Titre: Title:	Modélisation des erreurs de prévisions de précipitations
Auteur: Author:	Liliane Guilbault
Date:	2006
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Guilbault, L. (2006). Modélisation des erreurs de prévisions de précipitations [Master's thesis, École Polytechnique de Montréal]. PolyPublie. https://publications.polymtl.ca/7713/

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/7713/
Directeurs de recherche: Advisors:	Mario Lefebvre
Programme: Program:	Unspecified

UNIVERSITÉ DE MONTRÉAL

MODÉLISATION DES ERREURS DE PRÉVISIONS DE PRÉCIPITATIONS

LILIANE GUILBAULT
DÉPARTEMENT DE MATHÉMATIQUES ET
DE GÉNIE INDUSTRIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(MATHÉMATIQUES APPLIQUÉES)
AVRIL 2006

Your file Votre référence
ISBN: 978-0-494-17944-4
Our file Notre référence
ISBN: 978-0-494-17944-4

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

MODÉLISATION DES ERREURS
DE PRÉVISIONS DE
PRÉCIPITATIONS

présenté par: GUILBAULT Liliane
en vue de l'obtention du diplôme de: Maîtrise ès sciences appliquées
a été dûment accepté par le jury d'examen constitué de:

Mme LABBÉ Chantal, Ph.D., présidente
M. LEFEBVRE Mario, Ph.D., membre et directeur de recherche
M. TURGEON André, Ph.D., membre

REMERCIEMENTS

J'aimerais témoigner de ma reconnaissance à mon directeur de recherche, monsieur Mario Lefebvre, pour son aide précieuse, ses conseils avisés et sa supervision respectueuse. Je suis particulièrement heureuse d'avoir pu profiter des ses connaissances approfondies, toujours humblement transmises. Je tiens aussi à lui adresser mes remerciements pour son soutien financier. Je remercie également messieurs André Turgeon, Denis Tremblay et Marco Latraverse, de l'entreprise Hydro-Québec, pour la bourse de recherche octroyée et l'appui donné à ce projet. Enfin, qu'il me soit permis de souligner l'aide apportée par monsieur Stéphane Krau, du Groupe d'études et de recherche en analyse des décisions (GERAD).

RÉSUMÉ

On cherche des modèles statistiques pour les erreurs produites lors des prévisions de précipitations faites par Hydro-Québec pour la période débutant le 29 mai 2003 et se terminant le 28 mai 2004.

Pour ce faire, on analyse les erreurs de prévisions de précipitations de certaines des stations du bassin versant de la rivière Gatineau. Pour ces analyses, on considère les précipitations annuelles prévues et observées pour trois cas, soit pour le jour même, un et deux jours d'avance. On analyse également des sous-ensembles de ces données. Puis, pour chacun des cas, on cherche des modèles statistiques qui s'ajustent aux données considérées.

Les modèles semblant les plus appropriés sont les lois normale, lognormale et de Pareto généralisée. Ainsi, à condition d'enlever certaines données aberrantes, la loi normale pourrait être acceptable lorsqu'on prévoyait entre 5 et 10 mm de précipitations. Le modèle de la loi lognormale est particulièrement approprié dans les situations suivantes: celle où on ne prévoyait aucune précipitation et celle où on s'attendait à au moins 10 mm. Finalement, la loi de Pareto généralisée est plus adaptée lorsqu'on considère toutes les données et lorsqu'on prévoyait entre 0 et 5mm de précipitations.

ABSTRACT

We look for statistical models for the forecast errors made by Hydro-Québec when forecasting precipitation, for the one-year period beginning on May 29, 2003 and ending on May 28, 2004.

To do so, we analyze the forecast errors for the rainfall at some of the stations of the Gatineau river basin. For these analyses, we consider the rainfall, both forecasted and observed during the one-year period of interest, for the current day as well as for one and two days in advance. We also analyze subsets of the data set. Then, for each case, we look for statistical models which fit the data collected.

The appropriate models are the normal, lognormal, and generalized Pareto distributions. If we remove some outliers, the normal distribution is acceptable when between 5 and 10 mm of rainfall is forecasted. The lognormal distribution is more appropriate in the following situations: no precipitation at all, and more than 10 mm of precipitation is forecasted. Finally, the generalized Pareto distribution is the best model when we consider all the data and in the case when less than 5 mm of precipitation is forecasted.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES SIGLES ET ABRÉVIATIONS

c	Paramètre de la loi de Pareto généralisée
c_Z	Paramètre de la loi de Pareto généralisée, relié à la variable aléatoire Z
$c_{Z'}$	Paramètre de la loi de Pareto généralisée, relié à la variable aléatoire Z'
$\hat{c}_Z$	Paramètre estimé de la loi de Pareto généralisée, relié à la variable aléatoire Z
CNRM	Centre national de recherches météorologiques
D^2	Statistique de Pearson
ddl	Degré de liberté
ESP	Extended Streamflow Predicition
E[Z]	Espérance de Z
f_W	Fonction de densité de probabilité de la variable aléatoire W
f_X	Fonction de densité de probabilité de la variable aléatoire X
f_Z	Fonction de densité de probabilité de la variable aléatoire Z
$f_{Z'}$	Fonction de densité de probabilité de la variable aléatoire Z'
GERAD	Groupe d'études et de recherche en analyse des décisions
H_0	Hypothèse nulle
H_1	Hypothèse non nulle
k	Nombre de classes
m_i	Nombre d'observations que devrait compter le i^e intervalle
mm	Millimètre
n	Nombre de degrés de liberté
n	Nombre de données
n_i	Nombre d'observations dans le i^e intervalle
p	Prévisions

p_i	Probabilité qu'une observation se trouve dans le i^e intervalle si le modèle testé est le bon
P	Probabilité
r	Nombre de paramètres estimés
r_1	Coefficient de corrélation – Un jour d'écart
r_2	Coefficient de corrélation – Deux jours d'écart
s_W	Écart type de W
s_W^2	Variance de W
s_{W1}	Écart type de W – Jour 1
s_{W1}^2	Variance de W – Jour 1
s_{X1}	Écart type de X – Jour 1
s_{X1}^2	Variance de X – Jour 1
s_{X2}	Écart type de X – Jour 2
s_{X2}^2	Variance de X – Jour 2
s_{X3}	Écart type de X – Jour 3
s_{X3}^2	Variance de X – Jour 3
s_{Y1}	Écart type de Y – Jour 1
$s_{Y'1}$	Écart type de Y' – Jour 1
s_{Y1}^2	Variance de Y – Jour 1
$s_{Y'1}^2$	Variance de Y' – Jour 1
s_{Y2}	Écart type de Y – Jour 2
s_{Y2}^2	Variance de Y – Jour 2
s_{Y3}	Écart type de Y – Jour 3
s_{Y3}^2	Variance de Y – Jour 3
s_{Z1}	Écart type de Z – Jour 1
s_{Z1}^2	Variance de Z – Jour 1
W	Valeur de l'erreur de prévision
$\overline{w}_1$	Moyenne de W – Jour 1
X	Valeur absolue de l'erreur de prévision

$\bar{x}_1$	Moyenne de X – Jour 1
$\bar{x}_2$	Moyenne de X – Jour 2
$\bar{x}_3$	Moyenne de X – Jour 3
Y	Valeur absolue de l'erreur de prévision, étant donné qu'elle est supérieure à 0,2 mm
Y'	Valeur de l'erreur de prévision, étant donné qu'elle est supérieure à 0,2 mm et inférieure à 0,2 mm
$\bar{y}_1$	Moyenne de Y – Jour 1
$\bar{y}_2$	Moyenne de Y – Jour 2
$\bar{y}_3$	Moyenne de Y – Jour 3
Z	Valeur absolue de l'erreur de prévision, étant donné qu'elle est supérieure à 0,2 mm moins 0,2
Z	Variable aléatoire normale centrée réduite
$\bar{z}_1$	Moyenne de Z – Jour 1
Z'	Valeur de l'erreur de prévision, étant donné qu'elle est supérieure à 0,2 mm moins 0,2 et inférieure à 0,2 mm plus 0,2
λ_W	Estimateur de la loi exponentielle double, relié à la variable aléatoire W
$\hat{\lambda}_{\mathrm{W}}$	Paramètre estimé de la loi exponentielle double, relié à la variable aléatoire W
λ_Z	Estimateur de la loi exponentielle, relié à la variable aléatoire Z
$\hat{\lambda}_Z$	Paramètre estimé de la loi exponentielle, relié à la variable aléatoire Z
θ	Paramètre de la loi de Pareto généralisée
θ_Z	Paramètre de la loi de Pareto généralisée, relié à la variable aléatoire Z
$\theta_{Z'}$	Paramètre de la loi de Pareto généralisée, relié à la variable aléatoire Z'
α	Risque de première espèce

μ	Espérance de la loi normale et lognormale
σ^2	Variance de la loi normale et lognormale

INTRODUCTION

Dans le cadre d'un projet d'innovation technologique, pour lequel un des objectifs poursuivis consistait à améliorer la gestion des systèmes au fil de l'eau, Hydro-Québec a fait des prédictions pour les précipitations journalières. L'approche méthodologique utilisée, appelée bruitage, consistait à transformer aléatoirement les prévisions de précipitations dans un contexte de création de séries Extended Streamflow Predicition (ESP). Le bassin versant témoin utilisé, c'est-à-dire la surface délimitée pour la recherche, fut celui de la rivière Gatineau pour lequel des stations furent considérées selon une grille d'étude spécifique.

L'objectif visé consiste à analyser les prévisions fournies par Hydro-Québec pour en évaluer la précision. Les données fournies pour chacune des stations observées, soit les précipitations réelles et prévues, concernent la période allant du 29 mai 2003 au 28 mai 2004, inclusivement. Par conséquent, les données utilisées pour ce projet couvrent une année. Par la suite, il s'agit de trouver des modèles statistiques appropriés qui permettront de quantifier la précision de ces prévisions, et ce, pour le jour même et un et deux jours d'avance. On cherche donc à modéliser les erreurs de prévisions de précipitations. Les stations étudiées dans ce document ont été choisies à l'intérieur de la grille d'étude utilisée par Hydro-Québec. De plus, les stations pour lesquelles il n'y avait pas suffisamment de données pour certains mois n'ont pas été retenues.

Les chercheurs se sont jusqu'à maintenant surtout intéressés aux prévisions de précipitations. Au fil des ans, plusieurs méthodes pour produire les prévisions de précipitations ont été utilisées. Pensons entre autres aux méthodes de prévision statistique et de prévision numérique. Par contre, peu se sont intéressés à la modélisation des erreurs engendrées par ces prévisions. Parce que, comme montré par certains auteurs dont McBride et Ebert (2000), la vérification des prévisions des précipitations est déjà une tâche difficile. Tel que mentionné par Déqué (2003) du Centre national de recherches météorologiques (CNRM) de France, le caractère imprévisible des mouvements atmosphériques fait que les prévisions se limitent à quelques jours. Pour mesurer le succès d'une prévision, ce dernier parle de l'utilisation d'un coefficient de corrélation spatio-temporelle entre la prévision et l'observation. Dans le cadre d'un stage dans un centre de recherches portant sur les simulateurs de climat, Rousseau (2004) a cherché à modéliser les erreurs de prévisions de Météo-France en utilisant des modèles de régression linéaire pour déterminer des liens temporels entre les erreurs de prévisions. En ce qui nous concerne, nous n'avons pas considéré la possibilité que les erreurs de prévisions d'un jour donné pouvaient s'expliquer par celles du jour précédent. Nous allons plutôt trouver des modèles statistiques s'ajustant aux données et ce, en considérant des périodes spécifiques.

Le présent document se divise en trois parties. On examine les trois cas suivants, soit les prévisions pour la journée même, celles pour le lendemain et, finalement, celles à deux jours d'avance. Chacune des parties présente tout d'abord une analyse des données pour chaque cas considéré et pour les stations retenues. De plus, on considère également des sous-ensembles spécifiques de données. Ensuite, pour chacune des sections, et faisant suite à l'analyse des données, on cherche des modèles en effectuant des tests d'ajustement de ces modèles aux données. On présente alors des modèles statistiques généraux, et des modèles plus spécifiques pour les sous-ensembles, qui s'ajustent aux données de l'ensemble des stations retenues.

CHAPITRE 1: MODÈLES STATISTIQUES – JOUR MÊME

Le premier cas étudié est celui où des prédictions de précipitations sont faites pour le jour même. On cherche à modéliser les erreurs de prévisions de ces précipitations.

Posons

W : Valeur de l'erreur de prévision
X : Valeur absolue de l'erreur de prévision

Les caractéristiques de la différence W entre les valeurs de précipitations observées et les valeurs prévues et des valeurs absolues des erreurs de prévisions X qui suivent ont été compilées pour une année, soit du 29 mai 2003 au 28 mai 2004, inclusivement. Pour l'ensemble des données, on considère également les erreurs pour les périodes d'été et d'hiver.

1.1 Valeurs des erreurs de prévisions

Les tableaux suivants présentent les données statistiques obtenues pour W, soit les valeurs des erreurs de prévisions. Pour certains groupes de données, afin d'obtenir des erreurs qui soient très peu corrélées entre elles et de ce fait considérées indépendantes, des données ont été éliminées. Par exemple, on a pu éliminer les données numéros 5, 10, 15, etc. ou encore les données numéros 1, 4, 7, etc. Ces données statistiques, représentant les données non corrélées, sont montrées en caractères italiques.

Pour la station Wapus, on a 125 données au total. Cependant, les mois de décembre et suivants ne sont représentés que par huit données. Pour la station Parent, il y a 188 données mais neuf seulement au total pour les mois de novembre, décembre et janvier.

Ainsi, en raison d'un nombre de données insuffisant pour certains mois consécutifs, les stations Wapus et Parent ne seront pas étudiées.

Tableau 1.1: Données statistiques de W – Prévisions jour 1 (Toutes les données)

STATIONS	Moyenne ($\overline{w}_1$)	Variance (s_{W1}^2)	Écart type (s_{W1})	Nombre de données (n)
Brotkord	0.22	98.93	9.95	228
	0,78	*117,47*	*10,84*	*152*
Chouart	0.49	102.37	10.12	240
	0,17	*86,46*	*9,30*	*192*
High Falls	0.64	99.91	10.00	245
	1,33	*108,09*	*10,40*	*122*
La Pêche	0.53	92.84	9.64	239
	0,81	*91,66*	*9,57*	*192*
Maniwaki Airport	0.25	118.63	10.89	245
	-0,99	*112,53*	*10,61*	*163*
Mont St-Michel	0.30	116.96	10.81	221
	0,59	*104,21*	*10,21*	*177*
Wapus	Nombre de données insuffisant			
Parent	Nombre de données insuffisant			

Tableau 1.2: Données statistiques de W – Prévisions jour 1 : Été et hiver (Toutes les données)

STATIONS	**Été**				**Hiver**			
	Moyenne ($\overline{w}_1$)	Variance (s_{W1}^2)	Écart type (s_{W1})	Nombre de données (n)	Moyenne ($\overline{w}_1$)	Variance (s_{W1}^2)	Écart type (s_{W1})	Nombre de données (n)
Brotkord	0.02	124.69	11.17	153	0.64	47.11	6.86	75
	0,58	*160,94*	*12,69*	*102*	*1,20*	*29,99*	*5,48*	*50*
Chouart	0.77	122.86	11.08	182	-0,39	38.09	6.17	58
	0,30	*100,10*	*10,00*	*146*	*-0,24*	*44,22*	*6,65*	***46***
High Falls	0.71	119.29	10.92	189	0.41	35.42	5.95	56
	1,23	*130,58*	*11,43*	*94*	*1,69*	*34,43*	*5,87*	***28***
La Pêche	0.38	113.14	10.64	184	1.03	25.45	5.04	55
	0,65	*109,53*	*10,47*	*148*	*1,34*	*30,18*	*5,49*	***44***
Maniwaki Airport	0.16	146.73	12.11	189	0.52	24.65	4.97	56
	-1,42	*137,05*	*11,71*	*125*	*0,42*	*30,71*	*5,54*	***38***
Mont St-Michel	-0.03	158.06	12.57	149	0.99	32.62	5.68	72
	0,14	*136,77*	*11,69*	*120*	*1,54*	*35,53*	*5,96*	*57*

Dans le tableau précédent, on remarquera que le fait d'éliminer certaines données pour la période d'hiver (afin d'obtenir un ensemble de données moins corrélées) entraîne des ensembles possédant un nombre de données assez faible.

Dans le tableau 1.2 et les tableaux subséquents, un nombre de données inférieur à 50 est indiqué en caractère gras.

Tableau 1.3: Données statistiques de W – Prévisions jour 1 (Prévisions p = 0 mm)

STATIONS	Moyenne ($\overline{w}_1$)	Variance (s_{W1}^2)	Écart type (s_{W1})	Nombre de données (n)
Brotkord	-2.28	39.11	6.25	69
Chouart	-1.39	8.87	2.98	55
High Falls	-2.65	50.54	7.11	87
La Pêche	-2.47	41.32	6.43	89
Maniwaki Airport	-2.41	30.13	5.49	84
Mont St-Michel	-1.99	22.38	4.73	73

Tableau 1.4: Données statistiques de W – Prévisions p jour 1 (0<p<1)

STATIONS	Moyenne ($\overline{w}_1$)	Variance (s_{W1}^2)	Écart type (s_{W1})	Nombre de données (n)
Brotkord	-4.18	85.53	9.25	59
Chouart	-3.20	47.43	6.89	67
	-2,90	*40,26*	*6,35*	*54*
High Falls	-4.22	80.37	8.96	**43**
La Pêche	-4.40	67.91	8.24	**41**
Maniwaki Airport	-6.31	176.04	13.27	**45**
Mont St-Michel	-4.20	116.75	10.80	50
	-4,12	*128,36*	*11,33*	***40***

Tableau 1.5: Données statistiques de W – Prévisions p jour 1 (1≤p<5)

STATIONS	Moyenne ($\overline{w}_1$)	Variance (s_{W1}^2)	Écart type (s_{W1})	Nombre de données (n)
Brotkord	0,67	15.52	3.94	**49**
Chouart	-1.31	107.76	10.38	69
High Falls	0,31	16.52	4.06	61
La Pêche	0.13	25.77	5.08	59
Maniwaki Airport	-0.08	69.17	8.32	66
Mont St-Michel	-1.88	126.94	11.27	52

Si on regroupe les prévisions des deux derniers groupes, soit 0<p<1 et 1≤p<5, on obtient le tableau 1.6.

Tableau 1.6: Données statistiques de W – Prévisions p jour 1 (0<p<5)

STATIONS	Moyenne ($\overline{w}_1$)	Variance (s_{W1}^2)	Écart type (s_{W1})	Nombre de données (n)
Brotkord	-1.98	59.21	7.69	108
Chouart	-2.24	78.37	8.85	136
High Falls	-1.56	47.43	6.89	104
La Pêche	-1.73	47.54	6.89	100
Maniwaki Airport	-2.61	120.73	10.99	111
Mont St-Michel	-3.02	122.09	11.05	102

Tableau 1.7: Données statistiques de W – Prévisions p jour 1 (5≤p<10)

STATIONS	Moyenne ($\overline{w}_1$)	Variance (s_{W1}^2)	Écart type (s_{W1})	Nombre de données (n)
Brotkord	2.67	179.58	13.4	**26**
Chouart	4.12	72.00	8.49	**27**
High Falls	3.89	50.06	7.08	**30**
La Pêche	4.95	82.78	9.10	**26**
Maniwaki Airport	5.79	13.55	3.68	**24**
Mont St-Michel	4.18	18.63	4.32	**18**

Tableau 1.8: Données statistiques de W – Prévisions jour 1 (10≤p)

STATIONS	Moyenne ($\overline{w}_1$)	Variance (s_{W1}^2)	Écart type (s_{W1})	Nombre de données (n)
Brotkord	14.10	126.98	11.27	**25**
Chouart	17.65	166.40	12.90	**22**
High Falls	18.04	202.70	14.24	**24**
La Pêche	16.25	170.97	13.08	**24**
Maniwaki Airport	15.88	170.80	13.07	**26**
Mont St-Michel	15.89	107.10	10.35	**28**

Si on regroupe les prévisions des deux derniers groupes, soit 5≤p<10 et 10≤p, on obtient le tableau qui suit.

Tableau 1.9: Données statistiques de W – Prévisions jour 1 (5≤p)

STATIONS	Moyenne ($\overline{w}_1$)	Variance (s_{W1}^2)	Écart type (s_{W1})	Nombre de données (n)
Brotkord	8.27	184.08	13.57	51
	6.84	*180.70*	*13.44*	***41***
Chouart	10.19	158.04	12.57	49
	10.11	*184.37*	*13.58*	***32***
High Falls	10.18	165.77	12.88	54
La Pêche	10.38	155.02	12.45	50
Maniwaki Airport	11.03	119.45	10.93	50
Mont St-Michel	11.31	104.70	10.23	**46**

Les tableaux 1.10 et 1.11 fournissent les coefficients de corrélation entre les données et ces mêmes données mais avec un jour d'écart, soit r_1, et deux jours d'écart, soit r_2. Encore ici, les caractères italiques indiquent que des données ont été éliminées pour un ensemble de données.

Tableau 1.10: Coefficients de corrélation de W - Prévisions jour 1

STATIONS	Toutes		Prévisions (p)=0		0<p<1		1≤p<5		0<p<5		5≤p<10		10≤p		5≤p	
	r_1	r_2	r_1	r_2	r_1	r_2	r_1	r_2	r_1	r_2	r_1	r_2	r_1	r_2	r_1	r_2
Brotkord	-0.31 *-0.23*	0,05 *-0.10*	-0,02	-0,08	-0,09	0,09	0,05	0,04	0,02	0,05	0,00	0,03	-0,47	0,16	-0,29 *-0,11*	-0,06 *-0,07*
Chouart	-0.31 *-0,09*	0,02 *-0,04*	0,13	0,05	0,02 *0,00*	0,26 *-0,10*	-0,09	-0,03	-0,01	0,07	0,21	0,00	0,29	0,32	-0,34 *0,06*	0,12 *-0,09*
High Falls	-0,37 *-0,03*	0,03 *0,16*	-0,06	-0,02	-0,02	-0,16	0,08	-0,05	-0,01	0,08	0,29	0,14	-0,27	-0,21	-0,14	-0,12
La Pêche	-0,27 *-0,16*	-0,04 *-0,01*	-0,04	-0,06	-0,02	0,07	0,02	-0,05	-0,07	0,08	-0,09	-0,08	-0,34	-0,13	-0,20	-0,21
Maniwaki Airport	-0,39 *-0,02*	0,10 *-0,03*	0,10	0,09	0,22	-0,15	0,05	0,15	0,15	0,05	-0,19	-0,05	-0,36	0,07	-0,19	-0.08
Mont St-Michel	-0,38 *-0,07*	0,10 *-0,05*	-0,01	0,05	0,30 *0,02*	-0,07 *-0,02*	-0,04	-0,07	0,24	-0,04	0,19	-0,01	0,06	-0,12	-0,20	0,17

Tableau 1.11: Coefficients de corrélation de W – Été et hiver (Prévisions jour 1 - Toutes les données)

STATIONS	Été		Hiver	
	r_1	r_2	r_1	r_2
Brotkord	-0.26	-0.15	-0.34	0.10
	-0,23	*-0,05*	*-0,19*	*-0,16*
Chouart	-0.34	0.05	-0.06	-0.12
	-0,09	*-0,03*	*0,16*	*-0,04*
High Falls	-0.38	0.03	-0.33	-0.06
	-0,02	*0,16*	*-0,19*	*0,12*
La Pêche	-0.29	-0.04	0.06	-0.07
	-0,17	*-0,01*	*0,02*	*-0,04*
Maniwaki Airport	-0.39	0.11	-0.23	-0.14
	0,00	*-0,03*	*-0,46*	*-0,08*
Mont St-Michel	-0.45	0.11	0.24	-0.05
	-0,11	*-0,04*	*0,20*	*-0,13*

Pour les coefficients de corrélation dont les valeurs sont proches de zéro, l'hypothèse d'indépendance des données est plausible.

1.2 Valeurs absolues des erreurs de prévisions

Les tableaux suivants représentent les données statistiques obtenues pour X, soit les valeurs absolues des erreurs de prévisions.

Tableau 1.12: Données statistiques de X – Prévisions jour 1 (Toutes les données)

STATIONS	Moyenne ($\bar{x}_1$)	Variance (s_{X1}^2)	Écart type (s_{X1})	Nombre de données (n)
Brotkord	5.18	72.02	8.49	228
	4,73	*53,79*	*7,33*	*183*
Chouart	4.99	77.66	8.81	240
	4,53	*65,87*	*8,12*	*192*
High Falls	5.23	72.87	8.54	245
	5,53	*79,07*	*8,89*	*122*
La Pêche	5.00	68.02	8.25	239
	4,97	*67,05*	*8,19*	*192*
Maniwaki Airport	5.35	89.99	9.49	245
	4,95	*88,9*	*9,43*	*163*
Mont St-Michel	5.4	87.72	9.37	221
	5,30	*76,26*	*8,73*	*177*

Tableau 1.13: Données statistiques de X – Prévisions jour 1 : Été et hiver (Toutes les données)

STATIONS	Été				Hiver			
	Moyenne ($\bar{x}_1$)	Variance (s_{X1}^2)	Écart type (s_{X1})	Nombre de données (n)	Moyenne ($\bar{x}_1$)	Variance (s_{X1}^2)	Écart type (s_{X1})	Nombre de données (n)
Brotkord	6.05	87.86	9.37	153	3.41	35.72	5.98	75
	5,43	*62,60*	*7,91*	*123*	*3,30*	*33,37*	*5,78*	*60*
Chouart	5.76	90,12	9.49	182	2.56	31,57	5.62	58
	5,12	*73,79*	*8,59*	*146*	*2,65*	*37,09*	*6,09*	***46***
High Falls	5,92	84,52	9,19	189	2,88	27,13	5,21	56
	6,27	*92,37*	*9,61*	*94*	*3,04*	*27,83*	*5,28*	***28***
La Pêche	5.69	80.73	8.99	184	2.69	19.14	4.37	55
	5,56	*78,82*	*8,88*	*148*	*2,96*	*23,03*	*4,8*	***44***
Maniwaki Airport	6,13	108,96	10,44	189	2,70	17.50	4,18	56
	5,58	*107,66*	*10,38*	*125*	*2,85*	*22,55*	*4,75*	***38***
Mont St-Michel	6.37	117.25	10.83	149	3,41	21.66	4.63	72
	6,13	*98,95*	*9,95*	*120*	*3,58*	*24,92*	*4,99*	*57*

Tableau 1.14: Données statistiques de X – Prévisions jour 1 (Prévisions = 0 mm)

STATIONS	Moyenne ($\bar{x}_1$)	Variance (s_{X1}^2)	Écart type (s_{X1})	Nombre de données (*n*)
Brotkord	2.28	39.11	6.25	69
Chouart	1.39	8.87	2.98	55
High Falls	2.65	50.54	7.11	87
La Pêche	2.47	41.32	6.43	89
Maniwaki Airport	2.41	30.13	5.49	84
Mont St-Michel	1.99	22.38	4.73	73

Tableau 1.15: Données statistiques de X – Prévisions p jour 1 (0<p<1)

STATIONS	Moyenne ($\bar{x}_1$)	Variance (s_{X1}^2)	Écart type (s_{X1})	Nombre de données (*n*)
Brotkord	4.47	82.98	9.11	59
Chouart	3,54	45.13	6,72	67
	3,26	*38,02*	*6,17*	*54*
High Falls	4.55	77.41	8.8	**43**
La Pêche	4.6	66.03	8.13	**41**
Maniwaki Airport	6.52	173.23	13.16	**45**
Mont St-Michel	4.57	113.46	10.65	50
	4,53	*124,78*	*11,17*	***40***

Tableau 1.16: Données statistiques de X – Prévisions p jour 1 (1≤p<5)

STATIONS	Moyenne ($\bar{x}_1$)	Variance (s_{X1}^2)	Écart type (s_{X1})	Nombre de données (*n*)
Brotkord	2,96	7,04	2.65	**49**
Chouart	4.39	89.97	9.49	69
High Falls	2,92	7,97	2.82	61
La Pêche	3.13	15.82	3.98	59
Maniwaki Airport	3.36	57,74	7,6	66
Mont St-Michel	5.2	103.02	10.15	52

Si on regroupe les prévisions des deux derniers groupes, soit 0<p<1 et 1≤p<5, on obtient le tableau suivant.

Tableau 1.17: Données statistiques de X – Prévisions p jour 1 (0<p<5)

STATIONS	Moyenne ($\bar{x}_1$)	Variance (s_{X1}^2)	Écart type (s_{X1})	Nombre de données (*n*)
Brotkord	3,78	48,70	6.98	108
Chouart	3,97	67,56	8.22	136
High Falls	3,59	36.86	6,07	104
La Pêche	3,73	36.47	6.04	100
Maniwaki Airport	4,64	105.85	10,29	111
Mont St-Michel	4.89	107.16	10.35	102

Tableau 1.18: Données statistiques de X – Prévisions p jour 1 (5≤p<10)

STATIONS	Moyenne ($\bar{x}_1$)	Variance (s_{X1}^2)	Écart type (s_{X1})	Nombre de données (*n*)
Brotkord	8.58	110.39	10.51	**26**
Chouart	7.02	38.42	6.20	**27**
High Falls	7.32	10.32	3.21	**30**
La Pêche	7.93	42.81	6.54	**26**
Maniwaki Airport	6.42	5.52	2.35	**24**
Mont St-Michel	5.53	4.69	2.17	**18**

Tableau 1.19: Données statistiques de X – Prévisions jour 1 (10≤p)

STATIONS	Moyenne ($\bar{x}_1$)	Variance (s_{X1}^2)	Écart type (s_{X1})	Nombre de données (*n*)
Brotkord	15.69	77.69	8.81	**25**
Chouart	17.75	162.52	12.75	**22**
High Falls	19.05	163.8	12.80	**24**
La Pêche	16.49	163.03	12.77	**24**
Maniwaki Airport	16.88	136.76	11.69	**26**
Mont St-Michel	16.10	100.11	10.01	**28**

Afin d'avoir un nombre plus élevé de données, si on regroupe les prévisions des deux derniers groupes, soit 5≤p<10 et 10≤p, on obtient le tableau suivant.

Tableau 1.20: Données statistiques de X – Prévisions jour 1 (5≤p)

STATIONS	Moyenne ($\bar{x}_1$)	Variance (s_{X1}^2)	Écart type (s_{X1})	Nombre de données (*n*)
Brotkord	12.07	105.37	10.27	51
Chouart	11.84	121.00	11.00	**49**
High Falls	12.53	111.36	10.55	54
La Pêche	12.04	116.99	10.82	50
Maniwaki Airport	11.86	100.24	10.01	50
Mont St-Michel	11.96	89.02	9.44	**46**

Les tableaux 1.21 et 1.22 fournissent les coefficients de corrélation entre les données et ces mêmes données mais avec un jour d'écart, soit r_1, et deux jours d'écart, soit r_2.

Tableau 1.21: Coefficients de corrélation de X - Prévisions jour 1

STATIONS	Toutes		Prévisions (p)=0		0<p<1		1≤p<5		0<p<5		5≤p<10		10≤p		5≤p	
	r_1	r_2	r_1	r_2	r_1	r_2	r_1	r_2	r_1	r_2	r_1	r_2	r_1	r_2	r_1	r_2
Brotkord	0.32 *0,10*	-0.03 *0,06*	-0,02	-0.08	-0.09	0,08	0,13	-0,10	-0.06	0.06	-0,08	-0.13	-0,17	0,08	0.14	-0.04
Chouart	0.29 *0,04*	0.00 *0,02*	0,13	0,05	0,02 *0,01*	0.29 *-0,06*	-0,06	-0,09	-0,04	0,05	-0,26	-0.02	0.29	0.31	-0.07	0,04
High Falls	0.35 *0,02*	0,00 *0,05*	-0,06	-0,02	-0,04	-0,15	0,11	-0,10	-0,01	-0,02	-0,09	0.23	-0.29	-0,12	-0.12	-0,16
La Pêche	0,25 *0,20*	-0,01 *-0,06*	-0,04	-0,06	-0,01	0,08	0.14	-0,07	-0,03	0.07	0,00	0.00	-0.33	-0,14	-0,12	-0,13
Maniwaki Airport	0.36 *0,09*	0.07 *-0,02*	0,10	0,09	0.22	-0,15	0,00	0,18	0,12	0,01	-0,36	0,04	-0,30	-0.04	-0,15	-0,06
Mont St-Michel	0,38 *0,17*	-0,01 *-0,08*	-0,01	0,05	0,32 *0,03*	-0,07 *-0,02*	-0,07	-0,13	0,20	-0,09	0,05	0,14	0,10	-0,11	-0,17	0,19

Tableau 1.22: Coefficients de corrélation de X – Été et hiver (Prévisions jour 1- Toutes les données)

STATIONS	Été		Hiver	
	r_1	r_2	r_1	r_2
Brotkord	0.31	-0.03	0.36	-0.08
	0,05	*0,04*	*0,34*	*-0,15*
Chouart	0.29	-0.05	0.15	-0.07
	-0,01	*0,01*	*0,09*	*-0,10*
High Falls	0,32	-0.02	0.41	-0.08
	0,01	*0,04*	*-0,15*	*0,00*
La Pêche	0,25	-0.04	0.03	-0.07
	0,20	*-0,08*	*-0,01*	*-0,09*
Maniwaki Airport	0,34	0.05	0.38	-0.08
	0,06	*-0,03*	*0,32*	*-0,07*
Mont St-Michel	0,37	-0.03	0,28	-0.03
	0,15	*-0,10*	*0,20*	*-0,13*

Pour les coefficients de corrélation dont les valeurs sont proches de zéro, l'hypothèse d'indépendance des données est plausible. Pour la période d'hiver des stations Brotkord et Maniwaki Airport , il est assez difficile d'obtenir un sous-ensemble de données qui sont moins corrélées tel qu'on peut le voir dans le tableau 1.22.

1.3 Valeurs absolues des erreurs de prévisions avec un seuil imposé

On ne cherche pas un modèle continu pour les erreurs inférieures à un certain seuil car on arrondit souvent les données observées. On suppose un seuil de 0,2 mm.

On pose

Y: Valeur absolue de l'erreur de prévision, étant donné qu'elle est supérieure à 0,2 mm
Z = Y – 0,2

Le paramètre estimé $\hat{\lambda}_z$ est relié à la loi exponentielle tandis que les paramètres $\hat{\theta}_z$ et $\hat{c}_z$ sont reliés à la loi de Pareto généralisée. Ces paramètres seront définis ultérieurement.

Tableau 1.23: Données statistiques de Y – Prévisions jour 1 (Toutes les données supérieures à 0,2 mm)

STATIONS	Moyenne ($\bar{y}_I$)	Variance (s_{YI}^2)	Écart type (s_{YI})	Nombre de données (n)	$\hat{\lambda}_z$	$\hat{\theta}_z$	$\hat{c}_z$
Brotkord	7.267	86.413	9.296	162	0.1415	6.7388	0.0378
Chouart	6.377	90.964	9.538	187	0.1619	5.4450	0.0662
High Falls	7.180	86.459	9.298	178	0.1433	6.5820	0.0400
La Pêche	6.768	80.574	8.976	176	0.1523	6.3047	0.0461
Maniwaki Airport	7.232	108.303	10.407	181	0.1422	5.6804	0.0531
Mont St-Michel	7.303	105.298	10.261	163	0.1408	5.8398	0.0496

Tableau 1.24: Données statistiques de Y – Prévisions jour 1 (Toutes les données non corrélées, supérieures à 0,2 mm)

STATIONS	Moyenne ($\bar{y}_I$)	Variance (s_{YI}^2)	Écart type (s_{YI})	Nombre de données (n)	$\hat{\lambda}_z$	$\hat{\theta}_z$	$\hat{c}_z$
Brotkord	*6,682*	*63,430*	*7,964*	*129*	*0,1543*	*7,9243*	*0,0313*
Chouart	*5,775*	*77,290*	*8,791*	*150*	*0,1794*	*5,3452*	*0,0765*
High Falls	*7,022*	*80,530*	*8,219*	*142*	*0,1466*	*6,7384*	*0,0392*
La Pêche	*6,740*	*79,532*	*8,918*	*141*	*0,1529*	*6,3270*	*0,0460*
Maniwaki Airport	*6,825*	*110,184*	*10,487*	*118*	*0,1509*	*5,3241*	*0,0649*
Mont St-Michel	*7,196*	*90,467*	*9,511*	*130*	*0,1429*	*6,3574*	*0,0426*

Dans le cas suivant, les données de départ ne sont pas corrélées.

Tableau 1.25: Données statistiques de Y – Prévisions jour 1 (0<p<5 non corrélées, supérieures à 0,2 mm)

STATIONS	Moyenne ($\bar{y}_1$)	Variance (s_{Y1}^2)	Écart type (s_{Y1})	Nombre de données (n)	$\hat{\lambda}_z$	$\hat{\theta}_z$	$\hat{c}_z$
Brotkord	4.568	55.679	7.462	89	0.2289	5.0426	0.1121
Chouart	4.484	74.382	8.624	120	0.2334	4.6551	0.1410
High Falls	3.960	39.397	6.277	94	0.2660	5.1194	0.1255
La Pêche	4.426	40.468	6.361	84	0.2366	5.5798	0.0917
Maniwaki Airport	4.994	112.412	10.602	103	0.2086	4.5140	0.1378
Mont St-Michel	5.710	121.204	11.009	87	0.1815	4.6684	0.1088

1.4 Modélisation

On ne connaît pas à priori la loi de probabilité suivie par la variable aléatoire qui représente l'erreur de prévision. Pour les erreurs de prévisions, les modèles possibles considérés sont les suivants:

- Loi normale (Loi de Gauss)
- Loi lognormale
- Loi exponentielle double (Loi de Laplace)
- Loi exponentielle
- Loi de Pareto généralisée

Lors des tests d'ajustement permettant de vérifier si le modèle testé est le bon, on cherchera à vérifier les hypothèses qui suivent:

H_0 : L'ajustement entre les n_i et m_i est acceptable. Les différences entre ces fréquences ne sont pas significatives. Le modèle proposé est acceptable.

H_1 : Le modèle proposé n'est pas acceptable.

Noter que m_i représente le nombre de données que devrait compter le i^e intervalle et n_i est le nombre de données dans le i^e intervalle, utilisés pour les tests d'ajustement de

Pearson. De plus, pour ces derniers, bien qu'il soit préférable que la taille de l'échantillon soit d'au moins 50, certains échantillons dont la taille est un peu inférieure sont tout de même considérés. Mais le choix du nombre d'intervalles tient compte du fait que le nombre de données contenu dans les intervalles doit être au moins de cinq.

En ce qui concerne les tests de normalité effectués, on a choisi celui d'Anderson-Darling, lequel est disponible avec MINITAB, le logiciel utilisé. De plus, chaque valeur *p*, c'est-à-dire la plus petite valeur du seuil de signification du test pour laquelle le modèle est rejeté, est également calculée à l'aide de MINITAB.

1.4.1 Loi normale

La fonction de densité de probabilité de la variable aléatoire W est de la forme suivante, avec une espérance μ et une variance σ^2:

$$f_W(w;\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-(w - \mu)^2/2\sigma^2\right] \quad \text{pour } w \in \Re$$

$P(a < W < b) = P\left((a-\mu)/\sigma < Z < (b-\mu)/\sigma\right)$ où Z est une variable aléatoire normale centrée réduite

Pour vérifier si les données proviennent d'une loi normale, on effectue un test de normalité d'Anderson-Darling avec les valeurs des erreurs de prévisions W qui ne sont pas ou peu corrélées. On effectue le test lorsqu'on prévoyait entre 0 et 5 mm de précipitations, mais également sur les deux sous-groupes associés suivants: entre 0 et 1 mm et 1 et 5 mm. Remarquer que pour les tests subséquents, les catégories présentées dans le tableau 1.26 ne seront pas nécessairement les mêmes.

Tableau 1.26: Statistiques du test d'Anderson-Darling de W jour 1 – Loi normale

STATIONS	Toutes les données	Été	Hiver	p = 0 mm	0<p<1	1≤p<5	0<p<5	5≤p
Brotkord	10.25	6.13	3.04	15.77	9.55	3.07	12.51	3.47
Chouart	18.56	12.11	6.83	10.46	6.77	12.88	19.07	1.73
High Falls	9.88	6.89	2.74	19.25	6.62	4.21	10.99	3.08
La Pêche	15.42	11.33	3.44	18.15	8.24	5.14	9.83	3.79
Maniwaki Airport	17.17	12.95	3.22	17.03	7.52	12.28	18.30	2.26
Mont St-Michel	13.82	9.53	5.96	13.98	7.48	9.58	16.23	1.99

En conclusion, puisque toutes les valeurs *p* sont à peu près égales à zéro, la variable aléatoire W ne suit pas une loi normale, quelle que soit la station ou la catégorie considérée.

Si on effectue maintenant un test de normalité d'Anderson-Darling sur les valeurs de la variable aléatoire X dans le cas où on prévoyait entre 5 et 10 mm de précipitations, on obtient une valeur *p* de 0,204 pour Brotkord si on élimine une donnée aberrante, soit 58,63. On obtient une valeur *p* de 0,177 pour la station Chouart si on enlève la valeur aberrante 36,02. Mais, dans ce cas, le coefficient de corrélation est de –0,26. La loi normale pourrait donc être acceptable pour la valeur absolue de l'erreur dans le cas où on prévoyait entre 5 et 10 mm.

Vérifions la normalité par un test de normalité d'Anderson-Darling sur les valeurs de la variable aléatoire X dans le cas où on prévoyait au moins 5 mm de précipitations. On obtient des valeurs *p* proches de zéro. Dans ce cas-ci, la loi normale n'est pas acceptable.

Finalement, on réalise un test de normalité d'Anderson-Darling sur les valeurs de la variable aléatoire X dans le cas où on prévoyait au moins 10 mm de précipitations. Pour la station Chouart, la valeur *p* est de 0,170, ce qui est acceptable. De plus, le coefficient de corrélation est de 0,29. On ne peut conclure que la loi normale pourrait être acceptable, en général, pour la valeur absolue de l'erreur dans le cas où on prévoyait plus de 10 mm car pour les autres stations, la valeur p est près de zéro. Nous verrons toutefois, à la section suivante, qu'on peut faire beaucoup mieux en considérant la loi lognormale.

1.4.2 Loi lognormale

La fonction de densité de probabilité de la variable aléatoire X est de la forme suivante, avec une espérance μ et une variance σ^2:

$$f_X(x;\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-(\ln x - \mu)^2 / 2\sigma^2\right]$$

Pour vérifier si les erreurs significatives lorsqu'on ne prévoyait aucune précipitation proviennent d'une loi lognormale, on effectue un test de normalité d'Anderson-Darling avec le logarithme des valeurs Y, soit les valeurs absolues des erreurs de prévisions, étant donné qu'elles sont supérieures à 0,2 mm.

- Brotkord – valeur $p = 0{,}826$ (22 données)
- Chouart – valeur $p \approx 0$ (18 données)
- High Falls – valeur $p = 0{,}325$ (30 données)
- La Pêche – valeur $p \approx 0$ (43 données)
- Maniwaki Airport – valeur $p \approx 0$ (29 données)
- Mont St-Michel – valeur $p \approx 0$ (30 données)

Puisqu'on obtient de très bonnes valeurs p pour deux stations, on peut conclure que, dans le cas où on ne prévoyait aucune précipitation, la valeur absolue de l'erreur de prévision, étant donné qu'elle est supérieure à 0,2 mm, suit parfois une loi lognormale.

On effectue maintenant un test de normalité d'Anderson-Darling avec le logarithme des valeurs X (valeur absolue de l'erreur de prévision) lorsqu'on prévoyait de 5 à 10 mm de précipitations. Pour toutes les stations, la valeur p est à peu près égale à zéro.

Finalement, on effectue un test de normalité d'Anderson-Darling avec le logarithme des valeurs X dans le cas où on prévoyait au moins 10 mm de précipitations.

- Brotkord – valeur $p = 0{,}980$
- Chouart – valeur $p \approx 0$
- High Falls – valeur $p = 0{,}398$ (coefficient de corrélation = -0,29)
- La Pêche – valeur $p \approx 0$
- Maniwaki Airport – valeur $p \approx 0$
- Mont St-Michel – valeur $p \approx 0$

On a d'excellentes valeurs p pour les stations Brotkord et High Falls. On peut conclure que la valeur absolue de l'erreur de prévision, quand on prévoit au moins 10 mm de précipitations, suit parfois une loi lognormale.

1.4.3 Loi exponentielle double

La fonction de densité de probabilité de la variable aléatoire W (valeur de l'erreur de prévision) est de la forme suivante :

$$f_W(w\,;\lambda_w) = \tfrac{1}{2}\ \lambda_w \exp\left(-\lambda_W |w|\right) \quad \text{pour } w \in \Re \text{ et } \lambda_w > 0$$

$$P(a < W < b) = \int_a^b \tfrac{1}{2}\ \lambda_w \exp\left(-\lambda_W |w|\right) dw = \tfrac{1}{2}\ e^{-\lambda_W a} - \tfrac{1}{2}\ e^{-\lambda_W b} \quad \text{pour a et b} > 0$$

Pour estimer le paramètre λ_W, on pose

$$E[W^2] = \sum_{k=1}^{n} \frac{W_k^2}{n} \Leftrightarrow 2/\lambda_w^2 = \sum_{k=1}^{n} \frac{W_k^2}{n}$$

$$\Rightarrow \hat{\lambda}_w = \left(\frac{2n}{\sum_{k=1}^{n} W_k^2} \right)^{1/2}$$

On peut aussi écrire que

$$\mathrm{Var}[W] = s_w^2 \Leftrightarrow 2/\lambda_w^2 = s_w^2$$

$$\Rightarrow \hat{\lambda}_w = \frac{\sqrt{2}}{s_w}$$

Pour un nombre de données n assez grand, les deux estimateurs seront presque égaux. Par conséquent, on peut écrire que

$$\hat{\lambda}_w = \left(\frac{2n}{\sum_{k=1}^{n} W_k^2} \right)^{1/2} \approx \frac{\sqrt{2}}{S_W}$$

On effectue un test d'ajustement de Pearson avec les valeurs des erreurs de prévisions qui ne sont pas ou peu corrélées.

Pour les groupes non considérés, soit que le nombre de données disponibles est trop petit ou qu'il n'est pas possible de créer des intervalles ayant tous au moins cinq données.

Tableau 1.27: Valeurs estimées du paramètre λ_W de W jour 1 – Loi exponentielle double

STATIONS	Toutes les données	Été	Hiver	0<p<5	5≤ p
Brotkord	0,13	0,11	0,26		
Chouart	0,15	0,14	0,21		
High Falls	0,14	0,12		0,21	0,11
La Pêche	0,15	0,14	0,26	0,21	0,11
Maniwaki Airport	0,13	0,12		0,13	0,13
Mont St-Michel	0,14	0,12	0,24	0,13	0,14

Tableau 1.28: Test d'ajustement de Pearson de W – Prévisions jour 1 (Toutes les données) – Loi exponentielle double

STATIONS	i	(-∞, -15]	(-15, -5]	(-5, 0)	[0; 5)	[5; 15)	[15; ∞)	
Brotkord	n_i	9	10	31	68	25	9	
	p_i	0.0711	0.1899	0.2390	0.2390	0.1899	0.0711	
	m_i	10.81	28.86	36.32	36.32	28.86	10.81	$D^2 = 41{,}85$
Chouart	n_i	7	14	54	91	18	8	
	p_i	0.0527	0.1835	0.2638	0.2638	0.1835	0.0527	
	m_i	10.12	35.23	50.65	50.65	35.23	10.12	$D^2 = 54{,}98$
High Falls	n_i	5	8	33	50	16	10	
	p_i	0.0612	0.1871	0.2517	0.2517	0.1871	0.0612	
	m_i	7.47	22.82	30.71	30.71	22.82	7.47	$D^2 = 25{,}63$
La Pêche	n_i	8	15	48	87	26	8	
	p_i	0.0527	0.1835	0.2638	0.2638	0.1835	0.0527	
	m_i	10.12	35.23	50.65	50.65	35.23	10.12	$D^2 = 41{,}14$
Maniwaki Airport	n_i	8	16	36	79	18	6	
	p_i	0.0711	0.1899	0.2390	0.2390	0.1899	0.0711	
	m_i	11.60	30.95	38.95	38.95	30.95	11.60	$D^2 = 57{,}85$
Mont St-Michel	n_i	7	17	38	81	25	9	
	p_i	0.0612	0.1871	0.2517	0.2517	0.1871	0.0612	
	m_i	10.84	33.11	44.55	44.55	33.11	10.84	$D^2 = 42{,}28$

Calculs - Brotkord

Calculs des probabilités

Le modèle est symétrique par rapport à l'origine. Par conséquent,

- $P(-\infty < W \leq -15) = P(15 \leq W < \infty) = ½\ e^{-0,13*15} - ½\ e^{-\infty} = 0{,}0711$
- $P(-15 < W \leq -5) = P(5 \leq W < 15) = ½\ e^{-0,13*5} - ½\ e^{-0,13*15} = 0{,}1899$
- $P(-5 < W < 0) = P(0 \leq W < 5) = ½\ e^{0} - ½\ e^{-0,13*5} = 0{,}2390$

Calcul de la statistique D^2 de Pearson

$$D^2 = \frac{(n_1 - m_1)^2}{m_1} + \ldots + \frac{(n_6 - m_6)^2}{m_6}$$

$$= \frac{(9-10{,}81)^2}{10{,}81} + \ldots + \frac{(9-10{,}81)^2}{10{,}81}$$

$= 41{,}85 \Rightarrow$ L'ajustement est très mauvais.

Nombre de degrés de liberté

k : Nombre de classes = 6

r : Nombre de paramètres estimés = 1

n : Nombre de degrés de liberté = k – r – 1 = 4

Soit $P\left(\chi_n^2 > \chi_n^2(\alpha)\right) = \alpha$ où α représente le risque. Pour un risque α de 5%, c'est-à-dire un niveau de confiance de 95%, et quatre degrés de liberté, on trouve un seuil de décision $\chi_4^2(0{,}05)$ de 9,4877. Pour un risque α de 1%, c'est-à-dire un niveau de confiance de 99%, on trouve un seuil de décision $\chi_4^2(0{,}01)$ de 13,2767. Puisque D^2 = 41,85 > 13,2767, on rejette l'hypothèse H_0. Par conséquent, dans le cas où on considère l'ensemble des données, la variable aléatoire W ne suit pas une loi exponentielle double.

Tableau 1.29: Test d'ajustement de Pearson de W – Prévisions jour 1 (Été) – Loi exponentielle double

STATIONS	i	(-∞, -15]	(-15, -5]	(-5, 0)	[0; 5)	[5; 15)	[15; ∞)	
Brotkord	n_i	9	6	21	38	20	8	
	p_i	0.0960	0.1924	0.2115	0.2115	0.1924	0.0960	
	m_i	9.79	19.63	21.58	21.58	19.63	9.79	$D^2 = 22{,}38$
Chouart	n_i	6	12	40	63	18	7	
	p_i	0.0612	0.1871	0.2517	0.2517	0.1871	0.0612	
	m_i	8.94	27.31	36.75	36.75	27.31	8.94	$D^2 = 32{,}18$
High Falls	n_i	5	7	24	37	12	9	
	p_i	0.0826	0.1918	0.2256	0.2256	0.1918	0.0826	
	m_i	7.77	18.03	21.21	21.21	18.03	7.77	$D^2 = 22{,}07$
La Pêche	n_i	8	11	39	63	20	7	
	p_i	0.0612	0.1871	0.2517	0.2517	0.1871	0.0612	
	m_i	9.06	27.69	37.25	37.25	27.69	9.06	$D^2 = 30{,}66$
Maniwaki Airport	n_i	7	15	26	57	15	5	
	p_i	0.0826	0.1918	0.2256	0.2256	0.1918	0.0826	
	m_i	10.33	23.97	28.20	28.20	23.97	10.33	$D^2 = 40{,}12$
Mont St-Michel	n_i	6	14	27	49	18	6	
	p_i	0.0826	0.1918	0.2256	0.2256	0.1918	0.0826	
	m_i	9.92	23.01	27.07	27.07	23.01	9.92	$D^2 = 25{,}48$

Tableau 1.30: Test d'ajustement de Pearson de W – Prévisions jour 1 (Hiver) – Loi exponentielle double

STATIONS	i	(-∞, -3]	(-3, 0)	[0; 3)	[3; ∞)	
Brotkord	n_i	6	8	25	11	
	p_i	0.2292	0.2708	0.2708	0.2292	
	m_i	11.46	13.54	13.54	11.46	$D^2 = 14{,}59$
Chouart	n_i	5	12	23	6	
	p_i	0.2663	0.2337	0.2337	0.2663	
	m_i	12.25	10.75	10.75	12.25	$D^2 = 21{,}58$
La Pêche	n_i	5	8	21	10	
	p_i	0.2663	0.2337	0.2337	0.2663	
	m_i	11.72	10.28	10.28	11.72	$D^2 = 15{,}78$
Mont St-Michel	n_i	8	7	27	15	
	p_i	0.2434	0.2566	0.2566	0.2434	
	m_i	13.87	14.63	14.63	13.87	$D^2 = 17.02$

Tableau 1.31: Test d'ajustement de Pearson de W – Prévisions jour 1 (0<p<5) – Loi exponentielle double

STATIONS	i	(-∞, -3]	(-3, 0)	[0; 3)	[3; ∞)	
High Falls	n_i	22	20	51	11	
	p_i	0.2663	0.2337	0.2337	0.2663	
	m_i	27.69	24.31	24.31	27.69	$D^2 = 41{,}32$
La Pêche	n_i	23	13	51	13	
	p_i	0.2663	0.2337	0.2337	0.2663	
	m_i	26.63	23.37	23.37	26.63	$D^2 = 44{,}74$
Maniwaki Airport	n_i	24	20	53	14	
	p_i	0.3385	0.1615	0.1615	0.3385	
	m_i	37.58	17.92	17.92	37.58	$D^2 = 88{,}58$
Mont St-Michel	n_i	26	12	52	12	
	p_i	0.3385	0.1615	0.1615	0.3385	
	m_i	34.53	16.47	16.47	34.53	$D^2 = 94.67$

Pour toutes les stations des tableaux précédents, la valeur D^2 est grande. C'est donc que les ajustements sont mauvais. On rejette H_0. La variable aléatoire W ne suit pas une loi exponentielle double. Ainsi, le modèle exponentiel double n'est pas acceptable pour les données de chacune des stations, quel que soit le sous-ensemble de données considéré.

1.4.4 Loi exponentielle

Soit

Y : Valeur absolue de l'erreur de prévision, étant donné qu'elle est supérieure à 0,2 mm

$Z = Y - 0{,}2$

La fonction de densité de probabilité de la variable aléatoire Z est de la forme suivante:

$f_Z(z\,;\lambda_Z) = \lambda_Z e^{-\lambda_Z}$ pour $\lambda_Z > 0$

$P(a < Z < b) = \int_a^b \lambda_Z\, e^{-\lambda_Z} = e^{-\lambda_Z a} - e^{-\lambda_Z b}$ pour a et b > 0

$\hat{\lambda}_Z = \dfrac{1}{\bar{Z}_1}$

Noter que les valeurs de $\hat{\lambda}_Z$ utilisées se retrouvent dans les tableaux 1.23, 1.24 et 1.25.

En premier lieu, on effectue un test d'ajustement de Pearson sur toutes les données. Dans le tableau suivant, seule la station Brotkord obtient une valeur *p* pouvant être acceptable.

Tableau 1.32: Test d'ajustement de Pearson 1 de Z – Prévisions jour 1 (Toutes les données) – Loi exponentielle

STATIONS	i	(0; 3)	[3; 7,5)	[7,5; 15)	[15; ∞)	
Brotkord	n_i	66	45	30	21	
	p_i	0.3459	0.3081	0.2263	0.1197	
	m_i	56.04	49.91	36.66	19.40	$D^2 = 3{,}60$; $p = 0{,}1653$
Chouart	n_i	97	50	20	20	
	p_i	0.3847	0.3183	0.2088	0.0882	
	m_i	71.94	59.53	39.04	16.49	$D^2 = 20{,}29$
High Falls	n_i	78	50	28	22	
	p_i	0.3494	0.3092	0.2249	0.1166	
	m_i	62.19	55.03	40.03	20.75	$D^2 = 8{,}17$
La Pêche	n_i	80	51	25	20	
	p_i	0.3667	0.3141	0.2173	0.1019	
	m_i	64.53	55.29	38.25	17.93	$D^2 = 8{,}87$
Maniwaki Airport	n_i	83	51	24	23	
	p_i	0,3473	0.3085	0,2257	0.1185	
	m_i	62.86	55.84	40.86	21.44	$D^2 = 13{,}94$
Mont St-Michel	n_i	73	46	24	20	
	p_i	0.3445	0.3076	0.2269	0.1210	
	m_i	56.15	50.14	36.98	19.73	$D^2 = 9{,}95$

<u>Calculs</u> - Brotkord

Calculs des probabilités

$\hat{\lambda}_Z = \frac{1}{\bar{Z}_1} = (7{,}267 - 0{,}2)^{-1} = 0{,}1415$

- $P(0 < Z < 3) = 1 - e^{-0{,}1415*3} = 0{,}3459$
- $P(3 \leq Z < 7{,}5) = e^{-0{,}1415*3} - e^{-0{,}1415*7{,}5} = 0{,}3081$
- $P(7{,}5 \leq Z < 15) = e^{-0{,}1415*7{,}5} - e^{-0{,}1415*15} = 0{,}2263$
- $P(15 \leq Z < \infty) = e^{-0{,}1415*15} = 0{,}1197$

Calcul de la statistique D^2 de Pearson

$$D^2 = \frac{(n_1 - m_1)^2}{m_1} + \ldots + \frac{(n_4 - m_4)^2}{m_4}$$

$$= \frac{(66 - 56{,}04)^2}{56{,}04} + \ldots + \frac{(21 - 19{,}04)^2}{19{,}04}$$

$$= 3{,}60$$

Nombre de degrés de liberté

k : Nombre de classes = 4

r : Nombre de paramètres estimés = 1

n : Nombre de degrés de liberté = k – r – 1 = 2

Pour un risque α de 5% et 2 degrés de liberté, on trouve un seuil de décision $\chi^2_2(0{,}05)$ de 5,9914. Puisque $D^2 = 3{,}60 < 5{,}9914$, on ne rejette l'hypothèse H_0.

Valeur $p = P(D^2_{ddl} \geq D^2_{observé})$

$= P(D^2_2 \geq 3{,}60)$

$= 0{,}1653$ ce qui est acceptable

Dans le tableau suivant, on voit que si on change les bornes des intervalles, la valeur p pour la station High Falls est acceptable et bonne pour la station La Pêche.

Tableau 1.33: Test d'ajustement de Pearson 2 de Z – Prévisions jour 1 (Toutes les données) – Loi exponentielle

STATIONS	i	(0; 5)	[5; 8)	[8; 10)	[10; ∞)		
Chouart	n_i	124	26	8	29		
	p_i	0.5549	0.1712	0.0757	0.1981		
	m_i	103.77	32.02	14.16	37.05	$D^2 = 9{,}51$	
High Falls	n_i	103	28	9	38		
	p_i	0.5115	0.1707	0.0792	0.2387		
	m_i	91.04	30.38	14.10	42.48	$D^2 = 4{,}07$	$p = 0{,}1307$
La Pêche	n_i	104	28	11	33		
	p_i	0.5329	0.1713	0.0777	0.2182		
	m_i	93.79	30.14	13.67	38.40	$D^2 = 2{,}54$	$p = 0{,}2808$
Maniwaki Airport	n_i	108	30	9	34		
	p_i	0.5089	0.1706	0.0794	0.2412		
	m_i	92.10	30.87	14.36	43.66	$D^2 = 6{,}91$	
Mont St-Michel	n_i	97	24	10	32		
	p_i	0.5054	0.1704	0.0796	0.2447		
	m_i	82.37	27.78	12.97	39.88	$D^2 = 5{,}35$	$p = 0{,}0689$

Si on reprend les calculs avec des données non corrélées, on obtient en plus une très bonne valeur p pour le Mont St-Michel. On pourrait donc suggérer une loi exponentielle comme modèle pour Z lorsqu'on considère toutes les données. Mais on verra à la section suivante qu'on peut faire beaucoup mieux en considérant une loi de Pareto généralisée.

Tableau 1.34: Test d'ajustement de Pearson 2 de Z – Prévisions jour 1 (Toutes les données non corrélées) – Loi exponentielle

STATIONS	i	(0; 5)	[5; 8)	[8; 10)	[10; ∞)		
Chouart	n_i	105	19	4	22		
	p_i	0.5922	0.1697	0.0718	0.1663		
	m_i	88.82	25.46	10.77	24.95	$D^2 = 9.19$	
High Falls	n_i	82	24	5	31		
	p_i	0.5195	0.1710	0.0787	0.2309		
	m_i	73.77	24.28	11.17	32.79	$D^2 = 4.43$	p = 0,1092
La Pêche	n_i	85	20	9	27		
	p_i	0.5344	0.1713	0.0775	0.2167		
	m_i	75.36	24.15	10.93	30.56	$D^2 = 2.70$	p = 0.2592
Maniwaki Airport	n_i	72	23	4	19		
	p_i	0,5299	0.1712	0.0779	0.2210		
	m_i	62.52	20.20	9.19	26.08	$D^2 = 6.68$	
Mont St-Michel	n_i	75	20	9	26		
	p_i	0.5107	0.1706	0.0792	0.2395		
	m_i	66.39	22.18	10.30	31.13	$D^2 = 2.34$	p = 0.3104

Le test d'ajustement suivant est effectué pour le cas où on prévoyait entre 0 et 5 mm de précipitations.

Tableau 1.35: Test d'ajustement de Pearson de Z – Prévisions jour 1 (0<p<5) – Loi exponentielle

STATIONS	i	(0; 1)	[1; 2,5)	[2,5; 4)	[4; 10)	[10; ∞)	
Brotkord	n_i	24	26	16	16	7	
	p_i	0.2046	0.2312	0.1640	0.2989	0.1013	
	m_i	18.21	20.57	14.59	26.60	9.02	$D^2 = 8.08$
Chouart	n_i	43	28	22	16	11	
	p_i	0.2082	0.2339	0.1648	0.2962	0.0969	
	m_i	24.98	28.07	19.78	35.55	11.63	$D^2 = 24.03$
High Falls	n_i	23	34	17	12	8	
	p_i	0.2335	0.2521	0.1692	0.2752	0.0700	
	m_i	21.95	23.70	15.90	25.86	6.58	$D^2 = 12.34$
La Pêche	n_i	20	27	14	14	9	
	p_i	0.2107	0.2358	0.1654	0.2943	0.0938	
	m_i	17.70	19.81	13.89	24.72	7.88	$D^2 = 7.72$
Maniwaki Airport	n_i	28	34	19	14	8	
	p_i	0.1883	0.2181	0.1595	0.3100	0.1242	
	m_i	19.39	22.46	16.43	31.93	12.79	$D^2 = 22,01$
Mont St-Michel	n_i	22	25	15	17	8	
	p_i	0.1660	0.1988	0.1514	0.3210	0.1629	
	m_i	14.44	17.29	13.17	27.93	14.17	$D^2 = 14,61$

Dans le cas où on prévoyait entre 0 et 5 mm, comme les valeurs p sont à peu près égales à zéro, on rejette l'hypothèse H_0. On peut donc conclure que pour ce sous-groupe, la valeur aléatoire Z ne suit pas une loi exponentielle.

1.4.5 Loi de Pareto généralisée

Soit

Y : Valeur absolue de l'erreur de prévision, étant donné qu'elle est supérieure à 0,2 mm

Y': Valeur de l'erreur de prévision, étant donné qu'elle est supérieure à 0,2 mm et inférieure à 0,2 mm

Z = Y – 0,2

Z' = Y' – 0,2 pour Y' > 0 et Z' = Y' + 0,2 pour Y' < 0

La fonction de densité de probabilité des variables aléatoires Z' et Z est de la forme suivante:

$f_{Z'}(z',c_{z'},\theta_{z'}) = (c_{z'}/2)(\theta_{z'}-2)(1+c_{z'}|z'|)^{1-\theta_{z'}}$

$P(a \leq Z' \leq b) = \int_a^b (c_{z'}/2)(\theta_{z'}-2)(1+c_{z'}|z'|)^{1-\theta_{z'}}\,dz' = ½\,(1+c_{z'}a)^{2-\theta_{z'}} - ½\,(1+c_{z'}b)^{2-\theta_{z'}}$ -

pour a et b ≥ 0

$f_Z(z;c_z,\theta_z) = c_z(\theta_z-2)(1+c_z z)^{1-\theta_z}$ pour $z \geq 0$, $c_z > 0$ et $\theta_z > 2$

$P(a \leq Z \leq b) = \int_a^b c_z(\theta_z-2)(1+c_z z)^{1-\theta_z}\,dz = (1+c_z a)^{2-\theta_z} - (1+c_z b)^{2-\theta_z}$

<u>Paramètres</u>

Pour pouvoir estimer le paramètre θ d'une loi de Pareto généralisée par la méthode des moments, la moyenne ne doit pas être supérieure à l'écart type. Pour Z, on aura

$\hat{\theta}_z = 2 + (2s_{Z1}^2 / (s_{Z1}^2 - \bar{z}_1^2))$

Lorsque le paramètre θ_Z est estimé, on peut ensuite estimer le paramètre c_Z. Deux méthodes sont alors possibles.

Pour estimer le paramètre c_Z à partir de l'équation suivante, on utilise un centile de Z, soit le 75^e des données. Le 75^e centile des données, un compromis entre la médiane et le 95^e centile, permet généralement d'obtenir un bon test d'ajustement.

$$\hat{c}_z = \frac{1}{z_{0,75}}\left(\frac{1}{4^{1/(2-\hat{\theta}_z)}} - 1\right)$$

Mais plutôt que d'utiliser le 75^e centile des données pour estimer c_Z, on préfère se servir de l'équation suivante:

$E[Z] = 1/(\hat{c}_Z(\hat{\theta}_Z - 3)) \quad \Rightarrow \quad \hat{c}_Z = 1/(\bar{z}(\hat{\theta}_Z - 3))$ pour $\hat{\theta}_Z > 3$

Noter que les valeurs de $\hat{\theta}_z$ et $\hat{c}_z$ utilisées se retrouvent dans les tableaux 1.23, 1.24 et 1.25. On effectue en premier lieu un test d'ajustement de Pearson pour Z dans le cas où on considère toutes les données.

Tableau 1.36: Test d'ajustement de Pearson 1 de Z – Prévisions jour 1 (Toutes les données) – Loi de Pareto généralisée

STATIONS	i	(0; 3)	[3; 7,5)	[7,5; 15)	[15; ∞)		
Brotkord	n_i	66	45	30	21		
	p_i	0.3993	0.2947	0.1873	0.1188		
	m_i	64.68	47.74	30.34	19.24	$D^2 = 0,35$	p = 0,5541
Chouart	n_i	97	50	20	20		
	p_i	0.4643	0.2864	0.1564	0.0929		
	m_i	86.82	53.55	29.25	17.37	$D^2 = 4,75$	p = 0,0293
High Falls	n_i	78	50	28	22		
	p_i	0.4050	0.2944	0.1845	0.1161		
	m_i	72.09	52.40	32.84	20.66	$D^2 = 1,39$	p = 0,2384
La Pêche	n_i	80	51	25	20		
	p_i	0.4272	0.2941	0.1745	0.1042		
	m_i	75.19	51.75	30.71	18.34	$D^2 = 1,53$	p = 0,2161
Maniwaki Airport	n_i	83	51	24	23		
	p_i	0.4193	0.2892	0.1755	0.1159		
	m_i	75.90	52.35	31.77	20.98	$D^2 = 2,79$	p = 0,0949
Mont St-Michel	n_i	73	46	24	20		
	p_i	0.4128	0.2902	0.1788	0.1183		
	m_i	67.29	47.30	29.14	19.28	$D^2 = 1,45$	p = 0,2285

<u>Calculs</u> - Brotkord

<u>Calcul des paramètres</u>

Puisque la moyenne n'est pas supérieure à l'écart type, on peut estimer le paramètre θ_z d'une loi de Pareto généralisée par la méthode des moments. La déviation standard demeure la même.

$$\hat{\theta}_z = 2 + (2s_{Z1}^2 / (s_{Z1}^2 - \bar{z}_1^2)) = 2 + (2*86,413 / (86,413 - (7,267-0,2)^2)) = 6,7388$$

$$E[Z] = 1/\left(\hat{c}_z\left(\hat{\theta}_z - 3\right)\right) \Rightarrow \hat{c}_z = 1/\left(\bar{z}\left(\hat{\theta}_z - 3\right)\right) = \frac{1}{(7,267-0,2)(6,7388-3)} = 0,0378$$

<u>Calculs des probabilités</u>

$$P(a \le Z \le b) = \int_a^b \hat{c}_z\left(\hat{\theta}_z - 2\right)\left(1+\hat{c}_z z\right)^{1-\hat{\theta}_z} dz = \left(1+\hat{c}_z a\right)^{2-\hat{\theta}_z} - \left(1+\hat{c}_z b\right)^{2-\hat{\theta}_z}$$

$$= (1 + 0{,}0378a)^{-4{,}739} - (1 + 0{,}0378b)^{-4{,}739}$$

- $P(0 \leq Z < 3) = 1 - (1 + 0{,}0378*3)^{-4{,}7388} = 0{,}3993$
- $P(3 \leq Z < 7{,}5) = (1 + 0{,}0378*3)^{-4{,}7388} - (1 + 0{,}0378*7{,}5)^{-4{,}7388} = 0{,}2947$
- $P(7{,}5 \leq Z < 15) = (1 + 0{,}0378*7{,}5)^{-4{,}7388} - (1 + 0{,}0378*15)^{-4{,}7388} = 0{,}1873$
- $P(15 \leq Z < \infty) = (1 + 0{,}0378*15)^{-4{,}7388} = 0{,}1188$

Nombre de degrés de liberté

k : Nombre de classes = 4

r : Nombre de paramètres estimés = 2

n : Nombre de degrés de liberté = k – r – 1 = 1

Pour un risque α de 5% et 1 degré de liberté, on trouve un seuil de décision $\chi_1^2(0{,}05)$ de 3,8414. Puisque $D^2 = 0{,}35 < 3{,}8414$, on ne rejette pas l'hypothèse H_0.

Valeur $p = P(D^2_{ddl} \geq D^2_{observé})$

$= P(D^2_1 \geq 0{,}35)$

$= 0{,}5541$ ce qui est très bon

Tableau 1.37: Test d'ajustement de Pearson 1 de Z – Prévisions jour 1 (Toutes les données non corrélées) – Loi de Pareto généralisée

STATIONS	i	(0; 3)	[3; 7,5)	[7,5; 15)	[15; ∞)		
Brotkord	n_i	54	37	21	17		
	p_i	0.4127	0.3009	0.1844	0.1021		
	m_i	53.23	38.82	23.78	13.17	D^2 = 1,54	p = 0.2146
High Falls	n_i	62	42	21	17		
	p_i	0.4096	0.2956	0.1831	0.1117		
	m_i	58.16	41.98	25.99	15.86	D^2 = 1,29	p = 0.2560
La Pêche	n_i	64	40	21	16		
	p_i	0.4282	0.2942	0.1742	0.1034		
	m_i	60.37	41.49	24.56	14.58	D^2 = 0,93	p = 0,3349
Mont St-Michel	n_i	57	36	21	16		
	p_i	0.2364	0.1713	0.2933	0.0858		
	m_i	30.74	22.27	38.13	11.15	D^2 = 4,75	p = 0,0293

Si on regarde le premier des deux tableaux précédents, on remarque qu'on obtient de très bonnes valeurs p pour les stations Brotkord, High Falls, La Pêche et Maniwaki Airport. Par conséquent, dans le cas où on considère toutes les données, l'erreur Z pourrait suivre le modèle de la loi de Pareto généralisée. Le modèle proposé est accepté encore plus fortement si on considère des données de ce sous-ensemble qui sont très peu corrélées entre elles. Essayons de voir si on peut faire un test d'ajustement de Pearson encore meilleur.

Tableau 1.38: Test d'ajustement de Pearson 2 de Z – Prévisions jour 1 (Toutes les données) – Loi de Pareto généralisée

STATIONS	i	(0; 5)	[5; 8)	[8; 10)	[10; ∞)		
Chouart	n_i	124	26	8	29		
	p_i	0,6266	0.1421	0.0575	0.1737		
	m_i	117.18	26.58	10.76	32.48	D^2 = 1,49	p = 0,2222
High Falls	n_i	103	28	9	38		
	p_i	0.5663	0.1535	0.0662	0.2140		
	m_i	100.80	27.32	11.79	38.10	D^2 = 0,73	p = 0,3929
La Pêche	n_i	104	28	11	33		
	p_i	0.5903	0.1506	0.0634	0.1957		
	m_i	103,90	26,51	11.15	34.44	D^2 = 0,15	p = 0,6985
Maniwaki Airport	n_i	108	30	9	34		
	p_i	0.5793	0.1487	0.0632	0.2088		
	m_i	104.86	26,91	11.44	37.79	D^2 = 1,35	p = 0,2453
Mont St-Michel	n_i	97	24	10	32		
	p_i	0.5727	0.1500	0.0642	0.2131		
	m_i	93.35	24.45	10.47	34.73	D^2 = 0,39	p = 0,5323

Tableau 1.39: Test d'ajustement de Pearson 2 de Z – Prévisions jour 1 (Toutes les données non corrélées) – Loi de Pareto généralisée

STATIONS	i	(0; 5)	[5; 8)	[8; 10)	[10; ∞)		
Chouart	n_i	105	19	4	22		
	p_i	0.6615	0.1360	0.0530	0.1495		
	m_i	99.23	20.40	7.95	22.43	$D^2 = 2.40$	p = 0.1213
High Falls	n_i	82	24	5	31		
	p_i	0.5719	0.1536	0.0659	0.2086		
	m_i	81.20	21.82	9.36	29.62	$D^2 = 2,32$	p = 0.1277
La Pêche	n_i	85	20	9	27		
	p_i	0.5914	0.1506	0.0633	0.1947		
	m_i	83.39	21.24	8.92	27.45	$D^2 = 0.11$	p = 0.7401
Maniwaki Airport	n_i	72	23	4	19		
	p_i	0.6073	0.1438	0.0594	0.1895		
	m_i	71.66	16.97	7.01	22.36	$D^2 = 3,94$	p = 0,0472
Mont St-Michel	n_i	75	20	9	26		
	p_i	0,5687	0.1525	0.0656	0.2132		
	m_i	73.93	19.82	8.53	27.72	$D^2 = 0.15$	p = 0.6985

Le tableau 1.38 permet de conclure encore plus fortement que le modèle proposé, soit la loi de Pareto généralisée, est excellent, compte tenu qu'on obtient de très bonnes valeurs *p* pour toutes les stations, corroboré par les sous-ensembles de données qui sont très peu corrélées (tableau 1.39).

Vérifions maintenant si le modèle est aussi acceptable lorsqu'on considère les périodes d'été et d'hiver. Considérant la station Mont St-Michel, on obtient les tableaux suivants.

Tableau 1.40: Données statistiques de Y – Prévisions jour 1 (Été - Données supérieures à 0,2 mm)

STATION	Moyenne ($\bar{y}_1$)	Variance (s_{Y1}^2)	Écart type (s_{Y1})	Nombre de données (n)	$\hat{\theta}_z$	$\hat{c}_z$
Mont St-Michel	8.602	139,940	11.830	110	6.0360	0.0392

Tableau 1.41: Données statistiques de Y – Prévisions jour 1 (Été - Données non corrélées supérieures à 0,2 mm)

STATION	Moyenne ($\bar{y}_1$)	Variance (s_{Y1}^2)	Écart type (s_{Y1})	Nombre de données (n)	$\hat{\theta}_z$	$\hat{c}_z$
Mont St-Michel	*8,236*	*116,372*	*10,788*	*89*	*6,4936*	*0,0356*

Tableau 1.42: Test d'ajustement de Pearson de Z – Prévisions jour 1 (Été) – Loi de Pareto généralisée

STATION	i	(0; 5)	[5; 8)	[8; 10)	[10; ∞)		
Mont St-Michel	n_i	61	15	6	28		
	p_i	0.5144	0.1530	0.0694	0.2632		
	m_i	56.59	16.83	7.63	28.95	$D^2 = 0.92$	p = 0,3375

Tableau 1.43: Test d'ajustement de Pearson de Z – Prévisions jour 1 (Été – Données non corrélées) – Loi de Pareto généralisée

STATION	i	(0; 5)	[5; 8)	[8; 10)	[10; ∞)		
Mont St-Michel	n_i	48	13	6	22		
	p_i	0.5212	0.1547	0.0698	0.2543		
	m_i	46.39	13.77	6.21	22.64	$D^2 = 0.12$	p = 0,7290

Tableau 1.44: Données statistiques de Y – Prévisions jour 1 (Hiver - Données supérieures à 0,2 mm)

STATION	Moyenne ($\bar{y}_1$)	Variance (s_{Y1}^2)	Écart type (s_{Y1})	Nombre de données (*n*)	$\hat{\theta}_z$	$\hat{c}_z$
Mont St-Michel	4.606	23.724	4.871	53	13.0058	0.0227

Tableau 1.45: Données statistiques de Y – Prévisions jour 1 (Hiver - Données non corrélées supérieures à 0,2 mm)

STATION	Moyenne ($\bar{y}_1$)	Variance (s_{Y1}^2)	Écart type (s_{Y1})	Nombre de données (*n*)	$\hat{\theta}_z$	$\hat{c}_z$
Mont St-Michel	*4,939*	*28,107*	*5,302*	*41*	*11,9514*	*0,0236*

Tableau 1.46: Test d'ajustement de Pearson de Z – Prévisions jour 1 (Hiver) – Loi de Pareto généralisée

STATION	i	(0; 1)	[1; 2,5)	[2,5; 4,5)	[4,5; 7,5)	[7,5; ∞)		
Mont St-Michel	n_i	12	12	11	10	8		
	p_i	0.2187	0.2363	0.2018	0.1657	0.1774		
	m_i	11.59	12.52	10.70	8.78	9.40	$D^2 = 0.42$	p = 0.8106

Tableau 1.47: Test d'ajustement de Pearson de Z – Prévisions jour 1 (Hiver – Données non corrélées) – Loi de Pareto généralisée

STATION	i	(0; 1)	[1; 2,5)	[2,5; 4,5)	[4,5; 7,5)	[7,5; ∞)		
Mont St-Michel	n_i	10	8	9	7	7		
	p_i	0.2069	0.2274	0.1990	0.1688	0.1979		
	m_i	8.48	9.32	8.16	6.92	8.11	$D^2 = 0.70$	p = 0.7047

Ainsi, pour les périodes d'été et d'hiver, la loi de Pareto généralisée est un excellent modèle.

A l'aide des paramètres estimés pour la station du Mont St-Michel, soit $\hat{\theta}_Z$ = 5,8398 et $\hat{c}_Z$ = 0,0496, on teste maintenant l'hypothèse que la version de la loi de Pareto généralisée définie sur toute la droite réelle, soit de -∞ à ∞, est aussi un bon modèle pour les erreurs significatives elles-mêmes.

$$P(a \le Z' \le b) = \int_a^b (\hat{c}_Z/2)(\hat{\theta}_Z - 2)(1+\hat{c}_Z\,|z'|)^{1-\hat{\theta}_Z} dz' = \tfrac{1}{2}(1+\hat{c}_Z a)^{2-\hat{\theta}_Z} - \tfrac{1}{2}(1+\hat{c}_Z b)^{2-\hat{\theta}_Z}$$

$$= \tfrac{1}{2}(1+0{,}0496a)^{-3{,}8398} - \tfrac{1}{2}(1+0{,}0496b)^{-3{,}8398} \quad \text{pour a et b} \ge 0$$

<u>Calculs des probabilités</u>

Le modèle est symétrique par rapport à l'origine. Par conséquent,

- $P(-\infty < Z' \le -10) = P(10 \le Z' < \infty) = 0{,}1065$
- $P(-10 < Z' \le -5) = P(5 \le Z' < 10) = 0{,}1071$
- $P(-5 < Z' \le -3) = P(3 \le Z' < 5) = 0{,}0800$
- $P(-3 < Z' < 0) = P(0 \le Z' < 3) = 0{,}2065$

Tableau 1.48: Test d'ajustement de Pearson de Z' – Prévisions jour 1 (Toutes les données) – Loi de Pareto généralisée

STATION	i	(-∞; -10]	(-10;-5]	(-5;-3]	(-3; 0)	(0; 3)	[3; 5)	[5; 10)	[10; ∞)		
Mont St-Michel	n_i	12	17	11	29	44	13	17	20		
	p_i	0,1065	0,1071	0,0800	0,2065	0,2065	0,0800	0,1071	0,1065		
	m_i	17,36	17,45	13,03	33,66	33,66	13,03	17,45	17,36	D^2 = 6,22	p = 0,2854

<u>Nombre de degrés de liberté</u>

k : Nombre de classes = 8

r : Nombre de paramètres estimés = 2

n : Nombre de degrés de liberté = k – r – 1 = 5

Valeur $p = P(D^2_{ddl} \ge D^2_{observé})$

$= P(D^2{}_5 \geq 6{,}22)$

$= 0{,}2854$ ce qui est bon.

Pour les données non corrélées, les paramètres estimés sont les suivants: $\hat{\theta}_Z = 6{,}3574$ et $\hat{c}_Z = 0{,}0426$

Tableau 1.49: Test d'ajustement de Pearson de Z' – Prévisions jour 1 (Toutes les données non corrélées) – Loi de Pareto généralisée

STATION	i	(-∞; -10]	(-10;-5]	(-5;-3]	(-3; 0)	(0; 3)	[3; 5)	[5; 10)	[10; ∞)		
Mont St-Michel	n_i	9	14	8	21	36	10	15	17		
	p_i	0,1065	0,109	0,0805	0,2039	0,2039	0,0805	0,109	0,1065		
	m_i	13,85	14,18	10,47	26,51	26,51	10,47	14,18	13,85	$D^2 = 7{,}61$	$p = 0{,}1791$

Par conséquent, à la lueur des résultats obtenus, on peut dire que la loi de Pareto généralisée définie sur toute la droite réelle est aussi un bon modèle pour les erreurs de prévisions significatives lorsqu'on considère toutes les données. Les données non corrélées ne donnent par contre pas plus de valeur au modèle proposé.

Les tests d'ajustement suivants sont effectués pour le cas où on prévoyait entre 0 et 5 mm.

Tableau 1.50: Test d'ajustement de Pearson 1 de Z – Prévisions jour 1 (0<p<5) – Loi de Pareto généralisée

STATIONS	i	(0; 1)	[1; 2,5)	[2,5; 4)	[4; 10)	[10; ∞)		
Brotkord	n_i	24	26	16	16	7		
	p_i	0.2762	0.2522	0.1476	0.2225	0.1015		
	m_i	24.58	22.44	13.14	19.80	9.04	$D^2 = 2.39$	p = 0.3027
Chouart	n_i	43	28	22	16	11		
	p_i	0.2955	0.2560	0.1436	0.2082	0.0967		
	m_i	35.46	30.72	17.23	24.98	11.61	$D^2 = 6.43$	
High Falls	n_i	23	34	17	12	8		
	p_i	0.3084	0.2647	0.1458	0.2020	0.0792		
	m_i	28.99	24.88	13.70	18.99	7.44	$D^2 = 7.99$	
La Pêche	n_i	20	27	14	14	9		
	p_i	0.2696	0.2528	0.1509	0.2294	0.0973		
	m_i	22.65	21.24	12.68	19.27	8.17	$D^2 = 3.53$	p = 0.1712
Maniwaki Airport	n_i	28	34	19	14	8		
	p_i	0.2771	0.2477	0.1435	0.2184	0.1133		
	m_i	28.54	25.52	14.78	22.49	11.67	$D^2 = 8.40$	
Mont St-Michel	n_i	22	25	15	17	8		
	p_i	0.2408	0.2329	0.1449	0.2411	0.1403		
	m_i	20.95	20.26	12.61	20.98	12.20	$D^2 = 3.82$	p = 0.1481

Dans le cas du Mont St-Michel, si on enlève la valeur aberrante 63,83, la valeur D^2 passe à 2,55. Par conséquent la valeur *p* augmente à 0,2794, ce qui est beaucoup mieux.

On peut donc conclure que le modèle proposé de la loi de Pareto généralisée dans le cas où on prévoyait entre 0 et 5 mm de précipitations est assez bon.

En résumé, pour les prévisions pour la journée même, les modèles recommandés pour l'ensemble des stations considérées sont présentés dans le tableau suivant.

Tableau 1.51 : Modèles recommandés – Prévisions jour 1

Toutes les données	Prévisions p = 0	0<p<5	5≤p<10	10≤p
Loi de Pareto généralisée	Loi lognormale	Loi de Pareto généralisée	Loi normale	Loi lognormale

CHAPITRE 2 : MODÈLES STATISTIQUES – UN JOUR D'AVANCE

Pour les tests suivants, on ne considère que les stations ayant obtenu les meilleurs résultats pour les prévisions du jour même. Il s'agit maintenant de trouver des modèles pour les erreurs de prévisions lorsque les prévisions de précipitations sont faites pour le lendemain.

2.1 Valeurs absolues des erreurs de prévisions

Soit X, la valeur absolue de l'erreur de prévision.

Tableau 2.1: Données statistiques de X – Prévisions jour 2 (Toutes les données)

STATIONS	Moyenne ($\bar{x}_2$)	Variance (s_{X2}^2)	Écart type (s_{X2})	Nombre de données (*n*)
Brotkord	4.45	56.14	7.49	222
High Falls	4.33	46.82	6.84	243
	4,09	*40,67*	*6,38*	*195*
La Pêche	4.48	55.40	7.44	220
Mont St-Michel	4.56	61.08	7.82	221
	4,72	*67,22*	*8,20*	*177*

Tableau 2.2: Données statistiques de X – Prévisions jour 2 (Prévisions p = 0 mm)

STATIONS	Moyenne ($\bar{x}_2$)	Variance (s_{X2}^2)	Écart type (s_{X2})	Nombre de données (*n*)
Brotkord	2,17	45.24	6.73	51
High Falls	1,31	13.50	3.67	57
La Pêche	1,46	13.01	3.61	51
Mont St-Michel	1,94	25.53	5.05	52

Tableau 2.3: Données statistiques de X – Prévisions jour 2 (0<p<5)

STATIONS	Moyenne ($\bar{x}_2$)	Variance (s_{X2}^2)	Écart type (s_{X2})	Nombre de données (n)
Brotkord	3,34	38.9	6.24	126
High Falls	3,86	47.37	6.88	144
La Pêche	4.41	64.82	8.05	133
Mont St-Michel	3.56	60.11	7.75	127

Tableau 2.4: Données statistiques de X – Prévisions jour 2 ($5 \leq p<10$)

STATIONS	Moyenne ($\bar{x}_2$)	Variance (s_{X2}^2)	Écart type (s_{X2})	Nombre de données (n)
Brotkord	7.91	98.47	9.92	27
High Falls	6.82	9.66	3.11	24
La Pêche	5.36	5.11	2.26	21
Mont St-Michel	7.84	38.43	6.2	25

Pour les stations High Falls et La Pêche, il n'y a pas assez de données pour choisir des sous-ensembles de données moins corrélées.

Tableau 2.5: Données statistiques de X– Prévisions jour 2 ($10 \leq p$)

STATIONS	Moyenne ($\bar{x}_2$)	Variance (s_{X2}^2)	Écart type (s_{X2})	Nombre de données (n)
Brotkord	13.42	31.16	5.58	18
High Falls	14.37	62.42	7.90	18
La Pêche	14.09	68.19	8.26	15
Mont St-Michel	15.21	59.89	7.74	17

Pour la station Brotkord, il n'y a pas assez de données pour choisir un sous-ensemble de données moins corrélées.

Tableau 2.6: Coefficients de corrélation de X - Prévisions jour 2

STATIONS	Toutes		Prévisions (p)=0		$0<p<5$		$5\leq p<10$		$10\leq p$	
	r_1	r_2	r_1	r_2	r_1	r_2	r_1	r_2	r_1	r_2
Brotkord	0.17	-0,03	0,01	-0,02	0,00	-0,05	0,04	-0,02	-0,24	0,22
High Falls	0,26	0,00	0.10	0,05	-0,04	-0,06	0,42	0,36	-0,21	-0,03
	0,19	*0,01*								
La Pêche	0.19	-0,03	0,19	0,03	-0,05	-0,06	-0,45	0,07	-0,21	-0,13
Mont St-Michel	0,26	0,03	0,02	0,14	-0,04	-0,07	-0,03	0,30	-0,32	-0,09
	0,19	*0,03*								

Pour la majorité des données considérées, les coefficients de corrélation sont proches de zéro; l'hypothèse d'indépendance des données correspondantes est plausible.

2.2 Valeurs absolues des erreurs de prévisions avec un seuil imposé

Soit

Y : Valeur absolue de l'erreur de prévision, étant donné qu'elle est supérieure à 0,2 mm

Z = Y – 0,2

Tableau 2.7: Données statistiques de Y – Prévisions jour 2 (Toutes les données non corrélées, supérieures à 0,2 mm)

STATIONS	Moyenne ($\bar{y}_2$)	Variance (s_{Y2}^2)	Écart type (s_{Y2})	Nombre de données (n)	$\hat{\theta}_z$	$\hat{c}_z$
Brotkord	5.824	65.854	8,115	169	5.8483	0,0624
High Falls	*5,480*	*47,233*	*6,873*	*145*	*6,8808*	*0,0488*
La Pêche	6,013	65.750	8.109	163	6.1146	0.0552
Mont St-Michel	*6,028*	*78,502*	*8,860*	*138*	*5,5253*	*0,0679*

Tableau 2.8: Données statistiques de Y – Prévisions jour 2 (0 <p<5 non corrélées, supérieures à 0,2 mm)

STATIONS	Moyenne ($\bar{y}_2$)	Variance (s_{Y2}^2)	Écart type (s_{Y2})	Nombre de données (n)	$\hat{\theta}_z$	$\hat{c}_z$
Brotkord	3.853	43.083	6.564	109	4.8974	0.1443
High Falls	4.604	53.561	7.319	120	5.1354	0.1063
La Pêche	5.302	73.884	8.596	110	5.0879	0.0939
Mont St-Michel	4.354	70.868	8.418	103	4.6437	0.1465

2.3 Modélisation

2.3.1 Lois normale et lognormale

On effectue un test de normalité d'Anderson-Darling sur les valeurs de Y dans le cas où on ne prévoyait aucune précipitation. On obtient les valeurs p qui suivent pour chacune des stations.

- Brotkord – valeur $p \approx 0$ (15 données)
- High Falls – valeur $p \approx 0$ (15 données)
- La Pêche – valeur $p \approx 0$ (17 données)

- Mont St-Michel – valeur $p \approx 0$ (21 données)

Par contre, si on prend le logarithme de la valeur de Y, on obtient de bonnes valeurs p pour les stations High Falls et Mont St-Michel et des valeurs acceptables pour Brotkord et La Pêche. Donc, dans le cas où on ne prévoyait pas de précipitations, on peut conclure que la valeur absolue de l'erreur de prévision, étant donné qu'elle est supérieure à 0,2 mm, suit une loi lognormale.

- Brotkord – valeur $p = 0{,}160$
- High Falls – valeur $p = 0{,}351$
- La Pêche – valeur $p = 0{,}134$
- Mont St-Michel – valeur $p = 0{,}280$

Ensuite, on effectue un test de normalité d'Anderson-Darling sur les valeurs de X dans le cas où on prévoyait de 5 à 10 mm de précipitations. Pour chacune des stations, on obtient des valeurs p à peu près égales à zéro. Mais pour Brotkord, si on enlève une donnée aberrante, soit 56,67 mm, on obtient une très bonne valeur p égale à 0,566. Pour High Falls, si on enlève 16,9 mm, une valeur beaucoup plus grande que les autres, on obtient une valeur p de 0,202. Toutefois, en ce qui concerne cette station, on doit être prudent, car le coefficient de corrélation des données à un jour d'écart est de 0,42. On peut tout de même conclure qu'une loi normale est acceptable pour les données lorsqu'on prévoyait de 5 à 10 mm de précipitations. Noter que si on effectue un test de normalité d'Anderson-Darling avec le logarithme des valeurs X, pour toutes les stations, la valeur p est à peu près égale à zéro.

Finalement, on effectue un test de normalité d'Anderson-Darling sur les valeurs de X dans le cas où on prévoyait au moins 10 mm de précipitations.

- Brotkord – valeur $p = 0{,}233$

- High Falls – valeur $p \approx 0$
- La Pêche – valeur $p \approx 0$
- Mont St-Michel – valeur $p = 0{,}750$

Par contre, si on prend le logarithme de la valeur de X, on obtient une très bonne valeur p pour la station High Falls.

- Brotkord – valeur $p \approx 0$,
- High Falls – valeur $p = 0{,}520$
- La Pêche – valeur $p = 0{,}122$
- Mont St-Michel – valeur $p \approx 0$

Bien que la loi normale pour le cas où on prévoyait au moins 10 mm de précipitations est parfois un très bon modèle, la loi lognormale pourrait aussi être acceptable.

2.3.2 Loi de Pareto généralisée

On effectue tout d'abord un test d'ajustement de Pearson sur les valeurs de Z dans le cas où on considère toutes les données.

Tableau 2.9: Test d'ajustement de Pearson de Z – Prévisions jour 2 (Toutes les données) – Loi de Pareto généralisée

STATIONS	i	(0; 3)	[3; 7,5)	[7,5; 16,5)	[16,5; ∞)		
Brotkord	n_i	84	46	26	13		
	p_i	0.4835	0.2884	0.1626	0.0656		
	m_i	81.71	48.74	27.47	11.08	$D^2 = 0{,}63$	$p = 0{,}4274$
High Falls	n_i	74	41	19	11		
	p_i	0.4867	0.2951	0.1622	0.0560		
	m_i	70.57	42.79	23.52	8.11	$D^2 = 2.14$	$p = 0{,}1435$
La Pêche	n_i	83	42	25	13		
	p_i	0.4679	0.2919	0.1707	0.0696		
	m_i	76.26	47.58	27.82	11.34	$D^2 = 1{,}78$	$p = 0{,}1821$
Mont St-Michel	n_i	70	40	16	12		
	p_i	0.4800	0.2858	0.1635	0.0706		
	m_i	66.25	39.44	22.57	9.74	$D^2 = 2.66$	$p = 0{,}1029$

Pour la station Brotkord, la valeur p est très bonne et elle est acceptable pour les autres. Par conséquent, ce modèle est acceptable lorsqu'on considère toutes les données.

On effectue maintenant un test d'ajustement de Pearson sur les valeurs de Z lorsqu'on prévoyait entre 0 et 5 mm de précipitations.

Tableau 2.10: Test d'ajustement de Pearson de Z – Prévisions jour 2 (0<p<5) – Loi de Pareto généralisée

STATIONS	i	(0; 1)	[1; 3)	[3; 8,5)	[8,5; ∞)		
Brotkord	n_i	42	31	26	10		
	p_i	0.3233	0.3240	0.2544	0.0984		
	m_i	35.24	35.32	27.73	10.72	$D^2 = 1,98$	$p = 0,1594$
High Falls	n_i	36	38	31	15		
	p_i	0.2716	0.3087	0.2869	0.1328		
	m_i	32.59	37.04	34.43	15.94	$D^2 = 0,78$	$p = 0,3771$
La Pêche	n_i	36	31	27	16		
	p_i	0.2420	0.2932	0.3014	0.1634		
	m_i	26.62	32.25	33.15	17.98	$D^2 = 4,71$	
Mont St-Michel	n_i	35	30	29	9		
	p_i	0.3033	0.3149	0.2639	0.1179		
	m_i	31.23	32.44	27.18	12.14	$D^2 = 1,57$	$p = 0,2102$

Pour la station High Falls, ce modèle est très bon et il est acceptable pour les stations Brotkord et Mont St-Michel. On peut donc conclure que ce modèle est acceptable lorsqu'on prévoyait entre 0 et 5 mm.

En résumé, pour les prévisions pour le lendemain, les modèles recommandés pour l'ensemble des stations considérées sont présentés dans le tableau suivant.

Tableau 2.11 : Modèles recommandés – Prévisions jour 2

Toutes les données	Prévisions p = 0	0<p<5	5≤p<10	10≤p
Loi de Pareto généralisée	Loi lognormale	Loi de Pareto généralisée	Loi normale	Loi lognormale (La loi normale est aussi un bon modèle.)

CHAPITRE 3 : MODÈLES STATISTIQUES – DEUX JOURS D'AVANCE

Le dernier cas étudié est celui où des prédictions de précipitations sont faites pour le deuxième jour. Encore ici, on ne considère que les stations les plus fiables pour trouver des modèles statistiques appropriés qui s'ajustent bien aux données.

3.1 Valeurs absolues des erreurs de prévisions

Soit X, la valeur absolue de l'erreur de prévision.

Tableau 3.1: Données statistiques de X – Prévisions jour 3 (Toutes les données)

STATIONS	Moyenne ($\bar{x}_3$)	Variance (s_{X3}^2)	Écart type (s_{X3})	Nombre de données (*n*)
Brotkord	4.62	60.45	7.77	220
High Falls	4.61	50.11	7.08	243
La Pêche	4.46	54.31	7.37	221
Mont St-Michel	4.93	64.66	8.04	222

Tableau 3.2: Données statistiques de X – Prévisions jour 3 (Prévisions p = 0 mm)

STATIONS	Moyenne ($\bar{x}_3$)	Variance (s_{X3}^2)	Écart type (s_{X3})	Nombre de données (*n*)
Brotkord	2.27	24.66	4.97	51
High Falls	3.95	76.53	8.75	43
La Pêche	3.53	63.20	7.95	45
Mont St-Michel	2.34	25.27	5.03	38

Tableau 3.3: Données statistiques de X – Prévisions jour 3 (0<p<5)

STATIONS	Moyenne ($\bar{x}_3$)	Variance (s_{X3}^2)	Écart type (s_{X3})	Nombre de données (*n*)
Brotkord	3,99	71.23	8.44	129
High Falls	3.21	32.80	5.73	154
La Pêche	3.71	46.75	6.84	141
Mont St-Michel	4.14	69.93	8.36	143

Tableau 3.4: Données statistiques de X – Prévisions jour 3 (5 ≤p<10)

STATIONS	Moyenne ($\bar{x}_3$)	Variance (s_{X3}^2)	Écart type (s_{X3})	Nombre de données (n)
Brotkord	5.96	7.57	2.75	21
High Falls	6.41	10.27	3.20	29
La Pêche	5.26	9.26	3.04	22
Mont St-Michel	5.13	5.70	2.39	20

Pour les stations Brotkord, La Pêche et Mont St-Michel, il n'y a pas assez de données pour choisir des sous-ensembles de données moins corrélées.

Tableau 3.5: Données statistiques de X – Prévisions jour 3 (10 ≤ p)

STATIONS	Moyenne ($\bar{x}_3$)	Variance (s_{X3}^2)	Écart type (s_{X3})	Nombre de données (n)
Brotkord	13.68	44.67	6.68	19
High Falls	15.98	62.01	7.87	17
La Pêche	14.50	82.11	9.06	13
Mont St-Michel	14.86	45.03	6.71	21

Pour les stations High Falls, La Pêche et Mont St-Michel, il n'y a pas assez de données pour choisir des sous-ensembles de données moins corrélées.

Tableau 3.6: Coefficients de corrélation de X - Prévisions jour 3

STATIONS	Toutes		Prévisions p = 0		0<p<5		5≤p<10		10≤p	
	r_1	r_2	r_1	r_2	r_1	r_2	r_1	r_2	r_1	r_2
Brotkord	0.18	0.04	0.05	-0.17	0.03	-0.03	-0.02	-0.36	0.11	0.09
High Falls	0.20	-0.07	-0.12	-0.05	-0.07	-0.08	0.00	0.01	-0.42	0.19
La Pêche	0.16	-0.09	-0.12	-0.08	0.03	-0.06	0.06	-0.36	-0.59	0.10
Mont St-Michel	0.21	-0.12	0.01	-0.20	-0.10	-0.13	-0.31	0.02	-0.34	-0.17

Pour la majorité des données considérées, les coefficients de corrélation sont proches de zéro; l'hypothèse d'indépendance des données correspondantes est plausible.

3.2 Valeurs absolues des erreurs de prévisions avec un seuil imposé

Soit

Y : Valeur absolue de l'erreur de prévision, étant donné qu'elle est supérieure à 0,2 mm

Z = Y – 0,2

Tableau 3.7: Données statistiques de Y – Prévisions jour 3 (Toutes les données non corrélées, supérieures à 0,2 mm)

STATIONS	Moyenne ($\bar{y}_3$)	Variance (s_{Y3}^2)	Écart type (s_{Y3})	Nombre de données (n)	$\hat{\theta}_z$	$\hat{c}_z$
Brotkord	5.884	69.996	8.366	172	5.7145	0.0648
High Falls	5.964	57.223	7.565	187	6.7687	0.0460
La Pêche	5.745	62.981	7.936	171	5.9077	0.0620
Mont St-Michel	6.232	74.107	8.609	175	5.9291	0.0566

Tableau 3.8: Données statistiques de Y – Prévisions jour 3 (0<p<5 non corrélées, supérieures à 0,2 mm)

STATIONS	Moyenne ($\bar{y}_3$)	Variance (s_{Y3}^2)	Écart type (s_{Y3})	Nombre de données (n)	$\hat{\theta}_z$	$\hat{c}_z$
Brotkord	4.532	78.991	8.888	113	4.6232	0.1422
High Falls	3.976	38.156	6.177	123	5.1933	0.1207
La Pêche	4.446	53.243	7.297	117	5.0239	0.1164
Mont St-Michel	4.983	80.734	8.985	118	4.7908	0.1167

3.3 Modélisation

3.3.1 Lois normale et lognormale

On effectue un test de normalité d'Anderson-Darling sur les valeurs de Y dans le cas où on ne prévoyait pas de précipitations. On obtient des valeurs p à peu près égales à zéro pour chacune des stations.

Mais si on prend le logarithme de la valeur de Y, on obtient une bonne valeur p pour la station Brotkord et d'excellentes valeurs p pour les stations High Falls et Mont St-Michel. On peut donc conclure que la loi lognormale est un très bon modèle pour les erreurs significatives lorsqu'on ne prévoyait pas de précipitations.

- Brotkord – valeur p = 0,262 (19 données)
- High Falls – valeur p = 0,522 (19 données)

- La Pêche – valeur p = 0,093 (19 données)
- Mont St-Michel – valeur p = 0,437 (19 données)

On effectue maintenant un test de normalité d'Anderson-Darling sur les valeurs de X dans le cas où on prévoyait de 5 à 10 mm de précipitations. Pour Brotkord et La Pêche, la valeur p est pratiquement nulle. Pour High Falls, la valeur p de 0,110 est acceptable. Par contre, on obtient une excellente valeur p pour le Mont St-Michel, soit 0,734. Toutefois, en ce qui concerne le Mont St-Michel, on doit être prudent, car la valeur absolue du coefficient de corrélation des données à un jour d'écart est de 0,31. Par conséquent, on peut conclure qu'une loi normale est acceptable lorsqu'on prévoyait entre 5 et 10 mm de précipitations. Noter que si on effectue un test de normalité d'Anderson-Darling avec le logarithme des valeurs X, pour toutes les stations, la valeur p est à peu près égale à zéro.

Finalement, on effectue un test de normalité d'Anderson-Darling sur les valeurs de X dans le cas où on prévoyait au moins 10 mm de précipitations.

- Brotkord – valeur p = 0,403
- High Falls – valeur p = 0,574
- La Pêche – valeur p = 0,560
- Mont St-Michel – valeur p = 0,436

Puisqu'on obtient d'excellentes valeurs p pour toutes les stations, une loi gaussienne est un très bon modèle. Mais on peut faire encore mieux en considérant une loi lognormale. Ainsi, les données pourraient aussi provenir d'une loi lognormale, un excellent modèle pour les erreurs de prévisions lorsqu'on prévoyait au moins 10 mm de précipitations. Exception faite pour Brotkord, noter toutefois que les données pour les stations sont très corrélées.

- Brotkord – valeur p = 0,422
- High Falls – valeur p = 0,819
- La Pêche – valeur p = 0,749
- Mont St-Michel – valeur p = 0,612

3.3.2 Loi de Pareto généralisée

On effectue tout d'abord un test d'ajustement de Pearson sur les valeurs de Z pour toutes les données.

Tableau 3.9: Test d'ajustement de Pearson 1 de Z – Prévisions jour 3 (Toutes les données) – Loi de Pareto généralisée

STATIONS	i	(0; 3)	[3; 7,5)	[7,5; 16,5)	[16,5; ∞)	
Brotkord	n_i	84	47	27	14	
	p_i	0.4831	0.2873	0.1625	0.0671	
	m_i	83.10	49.41	27.94	11.54	$D^2 = 0,68$ | p = 0,4096
High Falls	n_i	94	45	32	16	
	p_i	0.4604	0.2965	0.1755	0.0676	
	m_i	86.09	55.45	32.82	12.63	$D^2 = 3.61$
La Pêche	n_i	93	37	27	14	
	p_i	0.4867	0.2886	0.1611	0.0637	
	m_i	83.22	49.34	27.55	10.89	$D^2 = 5.14$
Mont St-Michel	n_i	85	48	24	18	
	p_i	0.4600	0.2910	0.1741	0.0749	
	m_i	80.50	50.92	30.47	13.11	$D^2 = 3.62$

Pour la station Brotkord seulement, la valeur p est très bonne. Essayons de voir si on peut faire un meilleur test d'ajustement pour les autres des stations.

Tableau 3.10: Test d'ajustement de Pearson 2 de Z – Prévisions jour 3 (Toutes les données) – Loi de Pareto généralisée

STATIONS	i	(0; 5,5)	[5,5; 8)	[8; 10)	[10; ∞)	
Brotkord	n_i	123	10	10	29	
	p_i	0.6778	0.1103	0.0556	0.1563	
	m_i	116.58	18.98	9.56	26.89	$D^2 = 4.79$
High Falls	n_i	127	15	11	34	
	p_i	0.6591	0.1167	0.0598	0.1643	
	m_i	123.26	21.82	11.19	30.73	$D^2 = 2.59$ | $p = 0.1075$
La Pêche	n_i	122	12	10	27	
	p_i	0.6824	0.1105	0.0554	0.1517	
	m_i	116.69	18.89	9.47	25.94	$D^2 = 2.83$ | $p = 0.0925$
Mont St-Michel	n_i	122	11	10	32	
	p_i	0.6552	0.1143	0.0589	0.1717	
	m_i	114.66	20.00	10.30	30.04	$D^2 = 4.65$

Tableau 3.11: Test d'ajustement de Pearson 4 de Z – Prévisions jour 3 (Toutes les données) – Loi de Pareto généralisée

STATIONS	i	(0; 1)	[1; 2,5)	[2,5; 4)	[4; 8)	[8; 10)	[10; ∞)	
Brotkord	n_i	51	26	25	31	10	29	
	p_i	0.2081	0.2195	0.1477	0.2128	0.0556	0.1563	
	m_i	35.79	37.75	25.41	36.61	9.56	26.89	$D^2 = 11.18$
High Falls	n_i	49	38	25	30	11	34	
	p_i	0.1932	0.2120	0.1482	0.2225	0.0598	0.1643	
	m_i	36.12	39.65	27.71	41.60	11.19	30.73	$D^2 = 8.51$
La Pêche	n_i	50	36	15	33	10	27	
	p_i	0.2095	0.2211	0.1487	0.2135	0.0554	0.1517	
	m_i	35.83	37.81	25.43	36.51	9.47	25.94	$D^2 = 10.38$
Mont St-Michel	n_i	42	38	26	27	10	32	
	p_i	0.1945	0.2109	0.1460	0.2180	0.0589	0.1717	
	m_i	34.04	36.92	25.56	38.15	10.30	30.04	$D^2 = 5.30$ | $p = 0.1511$

On obtient une très bonne valeur p pour la station Brotkord et des valeurs acceptables pour les stations High Falls et Mont St-Michel. Par contre, on a dû effectuer des tests d'ajustement différents, ce qui implique que la conclusion à laquelle on pourrait arriver est moins forte. Mais comme le modèle de la loi de Pareto généralisée semble approprié pour au moins une station, soit Brotkord, il y a de bonnes chances que le modèle puisse aussi s'appliquer à d'autres stations.

On effectue maintenant un test d'ajustement de Pearson sur les valeurs de Z lorsqu'on prévoyait entre 0 et 5 mm de précipitations.

Tableau 3.12: Test d'ajustement de Pearson de Z – Prévisions jour 3 (0<p<5) – Loi de Pareto généralisée

STATIONS	i	(0; 1)	[1; 3)	[3; 8,5)	[8,5; ∞)		
Brotkord	n_i	45	29	29	10		
	p_i	0.2945	0.3118	0.2686	0.1251		
	m_i	33.27	35.23	30.36	14.13	$D^2 = 6.50$	
High Falls	n_i	42	39	28	14		
	p_i	0.3051	0.3222	0.2678	0.1049		
	m_i	37.53	39.64	32.94	12.90	$D^2 = 1.38$	$p = 0.2401$
La Pêche	n_i	40	35	28	14		
	p_i	0.2831	0.3125	0.2794	0.1250		
	m_i	33.13	36.56	32.69	14.62	$D^2 = 2.19$	$p = 0.1389$
Mont St-Michel	n_i	36	36	30	16		
	p_i	0.2652	0.3022	0.2865	0.1461		
	m_i	31.29	35.66	33.81	17.23	$D^2 = 1.23$	$p = 0.2674$

Pour la station La Pêche, ce modèle est acceptable et il est bon pour les stations High Falls et Mont St-Michel. On peut donc conclure que ce modèle est bon lorsqu'on prévoyait entre 0 et 5 mm de précipitations.

En résumé, pour les prévisions pour le surlendemain, les modèles recommandés pour l'ensemble des stations considérées sont présentés dans le tableau suivant.

Tableau 3.13 : Modèles recommandés – Prévisions jour 3

Toutes les données	Prévisions p = 0	0<p<5	5≤p<10	10≤p
Loi de Pareto généralisée	Loi lognormale	Loi de Pareto généralisée	Loi normale	Loi lognormale (La loi normale est aussi un bon modèle.)

CONCLUSION

Dans ce mémoire, nous avons cherché des modèles statistiques pour les erreurs de prévisions de précipitations d'une année, lesquelles furent produites par Hydro-Québec pour le bassin versant de la rivière Gatineau. Nous nous sommes concentrés sur les stations à l'intérieur d'une grille d'étude pour lesquelles nous possédions suffisamment de données représentatives. Nous avons cherché des modèles pour les trois cas suivants: les erreurs de prévisions du jour même, celles du lendemain, et finalement celles du deuxième jour. Les meilleures lois retenues sont les lois normale, lognormale et de Pareto généralisée.

Indépendamment de la période de prévisions considérée, on a constaté que les modèles trouvés sont à peu près les mêmes pour les trois cas étudiés. Donc, lorsqu'on utilise toutes les données, bien qu'on puisse suggérer une loi exponentielle comme modèle pour Z (valeur absolue de l'erreur de prévision, étant donné qu'elle est supérieure à 0,2 mm, moins 0,2) dans le cas des erreurs de prévisions de précipitations du jour même, on fait beaucoup mieux en considérant la loi de Pareto généralisée, et ce, pour les trois cas étudiés. Par contre, lorsqu'on ne prévoyait pas de précipitations, le modèle de la loi lognormale est beaucoup plus approprié. Dans le cas où on prévoyait entre 0 et 5 mm de précipitations, la loi de Pareto généralisée est le modèle le plus approprié pour les erreurs de prévisions du jour même, d'un jour d'avance et pour deux jours d'avance. Ensuite, lorsqu'on prévoyait entre 5 et 10 mm, si on enlève des données aberrantes pour les prévisions du jour même et le lendemain, une loi normale pourrait être acceptable pour les trois périodes. Finalement, lorsqu'on prévoyait plus de 10 mm de précipitations, la loi lognormale est un bon modèle pour les erreurs de prévisions du jour même, du lendemain et du surlendemain. Par contre, pour les erreurs de prévisions du lendemain et du surlendemain, la loi normale est aussi un bon modèle.

Le test d'ajustement de Pearson, effectué pour les périodes d'été et d'hiver (toutes les données pour les erreurs de prévisions du jour même) pour la station du Mont St-Michel, a permis de conclure que la loi de Pareto généralisée est toujours un excellent modèle pour la variable aléatoire Z. De même, toujours à partir de cette station témoin, la loi de Pareto généralisée définie sur toute la droite réelle est aussi un bon modèle pour les erreurs de prévisions significatives lorsqu'on considère toutes les données du jour même.

On suppose que les résultats obtenus s'appliqueraient aussi bien pour d'autres ensembles de données, soit des prévisions et des observations annuelles subséquentes. Il serait toutefois judicieux de vérifier si les modèles proposés sont toujours valables pour ces autres ensembles de données.

BIBLIOGRAPHIE

ACCADIA, C., CASAIOLI, M., MARIANI, S., LAVAGNINI, A., SPERANZA, A., DE VENERE, A., INGHILESI, et al. (2003). «Application of a statistical methodolgy for limited area model intercomparison using a bootstrap technique». *Il Nuovo Cimento.* 26:1. 61-77.

DÉQUÉ, M. (2003). «La prévision numérique à l'échelle saisonnière: que sait-on faire et que peut-on espérer ?». *La Météorologie.* 41. 20-29.

HINES, W. W., MONTGOMERY, D. C. (1990). *Probability and Statistics in Engineering and Management Science.* 3th ed. New York: Wiley. 732p.

JOHNSON, L. E., OLSEN, B. G. (1998). «Assessment of Quantitative Precipitation Forecasts». *Weather and Forecasting.* 13:1. 75-83.

LAUZON, N. (1995). *Méthodes de validation et de prévision à court terme des apports naturels.* 301p. Mémoire de maîtrise en génie civil, École Polytechnique de Montréal.

LAUZON, N., BIRIKUNDAVYI, S., GIGNAC, C., ROUSSELLE, J. (1997). «Comparaison de deux procédures d'amélioration des prévisions à court terme des apports naturels d'un modèle déterministe». *Revue canadienne de génie civil.* 24:5. 723-735.

LEFEBVRE, M. (2004). «Modélisation des erreurs pour le système de prévision PRÉVIS». *Revue canadienne de génie civil.* 31:5. 892-897.

MC BRIDE, J. L., EBERT, E. E. (2000). «Verification of Quantitative Precipitation Forecasts from Operational Numerical Weather Prediction Models over Australia». *Weather and Forecasting*. 15:1. 103-121.

RIBEIRO, J., LAUZON, N., ROUSSELLE, J., TRUNG, H.T., SALAS, J.D. (1998). «Comparaison de deux modèles de prévision journalière en temps réel des apports naturels». *Revue canadienne de génie civil*. 25:2. 291-304.

ROUSSEAU, V. (2004). *Étude d'un simulateur de climat – Modélisation statistique des erreurs de prévisions de MétéoFrance*. [En ligne]. 62p. Rapport de stage en mathématiques et statistique, Université René Descartes. http://www.inra.fr/Internet/Departements/MIA/garcia/Anticipation/AvancementProjet/Stages/2004/Vanessa/rapportVanessa2.pdf (Page consultée le 15 janvier 2006)