VIGIPOL

Pollution à l'ozone : mise en place d'un outil de vigilance par application des techniques des forêts aléatoires

VIGIPOL

Random forests for anticipation of pollution episodes

Pierre BEAL**, Emmanuel BUISSON**, Victor BRUYERE**,
Benjamin CHABANON**, Gwénaëlle HOURDIN*, BouAlem MESBAH*, David POULET**

Résumé

La région PACA est fortement exposée aux épisodes de pollution à l'ozone, notamment sur le pourtour de l'étang de Berre. Pour essayer de réduire la fréquence et l'intensité de ces épisodes, un ensemble de mesures a été mis en place. Parmi celles-ci, une action consiste à intervenir en planifiant les émissions des espèces primaires à l'origine de cette pollution à partir des résultats de prévisions des niveaux des polluants ciblés sur la région. Les résultats de cette action sont d'autant plus efficaces que le terme de la prévision donne aux industriels un temps suffisant de réaction.

VIGIPOL s'inscrit totalement dans cette logique de mise en vigilance « le plus tôt possible » des industriels. Il répond aux besoins des adhérents d'AIRFOBEP qui sont acteurs de la réduction des émissions par planification de la production de leur site.

L'objectif de VIGIPOL est de fournir une prévision des niveaux de pollution à l'ozone pour le jour même et les trois jours qui suivent. L'outil est basé sur la méthode statistique des forêts aléatoires pour la réalisation des prévisions. Il est actuellement opérationnel sur la zone depuis plus d'un an.

Mots clés

Ozone. Anticipation. Forêts aléatoires.

Abstract

The *"région de l'étang de Berre"* is a very important industrial area in the south of France. Several activities are realized there in huge industrial plants; namely, oil refining and chemistry, metal industry… etc. This is why, one frequently observes very important levels of air pollution in this area. Actually, the highest sulfur dioxide and ozone concentrations in France are measured there. AIRFOBEP air quality monitoring sensors often measure ozone concentrations as high as 240 µg/m^3.

To ovoid ozone pollution, it is necessary to reduce emissions of its precursors. In the *"région de l'étang de Berre",* ozone precursors are mainly produced by the industrial activity. Like for example emissions of VOC's from several industrial processes.

Reducing ozone precursors then means regulating industrial activities when whether situation is favorable to ozone production. But, in the case of heavy industries, this kind of regulation is a long term planning issue.

The aim of the project presented in this paper is to produce a useful information to perform such a planning. This is done by using a statistical modeling tool called VIGIPOL. Modeling in VIGIPOL is based on the well known random forests technique. This technique, makes VIGIPOL able to give a daily prediction of the ozone pollution risk level for the four coming days.

VIGIPOL operated during the last two years and shows good prediction performances. It became one of the main information sources for the area industries managers. They use VIGIPOL data when planning some specific activities during big ozone pollution period.

AIRFOBEP is the air quality monitoring network in the area of the *"région de l'étang de Berre".*

Keywords

Ozone. Anticipation. Random forests.

** AIRFOBEP – route de la Vierge – 13500 Martigues – France – E-mail : airfobep@airfobep.org*

*** NUMTECH – 6, allée Alan Turing – BP 30242 – parc technologique de La Pardieu – 63175 AUBIÈRE Cedex*
E-mail : numtech@numtech.fr

1. Introduction

Depuis plusieurs décennies, la pollution photochimique est un sujet de préoccupations, et ce à plus d'un titre.

En effet, ces polluants – et notamment l'ozone qui en est le traceur – sont des oxydants puissants. De fortes concentrations en ozone peuvent irriter les voies respiratoires, provoquer des difficultés pour cette fonction vitale, mais aussi des dégâts aux poumons après seulement quelques heures d'exposition. Selon l'Agence européenne pour l'environnement, jusqu'à 30 % de la population urbaine est exposée à des concentrations supérieures aux niveaux limites fixés par l'Union européenne, provoquant 30 millions de jours sous médications respiratoires et plus de 20 000 décès prématurés par an. Ces impacts sanitaires auxquels s'ajoutent des effets environnementaux ont des conséquences économiques encore difficiles à chiffrer.

Par ailleurs, des séries de mesure réalisées depuis 1874 dans l'hémisphère Nord montrent une multiplication par cinq des niveaux moyens d'ozone. En effet, l'essor de l'activité industrielle depuis la fin du XIXe siècle est associé à une augmentation des émissions d'oxydes d'azote (NO_x) et de composés organiques volatiles (COV) qui seraient responsables de cet accroissement des concentrations d'ozone troposphérique.

Les zones les plus exposées à cette pollution sont celles qui regroupent l'ensemble des facteurs qui contrôlent la production photochimique de l'ozone. La région Provence-Alpes-Côte d'Azur figure au nombre de celles-ci. En effet, il s'agit d'une zone urbaine fortement industrialisée où les émissions de précurseurs sont importantes et où les conditions anticycloniques, caractérisées par un ensoleillement fort et une dispersion des polluants faible, sont fréquentes.

Dans ce cadre, différentes études ont été menées ces dernières années sur la zone géographique couverte par AIRFOBEP. Elles ont permis, d'abord, d'identifier plus finement les précurseurs contrôlant la production d'ozone ainsi que les principales sources de ces précurseurs puis de proposer un outil adapté à la prévision de ces pics de pollution.

2. La pollution à l'ozone en région PACA

2.1. Une pollution par l'ozone importante et des spécificités locales propices

En France, la région PACA, et plus particulièrement la zone de l'étang de Berre, sont les plus touchées par la pollution à l'ozone, que ce soit en termes d'intensité ou de fréquence des épisodes. Par rapport aux autres régions de France, les concentrations relevées y sont nettement supérieures et environ trois fois plus d'épisodes de pollution y sont observés par an.

Sur ces dernières années, les concentrations mesurées sur les stations du pourtour de l'étang de Berre ont régulièrement été les plus élevées de la région PACA et de France :

- 417 µg/m³/h à Sausset-les-Pins (2003) ;
- 280 µg/m³/h à Rognac (2004) ;
- 313 µg/m³/h à Rognac (2005) ;
- 327 µg/m³/h à Vitrolles (2006) ;
- 308 µg/m³/h à Rognac (2007).

Ces valeurs dépassent de très loin le premier seuil réglementaire.

En effet, en cas de pics de pollution, la loi sur l'air et l'utilisation rationnelle de l'énergie prévoit l'information des populations selon deux seuils :

- le seuil d'information-recommandation (180 µg/m³/h) qui concerne les populations les plus sensibles (femmes enceintes, nourrissons, jeunes enfants, personnes âgées, fumeurs ou personnes souffrants de pathologies respiratoires) ;
- le seuil d'alerte pour l'ensemble de la population (240 µg/m³/3 h, 300 µg/m³/3 h, 360 µg/m³/h).

Ainsi, en 2003, le dernier seuil d'alerte a été dépassé. Il est préconisé en cas de dépassement d'un de ces seuils d'éviter les compétitions sportives, les activités physiques intenses, voire même les activités à l'extérieur dans certains cas. Sur le territoire de compétence d'AIRFOBEP, ce sont en moyenne 30 procédures d'information de la population qui sont déclenchées par an. En parallèle, un certain nombre de mesures de réduction des émissions sont mises en place.

De nombreuses spécificités locales contribuent à faire de cette région une zone propice à la pollution photochimique. On peut citer trois des phénomènes locaux importants :

- les phénomènes météorologiques locaux tels que les vents faibles ou l'alternance des brises thermiques (brise de terre/brise de mer et brises d'étang), qui contribuent à des recirculations particulières des masses d'air polluées ;
- l'existence de reliefs particuliers (falaise de Vitrolles, chaîne de la Fare et de La Nerthe), qui favorisent l'accumulation de polluants (Figure 1) ;
- les fortes émissions de précurseurs de l'ozone, dues au trafic routier ou aux activités industrielles.

2.2. La réduction de la pollution par l'ozone liée à la diminution des émissions de COV

Différentes mesures de réductions des émissions de précurseurs ont été mises en place par les pouvoirs publics, que ce soit à court ou moyen terme.

Les premières s'attaquent aux émissions de précurseurs de l'ozone de façon pérenne en tenant compte de la nature transfrontalière du phénomène. Ainsi, la directive européenne sur les plafonds d'émissions reprise par le gouvernement français

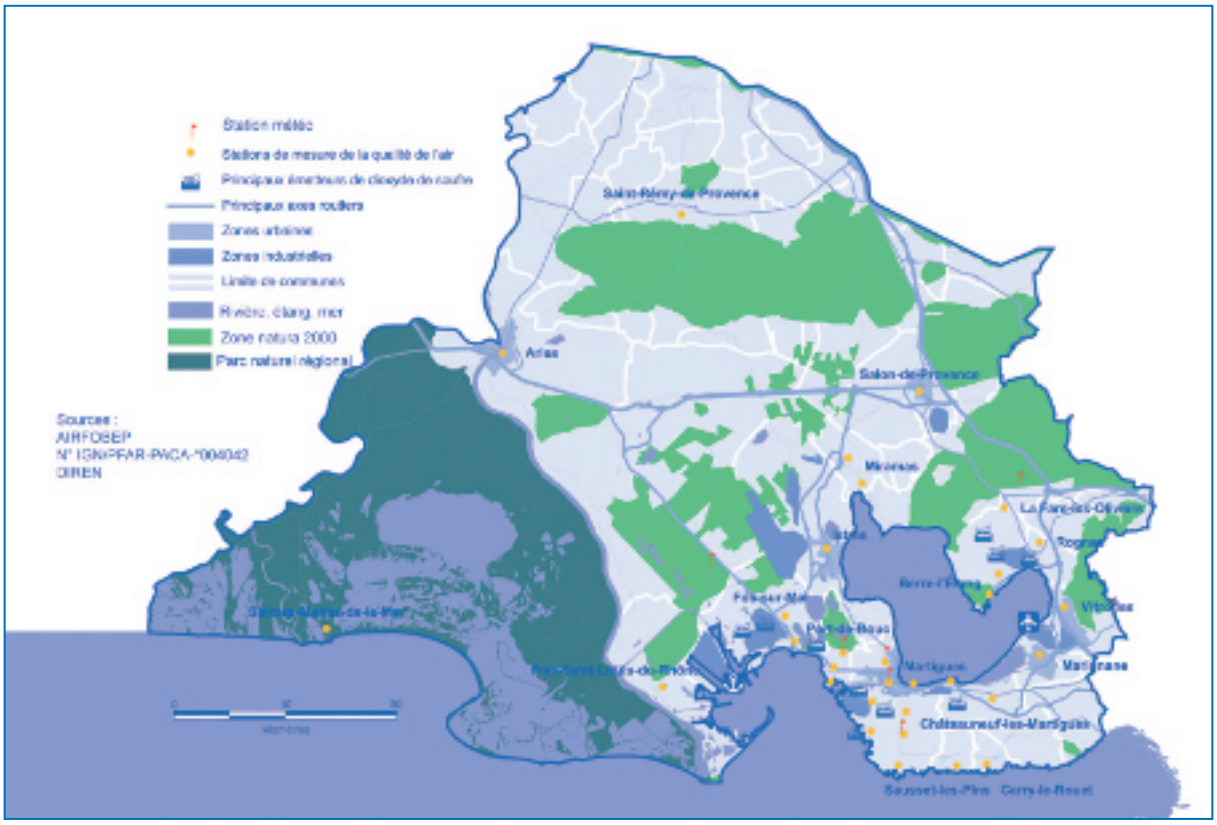


Figure 1.
Territoire de compétence d'AIRFOBEP : l'étang de Berre et l'ouest des Bouches-du-Rhône.
AIRFOBEP's area of intervention: Berre pond and the West of Bouches-du-Rhône.

prévoit une diminution de 40 % des émissions de précurseurs de l'ozone d'ici 2010 [1]. Concernant les émissions industrielles, l'action est engagée depuis plusieurs années par les DRIRE. C'est ainsi que les émissions de quelque 250 installations émettant plus de 250 tonnes annuelles de COV ont déjà été réduites de 20 % par rapport aux émissions initiales.

Dans les Bouches-du-Rhône, des mesures de réduction ponctuelles vont au-delà de ce que prévoit la réglementation et ce, dès le dépassement ou la prévision de dépassement du seuil d'information-recommandation. Il s'agit notamment de la réduction de la vitesse de 30 km/h sur les grands axes routiers ou du report de certaines activités industrielles, voire de l'instauration d'une circulation automobile alternée ou de l'arrêt de certaines activités industrielles en cas de dépassement du seuil d'alerte. Ces mesures d'urgence de court terme sont réglementaires.

Sur le territoire de compétence d'AIRFOBEP, en complément, il est proposé aux acteurs prépondérants des moyens d'anticipation de ces mesures, reposant sur une prévision statistique des épisodes.

Le développement de cet outil s'appuie sur des études précédentes qui ont permis d'identifier plus spécifiquement les précurseurs contrôlant la production d'ozone afin de proposer des actions de réduction d'émissions plus ciblées. Une cartographie statistique des régimes chimiques de production d'ozone sur les Bouches-du-Rhône a montré que le régime sur le pourtour de l'Étang de Berre était limité par les COV ou saturé en NO_x [2], grâce à l'utilisation de diagramme ISOPLETH [3]. Ce type de diagramme repose sur une évaluation de la concentration en ozone dans un système d'axes COV/NO_x, pour une zone déterminée. Il permet également d'évaluer la sensibilité de la production de l'ozone à la variation d'émissions et/ou des concentrations des précurseurs, toujours sur une zone déterminée. Dans la région de l'étang de Berre, une réduction des émissions de COV permettrait une diminution des concentrations relevées en ozone, notamment à proximité de ces sources d'émissions.

C'est d'abord sur ce levier de réduction des émissions des COV qu'intervient VIGIPOL. VIGIPOL est un outil de prévision des pics de pollution à l'ozone à J + 3 (c'est-à-dire à 72 heures) qui peut être utilisé selon deux logiques : une anticipation volontaire et/ou une planification des activités émettrices de ces composés. Il est à l'usage des principaux acteurs de la réduction des émissions de ces précurseurs.

Cependant, il est évident que ces mesures gagneraient en ampleur et en efficacité, notamment loin des sources d'émissions, avec des réductions d'émis-

sions pour les deux secteurs principaux que sont les grandes sources ponctuelles et le trafic. Il a en effet été montré qu'une telle combinaison d'actions permettait de diminuer plus efficacement le nombre de dépassements du seuil de 240 µg/m³ sur l'ensemble des Bouches-du-Rhône [4].

2.3. L'industrie et l'énergie : les deux principaux émetteurs de COV

Les activités industrielles et énergétiques sont responsables à 56 % des émissions de COV dans la région PACA [5]. Elles sont concentrées sur la zone de l'étang de Berre et du golfe de Fos-sur-Mer. Cette zone, qui représente 10 % du territoire régional, accueille le principal pôle industriel de la région, avec 85 % des sites SEVESO de Provence-Alpes-Côte d'Azur. De nombreuses activités y sont représentées, dont notamment le raffinage, la chimie et la pétrochimie, la production d'énergie, les produits minéraux, l'incinération de déchets industriels, la fabrication de papiers, de cartons, de peintures et de vernis, le stockage d'hydrocarbures, de produits chimiques et de gaz.

Ainsi, le secteur industriel est un acteur-clé de la réduction des émissions de COV et donc de la réduction des épisodes de pollution à l'ozone de la région de l'étang de Berre. Pour cela, il est indispensable de fournir aux industriels les outils opérationnels pour qu'ils puissent prendre, le plus tôt possible, les mesures adéquates. C'est pourquoi VIGIPOL fournit une prévision à J + 3 des concentrations en ozone.

3. Un outil de vigilance : VIGIPOL

3.1. L'anticipation, un enjeu majeur

Lors d'épisodes de pic de pollution à l'ozone, le préfet a à sa disposition des mesures d'urgence, qu'il peut décider de mettre en place. Celles-ci portent sur une réduction des émissions liées au trafic routier et aux activités industrielles. Elles peuvent être déclenchées suivant deux conditions de pollution : pollution constatée ou pollution prévue. Cette pollution, constatée ou prévue pour le jour J, est toujours associée à une prévision de pointe pour le jour J + 1. Elles sont donc déclenchées le jour J et sont applicables pour le jour J + 1 de 6 h à 21 h. Les prévisions utilisées pour

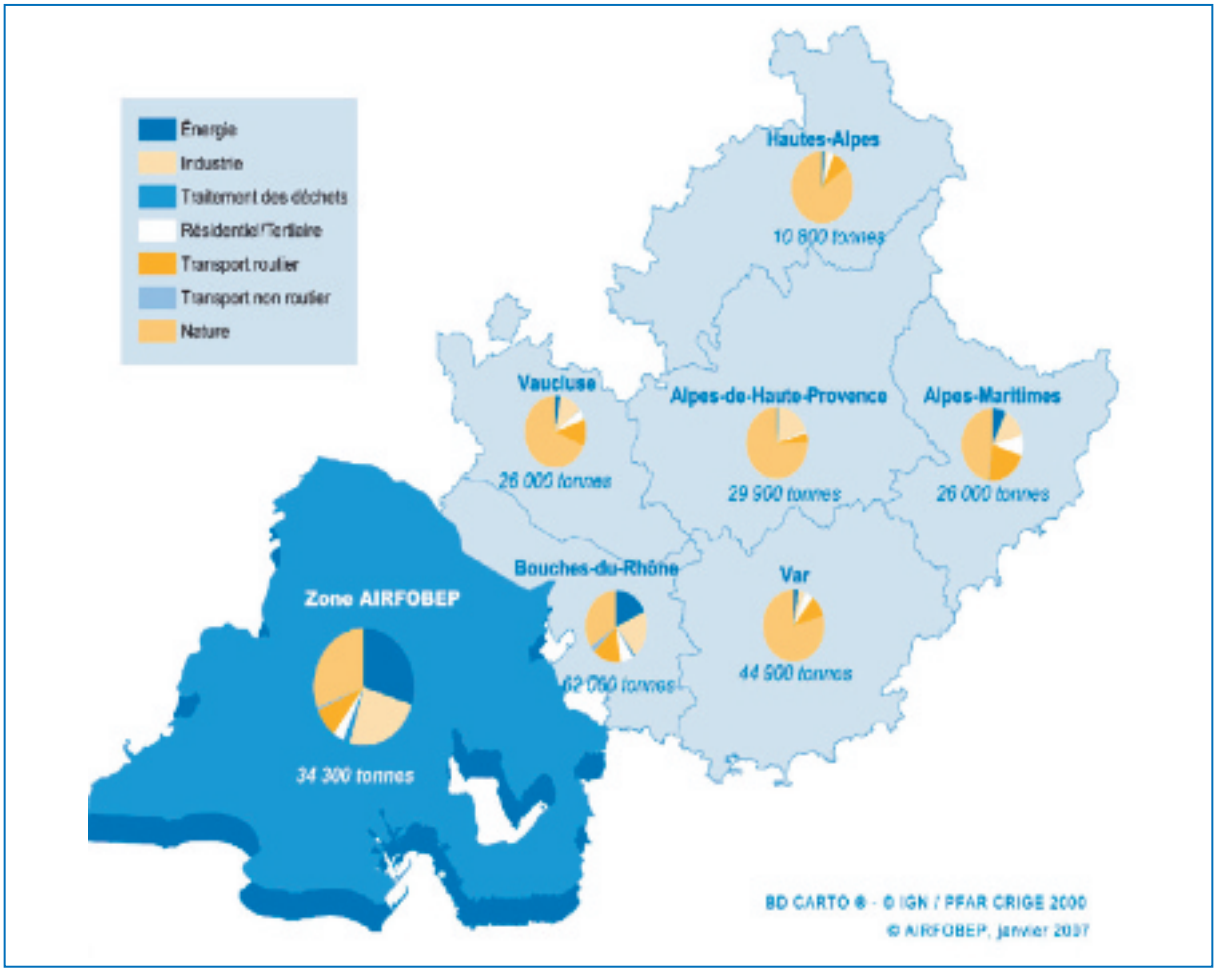


Figure 2.
Répartition des émissions de COV en région Provence-Alpes-Côte d'Azur.
VOC's emissions distribution in the region of Provence-Alpes-Côte d'Azur.

ces mesures d'urgence sont issues de la plate-forme inter-régionale AIRES* alimentée par la plate-forme nationale Prév'Air**. Comme tous les modèles, ce dernier a des limites [6, 7]. Les résultats ont été bien améliorés grâce à un processus d'adaptation statistique, notamment sur le territoire d'Airfobep [8].

Ces mesures sont également mises en œuvre dans un des départements limitrophes des Bouches-du-Rhône lorsque les conditions sont réunies pour qu'il y ait transfert de masses d'air vers l'un des départements voisins (Var, Vaucluse, Gard et Alpes-de-Haute-Provence). Ce processus permet d'anticiper le déplacement de masses d'air polluées dans les départements limitrophes et de réduire les émissions de polluants localement.

Les mesures imposées aux industriels, pour obtenir ces réductions d'émissions locales, sont graduées en fonction du seuil de pollution atteint. Quatre niveaux ont été définis qui vont du report des activités émettrices et maintien de la stabilité des procédés jusqu'à l'arrêt progressif des installations.

Pour le secteur industriel, l'anticipation des mesures de réduction est un enjeu majeur pour deux principales raisons. D'une part, une bonne anticipation permet de réduire l'intensité des pics, voire de les éviter, et d'autre part, une mise en vigilance plus rapide permet aux industriels d'anticiper plus tôt tout changement imposé aux activités émettrices. L'impact économique n'en est que meilleur.

3.2. Les objectifs de VIGIPOL

L'anticipation, qui permet de conduire à une planification des émissions industrielles, est donc un des moyens d'améliorer la qualité de l'air. Cette anticipation est d'autant plus efficace qu'elle se produit très en amont de l'épisode de pollution envisagé. L'outil VIGIPOL s'inscrit dans cette démarche de prise en compte, le plus tôt possible, d'un évènement de pollution intense.

Cet outil a pour objectif de fournir une prévision des niveaux de pollution à l'ozone pour le jour courant et les trois jours suivants. Il permet donc aux adhérents industriels d'AIRFOBEP, qui figurent au premier rang des acteurs d'une potentielle planification de leurs émissions de polluants primaires, d'anticiper les mesures d'urgence prises par la préfecture.

Les utilisateurs de VIGIPOL peuvent consulter en permanence, sur un site Internet, les résultats présentés sous la forme d'un tableau de bord synthétique des niveaux de pollution à l'ozone prévus pour les trois jours à venir.

Il est par contre à noter que, même si les industriels peuvent utiliser les résultats de VIGIPOL dans le cadre de la planification de leurs émissions, celui-ci ne s'inscrit pas dans le cadre réglementaire des procédures préfectorales. Son rôle se limite à la mise en vigilance.

4. Méthodologie

4.1. Forêts aléatoires

VIGIPOL est basé sur la méthode des forêts aléatoires [9]. Une forêt aléatoire est constituée d'un ensemble d'arbres simples de prévision, chacun étant capable de produire une réponse lorsqu'on lui présente un sous-ensemble de prédicteurs. Chaque arbre se présentant sous la forme d'un ensemble de tests hiérarchisés qui sont généralement destinés à établir une classification d'un ensemble d'individus (au sens statistique du terme).

En utilisant les ensembles d'arbres on obtient une amélioration significative de la prévision (c'est-à-dire une meilleure tendance à prévoir sur les nouvelles données), par rapport aux techniques classiques (CART, CHART…). La réponse de chaque arbre dépend du sous-ensemble de prédicteurs choisis indépendamment (avec remplacement) et avec la même distribution pour tous les arbres de la forêt.

Les forêts aléatoires peuvent s'appliquer sur des données à la fois nominales et continues.

Cette technique est idéale pour les problématiques où le nombre de catégories pour les variables est trop nombreux pour permettre l'utilisation de techniques classiques comme les arbres de décision, la régression logit…

De plus, sa faculté de pouvoir faire de la prévision sur un sous-ensemble aléatoire de prédicteurs apporte une solution idéale pour résoudre les problématiques où le nombre de variables à étudier dépasse un certain seuil de compréhension.

4.2. Arbre de décision dans le cas de VIGIPOL

Dans l'objectif de la prévision des épisodes d'ozone, il s'agit de déterminer si une journée donnée est à classer parmi celles susceptibles de voir survenir un dépassement du seuil considéré pour l'ozone. Pour ce faire, une série de tests portant sur les différents paramètres pris en compte doit être réalisée dans un ordre précis.

Un test à réaliser peut être par exemple : la température prévue pour cette journée est-elle supérieure à 30° ? La validité ou non de ce test induira le test suivant portant sur une autre variable ou une autre valeur de comparaison. Le processus se poursuit de proche en proche jusqu'à aboutir à l'une des extrémités de l'arbre de décision qui donnera finalement la probabilité d'occurrence du pic.

** www.aires-mediterranee.org*

*** www.prevair.org/fr/index.php*

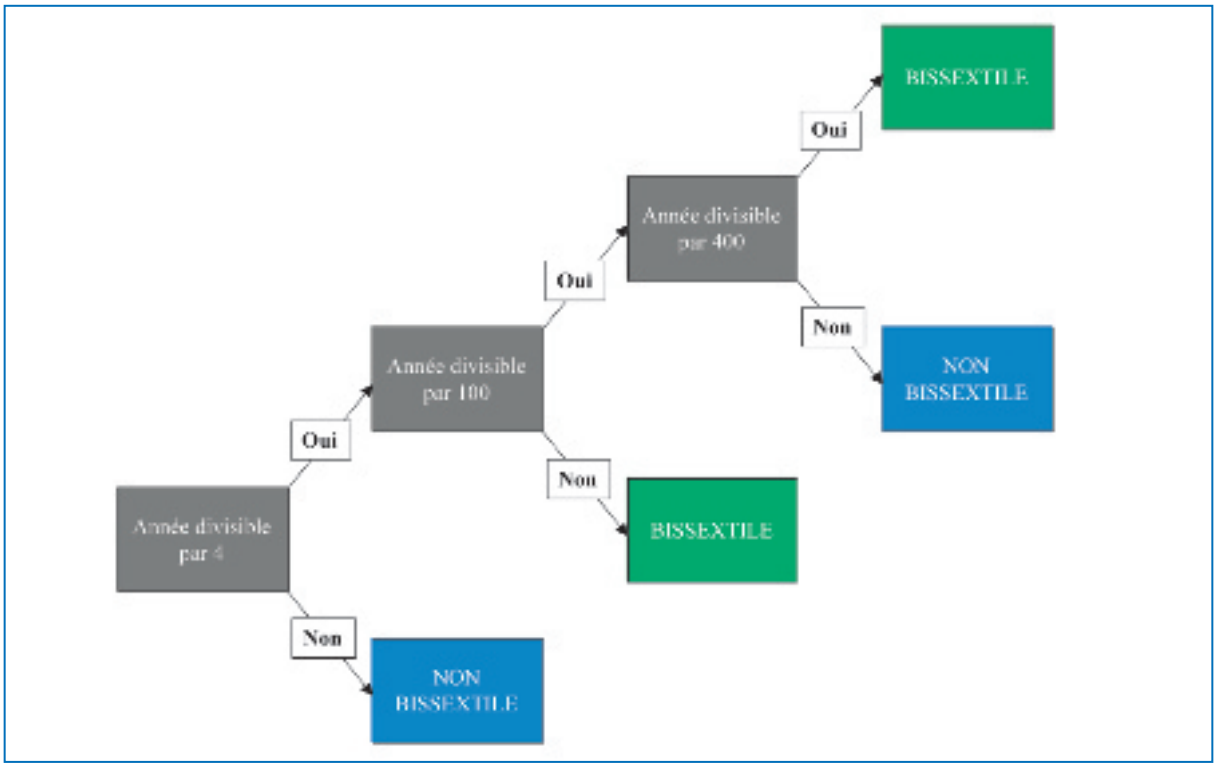


Figure 3.
Exemple d'arbre de décision « Une année donnée est-elle bissextile ? ».
Example of a decision tree « Is a given year a leap year ? ».

La Figure 3 donne un exemple d'arbre de décision.

De nombreuses méthodes permettent la construction d'un arbre de décision. Elles utilisent toutes un historique le plus complet possible, regroupant d'une part les valeurs des variables prédictives (c'est-à-dire celles qui interviendront dans les tests que contiendra l'arbre finalement obtenu) et d'autre part les valeurs de la variable à prédire (existence ou non d'un pic dans notre cas).

Ensuite, un ensemble de validation, composé des mêmes variables, est généralement utilisé pour éviter le surapprentissage (c'est-à-dire la détérioration du potentiel de prédiction de l'arbre induit par une complexification excessive et artificielle de l'arbre).

À partir de ces données, la méthode de construction procède de la façon suivante :

- évaluation du caractère discriminant (c'est-à-dire propre à distinguer les différentes modalités de la variable à prédire) d'un grand nombre de tests possible (un test étant composé du choix d'une variable prédictive et d'un niveau de comparaison). Cette estimation se fait à l'aide du calcul d'une quantité dépendante de la méthode de construction ;
- à partir du test donnant le meilleur résultat, l'échantillon de départ est subdivisé en deux sous-ensembles par application directe de ce même test ;
- la même procédure (recherche d'un test optimal) est alors appliquée à chacun des ensembles nouvellement créés ;
- le processus se poursuit jusqu'à vérification d'un critère d'arrêt.

La phase de construction est généralement suivie d'une phase d'élagage : elle consiste à détruire les parties de l'arbre dont l'existence n'est due qu'au hasard, c'est-à-dire à la présence d'une relation artificielle au sein de l'historique entre une variable prédictive et la variable à prédire. On utilise pour ce faire l'ensemble de validation qui permet de vérifier la pertinence de chacun des tests contenus dans l'arbre.

La méthode appliquée dans VIGIPOL suit le cheminement présenté ci-dessus et utilise, comme critère d'évaluation du pouvoir discriminant d'un test, l'indice de Gini qui évalue la pureté de chaque partition [5].

La façon d'évaluer ce pouvoir est réalisée lorsqu'un nouveau test est introduit dans l'arbre, en commençant par évaluer la pureté de chaque sous-ensemble nouvellement créé par la relation suivante :

$$G = 1 - \sum_{k=1}^{K} p(k)^2$$

où $p(k)$ est la proportion d'individus appartenant à la classe k dans le sous-ensemble considéré. K est le nombre de classes différentes (dans notre cas il existe deux classes : existence d'un pic durant la journée considérée ou non).

On remarquera que cette quantité G vaut 0 si le sous-ensemble est pur, c'est-à-dire si l'un des $p(k)$ est

égal à 1. Par ailleurs, G est maximal si toutes les classes sont également représentées dans le sous-ensemble, c'est-à-dire si $p(k) = 1/K$. Un test sera donc d'autant plus discriminant qu'il conduit à une valeur de G faible.

Pour estimer la pureté de la totalité des deux sous-ensembles créés par le test, on réalise une moyenne des valeurs de G calculées pour chaque subdivision pondérée par leur effectif respectif :

$$G_{total} = \sum_{n=1}^{N} E(n) \cdot G(n),$$

avec $G(n)$ l'indice de Gini estimé pour le sous-ensemble n. $E(n)$ est son effectif et N est le nombre de sous-ensembles (dans notre cas $N = 2$ car tous les tests utilisés sont binaires, c'est-à-dire qu'ils n'admettent comme réponse que deux possibilités, vrai ou faux). Le test finalement sélectionné sera celui conduisant à la valeur minimale de G_{total}.

4.3. Les forêts aléatoires dans le cas de VIGIPOL

La méthode des forêts aléatoires est fondée sur l'utilisation simultanée d'un grand nombre de ces arbres de décision. Chaque arbre est créé par l'une des méthodes classiques déjà évoquées précédemment. Toutefois, il n'y a pas de phase d'élagage. Dans le cadre de cette méthode, le taux d'erreur de la forêt est en effet dépendant de la qualité de chaque arbre, pris indépendamment, mais aussi de l'absence de corrélation entre ces derniers. La phase d'élagage, en plus d'exiger un temps de calcul non négligeable, conduit à améliorer le premier de ces deux points mais aussi à détériorer le second. Aussi, généralement, ce type de méthode ne comporte donc pas de phase d'élagage. Les arbres conservent leur extension maximale. La croissance d'un arbre n'est arrêtée que lorsque sa subdivision conduit à un nœud comportant cinq individus ou moins.

Par ailleurs, les forêts aléatoires introduisent deux originalités d'importance :

- L'historique sur lequel est construit chaque arbre, ainsi que l'ensemble de validation destiné à l'élagage de l'arbre, ne sont que des échantillons tirés aléatoirement au sein de l'historique global. Cette approche a pour but de corriger l'un des problèmes récurrents des arbres de décisions : une grande partie des arbres est très sensible à des variations même faibles de l'historique.

Ces modifications de la structure des arbres s'accompagnent d'une grande variabilité de leurs performances. Le ré-échantillonnage de l'historique permet de prendre en compte cette variabilité afin d'améliorer la robustesse et le pouvoir de prédiction du modèle.

- Les variables prédictives considérées lors de la construction d'un arbre sont elles-mêmes tirées aléatoirement parmi les variables prédictives disponibles. Dans notre cas, le nombre de variables sélectionnées est de cinq. Cette étape conduit à une amélioration de la technique précédente en garantissant une plus grande indépendance entre les arbres de la forêt et en permettant un balayage plus large du spectre de possibilité de structure des arbres.

Le nombre d'arbres inclus dans la forêt peut atteindre plusieurs milliers. Mais on peut vérifier que le taux d'erreur associé au modèle ne diminue plus significativement à partir de 1 000 arbres. Pour cela, on fait varier automatiquement le nombre d'arbres qui constitue la forêt, et on calcule l'erreur relative qui y est associée.

Dès lors, à partir d'un nouveau jeu de données qui contient les mêmes variables que l'historique de construction, la forêt aléatoire est susceptible de fournir une prédiction de la manière suivante :

- obtention d'une prédiction arbre par arbre en fonction des valeurs des variables prédictives ;
- estimation d'une prédiction globale par vote généralement à la majorité : la modalité finale donnée par la prédiction sera celle qui aura été la plus représentée parmi les réponses données par les différents arbres.

Signalons enfin que l'utilisation de cette technique réduit significativement l'influence de données manquantes (panne d'un capteur de mesure, transmission interrompue de donnée…) au sein du set nécessaire à la prévision. En effet, si un certain nombre d'arbres inclus dans la forêt ne peuvent pas être entièrement parcourus du fait de données manquantes, ils seront suppléés par d'autres qui, n'incluant pas ces paramètres, pourront toujours donner une prédiction plus précise.

Les forêts d'arbres ont donc été construites à partir d'un historique regroupant six années (1999-2004). Les variables d'entrées considérées dans cet historique regroupent :

- des variables météorologiques *mesurées :* température, vitesse et direction du vent, pression, humidité relative, rayonnement. Celles-ci sont obtenues par les réseaux d'AIRFOBEP et de Météo-France ;
- des variables météorologiques *calculées :* typologies météorologiques régionales, gradient vertical de température, minimum et maximum sur les stations de la température et de la vitesse du vent ;
- des variables chimiques *mesurées :* O_3, NO_2, NO_x, SO_2, PS et CO ;
- des variables chimiques *calculées :* minimum et maximum sur les stations de chaque polluant…
- autres variables : jour de la semaine, mois de l'année…

L'application VIGIPOL utilise par ailleurs, des valeurs de température, vitesse et direction prédites en trois points de la zone d'étude. Toutefois, ne disposant pas d'un historique de ces prévisions, les arbres ont été construits en considérant des valeurs mesurées plutôt que prédites ce qui revient à considérer que les prévisions sont « parfaites ». Toutefois, bien entendu, les résultats obtenus par l'application VIGIPOL et présentés dans le paragraphe 3.4, ont été estimés en utilisant de véritables données de prévision.

4.4. Résultats sur l'année 2006/2007

Pour mesurer le degré d'opérationnabilité de VIGIPOL, deux indicateurs différents ont été utilisés. Ils renseignent sur le taux de bonne prédiction et sur le taux de fausse alerte. Ces deux indicateurs sont respectivement l'indice α et l'indice β dont les définitions sont données ci-dessous :

$$\alpha = \frac{\textit{Nombre de bonnes prédictions de pics}}{\textit{Nombre de pics observés}} \times 100$$

$$\beta = \frac{\textit{Nombre de fausses alertes}}{\textit{Nombre de pics prévus}} \times 100$$

Si on considère l'indicateur de performance α comme étant égal à 100 % et l'indicateur β égal à 0 % cela signifie que le modèle a su déterminer tous les dépassements de seuil observés sans pour autant émettre de fausse alerte, c'est le modèle « idéal ».

Cette évaluation du modèle ne peut avoir lieu que si on suppose qu'aucune mesure de réduction des émissions n'est mise en place suite à son utilisation. Car, en mode opérationnel, on peut espérer que les industriels réagissent lorsque l'outil prévoit une situation très polluée. Il n'est donc pas totalement aberrant de penser qu'au fil du temps et de la prise en compte de ces prévisions par les émetteurs, les épisodes prévus pollués, mais mesurés non pollués, augmentent. Par conséquent les valeurs des paramètres α et β peuvent se dégrader.

Le test de VIGIPOL a conduit à obtenir une table de contingence à partir de laquelle les calculs d'indicateurs ont été effectués. La table de contingence est un moyen particulier de représenter simultanément deux caractères observés sur une même population (Tableau 1). Dans le tableau, « non pollués » signifie que le niveau de pollution étudié n'a pas été atteint et « pollués » signifie que le niveau de pollution étudié a été atteint.

Afin de permettre la distinction entre différents niveaux de pollution et ainsi d'autoriser la mise en place de mesures mieux adaptées à la situation, plusieurs forêts d'arbres ont été construites. Chacune d'elle considère comme variable à prédire un seuil donné de concentration : 110, 150 et 180 µg/m³. À partir des résultats fournis par ces différentes forêts, une stratégie de décision a été mise en place afin de définir un niveau de pollution :

- si la forêt associée au seuil 110 µg/m³ conclue à une absence de pic, l'application VIGIPOL signale une absence de pollution. Sinon :
- si la forêt associée au seuil 150 µg/m³ conclue à une absence de pic, l'application VIGIPOL signale un niveau moyen de pollution. Sinon :
- si la forêt associée au seuil 180 µg/m³ conclue à une absence de pic, l'application VIGIPOL signale un niveau élevé de pollution. Sinon :
- l'application VIGIPOL signale un niveau élevé de pollution.

À partir des différentes tables de contingence obtenues à chaque niveau de pollution et à chaque échéance, on obtient, pour l'année 2006, les résultats qui sont rassemblés dans le Tableau 2.

Tableau 1.
Table de contingence.
Table of contengency.

		Mesure	
		Non	**Oui**
Modèle	Non	Mesurés non pollués et prévus non pollués	Mesurés pollués et prévus non pollués
	Oui	Mesurés non pollués et prévus pollués	Mesurés pollués et prévus pollués

Tableau 2.
Résultat des paramètres de performances sur l'année 2006.
Results of parameters of performance on the year 2006.

Configuration de prévision	Niveau moyen	Niveau élevé	Niveau très élevé
Jour J	α = 97 % β = 1 %	α = 88 % β = 0 %	α = 80 % β = 7 %
Jour J + 1	α = 89 % β = 23 %	α = 76 % β = 19 %	α = 65 % β = 25 %
Jour J + 2	α = 91 % β = 25 %	α = 64 % β = 12 %	α = 62 % β = 16 %
Jour J + 3	α = 87 % β = 34 %	α = 56 % β = 18 %	α = 71 % β = 33 %

Tableau 3.
Résultat des paramètres sur la période janvier 2007 à septembre 2007.
Results of parameters from January to September 2007.

Différence	Niveau moyen	Niveau élevé	Niveau très élevé
Jour J	α = 97 % β = 1 %	α = 89 % β = 3 %	α = 63 % β = 0 %
Jour J + 1	α = 96 % β = 16 %	α = 79 % β = 30 %	α = 55 % β = 25 %
Jour J + 2	α = 91 % β = 30 %	α = 55 % β = 42 %	α = 45 % β = 25 %
Jour J + 3	α = 94 % β = 28 %	α = 35 % β = 50 %	α = 45 % β = 45 %

Quelle que soit la configuration de prévision, les résultats sont satisfaisants. La valeur des paramètres α étant majoritairement supérieure à 70 % et la valeur de l'indice β inférieure à 25 %.

Les résultats de fonctionnement de VIGIPOL, sur la période de janvier à septembre 2007, sont rassemblés dans le tableau ci-dessus :

Les performances de cette version sont dans l'ensemble moins bonnes sur l'année 2007 que sur l'année 2006.

Il est important, à cette étape, de garder à l'esprit que la météorologie de l'année 2007 a été atypique par rapport aux autres années et donc que les performances de VIGIPOL sur cette période sont à prendre avec recul.

Néanmoins, on peut constater que pour le niveau moyen le taux de bonne prévision α varie entre 97 et 91 % suivant les échéances et que le taux de fausse alerte passe de 1 % à 28 % entre les échéances J + 0 et J + 3.

Pour le niveau élevé, les performances sont correctes pour les échéances J + 0 et J + 1 et se dégradent sensiblement à partir de l'échéance J + 2 où l'on ne prévoit plus qu'un dépassement sur deux pour un taux de fausse alerte atteignant les 42 %. L'échéance J + 3 est celle dont les performances sont les plus faibles et où l'on atteint un taux de fausse alerte de 50 %.

C'est sur le niveau très élevé que les performances sont les moins significatives à cause d'un petit nombre de dépassements, qui en plus peuvent être inhabituels. Sur l'ensemble des échéances la moyenne de α est seulement de 52 %. D'un autre côté, les valeurs de β sont de 0 %, pour l'échéance J + 0, et de 45 %, pour l'échéance J + 3.

5. Le site Internet de VIGIPOL

Afin de permettre un accès rapide et facile, un module web a été intégré à l'outil VIGIPOL. Il est destiné à être consulté par différents intervenants (membres d'AIRFOBEP, industriels, service de l'État). Ce module est installé sur un serveur public, mais reste protégé par un système de *login* et de mots de passe. Les *logins* permettent d'identifier les visiteurs et de les rediriger sur différentes pages, selon leurs droits d'accès.



Figure 4.
Le site web. Website.

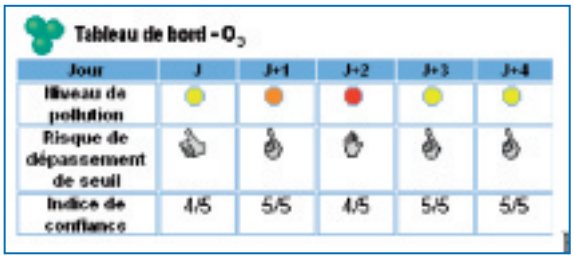


Figure 5.
Le tableau de bord.
Summary table.

Les informations affichées sont mises à jour aussi fréquemment que les calculs et les transferts ont lieu dans le module « processeur ». Elles sont triées par thème, dont les accès sont restreints selon les utilisateurs connectés :

- Variables explicatives ;
- Résultats des arbres ;
- Tableaux de bord journaliers ;
- Cartes de vigilance ;
- Statistiques.

Ce module affiche par défaut les dernières prévisions calculées en temps réel, mais un accès à l'historique est également disponible.

6. Conclusion

La région de l'étang de Berre est très sujette aux épisodes de pollution à l'ozone. Leur fréquence et leur intensité sont liées aux spécificités régionales. Les deux principales caractéristiques qui conduisent à cet état de fait sont :

- les sites industriels émetteurs, nombreux et importants ;
- la situation météorologique locale, liée à la topographie, la présence de masses d'eau importantes et bien sûr l'ensoleillement.

L'objectif du développement de l'application VIGIPOL est de fournir un outil d'anticipation aux industriels afin qu'ils puissent moduler leurs émissions en amont d'épisodes de pollution probables. Cet objectif a été atteint et le mode opérationnel fonctionne, avec de bons résultats théoriques, depuis l'été 2006.

Par contre, il est un peu tôt pour faire une réelle analyse de l'utilisation que les industriels font de cet outil. Un bilan sera réalisé dans le courant de l'année 2008 pour essayer de quantifier l'impact réel de l'outil sur la planification des émissions et donc son incidence sur une évolution globale de la qualité de l'air, et notamment de la pollution à l'ozone, sur la région de Berre.

Parallèlement, de nouveaux développements de l'outil sont en cours de réalisation. Un axe d'améliorations porte sur la pertinence des paramètres déterminants et donc celle de la forêt d'arbres. Par ailleurs, l'outil va être intégré plus globalement dans les divers systèmes de prévisions d'épisodes de pollution atmosphérique d'AIRFOBEP.

Enfin, un challenge ambitieux consistera en la transposition de ce modèle à d'autres polluants (notamment le dioxyde de soufre), dans l'objectif final de prévoir l'ensemble des épisodes de pollution qui touchent la région de l'étang de Berre.

Remerciements

Nous tenons à remercier M. Morgan JACQUINOT qui, à travers ses remarques pertinentes, a permis de faire avancer le projet.

Références

1. Arrêté ministériel, 8 juillet 2003.
2. Coll I, Fayet S, Lasry F. Production d'ozone et sensibilité aux émissions : premiers résultats de scénarios de réductions des émissions. Sixième rencontre ESCOMPTE, Marseille, Préfecture, 2-4 février 2005.
3. Exemple de diagramme isopleth in Camredon M, Aumont B. Modélisation chimique de l'ozone et des oxydants gazeux. *Pollution Atmosphérique* 2007 ; 193 : 51-60.
4. Coll I, Lasry F, Fayet S. Simulation de scénarios de réductions des émissions. Colloque de restitution Primequal, Avignon, Palais des Papes, 31 mars 2006.
5. Airfobep. Inventaire des émissions de polluants atmosphériques sur la zone d'AIRFOBEP (exploitation de l'inventaire PACA 1999), 2006.
6. Van Loon M, Vautard R, Shaap M, *et al.* Evaluation of long term ozone simulations from seven regional air quality models and their ensemble. *Atmospheric Environment* 2007 ; 41 : 173-88.
7. Vautard R, Builtjes PHJ, Thunis P, *et al.* Evaluation and intercomparison of Ozone and PM_{10} simulations by several chemistry transport models over four European cities within the CityDelta project. *Atmospheric Environment* 2007 ; 41 : 2083-97.
8. Honoré C, Meleux F. Travaux relatifs au système PREV'AIR, Ineris, novembre 2005.
9. Breiman L. Random forests. *Machine Learning* 2001 ; 45 : 5-32.