

Application de la théorie des fonctions de croyances à la surveillance de l'environnement

Application of belief functions theory to environment diagnosis

P. VANNOORENBERGHE¹

E. LEFEVRE²

O. COLOT²

¹ Heudiasyc, UMR CNRS 6599

Université de Technologie de Compiègne

BP 20529, 60205 Compiègne cedex

² Perception Systèmes Information, EA 2120

Université / INSA de Rouen

Place Emile Blondel, 76821 Mont Saint-Aignan cedex

Patrick.Vannoorenberghe@hds.utc.fr

Résumé

Dans ce papier, on se propose d'appliquer la théorie des fonctions de croyance au diagnostic de l'environnement. L'approche présentée consiste à fusionner les informations issues de mesures de pollution de manière à prédire l'apparition d'un pic. L'originalité de cette étude réside dans le fait que le contexte de l'apprentissage évolue dans le temps avec les données de pollution. Notre contribution réside dans l'utilisation de coefficients de fiabilité attribués à chaque source d'information afin de modéliser le plus précisément possible l'évolution dans le contexte de l'apprentissage. Ces coefficients sont définis par l'intermédiaire d'une mesure de ressemblance entre des approximations de lois de probabilités a priori inconnues. La méthode a donné des résultats satisfaisants sur un problème concret de prédiction de la pollution atmosphérique sur le réseau de surveillance de la ville de Lyon.

Mots Clef

Diagnostic, Fonctions de croyance, Apprentissage, Environnement.

Abstract

In this paper, we propose to apply the theory of belief functions to environment diagnosis. The proposed approach consists in aggregating information extracted from several pollutant measurements in order to predict a pollution episode. The originality of this study resides in the fact that the training context evolves with time due to an evolution of the pollution data. Our

contribution is based on the computation of reliability coefficients assigned as precisely as possible to every information source in order to model the evolution in the learning context. These coefficients are defined by means of a comparison between unknown probability laws approximations. The methodology gives satisfactory results on a problem of atmospheric pollution prediction using the surveillance network in Lyon.

Keywords

Diagnosis, Belief function, Learning, Environment.

1 Introduction

La place croissante occupée par les problèmes environnementaux dans notre société (LOI 96-1236 du 30 décembre 1996 sur l'Air) et le poids économique important de l'industrie de l'environnement ouvrent aujourd'hui un nouveau champ d'application aux techniques de surveillance. Diagnostiquer l'état de l'environnement pose des problèmes spécifiques liés aux difficultés rencontrées dans l'acquisition des informations et la modélisation des phénomènes [1] :

- Difficultés d'acquisition des informations : la mesure de certaines grandeurs nécessite des capteurs sophistiqués peu compatibles avec les contraintes du terrain, et la dispersion géographique des capteurs rend difficile une maintenance régulière des stations de mesures engendrant par ce fait un certain nombre de données manquantes.
- Difficultés de modélisation : la complexité des

phénomènes physiques (recombinaison de polluants dans l'atmosphère) rend difficile la mise au point de modèles analytiques prenant en compte de façon suffisamment précise le comportement du système observé.

Depuis, quelques années, une approche du type "Reconnaissance de Formes (RdF)" au problème de la surveillance et du diagnostic a émergé [2]. L'approche RdF du diagnostic consiste à construire une fonction de décision par apprentissage à partir d'un historique de mesures effectuées sur le système. Les méthodes utilisées sont celles de l'analyse de données et de l'apprentissage statistique mais peuvent s'adapter aux données imparfaites. Ainsi, elles permettent de :

- modéliser les informations imprécises et incertaines sur les capteurs (par les fonctions de croyance [3] ou les distributions de possibilité par exemple [4]),
- rendre compte de l'évolution du contexte dans le cadre de l'apprentissage.

Dans le cadre du diagnostic de l'environnement, la prise en compte de ces informations supplémentaires permet généralement d'augmenter les performances du système de surveillance [5, 6, 7]. Cette étude se focalise sur une approche de type crédibiliste avec une attention particulière sur le problème de l'apprentissage des fonctions de croyance.

Ce papier propose dans la section 2 quelques rappels théoriques sur les fonctions de croyance permettant ainsi d'introduire les notations utilisées. Nous présentons ensuite le problème des données environnementales à traiter (section 3) ainsi que la méthodologie permettant le diagnostic de l'environnement (section 4). Enfin, cet article se termine par quelques résultats expérimentaux sur les données extraites du réseau de surveillance de la ville de Lyon (cf. section 5).

2 Les fonctions de croyance

Nous rappelons dans cette section quelques éléments théoriques des fonctions de croyance en insistant sur la possibilité d'affaiblir une structure de croyance. La théorie de l'évidence fut initialement introduite par Dempster [8] lors de ses travaux sur les bornes inférieure et supérieure d'une famille de distributions de probabilités. A partir de ce formalisme mathématique, Shafer [9] a montré l'intérêt des fonctions de croyance pour la modélisation de connaissances incertaines. L'utilité des fonctions de croyance, comme alternative aux probabilités subjectives, a été démontrée plus tard de manière axiomatique par Smets [10, 11] au travers du *Modèle de Croyances Transférables*.

Soit Θ l'ensemble des N hypothèses solutions du problème. L'ensemble Θ , appelé *cadre de discernement*, est défini de la manière suivante :

$$\Theta = \{H_n\} \quad n \in [1, N]. \quad (1)$$

On suppose que les hypothèses sont exclusives et que le cadre de discernement est exhaustif. On définit une masse de probabilité élémentaire, appelée *masse de croyance*, qui caractérise la véracité d'une proposition \mathcal{H} pour une source d'information S_j donnée pour tout $j \in [1, Q]$. La masse associée à cette source S_j que nous noterons m_j est alors définie par :

$$m_j : 2^\Theta \rightarrow [0, 1] \quad (2)$$

et vérifie les propriétés suivantes : $m_j(\emptyset) = 0$ et $\sum_{\mathcal{H} \subseteq \Theta} m_j(\mathcal{H}) = 1$. Cette probabilité se différencie d'une probabilité au sens classique du terme par le fait que la totalité de la masse de croyance est répartie non seulement sur les hypothèses singletons H_n mais aussi sur les hypothèses combinées \mathcal{H} . La modélisation issue de la fonction m_j est appelée jeu de masses. A partir de la fonction m_j , on définit respectivement les fonctions de *crédibilité* Cr_j et de *plausibilité* Pl_j par :

$$Cr_j(\mathcal{H}) = \sum_{\mathcal{H}' \subseteq \mathcal{H}} m_j(\mathcal{H}') \quad (3)$$

$$Pl_j(\mathcal{H}) = \sum_{(\mathcal{H} \cap \mathcal{H}') \neq \emptyset} m_j(\mathcal{H}') = 1 - Cr_j(\overline{\mathcal{H}}) \quad (4)$$

où $\overline{\mathcal{H}}$ représente l'événement contraire de la proposition \mathcal{H} . La crédibilité $Cr_j(\mathcal{H})$ mesure la force avec laquelle on croit en la véracité de la proposition \mathcal{H} . La plausibilité $Pl_j(\mathcal{H})$, fonction duale de la crédibilité, mesure l'intensité avec laquelle on ne doute pas de \mathcal{H} . La principale difficulté consiste à modéliser les connaissances sur le problème en initialisant de manière adéquate [12] les fonctions de croyance m_j . Cette modélisation dépend généralement de l'application envisagée. Cependant, il est intéressant de remarquer que certaines modélisations [13] permettent de s'octroyer trois propriétés fondamentales, qui sont : la cohérence avec l'approche bayésienne dans le cas où les probabilités *a priori* sont connues, la séparabilité des hypothèses H_n et la cohérence avec l'association probabiliste des sources¹. Ces modèles ont été proposés par A. Appriou qui définit une masse de croyance m_{nj} à chaque couple (H_n, S_j) . De plus, des coefficients notés q_{nj} permettent d'affaiblir la masse de croyance accordée à chacun de ces couples. Ces modèles sont définis respectivement pour le modèle 1 par :

$$m_{nj}(H_n) = 0 \quad (5)$$

¹L'auteur intéressé trouvera la démonstration dans [14]

$$m_{nj}(\overline{H}_n) = q_{nj} - q_{nj} \cdot R_j \cdot p(x'_j/H_n) \quad (6)$$

$$m_{nj}(\Theta) = 1 - q_{nj} + q_{nj} \cdot R_j \cdot p(x'_j/H_n) \quad (7)$$

et le modèle 2 :

$$m_{nj}(H_n) = \frac{q_{nj} \cdot R_j \cdot p(x'_j/H_n)}{1 + R_j \cdot p(x'_j/H_n)} \quad (8)$$

$$m_{nj}(\overline{H}_n) = \frac{q_{nj}}{1 + R_j \cdot p(x'_j/H_n)} \quad (9)$$

$$m_{nj}(\Theta) = 1 - q_{nj}. \quad (10)$$

Dans ces équations, R_j est un facteur de normalisation contraint par :

$$R_j \in [0, \frac{1}{\max_{n \in [1, N]} p(x'_j/H_n)}] \quad (11)$$

où $p(x'_j/H_n)$ représente la densité de probabilité de la mesure x'_j issue de la source S_j sous les différentes hypothèses H_n . Ces densités pourront être déterminées à partir de la connaissance *a priori* de la distribution de probabilité ou à partir d'estimations calculées sur un ensemble d'apprentissage (estimateurs de densité tels que les noyaux). Celles-ci sont donc plus ou moins représentatives des densités réellement rencontrées. Les coefficients q_{nj} caractérisent ici le degré de représentativité. Lorsque les densités sont parfaitement représentatives de l'apprentissage alors les coefficients q_{nj} sont égaux à 1 et dans ce cas les structures de croyances ne sont pas affaiblies. A l'inverse, lorsque la distribution de probabilité est totalement méconnue, ce qui est caractérisé par un coefficient q_{nj} égal à 0, les structures de croyance deviennent élément neutre de l'opérateur de combinaison de Dempster. Appriou [13] fixe ces coefficients de fiabilité à 1 lorsque la confiance accordée aux sources est élevée et à 0.9 dans le cas contraire. Une autre approche [15] détermine la valeur de ces coefficients à l'aide d'une fonction linéaire de la dispersion des données relatives à chaque hypothèse et chaque source à valeur dans $[0.9, 1]$. Pour un écart-type minimum, le coefficient est fixé à 1 alors que dans le cas d'un écart-type maximal le coefficient est fixé à 0.9.

Dans le cadre de la théorie de l'évidence de Dempster-Shafer, la fusion des informations issues de sources distinctes est réalisée en utilisant la *loi de combinaison de Dempster*. Celle-ci, qui s'avère commutative et associative, est définie par :

$$\forall \mathcal{H} \in 2^\Theta \quad m(\mathcal{H}) = m_1(\mathcal{H}) \oplus \dots \oplus m_Q(\mathcal{H}) \quad (12)$$

où \oplus représente l'opérateur de combinaison. Dans un cas à deux sources notées S_i et S_j , la combinaison peut

se mettre sous la forme :

$$m(\mathcal{H}) = \frac{1}{1 - \kappa} \sum_{(\mathcal{H}' \cap \mathcal{H}'' = \mathcal{H})} m_i(\mathcal{H}') \cdot m_j(\mathcal{H}'') \quad (13)$$

où κ est défini par :

$$\kappa = \sum_{(\mathcal{H}' \cap \mathcal{H}'' = \emptyset)} m_i(\mathcal{H}') \cdot m_j(\mathcal{H}''). \quad (14)$$

Dans l'équation (13), le coefficient κ reflète le conflit existant entre les deux sources S_i et S_j . Lorsque ce facteur est égal à 1, les sources sont en conflit total et les informations ne peuvent être fusionnées. Au contraire, lorsque κ est nul, les sources sont en parfait accord. Cette règle de fusion, déduite de la règle de conditionnement [16], a été critiquée dans plusieurs travaux dont [17], en particulier dans le cas de sources en conflit total. Pour pallier cet inconvénient, Dubois et Prade [18] ont été amenés à définir les opérateurs de fusion conjonctive et disjonctive $\forall \mathcal{H} \in 2^\Theta$:

$$(m_i \cap m_j)(\mathcal{H}) = \sum_{\mathcal{H}' \cap \mathcal{H}'' = \mathcal{H}} m_i(\mathcal{H}') \cdot m_j(\mathcal{H}'') \quad (15)$$

et

$$(m_i \cup m_j)(\mathcal{H}) = \sum_{\mathcal{H}' \cup \mathcal{H}'' = \mathcal{H}} m_i(\mathcal{H}') \cdot m_j(\mathcal{H}''). \quad (16)$$

Une fois la masse résultante m ainsi obtenue, la décision peut alors être prise. Différentes règles de décision ont été définies, les plus courantes étant la règle du maximum de plausibilité et la règle du maximum de crédibilité ou de probabilité pignistique [19, 20].

3 Les problèmes liés aux données environnementales

Le problème envisagé concerne le diagnostic précoce de l'évolution de la pollution atmosphérique vers une situation du type événement anormal. On entend par événement une situation qui peut être profondément différente selon l'application envisagée. Par exemple, l'apparition d'un pic de pollution à l'ozone peut survenir avec un niveau de seuil particulier [7].

Dans ce cadre, nous nous sommes focalisés sur le réseau de surveillance de la ville de Lyon. COPARLY (Comité de Coordination pour le Contrôle de la Pollution atmosphérique en région Lyonnaise), le nom donné au réseau de surveillance de la pollution sur l'agglomération lyonnaise, couvre une zone d'environ 30×30 Kms. Le réseau est constitué de 49 sites de mesures permettant un échantillonnage de différentes concentrations de polluants. Plus précisément, 17 capteurs permettent une mesure horaire du monoxyde

d'azote (NO) et du dioxyde d'azote (NO₂) et 7 capteurs d'ozone (O₃). La température est également mesurée chaque heure et est prédite à un horizon de 24 heures. Enfin, une prévision journalière des conditions atmosphériques sur 36 heures est réalisée par l'intermédiaire d'un indice climatologique, appelé géopotential, proposé par P. Benichou à Météo-France [21]. Nous disposons d'un ensemble de 897 jours de données réparties sur les périodes du 1 Avril au 30 Septembre des années 1994 à 1998. L'événement à diagnostiquer, défini par COPARLY, est produit sous forme binaire comme la prédiction du jour J de l'absence ou de la présence d'une pollution au jour J+1.

Les informations délivrées par les capteurs de pollution sont de nature imparfaite c'est-à-dire imprécises, incertaines et incomplètes. En effet, une première imperfection réside dans l'incomplétude des bases de données. A titre d'exemple, sur les mesures de concentration en ozone, 24% des données sont manquantes. Plusieurs techniques d'analyse de données permettent de prendre en compte ce type d'imperfection soit en remplaçant les données manquantes ou tout simplement en les ignorant. De surcroît, la pollution étant un phénomène difficilement modélisable, il s'avère que le contexte du problème posé évolue en fonction du temps. A titre d'exemple, on a représenté sur la figure Fig. 1 l'évolution de la concentration moyenne journalière en ozone sur un capteur durant les cinq années d'étude. Les moyennes annuelles sur ce cap-

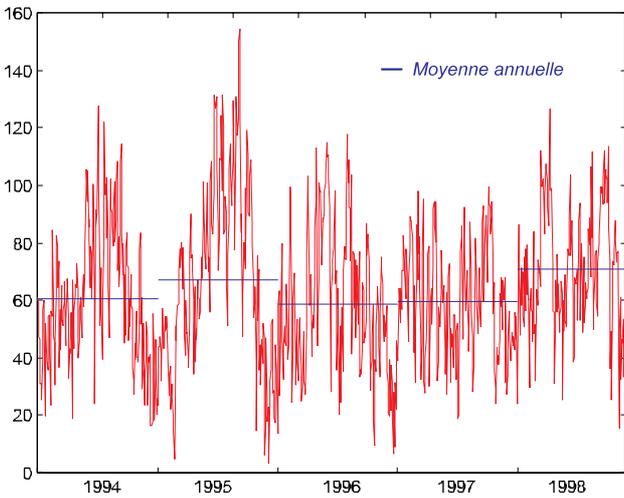


FIG. 1: Evolution de la concentration moyenne journalière en ozone sur un capteur donné

teur sont également représentées sur ce graphique. On s'aperçoit que ces moyennes varient d'une année sur l'autre rendant le problème de l'apprentissage d'autant plus délicat à appréhender. De plus, le problème

étudié s'oriente autour de la détection d'événements rares. En effet sur les 897 jours de données observés, seulement 58 jours sont considérés comme des situations anormales représentant uniquement 6.3% des données. Ces constatations permettent de mettre en oeuvre une méthodologie appliquée aux traitements de ce type de données.

4 Méthodologie

La méthodologie mise en oeuvre s'articule autour de trois principales étapes. Dans un premier temps, une modélisation des connaissances sur le problème est réalisée par l'intermédiaire de l'apprentissage des densités de probabilité $p(x/H_n)$. Ensuite, de manière à qualifier la bonne représentativité de cette modélisation, on réalise un apprentissage des coefficients q_{nj} permettant ainsi d'obtenir les jeux de masses m_{nj} . Enfin, une partie finale de fusion des informations permet de construire la fonction de croyance utilisée pour la prise de décision.

4.1 Modélisation des connaissances

La construction des jeux de masse m_{nj} (cf. équations (8) à (10)) passent par une première estimation des densités $p(x_j/H_n)$. De manière à estimer ces densités suffisamment précisément, on utilise une méthode d'estimation par noyau [22] sur les données de la base d'apprentissage. Cette technique permet d'approcher une fonction $p(x)$ comme une somme de T termes. Dans ce contexte, l'estimation $\hat{p}(x)$ de la fonction peut s'écrire :

$$\hat{p}(x) = \frac{1}{Th} \sum_{t=1}^T K\left(\frac{x_j - x_j^t}{h}\right) \quad (17)$$

où x_j^t décrit les valeurs des données dans la base d'apprentissage. Dans cette équation, la fonction $K(\cdot)$ est appelée noyau dont le plus classique est le noyau gaussien :

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp^{-\frac{1}{2}u^2}. \quad (18)$$

Les figures Fig. 2 et Fig. 3 proposent les densités estimées à partir de cette méthode pour deux jeux de données correspondant aux concentrations respectives d'ozone sur un capteur dans le cas d'épisode de pollution ou non. On a représenté sur ces figures les données relatives à un capteur pour la base d'apprentissage par un histogramme puis par l'estimation obtenue avec un noyau gaussien. A partir de ces estimations et des équations (8), (9) et (10) liées au modèle 2 proposé par A. Appriou dans [13], il nous reste à estimer les coefficients q_{nj} de manière à obtenir les différents jeux de masses. En ce qui concerne les coefficients R_j , nous avons choisi de les borner à droite.

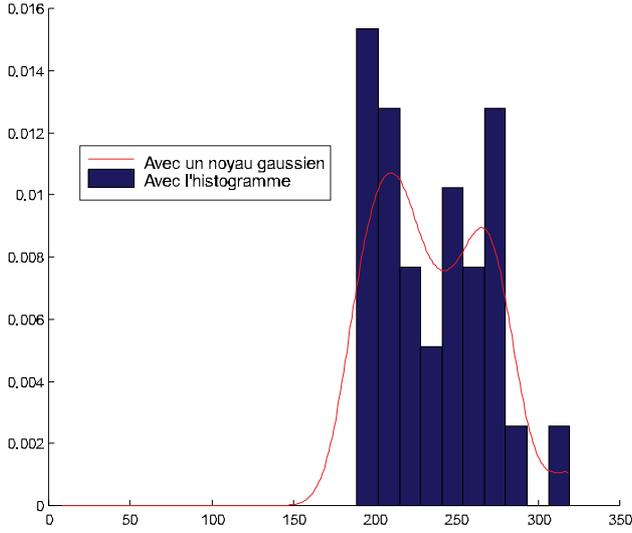


FIG. 2: Estimations de la densité de probabilité pour l'hypothèse "Jour pollué"

4.2 Apprentissage des coefficients q_{nj}

De manière à essayer de qualifier la représentativité de l'apprentissage, nous proposons d'utiliser une mesure d'information qui est calculée pour chaque couple (S_j, H_n) . Le but de la démarche est de permettre d'évaluer la fiabilité de la source S_j relativement à l'hypothèse H_n lorsque les conditions d'apprentissage ont évolué. Pour satisfaire ce principe, on s'appuie sur l'examen des données d'une base de validation et la comparaison entre ces données et les données de la base d'apprentissage. Plus précisément, nous proposons l'évaluation des q_{nj} par le calcul de la distance de Hellinger entre les données relatives à (S_j, H_n) sur la base d'apprentissage et les données relatives à (S_j, H_n) sur la base de validation. La comparaison entre ces deux jeux de données est réalisée en comparant l'histogramme optimal $\hat{\lambda}_{C_{opt}/H_n}^a$ calculé sur la base d'apprentissage et l'histogramme $\hat{\lambda}_{C_{opt}/H_n}^v$ calculé sur la base de validation. Ces histogrammes sont considérés comme optimaux au sens du principe de conservation de l'information [23]. Ainsi, $\hat{\lambda}_{C_{opt}/H_n}^a$ et $\hat{\lambda}_{C_{opt}/H_n}^v$ représentent respectivement les histogrammes optimaux pour un découpage en C_{opt} classes réalisés sur les données relatives à l'hypothèse H_n dans la base d'apprentissage a et la base de validation v . Ce découpage en C_{opt} classes, obtenu par le calcul d'un critère d'information du type Akaike, permet d'obtenir des estimations $\hat{\lambda}_{C_{opt}/H_n}^a$ et $\hat{\lambda}_{C_{opt}/H_n}^v$ des densités respectives quant à ces données [24]. Enfin, une mesure de dissemblance entre ces deux estimations de lois peut être cal-

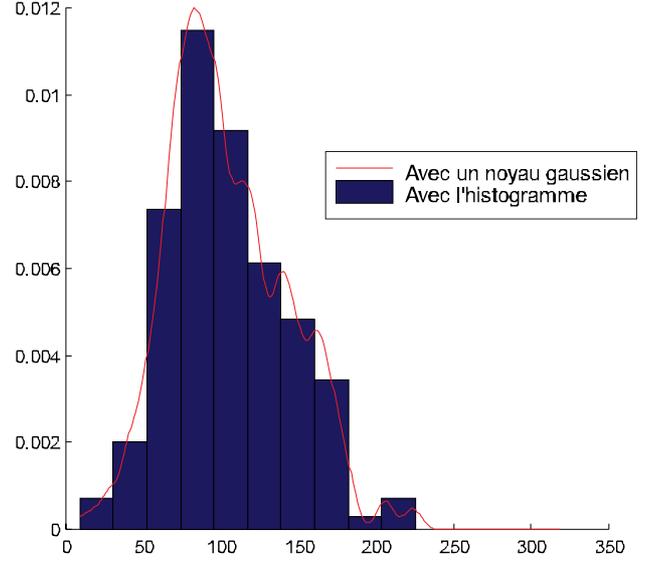


FIG. 3: Estimations de la densité de probabilité pour l'hypothèse "Jour non pollué"

culée en introduisant la distance de Hellinger [23, 25]. Lorsque les estimations de lois sont proches, c'est-à-dire lorsque les données de la base d'apprentissage sont proches des données de la base de validation, la distance tend vers 0. Au contraire, si les lois sont fortement dissemblables, alors la distance est proche de 1. Compte-tenu de cette remarque, on choisit de faire évoluer le coefficient q_{nj} en suivant l'équation suivante :

$$q_{nj} = 1 - \text{Hell}(\hat{\lambda}_{C_{opt}/H_n}^a; \hat{\lambda}_{C_{opt}/H_n}^v). \quad (19)$$

L'obtention de ces différents coefficients permet la construction des jeux de masses relatifs à chaque hypothèse H_n et à chaque source d'information S_j .

4.3 Fusion et décision

A partir des différents jeux de masse $m_{nj}(\cdot)$, on réalise une première étape de fusion de manière à obtenir un jeu de masses unique $m_j(\cdot)$ pour chaque source d'informations S_j . Cette étape conduit à utiliser l'opérateur de combinaison de Dempster qui peut s'écrire :

$$m_j(\mathcal{H}) = \bigoplus_{n \in [1, N]} m_{nj}(\mathcal{H}) \quad \forall \mathcal{H} \in 2^\Theta. \quad (20)$$

Une seconde étape de fusion consiste à agréger les informations de chacune des sources S_j de manière à obtenir le jeu de masse final $m(\cdot)$. Celui-ci est calculé par :

$$m(\mathcal{H}) = \bigoplus_{j \in [1, Q]} m_j(\mathcal{H}) \quad \forall \mathcal{H} \in 2^\Theta. \quad (21)$$

TAB. 1: Résultats sur la base d'apprentissage

Réalité	Décision	
	Pollution	Non Pollution
Pollution	100%	0%
Non Pollution	4.56%	95.44%

TAB. 2: Résultats sur la base de validation

Réalité	Décision	
	Pollution	Non Pollution
Pollution	68.75%	31.25%
Non Pollution	1.31%	98.69%

Enfin, la prise de décision se fait en utilisant le critère de maximum de plausibilité. Ce critère conduit à retenir l'hypothèse singleton H_n qui maximise la fonction de plausibilité parmi l'ensemble des hypothèses en compétition. Une autre règle de décision associant des coûts différents à une non détection et à une fausse alarme est à l'étude.

5 Résultats

Dans cette section, nous présentons quelques résultats sur un problème concret de prédiction de la pollution atmosphérique sur le réseau de surveillance de la ville de Lyon. L'événement à prédire est de type binaire conduisant ainsi à un cadre de discernement Θ à deux hypothèses. La constitution des bases de données a été faite de manière chronologique. Les années 1994 à 1996 (488 jours) ont été choisies comme base d'apprentissage correspondant à 458 jours non pollués et 30 jours pollués. Les bases de validation et de test correspondent respectivement aux années 1997 et 1998 représentant 245 (229 non pollués et 16 pollués) et 164 jours (152 non pollués et 12 pollués). En ce qui concerne les informations, nous avons choisi de travailler avec 3 sources qui sont respectivement le géopotential (S_1), la température (S_2) et la concentration en ozone (S_3) maximales calculées le jour précédent. Les tableaux Tab. 1, Tab. 2 et Tab. 3 représentent les pourcentages de classification sur chacune de ces bases de données. Ces résultats permettent d'obtenir un taux de bonne

TAB. 3: Résultats sur la base de test

Réalité	Décision	
	Pollution	Non Pollution
Pollution	100%	0%
Non Pollution	1.97%	98.03%

TAB. 4: Coefficients q_{nj} pour $n \in [1, N]$ et $j \in [1, Q]$; $N=2$ hypothèses, $Q=3$ sources

Hypothèse H_n	Source S_j		
	S_1	S_2	S_3
Non Pollution	0.999	0.997	0.970
Pollution	0.974	0.978	0.918

classification de 97.18% contre un taux de 92.98% dans le cas où les coefficients q_{nj} sont tous égaux à 1. Le tableau Tab. 4 présente les valeurs respectives de ces coefficients. Il met en évidence le fait que le coefficient relatif à la source S_3 dans le cas de l'hypothèse de jour pollué est plus faible. Ceci est imputable à une mauvaise représentativité de l'apprentissage car le nombre d'exemples relatifs à cette hypothèse est faible comparativement à l'hypothèse de jour non pollué. Ce fait est particulièrement sensible sur les données relatives à l'ozone (source S_3) où les mesures de concentration sont très imprécises. Ce résultat permet de montrer l'utilité d'un apprentissage suffisamment précis des coefficients d'affaiblissement des fonctions de croyance et d'obtenir ainsi une amélioration des performances du système de surveillance.

6 Conclusion

Dans cet article, nous avons proposé une application de la théorie des fonctions de croyance à la détection de situations anormales liées à la surveillance de l'environnement. L'approche présentée consiste à fusionner les informations (malheureusement imparfaites) issues des capteurs de mesures de pollution de manière à prédire l'apparition d'un pic. Ces informations ont été modélisées par des fonctions de croyance permettant ainsi la gestion de leur imprécision et leur incertitude. L'originalité de cette étude réside dans le fait que le contexte de l'apprentissage évolue au fil du temps suivant ainsi l'évolution de la pollution de fond. Pour pallier cet inconvénient, nous avons mis en oeuvre une méthode de validation de l'apprentissage basée sur la comparaison des données d'apprentissage avec les données extraites d'une base de validation. Nous avons montré que la méthode proposée s'avère efficace sur un problème concret de prédiction de la pollution atmosphérique sur le réseau de surveillance de la ville de Lyon. Les travaux futurs portent sur une sélection optimale des capteurs comme sources d'informations pour la fusion.

Remerciements

Les auteurs tiennent à remercier l'association CO-PARLY pour avoir aimablement mis à disposition

les informations concernant les mesures de pollution et les données météorologiques. Qu'il soit permis de citer l'Association ECRIN (Echange et Coordination Recherche-Industrie) qui par l'intermédiaire de son groupe de travail "Prévention des pics de pollution" du Club Ingénierie du Traitement de l'Information propose un lieu d'échange et de rencontres entre différentes communautés (<http://www.ecrin.asso.fr/>). Il est à noter que ces travaux sont co-financés par l'intermédiaire d'un projet de recherche soutenu par le GDR-PRC ISIS du Centre National de la Recherche Scientifique (CNRS).

Références

- [1] A. Ionescu, *Contribution à la compréhension de la dispersion des polluants atmosphériques en zone urbaine*. PhD thesis, Université de Paris XII - Val de Marne, 1998.
- [2] T. Denoeux, M. Masson, and B. Dubuisson, "Advanced pattern recognition techniques for system monitoring and diagnosis : A survey," *Journal européen des systèmes automatisés*, vol. 31, no. 9-10, pp. 1509–1539, 1998.
- [3] S. Petit-Renaud, *Application des systèmes flous et de la théorie des croyances à l'estimation fonctionnelle en présence d'informations incertaines ou imprécises*. PhD thesis, Université de Technologie de Compiègne, 1999.
- [4] P. Vannoorenberghe, R. Lherbier, G. Delmaire, and G. Roussel, "Possibilistic identification of pollutant source by multisensor data fusion," in *Air Pollution Modelling and Simulation, APMS'98*, vol. 2, (Champs-Sur-Marne), pp. 767–768, 1998.
- [5] M. Masson, S. Canu, Y. Grandvalet, and A. Lynggaard-Jensen, "Software sensor design based on empirical data," *Ecological Modelling*, vol. 120, pp. 131–139, 1999.
- [6] S. Petit-Renaud and T. Denoeux, "Application de la théorie des fonctions de croyance en régression," in *Rencontres francophones sur la Logique Floue et ses Applications*, (Valenciennes), pp. 169–176, Cépadués, 1999.
- [7] N. Peton, G. Dray, D. Pearson, M. Mesbah, and B. Vuillot, "Modelling and analysis of ozone episodes," in *APMS'98, Air Pollution and Modelling Simulation*, (Champs-sur-Marne), pp. 75–86, 1998.
- [8] A. Dempster, "Upper and lower probabilities induced by multivalued mapping," *Annals of Mathematical Statistics*, vol. AMS-38, pp. 325–339, 1967.
- [9] G. Shafer, *A Mathematical Theory of Evidence*. Princeton University Press, 1976.
- [10] P. Smets, "What is Dempster-Shafer's model?," in *Advances in the Dempster-Shafer Theory of Evidence* (R. Yager, M. Fedrizzi, and J. Kacprzyk, eds.), pp. 5–34, Wiley, 1994.
- [11] P. Smets and R. Kennes, "The transferable belief model," *Artificial Intelligence*, vol. 66, no. 2, pp. 191–234, 1994.
- [12] E. Lefevre, O. Colot, and P. Vannoorenberghe, "Using information criteria in dempster-shafer's basic belief assignment," in *International Conference on Fuzzy Systems, FUZZ-IEEE'99*, (Seoul, KOREA), 1999.
- [13] A. Appriou, "Probabilités et incertitude en fusion de données multi-senseurs," *Revue Scientifique et Technique de la Défense*, vol. 11, pp. 27–40, 1991.
- [14] A. Appriou, "Multisensor signal processing in the framework of the theory of evidence," in *Application of Mathematical Signal Processing Techniques to Mission Systems*, pp. (5–1)(5–31), Research and Technology Organization (Lecture Series 216), November 1999.
- [15] S. Mathevet, "Application de la théorie de l'évidence à la combinaison de segmentations en régions," in *Dix-Septième Colloque GRETSI (Vannes)*, pp. 635–638, Septembre 1999.
- [16] P. Smets, "The combination of evidence in the transferable belief model," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 12, no. 5, pp. 447–458, 1990.
- [17] R. Yager, "On the Dempster-Shafer framework and new combination rules," *Information Sciences*, vol. 41, pp. 93–138, 1987.
- [18] D. Dubois and H. Prade, "A set-theoric view of belief functions : Logical operations and approximations by fuzzy sets," *International Journal of General Systems*, vol. 12, pp. 193–226, 1986.
- [19] H. Xu, Y.-T. Hsia, and P. Smets, "The transferable belief model for decision making in the valuation-based system," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 26A, pp. 698–707, 1996.
- [20] T. Denoeux, "Analysis of evidence-theory decision rules for pattern classification," *Pattern Recognition*, vol. 30, no. 7, pp. 1095–1107, 1997.
- [21] P. Benichou, "Classification automatique de configurations météorologiques sur l'europe occidentale," Tech. Rep. Monographie 8, METEO-FRANCE, 1995.

- [22] G. Saporta, *Probabilités, Analyse des Données et Statistique*. Editions Technip, 1990.
- [23] O. Colot, C. Olivier, P. Courtellemont, A. El-Matouat, and D. de Brucq, "Information criteria and abrupt changes in probability laws," in *Signal Processing VII : Theories and Applications* (M. Holt, C. Cowan, P. Grant, and W. Sandham, eds.), pp. 1855–1858, EUSIPCO'94, September 1994.
- [24] E. Lefevre, O. Colot, and P. Vannoorenberghe, "A classification method based on the dempster-shafer's theory and information criteria," in *International Conference on Information Fusion, FUSION'99*, (San Franscisco, Silicon Valley, USA), pp. 1179–1184, 1999.
- [25] M. Basseville, "Distance measures for signal processing and pattern recognition," Rapport de recherche interne 899, INRIA, Rennes, 1988.