

ANALYSE DE LA CORRÉLATION ENTRE FRÉQUENCES DES SURVERSES ET PLUVIOMÉTRIE

Alain Mailhot, Guillaume Talbot

INRS-Eau, Terre et Environnement

Travail réalisé pour le
Ministère du Développement Durable, de l'Environnement, de la Faune et des Parcs

Rapport de recherche R1500

Version finale

3 mars 2014

Table des matières

Liste des figures	v
Liste des tableaux	vii
1 Description du mandat	1
2 Données sur les surverses : <i>Suivi des ouvrages municipaux d'assainissement des eaux (SOMAE)</i>	3
2.1 Modes d'enregistrement des surverses et données disponibles	4
2.2 Traitement des données manquantes et erronées	8
2.3 Données disponibles sur les surverses et fréquences des visites	8
3 Données pluviométriques utilisées	11
4 Modèle statistique d'occurrence des surverses	13
4.1 Estimation de la probabilité de jour avec surverse	13
5 Association entre pluies et surverses	17
5.1 Association ODS et stations météorologiques	17
5.2 Liens entre probabilité d'occurrence des surverses et seuil des pluies	18
5.3 Table de contingences des dépassements des seuils de pluies et d'occurrences de surverse	19
5.4 Procédure utilisée pour la comparaison des séries de pluie et de surverses	21
5.5 Estimation du pouvoir prévisionnel du modèle basé sur les dépassements des seuils de pluie	22
5.6 Impact de la distance entre station météorologique et ODS sur la performance du modèle basé sur le seuil de pluie	27
6 Conclusion	33
Références	37

Liste des figures

2.1	Carte localisant les stations d'épuration et les stations météorologiques considérées.	4
2.2	Histogramme du nombre d'ODS par station d'épuration (la dernière barre à droite du graphique regroupe tous les cas de stations d'épuration ayant plus de 60 ODS; voir le Tableau 2.1 pour une liste de ces stations).	5
2.3	Histogramme du nombre d'année de données disponibles sur les surverses aux différents ODS.	9
2.4	Histogramme du nombre moyen de visites par semaine aux différents ODS.	9
3.1	Histogramme du nombre d'années de données disponibles aux différentes stations météorologiques.	12
4.1	Histogramme du nombre annuel moyen de jours avec surverses aux différents ODS (la barre à l'extrême droite correspond à la fraction d'ODS ayant plus de 40 surverses par année en moyenne).	16
5.1	Histogramme des distances entre stations météorologiques et séries de surverses aux différents ODS (la barre à droite correspond aux cas de stations météorologiques situées à plus de 100 km de la station d'épuration; voir le Tableau 5.1 pour la liste de ces stations).	18
5.2	Histogrammes du nombre annuel moyen de jour avec surverse aux différents ODS pour les cas où : (a) $0 \geq m < 1$; (b) $m \geq 1$ et pouvoir prévisionnel non significatif à 99%; et (c) $m \geq 1$ et pouvoir prévisionnel significatif à 99%.	26
5.3	Exemple de valeurs du ORSS pour la station Jean-R.-Marcotte à Montréal (station 00065-1, ODS 63) et de la station de la Malbaie (station 15012-1, ODS 7) pour les différentes durées considérées.	27
5.4	Histogramme des durées optimales aux différents ODS correspondant à la durée avec la valeur du ORSS la plus élevée (cas $m \geq 1$).	28
5.5	Histogramme du nombre d'années communes entre séries aux stations pluviométriques et séries de surverses aux différents ODS lorsque la durée maximisant la valeur du ORSS est considérée à chaque ODS.	28

5.6	Valeurs du ORSS aux différents ODS en fonction du nombre annuel moyen de jours avec surverses pour les cas où $m \geq 1$. Les durées maximisant les valeurs du ORSS à chaque ODS sont considérées.	29
5.7	Histogramme des valeurs du ORSS aux différents ODS pour les durées maximisant les valeurs du ORSS (cas $m \geq 1$).	30
5.8	Nombre total de jour avec surverses estimé à partir des archives sur les surverses en fonction du nombre de jour avec dépassement du seuil de pluie à chaque ODS. Les durées maximisant les valeurs de HSS à chaque ODS sont considérées.	30
5.9	Box plots des valeurs du ORSS en fonction de la distance entre station météorologique et ODS (cas $m \geq 1$).	31
5.10	Fraction du nombre d'ODS où le pouvoir prévisionnel du modèle basé sur le seuil de pluie est non significatif en fonction de la distance entre station météorologique et ODS (cas $m \geq 1$).	31

Liste des tableaux

2.1	Liste des stations d'épuration ayant plus de 60 ODS associés.	6
2.2	Nombre d'ODS utilisant les différents modes d'enregistrement des surverses.	7
2.3	Nombre d'ODS dans le cas où plusieurs modes d'enregistrement ont été utilisés.	7
5.1	Stations d'épuration situées à plus de 100 km de stations météorologiques. .	19
5.2	Valeurs minimales du <i>Odds Ratio Skill Score</i> (ORSS) nécessaires pour avoir un pouvoir prévisionnel significatif à divers niveaux de confiance en fonction de la valeur de m (voir Tableau 5.4)	23
5.3	Nombre d'ODS et pourcentages du nombre total d'ODS appartenant à chacun des cas recensés lors de l'estimation du caractère significatif ou non de la valeur du ORSS.	25

Chapitre 1

Description du mandat

Dans le cadre de la mise en place de la *Stratégie pan-canadienne sur la gestion des effluents d'eaux usées municipales*, le *Ministère du Développement Durable, de l'Environnement, de la Faune et des Parcs* (MDDEFP) a développé une approche afin d'assister les directions régionales dans la mise en place et le suivi de cette stratégie. Cette approche permet : 1) d'établir un lien entre les fréquences de surverses et les caractéristiques de la pluie ; 2) vérifier l'impact sur les fréquences de surverses de diverses actions et modifications telles qu'une variation du débit de ruissellement moyen (suite au développement urbain), une réduction des eaux parasites (suite à une réfection du réseau), une augmentation des débits interceptés (modification des ouvrages d'interception), une variation des surfaces imperméables et l'ajout de bassin de rétention. Toute modification susceptible d'augmenter les volumes de ruissellement ou les volumes acheminés en réseau jusqu'aux ouvrages de surverse (ODS) risque d'entraîner une augmentation de la fréquence de surverse.

Il est important dans le contexte de mise en place de cette stratégie de déterminer les fréquences historiques des surverses à partir des données archivées et d'établir une relation entre cette fréquence et les données pluviométriques. Le présent mandat avait pour objectif de développer et d'appliquer une approche statistique permettant d'établir un lien entre la fréquence des surverses et la pluviométrie afin d'établir, à terme, une relation entre la fréquence des surverses et la fréquence des pluies. Tous les ODS dont les données sont disponibles dans la base de données *Suivi des ouvrages municipaux d'assainissement des eaux* (SOMAE) doivent être considérées.

L'approche retenue comporte deux étapes. La première consiste à développer et appliquer un modèle statistique simple permettant d'estimer la probabilité quotidienne d'occur-

rence de surverses à chaque ODS à partir des données d'archives de surverses de la base SOMAE. La seconde étape vise à comparer les séries de pluies disponibles aux stations météorologiques à proximité des ODS avec les archives de surverses afin de vérifier que les surverses surviennent lorsque la pluie totale sur une durée donnée dépasse un certain seuil.

Le présent rapport est structuré comme suit. Le Chapitre 2 présente une description des données sur les surverses (modes d'enregistrements, types de données disponibles, etc.) alors que le Chapitre 3 donne une description des données de précipitations disponibles. Le Chapitre 4 décrit le modèle statistique d'occurrence des surverses utilisé et présente les résultats de l'application du modèle statistique d'occurrence des pluies aux différents ODS. L'approche permettant d'estimer les seuils de pluie au delà de laquelle il y a surverse est décrite au Chapitre 5. La comparaison des séries de pluies et de surverses est également discutée à ce chapitre, ce qui inclut la présentation de l'indice (*score*) utilisé pour quantifier la performance du modèle liant les dépassements d'un seuil de pluie à l'occurrence de surverses. Finalement le Chapitre 6 présente un résumé des travaux et des principaux résultats ainsi que quelques perspectives de travaux futurs.

Il est important de noter qu'un fichier Excel joint au présent rapport donne une description détaillée des valeurs des différents variables et indices estimés dans le cadre de ces travaux pour l'ensemble des 4285 ODS à l'étude.

Chapitre 2

Données sur les surverses : *Suivi des ouvrages municipaux d'assainissement des eaux* (SOMAE)

Les données sur les surverses utilisées dans le cadre du présent projet proviennent de la base de données SOMAE (*Suivi des ouvrages municipaux d'assainissement des eaux*) qui est sous la responsabilité du *Ministère des Affaires Municipales, des Régions et de l'Occupation du Territoire* (MAMROT). Ce ministère est en effet en charge de recueillir les informations relatives au fonctionnement des ouvrages municipaux d'assainissement des eaux (stations d'épuration et ouvrages de débordement en réseau) afin de faire le suivi de ces ouvrages et de s'assurer qu'ils respectent les exigences environnementales prescrites. Les opérateurs des ouvrages des différentes municipalités sont tenus de transmettre ces informations au MAMROT. Concernant plus spécifiquement les ouvrages de surverse, les opérateurs doivent transmettre les données relatives aux dates de visites et d'occurrence des surverses aux différents ODS sur leur territoire.

Les ODS sont regroupés en fonction des stations d'épuration auxquelles ils sont rattachés. Seules les localisations (latitude, longitude) des stations d'épuration sont disponibles. Pour un ODS donné, les informations disponibles dans les archives des surverses de la base SOMAE sont : les dates des visites, le mode d'enregistrement, la présence ou l'absence du repère et la durée totale de surverse dans le cas des enregistreurs (voir section 2.1). Les données transmises par le MDDEFP dans le cadre du présent projet concernait un

total de 4 285 ODS. Le nombre correspondant de stations d'épurations est de 512. La carte de la Figure 2.1 localise l'ensemble de ces stations d'épuration de même que les stations météorologiques utilisées dans le cadre de ce projet (voir Chapitre 3 pour une description de ces données). Le nombre d'ODS par station d'épuration varie de 1 à 158 (2.2). Les stations d'épuration comportant plus de 60 ODS sont regroupées au Tableau 2.1.

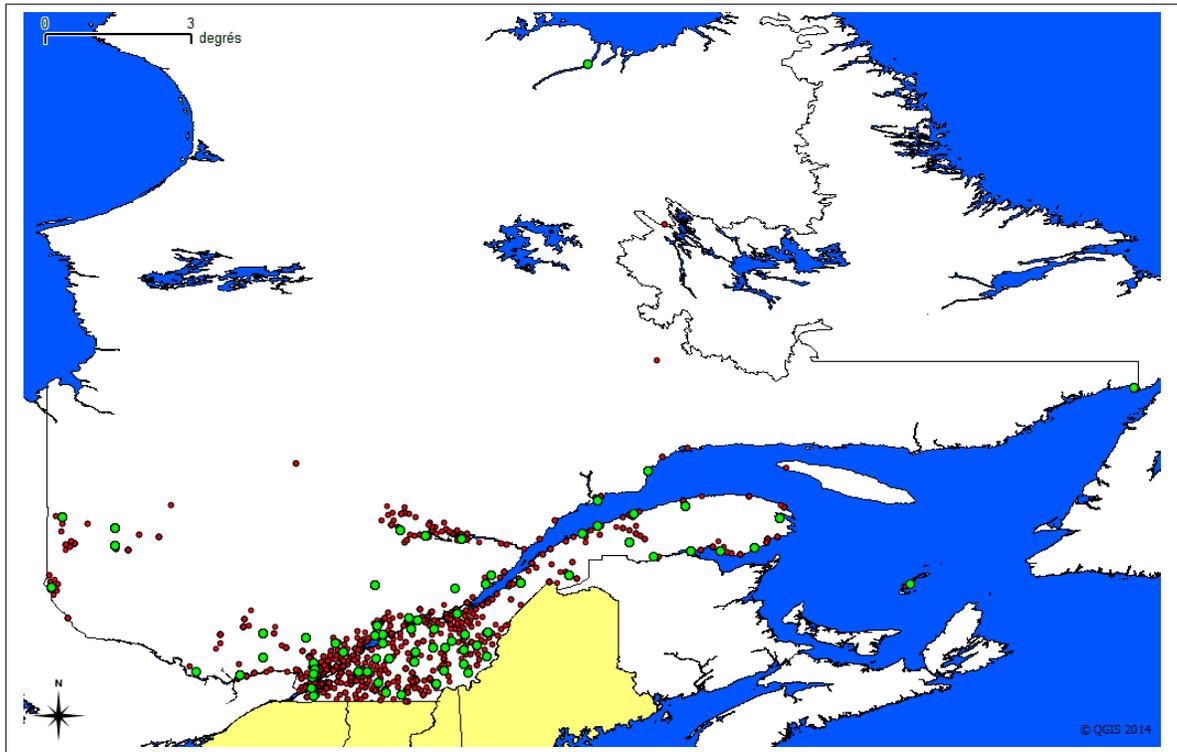


Figure 2.1: Carte localisant les stations d'épuration et les stations météorologiques considérées.

La section 2.1 aborde la question des modes d'enregistrements et des types d'enregistreurs des surverses et la section 2.2 décrit les données utilisées et donne un portrait global de la situation en termes d'ODS et de surverses.

2.1 Modes d'enregistrement des surverses et données disponibles

Trois modes d'enregistrement des surverses sont utilisés : les repères, les enregistreurs en continu et les enregistreurs automatiques. Les repères consistent en un objet déposé

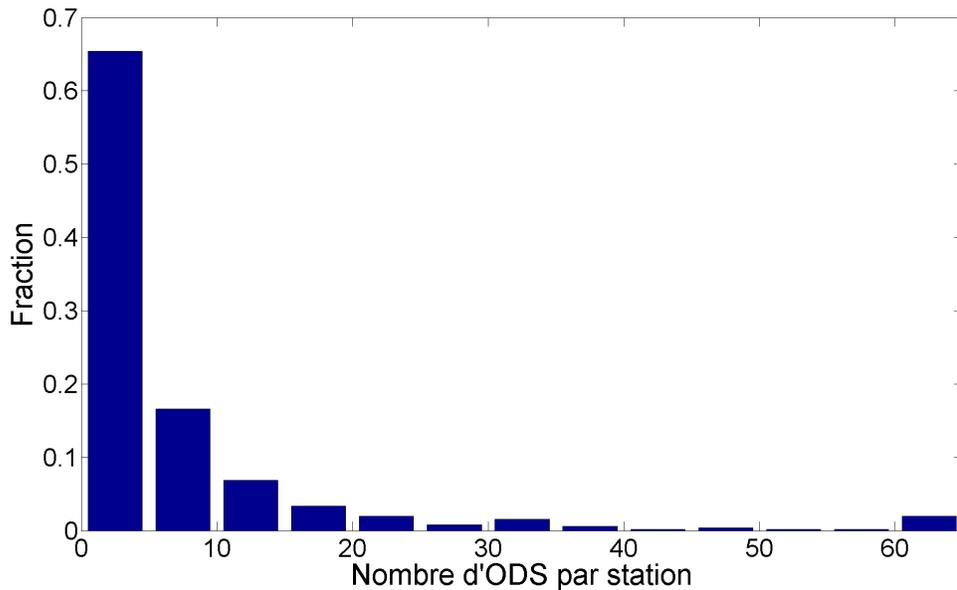


Figure 2.2: Histogramme du nombre d'ODS par station d'épuration (la dernière barre à droite du graphique regroupe tous les cas de stations d'épuration ayant plus de 60 ODS ; voir le Tableau 2.1 pour une liste de ces stations).

sur l'ouvrage de surverse qui se déplace lorsqu'il y a surverse (ce repère est retenu à l'ouvrage par une corde). L'opérateur visite régulièrement l'ODS et rapporte le déplacement ou non du repère à chacune de ses visites. Advenant son déplacement, il replace l'objet sur l'ouvrage. L'information ainsi fournie indique l'occurrence ou la non occurrence d'une ou plusieurs surverses durant la période depuis sa visite précédente sans indication précise sur le nombre de surverses et le nombre de jours où il y a eu surverse. Il est à noter que les visites de l'opérateur sont généralement hebdomadaires pour ce mode d'enregistrement. La visite hebdomadaire peut cependant s'effectuer à différent jour d'une semaine à l'autre. Les données enregistrées dans la base SOMAE sont, pour ce mode d'enregistrement, la date de la visite et l'observation du déplacement ou non du repère lors de la visite.

Les deux autres modes d'enregistrement sont instrumentés et consistent en des dispositifs permettant de mesurer la durée totale de surverses. Dans le cas des enregistreurs en continu, l'appareil mesure le nombre total d'heures de surverses entre deux relevés consécutifs. Lors de sa visite, l'opérateur note ce total et réinitialise l'appareil. La fréquence des enregistrements est donc conditionnée dans ce cas par la fréquence des relevés et des

Table 2.1: Liste des stations d'épuration ayant plus de 60 ODS associés.

Numéro de la station d'épuration	Nom de la station	Nombre d'ODS
00020-1	Québec (Est)	124
00020-2	Québec (Ouest)	144
00065-1	Montréal (Station Jean-R.-Marcotte)	158
00079-1	Gatineau	93
36350-1	Sherbrooke	101
41720-1	Drummondville	83
50730-1	Sorel-Tracy	61
56650-1	Longueuil	92
64500-3	Laval (Lapinière)	124
94300-1	Saguenay (Chicoutimi)	100

visites de l'opérateur. Les données archivées dans la base SOMAE sont, dans ce cas, les dates des visites et le nombre total d'heures de surverses depuis la visite précédente. Tout comme pour les repères, dans le cas de visites espacés de plusieurs jours, il n'est pas possible de savoir le nombre de surverses et le nombre exact de jours avec surverses.

Dans le cas d'enregistreur automatique, l'opérateur n'a pas à réinitialiser l'appareil et les cumuls enregistrés sont quotidiens. Les données disponibles sont alors le nombre total d'heures avec surverses chaque jour. Cependant il convient de noter que la période de 24 heures considérée pourra différer selon les ODS (communication personnelle de M. João Moreira de la Direction des infrastructures du MAMROT). Si dans plusieurs cas, on peut penser que la période couverte sera de minuit à minuit, il pourrait arriver qu'à d'autres ODS la période considérée soit, par exemple, de midi à midi.

L'interprétation des relevés est donc différente dans les trois cas. Dans le cas d'enregistreur automatique, les relevés peuvent être interprétés comme des indications de l'occurrence ou de la non occurrence de surverses chaque jour (la définition de "jour" pouvant cependant varier). Les relevés des enregistreurs en continu doivent être interprétés comme des indications d'occurrence de surverse durant la période entre deux relevés sans précision sur le nombre de jours où des surverses ont eu lieu si cette période entre deux visites couvre plusieurs jours. De même, les relevés des repères indiquent l'occurrence d'au moins un jour avec surverse durant la période entre deux relevés.

Le Tableau 2.2 présente le nombre d'ODS sous étude fonctionnant selon les différents

Table 2.2: Nombre d'ODS utilisant les différents modes d'enregistrement des surverses.

Mode d'enregistrements	Nombre	Pourcentage
Bloc repère	1944	45
Enregistreur en continu	368	4
Enregistreur quotidien	179	9
Combinaison	1757	41
Inconnu	37	1

Table 2.3: Nombre d'ODS dans le cas où plusieurs modes d'enregistrement ont été utilisés.

Modes d'enregistrement	Nombre	Pourcentage
Bloc repère et enregistreur en continu	1512	86.0
Bloc repère et enregistreur quotidien	207	11.8
Enregistreur en continu et enregistreur quotidien	8	1.7
Bloc repère, enregistreur en continu et enregistreur quotidien	30	0.5

modes d'enregistrement. Plusieurs ODS ont cependant connu plusieurs modes d'enregistrements au fil du temps. Le Tableau 2.3 donne la répartition des divers cas de figures dans ce dernier cas.

Les dates où les modes d'enregistrement ont été modifiés aux ODS du Tableau 2.3 sont cependant inconnues. Dans le cas des transitions d'un enregistrement à l'aide de bloc repère à des enregistreurs en continu ou automatique (ces cas représentent 97,8 % des cas répertoriés), cette date peut être estimée à partir des fichiers disponibles. En effet, dans ces cas, lors de la transition du mode bloc repère au mode enregistreur, les valeurs de la colonne "durée" des fichiers de données des ODS passent d'une valeur "NULL" à une valeur numérique. Il n'est cependant pas possible d'estimer ces dates dans les cas où l'on passe d'un enregistreur continu à un enregistreur automatique.

Les informations utilisées dans le cadre de la présente étude sont les dates des visites et la donnée relative à l'occurrence ou non d'au moins une surverse pendant la période depuis la visite précédente (les informations sur la durée des surverses n'ont pas été utilisées). Bien que l'interprétation de ces données diffère selon le mode d'enregistrement, le traitement de ces données demeurent le même quelque soit le mode d'enregistrement (ce point est discuté plus avant au Chapitre 4).

2.2 Traitement des données manquantes et erronées

Plusieurs sources d'erreur peuvent affecter les données sur les surverses. Les plus vraisemblables sont les erreurs de saisie, très difficiles à identifier. La fréquence des visites des opérateurs aux ODS peuvent également être irrégulières sans que cela n'ait de réel impact autre qu'une saisie plus irrégulière des données dans la base SOMAE. Les seules données manquantes qui peuvent être identifiées sont celles associées aux enregistreurs automatiques. Dans ce cas, la saisie des données devraient être journalières. Or il appert que certains jours sont manquants. Dans ce cas, la valeur enregistrée le jour suivant une séquence de jours non saisie est interprétée comme représentative de la période manquante. Ainsi si l'occurrence d'une surverse est indiquée, cela signifie qu'une ou plusieurs surverses est survenue pendant la période manquante. De même l'absence de surverse signifie qu'il n'y a pas eu de surverse pendant la période manquante (cette interprétation nous a été suggérée par M. João Moreira de la Direction des infrastructures du MAMROT).

2.3 Données disponibles sur les surverses et fréquences des visites

La Figure 2.3 présente l'histogramme du nombre d'années d'enregistrements des surverses aux différents ODS. Les historiques disponibles sont compris entre une et 12 années. Moins de 5 % des ODS ont un historique de surverses d'une année alors que plus de 60 % ont un historique de 12 ans.

La Figure 2.4 présente l'histogramme du nombre moyen de visite par semaine aux différents ODS. Comme on peut le constater, une majorité d'ODS est visitée quotidiennement ou possède des enregistrements quotidiens (valeurs près de sept) et plusieurs font l'objet de visites hebdomadaires (valeurs près de un).

Des 4285 ODS répertoriés, 767 ne présentent aucune surverse enregistrée pendant la période archivée.

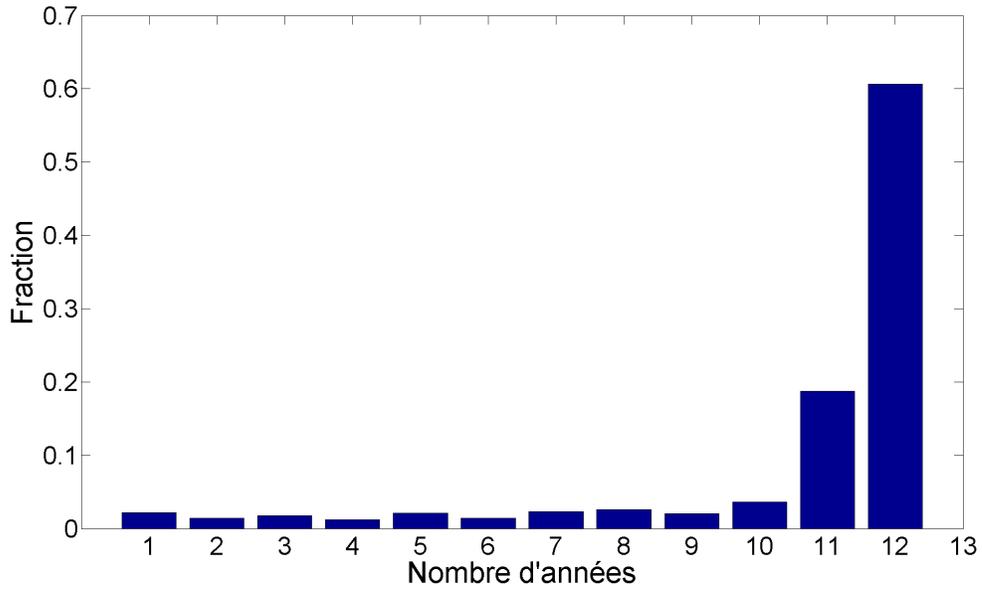


Figure 2.3: Histogramme du nombre d'année de données disponibles sur les surverses aux différents ODS.

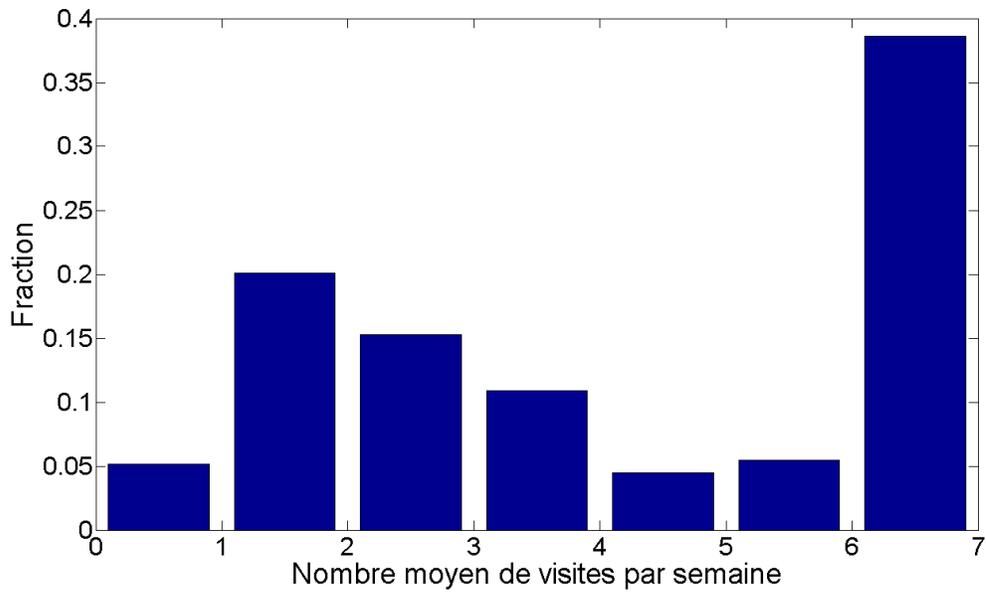


Figure 2.4: Histogramme du nombre moyen de visites par semaine aux différents ODS.

Chapitre 3

Données pluviométriques utilisées

Les données sur les précipitations utilisées proviennent des stations opérées par le *Ministère du Développement Durable, de l'Environnement, de la Faune et des Parcs* (MDDEFP) et d'*Environnement Canada* (EC). Ces données correspondent aux maxima journaliers pour les durées 5, 10, 15, 30 minutes, 1, 2, 4, 6, 12 heures. Les séries journalières de pluies aux stations du MDDEFP sont également utilisées. Seule la période de mai à octobre est considérée puisque les stations enregistrant les maxima journaliers ne sont généralement pas en opération pendant les mois d'hiver (la mention *annuel* dans ce qui suit fera référence, à moins d'indication contraire, à la période de mai à octobre). La Figure 3.1 présente l'histogramme du nombre d'années équivalentes de données disponibles aux différentes stations météorologiques. Ce nombre d'années équivalentes est estimé en faisant la somme du nombre de jours de données disponibles divisée par 182 (nombre de jours de la période de mai à octobre).

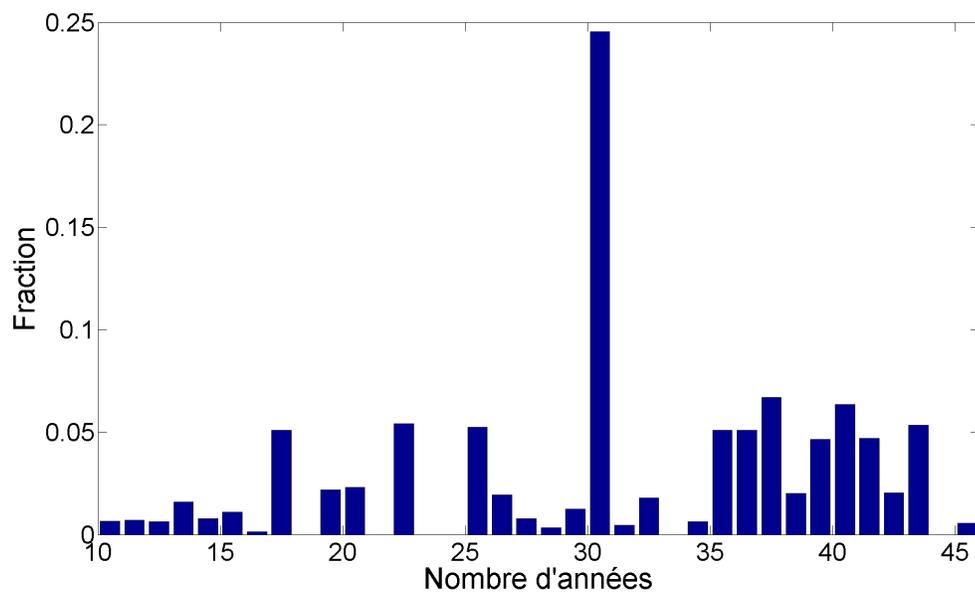


Figure 3.1: Histogramme du nombre d'années de données disponibles aux différentes stations météorologiques.

Chapitre 4

Modèle statistique d'occurrence des surverses

L'objectif est d'estimer la probabilité d'occurrence de jour sans surverse, θ , ou encore la fraction de jours sans surverse durant une période donnée (p. ex. du 1er mai au 31 octobre). Pour ce faire, on émet l'hypothèse que chacun de ces événements (occurrence ou non occurrence de surverse) pour chacun des jours de la période mai à octobre est statistiquement indépendant. Ce faisant la probabilité d'avoir k jours sans surverse durant les 184 jours de la période mai-octobre est donnée par la loi binomiale de paramètre θ . Ainsi la probabilité d'avoir k jours sans surverse sur une période de n jours, $p_n(k)$, est donné par :

$$p_n(k) = \binom{n}{k} \theta^k (1 - \theta)^{n-k} \quad (4.1)$$

La valeur moyenne du nombre de jours sans surverse est $n \theta$ et la variance $n \theta (1 - \theta)$.

4.1 Estimation de la probabilité de jour avec surverse

Rappelons que θ correspond à la probabilité d'avoir un jour sans surverse. On désigne par (l_i, k_i) les données des archives de surverses où l_i correspond à la date du i ème relevé et où $k_i = 1$ si une surverse est rapportée et $k_i = 0$ si aucune surverse n'est rapportée. On considère que l'occurrence des surverses est décrit par la loi de Bernoulli (probabilité $1 - \theta$ que $k = 1$, c'est-à-dire qu'il y ait surverse au jour considéré et probabilité θ que $k = 0$,

c'est-à-dire qu'il n'y ait pas de surverse au jour considéré). L'estimation de θ se fait en maximisant la fonction de vraisemblance, $p(l_i, k_i|\theta)$, de la séquence enregistrée des relevés qui s'exprime sous la forme :

$$p(l_i, k_i|\theta) = \prod_{\forall i, k_i=0} \theta^{(l_i-l_{i-1})} \prod_{\forall i, k_i=1} [1 - \theta^{(l_i-l_{i-1})}]$$

On définit $\delta_i = (l_i - l_{i-1})$ et on prend le logarithme de part et d'autre de l'équation :

$$\ln p(l_i, k_i|\theta) = \sum_{\forall i, k_i=0} \delta_i \ln \theta + \sum_{\forall i, k_i=1} \ln [1 - \theta^{\delta_i}]$$

On définit $N_0 = \sum_{\forall i, k_i=0} \delta_i$ qui correspond au nombre total de jours des périodes entre deux visites où aucune surverse n'est rapportée lors de la dernière visite. On a alors :

$$\ln p(l_i, k_i|\theta) = N_0 \ln \theta + \sum_{\forall i, k_i=1} \ln [1 - \theta^{\delta_i}]$$

On estime la dérivée par rapport à θ de cette fonction :

$$\frac{d \ln p(l_i, k_i|\theta)}{d\theta} = \frac{N_0}{\theta} - \sum_{\forall i, k_i=1} \delta_i \frac{\theta^{\delta_i-1}}{[1 - \theta^{\delta_i}]}$$

On pose l'égalité à zéro et on a :

$$\sum_{\forall i, k_i=1} \delta_i \frac{\theta^{\delta_i}}{[1 - \theta^{\delta_i}]} = N_0$$

On peut réécrire cette dernière expression sous la forme :

$$\sum_{k=1}^K N_k \frac{k\theta^k}{[1 - \theta^k]} - N_0 = 0 \quad (4.2)$$

où N_k est le nombre de visites séparées de k jours où une surverse est enregistrée lors de la deuxième visite et K est la valeur maximale des archives disponibles du nombre de jours séparant deux visites avec une surverse enregistrée lors de la deuxième visite. Une solution numérique de l'équation 4.2 doit être recherchée dans le cas général.

Dans le cas particulier où les relevés sont quotidiens, on a $\delta_i = 1, \forall i$ et donc :

$$\sum_{\forall i, k_i=1} \frac{N_1 \theta}{[1 - \theta]} = N_0$$

et donc :

$$\theta = \frac{N_0}{(N_0 + N_1)} \quad (4.3)$$

et N_0 correspond alors au nombre de jours sans surverse et N_1 au nombre de jours avec surverses.

Dans le cas un peu plus général où les relevés se font à tous les n jours (p. ex. à tous les sept jours dans le cas hebdomadaire), la solution est donnée par :

$$\theta = \left[\frac{N_0}{N_0 + nN_1} \right]^{\frac{1}{n}}$$

et N_0 correspond alors au nombre de relevés sans surverse et N_1 au nombre de relevés avec surverses.

Une fois estimée la probabilité d'occurrence des surverses à un ODS donné, qui correspond à $(1 - \theta)$, la probabilité d'occurrence d'un nombre k de surverse pendant une période de n jours est donnée par la loi binomiale (Eq. 4.1). La valeur moyenne du nombre de surverse annuel (période de mai à octobre comportant 182 jours) est $182(1 - \theta)$ et l'écart-type associé à cette valeur est $\sqrt{182 \theta (1 - \theta)}$. La Figure 4.1 présente l'histogramme du nombre moyen annuel de surverses aux différents ODS. La très grande majorité des ODS (environ 43 %) compte moins de une surverse en moyenne par année alors qu'environ 2 % présente plus de 40 surverses par année.

Il convient de noter que l'ODS 7 de la station d'épuration 63048-2 (Saint-Lin-Laurentides) a été visité seulement 10 fois entre le 2 mai et le 31 mai 2002. Des surverses ont été rapportées à chacune de ces visites ce qui implique que la probabilité de ne pas avoir de surverse est nulle et que conséquemment le nombre annuel de surverses est 182. Une possible explication dans ce cas est que l'ODS déverse même par temps sec, situation évidemment problématique.

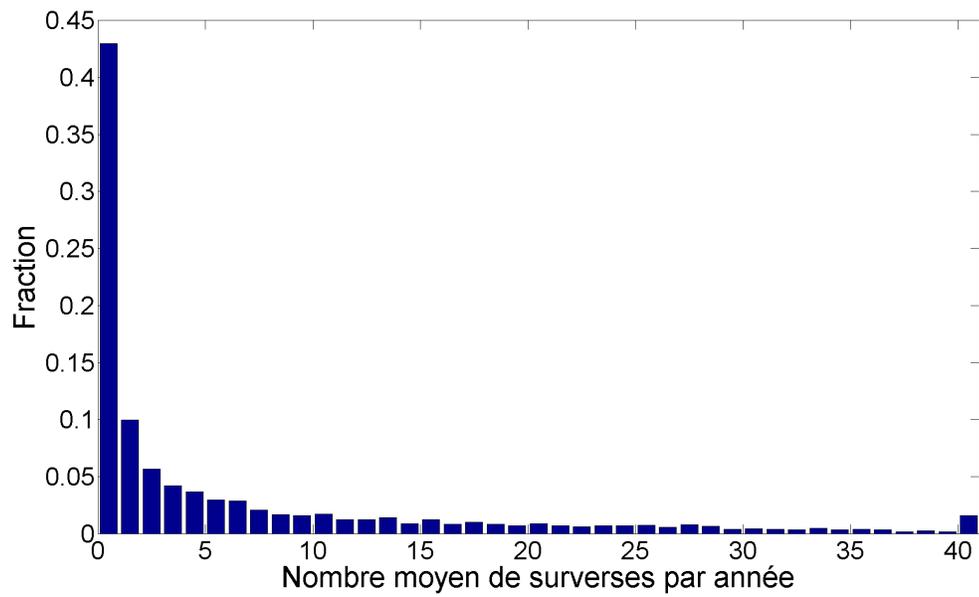


Figure 4.1: Histogramme du nombre annuel moyen de jours avec surverses aux différents ODS (la barre à l'extrême droite correspond à la fraction d'ODS ayant plus de 40 surverses par année en moyenne).

Chapitre 5

Association entre pluies et surverses

Le présent chapitre entend examiner le lien entre les séries pluviométriques et l'occurrence des surverses, l'hypothèse de base voulant qu'il y a surverse lorsque l'intensité de pluie sur une durée donnée dépasse un seuil. Plusieurs durées différentes sont considérées dans ce qui suit afin de voir si une durée spécifique à un ODS (qui pourrait correspondre au temps de concentration du bassin alimentant l'ODS par exemple) ne permet pas d'obtenir de meilleur résultat.

5.1 Association ODS et stations météorologiques

Les latitudes et longitudes des stations d'épuration associées aux différents ODS sont connues. La recherche de la station météorologique s'est donc faite en recherchant celle qui se trouve la plus près de la station d'épuration. Cette façon de procéder implique donc que les séries pluviométriques sont les mêmes pour tous les ODS associés à une même station d'épuration. La station météorologique ainsi identifiée est sélectionnée lorsque la période d'opération de cette dernière recoupe les périodes couvertes par les archives de surverses aux différents ODS (à noter que les séries sur les surverses peuvent couvrir des périodes différentes aux différents ODS). La règle suivante a été adoptée : la station météorologique retenue est celle qui est la plus près de la station d'épuration et qui a minimalement une année commune avec les séries de surverses à chaque ODS associé à cette station d'épuration. Les années communes correspondent aux années où la station météorologie est en opération et où les données d'archives de surverses sont disponibles. Cette association ne tient pas compte des données manquantes.

La Figure 5.1 présente l’histogramme des distances entre les ODS et les stations météorologiques qui leurs sont appariées. La grande majorité des ODS est associée à des stations météorologiques situées à moins de 35 km. Quatre stations d’épuration (situés en région éloignée) sont appariées à des stations météorologiques situées à plus de 100 km. Le Tableau 5.1 dresse la liste de ces ODS.

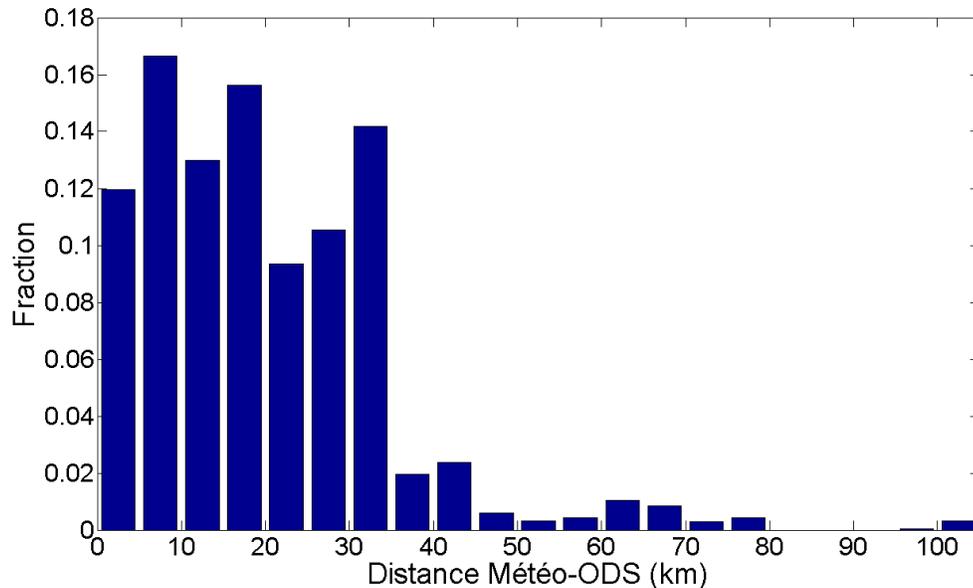


Figure 5.1: Histogramme des distances entre stations météorologiques et séries de surverses aux différents ODS (la barre à droite correspond aux cas de stations météorologiques situées à plus de 100 km de la station d’épuration ; voir le Tableau 5.1 pour la liste de ces stations).

5.2 Liens entre probabilité d’occurrence des surverses et seuil des pluies

On formule l’hypothèse que la probabilité d’occurrence des surverses est déterminée par la probabilité que la pluie dépasse un certain seuil chaque jour. Puisque nous considérons les séries des maxima journaliers de pluie, il faut préciser la durée considérée, l’hypothèse étant que la durée la plus appropriée est déterminée par le temps de concentration des bassins alimentant les différents ODS. Soit $\eta_{d,j}$ la valeur du seuil pour la durée d (p.ex. $d = 15, 30, 60$ minutes, 1, 2, 6, 12 et 24 heures) et l’ODS j , on pose donc :

Table 5.1: Stations d'épuration situées à plus de 100 km de stations météorologiques.

Nom de la station d'épuration	Numéro de la station d'épuration	Distance de la station météorologique (km)	Nombre d'ODS
Fermont	97035-1	253	2
Île-d'Anticosti (Port Menier)	98020-1	116	2
Schefferville	98450-1	379	5
Chibougamau	98200-1	219	5

$$\theta_j = F_d(\eta_{d,j})$$

où F_d est la fonction de répartition des valeurs des maxima journaliers pour la durée d et θ_j la probabilité journalière de ne pas avoir de surverses. On a donc :

$$\eta_{d,j} = F_d^{-1}(\theta_j)$$

La dernière expression établit donc un lien entre la hauteur de pluie et la probabilité de surverses. La hauteur de pluie dont la probabilité de dépassement est égale à la probabilité d'occurrence de surverses est estimée. Cette expression permet donc de déterminer les différents seuils de pluie pour les diverses durées qui généreront des surverses à chaque ODS.

5.3 Table de contingences des dépassements des seuils de pluies et d'occurrences de surverse

Les valeurs des seuils établies à la section précédente sont ensuite utilisées pour vérifier dans quelle mesure le dépassement de ces seuils entraînent l'occurrence de surverse. Considérons l'ODS j et la station météorologique située à proximité que nous identifions également par l'indice j . Les séries des maxima journaliers pour diverses durées T' (15, 30, 60 minutes, 1, 2, 6, 12 et 24 heures) sont disponibles à ces stations. Il faut noter que les durées 5 et 10 minutes, bien que disponibles n'ont pas été considérées dans cette étude. Considérons la période commune d'enregistrement des surverses et des données pluviométriques que nous

désignons par T_j . Un pas de temps journalier est considéré. Soit, pour la période commune T_j , $\{x_i(T')\}_j$ la série des maxima journaliers de pluie à la station j pour la durée T' et $(l_{m'}, k_{m'})_j$ la série des valeurs enregistrées à l'ouvrage de surverse j où $l_{m'}$ est la date du m' ième enregistrement à cet ODS et $k_{m'}$ est la valeur du relevé ($k = 1$ si une surverse est enregistrée alors que $k = 0$ si aucune surverse n'est enregistrée).

Supposons dans un premier temps, que les relevés sur les surverses sont quotidiens, c'est-à-dire que m' est équivalent à i . Dans ce cas, il s'agit de comparer les valeurs quotidiennes des pluies avec la valeur seuil et vérifier que lorsque ce seuil est dépassé, il y a surverse. On définit la table de contingence suivante pour l'ODS j et la durée T' (la notation utilisée pour les éléments de la table de contingence est celle communément utilisée dans la littérature) :

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}_{j,T'}$$

Les éléments de cette table sont définis comme suit :

- a : nombre de jours avec dépassement du seuil de pluie et avec surverse
- b : nombre de jours avec dépassement du seuil de pluie et sans surverse
- c : nombre de jours sans dépassement du seuil de pluie et avec surverse
- d : nombre de jours sans dépassement du seuil de pluie et sans surverse

Dans le cas plus général où les archives de surverses ne sont pas journalières, les éléments de matrice correspondent aux cas suivants :

- a : nombre de visites de l'ODS où le seuil de pluie a été dépassé au moins une fois pendant la période entre la présente visite et la précédente et où une surverse est rapportée lors de la seconde visite
- b : nombre de visites de l'ODS où le seuil de pluie a été dépassé au moins une fois pendant la période entre la présente visite et la précédente et où aucune surverse n'est rapportée lors de la seconde visite
- c : nombre de visites de l'ODS où le seuil de pluie n'a pas été dépassé pendant la période entre la présente visite et la précédente et où une surverse est rapportée lors de la seconde visite

- d : nombre de visites de l'ODS où le seuil de pluie n'a pas été dépassé pendant la période entre la présente visite et la précédente et où aucune surverse n'est rapportée lors de la seconde visite

5.4 Procédure utilisée pour la comparaison des séries de pluie et de surverses

Les archives sur les surverses ne permettent pas de déterminer le moment exact de la journée où il y a surverse tout comme les données de maxima journaliers de pluie n'indiquent pas le moment de la journée où cette pluie s'est produite. Il est donc possible que, dans les cas des enregistreurs en continu et des blocs repère, la valeur de pluie dépasse le seuil après la visite de l'ODS par l'opérateur auquel cas il y aura dépassement de seuil au cours de cette journée sans qu'une surverse ne soit enregistrée cette même journée, la surverse pouvant être enregistrée dans ce cas le lendemain ou lors de la prochaine visite de l'opérateur. La même situation se présente pour les enregistreurs automatiques puisque la période de 24 heures considérée pour ces enregistreurs pourra différer de celle utilisée pour le cumul sur 24 heures de pluie.

La procédure adoptée dans le cadre de la présente étude a été de comparer les données quotidiennes de pluie et l'occurrence des surverses le même jour. Ainsi dans le cas de relevé quotidien, le dépassement du seuil pour le jour considéré est vérifié, comparé à l'occurrence de surverses le même jour et la table de contingence mise à jour en fonction des différents cas de figures. Dans le cas où les visites à l'ODS sont séparées de plusieurs jours, on vérifie que le seuil est dépassé pendant la période depuis la dernière visite (en excluant le jour de cette visite) jusqu'au jour considéré (en incluant ce jour). Cette procédure ne permet donc pas de considérer les cas évoqués au paragraphe précédent où le dépassement du seuil surviendrait après la visite auquel cas la surverse, dans le cas de relevé quotidien, sera rapportée le lendemain. Ce point serait à reconsidérer dans le cadre de travaux futurs.

Les tables de contingence ne pourront être construites que si il existe une période commune entre les séries de pluies pour l'une des durées considérées et les archives de surverses. Dans le contraire, les tables de contingence ne pourront être construites. Ces derniers cas sont au nombre de 80 répartis comme suit :

- 51 ODS présentent des surverses et ne possèdent aucune période commune avec les séries de pluie aux stations météorologiques auxquelles ils étaient appariés pour au-

cune des durées ;

- 29 ODS ne présentent aucune surverse durant toute la période archivée et aucune période commune avec les séries de pluie aux stations météorologiques auxquelles ils étaient appariés pour aucune des durées.

5.5 Estimation du pouvoir prévisionnel du modèle basé sur les dépassements des seuils de pluie

Plusieurs mesures de l'exactitude relative de la table de contingence ont été proposées qui permettent d'évaluer la capacité d'un prédictant (ici le dépassement du seuil de pluie) à prévoir l'occurrence d'un événement donné (ici l'occurrence d'une surverse ; nous désignerons cette capacité dans ce qui suit comme le pouvoir prévisionnel du modèle). Plusieurs mesures (*score*) ont été proposées à cet effet (Wilks, 2011). Parmi celles-ci, le *threat score* (TS) est souvent utilisé dans le cas où la fréquence des événements que l'on cherche à prévoir est faible :

$$TS = \frac{a}{(a + b + c)} \quad (5.1)$$

Une autre mesure utilisée est le *Heidke Skill Score* (HSS) qui compare la performance de la prévision (basée sur le dépassement du seuil de pluie) et une prévision aléatoire utilisant la probabilité d'occurrence des surverses. Le HSS s'écrit comme suit (Wilks, 2011) :

$$HSS = \frac{2(ad - bc)}{(a + c)(c + d) + (a + b)(b + d)} \quad (5.2)$$

Une valeur du HSS égale à zéro signifie que la prévision basée sur le dépassement du seuil de pluie n'est pas meilleure qu'une prévision aléatoire basée sur la probabilité d'occurrence des surverses alors qu'une valeur de un signifie une prévision exacte à tous les coups. Bien que cet indice ait été considéré dans un premier temps, il appert qu'il n'existe pas à notre connaissance de référence permettant de juger de la qualité ou du caractère significatif du pouvoir prévisionnel d'un modèle utilisant les valeurs de HSS.

Un autre *score* doit être utilisé dans ce cas. Thornes & Stephenson (2001) indique que le *Odds Ratio Skill Score* (ORSS) peut être utilisé dans ce cas puisqu'un test a été développé qui permet de statuer sur le caractère significatif ou non du modèle prévisionnel. Le ORSS prend la forme suivante (Wilks, 2011) :

Table 5.2: Valeurs minimales du *Odds Ratio Skill Score* (ORSS) nécessaires pour avoir un pouvoir prévisionnel significatif à divers niveaux de confiance en fonction de la valeur de m (voir Tableau 5.4)

	Niveau de confiance		
m	95%	99%	99.9%
1	0.676	0.822	0.913
2	0.524	0.676	0.798
3	0.442	0.586	0.712
4	0.390	0.524	0.648
5	0.352	0.478	0.599
10	0.254	0.352	0.453
20	0.182	0.254	0.332

$$ORSS = \frac{ad - bc}{(ad + bc)} \quad (5.3)$$

Afin de déterminer si la valeur de ORSS indique un pouvoir prévisionnel significatif, la moyenne harmonique des éléments de la table de contingence est d'abord estimée (Thornes & Stephenson, 2001) :

$$\frac{1}{m} = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} \quad (5.4)$$

Les valeurs minimales du ORSS nécessaires pour conclure que le pouvoir prévisionnel est significatif pour différents niveaux de confiance sont présentées au Tableau 5.2 (tiré de l'annexe 1 de Thornes & Stephenson, 2001).

Pour les ODS où les tables de contingence ont pu être construites (soit un total de 2193 ODS), les éléments de ces tables pour chaque durée sont estimées. Trois cas de figures se présentent lors de l'estimation du caractère significatif ou non de la valeur du ORSS :

- Un ou plusieurs des éléments de la table de contingence est nulle ($m = 0$) et alors la valeur de ORSS ne peut être utilisée pour évaluer le pouvoir prévisionnel (Stephenson, 2000) ;
- La valeur de m est comprise entre zéro et un ($0 < m < 1$; correspond au cas où certains des éléments de la tableau de contingence sont petits) et la valeur de ORSS n'est pas une mesure fiable du pouvoir prévisionnel (Stephenson, 2000) ;

- La valeur de m est supérieure à un ($m \geq 1$) et alors la valeur de ORSS est calculée et comparée aux valeurs du Tableau 5.2.

Les cas ont donc été regroupés selon les valeurs de m estimées pour les tables de contingences associées aux diverses durées :

- Si $m < 1$ pour toutes les tables de contingences associées aux différentes durées alors le pouvoir prévisionnel du modèle ne peut être établi ;
- Si $m \geq 1$ pour une ou plusieurs durées, l'ORSS est estimé pour ces durées et la durée avec la valeur maximale du ORSS est sélectionnée (désignée dans ce qui suit par durée optimale) et le test établissant le caractère significatif ou non du modèle est déterminé pour cette durée. Dans le cas où les valeurs du ORSS sont égales pour plusieurs durées, la plus grande durée est sélectionnée.

Le Tableau 5.3 présente le nombre d'ODS appartenant à chacun de ces cas de figures. Des ODS pour lesquels le test peut être appliqué, 91,3 % présente un pouvoir prévisionnel significatif au seuil de 99%. La Figure 5.2 présente les histogrammes du nombre annuel moyen de jours avec surverses pour les différents cas recensés au Tableau 5.3. Cette figure montre clairement que la très grande majorité des cas où il n'est pas possible de statuer sur le pouvoir prévisionnel du modèle basé sur le seuil de pluie ($0 \leq m < 1$) correspondent à des ODS avec moins de 5 jours de surverses annuellement en moyenne. Dans les cas où le test peut être appliqué ($m \geq 1$), les cas de pouvoir prévisionnel non significatif sont dominés par des ODS avec moins de 20 jours avec surverse en moyenne par année alors que la distribution des valeurs dans le cas de liens significatifs entre seuil de pluie et occurrence de surverses comporte davantage de cas avec des nombres de jour avec surverses plus élevés. Les quelques cas classés non significatifs comportant un nombre plus important de jour avec surverses pourraient signifier qu'un nombre significatif de surverses ne sont pas causées par la pluie. Cette hypothèse demanderait toutefois à être vérifiée. Ces résultats montrent que, globalement, pour les ODS où le nombre de surverses enregistré n'est pas trop petit et où la période commune entre archives de surverses à cet ODS et séries sur les pluies à la station météorologique à proximité est suffisante, le modèle basé sur le seuil de pluie présente un pouvoir prévisionnel significatif.

La Figure 5.3 présente des exemples de valeurs du ORSS pour les différentes durées à deux ODS où il était possible de calculer les ORSS pour toutes les durées. Tel que mentionné plus haut, la durée optimale retenu pour établir le pouvoir prévisionnel du modèle

Table 5.3: Nombre d'ODS et pourcentages du nombre total d'ODS appartenant à chacun des cas recensés lors de l'estimation du caractère significatif ou non de la valeur du ORSS.

Cas	Nombre d'ODS	Pourcentage (%)
$0 \leq m < 1$	2012	47.9
$m \geq 1$ et ORSS significatif au seuil 99 %	2003	47.6
$m \geq 1$ et ORSS non significatif au seuil 99%	190	4.5

correspond à la durée pour laquelle la valeur du ORSS est maximale (12 heures et 1 heure dans le cas des ODS présentés à la Figure 5.3). Cette figure montre que, globalement, les valeurs du ORSS varient peu d'une durée à une autre. La Figure 5.4 présente l'histogramme des durées des pluies sélectionnées aux différents ODS qui montre que les plus longues durées sont plus souvent sélectionnées pour les ODS où $m \geq 1$).

La Figure 5.5 présente l'histogramme du nombre équivalent d'années où les séries aux stations météorologiques et séries de surverses aux différents ODS sont disponibles pour les cas où $m \geq 1$. La durée "optimale" est considérée dans ces cas (les données manquantes peuvent en effet différer selon les durées considérées). Ainsi si une donnée de pluie est manquante pendant une période entre deux visites de l'opérateur à l'ODS, la période entre ces deux visites est considérée manquante. Le nombre d'années équivalente est obtenu en faisant la somme des périodes (en jours) sans données manquantes divisée par 182 (nombre de jours de la période mai à octobre). Dans le cas de relevés quotidiens aux ODS, la journée est considérée manquante si le relevé de surverses ou la pluie est manquante.

La Figure 5.6 montre les valeurs d'ORSS maximales obtenues en fonction du nombre annuel moyen de surverses. Un grand nombre de cas où les valeurs d'ORSS sont près de un correspondent à des ODS où le nombre annuel moyen de jours avec surverses est faible (moins de cinq jours par année) et pour lequel le modèle basé sur le seuil de pluie présente un pouvoir prévisionnel significatif (points bleus de la Figure 5.6). Les cas classés non significatifs (points rouges de la Figure 5.6) sont majoritairement associés à des ODS avec moins de 10 jours de surverses par année en moyenne. La performance globale du modèle de prévision augmente lorsque le nombre de surverses augmente et les cas de pouvoir prévisionnel non significatif demeure peu nombreux aux ODS avec en moyenne 20 jours ou plus avec surverses par année.

La Figure 5.7 reprend les informations de la figure précédente et montre l'histogramme des valeurs de ORSS aux différents ODS pour les cas où $m \geq 1$. Ainsi, le mode de la

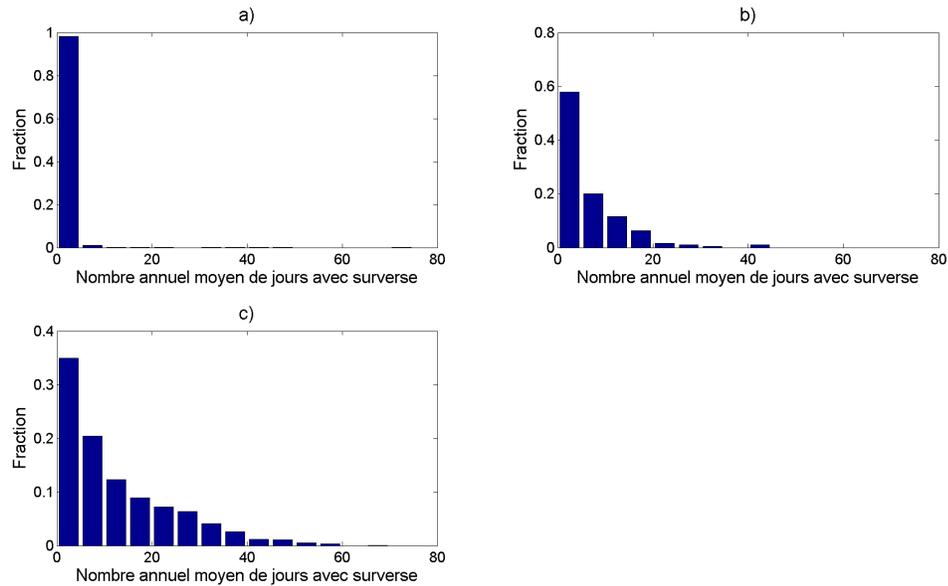


Figure 5.2: Histogrammes du nombre annuel moyen de jour avec surverse aux différents ODS pour les cas où : (a) $0 \geq m < 1$; (b) $m \geq 1$ et pouvoir prévisionnel non significatif à 99%; et (c) $m \geq 1$ et pouvoir prévisionnel significatif à 99%.

distribution de cette figure correspond à des valeurs du ORSS comprises entre 0.85 et 0.9 qui correspondent.

Afin de vérifier que le modèle basé sur le dépassement du seuil de pluie ne présente pas de biais important, le nombre de surverses estimé à partir des archives sur les surverses est comparé au nombre de surverses estimé à partir du dépassement du seuil de pluie à chaque ODS (correspondant à $a + b$). La Figure 5.8 montre que l'accord entre ces deux valeurs est généralement bon. Cependant le nombre de dépassement du seuil de pluie a tendance à être supérieur au nombre de surverse lorsque ce dernier est élevé. Ce dernier résultat suggère que le seuil de pluie devrait dans ces derniers cas être légèrement augmenté pour éliminer ce biais. Cette correction n'a pas été apportée considérant que globalement le biais demeure relativement faible. Une hypothèse pouvant expliquer ce biais pourrait être que les pluies de la période 2001-2012 utilisées pour la comparaison des archives de surverses et les séries pluviométriques sont globalement plus importantes que les périodes utilisées pour déterminer les seuils de pluie au delà desquels il y a surverse. Cette hypothèse n'a toutefois pas été vérifiée.

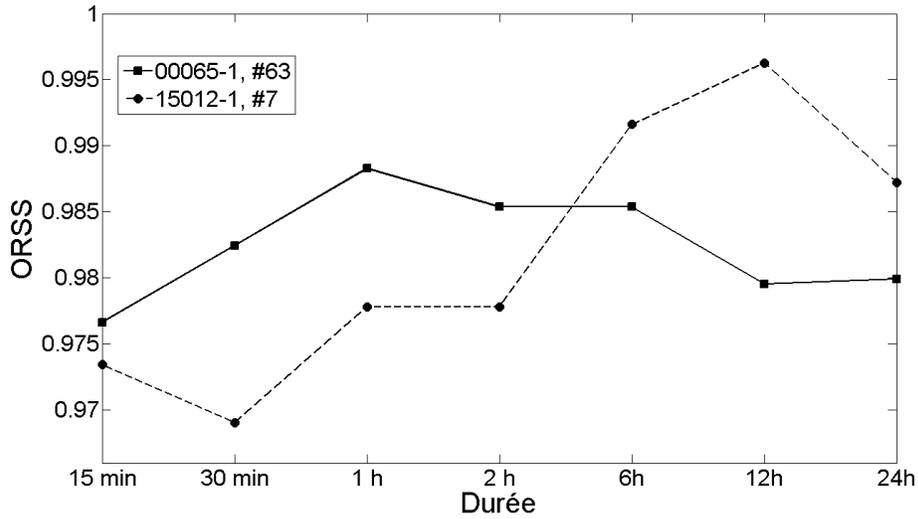


Figure 5.3: Exemple de valeurs du ORSS pour la station Jean-R.-Marcotte à Montréal (station 00065-1, ODS 63) et de la station de la Malbaie (station 15012-1, ODS 7) pour les différentes durées considérées.

5.6 Impact de la distance entre station météorologique et ODS sur la performance du modèle basé sur le seuil de pluie

Plusieurs facteurs peuvent avoir une incidence sur la qualité de la prévision basée sur le seuil de pluie. L'un d'eux concerne la représentativité de l'intensité de pluie mesurée à la station météorologique par rapport à la pluie réelle tombée sur le bassin associé à l'ODS considéré. On peut penser, si cette hypothèse s'avère fondée, que la performance du modèle basé sur le seuil de pluie sera d'autant moins bonne que la distance entre la station pluviométrique et l'ODS est grande et donc que le nombre de cas où le modèle basé sur le seuil de pluie a un pouvoir prévisionnel non significatif augmentera avec la distance entre ODS et station météorologique. La Figure 5.9 présente les box plots des valeurs d'ORSS en fonction de cette distance. On constate une légère diminution des valeurs d'ORSS lorsque la distance ODS-stations météorologiques dépasse 40 km.

La Figure 5.10 présente pour sa part le nombre annuel moyen de surverses en fonction de la distance dans les cas classés non significatifs. Cette fois l'impact de la distance est moins facilement discernable puisque les cas non significatifs dans les cas de grandes distances ODS-stations météorologiques (plus de 40 km) correspondent à des ODS avec un nombre

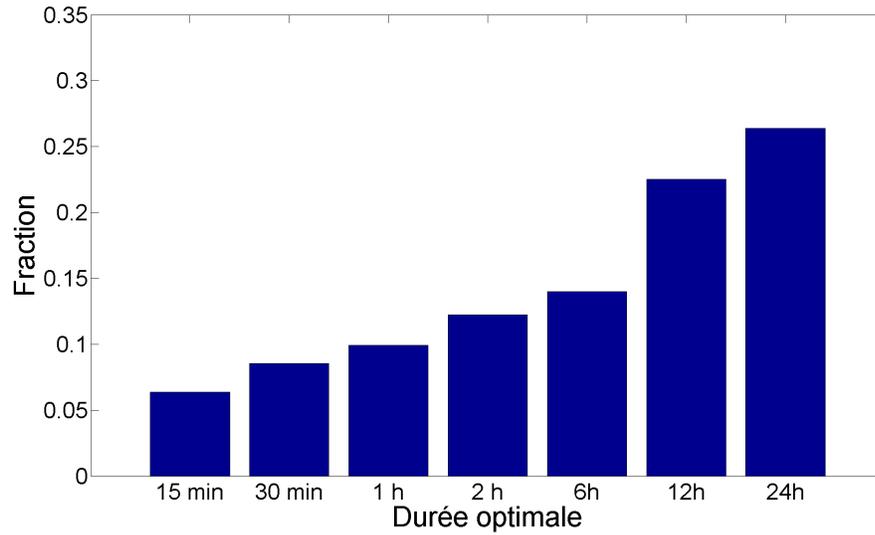


Figure 5.4: Histogramme des durées optimales aux différents ODS correspondant à la durée avec la valeur du ORSS la plus élevée (cas $m \geq 1$).

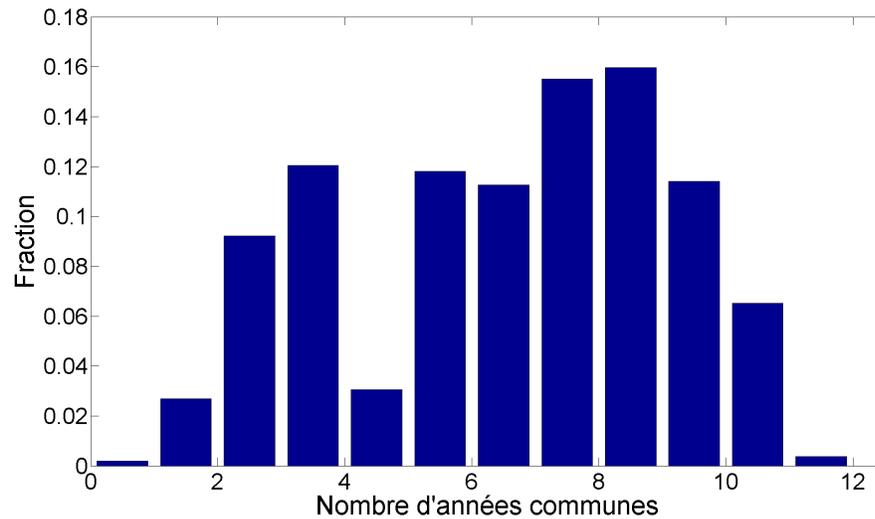


Figure 5.5: Histogramme du nombre d'années communes entre séries aux stations pluviométriques et séries de surverses aux différents ODS lorsque la durée maximisant la valeur du ORSS est considérée à chaque ODS.

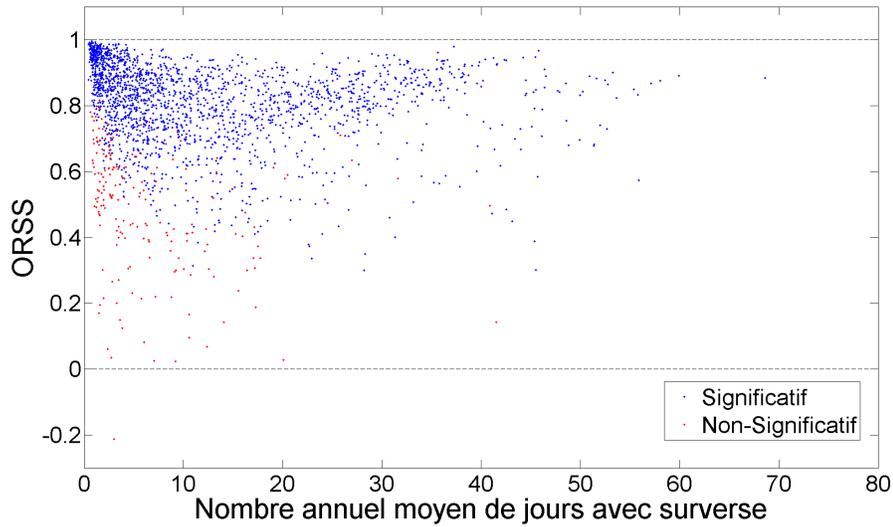


Figure 5.6: Valeurs du ORSS aux différents ODS en fonction du nombre annuel moyen de jours avec surverses pour les cas où $m \geq 1$. Les durées maximisant les valeurs du ORSS à chaque ODS sont considérées.

annuel moyen de jours avec surverses relativement faibles (moins de 15 par année). Ces résultats suggèrent que bien que la distance ODS-stations météorologiques puisse avoir un certain effet sur la relation entre dépassement de seuil de pluie et occurrence de surverse, il n'est pas aussi déterminant que le nombre annuel moyen de surverse par jour.

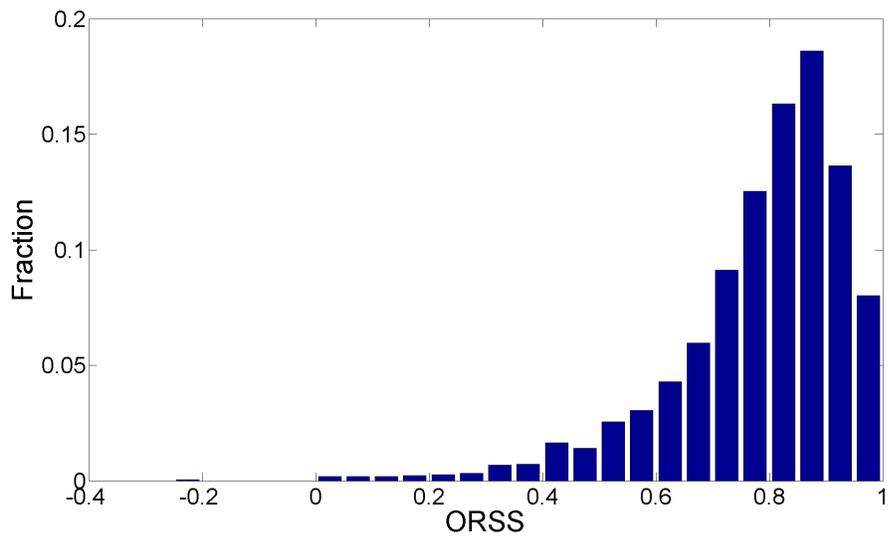


Figure 5.7: Histogramme des valeurs du ORSS aux différents ODS pour les durées maximisant les valeurs du ORSS (cas $m \geq 1$).

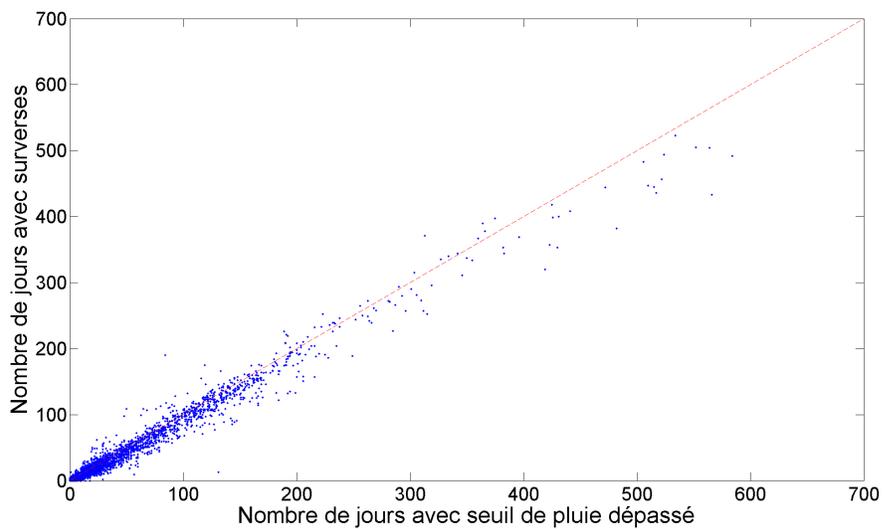


Figure 5.8: Nombre total de jour avec surverses estimé à partir des archives sur les surverses en fonction du nombre de jour avec dépassement du seuil de pluie à chaque ODS. Les durées maximisant les valeurs de HSS à chaque ODS sont considérées.

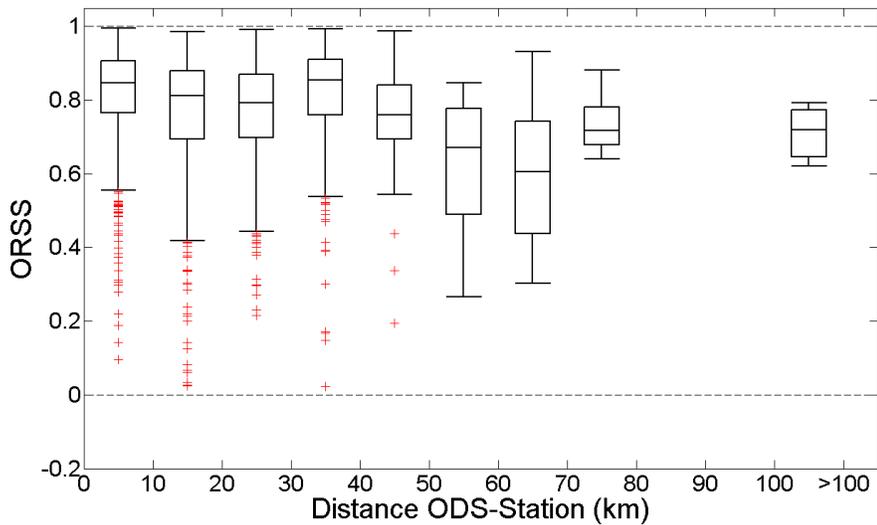


Figure 5.9: Box plots des valeurs du ORSS en fonction de la distance entre station météorologique et ODS (cas $m \geq 1$).

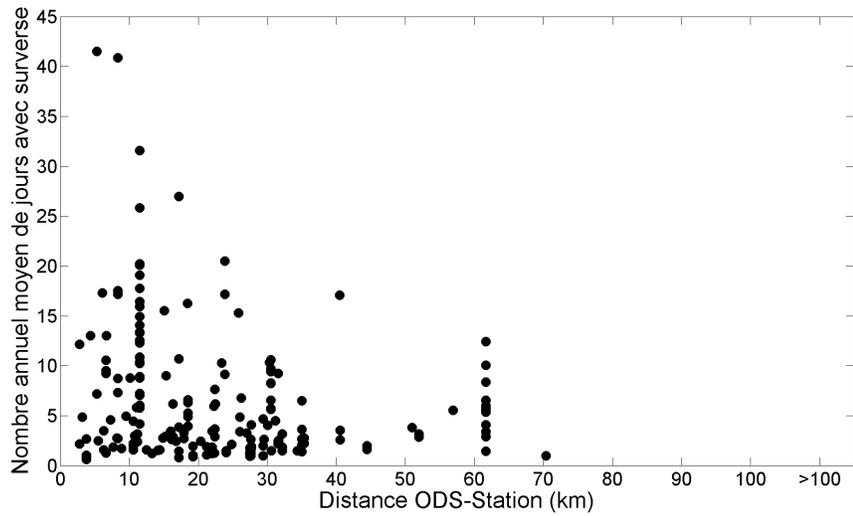


Figure 5.10: Fraction du nombre d'ODS où le pouvoir prévisionnel du modèle basé sur le seuil de pluie est non significatif en fonction de la distance entre station météorologique et ODS (cas $m \geq 1$).

Chapitre 6

Conclusion

Le présent rapport rend compte des travaux visant : 1) à développer une approche permettant d'estimer la fréquence d'occurrence annuelle (période de mai à octobre) des surverses à partir des archives de la base SOMAE ; 2) à développer une procédure permettant de déterminer le seuil de pluie correspondant, seuil de pluie au delà duquel il y a vraisemblablement surverse ; 3) à vérifier la validité de ce modèle à partir d'une comparaison des séries de surverses et de pluie. Les approches développées ont été appliquées à 4285 ODS répertoriés dans la base SOMAE. Les périodes couvertes par les archives de surverses aux différents ODS vont de 1 à 12 ans, avec une majorité d'ODS ($\approx 60\%$) pour laquelle on dispose d'une série de 12 ans. Les résultats de ces travaux sont regroupés dans un fichier Excel joint au présent rapport.

L'approche développée pour estimer le nombre annuel moyen de surverses aux différents ODS repose sur l'hypothèse que la probabilité d'occurrence des surverses est la même tous les jours et qu'il s'agit d'événements statistiquement indépendants à l'échelle journalière (i.e. que la probabilité d'occurrence à un jour donné n'est pas conditionnée par l'occurrence ou non de surverse les jours précédents). Cette hypothèse n'a pas été vérifiée mais demeure une bonne première approximation. Suivant ce modèle, la probabilité d'avoir un nombre donné de surverses au cours d'une période donnée est définie par la loi binomiale. De même le nombre moyen de surverses pendant une période donnée (ici la période de mai à octobre comprenant 182 jours) peut être estimé à partir de la probabilité quotidienne d'occurrence de surverses.

Les enregistrements des surverses de la base SOMAE ont été utilisés pour évaluer la probabilité d'occurrence quotidienne de surverses à chaque ODS. La méthode a été adaptée

en fonction du mode d'enregistrement utilisé à chaque ODS (bloc repère, enregistreur en continu ou automatique). L'estimation de cette probabilité utilise la méthode du maximum de vraisemblance. Des hypothèses supplémentaires ont été formulées afin de tenir compte des données manquantes dans le cas des enregistreurs automatiques. Les résultats montrent qu'un large pourcentage d'ODS (de l'ordre de 43%) connaît moins de une surverse en moyenne par année (période de mai à octobre) alors qu'environ 2 % connaît plus de 40 surverses par année.

Le modèle le plus simple que l'on puisse considérer pour expliquer l'occurrence des surverses est basé sur l'hypothèse que les surverses surviennent lorsque l'intensité de pluie dépasse un certain seuil. Les séries des maxima journaliers de pluie sur différentes durées ont donc été considérées de même que les séries de pluies journalières. Afin d'identifier les stations météorologiques les mieux à même de représenter les événements de pluie qui s'abattent sur les bassins alimentant les différents ODS, un critère de sélection a été défini visant à identifier les stations météorologiques les plus proches des stations d'épuration associées aux différents ODS (seules les coordonnées des stations d'épuration sont connues) et dont les séries recourent les périodes pour lesquels les enregistrements sur les surverses sont disponibles. Les fonctions de répartition empiriques à chaque station météorologique ont ensuite été construites et les valeurs des quantiles de pluie pour chaque durée correspondant à la probabilité quotidienne d'occurrence des surverses ont ensuite été déterminés. Ces quantiles correspondent donc au seuil de pluie pour une durée donnée au delà de laquelle il y aura surverse.

Ces valeurs de seuil ont ensuite été utilisées pour vérifier qu'il y a occurrence d'une surverse à un jour donné lorsqu'à ce même jour (ou pendant la période antérieure depuis la dernière visite de l'ODS) le seuil de pluie a été dépassé. Les tables de contingences des prévisions basées sur le dépassement du seuil et l'occurrence de surverses ont été construites à chaque ODS où il était possible de calculer cette table (les tables de contingence de 80 ODS n'ont pu être estimés). Le *Odds Ratio Skill Score* (ORSS) a été utilisé pour évaluer la qualité des prévisions basées sur le dépassement du seuil de pluie (autrement dit dans quelle mesure le dépassement du seuil entraîne une surverse) pour ces ODS. Une valeur du ORSS près de zéro signifie que le modèle basé sur les seuils de pluie n'est pas meilleur qu'un modèle aléatoire alors qu'une valeur près de un signifie un modèle dont les prévisions sont justes presque à tout coup. La durée de pluie dont la valeur du ORSS est la plus élevée a été retenue comme la durée "optimale" (*i.e.* celle qui offre les meilleures prévisions). À noter que les durées 5 et 10 minutes, bien que disponibles aux stations météorologiques n'ont pas

été considérées. A noter que les durées les plus longues sont plus souvent sélectionnées.

Les valeurs de ORSS obtenues pour les durées retenues varient d'un ODS à l'autre mais la grande majorité présente des valeurs supérieures à 0.8. Un test statistique a été appliqué afin de déterminer si le pouvoir prévisionnel du modèle basé sur le dépassement du seuil de pluie est significatif. Les résultats montrent que, pour les ODS où ce test peut être appliqué (51.2 % des ODS), 91.3 % des cas le modèle basé sur le seuil des pluies possède un pouvoir prévisionnel significatif. Les cas de modèle non significatif correspondent en grande majorité à des ODS où l'on dénombre moins de 10 jours de surverses par année en moyenne.

Le modèle basé sur le dépassement d'un seuil de pluie pour une durée donnée représente évidemment une simplification de la situation réelle. Plusieurs facteurs peuvent également déterminer l'occurrence ou non d'une surverse. Ainsi, à court terme, les conditions hydrologiques antérieures (p. ex. des événements de pluie antérieures de moindres intensités ayant modifiés les conditions d'humidité des sols et les volumes d'eau déjà présent en réseau) peuvent modifier le niveau de pluie nécessaire pour générer des surverses expliquant qu'une même intensité de pluie pourra, dans certains cas, générer une surverse alors qu'à d'autres occasions elle n'entraînera pas de surverses. De même, diverses modifications rapides ou à long terme de l'occupation du territoire (p. ex. développement urbain, imperméabilisation des surfaces) ou à caractère hydraulique (p. ex. modifications des diamètres de certaines conduites, reconfiguration de certaines parties du réseau) pourra entraîner une modification de la fréquence des surverses et des seuils de pluie nécessaire pour générer des surverses. De telles modifications pourraient entraîner une modification rapide ou à long terme des fréquences de surverses. L'approche utilisée repose sur l'hypothèse que de telles modifications, si elles ont eu cours, n'ont pas entraîné de modifications significatives de la fréquence des surverses durant la période archivée.

Plusieurs améliorations pourraient être apportées au modèle basé sur le dépassement de seuil proposé. Tel que mentionné précédemment, il est possible que les dépassements du seuil de pluie se produisent, un jour donné, après la visite de l'opérateur et, dans le cas de relevés quotidiens, faire en sorte que la surverse dans ce cas ne sera enregistrée que le lendemain. Un décompte exact de ce type de situations où il y a dépassement du seuil un jour donné et surverse le lendemain n'a pas été réalisé. Dans l'éventualité où ce type de situation s'avérerait fréquent, une modification des règles de construction des tables de contingence serait sans doute pertinente.

Bien que le seuil estimé à partir de la distribution empirique des précipitations de durée

donnée semble dans l'ensemble approprié, ce seuil pourrait être ajusté de façon à maximiser la valeur de ORSS à chaque ODS. Il s'agirait dans ce cas de modifier le seuil, de reconstruire la table de contingence et d'évaluer la nouvelle valeur de OrSS et de répéter cette séquence d'opérations jusqu'à obtenir la plus grande valeur de ORSS possible. L'amélioration des performances du modèle ainsi obtenue demeure cependant *a priori* difficile à évaluer et pourrait être marginale pour nombre d'ODS.

Aucune évaluation des incertitudes sur le nombre de jours avec surverses estimé n'a été réalisée. Cette évaluation permettrait de déterminer les intervalles de confiance des valeurs estimées. Ces intervalles varieraient sensiblement d'un ODS à l'autre puisque les nombres d'années d'enregistrements des surverses varient d'un ODS à l'autre. De même elles dépendront du nombre de surverses enregistré, les incertitudes sur l'estimation du nombre moyen de surverses étant plus élevées lorsque le nombre de surverses est faible. Enfin la fréquence des visites de l'opérateur et d'enregistrement des données sur les surverses auront un impact sur ces incertitudes.

Références

Stephenson D (2000). Use of the "odds ratio" for diagnosing forecast skill. *Weather and Forecasting*, 15(2):221–232.

Thornes J & Stephenson D (2001). How to judge the quality and value of weather forecast products. *Meteorological Applications*, 8(3):307–314.

Wilks D (2011). *Statistical Methods in the Atmospheric Sciences, third edition*. Elsevier.