

Société de Calcul Mathématique SA
Outils d'aide à la décision
depuis 1995



Amélioration des prévisions
de qualité de l'air au niveau régional

Note d'avancement n°1

adressée à

Lig'Air

par la

Société de Calcul Mathématique SA

en application du bon de commande no 2015/02/031 du 17/02/2015

Avril 2015

I. Introduction

L'objectif du travail confié à la SCM par Lig'Air est la définition et mise en œuvre d'une méthode statistique d'amélioration des prévisions de la qualité de l'air en Région Centre. Pour cela, Lig'Air a fourni à la SCM un volume important de données.

Cette note a pour but d'illustrer la méthode probabiliste SCM d'amélioration du modèle de prévision existant par le moyen d'un exemple simple. Par conséquent, nous ne présentons pas ici le travail d'analyse et de traitement de données réalisé sur l'ensemble des données fournies.

II. Présentation de la méthode SCM

La méthode repose sur la construction de tables de calibration, qui serviront à corriger la valeur donnée par le modèle de prévision.

La construction de la table de calibration repose sur les étapes suivantes :

1. construction de la loi de probabilité de l'erreur
2. hiérarchisation des paramètres influant sur l'erreur
3. sélection des paramètres prépondérants
4. construction de la table de calibration
5. validation de la table sur un autre jeu de données

Nous illustrerons chaque étape à partir de l'exemple simple suivant :

- polluant : O₃
- station de mesure : Blois Nord, code ISO :34061
- station météo : Blois
- année : 2014
- modèle : Esmeralda J+1

III. Loi de probabilité de l'erreur

A. Erreur réelle

Pour chaque mesure de concentration en O3 pour la station Blois Nord en l'année 2014, nous calculons l'erreur associée au modèle Esmeralda j+1 : il s'agit de la différence de la concentration modélisée par la concentration réelle.

Ensuite, nous construisons la loi de probabilité de l'erreur en remplissant le tableau des occurrences ci-dessous.

intervalle	Nombre d'occurrences	probabilité
[-70;-65[1	0%
[-65;-60[2	0%
[-60;-55[1	0%
[-55;-50[5	0%
[-50;-45[13	0%
[-45;-40[33	0%
[-40;-35[51	1%
[-35;-30[124	2%
[-30;-25[188	2%
[-25;-20[252	3%
[-20;-15[371	5%
[-15;-10[513	6%
[-10;-5[652	8%
[-5;0[748	9%
[0;5[813	10%
[5;10[775	10%
[10;15[737	9%
[15;20[650	8%
[20;25[548	7%
[25;30[491	6%
[30;35[386	5%
[35;40[285	4%
[40;45[210	3%
[45;50[111	1%
[50;55[74	1%
[55;60[61	1%
[60;65[23	0%
[65;70[11	0%
[70;75[7	0%
[75;80[4	0%
[80;85[2	0%

Tableau 1 Tableau d'occurrence de l'erreur

La loi de probabilité de cet erreur est donnée par l'histogramme ci-dessous.

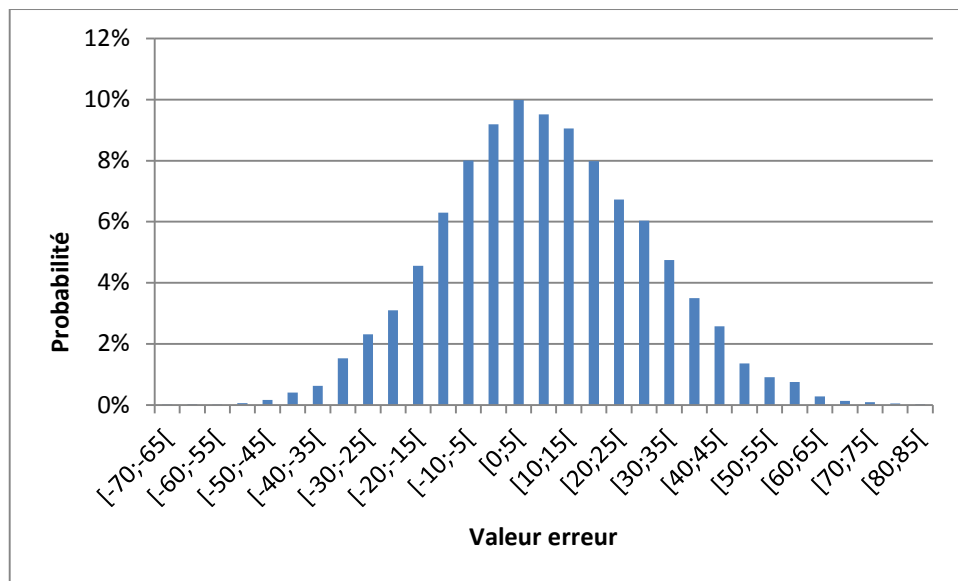


Figure 1 Histogramme de l'erreur

Le tableau ci-dessous résume les caractéristiques des lois de probabilité de l'erreur :

Minimum	-68,6
Maximum	82,3
Moyenne	7,7
Borne 5 %	-25,0
Borne 95%	41,7

B. Valeur absolue de l'erreur

A partir du tableau des occurrences, nous construisons également la loi de probabilité de la valeur absolue de l'erreur.

intervalle	Nombre d'occurrences	probabilité
[0;5[813	19%
[5;10[775	18%
[10;15[737	15%
[15;20[650	13%
[20;25[548	10%
[25;30[491	8%
[30;35[386	6%
[35;40[285	4%
[40;45[210	3%
[45;50[111	2%
[50;55[74	1%
[55;60[61	1%
[60;65[23	0%
[65;70[11	0%
[70;75[7	0%
[75;80[4	0%
[80;85[2	0%

Tableau 2 Tableau d'occurrence de la valeur absolue de l'erreur

L'histogramme est donné ci-dessous.

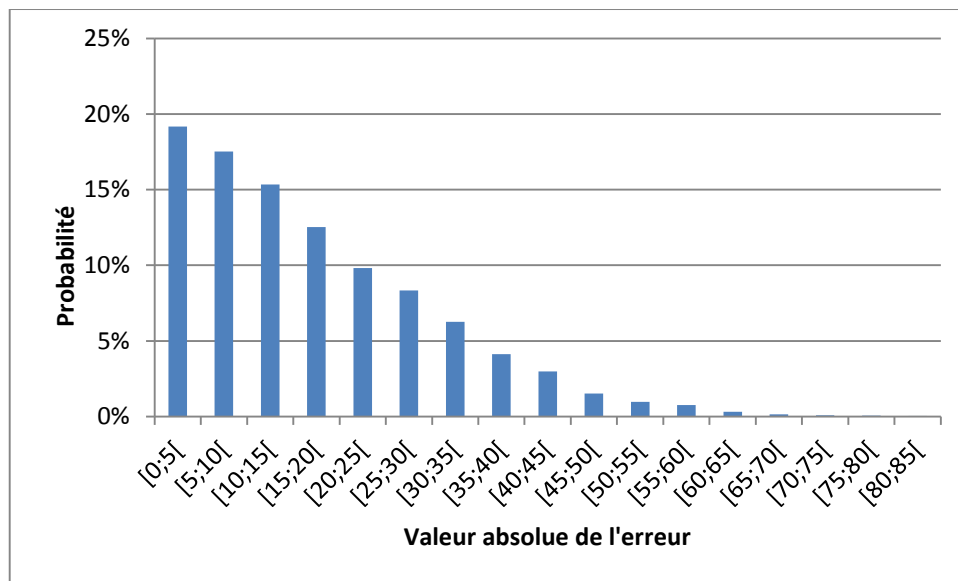


Figure 2 Histogramme de la valeur absolue de l'erreur

Le tableau ci-dessous résume les caractéristiques des lois de probabilité de la valeur absolue de l'erreur :

Minimum	0
Maximum	82,3
Moyenne	17,1
Borne 5 %	1,2
Borne 95%	42,5

IV. Hiérarchisation des paramètres

A. Principe de la méthode

La table de calibration est définie en prenant en compte les paramètres jouant un rôle majeur.

Afin de déterminer quels sont les paramètres qui influent le plus sur l'erreur, nous avons mis en œuvre une méthode probabiliste de hiérarchisation des paramètres : nous obtenons une classification des paramètres en fonction de leur influence. Les principes de la méthode utilisée sont présentés dans l'ouvrage suivant:

- Bernard Beauzamy : *Nouvelles Méthodes Probabilistes pour l'évaluation des risques*. Ouvrage édité et commercialisé par la Société de Calcul Mathématique SA. ISBN 978-2-9521458-4-8. ISSN 1767-1175, avril 2010.

Il est important de noter que cette méthode ne met pas en évidence des relations de cause à effet, mais seulement des relations pouvant exister entre les paramètres et les variables de sortie

B. Paramètres pris en compte

Les paramètres pris en compte sont les suivants :

- Paramètres météorologiques :
 - PRECIPITATION
 - TEMPERATURE
 - PRESSION STATION
 - VITESSE VENT
 - DIRECTION VENT
 - HUMIDITE RELATIVE
 - DUREE INSOLATION
 - RAYONNEMENT GLOBAL
 - NEBULOSITE
 - CODE TEMPS OBSERVE
 - VISIBILITE
 - PRESSION NIVEAU MER
- Autres paramètres :
 - Heure
 - Mois

Remarque

Pour l'instant aucun décalage temporel n'a été introduit : on regarde l'influence du paramètre mesuré au moment t sur l'erreur enregistrée au moment t . Il sera également intéressant d'évaluer l'effet de la météo mesurée au moment $t-\Delta t$ avec Δt correspondant à plusieurs heures ou jours.

C. Résultats

La classification des paramètres en compte est donnée dans le tableau ci-dessous.

Paramètres	Facteur d'influence
HEURE	0,28
VITESSE_VENT	0,24
HUMIDITE_RELATIVE	0,23
DIRECTION_VENT	0,20
CODE_TEMPS_OBSERVE	0,19
PRESSION_STATION	0,10
PRESSION_NIVEAU_MER	0,10
RAYONNEMENT_GLOBAL	0,05
MOIS	0,05
TEMPERATURE	0,04
DUREE_INSOLATION	0,04
PRECIPITATION	0,03
VISIBILITE	--

Tableau 3 Classification des paramètres en fonction de leur influence sur l'erreur

Le coefficient d'influence permet de classer les paramètres en fonction de leur influence; il est égal à 100% quand le paramètre correspond à la variable de sortie elle-même.

En tête du classement figure l'heure. Autrement dit, le modèle J+1 nécessite un calibrage par rapport à ce paramètre. Ceci peut être illustré par le graphique ci-dessous qui donne l'évolution des mesures réelles et modélisées pour le 1/01/2014 à Blois.

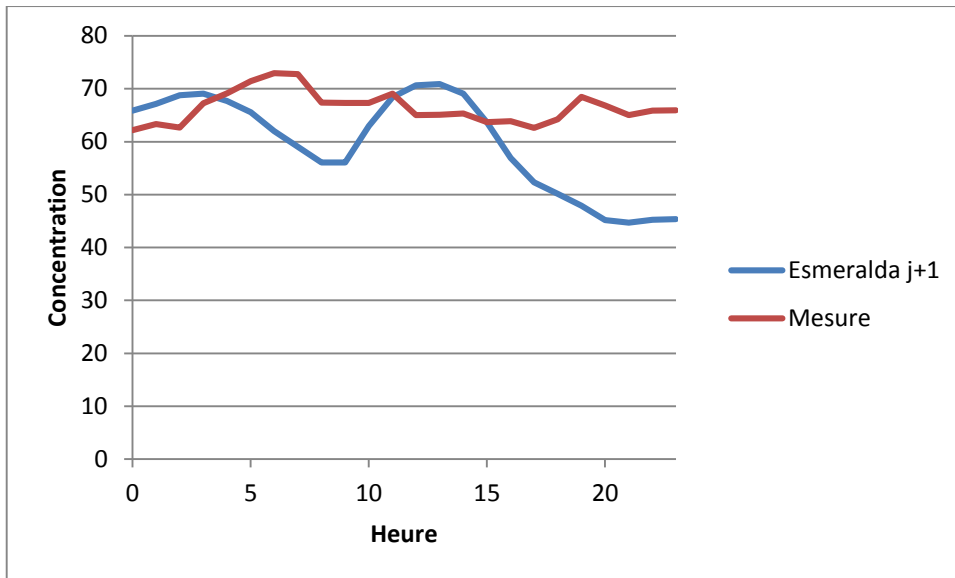


Figure 3 Concentrations horaires réelle et modélisée pour 1/01/2014

V. Construction de la table de calibration

Nous illustrons la construction de la table de calibration en prenant en compte le paramètre le plus influent, c'est-à-dire l'heure. Pour chaque heure, l'erreur moyenne est indiquée dans le tableau ci-dessous.

Heure	Erreur moyenne
0	5,6
1	7,2
2	8,2
3	9,2
4	10,8
5	12,8
6	15,7
7	17,8
8	18,1
9	15,9
10	13,3
11	10,1
12	7,9
13	5,5
14	3,4
15	1,6
16	0,4
17	-0,1
18	-0,3
19	1,4
20	3,7
21	5,2
22	5,7
23	5,0

Tableau 4 Table de calibration

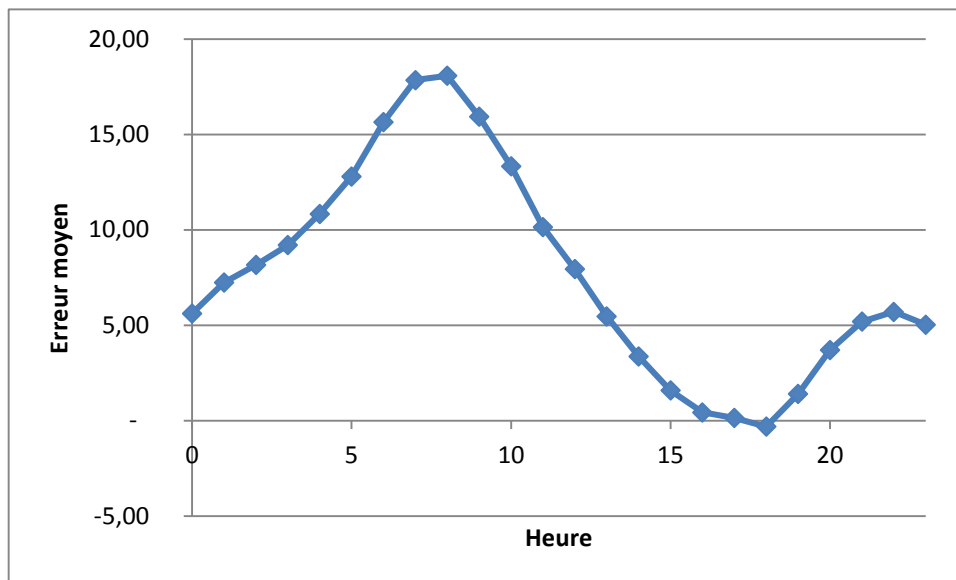


Figure 4 Evolution de l'erreur moyenne par heure

VI. Validation de la table de calibration

Afin de valider la table de calibration présentée, nous vérifions que :

- la correction entraîne une diminution globale de l'erreur ;
- elle permet d'améliorer les prévisions sur un autre jeu de données, externe.

Dans les parties ci-dessous nous vérifions ces deux points.

A. Comparaison des erreurs pour le même jeu de données

Nous corrigeons le modèle Esmeralda J+1 pour Blois en 2014 ainsi : pour chaque concentration horaire modélisée, nous retranchons l'erreur moyenne fournie par la table de calibration.

Nous obtenons une nouvelle base de données "corrigée" et nous calculons la nouvelle erreur.

La comparaison de la loi de probabilité de l'erreur avant et après correction est donnée ci-dessous.

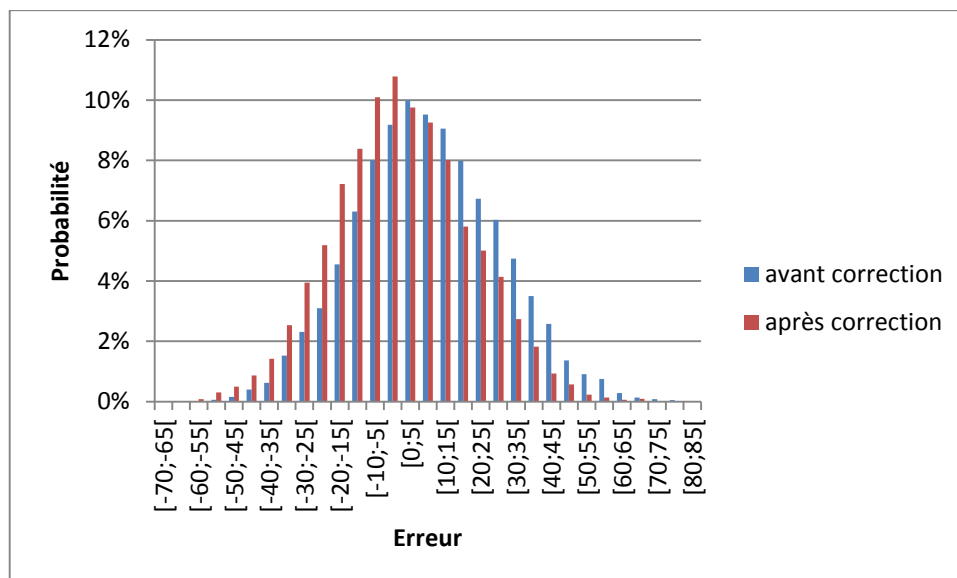


Figure 5 Loi de probabilité de l'erreur après correction simple

Le tableau ci-dessous résume la comparaison des lois de probabilité de l'erreur avant et après correction :

	Avant correction	Après correction
Minimum	-68,6	-68,3
Maximum	82,3	75,0
Moyenne	7,7	0,0
Borne 5 %	-25,0	-31,5
Borne 95%	41,7	32,8

La loi de probabilité de l'erreur après correction est décalée vers la gauche par rapport à la loi de probabilité de l'erreur avant correction.

Si l'on constate que la probabilité des valeurs élevées positives a diminué après correction, on constate également que la probabilité des valeurs élevées négatives a augmenté.

Nous comparons également les loi de probabilité de la valeur absolue de l'erreur avant et après correction, le graphique est donné ci-dessous.

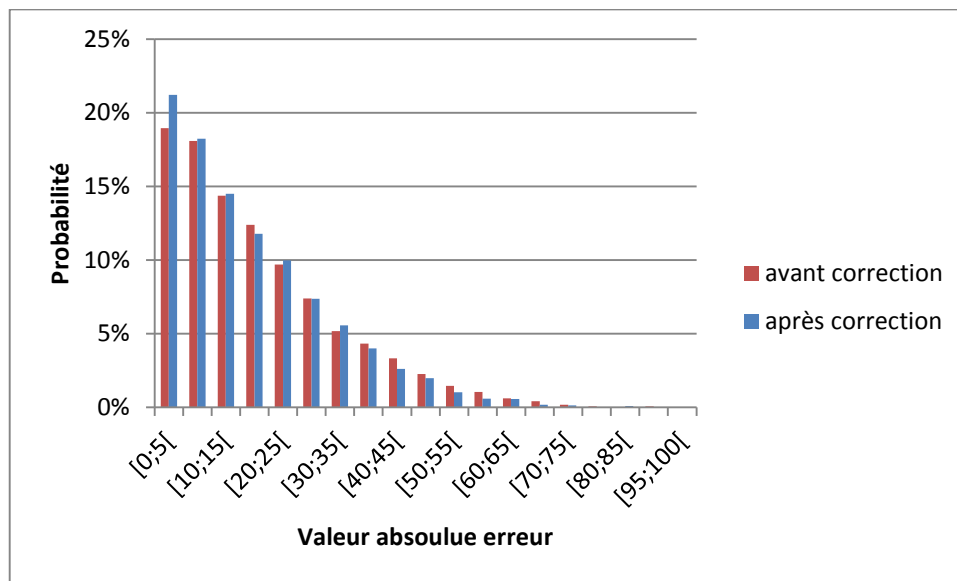


Figure 6 Lois de probabilité de la valeur absolue de l'erreur avant et après correction

Le tableau ci-dessous résume la comparaison des lois de probabilité de la valeur absolue de l'erreur avant et après correction :

	Avant correction	Après correction
Minimum	0	0
Maximum	82,3	75,0
Moyenne	17,1	15,5
Borne 5 %	1,2	1,3
Borne 95%	42,5	37,9

La comparaison des lois de probabilité de la valeur absolue de l'erreur permet de constater que, après correction :

- la proportions des erreurs ayant une valeur absolue faible (valeurs comprises entre 0 et 20) augmente : elle passe de 65 % à 69 %;
- à l'inverse, la proportions des erreurs ayant une valeur absolue élevée diminue (elle passe de 35 % à 31 %).

La correction apportée amène ainsi à une (faible) diminution de l'erreur.

B. Validation sur un jeu de données externe

1. Avant correction

Nous effectuons les mêmes corrections sur un autre jeu de données : Esmeralda J+1 pour Blois en 2012.

Loi de probabilité de l'erreur avant correction pour 2012 et pour 2014 est donnée ci-dessous.

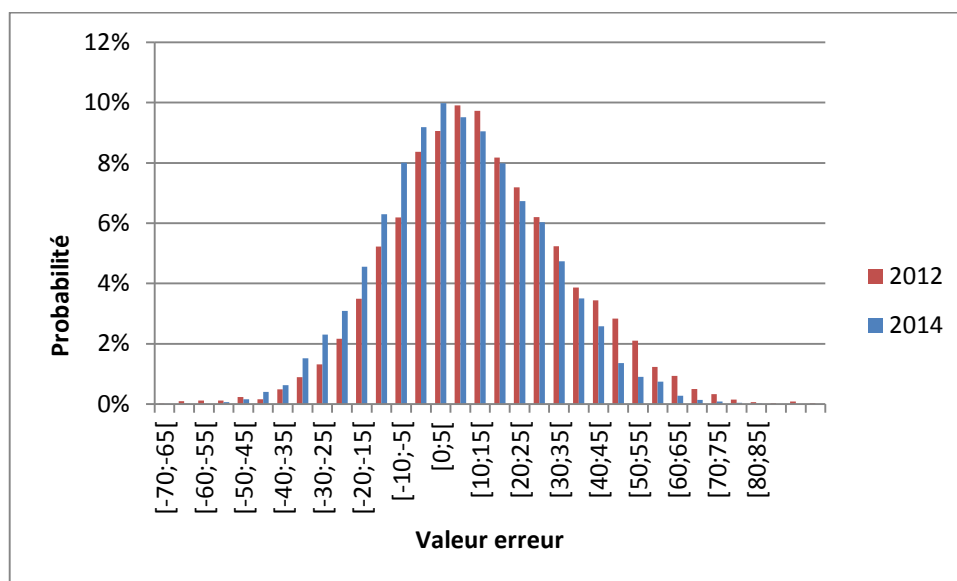


Figure 7 Histogramme de l'erreur

Le tableau ci-dessous résume la comparaison des lois de probabilité de l'erreur en 2012 et en 2014 :

	2012	2014
Minimum	-77,5	-68,6
Maximum	96,5	82,3
Moyenne	7,5	7,7
Borne 5 %	-26,1	-25,0
Borne 95%	45,8	41,7

La loi de probabilité de l'erreur pour Blois en 2012 est sensiblement proche de la loi de probabilité de l'erreur en 2014.

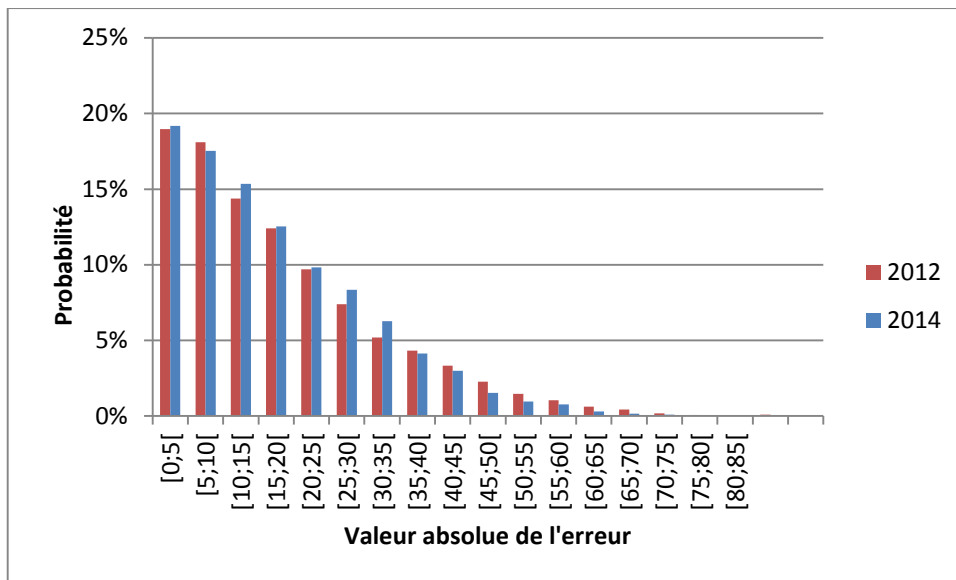


Figure 8 Histogramme de la valeur absolue de l'erreur

Le tableau ci-dessous résume la comparaison des lois de probabilité de la valeur absolue de l'erreur en 2012 et 2014 :

	2012	2014
Minimum	0	0
Maximum	96,5	82,3
Moyenne	18,1	17,1
Borne 5 %	1,2	1,2
Borne 95%	47,6	42,5

2. Après correction

Nous corrigeons le modèle Esmeralda J+1 pour l'année 2012 à partir de la table de calibration construite pour l'année 2014. La comparaison de loi de probabilité de la valeur absolue de l'erreur avant et après correction est donnée ci-dessous.

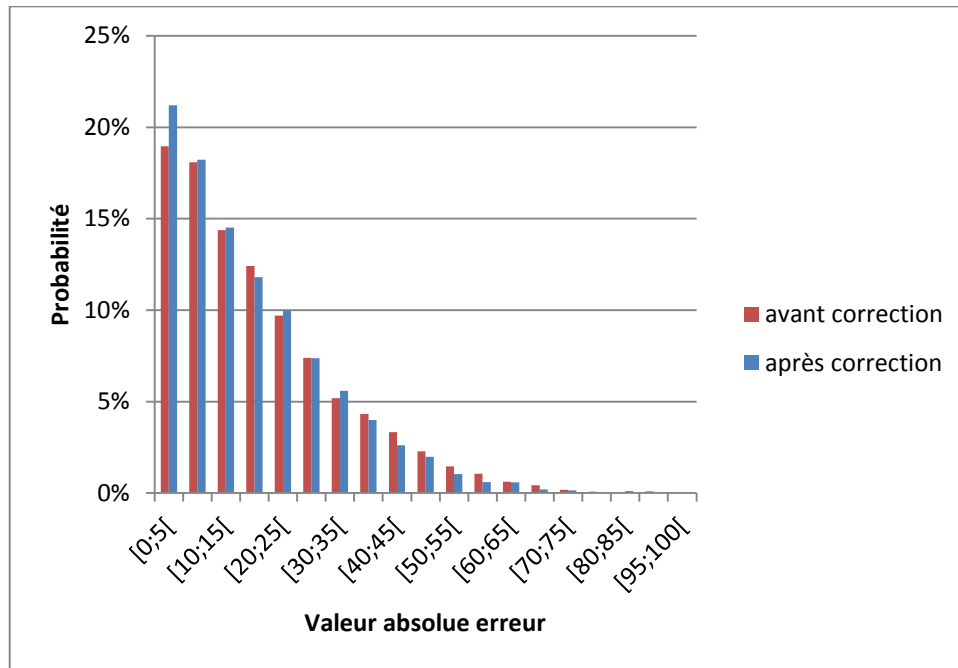


Figure 9 Lois de probabilité de la valeur absolue de l'erreur avant et après correction pour Esmeralda j+1 Blois 2012

Le tableau ci-dessous résume la comparaison des lois de probabilité de la valeur absolue de l'erreur avant et après correction :

	Avant correction	Après correction
Minimum	0	0
Maximum	96,5	88,6
Moyenne	18,1	16,7
Borne 5 %	1,2	1,2
Borne 95%	47,6	42,1

La comparaison des lois de probabilité de la valeur absolue de l'erreur permet de constater que, après correction :

- la proportions des erreurs ayant une valeur absolue faible (valeurs comprises entre 0 et 20) augmente : elle passe de 65 % à 73 %;

- à l'inverse, la proportions des erreurs ayant une valeur absolue élevée diminue (elle passe de 35 % à 27 %).

Nous en concluons que la table de calibration obtenue à partir des données 2014 permet d'améliorer le modèle Esmeralda J+1 de 2012.

VII. Conclusion

La table de calibration construite à partir des données Blois 2014 et Esmeralda j+1 permet d'améliorer les prévisions pour le jeu de données Blois 2014 et un jeu de données externe, Blois 2012.

Toutefois, les améliorations réalisées sont peu significatives (de l'ordre de 5 %). La table de calibration présentée pourra donc être améliorée en prenant en compte d'autres paramètres influents tels que la vitesse du vent ou l'humidité relative ou le paramètre émission, qui ne nous a pas été fourni, mais qui pourrait vraisemblablement jouer un rôle important sur la valeur l'erreur.