Les challenges pour les services d'eau potable sont dès lors multiples, qu'il s'agisse de répondre à une demande croissante en raison de la hausse des températures et des changements dans l'approvisionnement, de faire face à une plus grande variabilité du débit des cours d'eau en raison des changements de température et de précipitations, avec des dommages possibles aux infrastructures d'approvisionnement en eau en cas de fortes pluies ou de sécheresse, d'anticiper la probabilité que l'approvisionnement en eau et le traitement des eaux deviennent de plus en plus énergivores et coûteux, etc. (Sinisi et Aertgeerts 2011). Il s'agit ainsi de s'adapter à des conditions changeantes, alors que les réseaux ont été conçus pour des conditions stables. Or la transposition au niveau belge des scénarios du GIEC (Groupe d'experts intergouvernemental sur l'évolution du climat) fait craindre une augmentation de la fréquence de ces épisodes de sécheresse en été (IPCC 2023).

Dans ce contexte, le premier objectif de cette tâche est d'apprécier les préférences des consommateurs d'eau potable en matière de gestion patrimoniale des réseaux et d'estimer leurs consentements à payer (CAP) pour éviter certains impacts des sécheresses sur la performance de distribution d'eau du robinet (e.g. des coupures ponctuelles de l'eau de robinet ou une pression faible de l'eau)¹³. À notre connaissance,

¹³ Deux scénarios distincts ont été construits. Le premier décline les conséquences des sécheresses en termes de restrictions des usages et le second en termes de performance de service. Seuls les résultats concernant la performance sont présentés dans ce rapport. Les statistiques descriptives sur l'ensemble des enquêtés sont toutefois proposées en Annexe T08_01 (**Erreur ! Source du renvoi introuvable.**).

aucune étude n'a jusqu'à présent analysé les préférences des Wallons quant à la performance du service d'eau dans un contexte de changement climatique. Nous chercherons ainsi à mettre en évidence les bénéfices des mesures d'adaptation au risque des sécheresses et à ses conséquences sur le système de distribution d'eau du robinet, éléments qui pourraient permettre d'améliorer la gestion proactive des conséquences des sécheresses par la demande (Griffin et Mjelde 2000; Rossi, Nicolosi, et Cancelliere 2008; Sinisi et Aertgeerts 2011). Parmi les méthodes d'évaluation non marchande, la plus connue est l'évaluation contingente, qui s'appuie sur une enquête dans laquelle le répondant renseigne directement son CAP (Mitchell et Carson 1989). Au vu du caractère multidimensionnel de la performance des services d'alimentation en eau potable, nous avons choisi d'avoir recours à la méthode des choix multi-attributs (MCMA) ou « discrete choice experiment », qui en est une variation : le bien ou service est décomposé en plusieurs caractéristiques, l'enquête révélant à la fois son CAP pour ces dernières mais aussi les arbitrages réalisés entre elles.

Il s'agit aussi de comparer la vulnérabilité « objective » telle que définie par la SWDE et la vulnérabilité « subjective » telle que ressentie par les ménages. Nous faisons en effet l'hypothèse que la seconde ne correspond pas totalement à la première, et qu'elle a une influence sur les préférences des ménages en matière de performance de service rendu. En effet, la vulnérabilité subjective par rapport à un aléa climatique donné renvoie à la fois à la perception du risque chez l'individu (Slovic 1987) (elle-même influencée par différents facteurs comme l'expérience du risque ou le degré d'information quant à sa probabilité d'occurrence ou ses conséquences (Hellequin et al. 2013; Wang, Ge, et Gao 2018)), et à son niveau de concernement (Luís, Vauclair, et Lima 2018). Elle peut dès lors se traduire par des attitudes face au risque différentes selon les individus (Tapsuwan, Cook, et Moglia 2018), de sorte qu'étudier la sensibilisation ou la perception au risque de sécheresse et à son impact sur l'alimentation en eau potable et les attitudes déjà adoptées nous semble essentiel afin de mieux cerner les variables explicatives des CAP. Wilson et al. (2021) ont ainsi par exemple étudié les facteurs psychologiques et attitudinaux impactant le CAP des résidents pour éviter des restrictions d'usage de l'eau à Brisbane en Australie. Ces facteurs incluent des variables comportementales telles que le comportement des parents (et leur influence sur les décisions des enfants), la sensibilisation au système actuel de tarification ou encore les croyances religieuses. Des caractéristiques sociodémographiques semblent également influencer le CAP, telles que le nombre de personnes dans un ménage ou encore le niveau d'éducation du répondant. Cooper et al. (2019) et Price et al. (2019) ont également montré que les CAP diffèrent selon les régions et les contextes climatiques. Nos résultats permettront notamment de guider la conception et la mise en œuvre d'un plan efficace de lutte contre les effets de la sécheresse, garantissant une meilleure acceptabilité et donc une meilleure efficacité des mesures de prévention (Shao, Kam, et Cass 2023).

Cette présentation est organisée comme suit. Dans la section suivante, nous décrivons d'abord la méthode empirique utilisée à savoir la MCMA, suivie d'une revue de littérature dans la section 3. Dans la section 4, nous présentons le questionnaire et le protocole de collecte des données. Les résultats sont exposés dans la section 5 puis discutés dans la section 6.

8.2 La méthode des choix multi-attributs

La MCMA est une méthode d'évaluation des préférences déclarées, généralement utilisée dans l'évaluation des biens environnementaux non-marchands (Hanley et al., 2001). Son cadre conceptuel prend racine dans la théorie du consommateur de Lancaster (1966), qui suppose que les décisions du

consommateur sont déterminées par l'utilité procurée par les caractéristiques (appelées « attributs ») du bien plutôt que le bien lui-même. La MCMA est une variante de l'analyse conjointe, initialement développée par Louviere & Hensher (1982) et Louviere & Woodworth (1983) dans le marketing et les transports. Elle s'avère plus appropriée que d'autres méthodes telles que la méthode d'évaluation contingente dans le cas spécifique de l'évaluation des valeurs relatives à la performance du service d'approvisionnement en eau du fait de son caractère multidimensionnel (Blamey et al., 1999).

La MCMA repose sur un deuxième bloc théorique, à savoir la théorie de l'utilité aléatoire (Manski, 1977; McFadden, 1973), pour définir la fonction de maximisation de l'utilité. Considérons la fonction d'utilité suivante, où un individu i est confronté à un choix parmi J alternatives dans une série de choix t :

$$U_{ijt} = \beta_i x_{ijt} + \epsilon_{ijt} \quad \text{Équation 8}$$

Cette fonction est composée d'une part par la portion observable et déterministe de l'utilité, qui dépend des attributs x_{ijt} associés à chaque alternative et le vecteur β des paramètres estimés, et d'autre part, par ϵ_{ijt} la partie non observée de l'utilité. Le modèle d'utilité aléatoire se base sur l'hypothèse que l'individu cherche à maximiser son utilité. Ainsi, la probabilité que i choisisse l'alternative j plutôt que l'alternative g dans une série de choix t est la probabilité que U_{ij} soit plus importante que U_{ig} :

$$\pi_{ij} = \Pr(U_{ij} > U_{ig}) = \Pr(\beta' x_{ij} + \epsilon_{ij} > \beta' x_{ig} + \epsilon_{ig}) \quad \forall j \neq g \quad \text{Équation 9}$$

Une approche classique de spécification du modèle est le modèle logit conditionnel (McFadden, 1973), qui se base sur l'hypothèse d'une distribution homogène des préférences des individus. Dans ce cas :

$$\pi_{ij} = \frac{e^{\beta' x_{ij}}}{\sum_{g=1}^J e^{\beta' x_{ig}}} \quad \text{Équation 10}$$

D'autres spécifications permettent de traiter l'hétérogénéité des préférences des répondants, comme par exemple le logit à classes latentes. Celui-ci suppose l'existence de C classes dans l'échantillon des répondants. Le modèle est caractérisé par des préférences homogènes au sein des classes, et hétérogènes entre les classes. Il se base sur deux spécifications probabilistes distinctes. D'une part, on distingue la probabilité que le répondant n appartienne à la classe c , généralement spécifiée selon un ensemble de covariables socio-économiques (Boxall & Adamowicz, 2002; Provencher et al., 2002), et définie comme suit :

$$\pi_{ic} = \frac{e^{\lambda_c' z_i}}{\sum_{s=1}^C e^{\lambda_s' z_i}} \quad \text{Équation 11}$$

où z_n est l'ensemble des variables socio-économiques et λ_c est le vecteur des coefficients estimés, relatif à la classe c . D'autre part, on estime la probabilité que l'individu i , choisisse l'alternative j , sachant qu'il appartient à la classe de préférence c :

$$\pi_{ij|c} = \frac{e^{\beta_c' x_{ij}}}{\sum_{g=1}^J e^{\beta_c' x_{ig}}} \quad \text{Équation 12}$$

où β'_c représente le vecteur des paramètres estimés associés à c . Finalement, la probabilité inconditionnelle qu'un individu i choisisse l'alternative j est dérivée de l'Équation 11 et de l'Équation 12 :

$$\pi_{ij} = \sum_{s=1}^c \pi_{is} \pi_{ij|s} \quad \text{Équation 13}$$

Le CAP moyen de chaque niveau d'attribut est calculé sur la base des paramètres d'utilité estimés (Hanemann, 1984; Train, 2003), comme suit :

$$CAP = - \frac{b_{attribut}}{b_{coût}} \quad \text{Équation 14}$$

où $b_{attribut}$ est le coefficient estimé pour un niveau d'attribut donné et $b_{coût}$ le coefficient estimé pour l'attribut monétaire COUT.

8.3 Revue de littérature sur l'évaluation de la performance des services d'eau potable

Plusieurs études de MCMA ont exploré les préférences des ménages en matière de performance de l'approvisionnement en eau de robinet dans un contexte de pression sur la ressource. Une première catégorie de travaux intègre ainsi la continuité du service dans les attributs afin de considérer plus globalement la performance du service rendu. On citera par exemple Snowball et al. (2008) dont les attributs sont la qualité bactériologique de l'eau, la coloration de l'eau, la pression, la fréquence des interruptions de service et la fiabilité des compteurs, ou Latinopoulos (2014) qui mobilise la qualité de l'eau, la fréquence des interruptions de l'approvisionnement en eau et le niveau de création d'emplois agricoles. Ses résultats montrent que les répondants sont disposés à payer des montants élevés pour les niveaux élevés des deux premiers attributs mais pas pour le dernier, indiquant ainsi une priorité pour les usages domestiques de l'eau. Wang et al. (2018) montrent quant à eux que les ménages chinois sont prêts à payer leur eau plus cher en échange d'une amélioration de la pression de l'eau, de sa qualité, du délai d'entretien et d'information aux abonnés, et enfin d'une baisse de la fréquence des interruptions de l'approvisionnement. Enfin, Syuhada et al. (2020) estiment le CAP des ménages de Terebgganu, en Malaisie pour une amélioration de la qualité d'eau, de la pression d'eau fournie et une diminution des coupures d'eau, mettant en évidence des préférences fortes pour ces trois caractéristiques de la performance.

Un second groupe de travaux se focalise directement sur la continuité du service. Les interruptions de service sont ainsi décrites chez MacDonald et al. (2005) en termes de fréquence, durée, moyen de communication et alternative fournie. Cooper et al. (2019), également en Australie, et Price et al. (2019) au Canada adoptent la même perspective et font explicitement référence aux enjeux de sécheresse dans leur scénario. Hensher et al. (2006) s'intéressent quant à eux aux restrictions d'usage. Leur MCMA mobilise en effet comme attributs (i) la fréquence de restrictions des usages, (ii) leur durée, (iii) leur sévérité (i.e. les activités autorisées), (iv) le type de jours concernés (tous les jours vs des jours différents), et (v) l'apparence des espaces verts urbains selon leur niveau d'arrosage. Bennett et al. (2016) de leur côté évaluent les CAP pour de l'eau recyclée ou Tapsuwan et al. (2018) pour le recours à des réservoirs.

À notre connaissance, seuls Willis et al. (2005) s'intéressent donc aux conséquences des sécheresses sur la performance du service d'alimentation en eau potable. Ces auteurs mobilisent plusieurs attributs dont (i) la sécurité de l'approvisionnement, définie en partie comme le risque qu'un ménage soit sans eau courante pour une longue période (e.g. trois mois), (ii) le nombre de ménages affectés par une interruption temporaire de l'approvisionnement et la durée de celle-ci, et enfin (iii) le nombre de ménages affectés par une pression insuffisante de l'eau. Ainsi, à l'image du travail de Snowball et al. (2008), les attributs étudiés traitent de l'impact sur la communauté plutôt que sur le ménage du répondant, partant du postulat qu'il ne leur était pas possible d'identifier *a priori* les répondants ayant déjà été confrontés à des problèmes d'approvisionnement des autres. Il est également intéressant de noter que ce travail se place plutôt dans une démarche « statique » puisqu'il étudie les changements actuellement en cours et non prospectifs (Cooper et al., 2019), ne tenant ainsi pas compte de l'aggravation dans le temps des effets des sécheresses sur l'alimentation en eau potable.

8.4 Protocole d'enquête

8.4.1 Construction des séries de choix

La première étape d'une MCMA consiste à choisir les attributs permettant de décrire le bien ou service étudié (Bateman et al., 2002; Bennett & Blamey, 2001; Hanley et al., 2001; Hoyos, 2010), à savoir ici la performance du service d'eau potable dans un contexte de sécheresse. Dans notre cas, six attributs ont été sélectionnés, sur la base de la revue de la littérature et en accord avec les experts de la SWDE. Ils traduisent les mesures de prévention des conséquences de sécheresses, à différents niveaux d'ampleur, qui viseraient à :

- Réduire la durée des conséquences des épisodes de sécheresse qui pourrait augmenter de 10, de 15, de 20 voire de 25 jours par an par rapport à aujourd'hui ;
- Maintenir la pression de l'eau du robinet qui pourrait être moyenne, avec des conséquences en termes de confort (filet d'eau sous la douche...) et d'efficacité (allongement du temps d'arrosage...), voire basse (pas d'eau à l'étage, impossibilité d'utiliser les machines à laver le linge ou la vaisselle...) ;
- Moduler les acteurs visés par le partage de l'effort de réduction des consommations d'eau du robinet. Il pourrait s'agir soit des seuls ménages, soit des ménages et agriculteurs, soit des ménages et industriels, soit encore de tous ces acteurs ;
- Maintenir un accès à l'eau du robinet pour les ménages. Outre les restrictions, il est possible que les quantités d'eau à disposition soient rationnées, voire que les ménages aient à faire face à des coupures d'eau ponctuelles (par ex. un jour par semaine sur les mois de juillet et août) ;
- Réduire la probabilité d'être touché(e) par ces conséquences des sécheresses, qui pourra être de 10 %, de 20 %, de 30 % ou de 40 % (soit en moyenne 1, 2, 3 ou 4 années sur 10). Parmi ces six attributs, on distingue un attribut monétaire (ou Coût du scénario), défini par une augmentation de la facture d'eau annuelle du client (entre 12 € et 72 € par an).

Ces attributs sont ensuite combinés en alternatives combinant les attributs à des niveaux différents, pour illustrer des scénarios variés de performance du service d'approvisionnement en eau du robinet dans leur commune en période de sécheresses à l'horizon 2030. Les alternatives sont ensuite combinées en séries

de choix, le tout grâce au logiciel NGENE¹⁴. Etant donné le nombre d'attributs et de leurs niveaux, douze séries de choix ont été créées, qui ont ensuite été réparties par NGENE en deux groupes de six. Afin de constituer les séries de choix, une situation de référence leur a été ajoutée (un exemple est fourni sur la Figure 47). Cette option C reste inchangée d'une série à l'autre et correspond à un maintien de la gestion actuelle des sécheresses, i.e. sans mesure supplémentaire que celles mises en œuvre actuellement. Elle se traduirait donc par une dégradation de la performance du service fourni sur l'ensemble des critères. Elle est proposée à un coût nul (0 €) puisque le prix de l'eau fourni par la SWDE est resté inchangé ces dernières années ; nous conservons donc cette tendance.

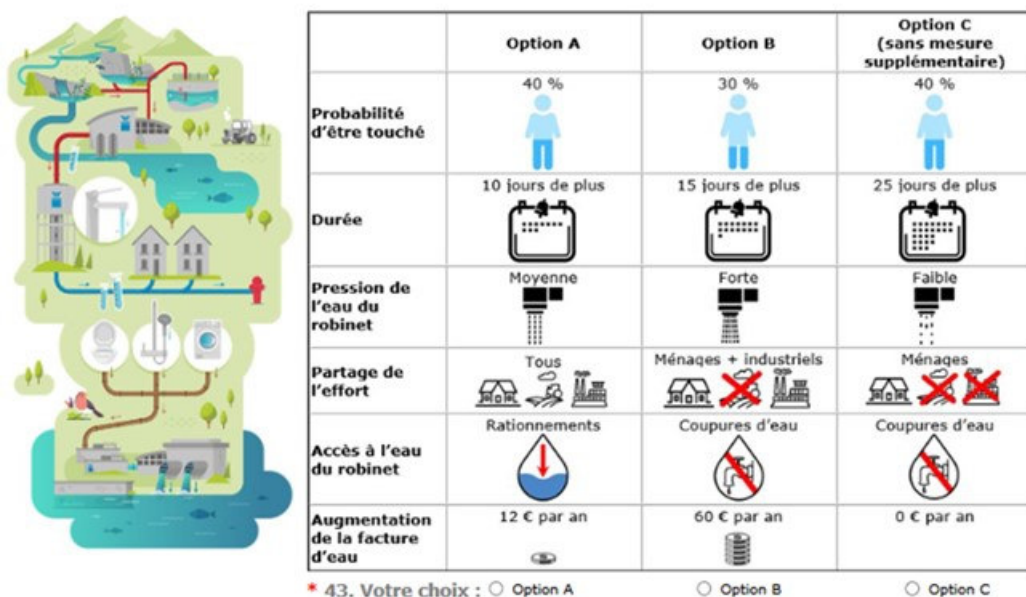


Figure 47 : Exemple de série de choix

Rappelons que les caractéristiques du service non mentionnés dans ces séries de choix (par exemple la qualité sanitaire de l'eau distribuée) étaient supposées rester identiques à la situation actuelle. Dans chacune des six cartes qui lui ont été présentées, le répondant devait choisir l'option qu'il préfère. Il est supposé que les mesures proposées s'imposeraient à l'ensemble des habitants de la commune afin d'éviter les comportements stratégiques (Carson & Groves, 2011).

Le texte accompagnant les séries de choix (Mariel et al., 2021) débutait par une présentation générale des conséquences des sécheresses puis des mesures mises en œuvre par la SWDE pour éviter ses conséquences sur l'approvisionnement en eau potable (remplissage des châteaux d'eau par des camions citernes ; pose de conduites d'interconnexion permettant la jonction entre les zones disposant d'importantes ressources et celles aux ressources plus limitées ; recherche de nouvelles sources d'eau

¹⁴ <https://www.choice-metrics.com/>

potable (nouveaux captages, dessalement d'eau de mer...); développement de projets de ressources alternatives (réutilisation des eaux de pluie ou des eaux usées traitées...); etc.). Ces mesures, mises en œuvre et coordonnées par la SWDE, la Région wallonne et les communes, étaient supposées s'accompagner d'une communication accrue auprès de l'ensemble des acteurs sur la nécessité de faire attention à sa consommation d'eau potable. Étaient ensuite décrites les différentes caractéristiques utilisées dans les séries de choix (Johnston et al., 2017). Enfin, la contrainte budgétaire est rappelée, ainsi que cela est préconisé dans la littérature (Mitchell & Carson, 1989), de même que le montant moyen annuel de la part « eau potable » d'un ménage wallon de trois personnes afin que le répondant dispose d'une référence (Martin-Ortega et al., 2011).

8.4.2 Structure du questionnaire

Le questionnaire était construit autour de six blocs de questions traitant :

1. Des aspects socio-économiques (commune de résidence, genre, âge, type de logement, catégories socio-professionnelles, revenus, etc.) ;
2. De la ressource en eau en général (enjeux autour de la gestion de l'eau, menaces sur la ressource selon le répondant, etc.) ;
3. Des usages en eau, avec notamment des questions sur les pratiques d'économies d'eau ;
4. Du service d'eau du robinet (origine de l'eau distribuée, organisme en charge de cette distribution, critères importants en termes de performance du service, appréciation de cette performance, etc.) ;
5. Du risque de sécheresse, afin d'approcher la perception de ce risque par les ménages mais aussi leur sensibilisation ;
6. Des séries de choix et questions de suivi associées qui ont pour objectif de s'assurer de la pertinence et de la crédibilité des scénarios proposés.

8.4.3 Protocole d'enquête

La population étudiée est composée des ménages habitant en région Wallonne et clients de la SWDE. Afin d'apprécier la vulnérabilité « objective » des répondants, et de pouvoir la comparer à leur vulnérabilité « subjective », trois groupes de communes wallonnes ont été définis (Tableau 14) :

1. Le premier regroupe des communes dans lesquelles les épisodes de sécheresse se sont traduits au moins une fois par des mesures de restriction des usages depuis janvier 2017 ;
2. Le second regroupe des communes dans lesquelles la sécheresse a des conséquences sur le fonctionnement de la SWDE (il est par exemple nécessaire d'approvisionner par camions des châteaux d'eau afin de garantir la continuité du service) mais sans conséquence importante pour les clients ;
3. Dans le troisième groupe, la sécheresse n'a aucune conséquence ou est inexistante (présence de barrages, de captages importants ou de liaisons par exemple). Ces communes ont été choisies à proximité de celle du groupe 1 afin que les contextes pluviométriques soient néanmoins relativement « proches ».

Une contrainte sur la sélection des communes a été imposée, à savoir que la totalité du territoire de chacune soit située dans la même Zone de Distribution en Eau (ZDE) afin que la qualité de l'eau puisse

être considérée comme uniforme et surtout provienne de la même source. Notons que l'ensemble des habitants de certaines communes n'est pas desservi par la SWDE, sans que ce soit problématique pour notre analyse étant donné le mode de passation adopté.

Tableau 14 : Liste des communes d'enquête par groupe

Groupe 1	Groupe 2	Groupe 3
Beauraing*	Andenne	Ath
Daverdisse	Assesse	Binche
Durbuy*	Aywaille	Braives
Gesves	Beloeil*	Crisnée
Houyet	Bertrix	Eupen
Jalhay	Braine-L'Alleud*	Grez-Doiceau
Lierneux	Braine-le-Comte	Herve
Ohey	Ciney*	La Louvière
Rochefort*	Dinant	Lessines
Sprimont	Ferrières	Leuze-en-Hainaut
Vielsalm	Gerpennes	Marche-en-Famenne
Waimes*	Hannut	Namur
	Mons	Plombières
	Paliseul	Raeren
	Rebecq	Remicourt
	Saint-Vith*	Soumagne
	Tournai	Thimister-Clermont
	Waremme	
	Wellin	
	Yvoir	

Note : * communes non entièrement desservies par la SWDE

Environ la moitié des adresses mail des clients, équitablement répartis sur son territoire, sont connues par les services de la SWDE. La passation du questionnaire s'est donc effectuée par internet, de manière auto-administrée. L'ensemble du protocole de collecte est conforme au Règlement Général sur la Protection des Données (RGPD). Le questionnaire a été testé auprès de 1 500 ménages (500 par groupe de communes) en Juin 2022 afin de s'assurer de sa bonne compréhension. Après des ajustements à la marge, le questionnaire final a été envoyé à 22 500 ménages sélectionnés de manière aléatoire (7 500 pour chaque groupe de commune). Ces derniers ont reçu de la part de la SWDE un message les invitant à répondre à l'enquête (Figure 48) le 20 Juillet 2022. Deux relances ont ensuite été effectuées, une le 1^{er} août et la 2^{ème} le 16 août. Le délai de réponse était fixé au 31 août. Les deux versions du questionnaire ont elles aussi été réparties aléatoirement entre les enquêtés selon leur rang (date-heure) de connexion. Le taux de réponse se situe autour de 15 % si l'on considère le nombre des mails ouverts. Étant donné que certaines personnes n'ont pas complété l'ensemble du questionnaire, 1 030 réponses (soit 6 180 observations) sont exploitables pour l'analyse économétrique.



Figure 48 : Mail envoyé par la SWDE à un échantillon aléatoire de clients habitant dans les communes d'enquête

8.5 Résultats

8.5.1 Analyse statistique descriptive

La Figure 49 présente la distribution géographique de l'échantillon par commune de la région Wallonne. La plus grande proportion (81 répondants soit 9 %) se trouve à Durbuy. La distribution par groupe de

commune est presque homogène : 33 % des répondants font partie du groupe 1, 38 % du groupe 2 et 29 % du groupe 3.

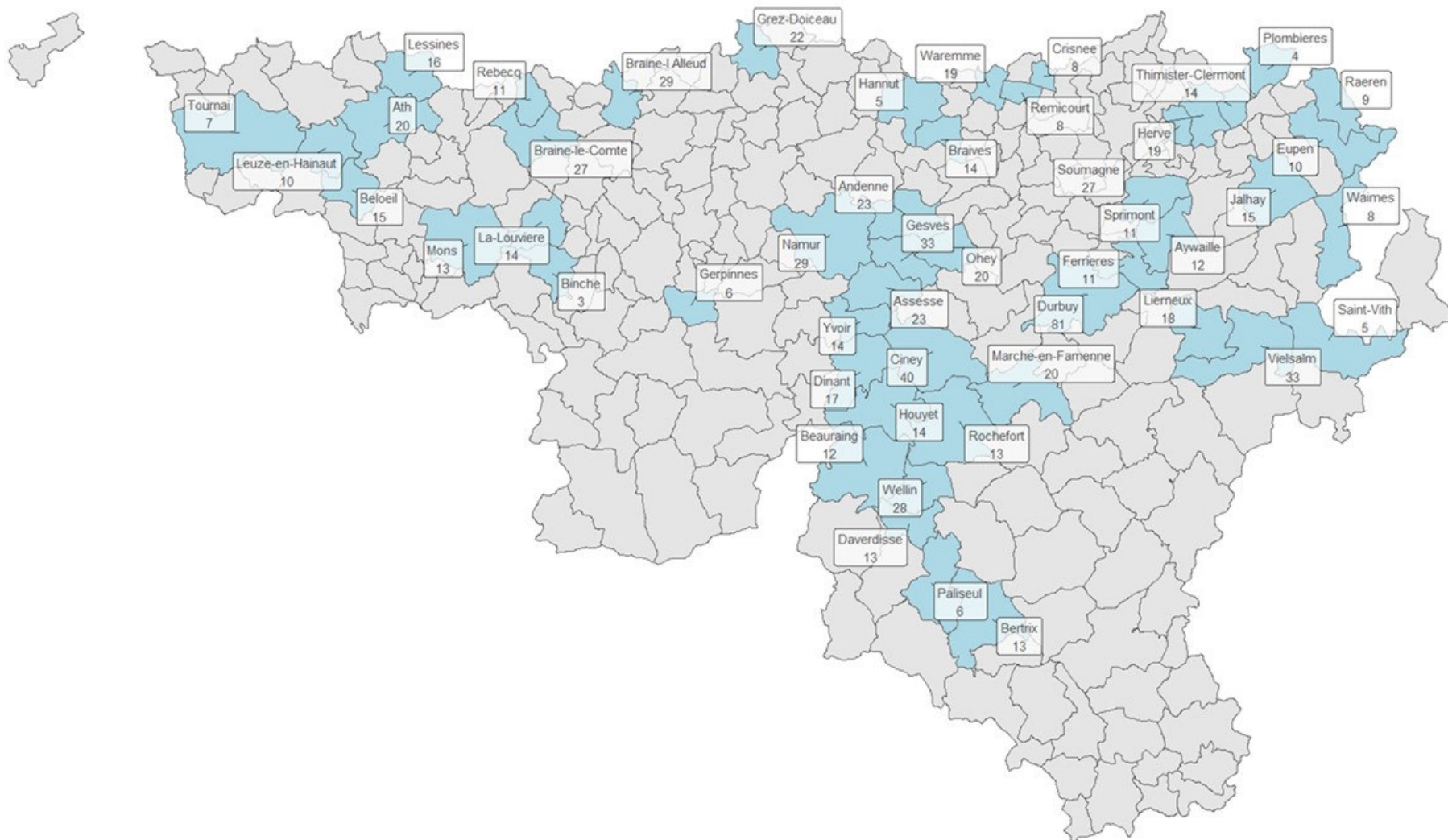


Figure 49 : Distribution des enquêtes par commune

On s'intéressera tout d'abord à l'analyse statistique des caractéristiques socio-géo-économiques de notre échantillon (Tableau 15). Le répondant-type est un homme âgé de 60 ans vivant en couple et habitant dans sa commune depuis environ 28 ans. Il exerce une activité professionnelle et son ménage a un revenu mensuel net compris entre 2 000 € et 3 999 € (ceci comprend les salaires, indemnités, retraites, rentes, allocations, pensions, revenus fonciers, etc.). On notera que notre échantillon est plus âgé que la population Wallonne et que les femmes y sont sous-représentées.

Tableau 15 : Analyse descriptive des caractéristiques socio-géo-économiques des répondants

Variable	Moyenne (écart-type) OU effectif (fréquence)	
	Echantillon	Région Wallonne
Taille de l'échantillon	1 030	3,644 millions*
Age	59,6 (13,2)	
Ancienneté dans la commune	27,6 (19,3)	
Taille du ménage	2,5 (1,4)	
Nombre d'enfants dans le ménage	0,5 (1,0)	
Femme	310 (30 %)	1,871 million* (51%)
Membre d'une association	222 (22 %)	
<i>Classe d'âge</i>		
20 – 39	82 (8 %)	0,911 million* (32 %)
40 – 59	402 (39 %)	0,974 million* (34 %)
60 – 79	489 (48 %)	0,759 million* (27 %)
> 80	43 (4 %)	0,185 million* (7 %)
<i>Niveau d'éducation</i>		
Aucun	13 (1 %)	
Bachelier	279 (27 %)	
CESS	297 (29 %)	
CEB	25 (2 %)	

CE1D	74 (7 %)	
Master et plus	297 (29 %)	
<i>Classe de revenus</i>		
Moins de 1 000 €	24 (2 %)	
De 1 000 € à 1 999 €	173 (17 %)	
De 2 000 € à 3 999 €	392 (38 %)	
De 4 000 € à 5 999 €	193 (19 %)	
De 6 000 € à 8 999 €	48 (5 %)	
De 9 000 € à 14 999 €	8 (0,8 %)	
15 000 € et plus	4 (0,4 %)	
<i>Statut professionnel</i>		
Activité professionnelle	528 (51 %)	
Au chômage	13 (1 %)	
Retraité ou en pré-retraite	412 (40 %)	
Autre (au foyer, étudiant, ...)	32 (3 %)	

* Source : Statbel (<https://statbel.fgov.be/fr/themes/population/structure-de-la-population>)

Notre échantillon est à 83 % propriétaire de son logement, ou accédant, et dans 93 % des cas, habite dans une maison. De plus, 85 % possèdent un jardin d'agrément, 43 % un potager et 47 % détiennent une citerne pour la récupération de l'eau de pluie.

8.5.2 Analyse bivariée par rapport au groupe de communes

Afin de comparer la vulnérabilité « objective » telle que définie par la SWDE et la vulnérabilité « subjective » telle que ressentie par les ménages Wallons, une analyse bivariée a été réalisée grâce au test d'indépendance du khi-deux (pour les variables qualitatives) et au test de Kruskal-Wallis¹⁵ (pour les variables quantitatives). Elle a pour objectif d'étudier l'existence de relations statistiques entre la variable « Groupe de commune », qui définit la répartition des communes selon leur vulnérabilité au risque de sécheresse, et les variables décrivant la vulnérabilité des clients (e.g. perception du risque de sécheresse).

¹⁵ Le test de Kruskal-Wallis est une alternative non paramétrique au test ANOVA (Analyse de Variance) à un facteur en principe utilisé pour étudier les différences de moyennes lorsqu'il y a plus de deux groupes à comparer. Le test de Kruskal-Wallis est en effet recommandé lorsque les hypothèses du test ANOVA à un facteur ne sont pas respectées (dans notre cas, l'hypothèse de la normalité n'est pas satisfaite).

Le Tableau 16 regroupe les variables pour lesquelles nous observons une corrélation statistiquement significative (i.e. la p-value est inférieure *a minima* à 10 %). Les colonnes reprennent les fréquences de chaque modalité (pour les variables qualitatives) ou moyennes (pour les variables quantitatives) pour chacun des groupes de communes ainsi que pour l'ensemble de l'échantillon. Le dictionnaire des variables avec une présentation détaillée de leur signification est proposé en Annexe TO8_01. Pour l'ensemble des autres variables non présentées dans ce tableau, aucune corrélation n'a donc été mise en évidence.

Tableau 16 : Résultats de l'analyse bivariée par groupe de communes

Variable → modalité	Groupe 1	Groupe 2	Groupe 3	Ensemble
<i>Caractéristiques socio-géo-économiques</i>				
ANC_COMMUNE***	25,3 ans	25,6 ans	31 ans	27,6 ans
<i>Inondations 2021</i>				
INOND_AMIS*** → oui	45 %	28 %	37%	37 %
<i>La gestion de l'eau</i>				
H_MENACE_RESS*** → Menacées	41 %	26 %	23 %	30 %
H_QUANTO_AUJ*** → Plutôt pas d'accord	28 %	15 %	15 %	19 %
H_QUANTO_5ANS*** → Inférieure aux besoins	48 %	33 %	40 %	41 %
<i>Vos usages de l'eau</i>				
CONSO_ADLT*** → De l'eau de robinet mais filtrée	17 %	16 %	10 %	14 %
<i>Votre service d'eau du robinet</i>				
H_QUALO_AUJ*** → Plutôt pas d'accord	17 %	14 %	10 %	13 %
PROVENANCE_AEP*** → Des nappes souterraines	47 %	47 %	36 %	43 %
H_QUAL_SERVICE*** → De mauvaise qualité	20 %	10 %	9 %	13 %
<i>Le risque de sécheresse</i>				
H_SECH_NBFUTUR_COMM** → Rester stables	11 %	17 %	19 %	16 %
H_SECH_RISK*** → Oui, plutôt	40 %	25 %	28 %	31 %
REST*** → Aucune	13 %	67 %	69 %	60 %
H_REST_GENE*** → Gênantes	10 %	5 %	5 %	6 %
SECH_CONF*** → La SWDE	42 %	52 %	57 %	51 %
<i>Scénarios</i>				
H_PROBA_SECH*** → Très probable	35 %	25 %	23 %	27 %

Note : * indique une significativité à 10 % (p-value inférieure à 0,10), ** à 5 % (p-value inférieure à 0,05) et *** à 1 % (p-value inférieure à 0,01). Afin d'en faciliter la lecture, n'apparaît dans ce tableau pour chaque variable que la modalité pour laquelle il existe la différence la plus significative entre les groupes de communes.

À partir de ces résultats, nous avons analysé la différence entre les groupes de communes par rapport aux variables ci-dessous. Nous nous intéressons plus particulièrement aux variables reflétant la vulnérabilité ressentie par les répondants au risque de sécheresse. Pour ce faire, nous nous rapportons au tableau des résidus, qui permet de voir si les effectifs observés pour chaque modalité sont supérieurs ou inférieurs à ceux attendus sous l'hypothèse d'indépendance. Les écarts dérivés des résidus sont considérés comme statistiquement significatifs si leur valeur est supérieure à 1,96 ou inférieure à -1,96, car les résidus tendent à suivre une loi normale centrée réduite.

Pour ce qui concerne les variables liées à la gestion de l'eau, les répondants du groupe 1 sont significativement plus nombreux à penser que les ressources en eau dans la commune sont aujourd'hui menacées (H_MENACE_RESS), que la quantité de la ressource en eau disponible dans leur commune aujourd'hui n'est pas suffisante pour satisfaire les besoins de tous les usagers (H_QUANTO_AUJ) et qu'il en sera de même dans cinq ans (H_QUANTO_5ANS). En termes de perception de risque également ces répondants se distinguent. Ils sont significativement plus nombreux à penser que le réseau d'eau du robinet de leur commune est actuellement concerné par un risque de sécheresse (H_SECH_RISK).

Logiquement en raison de la façon dont les groupes ont été construits, les répondants du groupe 1 sont surreprésentés parmi ceux qui déclarent avoir subi au moins une mesure de restrictions des usages de l'eau du robinet liées à la sécheresse au cours des 5 dernières années (REST). De même, ils sont proportionnellement plus nombreux à penser qu'il est très probable qu'une sécheresse se produise dans les cinq prochaines années dans leur commune (H_PROBA_SECH).

Enfin, les individus du groupe 3 se distinguent par une ancienneté de résidence dans la commune (ANC_COMMUNE) plus importante. Ceci pourrait avoir une influence sur leur expérience du risque de sécheresse par exemple.

Au final, l'ensemble de ces résultats laisse à penser qu'il existerait bien un lien entre vulnérabilité objective et subjective, qu'il s'agit maintenant d'étudier aussi en termes de préférences des répondants quant aux mesures de lutte contre les conséquences des sécheresses sur la performance du service d'eau potable.

8.5.3 Analyse économétrique des choix

8.5.3.1 Rappel des attributs et de leurs niveaux

Une description des attributs ayant servi à décrire les alternatives proposées dans les cartes de choix, et de leurs niveaux, est reprise dans le Tableau 17 (cf. rapport intermédiaire de 2022 pour une présentation détaillée du questionnaire et du protocole d'enquête).

Tableau 17 : Description des attributs et de leurs niveaux

Attribut	Description	Niveaux	
PROBA	Probabilité d'être touché(e) par les conséquences des sécheresses	1. 40 %	
		2. 30 %	
		3. 20 %	
		4. 10 %	
DUREE	Durée des conséquences des épisodes de sécheresse (en jours par an)	1. 25 jours de plus qu'aujourd'hui	
		2. 20 jours de plus qu'aujourd'hui	
		3. 15 jours de plus qu'aujourd'hui	
		4. 10 jours de plus qu'aujourd'hui	
PRESSION	Pression de l'eau	1. Faible	
		2. Forte	
		3. Moyenne	
PARTAGE	Acteurs concernées par un partage de l'effort de réduction des consommations d'eau du robinet	1. Ménages uniquement	
		2. Ménages + agriculteurs	
		3. Ménages + industriels	
		4. Ménages + agriculteurs + industriels	
DISPO	Disponibilité de l'eau du robinet	1. Coupures d'eau	
		2. Rationnements des quantités d'eau	
		3. Restrictions des usages	
COUT	Augmentation de la facture d'eau (en euros par an et par ménage)	0. 0 €	4. 48 €
		1. 12 €	5. 60 €
		2. 24 €	6. 72 €
		3. 36 €	

Dans la modélisation économétrique, les attributs PROBA, DUREE et COUT sont codés comme des variables quantitatives continues. Au contraire, PRESSION2, PRESSION3, PARTAGE2, PARTAGE3, PARTAGE4, DISPO2 et DISPO3 sont des variables dichotomiques qui prennent la valeur 1 lorsque ce niveau de l'attribut est présent dans l'alternative, 0 sinon.

8.5.3.2 Résultats du logit conditionnel

Le Tableau 18 présente les résultats du logit conditionnel, modèle qui n'explique les choix des répondants entre les différentes alternatives d'une même carte de choix que sur la base des attributs. L'ensemble des modélisations ont été réalisées avec le package Apollo de R (Hess & Palma, 2019).

Tableau 18 : Résultats du logit conditionnel

	Coefficients		CAP (en € de 2022)	
<i>Attributs</i>				
ASC	-0,2864***	(0,1067)	-22,31***	[6,88 – 37,73]
PROBA	0,0094***	(0,0017)	0,73***	[0,44 – 1,02]
DUREE	0,0078***	(0,0025)	0,61***	[0,22 – 1,00]
PRESSION2	0,2036***	(0,0470)	15,86***	[8,29 – 23,42]
PRESSION3	0,1787***	(0,0469)	13,91***	[6,54 – 21,28]
PARTAGE2	0,3664***	(0,0540)	28,53***	[18,98 – 38,09]
PARTAGE3	0,5542***	(0,0590)	43,16***	[31,78 – 54,54]
PARTAGE4	0,9193***	(0,0613)	71,59***	[56,80 – 86,38]
DISPO2	0,4936***	(0,0514)	38,44***	[28,64 – 48,24]
DISPO3	0,1468***	(0,0468)	11,43***	[3,94 – 18,91]
COUT	-0,0128***	(0,0010)		
<i>Paramètres du modèle</i>				
Nombre d'individus / d'observations	1 030 / 6 180			
Log de vraisemblance	-5 896,08			
Pseudo-R ² de McFadden	0,13			
AIC	11 814,15			
BIC	11 888,17			

Note : * indique une significativité à 10 % (p-value inférieure à 0,10), ** à 5 % (p-value inférieure à 0,05) et *** à 1 % (p-value inférieure à 0,01). Les valeurs entre parenthèses sont les standard errors robustes et celles entre crochets les intervalles de confiance à 95 % des CAP.

Les résultats montrent que les coefficients des attributs sont tous significativement différents de zéros. En d'autres termes, l'ensemble des niveaux des attributs a une influence sur les choix entre les alternatives d'une même série de choix. Le coefficient de COUT est négatif tel que cela est attendu, indiquant que les personnes interrogées sont sensibles à l'augmentation de la facture d'eau proposée dans une alternative : plus cette dernière est élevée, plus la probabilité que l'alternative en question soit

choisie sera faible (Latinopoulos, 2014; Syuhada et al., 2020; Wang et al., 2018). Les coefficients des autres niveaux d'attributs sont bien tous positifs. Le coefficient négatif de la constance spécifique à l'alternative (Alternative Specific Constant – ASC) indique que les répondants préfèrent éviter le scénario de référence. L'ASC a en effet été codée en référence à l'alternative C et prend donc la valeur 1 lorsque cette alternative est considérée, 0 sinon. Toutes choses égales par ailleurs, les ménages wallons enquêtés ont donc une préférence pour des alternatives dans lesquelles la performance du système d'approvisionnement en eau potable ne se dégrade pas à l'avenir du fait de l'amplification des épisodes de sécheresse.

La deuxième colonne du tableau affiche les résultats de l'estimation des CAP moyens, calculés grâce à la méthode Delta (Greene, 2003). Les intervalles de confiance à 95 % sont indiqués entre crochets. Le CAP moyen attribué à la réduction d'un point de la probabilité d'être touché par les conséquences des sécheresses (PROBA) est égal à 0,73 €. En d'autres termes, pour que la probabilité d'être touché par les conséquences des sécheresses soit réduite de 10 % (e.g. de 40 %, la valeur de référence, à 30 %, le niveau 2 de cet attribut), les ménages sont prêts à payer leur facture d'eau 7,31 € de plus par an. Le CAP pour réduire la durée des conséquences des épisodes de sécheresses (DUREE) est de 0,61 € par jour de moins, sur la base d'une valeur de référence de 25 jours. Ainsi, pour que l'augmentation de la durée des conséquences se limite à 10 jours au lieu de 25 jours, le CAP est de l'ordre de 9,15 € par ménage et par an.

De manière plus inattendue, les niveaux supérieurs des attributs PRESSION et DISPO ne bénéficient pas des CAP les plus élevés. En effet, les répondants sont prêts à payer 15,86 € pour avoir une pression moyenne (PRESSION2) plutôt qu'une pression faible et 13,91 € pour une pression forte (PRESSION3) plutôt qu'une pression faible. Ils préféreraient donc un niveau intermédiaire de ce critère de performance du service fourni. De même, leur CAP pour un rationnement de l'eau plutôt que des coupures d'eau ponctuelles (DISPO2) s'élève à 38,44 € mais celui d'un maintien de l'accès à l'eau (DISPO3) à 11,43 €. Il convient de rappeler que ces CAP s'inscrivent dans un contexte d'arbitrage entre les niveaux d'attributs et ne sont donc pas évalués de façon indépendante. Ils ne peuvent donc pas, pour un même attribut, être additionnés.

Enfin, les répondants sont fortement en faveur d'un partage de l'effort de réduction des consommations d'eau potable. En effet, le CAP pour un partage d'effort entre ménages et agriculteurs (PARTAGE2), par rapport au scénario où seuls les ménages seraient mobilisés, est de 28,53 € par an. Pour un partage d'effort entre ménages et industriels (PARTAGE3), il monte à 43,16 € et atteint même 71,59 € lorsque tous les acteurs sont concernés (PARTAGE4). Les ménages wallons attachent beaucoup d'importance à cette caractéristique des scénarios.

8.5.3.3 Résultats du modèle à classes latentes

Le modèle à classes latentes permet de prendre en compte l'hétérogénéité dans les préférences des ménages Wallons quant aux caractéristiques d'un service d'eau performant (Boxall & Adamowicz, 2002). La variable expliquée du modèle reste la probabilité de choisir une alternative plutôt qu'une autre dans une série de choix, mais les caractéristiques socio-géo-économiques des répondants sont maintenant prises en compte comme variables explicatives, en plus des niveaux d'attributs.

La première étape consiste à choisir le nombre de classes le plus adapté aux données. Pour faire ce choix, deux principaux paramètres sont généralement utilisés : le critère d'information d'Akaike (AIC) (Swait, 1994) et le critère d'information bayésien (BIC) (Roeder et al., 1999). On y ajoute un troisième, le CAIC (Consistent Akaike Information criterion), dérivé du AIC (Bozdogan, 1987), qui augmente la pénalité pour la complexité du modèle. La littérature préconise de choisir les modèles avec les AIC (et en l'occurrence les CAIC) et BIC les plus faibles (Lanza & Rhoades, 2013).

Le Tableau 19 résume les informations des modèles allant de 2 à 5 classes latentes. Bien que le modèle à 5 classes latentes ait les AIC, CAIC et BIC les plus faibles, nous optons pour celui à 4 classes car la baisse est, pour les trois critères d'information, bien plus importante lorsque l'on passe de 2 à 3 ou de 3 à 4 classes que pour une augmentation de 4 à 5 classes. Par ailleurs, le modèle à 4 classes présente l'avantage de contenir plus de paramètres significatifs et il est plus facile à interpréter qu'un modèle avec un nombre plus élevé de classes.

Tableau 19 : Comparaison entre les modèles selon le nombre de classes latentes

Nombre de classes	Nombre de paramètres	Log de vraisemblance	AIC	CAIC	BIC
2	23	-4 838,98	9 723,96	9 855,88	9 878,73
3	35	-4 709,46	9 488,92	9 689,67	9 724,44
4	47	-4 588,72	9 271,45	9 541,02	9 587,72
5	59	-4 534,13	9 186,26	9 524,67	9 583,28

Le Tableau 20 présente les résultats du modèle à 4 classes latentes¹⁶. Les covariables ayant une influence sur la probabilité d'appartenance aux classes sont reportées dans la partie du bas. Il s'agit pour certaines de caractéristiques socio-géo-économiques des répondants mais surtout de variables de perception ou de connaissance. Ainsi, les personnes dont le revenu mensuel net du ménage est supérieur à 3 500 €, la moyenne de l'échantillon (PLUS_REV_MOY), sont plus susceptibles d'appartenir à la classe 3 qu'à la classe 4 (utilisée comme référence), au contraire de celles citant la maîtrise du prix parmi les deux critères les plus importants pour un service d'eau du robinet performant (CRIT_PX) ou celles pensant que la ressource en eau représente surtout un enjeu économique (ENJEU_ECO). Penser que le prix de l'eau du robinet (hors assainissement) dépend avant tout des coûts liés à l'entretien des infrastructures de distribution augmente la probabilité d'appartenir aux classes 1 et 3 (COUT_AEP_INFRA) et être incité à réduire sa

¹⁶ Le nombre d'observations est plus faible que dans le logit conditionnel car certains répondants n'ont pas complété l'ensemble des questions relatives aux caractéristiques socio-géo-économiques et ne peuvent donc pas être intégrés à l'analyse.

consommation d'eau afin de laisser des ressources en eau en qualité et quantité suffisantes pour les générations futures (RAIS_GEN_FUT) celle d'appartenir aux classes 2 et 3.

Tableau 20 : Résultats du modèle à 4 classes latentes

	Classe 1		Classe 2		Classe 3		Classe 4	
<i>Attributs</i>								
ASC	-1,8183***	(0,2700)	-2,9297***	(1,0958)	-4,0144***	(1,2759)	2,1619**	(1,0518)
PROBA	0,0036	(0,0087)	-0,0021	(0,0208)	0,0977***	(0,0213)	0,0107	(0,0219)
DUREE	0,0110	(0,0097)	-0,0303	(0,0416)	-0,0063	(0,0063)	0,0553	(0,0398)
PRESSIION2	0,3408**	(0,1608)	1,2009	(0,8445)	-1,8589***	(0,4548)	-0,8880**	(0,3786)
PRESSIION3	0,1951	(0,1285)	0,7437	(0,5400)	0,2892	(0,1478)	-0,8582	(0,5254)
PARTAGE2	0,4933***	(0,1553)	-0,8129*	(0,4717)	1,3976***	(0,2750)	0,3566	(0,4982)
PARTAGE3	1,1425***	(0,1958)	-1,6408**	(0,7458)	-0,1468	(0,3920)	1,1898**	(0,5446)
PARTAGE4	1,3985***	(0,1755)	0,3469	(0,4023)	0,5274	(0,3346)	1,5507***	(0,5795)
DISPO2	0,1812	(0,1490)	-0,9437**	(0,4495)	2,9572***	(0,4236)	0,0027	(0,6248)
DISPO3	-0,5050***	(0,1523)	-0,2893	(0,4499)	2,6929***	(0,5739)	0,7261**	(0,3561)
COUT	-0,0389***	(0,0061)	0,0088	(0,0103)	-0,0073**	(0,0032)	-0,0402***	(0,0102)
<i>Fonctions d'appartenance aux classes</i>								
CONSTANTE	0,1545	(0,3378)	-1,0717*	(0,6490)	-0,1966	(0,3583)		
PLUS_REV_MOY	-0,2291	(0,2975)	-0,1565	(0,4860)	0,7296**	(0,2844)		

QUANT5_INF	0,9300***	(0,2966)	0,5999**	(0,3615)	0,7773**	(0,3193)
CRIT_PX	0,0362	(0,2757)	-0,1070	(0,3796)	-0,7784**	(0,3212)
RAIS_GEN_FUT	0,4940	(0,3012)	0,6544*	(0,3666)	0,8092***	(0,2841)
ENJ_AEP_PX	-0,4981*	(0,2792)	-0,2989	(0,4140)	-1,5306***	(0,3539)
FACTURE_PAYE_NSP	-10,3941***	(0,8735)	-10,4134***	(1,0364)	-9,6373***	(0,9468)
COUT_AEP_INFRA	0,7456***	(0,2772)	0,1148	(0,3645)	0,7701***	(0,2778)
ENJEU_ECO	-0,4345	(0,3837)	0,6063	(0,5159)	-0,8570**	(0,4656)
SECH_CONF_PERS	-0,8403**	(0,3662)	-1,1511**	(0,6388)	-0,4061**	(0,3550)
GROUPE_2	0,7312**	(0,2994)	1,1414**	(0,4470)	0,9730***	(0,3138)
GROUPE_3	0,3823	(0,3029)	0,5070	(0,4568)	0,6531**	(0,3087)
<i>Paramètres du modèle</i>						
Nombre d'individus / d'observations	842 / 5 052					
Log de vraisemblance	-3 657,43					
AIC	7 474,87					
BIC	7 997,07					

Note : * indique une significativité à 10 %, ** à 5 % et *** à 1 %. Les valeurs entre parenthèses sont les standard errors robustes.

Le coefficient lié au fait de penser que la quantité des ressources en eau disponible dans la commune dans cinq ans est inférieure aux besoins (QUANT15_INF) est négatif dans les trois classes, indiquant que les répondants de la classe 4 ne partagent pas cette perception. Cette classe est également caractérisée par des répondants estimant que la maîtrise du prix de l'eau sera un enjeu important concernant la gestion de l'eau du robinet dans leur commune dans les 5 prochaines années (ENJ_AEP_PX), qui ne savent pas à quel organisme ils payent leur facture d'eau (FACTURE_PAYE_NSP) et ne faisant confiance à personne dans la gestion du risque de sécheresse (SECH_CONF_PERS).

On notera enfin l'influence du groupe de communes sur la probabilité d'appartenance aux classes et donc sur l'hétérogénéité des préférences. La variable GROUPE_2 a un estimateur positif et significatif dans les trois classes. En d'autres termes, les individus habitant dans une commune dans laquelle la sécheresse a des conséquences sur le fonctionnement de la SWDE mais peu pour les clients ont une probabilité plus forte d'appartenir aux classes 1, 2 et 3 que ceux du groupe 1 qui ont connu au moins un épisode de sécheresse ayant entraîné des mesures de restriction des usages depuis janvier 2017. Habiter dans une commune du GROUPE_3 dans laquelle la sécheresse n'a aucune conséquence ou est inexistante augmente la probabilité d'appartenir à la classe 3. On peut donc en déduire que les habitants du groupe 1 se retrouvent plutôt dans la classe 4.

Le Tableau 20 montre aussi et surtout que les préférences pour les attributs sont différentes selon les classes. Ainsi, pour la classe 1 qui réunit la majorité des répondants (39 %), tous les niveaux du PARTAGE des efforts de réduction de consommation d'eau potable sont significatifs et positifs. Ces répondants sont par ailleurs indifférents à la réduction de la probabilité d'être touché par les conséquences des sécheresses (PROBA) et à la réduction de la durée des conséquences des épisodes de sécheresses (DUREE) et ont tendance à éviter les scénarios avec un maintien d'accès à l'eau de robinet (DISPO3). À l'inverse, un partage des efforts de réduction de consommation d'eau entre ménages et agriculteurs (PARTAGE2) et entre ménages et industriels (PARTAGE3) joue négativement sur la probabilité de choix de la classe 2. Ils sont également peu favorables à un rationnement des quantités d'eau disponibles (DISPO2). Enfin, cette classe est la seule pour laquelle l'attribut monétaire a un coefficient non significatif. Ceci indique que ces répondants, qui représentent environ 10 % de l'échantillon, ne tiennent pas compte du coût des mesures de prévention proposées lorsqu'ils font des choix.

Concernant la classe 3, les répondants sont les seuls à afficher une préférence significative pour une réduction de la probabilité d'être touché par les conséquences de sécheresse (PROBA). Ils sont également très nettement en faveur d'une pression forte de l'eau de robinet (PRESSION3), une pression moyenne (PRESSION2) jouant négativement sur leur probabilité de choix. Enfin, s'ils n'accordent d'importance qu'au partage des efforts de réduction de consommation d'eau entre ménages et agriculteurs, la continuité du service est une caractéristique essentielle de sa performance pour ces répondants, qui représentent 36 % de l'échantillon. Pour finir, la classe 4 qui réunit environ 15 % des répondants, est la seule pour laquelle le coefficient de l'ASC est positif, indiquant une préférence pour la situation de référence. Elle se caractérise également par un rejet d'une pression moyenne de l'eau et par une préférence pour un partage des efforts de réduction de consommation d'eau entre ménages et industriels ou entre tous les acteurs (i.e. ménages, agriculteurs et industriels). Un maintien de l'accès à l'eau joue significativement et positivement sur leur probabilité de choix.

8.5.4 Caractérisation des classes et estimation des consentements-à-payer

8.5.4.1 Caractérisation des classes

Sur la base des résultats du modèle à classes latentes, chaque répondant peut être assigné à la classe à laquelle il a la plus grande probabilité d'appartenir. Une analyse bivariée, dont les résultats sont résumés dans le Tableau 21, est ensuite réalisée sur la base des tests de khi-deux pour les variables qualitatives et de Kruskal-Wallis pour les variables quantitatives afin d'analyser l'existence de relations statistiques entre un certain nombre de caractéristiques socio-géo-économiques des répondants et la classe d'appartenance. Il est à noter que l'ensemble des variables qui ne servent pas déjà à expliquer l'appartenance aux classes ont été testées ici et que seules celles pour lesquelles la différence se révèle significative sont reportées dans le Tableau 21.

Tableau 21 : Résultats de l'analyse bivariée par classes

Variable → modalité	Classe 1	Classe 2	Classe 3	Classe 4	Total
<i>Caractéristiques socio-géo-économiques</i>					
NIV_ETUDE** → Master et plus	28 %	27 %	37 %	19 %	30 %
NB_PERS***	2,4 pers.	2,2 pers.	2,7 pers.	2,4 pers.	2,5 pers.
<i>La gestion de l'eau</i>					
VISION_RESS** → Un enjeu sanitaire	26 %	17 %	25 %	36 %	26 %
<i>Votre service d'eau du robinet</i>					
H_QUALO_AUJ** → Plutôt pas d'accord	13 %	17 %	11 %	16 %	13 %
H_QUAL_SERVICE*** → Plutôt de mauvaise qualité	10 %	15 %	8 %	19 %	11 %
<i>Le risque de sécheresse</i>					
H_SECH_AUJ*** → Non, plutôt pas	12 %	14 %	11 %	17 %	12 %
H_SECH_NBFUTUR*** → Rester stables	9 %	15 %	10 %	21 %	11 %
H_SECH_NBFUTUR_COMM*** → Rester stables	14 %	20 %	12 %	29 %	16 %
H_PROBA_SECH*** → Improbable	7 %	15 %	9 %	16 %	10 %

Note : * indique une significativité à 10 % (p-value inférieure à 0,10), ** à 5 % (p-value inférieure à 0,05) et *** à 1 % (p-value inférieure à 0,01). Afin d'en faciliter la lecture, n'apparaît dans ce tableau pour chaque variable que la modalité pour laquelle il existe la différence la plus significative entre les classes.

Les résultats mettent en évidence une surreprésentation des individus de la classe 3 parmi ceux avec un master et plus et une sous-représentation des individus de la classe 4 dans ce groupe. Par ailleurs, le nombre de personnes dans un ménage (NB_PERS) en moyenne est plus élevé pour la classe 3 (2,7) que pour les autres classes (2,5). Les individus de la classe 4 sont également significativement plus nombreux à considérer que la ressource en eau représente principalement un enjeu sanitaire (ENJ_SANIT), à ne pas estimer que la ressource en eau utilisée pour l'eau du robinet dans leur commune est de bonne qualité (H_QUALO_AUJ) ou encore à penser que le service d'eau du robinet actuellement fourni est de mauvaise qualité (H_QUAL_SERVICE). Au contraire, les répondants de la classe 3 sont significativement plus satisfaits du service fourni.

Nous observons aussi une disparité en termes de perception du risque entre les classes. Les répondants de la classe 4 sont ainsi significativement moins préoccupés par le risque de sécheresse en Wallonie (H_SECH_AUJ) que les autres et plus nombreux à estimer improbable qu'une sécheresse se produise dans les cinq prochaines années dans leur commune (H_PROBA_SECH). Ils sont également surreprésentés parmi ceux qui pensent que le nombre de sécheresse va rester stable à l'avenir en Wallonie (H_SECH_NBFUTUR) comme dans leur commune (H_SECH_NBFUTUR_COMM). En contrepartie, la classe 3 est sous-représentée dans ce dernier groupe.

En conclusion, les tendances suivies par la classe 4 diffèrent par rapport aux autres classes. Leur perception du risque est moins marquée et ils semblent plus optimistes quant au risque de sécheresse et à ses conséquences.

8.5.4.2 Représentation des liens entre groupes de communes et classes

Sur la base de cette réaffectation des individus aux classes, il est également possible de représenter le lien existant avec le groupe de communes¹⁷. Une distribution en classes par groupe de commune est ainsi présentée sur la Figure 50. Elle confirme que les individus habitant dans une commune du Groupe 1 sont plus présents dans la classe 4 (et dans une moindre mesure de la classe 1) et ceux du Groupe 2 dans la classe 2. Les répondants du Groupe 3 sont équitablement répartis dans les quatre classes.

¹⁷ Ces variables qui sont utilisées dans le modèle pour expliquer l'appartenance aux classes ne peuvent par construction être intégrées à l'analyse bivariée.

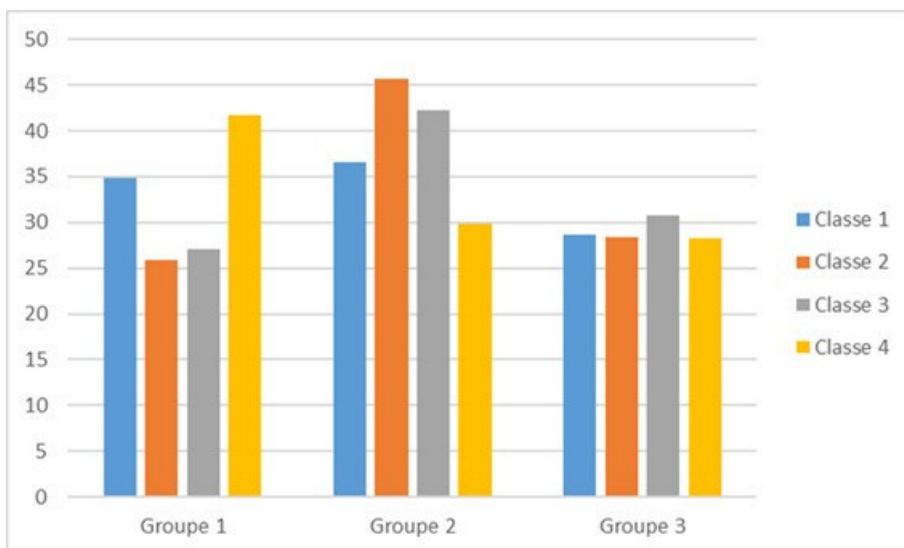


Figure 50 : Répartition en classes par groupe de commune

La représentation spatiale de la densité des enquêtés par groupe de commune et par classe (Figure 51, Figure 52, Figure 53 et Figure 54) souligne la même tendance.

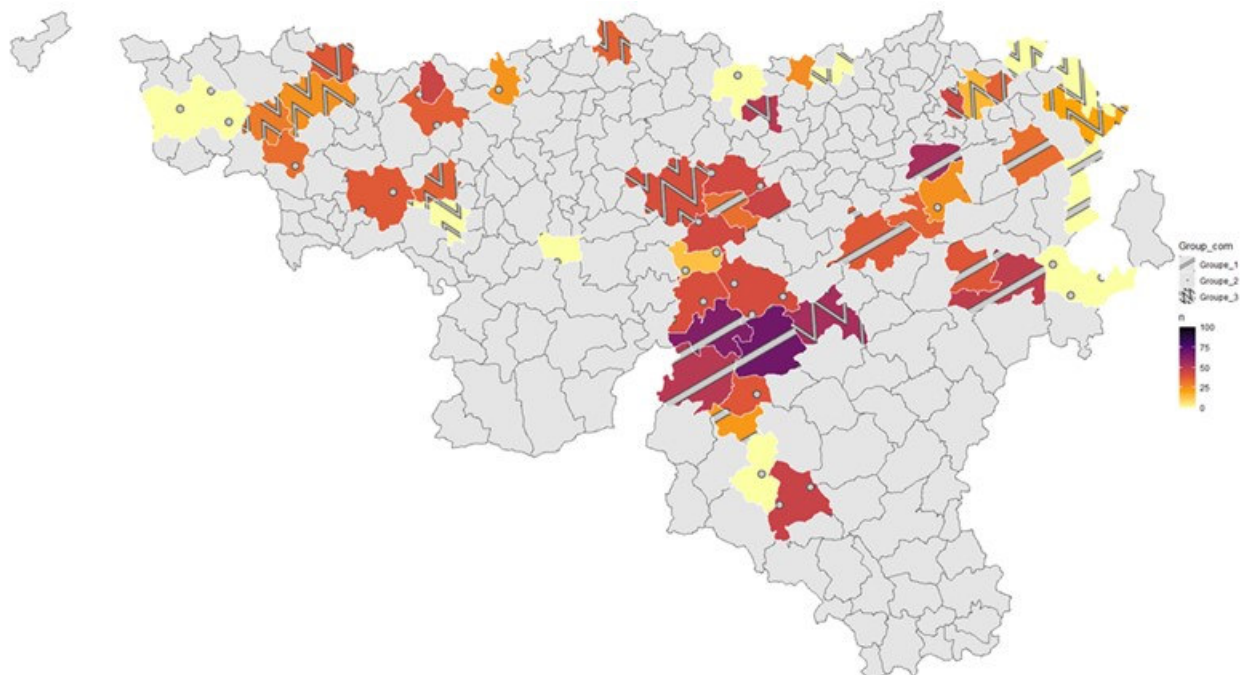


Figure 51 : Distribution spatiale de la densité des enquêtés par groupe de commune (Classe 1)

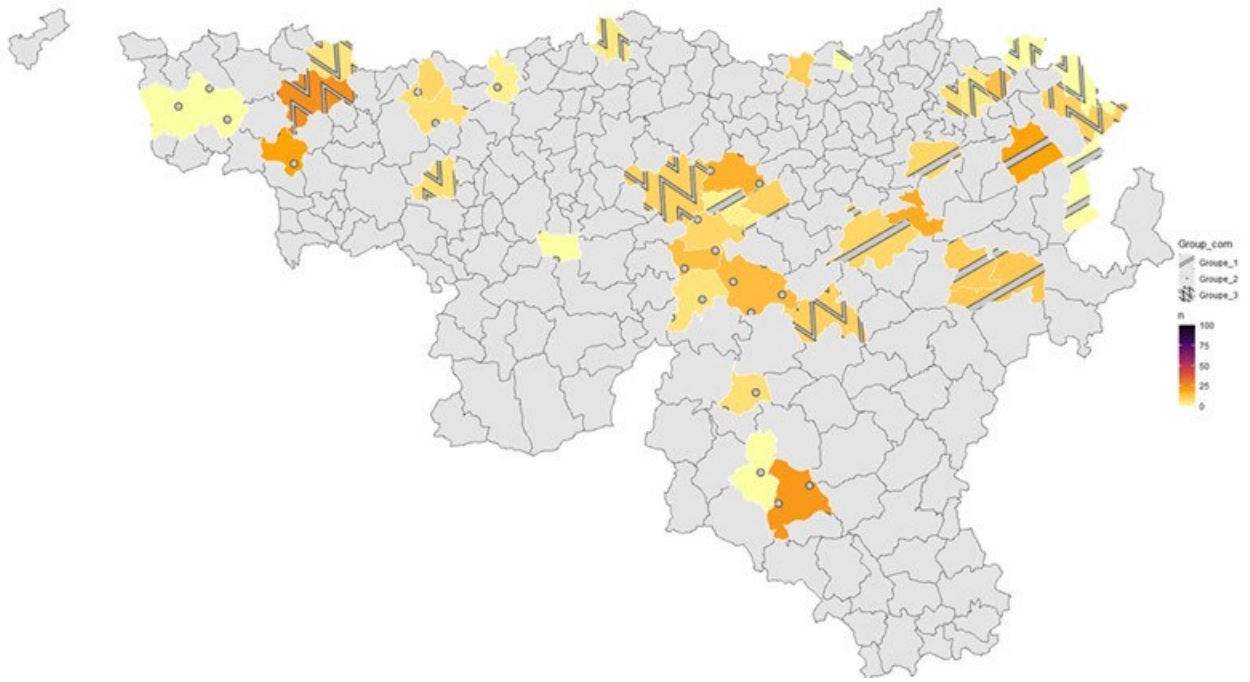


Figure 52 : Distribution spatiale de la densité des enquêtés par groupe de commune (Classe 2)

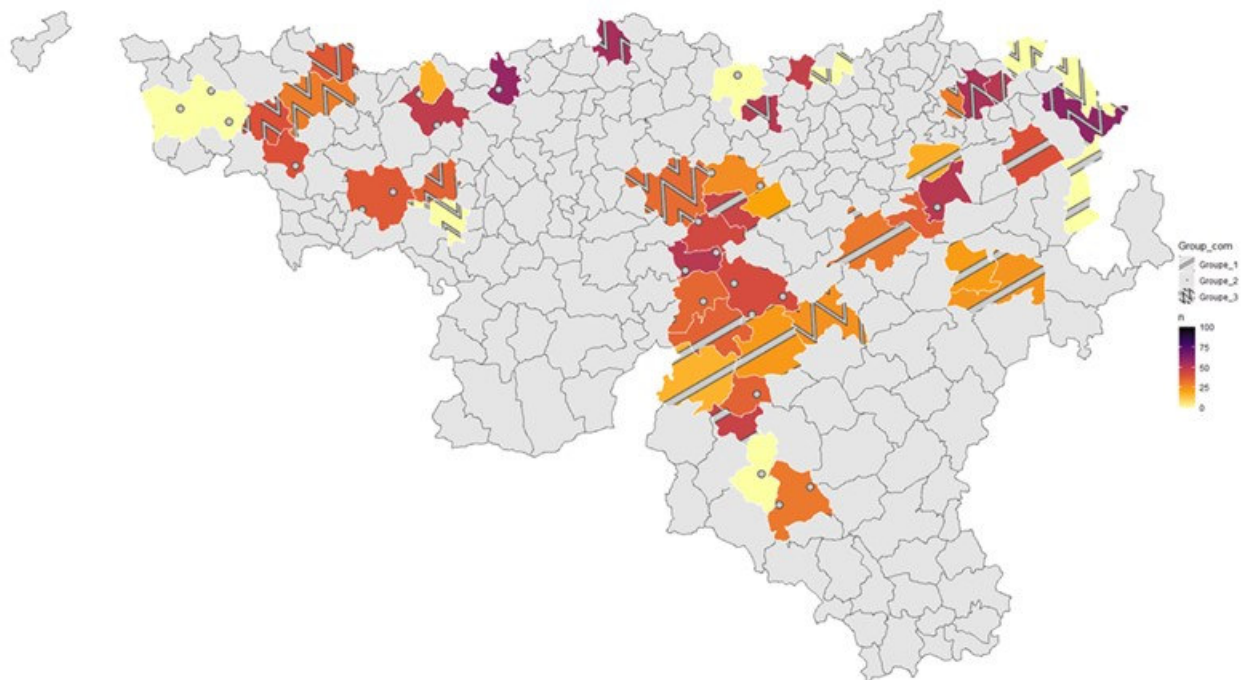


Figure 53 : Distribution spatiale de la densité des enquêtés par groupe de commune (Classe 3)

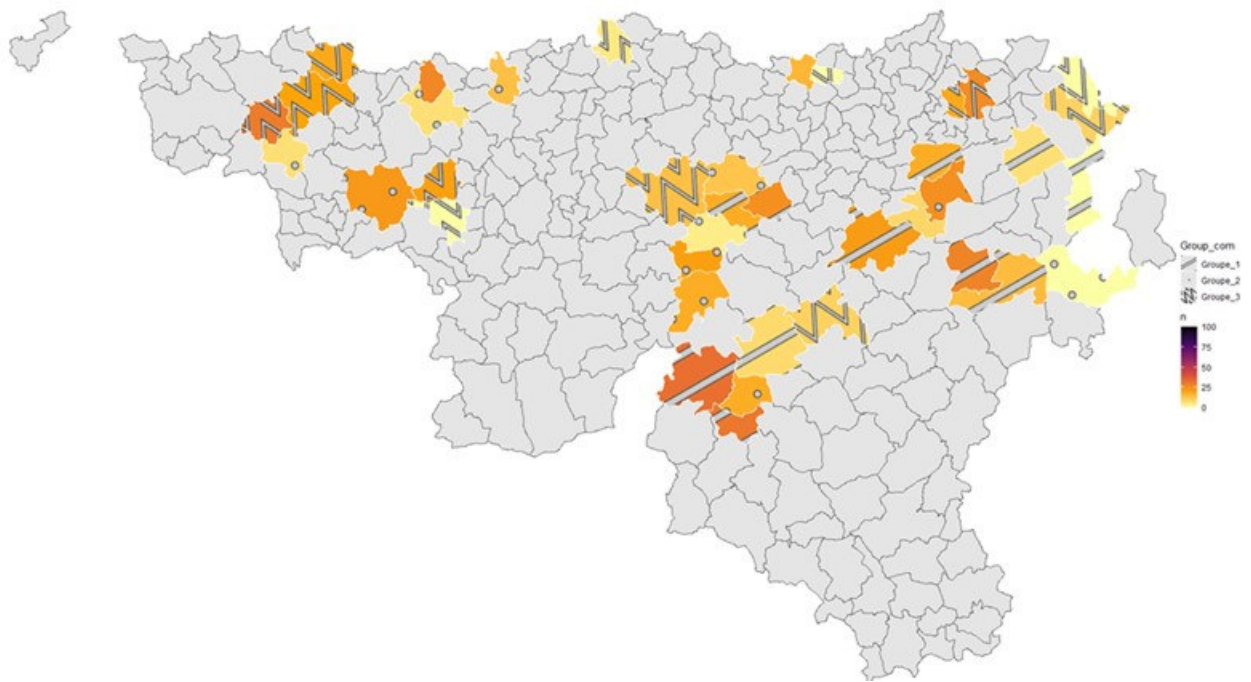


Figure 54 : Distribution spatiale de la densité des enquêtés par groupe de commune (Classe 4)

8.5.4.3 Estimation des consentements-à-payer

Le Tableau 22 affiche les résultats de l'estimation des CAP moyens issue du modèle à classes latentes et calculés grâce à la méthode Delta (Greene, 2003). À noter que la classe 2 n'apparaît pas ici car le coefficient de COUT n'était pas significatif pour cette classe (cf. Tableau 20) donc les CAP ne peuvent être calculés.

Tableau 22 : Estimation des CAP issus du modèle à classes latentes

	Classe 1	Classe 3	Classe 4
ASC	-46,80*** [-68,71 – -24,87]	-548,00** [-1124,87 – 28,86]	53,76 [-17,96 – 125,47]
PROBA	0,09 [-0,32 – 0,51]	13,33** [1,34 – 25,32]	0,27 [-0,89 – 1,42]
DUREE	0,29 [-0,19 – 0,75]	-0,86 [-2,79 – 1,06]	1,38 [-0,49 – 3,24]
PRESSION2	8,77* [-0,94 – 18,49]	-253,80** [-452,65 – -54,87]	-22,08** [-41,03 – -3,14]
PRESSIO N3	5,02 [-1,23 – 11,27]	39,47 [-15,30 – 94,24]	-21,34 [-51,51 – 8,83]
PARTAGE2	12,70*** [3,18 – 22,22]	190,80** [41,16 – 340,42]	8,87 [-15,93 – 33,67]
PARTAGE 3	29,40*** [18,02 – 40,79]	-20,04 [-125,17 – 85,09]	29,59 [-6,23 – 65,4]
PARTAGE4	35,99*** [21,04 – 50,94]	72,00 [-33,93 – 177,93]	38,56*** [0,99 – 76,13]
DISPO2	4,66 [-3,59 – 12,91]	403,70** [75,08 – 732, 29]	0,07 [-30,40 – 30,54]
DISPO3	-13,00*** [-21,05 – -4,94]	367,60** [50,32 – 684,89]	18,05 [-4,71 – 40,82]

Note : * indique une significativité à 10 %, ** à 5 % et *** à 1 %. Les valeurs entre crochets sont les intervalles de confiance à 95 % des CAP.

La prise en compte de l'hétérogénéité permet de mettre en évidence le fait que les classes ne valorisent pas de manière uniforme l'ensemble des attributs. Pour la classe 1, les CAP associés aux différents niveaux des attributs PRESSION2 et PARTAGE2-4 sont significatifs et positifs, et leurs valeurs sont légèrement inférieures à celles issues du logit conditionnel. Le CAP attribué à l'ASC est significatif et négatif, ce qui reflète une aversion pour la situation de référence chez ces répondants. Ils auraient une désutilité si un tel programme était mis en œuvre, quels que soient les attributs qu'il contient. Les

individus de cette classe semblent sensibilisés au risque de sécheresse et à ses conséquences, ayant une préférence pour le partage des efforts de réduction de consommation d'eau entre les acteurs concernés à différents niveaux d'engagement, ainsi que pour un niveau intermédiaire de pression de l'eau. Enfin, la classe 1 n'accorde pas de valeur à une réduction de la probabilité d'être touché par les conséquences des sécheresses, à leur durée, à un maintien d'une pression forte de l'eau plutôt que faible et enfin au rationnement des quantités d'eau à disposition plutôt qu'à des coupures d'eau. Ils attribuent même une valeur négative au fait de devoir subir des restrictions des usages, cette solution leur créant donc une réduction de bien-être par rapport à une situation de coupures d'eau ponctuelles (par exemple un jour par semaine sur les mois de juillet et août).

Les répondants de la classe 3 expriment des CAP significatifs pour ASC, PROBA, PARTAGE2, DISPO2 et DISPO3 et PRESSION2. À noter que les CAP pour ASC et PRESSION2 sont négatifs, le premier révélant une préférence pour la mise en place des mesures de prévention aux conséquences des sécheresses, et le deuxième un souhait de ne pas voir la pression de l'eau devenir « moyenne » plutôt que « faible », toutes choses égales par ailleurs. Et ce même si ces répondants ne sont pas disposés à payer pour que la pression atteigne un niveau « fort ». La classe 3 est ainsi composée d'individus souhaitant conserver un certain niveau de confort quant à l'accès à l'eau potable dans un contexte de sécheresses (forte préférence pour la disponibilité de l'eau) et exprimant une aversion pour les conséquences de ces dernières. Par ailleurs, les CAP de la classe 3 sont très élevés en valeur absolue par rapport à ceux des autres classes. À titre d'exemple, le CAP associé à une réduction de 1 point de PROBA est de l'ordre de 13,33 € par an et par ménage, soit, pour 10 points tel que cela était proposé dans le questionnaire, l'équivalent du montant de la facture d'eau pour un ménage de 3 personnes en Wallonie. Il faut toutefois ici de nouveau rappeler que les CAP pour les différents niveaux d'un même attribut ne peuvent être additionnés entre eux.

Enfin, la classe 4 affiche deux CAP significatifs : ceux relatifs à PRESSION2 et à PARTAGE4. Le CAP négatif pour PRESSION2 reflète les mêmes préférences que la classe 3 pour cet attribut. De plus, les individus de cette classe seraient prêts à payer 38,56 € par an et par ménage pour que l'effort de réduction des consommations d'eau potable soient partagés entre tous les acteurs concernés (ménages, agriculteurs et industriels). Cette classe représenterait donc les individus souhaitant principalement l'intégration du collectif dans les mesures à adopter pour la prévention des conséquences de sécheresses. Enfin, le coefficient négatif du COUT pour la classe 2 indique que les individus de cette classe sont indifférents quant aux niveaux de performance étudiés du système de distribution de l'eau dans un contexte de sécheresse.

8.6 Conclusion

Cette étude empirique, basée sur la MCMA, examine les arbitrages réalisés par les clients de la SWDE concernant les multiples dimensions de la performance future de leur service de distribution d'eau, dans un contexte de changement climatique engendrant une augmentation de la fréquence et de la sévérité des épisodes de sécheresse, ainsi que leurs CAP pour éviter les conséquences de ces dernières sur l'alimentation en eau potable. Les résultats de cette évaluation offrent des perspectives intéressantes. Nous constatons en effet que les ménages wallons valorisent l'ensemble des attributs proposés, lorsqu'ils sont considérés dans leur ensemble (logit conditionnel). Ils accordent notamment une grande importance au partage de l'effort de réduction des consommations d'eau potable, d'autant plus lorsque tous les acteurs sont concernés (ménages, agriculteurs et industriels).

Mais ce travail met aussi en évidence une forte hétérogénéité des préférences entre les ménages (logit à classes latentes), hétérogénéité qui peut être liée notamment à la vision qu'à le répondant de la ressource en eau (e.g. la voir comme un enjeu économique) ou du prix du service d'eau (estimer que le prix est un des deux critères les plus importants pour un service d'eau du robinet performant, que la maîtrise du prix de l'eau sera un enjeu important concernant la gestion de l'eau du robinet dans sa commune dans les 5 prochaines années ou encore que le prix de l'eau du robinet dépend avant tout des coûts liés à l'entretien des infrastructures de distribution mais également ne pas savoir à qui on paye sa facture d'eau). Mais l'appartenance aux différentes classes présentant des préférences différentes est aussi liée à des variables de perception de et de sensibilité au risque de sécheresse (réduire sa consommation d'eau afin de laisser des ressources en eau en qualité et quantité suffisantes pour les générations futures, ne faire confiance à personne dans la gestion du risque de sécheresse).

Enfin, en lien avec notre objectif de départ de comparaison entre la vulnérabilité « objective » telle que définie par la SWDE et la vulnérabilité « subjective » telle que ressentie par les ménages Wallons, on notera que le groupe de communes joue un rôle significatif dans l'appartenance aux classes. Ce résultat est encore renforcé par la caractérisation des classes, réattribuant les répondants à la classe à laquelle ils ont la plus grande probabilité d'appartenir. À titre d'exemple, la classe 4, regroupant 15 % de l'échantillon, est caractérisée par une surreprésentation des ménages appartenant au groupe de commune ayant subi des mesures de restriction des usages depuis janvier 2017 (groupe 1). Cette classe se distingue des autres classes, notamment par ce qui s'apparente à une perception du risque moins marquée et à un optimisme apparent quant au risque de sécheresse et à ses conséquences. Ceci pourrait être expliqué par la théorie de normalisation du risque (Lima, 2004; Lima et al., 2005; Richardson et al., 1987) selon laquelle des individus habitués au risque de sécheresse qui menacent leur commune tendent à en minimiser les conséquences, même dans le futur. Cette analyse appelle ainsi à une prise en compte différenciée des individus qui ne soit pas uniquement basée sur la vulnérabilité « objective » mais soit complétée également par une meilleure connaissance de la vulnérabilité « subjective » des ménages wallons, mais aussi et surtout par des programmes de communication autour des enjeux stratégiques de préservation des ressources en eau et d'entretien long-terme des infrastructures ou encore de formation du prix de l'eau.

Archive Tâche 08 :

Absence d'archive pour la Tâche 08.

9 Tâche 09 - Modélisation statistique couplée des fuites et des crashes

Objectifs :

- Modéliser les lois des durées de séjour dans les états sain et fuyard avant apparition de l'évènement terminal CRASH ;
- En déduire les durées de vie moyennes, les probabilités d'état en fonction du temps passé ;
- Déterminer les effets des caractéristiques sur ces quantités ;
- Evaluer l'impact des efforts de RF sans confondre avec la fiabilité réelle des tronçons.

L'implémentation de Casses dans la section 3 considère indistinctement les évènements de crash et fuite pour fournir un modèle ajusté sur les données SWDE a des fins d'analyse des facteurs prépondérants de précipitation d'une défaillance, et comme outil prédictif du taux de casses. Pourtant, ces évènements n'ont pas la même nature :

- Les crashes s'imposent au gestionnaire, leur date est supposée connue parfaitement (pas d'évènements masqués)
- Les fuites sont le résultat d'une opération de recherche, au périmètre identifié par le gestionnaire, et la qualité de la mesure peut être discutable, les bases de données d'enregistrement n'ayant pas été conçues à la base pour un traitement statistique. Par exemple, on ne connaît pas la liste exacte des tronçons qui ont effectivement été inspectés dans la cellule pendant une RAF (opération de Recherche Active des Fuites). Il est vraisemblable que la couverture soit partielle puisque l'objectif d'efficacité demande de chercher jusqu'à ce qu'une fuite soit identifiée.

La modélisation de la probabilité de fuite au niveau des tronçons de la tâche 07 vise à valoriser les enregistrements des campagnes de recherche de fuite pour donner un nouvel outil de prédiction de fuite, avant même qu'elle conduise à un crash, visible. Cette méthode ne prend en compte que les données de RF et ne propose pas une modélisation du processus graduel de dégradation vers un état de crash.

Dans cette section, nous présentons un nouveau modèle de couplage des deux aspects pour valoriser à la fois les données de RF et de crashes. Plusieurs pistes ont été explorées :

1. Le développement d'une variante du modèle LEYP
2. La mise en œuvre d'un modèle préexistant de type LEYP avec délai
3. L'application d'un modèle multi-états markovien

Les deux premières tentatives n'ayant pas été concluantes, elles seront présentées succinctement pour présenter ensuite la démarche choisie du modèle multi-état markovien.

9.1 Développement d'une variante de Casses et du modèle LEYP

Une première piste explorée a été de tenter d'insérer dans les données de crashes et fuites (les données d'entrée pour le calage de LEYP), une variable explicative d'intensité de RF concernant le tronçon : sur sa cellule, aux périodes où on sait qu'une campagne de RF est activée (quelques jours), il est naturel qu'un évènement ait plus de chance d'être observé. Tenir compte de cette augmentation ponctuelle du risque peut être envisagée en construisant une variable indicatrice, dépendant du temps et captant l'information de l'existence d'une campagne à des dates spécifiées et pour des durées spécifiées. Le logiciel Casses ne prend pas en compte cette possibilité, il est théoriquement possible de définir une

méthode de calage de modèle mais ceci a été laissé de côté pour des raisons de temps nécessaire au développement, sans garantie de succès.

9.2 Mise en œuvre du modèle de type LEYP avec délai.

Un modèle de type LEYP intégrant la possibilité d'apparition décalée des événements dans le temps a été proposé au niveau théorique dans Cha et Finkelstein 2021. Ce modèle qui semblait prometteur demandait un développement supplémentaire car la technique d'estimation statistique n'existe pas encore (seul le modèle stochastique est présenté, avec des calculs de caractéristiques de fiabilité). Cette piste n'a finalement pas la maturité suffisante pour être directement appliquée sur données réelles.

9.3 Application un modèle Multi-états Markovien pour la dégradation des tronçons

9.3.1 Généralités

Les modèles multi-états sont des modèles stochastiques décrivant l'évolution dans le temps de l'état des individus (ici les tronçons) où l'état est une quantité prenant ses valeurs dans un ensemble discret de valeurs possibles (état1, état2, ... étatK). Ce sont des modèles très utilisés en biostatistique et pour l'analyse de l'évolution des maladies (voir Beyersmann, Allignol, et Schumacher 2011 pour la théorie).

On appelle transition le fait de passer (à un moment t) d'un état E_k à un état $E_{k'}$ ($k \neq k'$). L'utilisateur identifie alors quelles transitions sont possibles (toutes ne le sont pas), et dans le cadre général, on décrit alors le modèle par le graphe des transitions possibles (réversibles ou non, Figure 55).

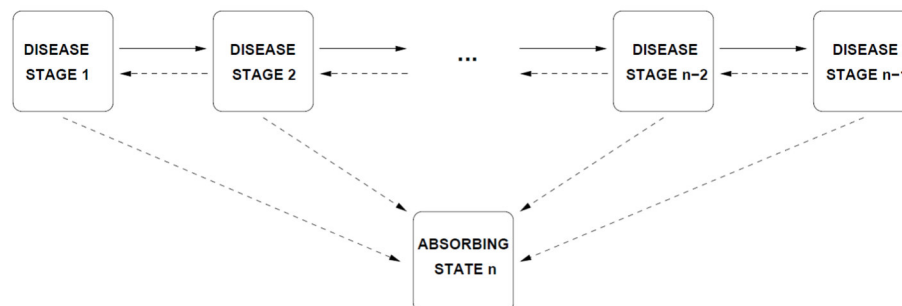


Figure 55. Exemple de graphe de Markov, tiré de Chris Jackson (2014).

Les paramètres d'un modèle Markovien multi-états sont alors les intensités de transitions entre états, regroupés dans la matrice des intensités (Figure 56). C'est une matrice carrée de taille $(n \times n)$ s'il y a n états possibles. Chaque valeur (hors la diagonale) q_{ij} est l'intensité de transition entre l'état i et l'état j . Quand un terme q_{ij} de la matrice est nul, il n'y a pas de transition possible de i vers j .

$$Q = \begin{pmatrix} q_{11} & q_{12} & 0 & 0 & \dots & q_{1n} \\ q_{21} & q_{22} & q_{23} & 0 & \dots & q_{2n} \\ 0 & q_{32} & q_{33} & q_{34} & \ddots & q_{3n} \\ 0 & 0 & q_{43} & q_{44} & \ddots & q_{4n} \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 \end{pmatrix}$$

Figure 56. Matrice des intensités du graphe de Markov.

Ces valeurs numériques sont assimilées à des paramètres de lois de durées de vie exponentielles, et tous les calculs sont possibles à partir de ces valeurs :

- Probabilité d'être dans un état donné à un moment donné
- Temps moyens de séjour dans chaque état
- Fonction de survie jusqu'à l'état terminal.

Dans le cadre du projet, on imagine naturellement que les tronçons sont dans un état neuf à la pose, sans fuite, puis pourront passer à un moment inconnu vers un état de fuite (tronçon fuyard), pendant lequel une recherche de fuite sur ce tronçon pourrait identifier l'état de fuite, et, si la fuite n'est pas réparée, iront (c'est une hypothèse) irrémédiablement vers un crash (Figure 57).

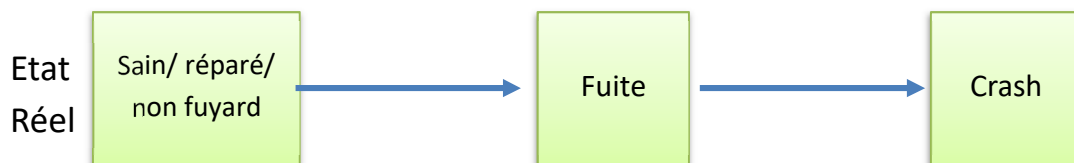


Figure 57. Graphe de Markov de succession des états. 2 transitions sont possibles : de sain vers fuite, puis de fuite vers crash.

On suppose donc ici que les fuites ne peuvent pas disparaître d'elles-mêmes, et que l'état de crash est un état terminal. Bien sûr, les transitions de retour vers un état non fuyard (ou réparé) sont possibles, mais on va alors considérer qu'il s'agit d'un nouvel individu statistique, avec une nouvelle trajectoire d'évolution et des caractéristiques identiques (matériau, âge, diamètre, ...), mais une variable du comptage de défaillances passées incrémentée de 1.

9.3.2 Intégration de covariables

Comme pour Casses et la modélisation de la tâche 07, il est possible d'intégrer la variabilité des comportements des tronçons en ajoutant des covariables au modèle. Chaque intensité de transition est alors formulée en fonction des covariables choisies, avec un jeu de paramètres spécifiques à estimer.

On peut aussi imaginer stratifier la modélisation en appliquant un modèle pour des sous-ensembles de tronçons (par ZF, par matériau, par nombre de défaillances passées...), comme cela a été fait lors de la tâche 03 (voir le paragraphe 3.1 page 42).

9.3.3 Le cas particulier des données de panel

Alors que les modèles les plus complexes sont facilement ajustables sur des jeux de données dont les dates de transitions sont connues pour un ensemble d'individus, l'ajustement statistique est néanmoins possible pour des observations de panel, c'est-à-dire des données apportant une connaissance partielle des trajectoires résumée dans la connaissance des états à quelques dates d'observation (Figure 58).

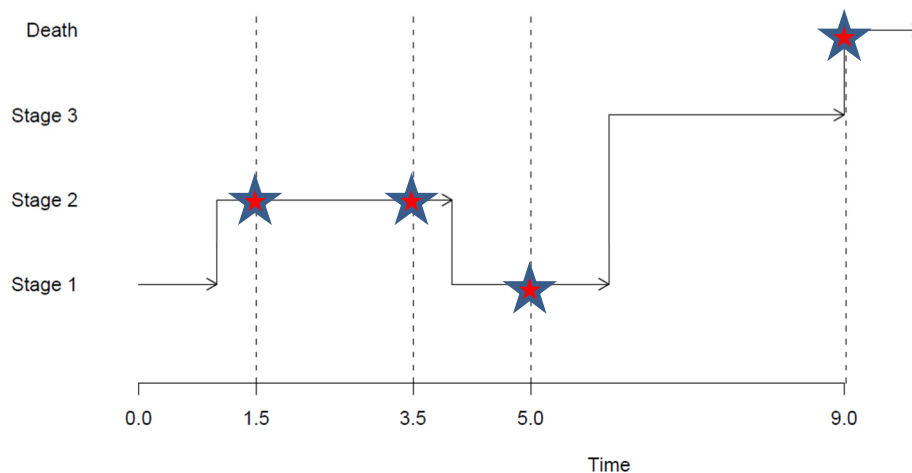


Figure 58. Exemple de trajectoire. La trajectoire se compose de 4 transitions non observées, mais 4 états observés aux dates 1.5 (state2), 3.5 (state 2), 5 (state 1), 9 (Death, état terminal).

La Figure 58 illustre une trajectoire observée avec des données de panel, qui comprennent 3 dates d'observation donnant la succession d'états 2-2-3, puis l'observation exacte de la date de mort (transition vers état Death).

La plupart du temps, les modèles avec transitions irréversibles sont ajustables à partir de données de panel.

Dans notre cas, les dates d'observation seront de deux types : pour un tronçon donné,

- Les dates de RF sur la cellule contenant le tronçon : donnée standard de type panel
- Les dates de crashes observés sur le tronçon : donnée exacte

9.3.4 Le cas particulier des données d'inspection générant une possibilité de mauvaise classification : modèle HMM

On peut en outre modéliser la possibilité d'erreur d'observation et admettre qu'une séquence d'états observée (par exemple 1->1->1->2->3->Death) comporte des erreurs. Des paramètres supplémentaires sont alors à estimer : pour chaque état réel, les probabilités d'observations de chaque état. Elles sont notifiées dans une matrice des probabilités d'erreurs, à lire en ligne (Figure 59).

Par exemple pour le projet SWDE, on peut supposer qu'une RF sur une cellule conduise pour les tronçons considérés à une erreur possible d'observation d'état « sain » alors qu'il est fuyard : c'est en fait la probabilité que le tronçon ait vraiment été inspecté pendant la RF. Nous ne revenons pas ici sur la structure du jeu de données des recherches de fuites (voir Tâche 07, page 74), qui génère un très grand nombre d'observations sur les tronçons concernés par une recherche quand :

- Les tronçons sont positionnés dans une cellule ayant fait l'objet d'une RF à la date t,
- A la date t, 0, 1 ou plus, fuites ont été identifiées dans cette cellule sur quelques tronçons donc tous les autres sont **par défaut** déclarés « non fuyard ». On sait que la campagne n'inspecte pas exhaustivement l'ensemble des tronçons, il y a donc un fort taux de **faux négatifs** (un tronçon non inspecté, mais fuyard).

Puisqu'on suppose qu'il n'y a pas de faux positifs dans les données finalement utilisées dans le modèle (« trou pour rien », le tronçon est déclaré fuyard à tort), la matrice de mauvaise classification est :

		Observations		
		Sain	fuyard	crash
Etats Réels	Sain	1	0	0
	Fuyard	1-p	p	0
	crash	0	0	1

Figure 59. Matrice de probabilité d'erreur de classification. Le paramètre p est à estimer par le modèle.

9.3.5 Synthèse du formatage des données SWDE pour le calage du modèle HMM

La Figure 60 résume le principe de formatage des données. Si la RF inspecte vraiment le tronçon, son vrai état est remonté (sain ou fuyard). Si la RF n'inspecte pas le tronçon, l'état « sain » est remonté, mais c'est peut-être un faux négatif (si le tronçon est fuyard). Nous appellerons alors tous les états différents de fuyard comme « incertain ».

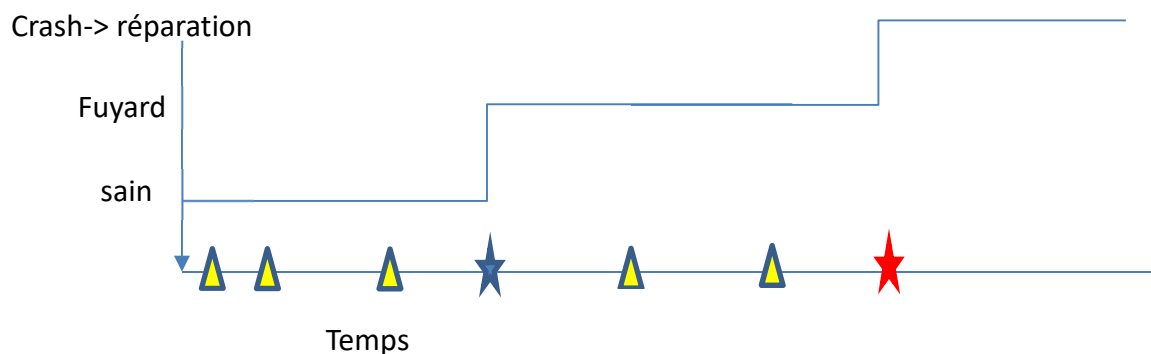


Figure 60. Trajectoire d'un tronçon. Les dates de RF de sa cellule sont indiquées par des triangles jaunes. Les 3 premières sont faites sur un tronçon sain, les deux suivantes auraient pu révéler la fuite mais il n'a pas été inspecté, finalement, le tronçon expérimente un crash (étoile rouge).

Nous avons construit une table de données regroupant l'ensemble des tronçons, et l'état supposé (ou connu) à chaque date de RF, ou de crash, assortie des variables de caractéristiques pouvant impacter les intensités de transitions (Figure 61 à Figure 63). Les modélisations précédentes permettent de sélectionner les variables potentielles : nombre de défaillances passées, longueur (ou son Log), matériau, diamètre, pourcentage de route ou d'argile, CLC1, nombre de raccordement.

IDT	date	etat	MAT3	STATUT	LNG	NB_RACCT	DIA	DDP	DHS	Proute	Pargile	CLC1	Lcell_km	
1	100000366	2016-03-11	incertain	FONTE GRISE	ES	5.530752	5	60	1948	NA	0.0000000	0	0	19.71779
2	100000366	2018-05-22	incertain	FONTE GRISE	ES	5.530752	5	60	1948	NA	0.0000000	0	0	19.71779
3	100000366	2019-03-29	incertain	FONTE GRISE	ES	5.530752	5	60	1948	NA	0.0000000	0	0	19.71779
4	100000366	2019-06-27	incertain	FONTE GRISE	ES	5.530752	5	60	1948	NA	0.0000000	0	0	19.71779
5	100000366	2021-03-18	incertain	FONTE GRISE	ES	5.530752	5	60	1948	NA	0.0000000	0	0	19.71779
6	100000366	2021-12-31	fin_obs	FONTE GRISE	ES	5.530752	5	60	1948	NA	0.0000000	0	0	19.71779

Figure 61.Exemple d'une trajectoire vierge d'évènement.

IDT	date	etat	MAT3	STATUT	LNG	NB_RACCT	DIA	DDP	DHS	Proute	Pargile	CLC1	Lcell_km	Ndef0	IDT2	
154	100014330	2016-10-13	incertain	PVC	ES	5.718074	31	110	2012	NA	0.3380888	0	0	11.16738	0	100014330.0
155	100014330	2016-10-20	incertain	PVC	ES	5.718074	31	110	2012	NA	0.3380888	0	0	11.16738	0	100014330.0
156	100014330	2016-11-22	incertain	PVC	ES	5.718074	31	110	2012	NA	0.3380888	0	0	11.16738	0	100014330.0
157	100014330	2017-01-11	incertain	PVC	ES	5.718074	31	110	2012	NA	0.3380888	0	0	11.16738	0	100014330.0
158	100014330	2017-01-27	incertain	PVC	ES	5.718074	31	110	2012	NA	0.3380888	0	0	11.16738	0	100014330.0
159	100014330	2017-12-13	incertain	PVC	ES	5.718074	31	110	2012	NA	0.3380888	0	0	11.16738	0	100014330.0
160	100014330	2017-12-18	incertain	PVC	ES	5.718074	31	110	2012	NA	0.3380888	0	0	11.16738	0	100014330.0
161	100014330	2019-01-28	crash	PVC	ES	5.718074	31	110	2012	NA	0.3380888	0	0	11.16738	0	100014330.0
162	100014330	2019-01-28	repaired	PVC	ES	5.718074	31	110	2012	NA	0.3380888	0	0	11.16738	1	100014330.1
163	100014330	2019-03-01	incertain	PVC	ES	5.718074	31	110	2012	NA	0.3380888	0	0	11.16738	1	100014330.1
164	100014330	2019-07-23	incertain	PVC	ES	5.718074	31	110	2012	NA	0.3380888	0	0	11.16738	1	100014330.1

Figure 62.Exemple d'une trajectoire avec un crash (2 trajectoires créées).

IDT	date	etat	MAT3	STATUT	LNG	NB_RACCT	DIA	DDP	DHS	Proute	Pargile	CLC1	Lcell_km	Ndef0	IDT2	
420	100014372	2020-12-18	incertain	FONTE GRISE	ES	5.681306	59	60	1950	NA	0.7277990365	0	0	11.16738	0	100014372.0
421	100014372	2020-12-28	incertain	FONTE GRISE	ES	5.681306	59	60	1950	NA	0.7277990365	0	0	11.16738	0	100014372.0
422	100014372	2021-01-13	fuite	FONTE GRISE	ES	5.681306	59	60	1950	NA	0.7277990365	0	0	11.16738	0	100014372.0
423	100014372	2021-01-13	repaired	FONTE GRISE	ES	5.681306	59	60	1950	NA	0.7277990365	0	0	11.16738	1	100014372.1
424	100014372	2021-10-26	incertain	FONTE GRISE	ES	5.681306	59	60	1950	NA	0.7277990365	0	0	11.16738	1	100014372.1
425	100014372	2021-10-29	incertain	FONTE GRISE	ES	5.681306	59	60	1950	NA	0.7277990365	0	0	11.16738	1	100014372.1
426	100014372	2021-11-05	incertain	FONTE GRISE	ES	5.681306	59	60	1950	NA	0.7277990365	0	0	11.16738	1	100014372.1
427	100014372	2021-11-18	incertain	FONTE GRISE	ES	5.681306	59	60	1950	NA	0.7277990365	0	0	11.16738	1	100014372.1
428	100014372	2021-11-24	incertain	FONTE GRISE	ES	5.681306	59	60	1950	NA	0.7277990365	0	0	11.16738	1	100014372.1
429	100014372	2021-12-30	incertain	FONTE GRISE	ES	5.681306	59	60	1950	NA	0.7277990365	0	0	11.16738	1	100014372.1
430	100014372	2021-12-31	fin_obs	FONTE GRISE	ES	5.681306	59	60	1950	NA	0.7277990365	0	0	11.16738	1	100014372.1

Figure 63. Exemple d'une trajectoire avec fuite trouvée par RF, puis réparée (2 trajectoires).

9.3.6 Ajustement et modélisation par le package msm de R

Le package *msm* permet un ajustement des modèles Markovien multi-états avec erreur possible de classification, données de panel, et état terminal parfaitement observé (Christopher Jackson 2011). Il inclut la possibilité de covariables sur les intensités, sur la probabilité d'erreur de classification. Les paramètres des variables peuvent être spécifiques à chaque transition, ou supposés communs.

L'estimation repose sur une maximisation de la vraisemblance, qui nécessite donc un choix des valeurs initiales, et n'est pas assurée de converger. Si la convergence est faite, on a accès à une matrice hessienne qui permet le calcul des intervalles de confiance.

Nous avons exploré un grand nombre de modèles possibles selon les variables choisies, selon les ZF, selon le nombre de défaillances passées (modèle A, B, C ...). Tous ne seront pas présentés.

- La variable « *nombre de défaillances passées* » dans les covariables échoue à donner un modèle correctement ajusté, on obtient un énorme coefficient de Ndef dans les hazard ratios (voir paragraphe 9.3.7.3 page 136), sans doute dû à la structure des données (trajectoires réinitialisées à chaque fuite ou crash) et surtout à la définition de Ndef0 : nombre de défaillance observées depuis 2016. Nous n'avons pas la connaissance suffisante du passé et la valeur Ndef=1 est possible pour un tronçon avec 10 défaillances depuis 30 ans (mais une seule depuis 2016), ou un tronçon neuf dont c'est la première défaillance.
- L'estimation finale de p (1-p = probabilité de mauvais classement sain | fuite c'est-à-dire de non-inspection d'un tronçon fuyard) est de l'ordre de 10% à 14%, ce qui semble beaucoup quand on tente une comparaison avec les résultats de la tâche 07).
- Les coefficients des variables semblent cohérents et robustes selon les modèles.

Selon les modèles, il faut entre 30 minutes et 4 heures pour obtenir l'ajustement.

9.3.7 Résultats

9.3.7.1 Préparation des données

Le modèle a été ajusté sur un sous-échantillon de tronçons de plus de 6 mètres, en filtrant selon les trois zones fonctionnelles MO01, VE06 et NA03, donnant 1396345 observations avant le 31/12/2021 (inspections datées sur cellule, fuites et crashes) sur 63791 tronçons. Les observations de crashes trop rapprochées dans le temps sur chaque tronçon ont été agrégées s'il y a moins de 8 jours entre deux dates, considérant qu'il s'agit certainement du même évènement en cours de réparation. Par exemple :

IDT 200002870 : crashes le 11/11/2021 et le 12/11/2021

IDT 200005300 : crashes le 30/01/2020, le 31/01/2020 et le 01/02/2020

IDT 200005300 : crashes les 13, 23 et 24/12/2021

IDT 200018566 : crashes les 6 et 7/09/2016

IDT 200020019 : crashes les 22, 23, 25/05/2018

IDT 200024300 : crashes les 22 et 23/03/2018

IDT 200029641 : crashes les 2,4,5 et 7/10/2021

STARAPIC-176477 : crashes les 01 et 02/11/2021

Ce prétraitement des données est crucial car la modélisation cherche à estimer des paramètres associés aux durées de séjour dans les états sain et fuyard. Les durées entre évènements, surtout avérés de crash, ont donc un impact important sur le calage du modèle. Une discussion devra être engagée sur la qualité des données et les résultats doivent être considérés avec prudence. Le choix du seuillage à 8 jours est subjectif.

Sur les trois zones fonctionnelles sélectionnées, évidemment la majorité concerne les « transitions » incertain-incertain qui résume l'information que deux campagnes de RF successives ont été effectuées sur une cellule, et qu'un grand nombre de tronçons n'a pas été l'objet d'une détection de fuite (Tableau 23). « Fin_obs » correspond au 31/12/2021, « repaired » correspond à la première date

d'une nouvelle trajectoire d'un tronçon virtuel ayant fait l'objet d'une fuite ou d'un crash (Figure 60, Figure 62 et Figure 63).

Le nombre de réparations (somme du nombre de fuites et du nombre de crashes) correspond finalement au nombre de trajectoires démarrant après une réparation (1335+850+77+339+3707+196+730= 7234). Cela donne un effectif vis-à-vis des trajectoires sans défaillance initiale (trajectoire à la première observation « incertain ») et on verra que celles-ci ne pourront pas être utilisées pour aboutir à un bon modèle ajusté.

Tableau 23. Effectifs des transitions observées des 3ZF après suppressions des double crashes (écarts en jours <=8 jours)

ACIER

	to					
from	crash	fin_obs	fuite	incertain	repaired	tot
incertain	695	5988	326	172675	0	179684
	0,39%	3,33%	0,18%	96,10%	0,00%	100,00%
repaired	105	156	50	1024	0	1335
	7,87%	11,69%	3,75%	76,70%	0,00%	100,00%
	800	6144	376	173699	0	181019

AMIANTE

	to					
from	crash	fin_obs	fuite	incertain	repaired	tot
incertain	488	4940	194	94418	0	100040
	0,49%	4,94%	0,19%	94,38%	0,00%	100,00%
repaired	63	90	23	674	0	850
	7,41%	10,59%	2,71%	79,29%	0,00%	100,00%
	551	5030	217	95092	0	100890

AUTRE

	to					
from	crash	fin_obs	fuite	incertain	repaired	tot
incertain	29	603	34	12390	0	13056
	0,22%	4,62%	0,26%	94,90%	0,00%	100,00%
repaired	5	8	1	63	0	77
	6,49%	10,39%	1,30%	81,82%	0,00%	100,00%
	34	611	35	12453	0	13133

FONTE_DUCTILE

	to					
from	crash	fin_obs	fuite	incertain	repaired	tot
Incertain	169	9312	84	193672	0	203237
	0,08%	4,58%	0,04%	95,29%	0,00%	100,00%
Repaired	25	47	3	264	0	339
	7,37%	13,86%	0,88%	77,88%	0,00%	100,00%
	194	9359	87	193936	0	203576

FONTE_GRISE

from	to crash	fin_obs	fuite	incertain	repaired	tot
Incertain	1880	22279	1063	447601	0	472823
	0,40%	4,71%	0,22%	94,67%	0,00%	100,00%
Repaired	185	396	117	3009	0	3707
	4,99%	10,68%	3,16%	81,17%	0,00%	100,00%
	2065	22675	1180	450610	0	476530

PE

from	to crash	fin_obs	fuite	incertain	repaired	tot
incertain	97	1983	54	41453	0	43587
	0,22%	4,55%	0,12%	95,10%	0,00%	100,00%
repaired	12	17	16	151	0	196
	6,12%	8,67%	8,16%	77,04%	0,00%	100,00%
	109	2000	70	41604	0	43783

PVC

from	to crash	fin_obs	fuite	incertain	repaired	tot
incertain	416	17871	158	358239	0	376684
	0,11%	4,74%	0,04%	95,10%	0,00%	100,00%
repaired	31	101	13	585	0	730
	4,25%	13,84%	1,78%	80,14%	0,00%	100,00%
	447	17972	171	358824	0	377414

9.3.7.2 Choix des variables

Les variables considérées qui peuvent impacter les intensités de transition sont :

- Le nombre de défaillance passées **Ndef0** : sans autre information, seul le nombre d'évènements (fuite détectée par RF ou crash) depuis le 01/01/2016 est disponible dans le jeu de données.
- l'âge du tronçon au 01/01/2016 **age2016**, en siècle.
- le logarithme de la longueur **lnL**
- le nombre de raccordements **NB_RACCT**,
- le matériau : **MAT3** (ACIER, AMIANTE CIMENT, AUTRE, FONTE DUCTILE, FONTE GRISE, PE, PVC)
- le diamètre **DIA**
- le pourcentage sous la route **Proute**

Les autres variables construites (Pargile, Plimon, CLC1, pression) n'ont pas été utilisées, pour des raisons de temps de calcul, mais aussi de valeurs manquantes en grand nombre.

Pour la probabilité d'inspection (liée à la probabilité de mauvais classement), nous avons testé des modèles sans variable et avec la longueur de la cellule en kilomètres **Lcell_km**.

Ces variables n'ont pas de données manquantes.

Remarque : le cas de Ndef0

Bien que le nombre de défaillances, utilisé comme variable à modéliser dans Casses, ou comme covariable pour la prédiction de l'état de fuite dans la tâche 07 du modèle ZIB, soit une variable a priori impactante pour la dégradation des tronçons, et donc pour modéliser les intensités de transition vers des états de fuite ou de crash, nous avons constaté que les modèles incluant Ndef0 en covariable donnaient des résultats non exploitables. Les coefficients de Ndef0 étaient anormalement grands, les probabilités prédites, les durées de séjour dans des états sain ou fuyard extrêmement dépendantes du fait que les trajectoires soient « initiales » (ie Ndef0=0), premières trajectoires de chaque tronçon, ou « après défaillance » (ie trajectoire d'un tronçon avec Ndef=1,2 ou plus). Ce problème a été identifié mais aucune solution envisagée n'a pu aboutir.

En conséquence, deux options sont possibles : ajuster le modèle sur les fractions de trajectoires initiales (celles pour Ndef0=0, ce qui supprime une très faible fraction des trajectoires), ou considérer l'ensemble des trajectoires, sans intégrer la covariable Ndef0. Les résultats ci-dessous concerne la première option, la seconde a ensuite été appliquée.

9.3.7.3 Résultats du calage par msm, pour les trajectoires initiales Ndef0=0 : modèle F

La modélisation peut être résumée par l'ensemble des paramètres estimés (coefficients de chaque variable liée à chaque transition), ou, comme en régression logistique, par l'effet « hazard ratio » sur chaque transition. Le hazard ratio (HR) est le rapport des intensités (pour une transition $i \rightarrow j$ donnée) pour deux individus dont les caractéristiques ne changent que d'une unité sur la variable X_p . Cet effet est ici aussi un coefficient de proportionnalité sur les durées de séjour car les transitions entre états sont unidirectionnelles (états de dégradation successifs de « sain » vers « fuyard » puis « crash »).

$$HR(\text{variable } X_p, \text{transition } i \rightarrow j) = \frac{q_{i,j}(X_1, \dots, X_p = \mathbf{x} + \mathbf{1})}{q_{i,j}(X_1, \dots, X_p = \mathbf{x})} = \exp(\beta_p^{ij})$$

Il y a donc dans notre cas, pour deux transitions à modéliser (sain-fuyard puis fuyard-crash), et 5 variables numériques, une variable catégorielle à 7 modalités :

$$2 * (5 + [7 - 1]) \text{ coefficients à estimer} + 2 \text{ intercepts} = 24 \text{ paramètres}$$

Et concernant les paramètres de mauvais classement : une probabilité

$$1 - p = \text{Proba}(\text{classement sain} \mid \text{état réel fuyard})$$

le paramètre p étant finalement la probabilité de détection d'une fuite sur un tronçon fixé et concerné par une campagne de RF (Figure 59)

Ci-dessous un exemple de sortie brute du package msm (Figure 64).

```

Call:
msm(formula = state ~ time, subject = IDT2, data = df.msm.ndef0, qmatrix = q, ematrix = ematrix2, obstype = obstype, obstrue = obstrue, covariates = list('1-2' = ~age.2016 + lnL + NB_RACCT + DIA + MAT3, '2-3' = ~age.2016 + lnL + NB_RACCT + DIA + MAT3), initprobs = c(0.99, 0.01, 0), est.initprobs = TRUE, deathexact = 3, censor = 99, method = "BFGS", control = list(trace = 2, REPORT = 1, fnscale = 30000, maxit = 200, reltol = 1e-06))

Maximum likelihood estimates
Baselines are with covariates set to their means

Transition intensities with hazard ratios for each covariate
      Baseline
sain - sain  -1.994e-05 (-2.109e-05,-1.884e-05)
sain - fuite  1.994e-05 ( 1.884e-05, 2.109e-05) 0.9702 (0.7594,1.240) 2.038 (1.963,2.115) 1.027 (1.024,1.030) 0.9977 (0.9970,0.9983) 0.8699 (0.7682,0.9852) 1.171 (0.8532,1.608) 0.2828 (0.2304,0.347) 0.9777 (0.8865,1.0783) 0.6496 (0.5010,0.8423) 0.2766 (0.2381,0.3212)
fuite - fuite -2.617e-03 (-3.163e-03,-2.165e-03)
fuite - crash 2.617e-03 ( 2.165e-03, 3.163e-03) 1.0485 (0.6518,1.687) 1.244 (1.150,1.346) 1.013 (1.006,1.021) 0.9999 (0.9986,1.0012) 0.8425 (0.6416,1.1064) 0.711 (0.3809,1.327) 0.6955 (0.4690,1.031) 0.5739 (0.4693,0.7019) 0.8562 (0.5101,1.4372) 0.7058 (0.5278,0.9438)

Misclassification probabilities
      Baseline
Obs State 1 | State 2 0.8622 (0.8415,0.8806)
Obs State 2 | State 2 0

Initial state occupancy probabilities
      Estimate      LCL      UCL
State 1 0.994044192 0.992702272 0.995138566
State 2 0.005955808 0.004861434 0.007297728
State 3 0.000000000 0.000000000 0.000000000

-2 * log-likelihood: 84109.13

```

Figure 64. Exemple de sortie du package msm.

Une meilleure visualisation des effets est donnée par le tableau des rapports d'intensité (Tableau 24)

Tableau 24. Résultats du modèle F : rapports des risques par variable et par transition.

Modèle F			
traj	1389455		
Ndef==0			
tous matériaux			
age,2016 + lnL + NB_RACCT + MAT3 + DIA + Proute			
-2 * log-likelihood:	80204,67		
miscclassification : Lcell_km			
hazard,msm(fit6,Ndef0,f)			
\$age,2016	HR	L	U
sain-fuite	0,849	0,646	1,116
fuite-crash	1,467	0,795	2,709
\$lnL	HR	L	U
sain-fuite	1,874	1,796	1,956
fuite-crash	1,874	1,699	2,067
\$NB_RACCT	HR	L	U
sain-fuite	1,025	1,022	1,029
fuite-crash	1,013	1,002	1,025
\$MAT3ACIER	1	baseline	
\$`MAT3AMIANTECIMENT	HR	L	U
sain-fuite	0,977	0,860	1,108
fuite-crash	0,295	0,199	0,438
\$MAT3AUTRE	HR	L	U
sain-fuite	1,667	0,923	3,009
fuite-crash	0,071	0,028	0,182
\$`MAT3FONTEDUCTILE	HR	L	U
sain-fuite	0,348	0,280	0,433
fuite-crash	0,186	0,107	0,324
\$`MAT3FONTEGRISE	HR	L	U
sain-fuite	1,261	1,137	1,398
fuite-crash	0,088	0,067	0,117
\$MAT3PE	HR	L	U
sain-fuite	0,661	0,494	0,884
fuite-crash	0,183	0,100	0,333
\$MAT3PVC	HR	L	U
sain-fuite	0,316	0,270	0,369
fuite-crash	0,198	0,134	0,294
\$DIA	HR	L	U
sain-fuite	0,9975	0,9967	0,9982
fuite-crash	0,9996	0,9983	1,0010
\$Proute	HR	L	U
sain-fuite	1,322	1,204	1,451
fuite-crash	0,788	0,636	0,978

Par exemple :

- Le HR de la transition de sain vers fuite pour le matériau fonte grise est **1,261** : la fonte grise a une intensité de transition « sain-fuite) 26,1% plus grande que celle de l'acier. La durée de séjour est donc 21,6% plus petite.
- Le HR de la variable $\ln L$ est de **1,874** pour les deux intensités : une augmentation de 1 du $\ln L$ (ie passer de $\ln(55m)^4$ à $\ln(150m)^5$ ou de $\ln(150m)^5$ à $\ln(400m)^6$) conduit à une intensité de transition augmentée de 87,4%.

Une interprétation rapide permet de dire que :

- L'âge en 2016 (comptée en siècle) et le diamètre n'ont pas d'impact significatif sur la modélisation.
- La longueur et le nombre de raccordements ont des effets néfastes sur les intensités : plus le tronçon est long (ou plus il a de raccordements), plus les intensités augmentent, donc plus les durées de séjour raccourcissent.
- Par rapport à l'acier, tous les matériaux ont des intensités réduites (le HR est <1), et donc des durées de séjour plus longues avant apparition d'une fuite puis d'un crash (hormis la transition sain-fuite des catégories AUTRES et FONTE GRISE).
- La probabilité de mauvais classement dépend de la longueur de la cellule dans le modèle F, mais est mal ajustée. Son coefficient est très grand, ce qui donne des prévisions de probabilité de mauvais classement soit très forte (pour les cellules de longueur inférieure à 21km, la probabilité estimée vaut 1), soit très faible (pour les cellules de longueur supérieur à 22km, la probabilité estimée vaut presque 0). Les autres modèles précédemment testés (sans covariable dans la probabilité de mauvais classement), donnait des estimations de l'ordre de 10% à 14%.

9.3.7.4 Résultats du calage par msm sur toutes les trajectoires : modèle G

En intégrant aussi les portions de trajectoires après fuite ou crash, on obtient les résultats suivants (Figure 65)


```
> fit6.Ndef0.g
```

```
Call:
```

```
msm(formula = state ~ time, subject = IDT2, data = df.msm.ndef0, qmatrix = q, ematrix = ematrix2, obstype = obstype, obstrue = obstrue, covariates = list(`1-2` = ~age.2016 + lnL + NB_RACCT + MAT3 +  
DIA + Proute, `2-3` = ~age.2016 + lnL + NB_RACCT + MAT3 + DIA + Proute), misccovariates = ~1, initprobs = c(0.99, 0.01, 0), est.initprobs = TRUE, deathexact = 3, censor = 99, method =  
"BFGS", control = list(trace = 2, REPORT = 1, fnscale = 30000, maxit = 200, reltol = 1e-06))
```

```
Maximum likelihood estimates
```

```
Baselines are with covariates set to their means
```

```
Transition intensities with hazard ratios for each covariate
```

	Baseline	age.2016	lnL	NB_RACCT	MAT3AMIANTE CIMENT	MAT3AUTRE	MAT3FONTE DUCTILE
sain - sain	-1.995e-05 (-2.110e-05,-1.886e-05)						
sain - fuite	1.995e-05 (1.886e-05, 2.110e-05)	0.9192 (0.7187,1.176)	2.055 (1.979,2.134)	1.026 (1.023,1.030)	0.8803 (0.7770,0.9973)	1.1687 (0.8495,1.608)	0.2870 (0.2340,0.3519)
fuite - fuite	-2.675e-03 (-3.231e-03,-2.215e-03)						
fuite - crash	2.675e-03 (2.215e-03, 3.231e-03)	1.0297 (0.6395,1.658)	1.251 (1.154,1.355)	1.014 (1.006,1.022)	0.8425 (0.6411,1.1071)	0.6888 (0.3692,1.285)	0.6964 (0.4687,1.0346)
	MAT3FONTE GRISE	MAT3PE	MAT3PVC	DIA	Proute		
sain - sain							
sain - fuite	0.9888 (0.8963,1.0908)	0.6287 (0.4837,0.8171)	0.2786 (0.2398,0.3237)	0.9975 (0.9969,0.9982)	1.200 (1.104,1.303)		
fuite - fuite							
fuite - crash	0.5545 (0.4533,0.6784)	0.8431 (0.5021,1.4157)	0.7141 (0.5336,0.9556)	1.0000 (0.9987,1.0013)	1.002 (0.846,1.186)		

```
Misclassification probabilities
```

```
Baseline  
Obs State 1 | State 2 0.8601 (0.8392,0.8786)  
Obs State 2 | State 2 0
```

```
Initial state occupancy probabilities
```

	Estimate	LCL	UCL
State 1	0.99408461	0.992719692	0.995182627
State 2	0.00591539	0.004817373	0.007280308
State 3	0.00000000	0.000000000	0.000000000

```
-2 * log-likelihood: 84089.82
```

Figure 65. Résultats du calage par msm sur toutes les trajectoires : modèle G

Une meilleure visualisation des effets est donnée par le tableau des rapports d'intensité (Tableau 25).

Tableau 25. Résultats du modèle G : rapports des risques par variable et par transition.

Modèle G			
traj	1389455		
Ndef quelconques			
tous matériaux			
age,2016 + InL + NB_RACCT + MAT3 + DIA + Proute			
-2 * log-likelihood:	84089,82		
miscclassification :~1			
hazard,msm(fit6GNdef0.g)			
\$age.2016	HR	L	U
sain-fuite	0,919	0,719	1,176
fuite-crash	1,030	0,640	1,658
\$InL	HR	L	U
sain-fuite	2,055	1,979	2,134
fuite-crash	1,251	1,154	1,355
\$NB_RACCT	HR	L	U
sain-fuite	1,026	1,023	1,030
fuite-crash	1,014	1,006	1,022
\$MAT3ACIER			
\$`MAT3AMIANTECIMENT	HR	L	U
sain-fuite	0,880	0,777	0,997
fuite-crash	0,842	0,641	1,107
\$MAT3AUTRE	HR	L	U
sain-fuite	1,169	0,850	1,608
fuite-crash	0,689	0,369	1,285
\$`MAT3FONTEDUCTILE	HR	L	U
sain-fuite	0,287	0,234	0,352
fuite-crash	0,696	0,469	1,035
\$`MAT3FONTEGRISE	HR	L	U
sain-fuite	0,989	0,896	1,091
fuite-crash	0,555	0,453	0,678
\$MAT3PE	HR	L	U
sain-fuite	0,629	0,484	0,817
fuite-crash	0,843	0,502	1,416
\$MAT3PVC	HR	L	U
sain-fuite	0,279	0,240	0,324
fuite-crash	0,714	0,534	0,956
\$DIA	HR	L	U
sain-fuite	0,9975	0,9969	0,9982
fuite-crash	1,0000	0,9987	1,0013
\$Proute	HR	L	U
sain-fuite	1,200	1,104	1,303
fuite-crash	1,002	0,846	1,186

Dans ce modèle, l'estimation de la matrice des probabilités de mauvais classement est de 14% : selon ce modèle, lors d'une RF, un tronçon effectivement fuyard a une probabilité de 14% d'être déclaré fuyard.

9.3.7.5 Tableaux de sorties des prédictions (modèle G)

Le modèle retenu pour la présentation des résultats finaux est le modèle G (Tableau 25), ses coefficients sont globalement robustes et proches de ceux du modèle F, sans le défaut d'interprétation de la variable Lcell_km.

Un tableau synthétique des durées de séjour dans l'état sain, puis dans l'état fuyard a été construit pour l'ensemble des croisements possibles des variables. Cela représente un grand nombre de croisements, accessible dans l'Annexe T09_01 avec les outils de filtre. Les durées avant fuite sont de l'ordre de plusieurs dizaines d'année mais la durée entre l'apparition de la fuite et le crash de l'ordre de 6 mois, ceci pouvant varier selon les caractéristiques (Tableau 26):

Tableau 26. Extrait de l'Annexe T09_01.

age	LONG	NB_RACC	DIA	MAT3	Proute	sain	sain.lower	sain.upper	fuite	fuite.lower	fuite.upper
0,1	100	40	100	FONTE DUCTILE	0,4	62,05	51,10	75,35	0,54	0,36	0,80

On constate que toutes les durées de séjour prédites dans l'état fuyard sont relativement courtes (en semaines, mois, ou au plus 1 an de fuite) relativement au séjour dans l'état sain (plusieurs dizaines d'années).

Il est possible de visualiser les évolutions des probabilités d'état selon toutes les configurations des caractéristiques. On en donne quelques exemples ci-dessous. On peut alors voir par exemple l'effet du matériau (ACIER contre FONTE DUCTILE) et des longueurs de tronçons (55m=exp(4), 400m=exp(6), 1100m=exp(7)) sur les durées avant apparition d'une fuite ou d'un crash (les temps sont donnés en jours) : la fonte ductile semble plus fiable que l'acier, un tronçon en fonte ductile de 400m a globalement la même fiabilité qu'un autre en acier de 55m. Bien sûr, ces interprétations devraient être étayées et commentées par un expert métier (Tableau 27, Figure 65 et Figure 66).

Tableau 27. Répartition selon matériau et longueur des 63791 tronçons considérés dans les 3 ZF testées.

effectifs (ex :200 : [200-300])	0	100	200	300	400	500	600	700	800	900	1000	1100	1200	1300	1400	1500	1600	1700	1800	1900	2600	
ACIER	2295	2039	913	454	221	118	51	23	18	8	1	3	0	0	0	0	0	0	0	0	0	0
AMIANTE CIMENT	1945	1620	731	353	177	105	47	24	18	4	2	1	1	0	1	1	0	0	0	0	0	0
AUTRE	332	190	52	18	7	7	2	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0
FONTE DUCTILE	4897	2966	862	323	150	60	45	18	14	6	8	5	2	1	0	0	1	0	1	0	0	0
FONTE GRISE	9066	7916	3175	1267	616	287	159	79	39	29	24	10	2	1	1	1	0	1	0	1	1	1
PE	1343	406	138	53	25	18	8	2	5	0	0	0	1	0	0	1	0	0	0	0	0	0
PVC	8262	6303	2013	757	314	144	82	44	21	17	8	2	0	0	2	3	0	0	0	0	0	0

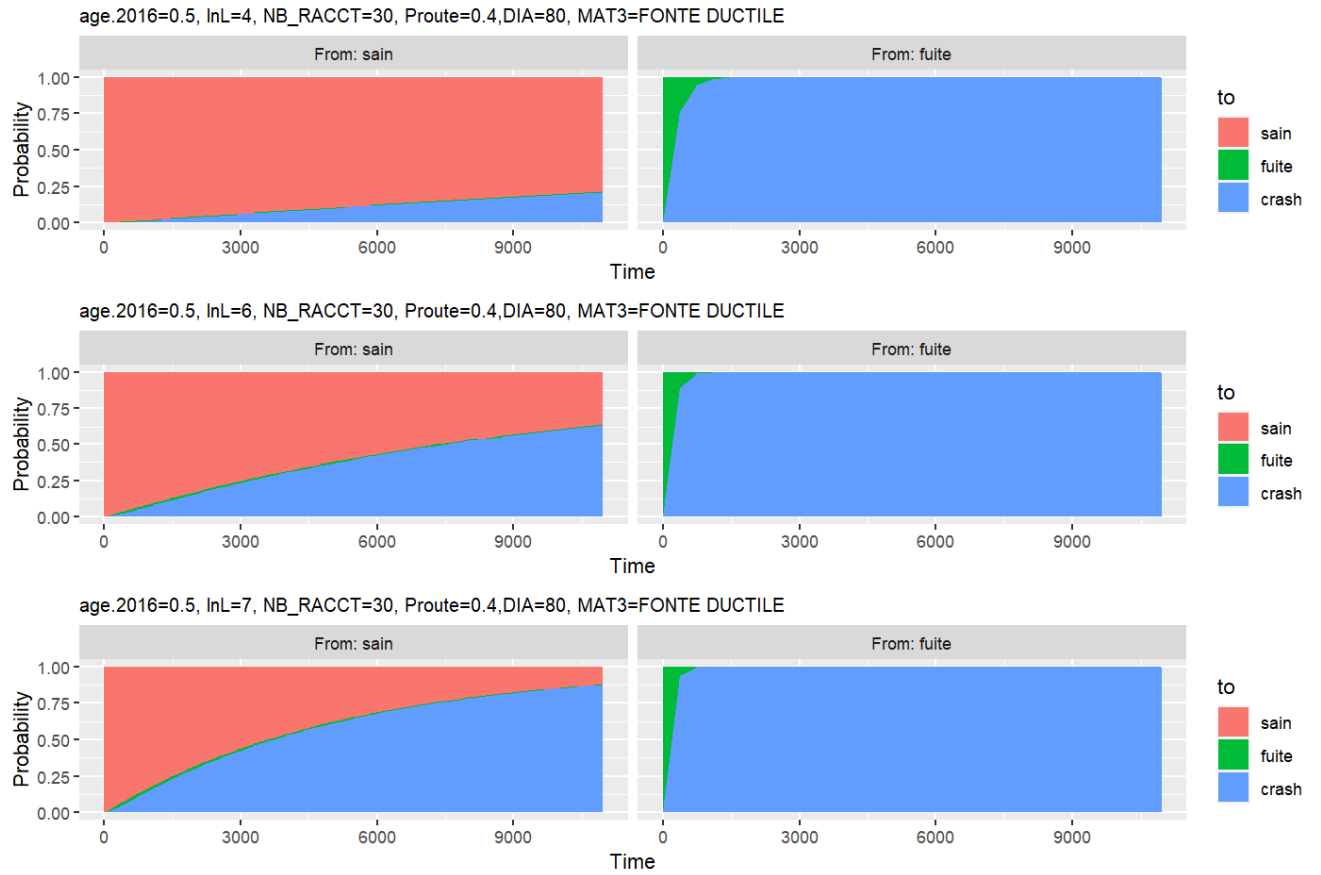


Figure 66. Probabilité de transition entre les états en fonction du temps pour un tronçon en fonte ductile de caractéristiques autres connues.

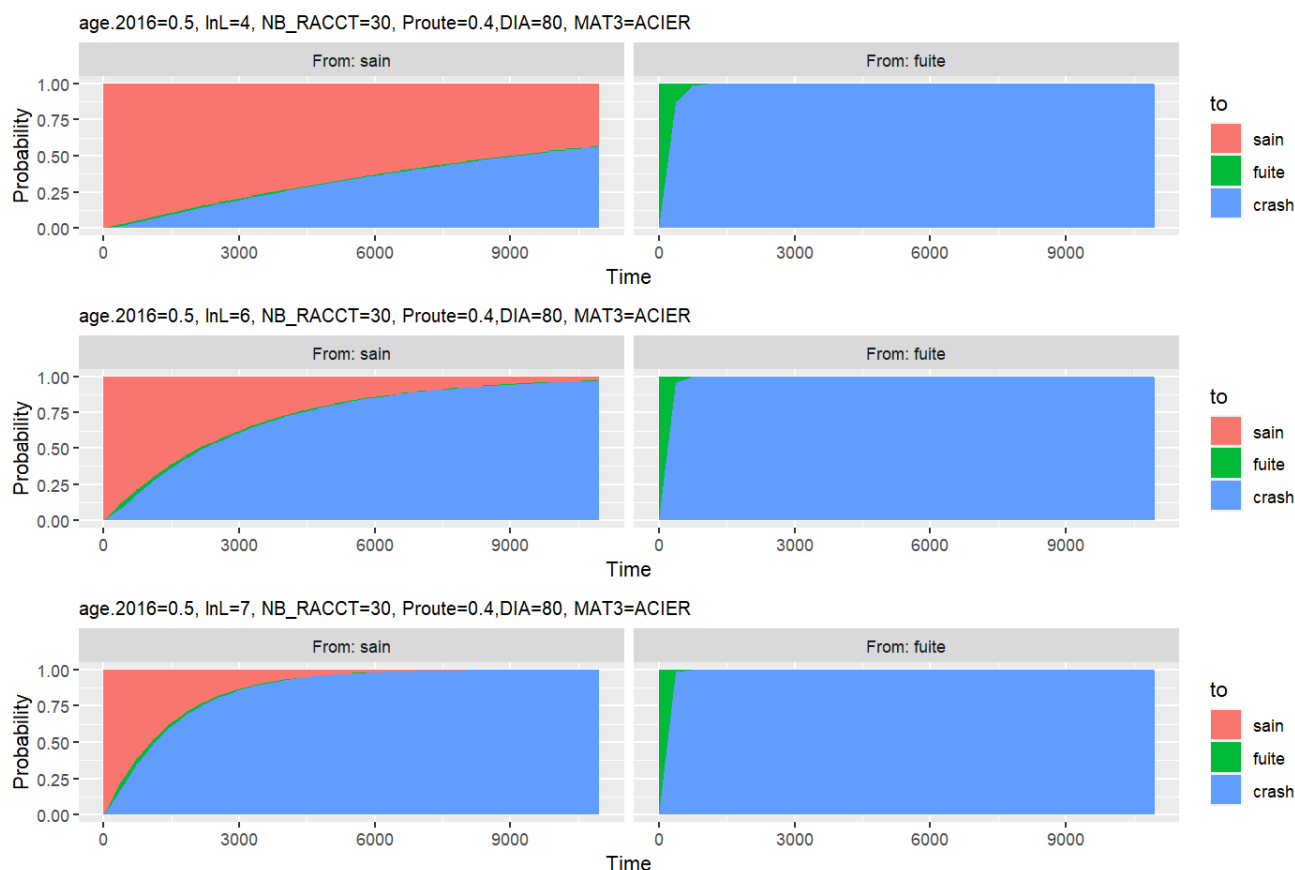


Figure 67. Probabilité de transition entre les états en fonction du temps pour un tronçon en acier de caractéristiques autres connues.

9.3.7.6 Conclusion et perspectives

- Le modèle fourni (et ses concurrents, résumé dans l'Annexe T09_02) a été estimé par le package `msm` (version 1.7) de R (4.2.2). Il donne des moyens intéressants d'interpréter les effets des caractéristiques des tronçons sur les durées avant apparition d'une fuite, puis d'un crash, à partir de données qui ne sont à la base ni adaptées, ni faites pour un traitement statistique. La grande quantité de données permet d'estimer des phénomènes invisibles (apparition d'une fuite), ainsi que des probabilités d'erreur de classification pendant la RF.
- Un effort sur la préparation des données pourrait augmenter la confiance dans l'interprétation des résultats, le traitement des doubles crashes mérite à lui seul une validation par des experts métier.
- Le modèle suppose des intensités qui ne varient pas dans le temps (elles peuvent dépendre de l'âge mais n'ont pas d'effet saisonnier).
- Le modèle est long à caler, l'ajustement a été fait sur une portion des données de la SWDE, on peut imaginer faire un modèle par zone fonctionnelle, ou pour l'ensemble des ZF. De même, la variable matériau dans le modèle peut en être extraite pour ajuster un modèle par matériau (si données en nombre suffisant).
- Les effets des variables ont été estimés sans interaction, les modèles avec interaction peuvent fournir des paramètres spécifiques (de `Proute` par exemple) pour chaque valeur d'une autre variable (matériau, longueur). Ceci est possible mais très chronophage. Toute nouvelle exploration de sélection de variables et recalage pourrait être supportée par un calcul sur serveur de calcul adapté.

- L'interprétation et une modélisation plus poussée des matrices d'erreur de classement peut être envisagée. Actuellement, on retient seulement que le coefficient estimé de 14% de mauvais classement permet un bon ajustement du modèle (et des autres paramètres), mais dire que 14% des tronçons inspectés auraient pu être la source d'une découverte de fuite est erroné. Ce chiffre peut être comparé aux probabilités moyennes de fuite tirée de la tâche 07.

Archive Tâche 09 :

- Dossier livrésT09 : Code, données et résultats du modèle multi état markovien
- Transition par matériaux : fichier reprenant le Tableau 23

10 Tâche 10 – Outil multicritère de programmation court terme

10.1 Objectif de la tâche et décisions stratégiques

Cette tâche vise à fournir à la SWDE une méthode de hiérarchisation (classement) des tronçons de canalisation pour le renouvellement dans le court terme (dans les 5 prochaines années environ), selon plusieurs critères de performance (ex : continuité, pertes) et d'opportunité (ex : coûts, coordination) pris simultanément.

Les premières réunions de cadrage de fin 2022 à début 2023 ont permis de comprendre le fonctionnement de l'ancien outil multicritère (utilisé depuis 2017, voir Figure 68) et de bien définir les attentes de la SWDE envers le nouvel outil :

- Pas d'outil « boîte noire », la méthode doit être facile à comprendre et expliquer aux agents, et le rang d'un tronçon dans le classement doit être facile à comprendre grâce à des statistiques sur son score pour les différents critères ;
- Tous les tronçons doivent faire partie du classement, contrairement à l'ancien outil pour lequel les tronçons classés sont seulement ceux remontés par les agents ;
- Deux hiérarchies parallèles : une pour les conduites de distribution, et une pour les conduites « critiques », c'est-à-dire les adductions ;
- Contrairement à l'outil précédent, pas de distinction des types d'investissements (ex : travaux urgents, travaux pour amélioration de service vs rendement) et de répartition cible de l'enveloppe budgétaire ;
- Centralisation de l'outil au « Processus Transformation et Données » de la SWDE (Couillet).

	A) Enveloppes des "travaux urgents et divers"	B) Travaux pour amélioration rendement	C) Travaux pour amélioration service	D) Travaux pour renouvellement de vannes, BI, raccords, vétustes,...
CRITERE				
AGE/NATURE CONDUITE		20%	20%	
RENDEMENT		35%	0%	
TVX TIERS		0%	0%	
Diamètre (criticité)	OUTIL PAS APPLICABLE	0%	25%	OUTIL PAS APPLICABLE
Qualité d'eau		0%	25%	
Coût		15%	10%	
NB fuites		20%	20%	
Montant fuite		10%	0%	
		100%	100%	

	A) Enveloppes des "travaux urgents et divers"	B) Travaux pour amélioration rendement	C) Travaux pour amélioration service	D) Travaux pour renouvellement de vannes, BI, raccords, vétustes,...
CRITERE				
COTATION PRECEDENTE RAMENEE A 85 %		85%	85%	
Travaux d'époutage (conduite en voirie)		15%	15%	
Travaux en voirie (nouveau coffre, rond-point), pose de gaz	OUTIL PAS APPLICABLE	12%	12%	OUTIL PAS APPLICABLE
Travaux de Nouveaux trottoirs		8%	8%	
Travaux de surfacage, FE, îlot directionnel		4%	4%	
COTATION FINALE		100%	100%	

Figure 68. Outil multicritère SWDE en place depuis 2017.

Les avancées de la tâche ont été suivies grâce à la mise en place d'un groupe de travail spécifique SWDE-INRAE qui s'est réuni 7 fois de mai à juin 2023. Le recrutement d'un stagiaire par INRAE prévu pour fournir un appui sur ce travail pour une période de 6 mois en 2023 n'a pas pu avoir lieu par manque de candidatures valables.

Deux décisions importantes ont été prises par le groupe de travail pendant l'avancée du développement de l'outil :

- Le calcul du score final d'un tronçon dans la hiérarchie en fonction des différents critères se fera par une somme pondérée de ses scores (ses indicateurs) normalisés pour chacun des critères (au détriment des méthodes plus complexes comme la famille Electre) ;

- La coordination des travaux avec les gestionnaires de voirie (GDV) ou autres gestionnaires de câbles ou conduites (GDC) n’aura qu’un caractère informatif (indicateur binaire ou à valeurs discrètes) dans l’outil et ne devra pas intervenir directement sur le classement des tronçons.

Enfin, seuls les tronçons en service d’au moins 6 mètres de longueur ont été inclus dans le classement multicritère. En effet, l’inclusion dans les chantiers de renouvellement des tronçons de moins de 6 mètres peut se décider a posteriori du classement suivant la localisation de ces derniers par rapport aux tronçons longs.

L’outil développé dans ce projet est sous la forme d’un code R.

10.2 Objectifs, critères, et indicateurs de performance et d’opportunité

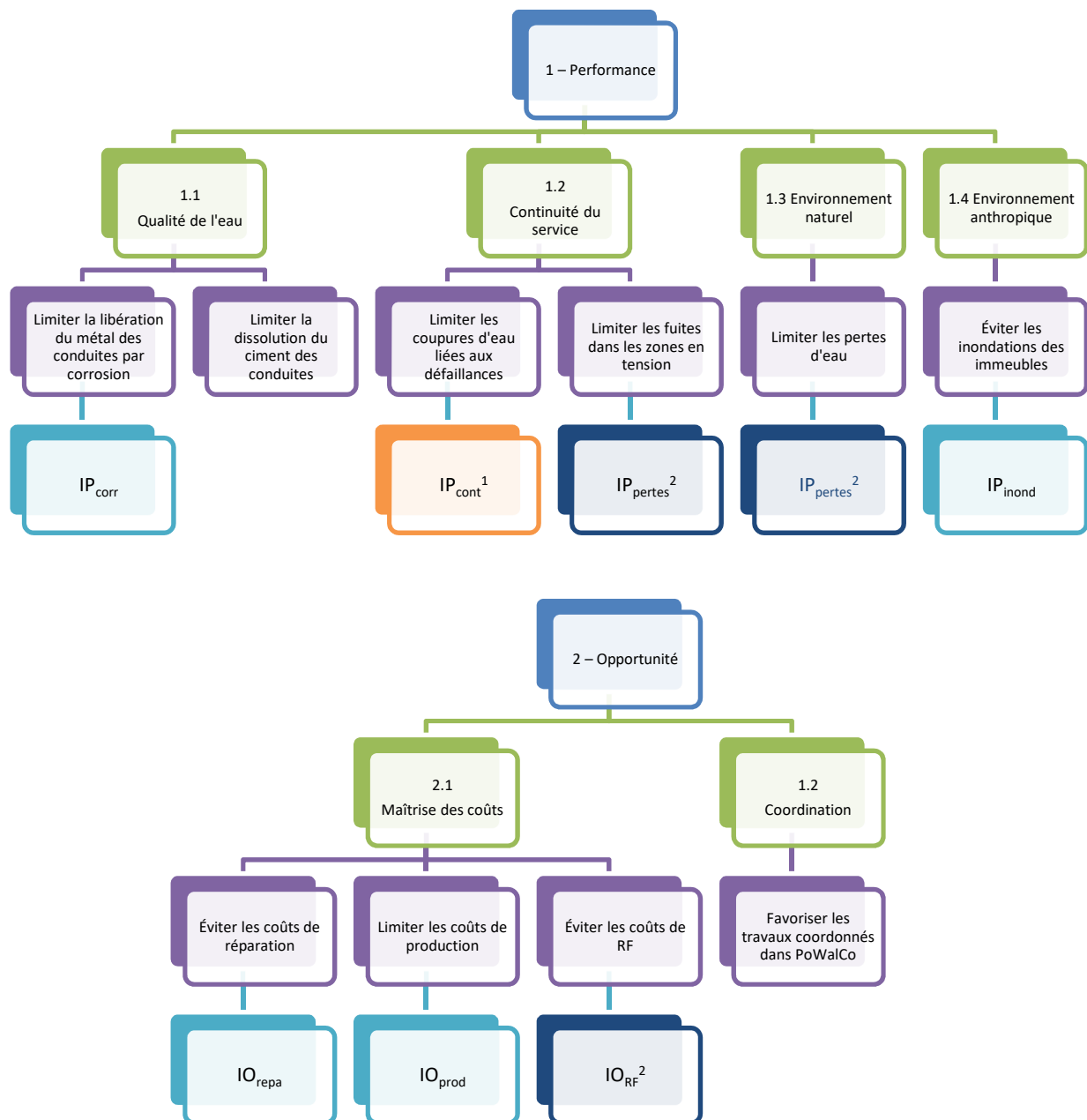


Figure 69. Objectifs, critères, et indicateurs retenus pour l’outil. ¹Formules différentes entre distribution et adduction. ²Indicateur applicable seulement à la distribution.

Les remarques suivantes peuvent être faites :

- On distingue deux catégories d'objectifs, critères, et indicateurs selon qu'il s'agit de la performance ou de l'opportunité. 6 critères de performance associés à 5 indicateurs et 4 critères d'opportunité associés à 3 indicateurs ont été retenus (Figure 69). Ces choix ont été fait en concertation avec la SWDE en prenant en compte les données disponibles.
- Le critère visant à limiter la dissolution du ciment des conduites, et par conséquent la présence de fibres d'amiantes pour les conduites en amiante ciment, comme pour la coordination, ne possède pas un indicateur continu intervenant dans le calcul du score final des tronçons. Il s'agit uniquement d'un indicateur binaire à caractère informatif (0/1, voir 10.3.2).
- Il existe deux indicateurs distincts associés à la continuité selon qu'il s'agit de conduites d'adduction (IP_{cont-a}) ou de distribution (IP_{cont-d}).
- IP_{pertes} intervient à la fois pour la continuité du service pour les zones en tension mais aussi pour la limitation des pertes d'eau en général sur l'ensemble du réseau (voir 10.3.1). Cet indicateur n'est calculable que pour les conduites de distribution avec les données actuelles (peut évoluer dans le futur).
- IO_{RF} n'est calculable que pour les conduites de distribution avec les données actuelles (peut évoluer dans le futur).
- Pour l'opportunité, le critère supplémentaire visant à limiter l'obsolescence du réseau (éradication des conduites intrinsèquement problématiques, remplacement des conduites sous-dimensionnées) n'est pas pris en compte car les données actuelles de la SWDE ne permettent pas de calculer des indicateurs adéquats.

10.3 Calcul des indicateurs

Deux types d'indicateurs sont considérés : les indicateurs qui varient de façon continue et qui interviennent dans le calcul du score final d'un tronçon, et les indicateurs à valeurs discrètes qui n'ont qu'un caractère informatif et n'interviennent pas dans le classement.

Notons qu'une dernière complétion de données a été effectuée dans le but de n'avoir aucune valeur manquante des indicateurs servant à calculer le score des tronçons (détails en Annexe T01_01). Il s'agit de :

1. Données de base MAT, DIA, DDP non complétées par les méthodes précédentes (Rodriguez et al., 2021) et essentielles pour le calcul des variables et indicateurs ;
2. Taux de casses pour les tronçons exclus de Casses (pour le point (1) notamment) ;
3. Probabilité d'état fuyard des tronçons exclus du modèle ZIB (soit pour le point (1), soit car jamais inspectés par la RF).

10.3.1 Indicateurs continus utilisés pour le rang des tronçons

Tableau 28. Synthèse des indicateurs continus et des principaux facteurs les influençant.

Indicateur	Catégorie (agrégé avec)	Applicable à (A=adduction, D=distribution)	Augmente avec
IP_{corr}	Performance	A+D	Corrosivité de l'eau Âge des conduites Diamètre <u>décroissant</u> Nombre de logements raccordés à la conduite
IP_{cont-a}	Performance	A	Taux de casses Diamètre
IP_{cont-d}	Performance	D	Taux de casses Nombre de logements raccordés à la conduite Densité d'abonnés de la cellule Longueur de conduites entre vannes de sectionnements de la cellule Proportion d'abonnés commerciaux ou industriels de la cellule Nombre de clients critiques de la cellule Proportion de linéaire de la cellule en tension (restrictions d'usages)
IP_{pertes}	Performance	D	Probabilité du tronçon d'être fuyard Écart de l'ILP à sa référence pour la cellule Dérive de l'ILP de la cellule dans le temps Tension pour l'alimentation en eau potable de la cellule
IP_{inond}	Performance	A+D	Taux de casses Proximité conduite-immeubles Urbanisation ou industrialisation des environs de la conduite
IO_{repa}	Opportunité	A+D	Taux de casses Coût moyen de réparation de ce type de conduite
IO_{prod}	Opportunité	A+D	Nécessité d'acheter de l'eau ou de l'amener par camions pour la ZDE
IO_{RF}	Opportunité	D	Nombre d'heures de RF nécessaires par km de conduites pour cette cellule Dérive de l'ILP de la cellule dans le temps

Ces indicateurs sont formulés dans la philosophie d'avoir dans la mesure du possible le produit d'un terme d'aléa avec un terme de conséquence, éventuellement multiplié par un dernier terme de gravité. Par exemple, pour IP_{cont-d} , l'idée sous-jacente est de minimiser le nombre de casses ramené au km (aléa) multiplié par le volume non distribué pendant les réparations (conséquence), multiplié par le nombre ou la proportion de clients importants, sensibles, ou critiques (gravité).

Le calcul de ces indicateurs repose sur des hypothèses, des nouvelles données nécessaires (Tableau 29), et des variables calculées qui sont reprises en détail dans une annexe séparée de ce rapport

(Annexe T10_01). Cette annexe précise également les améliorations possibles du calcul de ces indicateurs lorsque de nouvelles données seront disponibles.

Tableau 29. Nouvelles données demandées et reçues pour le développement de l'outil multicritère.

Donnée nécessaire	Correspondant SWDE	Date de transmission
Coûts de réparation précédents des conduites	Gaëtan Jacques	17/05/2023
Emprise des chantiers PoWalCo	Jean-Guy Baudoin	15/06/2023
Abonnés critiques de chaque cellule	Valentin Pierson	16/06/2023
Liste définitive des communes SWDE en tension entre 2017 et 2022	Thierry Goffin	03/07/2023
Indice AI de risque de dissolution du ciment	Sabine Denooz	12/07/2023
Débits de cellule en 2022 avec précision 0.01 m ³	Guy Deschuyteneer (via Olivier Hautier)	Non transmis (transmis le 19/04/2023 avec précision 1 m ³ insuffisante)

Le Tableau 30 détaille les variables calculées intervenant dans les indicateurs. Il ne reprend pas les variables évidentes comme le matériau, le diamètre des conduites (etc.), ni les variables dont le calcul est détaillé dans les rapports intermédiaires précédents.

Tableau 30. Variables utilisées dans les indicateurs.

Symbole	Code variable	IP _{cont-d}	IP _{perdes}	IP _{corr}	IP _{cont-a}	IP _{inond}	IO _{repa}	IO _{prod}	IO _{RF}
$TxCP$	VP01								
N_{logt}	VP02								
$p(Y=1)$	VP03								
ILP	VP04								
ILP_{ref0}	VP05								
ΔILP	VP06								
P	VP07								
T	VP08								
V_{corr}	VP09								
CLC	VP10								
C_{repa}	VO01								

10.3.2 Indicateurs discrets à caractère informatif

En parallèle des indicateurs continus qui sont utilisés pour le classement, trois indicateurs discrets sont ajoutés à titre informatif :

- Conduite de distribution critique car en entrée/sortie de cellule (0/1) ;
- Conduite présentant un risque de libération de fibres d'amiante ciment (0/1) ;
- Conduite dans un ou plusieurs chantiers PoWalCo ($n \in \mathbb{N}$ chantiers).

Les conduites de distribution qui ont déjà cassé de façon certaine (défaillance connue du SIG) entre 2011 et 2022 inclus, qui font au moins 6 mètres, et qui portent un compteur de sectorisation ont été identifiées dans le classement (indicateur=1) si elles ont au moins 0.5 casses prévues dans les 5 prochaines années. En effet, ces conduites sont des entrées ou sorties de cellule, donc leurs défaillances ont le potentiel de priver d'eau une ou plusieurs cellules entières pendant les réparations.

La dissolution possible du ciment des conduites et la libération potentielle de fibres d'amiante est basée sur la valeur de l'indice AI fourni par la SWDE à l'échelle des ZDE :

$$AI = pH + \log_{10}(A \times H)$$

où A est l'alcalinité (TAC) exprimée en mg/L de CaCO₃, et H est la dureté calcique exprimée en mg/L de CaCO₃. Notons que les valeurs de H marquées comme < 3 mg/L ont été remplacées par la valeur 1 dans le calcul afin de ne pas avoir de valeur manquante pour les ZDE concernées (6 ZDE concernées : 1078, 16674, 16675, 17835, 17836, 16995). La procédure suivante a été utilisée pour définir un indicateur binaire (0 ou 1) reporté auprès de chaque conduite dans le classement :

1. Par défaut, il est supposé que toutes les conduites ne présentent aucun risque (prend en compte les conduites pour lesquelles les données ne sont pas disponibles) -> indicateur = 0 ;
2. Les conduites en amiante ciment exposées à de l'eau dont AI < 12 sont identifiées comme étant à risque -> indicateur =1.

Concernant la coordination, à côté de chaque conduite dans le classement est reporté le nombre de chantiers PoWalCo programmés à partir de 2024 qui concernent la conduite (déterminé par intersection des polygones des chantiers avec les conduites). Les chantiers concernés sont soit programmés avec une date connue à partir de 2024 quelle que soit la date de saisie dans PoWalCo, soit programmés avec une date inconnue mais saisis dans PoWalCo en 2023. Par défaut, la Zone de Périmètre des chantiers est choisie (Figure 70), sauf si seule la Zone de Chantier est disponible (dans ce cas cette dernière est prise).

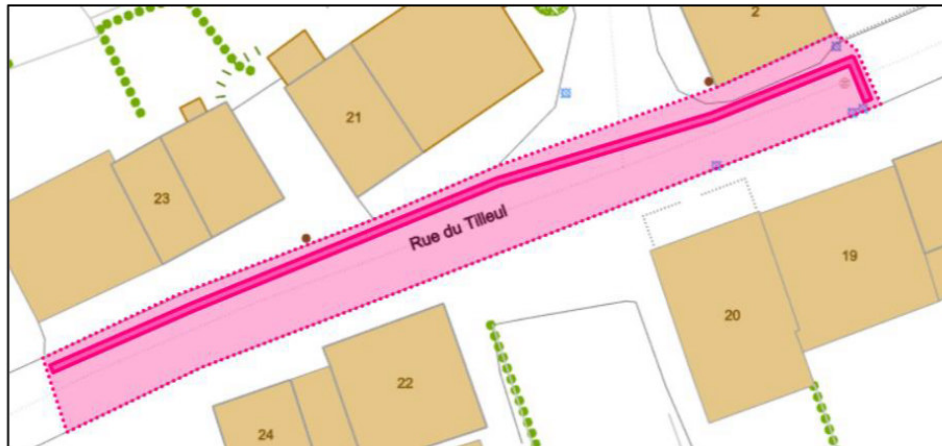


Figure 70. Zone de Chantier (trait épais) et Zone Périmètre (polygone) dans PoWalCo.

10.4 Agrégation des indicateurs

La procédure suivante a été adoptée. Elle utilise la fonction de normalisation $u : x \mapsto \frac{x - \min(x)}{\max(x) - \min(x)}$ entre 0 et 1 pour un vecteur x . Le traitement se fait en parallèle pour l'adduction et la distribution.

1. Calcul d'un score de performance normalisé : $IP = u\left(\sum_{j=1}^M w_j u(IP_j)\right)$, avec $M=3$ pour l'adduction et $M=4$ pour la distribution (voir Tableau 28), où w_j est le poids de l'indicateur de performance IP_j
2. Calcul d'un score d'opportunité normalisé : $IO = u\left(\sum_{k=1}^N w_k u(IO_k)\right)$, avec $N=2$ pour l'adduction et $N=3$ pour la distribution (voir Tableau 28), où w_k est le poids de l'indicateur d'opportunité IO_k
3. Calcul du score final $I = u(w_p IP + w_o IO)$, où w_p est le poids donné à la performance, et w_o le poids donné à l'opportunité.

Les poids w_k sont calculés avec la formule $w_k = \frac{\omega_k p_k}{\sum_j \omega_j p_j}$ en fonction de deux facteurs :

- Les préférences SWDE, exprimées comme une liste de nombres $p_j \geq 0$ dont la somme est strictement positive (pas nécessairement 1). Par exemple, une liste de préférences $\{2 ; 1 ; 0.5\}$ aura tendance à donner deux fois plus de poids au premier critère de la liste qu'au second, et 4 fois plus au premier qu'au dernier ;
- Les poids « équitables » $\omega_j > 0$ (dont la somme vaut 1), qui tiennent compte de la distribution des indicateurs qui sont agrégés en un score. En effet, pour certains indicateurs, environ 1 % du linéaire (voire moins) seulement possède des valeurs non négligeables (ex : valeurs > 0.1), ce qui veut dire que lors de l'agrégation des divers indicateurs, ces derniers n'auront que peu ou pas d'influence dans le score sans une augmentation appropriée de leur poids.

Le calcul des poids équitables ω_k est effectué de manière automatique en considérant simultanément les distributions cumulées selon la longueur des tronçons de tous les indicateurs qui doivent être agrégés (Figure 71), en particulier la valeur $q_{k,99}$ dépassée par seulement 1 % du linéaire pour l'indicateur I_k :

$$\omega_k = \frac{1}{q_{k,99}} \frac{1}{\sum_j \frac{1}{q_{j,99}}}$$

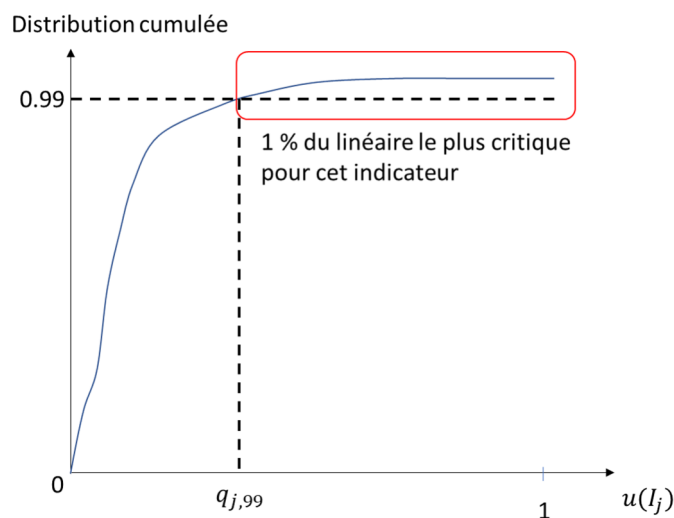


Figure 71. Distribution normalisée cumulée en longueur des tronçons d'un indicateur I_j .

Ainsi, plus un indicateur « chute » rapidement vers 0 pour son 1^{er} % du linéaire le plus critique par rapport aux autres indicateurs, plus q_{99} est faible, et plus son poids équitable ω aura tendance à être élevé pour compenser la chute rapide des valeurs de l'indicateur dans le calcul du score agrégé.

Les préférences utilisées et les poids obtenus sont détaillés dans le Tableau 31. On remarque par exemple que IP_{inond} chute le plus rapidement vers 0, lui donnant le poids équitable le plus élevé. En revanche, les préférences données en exemple, suggérant de le rendre 20 à 40 fois moins important que la continuité, font que son poids final est relativement faible.

Tableau 31. Poids obtenus

Indicateurs	Distribution			Adduction		
	Poids équitables ω_k	Préférences p_k	Poids w_k	Poids équitables ω_k	Préférences p_k	Poids w_k
IP _{corr}	0.066	0.05	0.004	0.246	0.01	0.004
IP _{cont}	0.401	2	0.861	0.306	2	0.961
IP _{pertes}	0.082	1	0.095			
IP _{inond}	0.451	0.10	0.05	0.449	0.05	0.035
IO _{repa}	0.817	2.5	0.887	0.776	2.5	0.896
IO _{prod}	0.107	1	0.0462	0.224	1	0.104
IO _{RF}	0.077	2	0.0666			
Performance	0.586	0.85	0.889	0.554	0.85	0.876
Opportunité	0.414	0.15	0.111	0.446	0.15	0.124

10.5 Format des résultats de l'outil

L'exécution de l'outil donne en résultat deux fichiers CSV (un pour l'adduction et un pour la distribution) ayant la structure suivante (Tableau 32). Dans ce fichier les tronçons sont déjà classés par score final décroissant, et le coût cumulé pour les renouvellements en suivant le classement est affiché. Ces coûts sont calculés grâce à la formule d'estimation de l'ancien outil multicritère SWDE à laquelle une révision des prix a été appliquée (Annexe T10_02). Les résultats sont en annexe séparée (Adduction : Annexe T10_03 et Distribution : Annexe T10_04).

Pour simplifier l'interprétation des scores des tronçons, chaque indicateur et chaque score (agrégation d'indicateurs) est associé à une position dans le classement, relative uniquement à cet indicateur ou à ce score. Par exemple, pour IP_{corr}, « top 2 % » veut dire que le tronçon est dans le top 2 % (entre 1 % et 2 % car sinon il serait top 1 %) du linéaire si on trie les conduites par IP_{corr} décroissant.

Tableau 32. Format des résultats de l'outil multicritère.

Colonne	Signification
IDT	ID de tronçon
pourcentLcumu	% de la longueur cumulée
BudgetCumule_ME	Budget cumulé pour les renouvellements en millions d'euros
score	score global
pos_score	position du tronçon dans le classement par cet indicateur seul
score_perf	score de performance
pos_score_perf	position du tronçon dans le classement par cet indicateur seul
score_opp	score d'opportunité
pos_score_opp	position du tronçon dans le classement par cet indicateur seul
NchantiersPrevus	nombre de chantiers concernant ce tronçon prévus par d'autres opérateurs à partir de 2024
TxCP	taux de casses prédit en 2022 (u/km/an)
proba_fuyard*	probabilité d'être fuyard
Ncasses_2011_2022	nombre de casses subies entre 2011 et 2022
NUM_CEL	numéro de cellule
NOM_SECT	nom du secteur
ZF	nom de la ZF
LNG	longueur en m
DIA	diamètre nominal
MAT	matériau (après regroupement)
DDP	année de pose
IP_corr_norm	indicateur performance de qualité de l'eau - corrosion métaux - normalisé
pos_IP_corr_norm	position du tronçon dans le classement par cet indicateur seul
risque_fibre_amiant e	indicateur 0/1 du risque de libération de fibres d'amiantes dans l'eau
IP_cont_norm	indicateur performance de continuité du service - normalisé
pos_IP_cont_norm	position du tronçon dans le classement par cet indicateur seul
compteur_secto_cas ses*	0/1 : conduite longue (plus de 6 m) porteuse de compteur de sectorisation, ayant déjà cassé de façon certaine et ayant plus de 0.5 casses prévues dans les 5 prochaines années
IP_pertes_norm*	indicateur performance de pertes - normalisé
pos_IP_pertes_norm	position du tronçon dans le classement par cet indicateur seul
IP_inond_norm	indicateur performance d'inondations des immeubles voisins - normalisé
pos_IP_inond_norm	position du tronçon dans le classement par cet indicateur seul
IO_prod_norm	indicateur opportunité de coûts de production - normalisé
pos_IO_prod_norm	position du tronçon dans le classement par cet indicateur seul
IO_repa_norm	indicateur opportunité de coûts de réparation - normalisé
pos_IO_repa_norm	position du tronçon dans le classement par cet indicateur seul
IO_RF_norm*	indicateur opportunité de coûts de RF - normalisé
pos_IO_RF_norm*	position du tronçon dans le classement par cet indicateur seul
methode_TxCP	méthode utilisée pour estimer TxCP
methode_pY	méthode utilisée pour estimer proba_fuyard
methode_pYL	méthode utilisée pour estimer proba_fuyard par unité de longueur
donnees_completees pour_outil	données complétées parmi MAT, DDP, DIA, spécifiquement pour le calcul des indicateurs de l'outil court terme
	*disponible uniquement pour les conduites de distribution

Archive Tâche 10 :

- Dossier court_terme : Données, codes et résultats de l'outil multicritère court terme
- Dossier reunions : compte rendus / présentations faites lors des réunions hebdomadaires pour la Tâche 10 depuis fin 2022

11 Tâche 11 - Efficacité des actions de recherche de fuite

11.1 Objectifs et mise en œuvre

La tâche 11 du projet GePaME s'est attachée à étudier les déterminants de l'efficacité des actions de recherche et de réparation des fuites dans l'objectif de mieux cibler les modes d'intervention en fonction des débits et des caractéristiques des cellules. Pour ce faire, une expérimentation a été conduite de mars à juin 2023 pour comparer les résultats obtenus grâce aux trois méthodes usuelles de pré-localisation (l'écoute systématique sur appareils, le tronçonnage et le déploiement d'enregistreurs acoustiques) sur différents types de cellules.

A partir des données des cellules obtenues dans le cadre de la tâche 5 complétées par les informations d'ILP et de débit nocturne de la semaine du 16 au 22 janvier 2023, 35 cellules ont été sélectionnées sur la base des critères suivants (Tableau 33 et Annexe T11_01) :

- La cellule dispose d'au moins 1000 jours avec des données de débits valides sur la période 2017-2021 (calcul de $Q_{moy17-21}$ fiable)
- La cellule dispose d'au moins 4 jours avec des données de débits valides sur la semaine du 16 au 22 janvier 2023 (calcul de Q_{23} fiable)
- La longueur et le nombre de raccordement n'ont pas varié de plus de 2% entre fin 2021 et début 2023 (cellules stables)
- Entité fonctionnelle = MONS ou NAMUR ou VERVIERS-MARCHE
- $ILP_{23} - ILP_{moy17-21} > 0 \text{ m}^3/\text{km}/\text{j}$ (la cellule a une performance moins bonne que celle constatée historiquement)
- $ILP_{23} - ILP_{ref} > 9 \text{ m}^3/\text{km}/\text{j}$ (la cellule a une performance dégradée par rapport à la valeur de référence)
- $Q_{23} - Q_{ref} > 4 \text{ m}^3/\text{h}$ (le débit de fuite à trouver correspond à au moins une grosse fuite)
- $Q_{23} - Q_{moy17-21} > 2 \text{ m}^3/\text{h}$ (le débit actuel est significativement supérieur à la moyenne historique)

Pour caractériser l'état fuyard actuel des cellules, les ILP et débits nocturnes constatés lors de la semaine du 16 au 22 janvier 2023 (ILP_{23} et Q_{23}) sont comparés aux valeurs de références calculées dans Perf'O (ILP_{ref} et Q_{ref}) et aux valeurs moyennes constatées sur la période 2017-2021 ($ILP_{moy17-21}$ et $Q_{moy17-21}$). Le débit nocturne est le minimum de la moyenne horaire glissante de 1h à 5 h, $Q_{ref} = ILP_{ref} \times L / 24$ et $Q_{23} = ILP_{23} \times L / 24$

Un protocole d'expérimentation a ensuite été établi (Annexe T11_02), les techniques de pré-localisation à privilégier étant attachées aux entités fonctionnelles :

- EF de MONS (responsable RF Malory Urbain) : Déploiement d'enregistreurs acoustiques
- EF de NAMUR (responsable RF Rudi Deckers) : Tronçonnage
- EF de VERVIERS-MARCHE (responsable RF Jérôme Schaus) : Ecoute systématique sur appareils

Pour chaque cellule concernée, une fiche de suivi a été établie pour être complétée au fur et à mesure de l'avancement de l'expérimentation par les responsables RF (Annexe T11_03), des réunions d'échange hebdomadaires se sont tenues entre les responsables RF et les autres personnes concernées au sein de SWDE et INRAE. En pratique, l'expérimentation s'est déroulée de mi-mars à fin juin 2023. L'évolution des actions mises en œuvre et des résultats obtenus sur les cellules a été suivie

à partir d'une part, des fiches actualisées par les responsables RF et, d'autre part, des données SAP pertinentes (fournies par Gaëtan Jaques) et des données de débit issues de Perf'O (fournies par Guy Deschuyteneer).

Tableau 33. Répartition et caractéristiques cumulées par secteur des cellules sélectionnées. Qhisto et Q0.01 sont respectivement le débit nocturne moyen et quantile 1% des débits nocturnes sur la période 2017 - 2021

Secteur	Verviers	Mons	Namur	Total
Responsable	Jérôme	Malory	Rudi	
Tech Préloc	Ecoute	Enregistreurs	Tronçonnage	
Nb Cell	11	10	14	35
L (km)	85.3	137.9	138.2	361.5
NbRacc	4 716	7 995	7 538	20 249
Qref (m³/h)	15.9	26.9	25.9	68.8
Q0.01 (m³/h)	27.9	38.5	32.0	98.5
Qhisto (m³/h)	61.0	71.9	68.8	201.7

En pratique, seules 26 des 35 cellules ont fait l'objet d'investigations valorisables (Tableau 34) :

- 2 cellules avaient en mars des débits proches de l'objectif du fait de réparations réalisées avant le début de l'expérimentation (la sélection des cellules était assise sur des débits de janvier) ;
- 1 cellule qui comporte un gros consommateur nocturne qui explique l'écart par rapport au débit de référence ;
- 1 cellule présente des variations erratiques du débit de nuit en l'absence d'interventions ;
- 1 cellule présente des valeurs de débit de nuit aberrantes du fait d'un problème d'échelle sur les comptages d'entrée et de sortie d'un réservoir ;
- 4 cellules du secteur de Mons n'ont pas pu être explorées ou exploitées du fait de contingences matérielles

A noter que, sur le secteur de Mons, il est apparu que les enregistreurs utilisés ne sont pas adaptés à un usage mobile (difficultés de paramétrages), ainsi, sur les cellules de ce secteur de , les informations recueillies sont incomplètes. Elles n'ont donc pas été prises en compte pour une partie de l'exploitation des résultats.

Tableau 34. Répartition et caractéristiques cumulées par secteur des cellules effectivement étudiées

Secteur	Verviers	Mons	Namur	Total
Nb Cell	9	5	12	26
L (km)	65.3	57.9	121.4	244.7
NbRacc	3 642	3 679	6 272	13 593
Qref (m³/h)	12.2	12.1	21.9	46.2
Q0.01 (m³/h)	20.7	21.7	29.7	72.0
Qhisto (m³/h)	47.8	34.2	56.2	138.2

Avertissement : L'échantillon des cellules étudiées étant restreint, les résultats présentés dans la suite n'ont pas de valeur statistique et ne produisent donc pas des références ou des ratios qui peuvent être généralisés. Ils permettent néanmoins de comparer les effets des différentes techniques utilisées dans

le contexte des cellules étudiées, et ainsi constituer une base de réflexion pour améliorer à l'avenir le recueil des informations utiles pour adapter les actions aux objectifs poursuivis.

11.2 Résultats par secteur et par technique

11.2.1 Pré-localisation

En premier lieu, il doit être noté que bien qu'à chaque secteur était assignée une technique de pré-localisation prédéfinie, en pratique, les opérateurs ont été conduit à utiliser d'autres méthodes, notamment quand la méthode initialement prévue ne pouvait pas facilement être mise en œuvre (problèmes d'accès aux vannes en zone urbanisée, organes déficients, etc.) ou quand la mise en œuvre de la technique prédéfinie n'apportait pas de résultats. Le Tableau 35 présente la répartition des 61 campagnes réalisées par type et par secteur. On constate, que pour le secteur de Verviers l'écoute systématique a bien été privilégiée mais qu'il a été recouru également au tronçonnage (2 cas) et aux enregistreurs (1 cas), et que pour le secteur de Namur qui privilégiait le tronçonnage, des enregistreurs ont été utilisés à 8 reprises.

Tableau 35. Campagnes de pré-localisation par type et par secteur

Type Pré-localisation	Verviers	Mons	Namur	Total
Contrôles	1		2	3
Ecoute systématique	21			21
Enregistreurs	1	7	8	16
Tronçonnage	2		19	21
Total général	25	7	29	61

Le Tableau 36 récapitule, par type de pré-localisation les cumuls des temps d'inspection (en heures/hommes) et de longueurs inspectées. Il apparaît que le temps moyen par kilomètre inspecté est similaire pour l'écoute systématique et le tronçonnage, tandis qu'il est sensiblement supérieur pour les enregistreurs. La Figure 72 montre que la dépendance entre le temps passé et la longueur inspectée pour chaque cellule, bien que tendanciellement croissante, est très dispersée (pour une même longueur le temps peut varier de 1 à 10). Cela est moins vrai pour la pose d'enregistreurs, technique pour laquelle il y a une proportionnalité marquée entre le temps passé et le linéaire inspecté.

Tableau 36. Temps et longueurs inspectées par type de pré-localisation

Type Prélocalisation	Nb Camp	Temps (h)	L_insp(Km)	T/L (h/km)
Contrôles	3	14.0	2.0	
Ecoute systématique	21	163.0	118.3	1.4
Enregistreurs	16	215.0	110.7	1.9
Tronçonnage	21	227.0	155.0	1.5
Total général	61	619.0	386.0	1.6

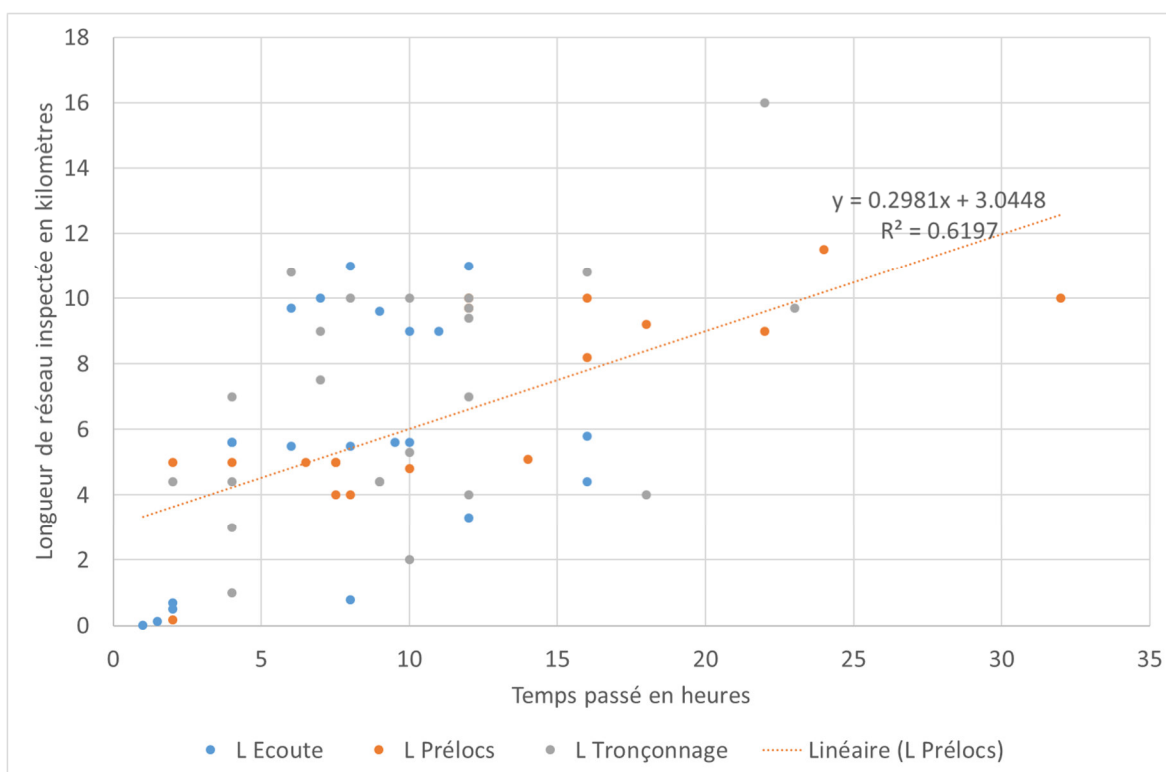


Figure 72. Longueur inspectée des cellules en fonction du temps passé selon la technique

11.2.2 Localisation

Pour la localisation des fuites, trois familles de techniques sont utilisées, la corrélation acoustique, l'écoute (au sol et/ou sur appareil) et l'injection de gaz traceur (avec écoute et/ou détection). La corrélation acoustique est la technique la plus largement utilisée (plus de 6 fois sur 10), l'emploi du gaz traceur est réservée aux cas où les autres techniques ont échoué (Tableau 37). Cela est cohérent avec les temps moyens de localisation constatés en fonction des techniques (Tableau 38). En effet, tandis que les temps moyens de localisation par corrélation ou par écoute sont similaires (de l'ordre de 3 heures par fuite), la localisation par gaz nécessite plus du double (6h30).

Tableau 37. Techniques de localisation des fuites par secteur

Technique	Verviers	Mons	Namur	Total	%
Corrélation acoustique	14	5	17	36	63%
Ecoute	6		10	16	28%
Gaz traceur	3		2	5	9%
Total général	23	5	29	57	100%

Tableau 38. Durées de localisation selon la technique utilisée

Technique	Nombre	Temps (h)	T/Nb (h)
Corrélation	36	110.5	3.1
Ecoute	16	50.5	3.2
Gaz	5	32.5	6.5
Total général	57	193.5	3.4

11.2.3 Réparation

Les réparations ici prises en compte concernent une période plus importante que celle dédiée à l'expérimentation (de janvier à la date de fin de l'expérimentation, qui varie selon les cellules étudiées).

Les réparations concernent en premier lieu les raccordements, leur nombre pour les conduites, un peu plus faible, reste dans des proportions voisines. Les trous pour rien (ou TPR, travaux engagés suite à une localisation qui n'aboutissent pas à la réparation d'une fuite) concernent 1 chantier sur 6 (Tableau 39).

Tableau 39. Réparations par secteur et par type d'organe

Organe	Verviers	Mons	Namur	Total	%
Conduite	14	1	10	25	27%
Raccordement	11	3	15	29	31%
BI/PI	1	3	11	15	16%
Autre	1	1	4	5	5%
Inconnu	2	1	1	4	4%
TPR	4	2	10	16	17%
Total général	33	10	51	94	100%

Les temps moyens de réparation sont très peu liés à l'organe réparé ou au type de défaillance (y compris TPR) en restant toujours voisin de 8 heures par réparation (ces moyennes cachent des valeurs très différentes allant de moins de deux heures à plus de 24 heures dans les cas extrêmes) (Tableau 40).

Tableau 40. Temps de réparations par organe puis par type de défaillance

Organe	NbRep	TpsRep	Tps/repa
Conduite	25	190	7.6
Raccordement	29	260	9.0
BI/PI	15	136	9.0
Autre	5	27	5.3
Inconnu	4	27	6.7
TPR	16	134	8.3
Type de défaillance			
TPR	16	134	8.3
Crash	25	193	7.7
Fuite	53	445	8.4
Total général	94	771	8.2

Le suivi des débits nocturnes des cellules a été réalisé en calculant la moyenne glissante sur 7 jours (de j-6 à j inclus). Ce choix a été fait pour limiter l'impact des variations de la consommation nocturne des usagers qui généralement subit un « effet weekend ». L'effet d'une réparation sur le débit nocturne de la cellule, c'est à dire le débit de la fuite réparée, estimé par différence entre la moyenne glissante calculée le jour j de la réparation (le débit nocturne du jour j, mesuré entre 1h et 5h, n'est pas affecté par la réparation) et celle calculée le jour j+7 (tous les débits intervenant dans le calcul de la moyenne

glissante sont postérieurs à la réparation). Les résultats moyens ainsi obtenus par organe et par type de défaillance sont présentés dans le Tableau 41. Ceux-ci montrent que le débit moyen gagné par réparation est de l'ordre de 1 m³/h et que, tous organes confondus, il est sensiblement équivalent pour les deux types de défaillances (crashes et fuites trouvées par la RF). On note, sans que cela puisse être généralisé au-delà de notre terrain d'étude, que les crashes sur conduite permettent un gain deux fois plus faible que la moyenne, tandis qu'à l'inverse, les crashes sur raccordement occasionnent un gain supérieur à la moyenne générale.

Tableau 41. Débits moyens gagnés par les réparations selon l'organe réparé et le type de défaillances (m³/h)

Organe	NbFui	NbCra	QgainFui	QgainCra	Qgain/fui	Qgain/Cra	Qgain/Repa
Conduite	18	6	18.9	2.9	1.05	0.48	0.91
Raccordement	16	13	17.1	17.0	1.07	1.31	1.18
BI/PI	11	3	7.3	0.4	0.67	0.15	0.56
Autre	3	2	0.8	0.3	0.47	0.27	0.22
Inconnu	3	1	4.8	1.2	1.61	1.20	1.51
Total général	51	25	49.1	21.8	0.96	0.87	0.93

Pour lier le gain en débits des réparations à la technique de pré-localisation, il a été choisi d'exclure les cellules du secteur de Mons qui, comme précisé plus haut, ont été inspectées avec des enregistreurs inappropriés qui ne sont pas représentatif des pratiques usuelles (Tableau 42). Les résultats montrent que les fuites trouvées suite à une pré-localisation par enregistreurs occasionnent en moyenne une réduction des pertes deux fois plus faible que celle constatées pour les fuites trouvées suite à une pré-localisation par écoute systématique ou par tronçonnage (ces deux dernières techniques produisant des résultats identiques). Deux raisons concomitantes semblent pouvoir expliquer ce résultat, en premier lieu, sur les secteurs de Namur et Verviers, les enregistreurs n'ont été déployés que lorsque les autres techniques étaient inappropriées ou infructueuses, en second lieu, les enregistreurs fournissent généralement plus de points de suspicions de fuite que les autres techniques.

Tableau 42. Débits moyens gagnés par les réparations selon le type de pré-localisation (hors Mons) (m³/h)

Type campagne	NbRep	Qgain	Qgain/Repa
Crash	24	21.5	0.90
Ecoute systématique	13	14.8	1.14
Enregistreurs	14	8.9	0.63
Tronçonnage	19	21.6	1.14
Total général	70	66.8	0.95

11.2.4 Débits

L'analyse des débits des cellules en début et en fin d'expérimentation permet d'estimer l'efficacité des actions mises en œuvre pour réduire les pertes. Le débit cible Q_{cible} (m³/h) est défini comme le maximum des deux valeurs Q_{ref} (débit de référence Perf'O) et $Q_{0.01}$ (valeur en dessous de laquelle le débit de la cellule (moyenne glissante sur 7 jours) a été inférieur pendant 1% du temps sur la période 2017-2021). Cette façon de procéder permet de viser un objectif plus réaliste lorsque Q_{ref} n'a historiquement jamais été atteint (ce qui est par exemple le cas lorsque la cellule connaît des

consommations nocturnes récurrentes importantes). Le Tableau 43 présente de façon globale et à l'échelle des secteurs (hors Mons) le bilan des débits et des *ILP* au cours de l'expérimentation.

Tableau 43. Bilan cumulé des débits (m³/h) et des *ILP* (m³/j.km) hors secteur de Mons

Secteur	Verviers	Namur	Total
<i>Nb Cell</i>	9	12	21
<i>L Cell</i>	65.3	121.4	186.7
<i>Qcible</i>	21.8	31.5	53.3
<i>Qdebut</i>	63.1	77.8	140.9
<i>Qfin</i>	38.3	46.7	84.9
<i>Qdebut - Qcible</i>	41.3	46.3	87.6
<i>GainQ (Qdebut – Qfin)</i>	24.9	31.1	56.0
<i>Qfin - Qcible</i>	16.5	15.2	31.6
<i>GainQ/(Qdebut - Qcible)</i>	60%	67%	64%
<i>GainQrep</i>	32.9	31.4	64.2
<i>GainQ-GainQrep</i>	-8.0	-0.2	-8.3
<i>GainQfui</i>	16.8	28.5	45.3
<i>GainQCra</i>	16.1	2.9	19.0
<i>GainQfui/GainRep</i>	51%	91%	70%
<i>GainQfui/(Qdebut - Qcible)</i>	41%	62%	52%
<i>ILPcible</i>	8.0	6.2	6.8
<i>ILPdebut</i>	23.2	15.4	18.1
<i>ILPfin</i>	14.0	9.2	10.9
<i>ILPdebut - ILPcible</i>	15.2	9.2	11.3
<i>GainILP (ILPdebut – ILPfin)</i>	9.1	6.2	7.2
<i>ILPfin - ILPcible</i>	6.1	3.0	4.1

Pour les 21 cellules traitées sur les secteurs de Namur et Verviers, l'expérimentation a permis de réduire de 64% le débit de perte estimé (56 m³/h) ce qui représente une baisse de l'*ILP* moyen de 7.2 m³/j.km (réduction de 18.1 à 10.9). En fin d'expérimentation le débit de fuite résiduel global est de 31.3 m³/h ce qui correspond à un potentiel résiduel de réduction de l'*ILP* de 4.1 m³/j.km. Les taux de réduction des pertes pour chacun des secteurs sont comparables (60% pour Verviers et 67% pour Namur), néanmoins la réduction est d'avantage due aux fuites trouvées par la RF à Namur qu'à Verviers (respectivement 62% et 41%). Ceci est en parti expliqué par une apparition de nouvelles fuites sur la période d'expérimentation plus importante sur le secteur de Verviers que sur celui de Namur. En effet, la comparaison du débit cumulé gagnés par les réparations et de l'amélioration du débit sur la période d'expérimentation montre que sur le secteur de Verviers 8 m³/h gagnés par la réparation des fuites ont été reperdus par l'apparition de nouvelles fuites tandis que sur le secteur de Namur, ce phénomène ne représente que 0.2 m³/h.

11.3 Résultats par cellules (secteurs de Namur et Verviers)

11.3.1 Données des cellules retenues

Le Tableau 44 fournit un extrait des principales données associées aux cellules retenues. Le type de campagne (*TypCam*) est défini comme suit :

- « Ecoute » si toutes les campagnes réalisées sur la cellule sont de l'écoute systématique
- « Tronçonnage » si toutes les campagnes réalisées sur la cellule sont du tronçonnage
- « Mixte » si plusieurs techniques ont été utilisées. Cela concerne 6 cellules, pour 5 d'entre elles le tronçonnage est associé à la pose d'enregistreurs, pour la dernière les trois techniques de pré-localisation ont été utilisées

Il est à noter que dans cette analyse, les campagnes de type « contrôles » (vannes ou débitmètres) sont exclues.

Tableau 44. Extrait des données recueillies sur les cellules retenues

Cellule	Sect	L	D	NbCam	TypCam	TpsCam	KmCam	NbLoc	NbFui	GainQfui	Qcible	Qdeb	Qfin
53100018	NAM	5.3	39	3	Mixte	16.0	6.5	1	1	0.1	2.5	3.4	2.9
53100055	NAM	16.3	29	2	Tronçonnage	17.0	19.0	2	4	8.0	2.2	3.9	0.5
55103071	VERV	5.6	76	5	Ecoute	29.5	22.9	3	3	2.2	2.4	6.2	3.7
55103239	VERV	3.3	95	2	Ecoute	13.5	3.4	1	2	5.6	0.9	9.4	2.9
55103244	VERV	1.9	179	2	Ecoute	3.0	0.7	1	2	4.0	1.2	7.4	3.6
55103266	VERV	5.8	55	3	Ecoute	40.0	11.0	3	1	2.3	1.1	6.8	2.4
56102038	NAM	7.5	39	1	Tronçonnage	7.0	7.5	1	1	3.6	1.4	5.8	2.2
56102039	NAM	9.7	53	2	Mixte	35.0	19.7	2	4	0.3	1.8	4.7	4.6
56102103	NAM	9.7	62	2	Mixte	24.0	19.4	1	2	0.8	2.0	6.2	6.6
56102115	NAM	13.9	53	6	Mixte	63.0	45.0	7	7	7.7	2.5	12.0	4.5
56102118	NAM	9.4	43	1	Tronçonnage	12.0	9.4	1	1	0.9	1.6	5.2	2.0
56104359	NAM	9.3	31	1	Tronçonnage	12.0	7.0	5	2	0.3	1.6	5.1	2.4
56105075	NAM	10.8	16	3	Tronçonnage	20.0	19.8	3	4	1.5	1.8	5.9	4.4
57100010	NAM	9.2	45	2	Tronçonnage	18.0	8.8	2	1	0.2	2.9	4.3	3.9
57100055	NAM	16.3	101	2	Tronçonnage	12.0	13.0	3	3	5.1	9.3	16.9	9.6
57100081	NAM	4.0	134	2	Mixte	20.0	8.0	1	1	0.0	1.9	4.2	3.1
58205203	VERV	11.0	46	1	Ecoute	10.0	9.0	1	1	0.5	3.4	5.1	3.7
58205219	VERV	9.7	56	3	Ecoute	26.0	28.3	6	2	0.1	4.0	8.0	7.7
58205237	VERV	5.5	81	2	Ecoute	14.0	11.0	1	0	0.0	3.6	5.1	4.4
58205380	VERV	11.1	48	2	Ecoute	20.0	22.0	5	2	0.2	2.9	8.5	6.7
58217155	VERV	11.4	20	3	Mixte	57.0	24.0	3	2	2.0	2.3	6.5	3.3
Total		186.7	53	50		469.0	315.4	53	46	45.3	53.3	140.9	84.9

Ces données mettent en évidence la grande variété des situations, de 1 à 6 campagnes par cellules avec des durées cumulées de 3 à 63 heures et un nombre de fuites trouvées par cellule allant de 0 à 7.

11.3.2 Regroupement des cellules par type de campagne

Le Tableau 45 présente les données principales et des ratios qui leur sont associés pour les regroupements de cellules par type de campagne.

Pour les différents ratios, des écarts importants existent entre les trois groupes et à ce stade il est important de préciser que ces écarts ne peuvent pas, sur l'échantillon étudié, être imputés uniquement à la technique utilisée. En effet les techniques sont associées à des secteurs différents (territoires et réseaux différents) et à des responsables RF différents (disponibilités et stratégies différentes), et par ailleurs la technique « mixte » résulte d'un contexte ou de difficultés qui ont rendu peu opérante la technique initialement prévue.

Par exemple, le nombre de fuites trouvé par kilomètre de réseau inspecté ($N_{fui}/KmCam$) est fortement influencé par l'effort de recherche consenti pour se rapprocher du débit cible. Un indicateur de cet effort est le rapport entre le linéaire inspecté et la longueur de la cellule ($KmCam/L$) qui peut aller du simple au double.

Tableau 45. Données et ratios des cellules regroupées par type de campagne

TypCam	Nam	Verv	L	D	NbCam	TpsCam	KmCam	NbLoc	TpsLoc	NbFui	NbTPR	GainQfui
Ecoute	0	8	54	63	20	156	108	21	80	13	3	15
Tronçonnage	7	0	79	47	12	98	85	17	50	16	7	20
Mixte	5	1	54	52	18	215	123	15	49	17	4	11
Total	12	9	187	53	50	469	315	53	179	46	14	45

Type Cam	KmCam/L	TpsCam/ KmCam	NbFui/L	NbFui/ KmCam	NbTPR/ KmCam	GainQfui/nbFui
Ecoute	2.01	1.44	0.24	0.12	0.028	1.14
Tronçonnage	1.07	1.16	0.20	0.19	0.083	1.22
Mixte	2.27	1.75	0.32	0.14	0.033	0.64
Total	1.69	1.49	0.25	0.15	0.044	0.98

11.3.3 Prise en compte de la typologie des cellules

Les données sélectionnées à l'origine de l'expérimentation pour caractériser les cellules sont :

- Le rapport du débit nocturne de la cellule en début d'expérimentation et du débit cible ($Q_{début}/Q_{cible}$), adimensionnel et d'autant plus élevé que la présence de fuites importantes est suspectée ;
- La densité de raccordement par kilomètre de réseau (D) en nombre de raccordements par km ;
- Le pourcentage du linéaire de la cellule constitué de canalisations métalliques ;
- Le pourcentage du linéaire de la cellule situé sous chaussée ;
- La corrosivité de l'eau transportée par les conduites (-1 = Entartrante, 0 = neutre, 1 = corrosive, 2 = très corrosive).

Chacune de ces variables a été scindée en deux classes concernant chacune environ la moitié des 21 cellules étudiées.

L'effet de la typologie de la cellule sur l'efficacité de la recherche de fuites pourrait être examiné selon plusieurs indicateurs basés sur les données disponibles, toutefois, en l'absence de significativité statistique de l'échantillon de cellules analysé (du fait de son faible effectif pour un nombre important de variables), il est proposé d'examiner, à titre d'exemple, un unique indicateur choisi selon une approche métier. Il s'agit du rapport entre le débit cumulé des fuites réparées et la durée cumulée des campagnes de pré-localisation en jours (GQF/JC , le nombre de journées de campagne est le temps de campagne en heures divisé par 8). Cet indicateur, qui mesure des « m³/h gagnés par journée de recherche de fuite » pourraient en effet être une information pertinente pour identifier les cellules prioritaires pour des actions de RF.

Le Tableau 46 récapitule, associé au nombre de cellules concernées, les valeurs de cet indicateur par technique de pré-localisation et par classes des variables sélectionnées.

Tableau 46. Gain en débit des fuites réparées par journée de campagne (par technique et par classes de variables)

Critère	Classe	Variable	Ecoute	Tronçonnage	Mixte	Toutes Tech
Ensemble des cellules		NbCel	8	7	6	21
		GQF/JC	0.76	1.60	0.40	0.77
Classe Qdebut/Qcible	[0-3]	NbCel	5	3	4	12
		GQF/JC	0.23	2.26	0.14	0.54
	[3 et +]	NbCel	3	4	2	9
		GQF/JC	1.69	0.99	0.79	1.10
Classe D (Racc/km)	[0-50]	NbCel	2	6	2	10
		GQF/JC	0.16	1.35	0.23	0.73
	[50 et +]	NbCel	6	1	4	11
		GQF/JC	0.90	3.38	0.50	0.80
Classe % conduites métalliques	[0-70%]	NbCel	2	4	3	9
		GQF/JC	0.52	1.95	0.18	0.77
	[70% et +]	NbCel	6	3	3	12
		GQF/JC	0.95	1.02	0.58	0.78
Classe % conduites sous chaussée	[0-40%]	NbCel	1	5	4	10
		GQF/JC	0.37	1.48	0.48	0.76
	[40% et +]	NbCel	7	2	2	11
		GQF/JC	0.79	1.98	0.02	0.79
Classe corrosivité de l'eau	[-1-0]	NbCel	0	7	4	11
		GQF/JC		1.60	0.48	0.88
	[1-2]	NbCel	8	0	2	10
		GQF/JC	0.76		0.02	0.62

Compte tenu de la faiblesse de l'échantillon, certaines des valeurs de *GQF/JC* concernent très peu de cellules voire ne sont pas calculables (classes de corrosivité de l'eau). Néanmoins, quelques constats peuvent être faits :

- Toutes techniques confondues, la seule variable qui a un impact net sur *GQF/JC* est *Qdebut/Qcible*, les cellules de la classe la plus élevée ([3 et +]) présentent des bénéfices valant en moyenne plus du double que pour l'autre classe, à noter que pour le « tronçonnage » la hiérarchie est inversée (pas de raison « métier » évidente pour expliquer cela) ;
- La proportion de conduites métalliques impacte l'efficacité de l'écoute et des méthodes mixtes mais pas celle du tronçonnage (conforme à la pratique, les méthodes acoustiques sont moins efficaces sur les conduites en plastique).

11.4 Conclusion

L'expérimentation conduite dans le cadre de la Tâche 11 du projet GePaME a été conçue avec des visées exploratoires pour mieux connaître les déterminants de l'efficacité des actions de recherche et de réparation des fuites, selon les techniques employées et en lien avec les débits et les caractéristiques des cellules.

Initialement lancée sur un échantillon restreint de 35 cellules sélectionnées dans trois secteurs (Mons, Namur, Verviers), l'étude n'a finalement produit des résultats complets que sur 21 cellules. La raison principale de ce décalage est l'inadéquation des enregistreurs acoustiques utilisés sur le secteur de Mons à un déploiement mobile (difficultés de paramétrage) qui concerne 10 cellules.

Ce travail n'a pas produit de règles ou de ratios généralisables à l'ensemble de la SWDE. Il a en revanche permis d'expérimenter une méthode pour concevoir et relier entre eux des indicateurs qui caractérisent ou mesurent la nature et l'intensité des actions menées, la typologie des cellules et les résultats obtenus.

Un élément central de la méthode utilisée est de combiner les débits nocturnes (issus de Perf'O) aux actions menées, celles-ci étant localisées, datées et caractérisées (Figure 73).

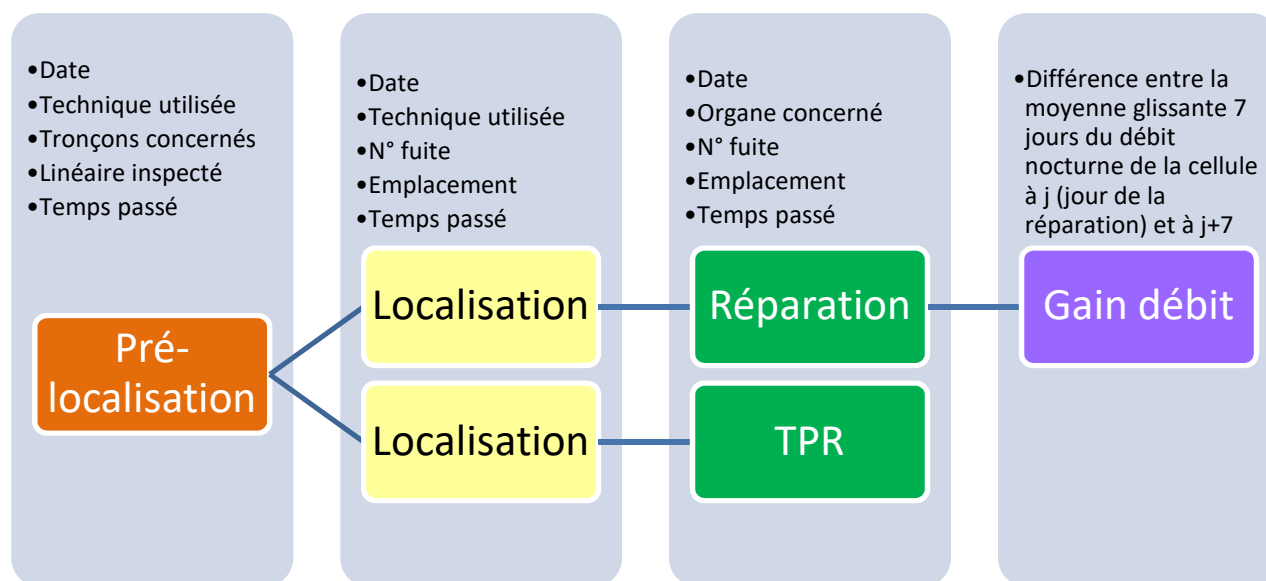


Figure 73 : Chaîne de combinaison des données associées aux actions et aux résultats

Une partie des données utilisées sont d'ores et déjà prises en charge par le système d'information de la SWDE (SAP, SIG, Perf'O), néanmoins, au cours de l'expérimentation, des écarts significatifs ont été constatés entre les données présente dans le SI et celles directement reportées dans les fiches de suivi par les opérateurs. Certaines autres informations ne sont pas aujourd'hui collectées systématiquement : conduites et linéaire concernés par une pré-localisation, techniques utilisées pour la pré-localisation et pour la localisation).

A partir de ces données, différents indicateurs de résultats peuvent être calculés. Nous proposons, par exemple, le débit de fuite économisé rapporté au temps de pré-localisation (le temps de localisation pourrait également être ajouté au dénominateur).

Si une démarche de collecte systématique et de fiabilisation des données était mise en œuvre, l'analyse de l'impact de différentes caractéristiques des actions et des cellules (techniques utilisées, urbanisation, nature des matériaux, écart à l'ILP de référence, etc.) sur la valeur des indicateurs de résultats nous paraît à même de fournir des éléments pertinents pour orienter ou optimiser les actions de RF.

Archive Tâche 11 :

- Traitement final des données et résultats
- Dossier ArchiveT11-02_FichesSuiviFinales : Fiches de suivi

12 Tâche 12 – Etude de scénarios long terme et simulation

12.1 Objectif de la tâche et déroulement

L'objectif de cette tâche est de fournir à la SWDE une méthode de simulation de l'évolution du réseau dans le long terme (une à plusieurs décennies) dans divers scénarios de stratégies de renouvellement du réseau selon la méthode multicritère décrite en section 10. Ces scénarios peuvent porter par exemple sur le montant alloué pour les renouvellements, ou sur le choix des préférences pour les différents critères de renouvellement (voir le paragraphe 10.4). Le suivi de l'évolution du réseau peut se faire par exemple grâce à des indicateurs similaires à ceux de l'outil multicritère court terme.

Compte tenu du développement de l'outil multicritère en fin de projet et de l'impossibilité matérielle à entreprendre de nouveaux recueils et traitements de données spécifiques, l'avancée de cette tâche a été limitée et seuls des résultats préliminaires sont disponibles. Ces derniers sont présentés dans les sections suivantes, suivis des pistes d'améliorations envisagées.

12.2 Principe de la simulation long terme (SLT)

Essentiellement, la SLT effectue une boucle sur l'outil court terme. Pour chaque année A future, successivement :

- Les indicateurs de performance et d'opportunité sont calculés en fonction des variables de l'année A et un score global est affecté à chaque tronçon du réseau ;
- Les tronçons en tête de classement sont « choisis » pour le renouvellement jusqu'à consommation totale du budget disponible (leur probabilité de renouvellement p_R est fixée à 1) ;
- Tous les tronçons sont dupliqués, le premier doublon est renouvelé de A à A+1 (probabilité p_R), le deuxième doublon reste en service de A à A+1 (probabilité $1 - p_R$) ;
- La probabilité d'existence des tronçons p_E l'année A+1 est mise à jour selon la probabilité de renouvellement entre A et A+1 : p_E l'année A devient $p_E(1 - p_R)$ l'année A+1 pour les tronçons non renouvelés, et devient $p_E p_R$ l'année A+1 pour les tronçons renouvelés ;
- Certaines variables évoluent de A à A+1. Des défaillances sont prédites sur tous les tronçons pour l'année A+1 grâce au LEYP. Pour les tronçons renouvelés, le matériau est éventuellement modifié (ex : FG -> FD), l'année de pose devient A+1, et le nombre de défaillances passées est remis à 0. Pour les tronçons non renouvelés, l'âge augmente d'un an ;
- Toutes les trajectoires d'un tronçon donné menant à un renouvellement entre A et A+1 sont fusionnées (ceci évite une croissance exponentielle de la taille des données simulées), ce qui revient à ne conserver qu'une copie des possibles doublons de (ID ; année de pose) et à conserver comme probabilité d'existence pour ce dernier la somme des p_E de toutes les répétitions éventuelles de ce doublon ;
- A devient A+1, retour au premier point.

Ainsi, un tronçon n'est plus seulement identifié par son ID, il faut également ajouter une année de pose. Les tronçons (ID ; 2030) et (ID ; 2031) ne sont donc plus les mêmes, puisque contrairement au premier, le second a été renouvelé en 2031. Cependant, (ID ; 2031) n'est qu'un seul et unique tronçon, peu importe qu'il existe diverses trajectoires où ce tronçon a été renouvelé 1, 2, ou n fois avant 2031 (Figure 74).

Notons que l’algorithme original de SLT permettait de simuler de façon probabiliste l’occurrence d’opportunités de coordinations avec d’autres gestionnaires et l’occurrence de renouvellements obligatoires imposés par les autres gestionnaires. Ces évènements avaient la capacité de modifier le coût de renouvellement des tronçons concernés chaque année. Une part du budget était alors mise de côté pour ces renouvellements, et consommée de façon probabiliste selon la position du tronçon dans le classement multicritère. L’inclusion de la coordination et des chantiers obligatoires nécessite cependant d’utiliser le modèle LEYP2s et non le LEYP. En négligeant cette possibilité, p_E et p_R reviennent à des variables binaires (simulation 100 % déterministe basée sur les renouvellements ciblés), et le LEYP suffit. Il est important de noter qu’avec la coordination et les renouvellements obligatoires, ce n’est plus le cas. Ainsi, dans les résultats présentés ci-dessous, les renouvellements sont déterminés uniquement par les critères de renouvellement déterministes sans composante aléatoire.

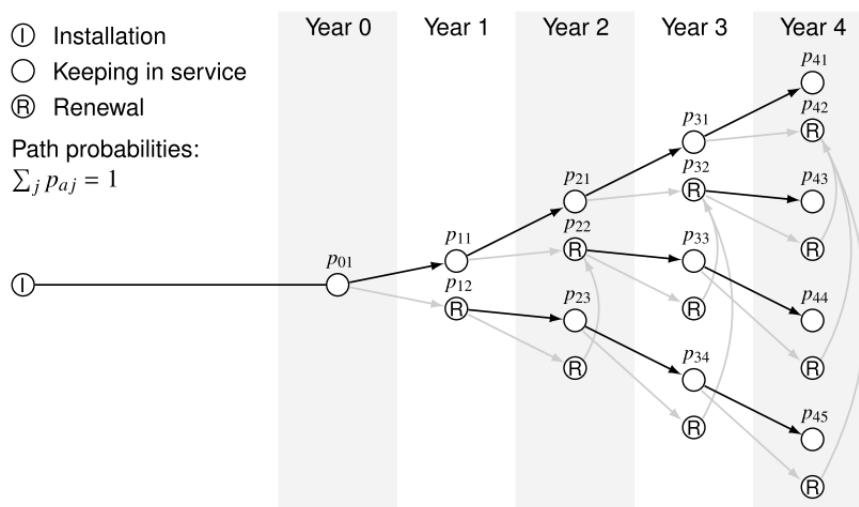


Figure 74. Trajectoire d'un tronçon dans la simulation long terme.

12.3 Hypothèses utilisées

Les simulations ont été faites uniquement sur le réseau d’adduction (3 645 km) pour les deux raisons suivantes. Premièrement, les données sont de taille bien plus restreinte, ce qui allège les temps de calcul. Deuxièmement, en distribution, il est plus difficile de prévoir l’évolution de certaines variables intervenant dans le calcul des indicateurs. Par exemple, concernant IP_{pertes} , il n’existe pas de mécanisme simple de prédiction des tronçons qui seront inspectés par la RF, et parmi ceux-ci, lesquels seront trouvés fuyards d’une année à l’autre (alors que les défaillances = fuites ou crashes peuvent être prédites par le LEYP simplement). Il n’existe pas non plus de moyen simple de prédire l’évolution de l’ILP d’une cellule d’une année à l’autre en fonction des tronçons renouvelés. Enfin, il est difficile de prévoir l’évolution de la tension sur l’approvisionnement en eau des cellules car cela prend en compte l’évolution des pertes, ainsi que le changement climatique et la mise en place d’infrastructures pour lutter contre ses effets (ex : sécheresse et schéma régional des ressources en eau).

Les hypothèses suivantes ont été utilisées, dans un souci de simplification pour les premiers résultats :

- Pas de prise en compte des tronçons « courts » (de moins de 6 mètres) ;

- Les indicateurs utilisés dans le classement chaque année et les poids sont directement issus de l'outil multicritère court terme (voir section 10.4). Le suivi long terme se fait pour un critère donné en sommant simplement la valeur de son indicateur (non normalisé) sur l'ensemble du réseau chaque année.
- Pas d'évolution de la qualité de l'eau et des coûts de production au sein des ZDE, ni des coûts de réparation des tronçons ;
- Paramètres du LEYP issus de Casses (modèle 2011-2021, strate adductions) ;
- Renouvellements des conduites en PVC et PE par du PVC (75 % de chances) ou du PE (25 % de chances), et des autres matériaux par de la FD (85 % de chances) ou de l'acier (15 % de chances), au diamètre identique. Ces probabilités sont issues des données de renouvellements depuis 2000 ;
- Coûts de renouvellement des tronçons en fonction du diamètre et du nombre de branchements (Annexe T10_02) ;
- 0.85 % du linéaire d'adduction renouvelé par an (moyenne 2000-2021), soit 9.402 M€ par an. **Il s'agit d'une estimation optimiste car elle comprend les poses de nouveaux réseaux en plus des renouvellements ;**
- Pas de coordination des travaux ni de renouvellements obligatoires, 100 % du budget pour des renouvellements ciblés ;
- Simulation de 2022 à 2070 inclus.

12.4 Résultats

Les indicateurs suivis chaque année sont divisés par la valeur au début de la simulation (2022). Le nombre de casses chaque année est affiché en complément (courbe noire).

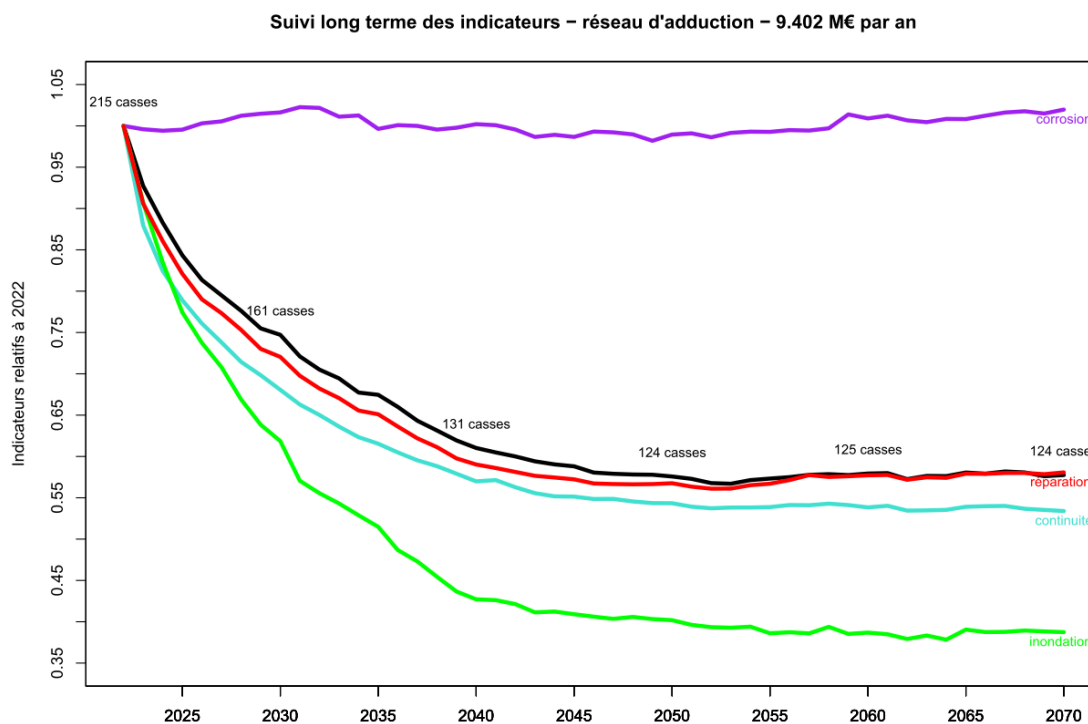


Figure 75. Simulation long terme avec un budget probablement surestimé.

Suivi long terme des indicateurs – réseau d'adduction – 4.701 M€ par an

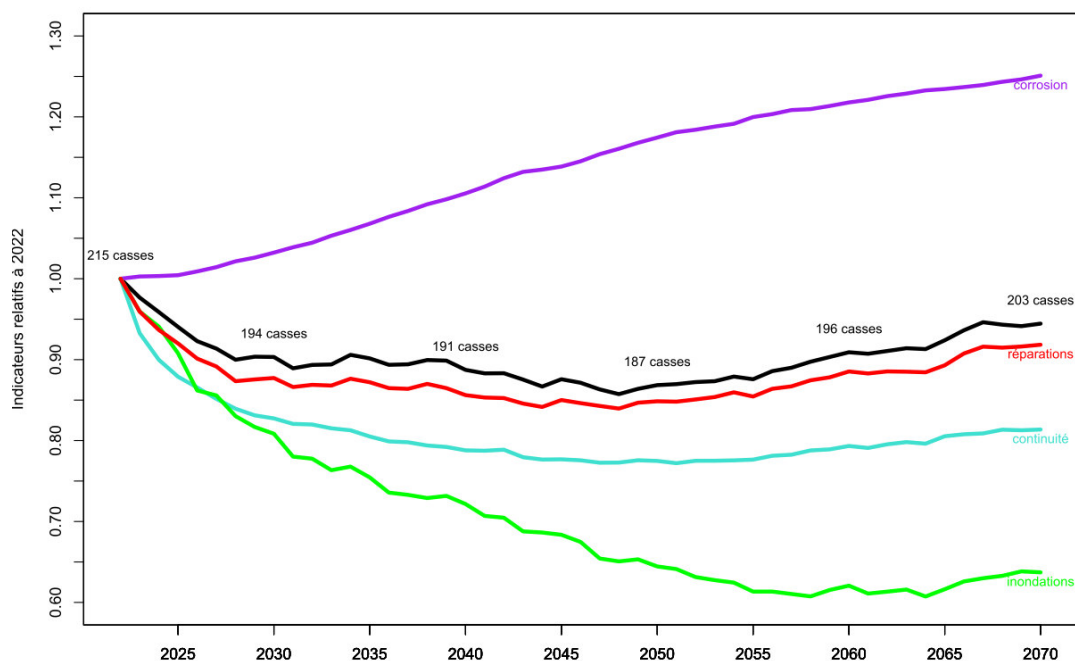


Figure 76. Simulation long terme avec un budget probablement sous-estimé.

Avec 9.402 M€ par an (probablement supérieur au budget actuel, Figure 75), la stratégie de renouvellement, basée sur les préférences du Tableau 31, semble convenir, sauf pour l'indicateur de corrosion qui ne s'améliore pas même sur le très long terme. La situation n'est pas satisfaisante avec seulement la moitié de ce budget (Figure 76), où les indicateurs ne montrent pas une amélioration rapide, et au contraire, une dégradation progressive après 2030 (sauf pour l'indicateur d'inondations des immeubles).

12.5 Prochaines étapes

La simulation long terme est un exercice difficile car il est nécessaire de modéliser par des hypothèses probabilistes l'évolution de nombreuses variables. Nous pouvons tout de même avancer les prochaines pistes d'amélioration de la méthode :

- Définir d'autres indicateurs non basés sur ceux de l'outil court terme (ne pas suivre que ceux utilisés pour classer les tronçons chaque année) ;
- Faire évoluer les variables et indicateurs gardés fixes dans un premier temps (coûts de production, coûts de réparation, qualité de l'eau, etc.). Ceci requiert des hypothèses probabilistes réalistes, qui peuvent demander des données difficiles à recueillir.
- Inclure la coordination et les renouvellements obligatoires et rendre la simulation des renouvellements réellement probabiliste ($p_E \in]0; 1[$ au lieu d'une variable binaire) ;
- Tester de nombreux autres scénarios basés sur des poids différents, des budgets différents, et un mélange des deux approches ;
- Utiliser l'outil comme un moyen de fixer les objectifs de performance et de calculer le budget nécessaire pour atteindre ces objectifs ;
- Appliquer l'outil pour les conduites de distribution (voir remarques en section 12.3).

Archive Tâche 12 :

- Dossier long-terme : données et code de l'outil multicritère long terme

13 Tâche 13 - Coordination du Projet et valorisation

13.1 Coordination du projet

La gouvernance du projet est assurée par deux instances :

- Le comité de pilotage stratégique (CPS) : composé de représentants des directions de SWDE et de l'unité Environnement, Territoires en Transition, Infrastructures, Sociétés (ETTIS) d'INRAE, il veille au bon déroulement du projet et peut proposer des réorientations des recherches et/ou des aménagements du contrat et se réunit une fois par an ;
- Le comité de suivi opérationnel (CSO) : composé des responsables de l'exécution du projet, il assure le suivi et l'animation du programme scientifique et se réunit une fois par mois.

Le CPS s'est réuni à deux reprises, les 08/12/21 et 13/12/22. Au cours de l'année 2023, une instance associant une partie de la direction de SWDE et le porteur de projet INRAE a pris le relai pour réaliser des points d'avancement réguliers (4 réunions). Les diaporamas supports de ces réunions sont fournis en archive (ArchiveT13-01_Pptx-ReunionsCPS&DirSWDE.zip).

Le CSO a tenu 23 réunions, les comptes rendus et les diaporamas relatifs à ces réunions sont également fournis en archive (ArchiveT13-02_CR&Pptx-ReunionsCSO.zip).

En complément de ces deux instances, des groupes thématiques associant les personnes plus particulièrement concernées au sein de SWDE et INRAE se sont réunis régulièrement en fonction de l'avancement des tâches concernées (tâche 4 « Intégration des outils dans le système d'information », tâche 10 « Outil multicritère de programmation court terme » et tâche 11 « Efficacité des actions de recherche de fuite »).

Le planning de réalisation du projet a été ajusté régulièrement par le CSO en fonction des contraintes et des aléas qui se sont présentés, la Figure 77 présente le planning final de déroulement du projet.

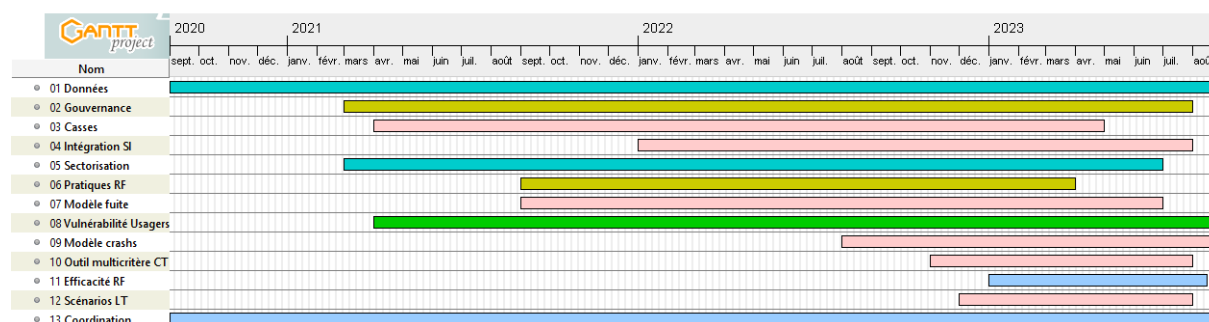


Figure 77 : Planning final de réalisation du projet GePaME

13.2 Valorisations

Pendant sa réalisation, le projet a fait l'objet de plusieurs valorisations à destination de publics scientifiques et techniques sous forme de présentation à des conférences ou d'articles, plusieurs valorisations sont également engagées dans les mois à venir (Tableau 47).

Tableau 47 : Communications des résultats de GePaME dans des conférences

Conférence	Titre	Tâche	1 ^{er} auteur
11-13/05/2022 IWA-LESAM (Leading-Edge Strategic Asset Management) Bordeaux	Understanding the leakage process for multi-scale water infrastructure asset management: necessity for a dialogue between sociological and data sciences	1, 6, 7	Marie Collet
14-16/06/2022 Congrès de l'ASTEE, Dunkerque, France	Méthodes automatisées de préparation des données pour la gestion patrimoniale du réseau de la SWDE	1	Nicolas Rodriguez
02-08/06/2023 Congrès de l'ASTEE, Nice, France	Orienter la recherche de fuites par l'estimation de la probabilité d'état fuyard des conduites	7	Marie Collet
02-08/06/2023 Congrès de l'ASTEE, Nice, France	Vulnérabilité des consommateurs wallons au risque de sécheresse et consentements à payer pour l'atténuer	8	Bénédicte Rulleau
30/05-02/06/2023 Conference on Mathematical Methods in Reliability (MMR), Murcia, Spain	A doubly binary regression for zero inflated observations in the prediction of pipes leakage	7	Vincent Couallier
03-07/07/2023 Journée de Statistiques (JdS), Bruxelles, Belgique	Vrais et faux zéros : double régression logistique à variables latentes pour modéliser les observations manquantes dans un modèle de présence / absence	7	Marie Collet
13-15/09/2023 IWA-EFFICIENT, Bordeaux, France	Improving data for the asset management of the water supply network of the Walloon Water Company	1	Nicolas Rodriguez
13-15/09/2023 IWA-EFFICIENT, Bordeaux, France	A new statistical method to predict the leaking condition of pipes	7	Marie Collet
13-15/09/2023 IWA-EFFICIENT, Bordeaux, France	Vulnerability of Walloon drinking water consumers to droughts and willingness to pay to mitigate their consequences	8	Bénédicte Rulleau
13-15/09/2023 IWA-EFFICIENT, Bordeaux, France	Applied research for multi-scale asset management of the Walloon Water Company's drinking water supply network	Toutes	Eddy Renaud
07-08/11/2023 IWA_PI 2023, Amsterdam, Netherlands	Development of a PI to take into account water loss reduction in the SWDE infrastructure management strategy	5, 7, 10	Eddy Renaud

A l'issue de la conférence ASTEE 2022 à Dunkerque, un article des mêmes auteurs et portant le même titre a été publié dans la revue TSM (Techniques Sciences Méthodes) :

Nicolas Rodriguez, Alain Husson, Yves Le Gat, Eddy Renaud. Méthodes automatisées de préparation des données pour la gestion patrimoniale des réseaux d'eau potable. Techniques Sciences Méthodes, 2022, 10, pp.71-87. (10.36904/tsm/202210071). (hal-03836560)

Dans la même logique, des articles seront soumis à des revues de l'IWA à l'issue de la conférence EFFICIENT 2023. D'autres valorisations sont envisagées sous forme d'article dans des revues spécialisées, notamment en ce qui concerne les tâches 6 « Analyse des pratiques de recherche de fuites » et 9 « Modélisation statistique couplée des fuites et des crashes ».

Enfin, le poster réalisé conjointement par INRAE et SWDE pour la communication « Applied research for multi-scale asset management of the Walloon Water Company's drinking water supply network »

à la conférence EFFICIENT 2023, fera l'objet d'une adaptation en français pour servir la communication interne et externe autour du projet.

Les résumés, articles et support de présentation des différentes valorisations sont joints en annexe (fichier Annexe T13_01).

Archive Tâche 13 :

- Dossier ArchiveT13-01_Pptx-ReunionsCPS&DirSWDE : présentations faites lors des CPS et réunions de directeurs sur le projet GePaME
- Dossier ArchiveT13-02_CR&Pptx-ReunionsCSO : présentations et comptes rendus des CSO mensuels du projet GePaME.

Conclusion générale

Principaux résultats

Le projet GePaME a produit des avancées scientifiques et des méthodes et outils opérationnels.

Les principales avancées scientifiques sont :

- Des méthodes et algorithmes de traitement et de complétion des données permettant d'alimenter les modèles (tâche 1) ;
- Des modèles statistiques permettant d'estimer la probabilité d'un tronçon d'être fuyard conjointement à celle d'être inspecté par les équipes de RF (tâche 7) et d'appréhender conjointement les phénomènes de fuites détectables et de casses franches du réseau (tâche 9), des méthodes d'analyse historique des débits minimum de nuit des cellules permettant de fixer des valeurs cibles et d'identifier les cellules critiques (tâche 5) ;
- Une explicitation sociologique des rapports homme-outils-décisions-actions à l'échelle de la SWDE (tâche 2) et sous l'angle particulier de la mise en œuvre de la recherche des fuites (tâche 6) ;
- Une analyse économétrique des consentements-à-payer des usagers pour des mesures de lutte contre les restrictions d'usage de l'eau liées au risque de sécheresse (tâche 8).

La valorisation opérationnelle par SWDE de ces avancées scientifiques concerne :

- L'amélioration de l'interopérabilité des bases de données alimentant le système d'information, la rationalisation de l'enregistrement des défaillances du réseau et de l'automatisation des méthodes de traitement des données à des fins de gestion patrimoniale (tâche 1 et 4) ;
- L'amélioration de l'organisation de la recherche de fuite en matière de priorisation des secteurs, de choix des types d'action et de gestion de l'information (tâches 5, 6 et 11)
- La mise en place d'un nouvel outil multicritère d'aide à la décision de renouvellement des canalisations, basé sur une prédiction des casses futures (tâche 3) et permettant de hiérarchiser l'ensemble du réseau à l'échelle du tronçon (tâche 10 alimentée par la plupart des autres tâches)
- L'expérimentation d'un simulateur permettant d'évaluer et comparer la performance à long terme du système résultant de différents scénarios d'investissement (tâche 12).

Autres enseignements

Le projet a été conduit sous une forme partenariale, multi-échelle et interdisciplinaire.

Grâce à ce partenariat INRAE a eu accès à une masse importante de données structurées et a bénéficié du savoir-faire et de l'expertise de la SWDE et a pu ainsi renouveler ses questions de recherche pour être en phase avec les besoins du terrain. En parallèle, SWDE a pu s'inscrire dans une démarche de progrès en ajustant ses pratiques pour valoriser les potentialités offertes par les innovations du projet.

L'approche multi-échelles a permis d'une part de prendre en compte les interactions entre l'organisation, les territoires et les infrastructures et d'autre part de renforcer la cohérence entre la programmation à court terme du renouvellement des infrastructures et les stratégies de performance du service à long terme.

Enfin, la mobilisation conjointe des sciences techniques et des sciences humaines et sociales, a permis une prise en compte du contexte humain (organisation et pratiques internes de la SWDE, point de vue des usagers) dans la construction des propositions techniques.

Les outils et méthodes développées dans le cadre du projet constituent une avancée pour améliorer la rationalisation et le décloisonnement des pratiques et stratégies de gestion patrimoniale des infrastructures et de lutte contre les pertes en eau. Après une phase nécessaire d'appropriation et d'évaluation, ils ouvrent des perspectives pour s'intégrer dans une vision plus globale de l'efficacité et de la performance du service en abordant notamment les questions des risques, de l'adaptation au changement climatique, de l'évolution des technologies et de leurs impacts sur les arbitrages stratégiques en termes de renouvellement des compétences humaines et d'investissement.

Références

- Abbott, Andrew. 1988. *The system of professions: An essay on the division of expert labor*. Chicago, IL, US: The University of Chicago Press.
- Akaike, H. 1974. « A new look at the statistical model identification ». *IEEE Transactions on Automatic Control* 19(6): 716-23.
- AquaWal. 2022. *Statistiques de l'eau potable et de l'assainissement des eaux usées en Wallonie. Rapport 2022*. <https://www.aquawal.be/servlet/Repository/sa-aquawal-statistiques-2022.pdf?ID=17583&saveFile=true>.
- Baati, Selma, et Kevin Caillaud. 2023. *Analyse des pratiques de recherche de fuites dans les réseaux d'eau potable. Le cas de la Société Wallonne Des Eaux*. INRAE. Rapport scientifique dans le cadre du projet Gestion Patrimoniale Multi-échelle des réseaux d'eau potable.
- Barbier, R. et al. 2016. *Les territoires de l'eau potable : chronique d'une transformation silencieuse (1970-2015)*. Quae. <https://hal.inrae.fr/hal-02603599> (29 août 2023).
- Barraqué, Bernard. 1995. « Les Politiques De L'eau En Europe ». *Revue française de science politique* 45(3): 420-53.
- Bastin, O. et al. 2022. « Risques et vulnérabilités climatiques des ouvrages du secteur de l'eau en Wallonie: Climate risk and vulnerability assessment of water sector infrastructure in Wallonia ». *Techniques Sciences Méthodes* 3: 33-48.
- Bateman, Ian J. et al. 2002. *Economic Valuation with Stated Preference Techniques: A Manual*. Cheltenham. <https://ueaeprints.uea.ac.uk/id/eprint/28663/> (24 août 2023).
- Bennett, J., B. McNair, et J. Cheesman. 2016. « Community preferences for recycled water in Sydney ». *Australasian Journal of Environmental Management* 23(1): 51-66.
- Bennett, Jeff, et Russell Blamey. 2001. *The Choice Modelling Approach to Environmental Valuation*. Edward Elgar Publishing.
- Berlin, Isaiah. 2011. *Le sens des réalités*. Les Belles Lettres, Le Goût des Idées. <https://cir.nii.ac.jp/crid/1130000796806302080> (29 août 2023).
- Bernoux, Philippe. 2004. « Sociologie du changement dans les entreprises et les organisations ». *Le Seuil*.
- Beyersmann, Jan, Arthur Allignol, et Martin Schumacher. 2011. *Competing Risks and Multistate Models with R*. Springer Science & Business Media.
- Blamey, Russell, Jenny Gordon, et Ross Chapman. 1999. « Choice Modelling: Assessing the Environmental Values of Water Supply Options ». *Australian Journal of Agricultural and Resource Economics* 43(3): 337-57.
- Blasco-Moreno, Anabel et al. 2019. « What does a zero mean? Understanding false, random and structural zeros in ecology ». *Methods in Ecology and Evolution* 10(7): 949-59.

- Boussard, Valérie. 2008. *Sociologie de la gestion, les faiseurs de performance*. Paris, Belin (collection Perspectives Sociologiques).
- Boxall, Peter C., et Wiktor L. Adamowicz. 2002. « Understanding Heterogeneous Preferences in Random Utility Models: A Latent Class Approach ». *Environmental and Resource Economics* 23(4): 421-46.
- Bozdogan, Hamparsum. 1987. « Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytical Extensions ». *Psychometrika* 52(3): 345-70.
- Bruno, Isabelle, et Emmanuel Didier. 2015. *Benchmarking: L'État sous pression statistique*. La Découverte.
- Caillaud, Kévin. 2017. « La réorganisation des services urbains environnementaux ». *Métropolitiques*: <https://www.metropolitiques.eu/La-reorganisation-des-services-urbains-environnementaux.html>.
- Callon, Michel. 1986. « ÉLÉMENTS POUR UNE SOCIOLOGIE DE LA TRADUCTION: La domestication des coquilles Saint-Jacques et des marins-pêcheurs dans la baie de Saint-Brieuc ». *L'Année sociologique (1940/1948-)* 36: 169-208.
- Carson, Richard T., et Theodore Groves. 2011. « Incentive and Information Properties of Preference Questions: Commentary and Extensions ». In *The International Handbook on Non-Market Environmental Valuation*, éd. Jeff Bennett. Edward Elgar Publishing. <https://www.elgaronline.com/view/edcoll/9781848444256/9781848444256.00020.xml> (24 août 2023).
- Cha, Ji Hwan, et Maxim Finkelstein. 2021. « On Some General Survival Models with Delayed Failures ». *Communications in Statistics - Theory and Methods*: 1-18.
- Chiapello, Ève et al. 2015. « À propos de l'emprise du chiffre ». *Entreprises et histoire* 79(2): 174-87.
- Cochoy, Franck, Jean-Pierre Garel, et Gilbert de Terssac. 1998. « Comment l'écrit travaille l'organisation: le cas des normes Iso 9000 ». *Revue française de sociologie* 39(4): 673-99.
- Cooper, Bethany, Michael Burton, et Lin Crase. 2019. « Willingness to Pay to Avoid Water Restrictions in Australia Under a Changing Climate ». *Environmental and Resource Economics* 72(3): 823-47.
- Craipeau, Sylvie, et Jean-Luc Metzger. 2011. « La gestionnarisation de la société : de la volonté de maîtrise à la déréalisation ». *Mana : revue de sociologie et d'anthropologie* (17-18): 23.
- CRA-W. 2022. « La sécheresse de 2022 en Wallonie en trois questions ». CRA-W. <https://www.cra.wallonie.be/fr/la-secheresse-de-2022-en-wallonie-en-trois-questions> (28 juillet 2023).
- Denis, Jérôme. 2006. « Préface : Les nouveaux visages de la performativité ». *Études de communication. langages, information, médiations* (29): 8-24.
- Dessai, S., & Sims, C. (2010). *Public perception of drought and climate change in southeast England*. <https://doi.org/10.3763/ehaz.2010.0037>

- Desrosières, Alain. 2014. *Prouver et gouverner: Une analyse politique des statistiques publiques*. La Découverte.
- Fidelle, Simon. 2022. *La gouvernance territoriale de la gestion patrimoniale des infrastructures d'eau potable. Étude de cas sur la Société Wallonne Des Eaux*. INRAE-Université Toulouse Jean-Jaurès. Mémoire de stage M2.
- Greene, William H. 2003. *Econometric Analysis*. 5th ed. Upper Saddle River, N.J: Prentice Hall.
- Griffin, Ronald C., et James W. Mjelde. 2000. « Valuing Water Supply Reliability ». *American Journal of Agricultural Economics* 82(2): 414-26.
- Hanemann, W. Michael. 1984. « Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses ». *American Journal of Agricultural Economics* 66(3): 332-41.
- Hanley, Nick, Susana Mourato, et Robert E. Wright. 2001. « Choice Modelling Approaches: A Superior Alternative for Environmental Valuation? - Hanley - 2001 - Journal of Economic Surveys - Wiley Online Library ». <https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-6419.00145> (7 juin 2023).
- Hellequin, Anne-Peggy, Hervé Flanquart, Catherine Meur-Ferec, et Bénédicte Rulleau. 2013. « Perceptions du risque de submersion marine par la population du littoral languedocien : contribution à l'analyse de la vulnérabilité côtière ». *Natures Sciences Sociétés* 21(4): 385-99.
- Hensher, David, Nina Shore, et Kenneth Train. 2006. « Water Supply Security and Willingness to Pay to Avoid Drought Restrictions* ». *Economic Record* 82(256): 56-66.
- Hess, Stephane, et David Palma. 2019. « Apollo: A Flexible, Powerful and Customisable Freeware Package for Choice Model Estimation and Application ». *Journal of Choice Modelling* 32: 100170.
- Hoyos, David. 2010. « The state of the art of environmental valuation with discrete choice experiments ». *Ecological Economics* 69(8): 1595-1603.
- Husson, Alain et al. 2022. *Gestion Patrimoniale Multi-Echelles des réseaux d'eau potable (GePaME) - Rapport intermédiaire 2022*. <https://hal.inrae.fr/hal-03836891> (29 août 2023).
- IPCC. 2023. *IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (Eds.)]*. IPCC, Geneva, Switzerland. First. Intergovernmental Panel on Climate Change (IPCC). <https://www.ipcc.ch/report/ar6/syr/> (22 août 2023).
- Jackson, Chris. 2014. « Multi-state modelling software, and encouraging statistical software development. » Présenté à MRC Biostatistics Unit Centenary Conference, Cambridge, UK.
- Jackson, Christopher. 2011. « Multi-State Models for Panel Data: The Msm Package for R ». *Journal of Statistical Software* 38: 1-28.
- Johnston, Robert J. et al. 2017. « Contemporary Guidance for Stated Preference Studies ». *Journal of the Association of Environmental and Resource Economists* 4(2): 319-405.

- Kanyoka, P., S. Farolfi, et S. Morardet. 2008. « Households' Preferences and Willingness to Pay for Multiple Use Water Services in Rural Areas of South Africa: An Analysis Based on Choice Modelling ». *Water SA* 34(6): 715-23.
- Kim, S., Shao, W., & Kam, J. (2019). Spatiotemporal patterns of US drought awareness. *Palgrave Communications*, 5(1), Article 1. <https://doi.org/10.1057/s41599-019-0317-7>
- Koss, Patricia, et M. Sami Khawaja. 2001. « The Value of Water Supply Reliability in California:: A Contingent Valuation Study ». *Water Policy* 3(2): 165-74.
- Lambert, Diane. 1992. « Zero-Inflated Poisson Regression, With an Application to Defects in Manufacturing ». *Technometrics* 34(1): 1-14.
- Lancaster, Kelvin J. 1966. « A New Approach to Consumer Theory ». *Journal of Political Economy* 74(2): 132-57.
- Lanza, Stephanie T., et Brittany L. Rhoades. 2013. « Latent Class Analysis: An Alternative Perspective on Subgroup Analysis in Prevention and Treatment ». *Prevention Science* 14(2): 157-68.
- Latinopoulos, Dionysis. 2014. « Using a choice experiment to estimate the social benefits from improved water supply services ». *Journal of Integrative Environmental Sciences* 11(3-4): 187-204.
- Lima, Maria Luisa. 2004. « On the Influence of Risk Perception on Mental Health: Living near an Incinerator ». *Journal of Environmental Psychology* 24(1): 71-84.
- Lima, Maria Luisa, Julie Barnett, et Jorge Vala. 2005. « Risk Perception and Technological Development at a Societal Level ». *Risk Analysis: An Official Publication of the Society for Risk Analysis* 25(5): 1229-39.
- Louviere, Jordan J., et David A. Hensher. 1982. « DESIGN AND ANALYSIS OF SIMULATED CHOICE OR ALLOCATION EXPERIMENTS IN TRAVEL CHOICE MODELING ». *Transportation Research Record* (890). <https://trid.trb.org/view/189334> (6 juillet 2023).
- Louviere, Jordan J., et George Woodworth. 1983. « Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data ». *Journal of Marketing Research* 20(4): 350-67.
- Luís, Sílvia, Christin-Melanie Vauclair, et Maria Luísa Lima. 2018. « Raising awareness of climate change causes? Cross-national evidence for the normalization of societal risk perception of climate change ». *Environmental Science & Policy* 80: 74-81.
- MacDonald, Darla Hatton et al. 2005. « Using a Choice Modelling Approach for Customer Service Standards in Urban Water1 ». *JAWRA Journal of the American Water Resources Association* 41(3): 719-28.
- Manski, Charles F. 1977. « The structure of random utility models ». *Theory and decision* 8(3): 229.
- Mariel, Petr et al. 2021. *Environmental Valuation with Discrete Choice Experiments: Guidance on Design, Implementation and Data Analysis*. Springer Nature. <https://library.oapen.org/handle/20.500.12657/43295> (30 août 2023).

- Martin-Ortega, Julia, Giacomo Giannoccaro, et Julio Berbel. 2011. « Environmental and Resource Costs Under Water Scarcity Conditions: An Estimation in the Context of the European Water Framework Directive ». *Water Resources Management* 25(6): 1615-33.
- McFadden, Daniel. 1973. « Conditional logit analysis of qualitative choice behavior ».
- Mitchell, Robert Cameron, et Richard T. Carson. 1989. « Using Surveys to Value Public Goods: The Contingent Valuation Method - Robert Cameron Mitchell, Richard T. Carson - Google Livres ». [https://books.google.fr/books?hl=fr&lr=&id=1R75c1UxVE0C&oi=fnd&pg=PR17&dq=\(12\)+Mitchell+R.C.+et+Carson+R.T.+\(1989\),+Using+surveys+to+value+public+goods:+The+contingent+valuation+method,+Washington+D.C.:+Resources+for+the+Future,+463+pages.&ots=78Yp1aYc5o&sig=3WRt4FUivQioNSDcRYY6ezN5Nr0&redir_esc=y#v=onepage&q&f=false](https://books.google.fr/books?hl=fr&lr=&id=1R75c1UxVE0C&oi=fnd&pg=PR17&dq=(12)+Mitchell+R.C.+et+Carson+R.T.+(1989),+Using+surveys+to+value+public+goods:+The+contingent+valuation+method,+Washington+D.C.:+Resources+for+the+Future,+463+pages.&ots=78Yp1aYc5o&sig=3WRt4FUivQioNSDcRYY6ezN5Nr0&redir_esc=y#v=onepage&q&f=false) (31 mai 2023).
- Moriña, David, Pedro Puig, et Albert Navarro. 2021. « Analysis of zero inflated dichotomous variables from a Bayesian perspective: application to occupational health ». *BMC Medical Research Methodology* 21(1): 277.
- Mullahy, John. 1986. « Specification and Testing of Some Modified Count Data Models ». *Journal of Econometrics* 33(3): 341-65.
- Muller, Pierre. 2005. « Esquisse d'une théorie du changement dans l'action publique. Structures, acteurs et cadres cognitifs ». *Revue française de science politique* 55(1): 155-87.
- Offner, Jean-Marc. 2006. « Les territoires de l'action publique locale. Fausses pertinences et jeux d'écart ». *Revue française de science politique* 56(1): 27-47.
- Ogien, Albert. 1995. « L'esprit gestionnaire(une analyse de l'air du temps) ». *Recherches d'histoire et de sciences sociales*.
- Price, James I., Patrick R. Lloyd-Smith, Diane P. Dupont, et Wiktor L. Adamowicz. 2019. « Floods and Water Service Disruptions: Eliciting Willingness-to-Pay for Public Utility Pricing and Infrastructure Decisions ». *Water Economics and Policy* 05(02): 1850021.
- Provencher, Bill, Kenneth A. Baerenklau, et Richard C. Bishop. 2002. « A Finite Mixture Logit Model of Recreational Angling with Serially Correlated Random Utility ». *American Journal of Agricultural Economics* 84(4): 1066-75.
- Région Wallonne, et SWDE. 2012. *Contrat de gestion 2012-2017*.
- . 2022. *Contrat de gestion 2022-2027*.
- Eddy Renaud, Alain Husson, Patrick Eisenbeis, Bruno de Grissac. 2022. "Impacts de la densité urbaine sur les pertes en eau des réseaux d'eau potable". *Techniques Sciences Méthodes*, 2022, 1/2, pp.53-63. {10.36904/tsm/202201053}. {hal-03720491}
- Richardson, Brad, John Sorensen, et E. Jon Soderstrom. 1987. « Explaining the Social and Psychological Impacts of a Nuclear Power Plant Accident1 ». *Journal of Applied Social Psychology* 17(1): 16-36.
- Rodriguez, Nicolas et al. 2021. *Gestion Patrimoniale Multi-Echelles des réseaux d'eau potable (GePaME) - Rapport intermédiaire 2021*. <https://hal.inrae.fr/hal-03836820> (29 août 2023).

- Roeder, Kathryn, Kevin G. Lynch, et Daniel S. Nagin. 1999. « Modeling Uncertainty in Latent Class Membership: A Case Study in Criminology ». *Journal of the American Statistical Association* 94(447): 766-76.
- Rossi, G, V Nicolosi, et A Cancelliere. 2008. « Recent Methods and Techniques for Managing Hydrological Droughts ». (80).
- Salais, Robert. 2010. « Usages et mésusages de l'argument statistique : le pilotage des politiques publiques par la performance ». *Revue française des affaires sociales* (1-2): 129-47.
- Schwalm, Christopher R. et al. 2017. « Global Patterns of Drought Recovery ». *Nature* 548(7666): 202-5.
- Shao, Wanyun, Jonghun Kam, et Evan Cass. 2023. « Public Awareness and Perceptions of Drought: A Case Study of Two Cities of Alabama ». *Risk, Hazards & Crisis in Public Policy* 14(1): 27-44.
- Silva, S. A., & Lima, M. L. (1997). Positive illusions related to dam risks.
- Sinisi, L, et R Aertgeerts. 2011. « Guidance on Water Supply and Sanitation in Extreme Weather Events ».
- Slovic, Paul. 1987. « Perception of Risk ». *Science* 236(4799): 280-85.
- Snowball, J.d., K.g. Willis, et C. Jeurissen. 2008. « Willingness to Pay for Water Service Improvements in Middle-Income Urban Households in South Africa: A Stated Choice Analysis ». *South African Journal of Economics* 76(4): 705-20.
- Stroobants, Marcelle. 2009. *Sociologie du travail*. Armand Colin.
- Swait, Joffre. 1994. « A structural equation model of latent segmentation and product choice for cross-sectional revealed preference choice data ». *Journal of Retailing and Consumer Services* 1(2): 77-89.
- SWDE. 2014. « Eaux plurielles. 100 ans d'une histoire singulière. »
- . 2019. *Rapport annuel d'activité*. <https://rapportannuelswde.be/>.
- . 2020. « Organisation géographique ».
- . 2022. *Rapport annuel d'activité*. <https://rapportannuelswde.be/>.
- . Date inconnue. « De l'eau pour tous aujourd'hui et demain. Notre stratégie 2022-2027. »
- Syuhada, C. I. Nur, K. Mahirah, et M. A. Roseliza. 2020. « Dealing with attributes in a discrete choice experiment on valuation of water services in East Peninsular Malaysia ». *Utilities Policy* 64: 101037.
- Tapsuwan, Sorada, Stephen Cook, et Magnus Moglia. 2018. « Willingness to Pay for Rainwater Tank Features: A Post-Drought Analysis of Sydney Water Users ». *Water* 10(9): 1199.
- Thibaut, Kevin, et Pierre Ozer. 2021. « Les sécheresses en Wallonie, un nouveau défi du changement climatique ? Quelques pistes pour améliorer la gestion de ce phénomène. » *Geo-Eco-Trop: Revue Internationale de Géologie, de Géographie et d'Écologie Tropicales* 45(3). <https://orbi.uliege.be/handle/2268/266096> (6 juillet 2023).

- Train, Kenneth. 2003. *Discrete Choice Methods with Simulation*. New York: Cambridge University Press.
- Van Loon, Anne F. et al. 2016. « Drought in a Human-Modified World: Reframing Drought Definitions, Understanding, and Analysis Approaches ». *Hydrology and Earth System Sciences* 20(9): 3631-50.
- Wang, Jia, Jiaju Ge, et Zhifeng Gao. 2018. « Consumers' Preferences and Derived Willingness-to-Pay for Water Supply Safety Improvement: The Analysis of Pricing and Incentive Strategies ». *Sustainability* 10(6): 1704.
- Willis, Kenneth G., Riccardo Scarpa, et Melinda Acutt. 2005. « Assessing Water Company Customer Preferences and Willingness to Pay for Service Improvements: A Stated Choice Analysis ». *Water Resources Research* 41(2).
<https://onlinelibrary.wiley.com/doi/abs/10.1029/2004WR003277> (6 juillet 2023).
- Wilson, Clevo et al. 2021. « Willingness to Pay to Ensure a Continuous Water Supply with Minimum Restrictions ». *Empirical Economics* 61(3): 1519-37.