

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



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

Rapport 2

adressé à

Lig'Air

par la

Société de Calcul Mathématique SA

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

Juillet 2015

# Résumé opérationnel

## ***Objectifs***

L'association Lig'Air est responsable de la surveillance de la qualité de l'air en région Centre. En cas de dépassement des seuils réglementaires, Lig'Air doit alerter les préfetures départementales le plus tôt possible afin que les mesures d'urgences nécessaires soient mises en œuvre.

Pour cela, deux modèles de prévision sont disponibles : un modèle régional, Esmeralda, et un modèle national, Prev'Air. Ces modèles s'appuient des données d'émission et sur les modèles de prévision de météo.

Bien que les prévisions fournies par ces modèles soient jugées satisfaisantes la majorité du temps, il existe des cas où des pics de pollution interviennent de façon imprévisible : Lig'Air a demandé à la SCM de définir une méthode probabiliste permettant d'anticiper ces types d'évènements et permettre ainsi de réagir rapidement.

## ***Traitement et analyse des données***

Les données fournies par Lig'Air sont de nature différente :

- données météo : réelles et modélisées ;
- données de mesures de concentration en Ozone et PM10 ;
- résultats des modèles de prévision.

Il s'agit de données horaires sur la période 2008-2014, pour 12 stations de mesure et 7 stations météo.

Un travail important de traitement des données est réalisé. Au final nous obtenons une base de données unique regroupant l'ensemble de l'information.

## ***Résultats de prévision actuels***

Notre analyse de données met en évidence le taux de détection de dépassement du seuil réglementaire pour chaque modèle de prévision. Les résultats obtenus pour les PM10 sont les suivants :

- pour le modèle Prev'Air :
  - dans 95 % des cas le modèle Prev'air ne détecte pas le dépassement du seuil réglementaire pour le jour même ;
  - dans 97 % des cas le modèle Prev'air ne détecte pas le dépassement du seuil réglementaire pour le lendemain ;
  - dans 0.1 % des cas le modèle Prev'air annonce un dépassement du seuil de façon erronée (fausses alertes) ;

- pour le modèle Esmeralda :
  - dans 99 % des cas le modèle Esmeralda ne détecte pas le dépassement du seuil réglementaire pour le jour même ;
  - dans 100 % des cas le modèle Esmeralda ne détecte pas le dépassement du seuil réglementaire pour le lendemain ;
  - dans 0.1 % des cas le modèle Esmeralda annonce un dépassement du seuil de façon erronée (fausses alertes).

Nous en concluons que le pourcentage d'échec est très élevé. Une amélioration des modèles existants est bien nécessaire. Une analyse de données sur les périodes "à risque" permet également de constater que, quand les pics de pollution ne sont pas détectés par les modèles, les modèles météo s'écartent aussi de façon significative de la réalité. Par conséquent, il n'est pas possible d'améliorer la détection des dépassements de seuil réglementaire en prenant en compte la météo modélisée (à moins d'améliorer les modèles de prévision météo), contrairement à ce qui avait été proposé dans le rapport n°1.

### ***Méthode d'amélioration proposée***

Après avoir testé plusieurs méthodes d'amélioration possibles reposant sur de nouveaux indicateurs et en traitant l'exemple des PM10, nous avons défini une méthode simple et robuste.

Nous constatons qu'en prenant en compte la mesure du jour même à 10h00 les dépassements de seuil réglementaire peuvent être détectés avec un taux de réussite satisfaisant.

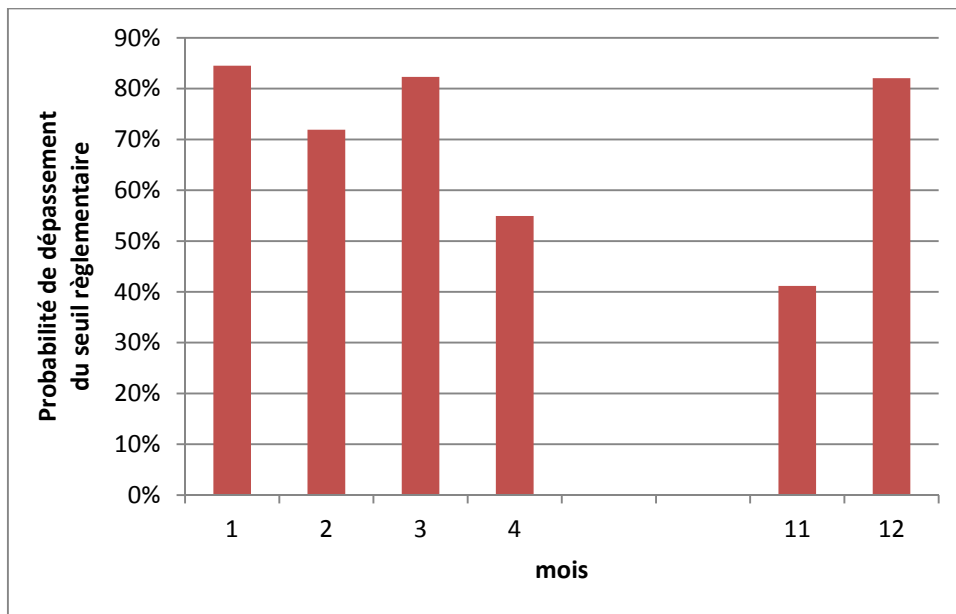
En exploitant un historique de 6 ans (2008-2014) pour 13 stations de mesure et 6 stations météo nous déterminons le risque que le seuil réglementaire soit dépassé en fonction de la teneur en PM10 à 10h00 du jour même.

Teneur à en PM10 à 10h00 supérieure à :	Probabilité de dépassement du seuil réglementaire
40 $\mu\text{g}/\text{m}^3$	32 %
50 $\mu\text{g}/\text{m}^3$	48 %
60 $\mu\text{g}/\text{m}^3$	71 %
70 $\mu\text{g}/\text{m}^3$	83 %

**Tableau 1 Risque de dépassement en fonction de la teneur en PM10 à 10h00**

En utilisant un seuil de 60  $\mu\text{g}/\text{m}^3$ , la probabilité de dépasser le seuil réglementaire est de 70 %.

Ce résultat peut aussi être affiné par mois, ce qui est donné ci-dessous.



**Figure 1** Probabilité de dépasser le seuil réglementaire en fonction du mois, quand la teneur en PM10 à 10h00 est supérieure à 60  $\mu\text{g}/\text{m}^3$

Au final, nous suggérons à Lig'Air de prendre en compte pour chaque jour des périodes "à risque", pour chaque station de mesure, la mesure réelle de pollution en PM10 à 10h00.

## Sommaire

Résumé opérationnel.....	2
I. Introduction.....	6
II. Traitement et analyse des données disponibles.....	7
A. Bases de données exploitées.....	7
B. Traitement et analyse des données.....	7
1. Analyse des données.....	8
2. Croisement de l'information.....	9
3. Création d'une base de données journalière.....	11
III. Analyse de l'erreur.....	12
A. Distribution.....	12
1. Erreur associée au modèle Prev'Air pour le jour même.....	12
2. Erreur associée au modèle Esmeralda pour le jour même.....	13
B. Comparaison des deux erreurs.....	14
IV. Analyse des situations "à risque".....	15
A. Modèle Prev'air.....	15
B. Modèle Esmeralda.....	15
C. Résultats hiérarchisation pour le modèle Prev'air.....	16
1. Paramètres pris en compte.....	16
2. Résultats de la hiérarchisation des paramètres.....	18
V. Prise en compte de nouveaux indicateurs.....	24
VI. Détection des dépassements de seuil en exploitant les mesures réelles à 10H00.....	28
A. Risque de dépassement en fonction de la teneur en PM10 à 10h00.....	28
B. Amélioration de la prévision en prenant en compte le mois.....	29
1. Profil horaire journalier.....	29
2. Probabilité de dépassement en fonction du mois.....	30

## I. Introduction

Lig'Air veut améliorer ses prévisions de qualité de l'air à échelle régionale, et en particulier le déclenchement des alertes relatives à des épisodes de pollution.

Deux modèles sont utilisés : ESMERALDA, plate-forme regroupant dix régions, et PREVAIR, qui est un modèle national.

Le déclenchement des alertes relatives aux épisodes de pollution est à la charge de Lig'Air ; il est fait par prévision, à partir des informations données par ces deux modèles. Elles se font actuellement sur les échéances J0 (aujourd'hui), J+1 (demain) pour PM10 et l'ozone O3.

La difficulté rencontrée par Lig'Air est que ces modèles sont peu pertinents pour prévoir les dépassements. Il faut agir vite (avant midi) et les indications fournies par les modèles sont peu appropriées.

Dans la note 1, nous avons illustré la méthode probabiliste SCM d'amélioration du modèle de prévision existant par le moyen d'un exemple simple de l'ozone dans une station. Nous avons construit une table de calibration permettant de corriger la teneur horaire en ozone.

Dans ce rapport, nous prenons en compte le cas des PM10 sur l'ensemble de l'historique. L'idée est celle de s'intéresser aux situations "à risque" et de définir une méthode permettant d'agir vite. Après avoir élaboré et testé plusieurs méthodes, nous présentons celle qui amène aux résultats les plus satisfaisants.

Dans un premier temps nous présentons les données traitées et exploitées.

La partie II présente une analyse de l'erreur des modèles de prévision (Esmeralda et Prev'air).

Dans la partie III nous donnons une analyse des situations "à risque" qui n'ont pas été correctement détectées. Suit une introduction des indicateurs envisagés afin d'améliorer la méthode.

Au final est proposée la méthode simple et robuste de détection des dépassements de seuils réglementaires.

## II. Traitement et analyse des données disponibles

### A. Bases de données exploitées

Notre travail repose sur l'exploitation des bases de données suivantes :

- données météo mesurée ;
- données météo modélisée ;
- mesures PM10 ;
- concentration PM10 modélisée par les deux modèles :Esmeralda, Prev'air pour le jour J et le jour J+1 ;

### B. Traitement et analyse des données

Un travail très important de traitement de données a été réalisé, nous avons :

1. analysé les données afin d'identifier les éventuelles données aberrantes ;
2. croisé l'information de chaque base ;
3. synthétisé l'information horaire en information journalière.

## 1. Analyse des données

Nos analyses montrent que, globalement, les données sont de bonne qualité. La part des données manquantes est non négligeable. Par ailleurs, le nombre de stations renseignées au cours du temps n'est pas constant. Nous donnons ci-dessous, sous forme de tableau, le nombre de stations pour lesquelles les données sont disponibles sur la période 2008-2014.

		2008	2009	2010	2011	2012	2013	2014
Préfecture	34011	X						
La Source	34012	X	X	X	X	X	X	
Gambetta	34013	X	X	X	X		X	X
Saint Jean de Braye	34014	X	X	X	X	X	X	
Montargis	34018	X	X	X	X	X	X	X
Joué les Tours	34024	X						
La Bruyère	34025	X	X	X	X	X	X	X
Pompidou	34028	X	X	X	X	X	X	X
La Source CNRS	34029							X
Baffier	34030							X
Leblanc	34032	X	X	X	X	X	X	X
Vierzon	34034	X	X	X	X	X	X	X
Verneuil	34038		X	X	X	X		
Lucé	34042	X	X	X	X	X	X	X
Dreux Centre	34046	X	X	X	X	X	X	X
St Remy	34047							X
Châteauroux Sud	34051	X	X	X	X	X	X	X
Blois Nord	34061	X	X	X	X	X	X	X

**Tableau 2 Stations renseignées pour chaque année**



## 2. Croisement de l'information

Pour croiser les données nous avons été confrontés au problème suivant :

- le format des dates est différent pour chaque base ;
- le nombre de dates renseignées est différent selon la base.

De ce fait, nous avons créé un nouveau code texte renseignant sur la date et réalisé une nouvelle base correspondant à l'intersection de chaque base.

Pour ce qui est du croisement des données météo et des données de concentration, la correspondance est donnée dans le tableau ci-dessous : pour chaque station de mesure de concentration nous choisissons la station météo la plus proche.

<b>Station mesure</b>	<b>Station météo la plus proche</b>
Chambord	Blois
Préfecture	Orleans
La Source	Orleans
Saint Jean de Braye	Orleans
Marigny les Usages	Orleans
Montargis	Amilly
Joué les Tours	Tours
La Bruyère	Tours
Tours Périurbaine	Tours
Pompidou	Tours
La Source CNRS	Orleans
Baffier	Bourges
Leblanc	Bourges
Vierzon	Romorantin
Verneuil	Chateauroux
Fulbert	Chartres
Lucé	Chartres
Oysonville	Chartres
Châteauroux Sud	Chateauroux
Faverolles	Romorantin
Issoudun	Chateauroux
Blois Nord	Blois
Blois Centre	Blois

**Tableau 3 Correspondance entre les stations météo et les stations de mesure**

La localisation géographique des stations de mesure de concentration et des stations météo est donnée ci-dessous.



Figure 2 Localisation des les stations de mesure météo et des stations de mesure de pollution

### 3. Création d'une base de données journalière

Les bases de données à notre disposition sont horaires. Il faut synthétiser l'information au niveau journalier. Pour chaque paramètre nous calculons la moyenne journalière.

Pour les données concentration nous calculons également le maximum journalier.

Enfin, pour le paramètre météo "Direction du vent" nous avons calculé 9 indicateurs :

- Dvnul : somme des heures où la direction du vent est nulle ;
- DV\_1 : somme des heures où la direction du vent est comprise entre  $0^\circ$  et  $45^\circ$  ;
- DV\_2 : somme des heures où la direction du vent est comprise entre  $45^\circ$  et  $90^\circ$  ;
- DV\_3 : somme des heures où la direction du vent est comprise entre  $90^\circ$  et  $135^\circ$  ;
- DV\_4 : somme des heures où la direction du vent est comprise entre  $135^\circ$  et  $180^\circ$  ;
- DV\_5 : somme des heures où la direction du vent est comprise entre  $180^\circ$  et  $225^\circ$  ;
- DV\_6 : somme des heures où la direction du vent est comprise entre  $225^\circ$  et  $270^\circ$  ;
- DV\_7 : somme des heures où la direction du vent est comprise entre  $270^\circ$  et  $315^\circ$  ;
- DV\_8 : somme des heures où la direction du vent est comprise entre  $315^\circ$  et  $360^\circ$ .

La nouvelle de base de données créée est fournie avec ce rapport sous forme de fichier Excel. Il existe un onglet par modèle de prévision pris en compte. Les résultats donnés dans les parties ci-dessous résultent de l'exploitation directe de cette base. Afin de garantir la robustesse des conclusions, l'ensemble des données a été pris en compte.

### III. Analyse de l'erreur

#### A. Distribution

Nous donnons ci-dessous la loi de probabilité de l'erreur des deux modèles, ils résultent de la différence de la valeur réelle et de la valeur modélisée.

##### 1. Erreur associée au modèle Prev'Air pour le jour même

Nous disposons en tout de environ 23 000 enregistrements. La loi de probabilité et la fonction de répartition de l'erreur sont données ci-dessous.

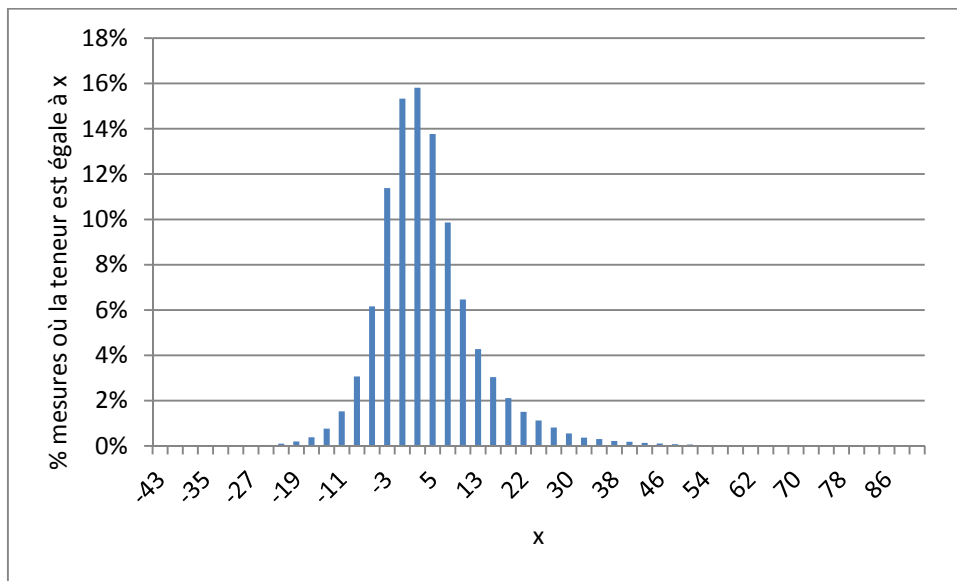


Figure 3 Loi de probabilité de l'erreur du modèle Prev'air pour le jour j

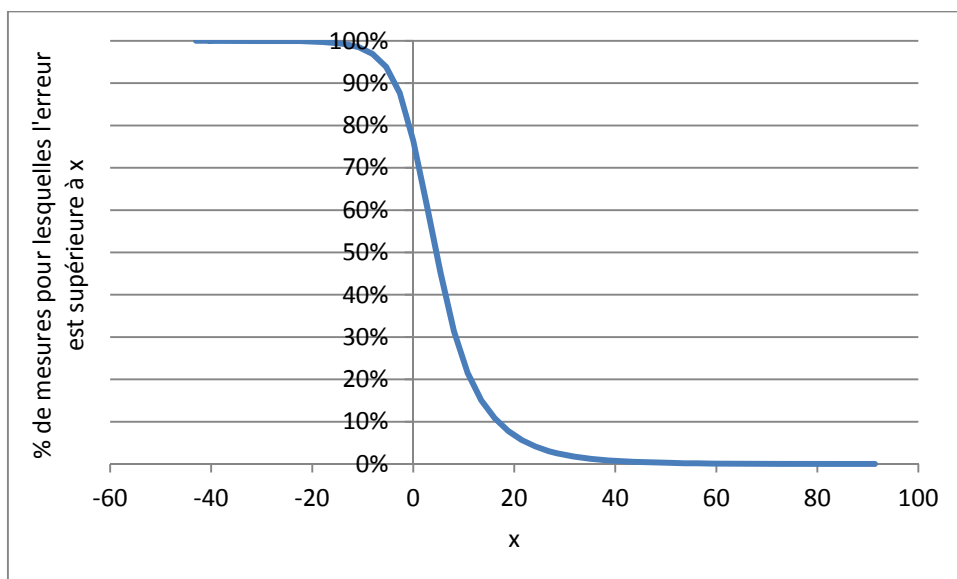


Figure 4 Distribution de l'erreur du modèle Prev'air pour le jour j

Le tableau ci-dessous résume les principales caractéristiques de l'erreur du modèle Prev'air.

<b>médiane</b>	4,5
<b>borne 95%</b>	22,6
<b>borne 5%</b>	-6,1
<b>moyenne</b>	5,8

Tableau 4 Principales caractéristiques de l'erreur du modèle Prev'air pour le jour j

## 2. Erreur associée au modèle Esmeralda pour le jour même

Nous disposons en tout de environ 4 000 enregistrements. La loi de probabilité et la fonction de répartition de l'erreur sont données ci-dessous.

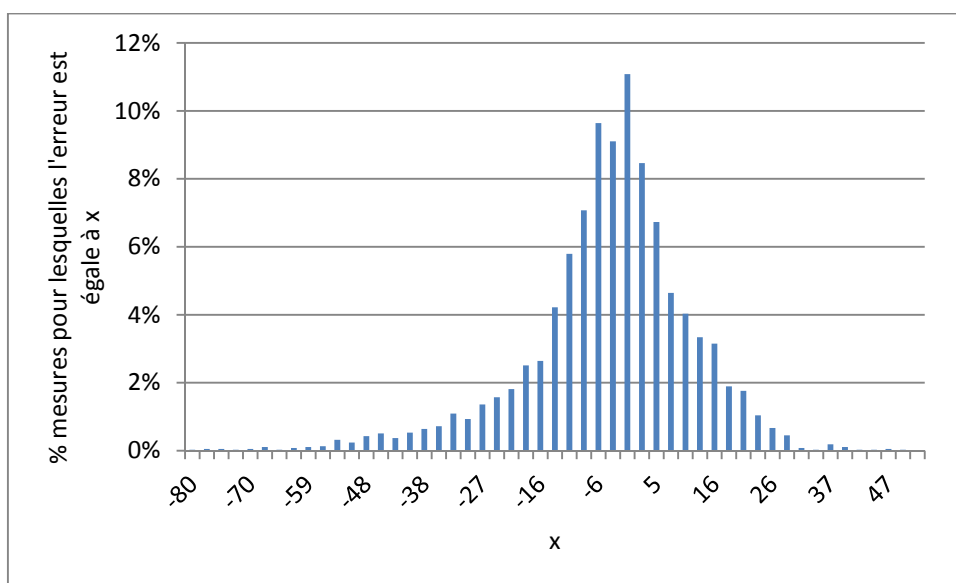


Figure 5 Loi de probabilité de l'erreur du modèle Esmeralda pour le jour j

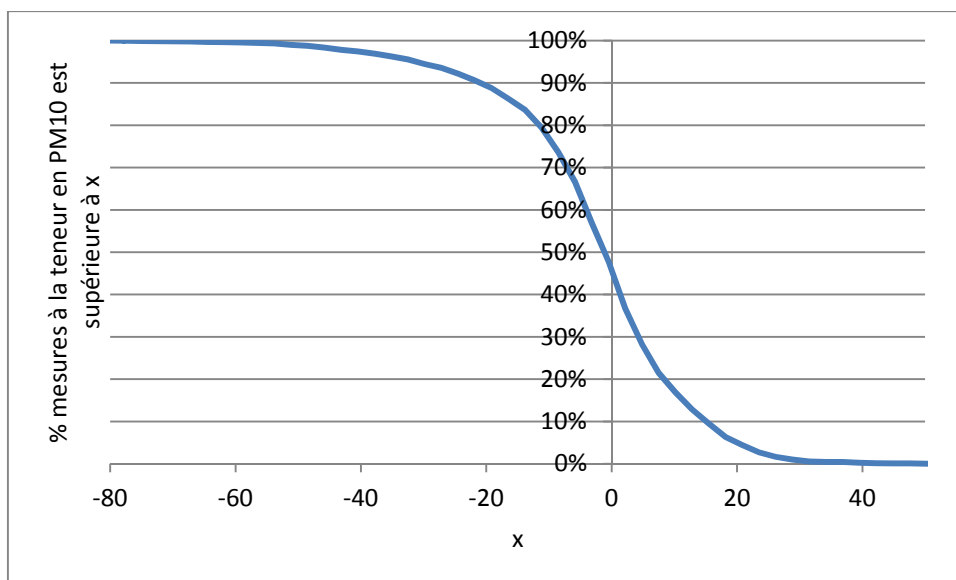


Figure 6 Distribution de l'erreur du modèle Esmeralda pour le jour j

Le tableau ci-dessous résume les principales caractéristiques de l'erreur du modèle Esmeralda.

<b>médiane</b>	-1,1
<b>borne 95%</b>	19,8
<b>borne 5%</b>	-31,0
<b>moyenne</b>	-2,5

Tableau 5 Principales caractéristiques de l'erreur du modèle Esmeralda pour le jour j

### B. Comparaison des deux erreurs

En rouge, l'histogramme de l'erreur du modèle Prev'air, en bleu, l'histogramme de l'erreur du modèle Esmeralda.

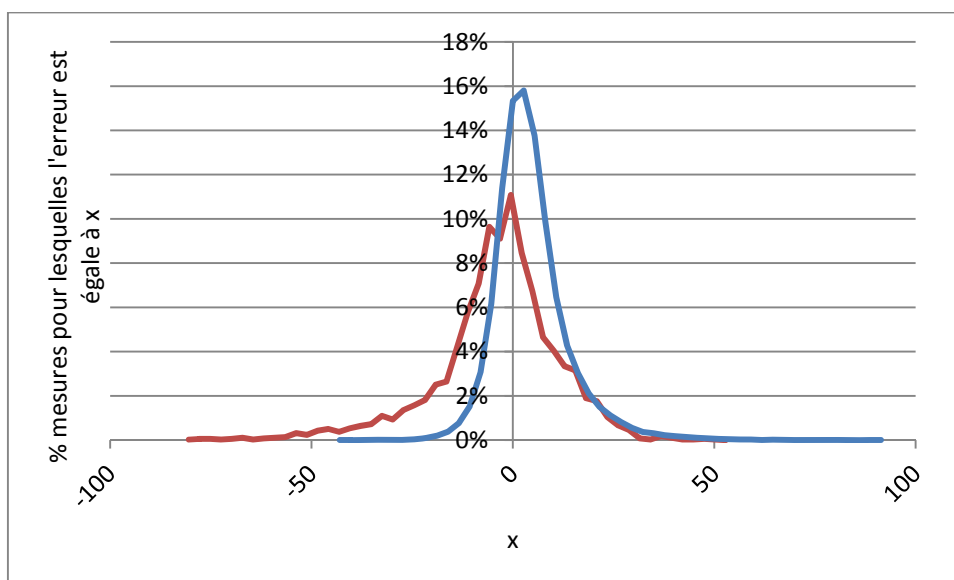


Figure 7 Comparaison des lois de probabilité de l'erreur des deux modèles

Nous constatons que les modèles ont un comportement différent.

L'erreur moyenne associée au modèle Prev'air est positive : le modèle sous-estime la teneur en polluant, la loi de probabilité est plutôt concentrée.

L'erreur moyenne associée au modèle Esmeralda est négative : le modèle surestime la teneur en polluant, la loi de probabilité est très étalée.

En utilisant la méthode proposée dans la note 1, nous pourrions construire des tables de calibration permettant de diminuer l'erreur moyenne en centrant en zéro la loi de probabilité.

Dans ce rapport nous nous intéressons plus particulièrement aux situations "à risque" : les situations où le seuil réglementaire PM10 a été dépassé.

Les questions que nous nous posons sont les suivantes :

- Est-ce que les modèles de prévision permettaient d'anticiper ce dépassement de seuil? Si ce n'est pas le cas : à quoi cela est-il dû ?
- Comment peut-on améliorer les modèles afin d'éviter ces cas de figure ?
- Quelle méthode complémentaire peut aider Lig'air à anticiper ces événements?

#### IV. Analyse des situations "à risque"

Si nous prenons en compte les situations où seuil réglementaire d'information de  $50 \mu\text{g m}^{-3}$  a été dépassé nous obtenons les résultats ci-dessous.

##### A. Modèle Prev'air

Nombre de fois où la teneur a dépassé le seuil d'information	791
Nombre de fois où le modèle Prev'air a anticipé le dépassement de ce seuil en j0 correctement	38
Nombre de fois où le modèle Prev'air a anticipé le dépassement de ce seuil en j1 correctement	30
Nombre de fois où le modèle Prev'air a lancé une fausse alerte de dépassement du seuil en j0	39
Nombre de fois où le modèle Prev'air a lancé une fausse alerte de dépassement du seuil en j1	37

**Tableau 6 Comportement du modèle Prev'air dans les situations "à risque"**

En conclusion :

- dans 95 % des cas le modèle Prev'air ne détecte pas le dépassement du seuil en j0 ;
- dans 97 % des cas le modèle Prev'air ne détecte pas le dépassement du seuil en j1 ;
- dans 0.1 % des cas le modèle Prev'air annonce un dépassement du seuil de façon erronée (fausse alerte).

##### B. Modèle Esmeralda

Nombre de fois où la teneur a dépassé le seuil d'information	141
Nombre de fois où le modèle Esmeralda a anticipé le dépassement de ce seuil en j0 correctement	2
Nombre de fois où le modèle Esmeralda a anticipé le dépassement du seuil en j1 correctement	0
Nombre de fois où le modèle Esmeralda a lancé une fausse alerte de dépassement du seuil en j0	2
Nombre de fois où le modèle Esmeralda a lancé une fausse alerte de dépassement du seuil en j1	0

**Tableau 7 Comportement du modèle Esmeralda dans les situations "à risque"**

## Remarque

Le modèle Esmeralda nous a été donné pour les années 2012-2014, c'est pour cette raison que le nombre de situation à risque est plus faible que pour les données Prev'air.

En conclusion :

- dans 99 % des cas le modèle Esmeralda ne détecte pas le dépassement du seuil en  $j_0$  ;
- dans 100 % des cas le modèle Esmeralda ne détecte pas le dépassement du seuil en  $j_1$ ;
- dans 0.1 % des cas le modèle Esmeralda annonce un dépassement du seuil de façon erronée (fausse alerte).

Autant pour le modèle Prev'air que pour le modèle Esmeralda, nous constatons que les pourcentages d'échec du modèle sont très élevés. Afin d'en comprendre la raison nous effectuons une hiérarchisation des paramètres sur l'erreur du modèle pour ces 731 valeurs.

### *C. Résultats hiérarchisation pour le modèle Prev'air*

Le principe de la méthode de hiérarchisation des paramètres dont les principes 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.

#### **1. Paramètres pris en compte**

Les paramètres pris en compte sont les suivants :

Météo réelle :

- PRECIPITATION ;
- TEMPERATURE ;
- PRESSION\_STATION ;
- VITESSE\_VENT ;
- DIRECTION\_VENT ;
- HUMIDITE\_RELATIVE ;
- DUREE\_INSOLATION ;
- RAYONNEMENT\_GLOBAL ;
- NEBULOSITE ;
- CODE\_TEMPS\_OBSERVE ;
- VISIBILITE ;
- PRESSION\_NIVEAU\_MER ;
- Dvnul ;
- DV\_1 ;
- DV\_2 ;
- DV\_3 ;
- DV\_4 ;
- DV\_5 ;
- DV\_6 ;



- DV\_7 ;
- DV\_8 ;

Météo modélisée j0 et j+1 :

- Température ;
- humidité\_à\_2m ;
- vitesse\_du\_vent\_à\_10m ;
- direction du vent\_à\_10m ;
- Radiation onde courte vers le sol ;
- NA ;
- Flux de chaleur sensible ;
- Flux de chaleur latente ;
- Radiation onde longue vers le sol ;
- Epaisseur optique basse ;
- Epaisseur optique moyenne ;
- Epaisseur optique haute ;
- Colonne intégrée d'eau ;
- Colonne intégrée neige ;
- Colonne intégrée grésil ;
- Colonne intégrée glace ;
- Colonne intégrée vapeur ;
- Précipitations\_totales ;
- hauteur\_de\_couche\_limite ;
- Longueur\_de\_Monin\_Obukhov ;
- Radiation d'onde longue sortante ;
- Epaisseur optique totale (méthode 1) ;
- Epaisseur optique totale (méthode 2) ;
- Vitesse\_du\_vent\_moyenne\_sur\_les\_1500m ;
- Direction\_du\_vent\_moyenne\_sur\_les\_1500m ;
- Humidité\_du\_sol ;
- Pression\_en\_surface ;
- Delta\_température\_entre\_le\_niveau\_100m\_et\_le\_sol ;
- Dvnul ;
- DV\_1 ;
- DV\_2 ;
- DV\_3 ;
- DV\_4 ;
- DV\_5 ;
- DV\_6 ;
- DV\_7 ;
- DV\_8.

## 2. Résultats de la hiérarchisation des paramètres

Nous hiérarchisons les paramètres influant sur l'erreur du modèle de prévision.

Modèle Prev'air

Les résultats obtenus pour le modèle Prev'air sont donnés dans le tableau ci-dessous.

Le coefficient d'influence renseigne de façon qualitative sur l'influence du paramètre sur la variable de sortie (l'erreur). Le sens de variation informe si l'influence du paramètre est croissante (+) ou décroissante (-).

<i>Paramètre</i>	<i>Coefficient d'influence</i>	<i>Sens de variation</i>
erreur	100%	+
<b>météo_modélisée_j0_direction du vent _à_10m</b>	<b>49%</b>	<b>+</b>
<b>HUMIDITE_RELATIVE</b>	<b>38%</b>	<b>-</b>
<b>TEMPERATURE</b>	<b>35%</b>	<b>+</b>
<b>météo_modélisée_j0_vitesse du vent _à_10m</b>	<b>31%</b>	<b>-</b>
<b>RAYONNEMENT_GLOBAL</b>	<b>27%</b>	<b>+</b>
<b>météo_modélisée_j0_Température</b>	<b>24%</b>	<b>-</b>
<b>VITESSE_VENT</b>	<b>24%</b>	<b>+</b>
<b>météo_modélisée_j0_humidité_à_2m</b>	<b>23%</b>	<b>+</b>
meteo_reelle_DV_1	22%	+
VISIBILITE	21%	+
DUREE_INSOLATION	20%	+
<b>DIRECTION_VENT</b>	<b>20%</b>	<b>-</b>
PRESSION_STATION	16%	-
meteo_reelle_DV_8	15%	+
PRESSION_NIVEAU_MER	14%	-
CODE_TEMPS_OBSERVE	5%	-
meteo_reelle_DV_2	2%	+
météo_modélisée_j1_DV_8	0%	-
météo_modélisée_j1_DV_7	0%	-
météo_modélisée_j1_DV_6	0%	-
météo_modélisée_j1_DV_5	0%	-
météo_modélisée_j1_DV_4	0%	-
météo_modélisée_j1_DV_3	0%	-
météo_modélisée_j1_DV_2	0%	-
météo_modélisée_j1_DV_1	0%	-
météo_modélisée_j1_Dvnul	0%	-
météo_modélisée_j1_Delta_température_entre_le_niveau_100m_et_le_sol	0%	-
météo_modélisée_j1_Pression_en_surface	0%	-
météo_modélisée_j1_Humidité_du_sol	0%	-
météo_modélisée_j1_Direction_du_vent_moyenne_sur_les_1500m	0%	-
météo_modélisée_j1_Vitesse_du_vent_moyenne_sur_les_1500m	0%	-
météo_modélisée_j1_Epaisseur_optique_totale_(méthode_2)	0%	-
météo_modélisée_j1_Epaisseur_optique_totale_(méthode_1)	0%	-

météo_modélisée_j1_Radiation d'onde longue sortante	0%	-
météo_modélisée_j1_Longueur_de_Monin_Obukhov	0%	-
météo_modélisée_j1_hauteur_de_couche_limite	0%	-
météo_modélisée_j1_Précipitations_totales	0%	-
météo_modélisée_j1_Colonne intégrée vapeur	0%	-
météo_modélisée_j1_Colonne intégrée glace	0%	-
météo_modélisée_j1_Colonne intégrée grésil	0%	-
météo_modélisée_j1_Colonne intégrée neige	0%	-
météo_modélisée_j1_Colonne intégrée d'eau	0%	-
météo_modélisée_j1_Epaisseur optique haute	0%	-
météo_modélisée_j1_Epaisseur optique moyenne	0%	-
météo_modélisée_j1_Epaisseur optique basse	0%	-
météo_modélisée_j1_Radiation onde longue vers le sol	0%	-
météo_modélisée_j1_Flux de chaleur latente	0%	-
météo_modélisée_j1_Flux de chaleur sensible	0%	-
météo_modélisée_j1_NA	0%	-
météo_modélisée_j1_Radiation onde courte vers le sol	0%	-
météo_modélisée_j1_direction du vent_à_10m	0%	-
météo_modélisée_j1_vitesse du vent_à_10m	0%	-
météo_modélisée_j1_humidité_à_2m	0%	-
météo_modélisée_j1_Température	0%	-
météo_modélisée_j0_DV_8	0%	-
météo_modélisée_j0_DV_7	0%	-
météo_modélisée_j0_DV_6	0%	-
météo_modélisée_j0_DV_5	0%	-
météo_modélisée_j0_DV_4	0%	-
météo_modélisée_j0_DV_3	0%	-
météo_modélisée_j0_DV_2	0%	-
météo_modélisée_j0_DV_1	0%	-
météo_modélisée_j0_Dvnul	0%	-
météo_modélisée_j0_Delta température entre le niveau 100m et le sol	0%	-
météo_modélisée_j0_Pression en surface	0%	-
météo_modélisée_j0_Humidité du sol	0%	-
météo_modélisée_j0_Direction du vent moyenne sur les 1500m	0%	-
météo_modélisée_j0_Vitesse du vent moyenne sur les 1500m	0%	-
météo_modélisée_j0_Epaisseur optique totale (méthode 2)	0%	-
météo_modélisée_j0_Epaisseur optique totale (méthode 1)	0%	-
météo_modélisée_j0_Radiation d'onde longue sortante	0%	-
météo_modélisée_j0_Longueur_de_Monin_Obukhov	0%	-
météo_modélisée_j0_hauteur_de_couche_limite	0%	-
météo_modélisée_j0_Précipitations_totales	0%	-
météo_modélisée_j0_Colonne intégrée vapeur	0%	-
météo_modélisée_j0_Colonne intégrée glace	0%	-
météo_modélisée_j0_Colonne intégrée grésil	0%	-
météo_modélisée_j0_Colonne intégrée neige	0%	-
météo_modélisée_j0_Colonne intégrée d'eau	0%	-

météo_modélisée_j0_Epaisseur optique haute	0%	-
météo_modélisée_j0_Epaisseur optique moyenne	0%	-
météo_modélisée_j0_Epaisseur optique basse	0%	-
météo_modélisée_j0_Radiation onde longue vers le sol	0%	-
météo_modélisée_j0_Flux de chaleur latente	0%	-
météo_modélisée_j0_Flux de chaleur sensible	0%	-
météo_modélisée_j0_NA	0%	-
météo_modélisée_j0_Radiation onde courte vers le sol	0%	-
meteo_reelle_DV_7	0%	-
meteo_reelle_DV_6	0%	-
meteo_reelle_DV_5	0%	-
meteo_reelle_DV_4	0%	-
meteo_reelle_DV_3	0%	-
meteo_reelle_Dvnul	0%	-
NEBULOSITE	0%	-
PRECIPITATION	0%	-

**Tableau 8 Résultats de la hiérarchisation des paramètres pour l'erreur associée au modèle Prev'air**

Que ce soit la météo réelle ou la météo modélisée, les paramètres jouant un rôle prépondérant sont les suivants :

- direction du vent ;
- vitesse du vent ;
- humidité ;
- température ;
- rayonnement global.

Par ailleurs, le rôle joué par le même paramètre météorologique est opposé quand il s'agit de la météo réelle que quand il s'agit de la météo modélisée. Par exemple : l'erreur augmente quand la température réelle augmente et quand la température modélisée diminue. Ou, encore, l'erreur augmente quand la vitesse du vent réelle augmente, et quand la vitesse du vent modélisée diminue.

Autrement dit l'erreur du modèle est étroitement liée à une erreur de prévision météo. Corriger le modèle dans les situations à risque reviendrait ainsi à corriger les modèles météo, ce qui est complexe.

Les résultats obtenus pour le modèle Esmeralda sont donnés dans le tableau ci-dessous.

<i>Paramètre</i>	<i>Coefficient d'influence</i>	<i>Sens de variation</i>
erreur modele 2	100%	+
MoyenneDemesure pm 10	81%	-
MaxDemesure pm 10	76%	-
mesure pm 10 à 10h00	71%	-
Données_veille_MoyenneDemesure pm 10	65%	-
Données_veille_MaxDemesure pm 10	63%	-
j0e_MoyenneDeTeneur_modelisee	58%	+
j0e_MaxDeTeneur_modelisee	56%	+
j0_MoyenneDeTeneur_modelisee	53%	-
j1e_MaxDeTeneur_modelisee	52%	+
j0_MaxDeTeneur_modelisee	51%	-
j1e_MoyenneDeTeneur_modelisee	50%	+
Données_veille_j0_MaxDeTeneur_modelisee	48%	-
météo_modélisée_j1_traite_Epaisseur optique totale (méthode 1)	48%	+
météo_modélisée_j1_traite_Radiation onde longue vers le sol	46%	+
météo_modélisée_j0_traite_Epaisseur optique totale (méthode 1)	46%	+
météo_modélisée_j0_traite_Radiation onde longue vers le sol	45%	+
j1_MaxDeTeneur_modelisee	44%	-
météo_modélisée_j1_traite_Epaisseur optique haute	44%	+
météo_modélisée_j1_traite_Radiation d'onde longue sortante	44%	+
j1_MoyenneDeTeneur_modelisee	43%	-
Données_veille_j0_MoyenneDeTeneur_modelisee	43%	-
météo_modélisée_j1_traite_Vitesse du vent moyenne sur les 1500m	43%	+
météo_modélisée_j0_traite_vitesse du vent à 10m	43%	+
meteo_reelle_traite_DV_5	42%	+
météo_modélisée_j1_traite_DV_6	42%	+
PRECIPITATION	42%	+
météo_modélisée_j0_traite_Vitesse du vent moyenne sur les 1500m	42%	+
météo_modélisée_j0_traite_DV_6	41%	+
météo_modélisée_j1_traite_vitesse du vent à 10m	40%	+
météo_modélisée_j1_traite_Flux de chaleur latente	40%	+
meteo_reelle_traite_DV_1	40%	-
météo_modélisée_j0_traite_Radiation d'onde longue sortante	40%	+
météo_modélisée_j0_traite_Epaisseur optique haute	40%	+
Données_veille_j1_MaxDeTeneur_modelisee	39%	-
météo_modélisée_j0_traite_Colonne intégrée glace	38%	+
meteo_reelle_traite_DV_6	38%	+
météo_modélisée_j1_traite_Colonne intégrée glace	38%	+
DIRECTION_VENT	38%	+
météo_modélisée_j0_traite_Flux de chaleur latente	37%	+

météo modélisée_j0_traite_Epaisseur optique basse	37%	+
météo modélisée_j1_traite_Epaisseur optique basse	36%	+
meteo_reelle_traite_DV_2	35%	-
Données_veille_j1_MoyenneDeTeneur_modelisee	34%	-
VISIBILITE	34%	+
TEMPERATURE	33%	+
météo modélisée_j1_traite_DV_5	32%	+
Données_veille_PRECIPITATION	32%	+
Données_veille_meteo_reelle_traite_DV_2	32%	-
météo modélisée_j0_traite_DV_5	31%	+
Données_veille_DIRECTION_VENT	31%	+
météo modélisée_j0_traite_Température	30%	+
Données_veille_meteo_reelle_traite_DV_1	30%	-
Données_veille_TEMPERATURE	29%	+
NEBULOSITE	29%	+
météo modélisée_j1_traite_Colonne intégrée d'eau	29%	+
météo modélisée_j1_traite_Colonne intégrée vapeur	28%	+
météo modélisée_j0_traite_Colonne intégrée d'eau	28%	+
Données_veille_VISIBILITE	27%	+
météo modélisée_j1_traite_Précipitations_totales	27%	+
météo modélisée_j1_traite_Température	27%	+
DUREE_INSOLATION	26%	-
PRESSION_NIVEAU_MER	26%	-
météo modélisée_j0_traite_Précipitations_totales	26%	+
Données_veille_PRESSION_NIVEAU_MER	25%	-
PRESSION_STATION	24%	-
météo modélisée_j1_traite_Colonne intégrée neige	23%	+
météo modélisée_j0_traite_Colonne intégrée neige	22%	+
météo modélisée_j0_traite_Colonne intégrée vapeur	21%	+
Données_veille_PRESSION_STATION	21%	-
Données_veille_NEBULOSITE	21%	+
météo modélisée_j1_traite_Colonne intégrée grésil	20%	+
météo modélisée_j1_traite_direction du vent_à_10m	19%	-
météo modélisée_j1_traite_Epaisseur optique moyenne	19%	+
météo modélisée_j0_traite_Colonne intégrée grésil	19%	+
météo modélisée_j0_traite_Epaisseur optique moyenne	18%	+
Données_veille_DUREE_INSOLATION	18%	-
meteo_reelle_traite_DV_7	18%	+
meteo_reelle_traite_DV_8	17%	-
meteo_reelle_traite_DV_4	17%	+
VITESSE_VENT	17%	+
HUMIDITE_RELATIVE	17%	+
météo modélisée_j0_traite_direction du vent_à_10m	16%	-
météo modélisée_j1_traite_Direction du vent_moyenne_sur_les_1500m	15%	+
Données_veille_HUMIDITE_RELATIVE	15%	+
CODE_TEMPS_OBSERVE	15%	+

météo modélisée_j0_traite_Direction du vent moyenne sur les 1500m	14%	+
Données veille_VITESSE_VENT	12%	+
météo modélisée_j0_traite_Longueur de Monin Obukhov	10%	-
Données veille_CODE TEMPS OBSERVE	10%	+
météo modélisée_j1_traite_hauteur de couche limite	9%	-
météo modélisée_j0_traite_NA	7%	-
météo modélisée_j0_traite_Epaisseur optique totale (méthode 2)	5%	+
météo modélisée_j1_traite_Longueur de Monin Obukhov	5%	-
météo modélisée_j1_traite_Epaisseur optique totale (méthode 2)	5%	+
météo modélisée_j1_traite_NA	4%	-
météo modélisée_j1_traite_Flux de chaleur sensible	4%	+
météo modélisée_j0_traite_hauteur de couche limite	3%	-
météo modélisée_j1_traite_humidité à 2m	3%	+
météo modélisée_j0_traite_humidité à 2m	2%	+
RAYONNEMENT GLOBAL	2%	-
météo modélisée_j0_traite_Flux de chaleur sensible	2%	+
Données veille_RAYONNEMENT GLOBAL	2%	-
Données veille_meteo_reelle_traite_Dvnul	0%	-
météo modélisée_j1_traite_DV_8	0%	-
météo modélisée_j1_traite_DV_7	0%	-
météo modélisée_j1_traite_DV_4	0%	-
météo modélisée_j1_traite_DV_3	0%	-
météo modélisée_j1_traite_DV_2	0%	-
météo modélisée_j1_traite_DV_1	0%	-
météo modélisée_j1_traite_Dvnul	0%	-
météo modélisée_j1_traite_Delta température entre le niveau 100m et le sol	0%	-
météo modélisée_j1_traite_Pression en surface	0%	-
météo modélisée_j1_traite_Humidité du sol	0%	-
météo modélisée_j1_traite_Radiation onde courte vers le sol	0%	-
météo modélisée_j0_traite_DV_8	0%	-
météo modélisée_j0_traite_DV_7	0%	-
météo modélisée_j0_traite_DV_4	0%	-
météo modélisée_j0_traite_DV_3	0%	-
météo modélisée_j0_traite_DV_2	0%	-
météo modélisée_j0_traite_DV_1	0%	-
météo modélisée_j0_traite_Dvnul	0%	-
météo modélisée_j0_traite_Delta température entre le niveau 100m et le sol	0%	-
météo modélisée_j0_traite_Pression en surface	0%	-
météo modélisée_j0_traite_Humidité du sol	0%	-
météo modélisée_j0_traite_Radiation onde courte vers le sol	0%	-
meteo_reelle_traite_DV_3	0%	-
meteo_reelle_traite_Dvnul	0%	-

**Tableau 9 Résultats de la hiérarchisation des paramètres pour l'erreur associée au modèle Esmeralda**

Les résultats obtenus pour le modèle Esmeralda sont différents de ceux obtenus avec le modèle Prev'air. Ceci est en accord avec le résultat prouvé dans le III où l'on a pu constater que l'erreur diffère en fonction du modèle pris en compte.

## V. Prise en compte de nouveaux indicateurs

Nous arrivons à la conclusion suivante : l'information météo ne suffit pas pour améliorer le modèle. La définition de nouveaux indicateurs paraît indispensable.

Nous avons calculé :

- l'erreur de la veille ;
- la teneur moyenne du polluant la veille ;
- la teneur du polluant à 10h00 du matin : cette information est en effet disponible avant de déclencher les alertes à 12h00 ;

En rajoutant ces nouveaux paramètres, nous effectuons une hiérarchisation des paramètres sur la teneur du polluant. Les résultats sont donnés dans le tableau ci-dessous.

Paramètre	Coefficient d'influence	Sens de variation
Teneur moyenne pm 10	100%	+
Teneur maximale pm 10	94%	+
mesure pm 10 à 10h00	90%	+
Données veille Teneur moyenne pm 10	77%	+
Données veille Teneur maximale pm 10	72%	+
j0_Moyenne Modèle	67%	+
erreur	65%	+
erreur veille	55%	+
Données veille j0_Max Modèle	48%	+
Données veille j0_Moyenne Modèle	46%	+
météo_modélisée_j1__traite_vitesse_du_vent_à_10m	45%	-
PRECIPITATION	44%	-
météo_modélisée_j1__traite_Colonne intégrée glace	44%	-
météo_modélisée_j0__traite_vitesse_du_vent_à_10m	44%	-
VISIBILITE	44%	-
météo_modélisée_j0__traite_Colonne intégrée glace	44%	-
météo_modélisée_j1__traite_Vitesse_du_vent_moyenne_sur_les_1500m	44%	-
météo_modélisée_j0__traite_Vitesse_du_vent_moyenne_sur_les_1500m	44%	-
météo_modélisée_j0__traite_Flux de chaleur latente	43%	-
meteo_reelle_traite_DV_1	43%	+
météo_modélisée_j0__traite_Radiation d'onde longue sortante	43%	-
météo_modélisée_j1__traite_Flux de chaleur latente	43%	-
météo_modélisée_j0__traite_Epaisseur optique haute	43%	-
météo_modélisée_j0__traite_Précipitations_totales	42%	-



DIRECTION_VENT	42%	-
météo_modélisée_j1__traite_Précipitations_totales	42%	-
météo_modélisée_j0__traite_Epaisseur optique totale (méthode 1)	42%	-
météo_modélisée_j0__traite_Radiation onde longue vers le sol	42%	-
météo_modélisée_j1__traite_Epaisseur optique totale (méthode 1)	41%	-
météo_modélisée_j1__traite_Radiation d'onde longue sortante	41%	-
météo_modélisée_j1__traite_Radiation onde longue vers le sol	41%	-
météo_modélisée_j1__traite_Epaisseur optique haute	41%	-
météo_modélisée_j1__traite_DV_6	40%	-
meteo_reelle_traite_DV_6	40%	-
météo_modélisée_j0__traite_DV_6	40%	-
meteo_reelle_traite_DV_2	38%	+
Données veille TEMPERATURE	36%	-
meteo_reelle_traite_DV_5	36%	-
TEMPERATURE	36%	-
PRESSION_STATION	35%	+
Données veille VISIBILITE	34%	-
météo_modélisée_j0__traite_Epaisseur optique basse	33%	-
Données veille j1_MoyenneDeTeneur_modelisee	31%	+
PRESSION_NIVEAU_MER	31%	+
Données veille j1_MaxDeTeneur_modelisee	31%	+
météo_modélisée_j1__traite_Epaisseur optique basse	30%	-
m	30%	-
Données veille PRESSION_STATION	27%	+
météo_modélisée_j0__traite_DV_5	27%	-
Données veille DIRECTION_VENT	27%	-
météo_modélisée_j1__traite_DV_5	27%	-
Données veille PRECIPITATION	26%	-
Données veille PRESSION_NIVEAU_MER	24%	+
météo_modélisée_j1__traite_Colonne intégrée neige	24%	-
Données veille meteo_reelle_traite_DV_1	24%	+
VITESSE_VENT	24%	-
météo_modélisée_j0__traite_Colonne intégrée d'eau	23%	-
Données veille meteo_reelle_traite_DV_2	23%	+
météo_modélisée_j0__traite_Colonne intégrée neige	23%	-
météo_modélisée_j1__traite_Colonne intégrée d'eau	22%	-
météo_modélisée_j0__traite_Epaisseur optique totale (méthode 2)	20%	-
météo_modélisée_j1__traite_Epaisseur optique totale (méthode 2)	19%	-
météo_modélisée_j0__traite_Température	19%	-
météo_modélisée_j1__traite_Flux de chaleur sensible	19%	-
DUREE_INSOLATION	18%	+
meteo_reelle_traite_DV_7	18%	-
météo_modélisée_j0__traite_humidité à 2m	17%	-
météo_modélisée_j1__traite_humidité à 2m	17%	-
météo_modélisée_j0__traite_Flux de chaleur sensible	17%	-
météo_modélisée_j1__traite_Epaisseur optique moyenne	17%	-

météo_modélisée_j1__traite_Colonne intégrée grésil	17%	-
météo_modélisée_j0__traite_Colonne intégrée grésil	17%	-
météo_modélisée_j1__traite_Température	16%	-
météo_modélisée_j0__traite_Epaisseur optique moyenne	16%	-
meteo_reelle_traite_DV_8	16%	+
mé- téo_modélisée_j1__traite_Direction du vent moyenne sur les 1500m	14%	-
mé- téo_modélisée_j0__traite_Direction du vent moyenne sur les 1500m	14%	-
meteo_reelle_traite_DV_4	14%	-
Données veille RAYONNEMENT_GLOBAL	13%	-
météo_modélisée_j1__traite_hauteur de couche limite	12%	+
a	12%	-
météo_modélisée_j0__traite_hauteur de couche limite	11%	+
RAYONNEMENT_GLOBAL	8%	-
Données veille VITESSE_VENT	8%	-
veille	8%	-
HUMIDITE_RELATIVE	6%	-
Données veille DUREE_INSOLATION	4%	+
météo_modélisée_j1__traite_Longueur de Monin Obukhov	4%	+
météo_modélisée_j0__traite_Longueur de Monin Obukhov	3%	+
météo_modélisée_j0__traite direction du vent _à_10m	3%	+
météo_modélisée_j1__traite direction du vent _à_10m	3%	+
météo_modélisée_j1__traite_NA	1%	-
météo_modélisée_j0__traite_NA	1%	+
Données veille HUMIDITE_RELATIVE	0%	-
Données veille meteo_reelle_traite_Dvnul	0%	-
Données veille CODE_TEMPS_OBSERVE	0%	-
Données veille NEBULOSITE	0%	-
météo_modélisée_j1__traite_DV_8	0%	-
météo_modélisée_j1__traite_DV_7	0%	-
météo_modélisée_j1__traite_DV_4	0%	-
météo_modélisée_j1__traite_DV_3	0%	-
météo_modélisée_j1__traite_DV_2	0%	-
météo_modélisée_j1__traite_DV_1	0%	-
météo_modélisée_j1__traite_Dvnul	0%	-
météo_modélisée_j1__traite_Delta température entre le niveau 100m et le sol	0%	-
météo_modélisée_j1__traite_Pression en surface	0%	-
météo_modélisée_j1__traite_Humidité du sol	0%	-
météo_modélisée_j1__traite_Colonne intégrée vapeur	0%	-
météo_modélisée_j1__traite_Radiation onde courte vers le sol	0%	-
météo_modélisée_j0__traite_DV_8	0%	-
météo_modélisée_j0__traite_DV_7	0%	-
météo_modélisée_j0__traite_DV_4	0%	-
météo_modélisée_j0__traite_DV_3	0%	-
météo_modélisée_j0__traite_DV_2	0%	-
météo_modélisée_j0__traite_DV_1	0%	-

météo_modélisée_j0__traite_Dvnul	0%	-
météo_modélisée_j0__traite_Delta_température_entre_le_niveau_100m_et_le_sol	0%	-
météo_modélisée_j0__traite_Pression_en_surface	0%	-
météo_modélisée_j0__traite_Humidité_du_sol	0%	-
météo_modélisée_j0__traite_Colonne_intégrée_vapeur	0%	-
météo_modélisée_j0__traite_Radiation_onde_courte_vers_le_sol	0%	-
meteo_reelle_traite_DV_3	0%	-
meteo_reelle_traite_Dvnul	0%	-
NEBULOSITE	0%	-

**Tableau 10 Résultats de la hiérarchisation des paramètres pour la teneur moyenne pm10 en prenant en compte les nouveaux indicateurs**

Les paramètres qui influent le plus sur la teneur moyenne journalière en PM10 sont (dans l'ordre décroissant) :

- teneur maximale pm 10 ;
- mesure pm 10 à 10h00 ;
- données veille Teneur moyenne pm 10 ;
- données veille Teneur maximale pm 10 ;
- j0\_Moyenne Modèle.

Autrement dit, l'indicateur qui nous renseigne le mieux sur la teneur moyenne journalière en polluant est la teneur à 10h00. Nous prenons en compte cet indicateur pour la définition de la méthode probabiliste de dépassement des seuils réglementaires.

## VI. Détection des dépassements de seuil en exploitant les mesures réelles à 10H00

### A. Risque de dépassement en fonction de la teneur en PM10 à 10h00

Les résultats précédents illustrent que souvent quand les seuils de pollution sont dépassés, les modèles météo ne sont pas pertinents. Or le modèle s'appuie essentiellement sur les modèles météo. Comme nous disposons également des mesures du polluant à 10h00 nous établissons une règle simple permettant de déterminer un risque de dépassement du seuil en connaissant la teneur du polluant à 10h00.

En prenant en compte l'ensemble des données (23000 : il s'agit des mesures effectuées pour 13 stations sur la période 2008-2015) nous effectuons l'analyse suivante :

Teneur à 10 h00 supérieure à :	Nombre de fois où ce seuil a été dépassé	Nombre de fois où ce seuil a été dépassé et le dépassement du seuil d'alerte a été atteint	Probabilité de dépassement du seuil d'alerte
40 $\mu\text{g}/\text{m}^3$	1568	496	32 %
50 $\mu\text{g}/\text{m}^3$	711	339	48 %
60 $\mu\text{g}/\text{m}^3$	301	213	71 %
70 $\mu\text{g}/\text{m}^3$	313	261	83 %

Tableau 11 Risque de dépassement en fonction de la teneur en PM10 à 10h00

A partir de ce tableau très simple nous pouvons connaître pour une station donnée le risque de dépassement de seuil réglementaire en connaissant la teneur en PM10 à 10h00 pour cette même station.

## B. Amélioration de la prévision en prenant en compte le mois

### 1. Profil horaire journalier

Si nous regardons le profil horaire journalier par mois, nous nous rendons compte que 10h00 ne correspond pas systématiquement à l'heure la plus polluée, mais pas non plus à l'heure où la teneur en polluant est la plus faible, ceci est représenté dans le graphique ci-dessous.

Par ailleurs le profil horaire journalier varie en fonction du mois : autrement dit pour une même teneur de pm10 à 10h00 le risque de dépassement réglementaire diffère que l'on soit en janvier ou en mars.

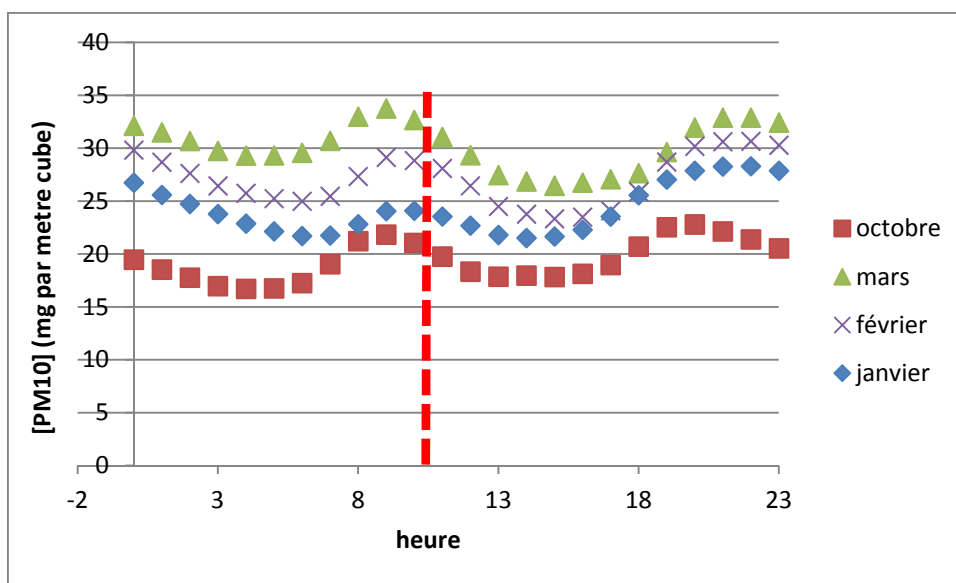


Figure 8 Teneur horaire moyenne en PM10 en fonction du mois

## 2. Probabilité de dépassement en fonction du mois

Par conséquent, nous pouvons améliorer la méthode de détection de dépassement des seuils réglementaires en prenant en compte un deuxième critère : le mois. Ceci nous amène aux résultats suivants pour les seuils de 50 et 60  $\mu\text{g}/\text{m}^3$ .

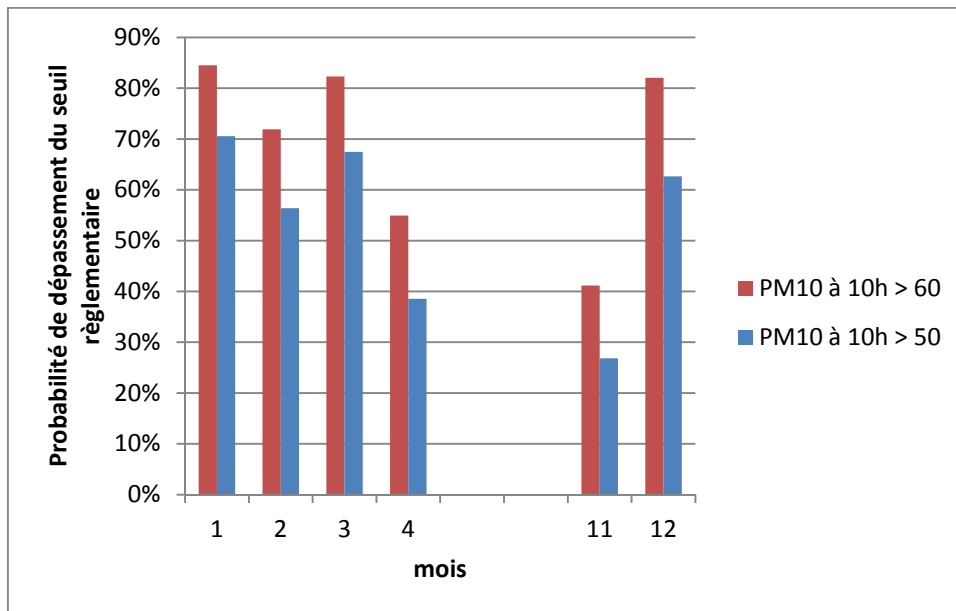


Figure 9 Probabilité de dépasser le seuil réglementaire en fonction du mois et de la teneur en PM10 à 10h00

En janvier, si à 10h00 la teneur en PM10 dépasse 60  $\mu\text{g}/\text{m}^3$ , le risque de dépassement réglementaire est de 85 %. Dans le même cas de figure au mois d'octobre le risque de dépassement égal à 40 %.