

# Apports des réseaux Bayésiens dynamiques et des modèles de mélange Gaussiens pour la classification des points singuliers de l'infrastructure ferroviaire

## Classification of rail singular points using dynamic Bayesian networks and Gaussian mixture models

Allou SAME\*, Laurent BOUILLAUT\* et Abdeljabbar BEN SALEM\*\*

\*INRETS (LTN)  
2 avenue du Général Malleret-Joinville  
94114 Arcueil Cedex

<sup>†</sup>Nancy université, CNRS (CRAN-UMR 7039)  
2 rue Jean Lamour  
54519 Vandoeuvre les Nancy Cedex

Didier LEVY.  
RATP (EST Voie)  
40b rue R. Salengro  
94724 Fontenay-sous-Bois

### Résumé

Cet article présente l'implémentation d'un Réseau Bayésien Dynamique (RBD) à variable exogène continue, dans un contexte de classification de points singuliers de l'infrastructure ferroviaire. Les tables de probabilités conditionnelles de ce modèle sont estimées à l'aide d'une stratégie s'appuyant sur des mélanges de lois gaussiennes estimées par l'algorithme EM. Les tests effectués sur une base de données énumérant les points singuliers des rails du réseau de métro parisien, ont mis en évidence de bonnes performances de la méthode proposée, en comparaison avec l'approche d'un réseau bayésien dynamique à variable exogène discrète. Outre le bon comportement de la méthode proposée, la parcimonie de l'approche est démontrée dès que l'ordre du modèle est supérieur ou égal à deux. Cette implémentation permet la fusion efficace de la décision d'un capteur spécifique de détection de défaut de rail avec celle du réseau bayésien et de lever ainsi certaines confusions du capteur.

### Summary

This paper addresses the implementation of a Dynamic Bayesian Network (DBN) with continuous exogenous input. The motivation for this study is the improvement of the classification of railway infrastructure singular points. Gaussian mixture distributions estimated via the EM algorithm are used to train the conditional probability tables of the DBN. The experimental study on a database recording the singular points of the Parisian subway rails shows that the proposed approach generates results comparable to those obtained with a discrete exogenous variable DBN. In addition to the good performances of the proposed method, its parsimony is proved as soon as the order of the model is greater than one. This continuous implementation allows the effective fusion with the decision provided by a specific rail defect sensor.

### Introduction

L'évolution des infrastructures ferroviaires, au cours de ces dernières décennies, a eu pour principale conséquence une très forte augmentation des contraintes exercées sur les voies, que ce soit en termes de fréquence de passage, de vitesse de circulation ou de charge à l'essieu. De nouveaux problèmes de sûreté de fonctionnement sont alors apparus. Parmi eux, l'un des plus critiques est certainement la prévention des ruptures de rails et donc, la détection précoce des défauts mineurs conduisant à ces ruptures. Dans ce cadre, un multicapteur à courants de Foucault a été réalisé [8] en partenariat avec la RATP afin de détecter précisément toute anomalie de rail et améliorer ainsi la disponibilité et la sécurité du système, en particulier dans le contexte des métros à conduite automatisée. Ce capteur permet de détecter les ruptures de rails mais également certains défauts mineurs (fissures, écailles, usures ondulatoires...) pouvant parfois conduire à ces ruptures. Leur détection permet de mettre en place des politiques de maintenance prévisionnelles ou prédictives des infrastructures. Toutefois, le taux de bonnes classifications des défauts mineurs n'est pas optimal et des confusions subsistent entre certains défauts mineurs et de simples inhomogénéités du rail, appelées points singuliers ; en effet, étant sensible à toutes modifications des caractéristiques électromagnétiques de la cible, le capteur détecte également les soudures de rail, les éclisses, les appareils de voie [1].

Pour lever ce problème, des travaux concernant l'amélioration du taux de reconnaissance des points singuliers à l'aide de modélisations probabilistes, par réseaux bayésiens dynamiques à variable exogène, ont été réalisés dans le cadre d'une variable exogène discrète [2]. Ces modèles probabilistes particuliers ont l'avantage de prendre en compte, par le biais de la variable exogène (ou variable d'entrée), des informations a priori sur la succession à la voie des points singuliers. Pour notre application, le vecteur des distances séparant un point singulier de ses prédécesseurs, souvent riche en information, a constitué cette variable exogène [2]. La décision fournie par ce classifieur, une fois fusionnée avec une décision générée par le traitement des signaux d'un capteur a permis d'améliorer significativement la décision finale [4].

Cependant, l'étude précédemment menée [2] s'appuie sur un vecteur de distances discret, où les distances ont une précision de l'ordre du mètre alors que le classifieur local peut fournir une décision quasi continue (une mesure tous les millimètres). La fusion de ces deux décisions ne peut donc être réalisée que tous les mètres, conduisant à une grande perte d'informations. Cet article propose donc, sans changer le modèle de base choisi, une nouvelle stratégie permettant d'utiliser la variable exogène sous forme continue, permettant ainsi la mise en place d'une méthode de décision capable de répondre à tout instant.

La section suivante définit les réseaux bayésiens dynamiques à variable exogène retenus puis décrit la stratégie utilisée pour apprendre ces modèles quand la variable exogène est continue. Une étude expérimentale sur des données réelles est effectuée dans la troisième section.

### Réseaux bayésiens dynamiques à variable exogène continue

Dans cette section, nous rappelons brièvement les réseaux bayésiens dynamiques à variable exogène déjà introduits dans l'article [2], puis détaillons la stratégie utilisée pour apprendre ces modèles dans le cas d'une variable exogène continue.

### Réseaux bayésiens dynamique à variable exogène

Nous supposons dans ce travail que  $(x_t)$  est la séquence d'événements observée, chaque variable de cette séquence pouvant prendre n'importe quel état dans l'ensemble  $\Omega = \{1, \dots, N\}$ . On parlera alors de réseau bayésien dynamique à variable exogène d'ordre  $m$  si l'état d'une réalisation  $x_t$  du processus étudié est influencé uniquement par les  $m$  événements observés précédemment et par une variable exogène  $u_t$  qui dans notre situation [2] désigne le vecteur  $(d_{t-1}, \dots, d_{t-m})$  où  $d_{t-i}$  est la distance entre les détections  $x_{t,i}$  et  $x_{t,i-1}$ . Nous considérons uniquement les réseaux bayésiens dynamiques à variable exogène d'ordre 1 et d'ordre 2 qui ont été retenus lors de l'étude précédente [2] (voir figures 1 et 2).

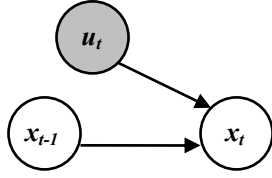


Fig. 1. Réseau bayésien dynamique d'ordre 1 à variable exogène

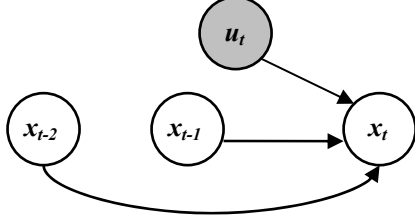


Fig. 2. Réseau bayésien dynamique d'ordre 2 à variable exogène

L'utilisation d'un espace discret pour les distances séparant deux événements consécutifs de notre séquence limite l'utilisation du modèle. Souhaitant, dans le cadre de notre application, faire collaborer ce modèle de décision avec celui utilisant les signaux d'un capteur, il s'avère nécessaire que les distances puissent appartenir à un espace continu.

Nous présentons dans la section suivante une méthode d'apprentissage permettant d'étendre l'approche discrète [2] au cas d'une variable exogène continue.

### Apprentissage des tables de probabilité conditionnelles

Dans cette partie, nous rappelons d'abord comment les tables de probabilité conditionnelles sont estimées dans le cas d'une variable exogène discrète puis proposons une méthode pour estimer ces tables dans le cas d'une variable exogène continue. Nous n'explicitons les estimations que pour le modèle d'ordre 1, le cas du modèle d'ordre 2 étant similaire.

L'apprentissage des tables de probabilités conditionnelles (TPC) s'effectue par la méthode du maximum de vraisemblance, à partir d'une séquence d'événements de longueur fixée. Il nécessite l'estimation des probabilités conditionnelles

$$A_{ij}(\mathbf{u}) = P(x_t = i \mid x_{t-1} = j, u_t = \mathbf{u}), \quad [1]$$

où  $i, j \in \Omega$ .

Dans le cas où la variable d'entrée est discrète, ces estimations sont données par

$$\hat{A}_{ij}(\mathbf{u}) = \frac{n_{ij}(\mathbf{u})}{\sum_{i=1}^N n_{ij}(\mathbf{u})} \quad [2]$$

où  $n_{ij}(\mathbf{u})$  désigne le nombre d'occurrences du couple  $(x_t=i, x_{t-1}=j)$  dans la séquence d'apprentissage, vérifiant la contrainte  $u_t=\mathbf{u}$ . Il faut noter que cette approche ne peut être implémentée pratiquement que dans d'une variable  $u_t$  discrète, à valeurs dans un espace au cardinal limité.

Pour une variable exogène continue, l'estimation de  $A_{ij}(\mathbf{u})$  dépend de la nature de la densité de  $u_t$ , conditionnellement à  $x_{t-1}=i$  et  $x_{t-1}=j$ . En se référant à la règle de Bayes,

$$\begin{aligned} A_{ij}(\mathbf{u}) &= P(x_t = i \mid x_{t-1} = j, u_t = \mathbf{u}) \\ &= \frac{P(x_t = i \mid x_{t-1} = j)P(\mathbf{u} \mid x_t = i, x_{t-1} = j)}{\sum_i P(x_t = i \mid x_{t-1} = j)P(\mathbf{u} \mid x_t = i, x_{t-1} = j)}, \end{aligned} \quad [3]$$

où  $P(\mathbf{u} \mid x_t = i, x_{t-1} = j)$  est la densité de  $u_t$  conditionnellement à  $x_t = i$  et  $x_{t-1} = j$ , nous proposons l'estimation

$$\hat{A}_{ij}(\mathbf{u}) = \frac{n_{ij}\hat{P}(\mathbf{u} \mid x_t = i, x_{t-1} = j)}{\sum_i n_{ij}\hat{P}(\mathbf{u} \mid x_t = i, x_{t-1} = j)}, \quad [4]$$

où  $n_{ij}$  désigne le nombre d'occurrences du couple  $(x_t=i, x_{t-1}=j)$  et  $\hat{P}(\mathbf{u} \mid x_t = i, x_{t-1} = j)$  est une estimation de la densité de  $u_t$

conditionnellement à  $x_t=i$  et  $x_{t-1}=j$ . Cette dernière estimation peut être obtenue, par exemple, par une méthode d'estimation paramétrique de densités. L'approche qui vient d'être décrite nécessite donc l'estimation séparée de  $N^2$  densités pour le modèle d'ordre 1 et plus généralement l'estimation de  $N^{m+1}$  densités pour les modèles d'ordre  $m$ .

Nous décrivons maintenant la méthode utilisée pour l'estimation des densités conditionnelles de la variable exogène.

### Estimation des densités conditionnelles de la variable d'entrée

Dans le cadre de notre application, la présence de plusieurs pics observés sur les différents histogrammes de la variable exogène (cf. figure 3 et 5) nous a conduit à considérer qu'elle pouvait être distribuée suivant un mélange de densités gaussiennes [7]. Notons

$$P(\mathbf{u}; \Theta) = \sum_{k=1}^K \pi_k P_k(\mathbf{u}; \mu_k, \Sigma_k) \quad [5]$$

la densité de  $u_t$  conditionnellement à  $x_t=i$  et  $x_{t-1}=j$ , où  $K$  est le nombre de composants du mélange, les  $\pi_k$  sont les proportions du mélange et  $P_k$  est la densité gaussienne de moyenne  $\mu_k$  et de matrice de covariance  $\Sigma_k$ ;  $\Theta = (\pi_1, \dots, \pi_K, \mu_1, \dots, \mu_K, \Sigma_1, \dots, \Sigma_K)$  désigne le vecteur paramètre du modèle. L'estimation du vecteur paramètre  $\Theta$  s'effectue ensuite à partir des observations  $u_t$  pour lesquelles  $x_t=i$  et  $x_{t-1}=j$  que l'on note, pour simplifier,  $u_1, \dots, u_n$ . La maximisation de la fonction de log-vraisemblance

$$L(\Theta) = \log \prod_{t=1}^n P(u_t; \Theta) \quad [6]$$

étant difficile à effectuer analytiquement, nous utilisons l'algorithme EM [5] pour effectuer cette maximisation. Cet algorithme s'appuie alors sur des données dites 'complétées' dont l'observation aurait simplifié la maximisation de la vraisemblance. On utilise, dans la situation des modèles de mélange, le vecteur  $((u_1, z_1), \dots, (u_n, z_n))$ , où les  $z_t$  sont les labels d'origine des  $u_t$ , comme données complétées. Chaque itération de l'algorithme EM se résume alors en deux étapes qui sont les suivantes.

**L'étape E**: qui consiste à calculer l'espérance de la vraisemblance complétée conditionnellement aux données observées et au vecteur paramètre courant  $\Theta^{(q)}$

$$Q(\Theta; \Theta^{(q)}) = E[L_C(\Theta) \mid u_1, \dots, u_n; \Theta^{(q)}], \quad [7]$$

où  $L_C(\Theta)$  est la fonction de vraisemblance des données complétées. Cette étape nécessite simplement de calculer les probabilités a posteriori

$$r_{tk}^{(q)} = \frac{\pi_k^{(q)} P_k(u_t; \mu_k^{(q)}, \Sigma_k^{(q)})}{\sum_k \pi_k^{(q)} P_k(u_t; \mu_k^{(q)}, \Sigma_k^{(q)})}. \quad [8]$$

**L'étape M**: qui consiste à calculer le vecteur paramètre  $\Theta^{(q+1)}$  qui maximise l'espérance  $Q(\Theta; \Theta^{(q)})$ . Cette maximisation fournit les paramètres

$$\begin{aligned} \pi_k^{(q+1)} &= \frac{\sum_{t=1}^n r_{tk}^{(q)}}{n}, \\ \mu_k^{(q+1)} &= \frac{\sum_{t=1}^n r_{tk}^{(q)} u_t}{\sum_{t=1}^n r_{tk}^{(q)}}, \\ \Sigma_k^{(q+1)} &= \frac{\sum_{t=1}^n r_{tk}^{(q)} (u_t - \mu_k^{(q+1)})(u_t - \mu_k^{(q+1)})^T}{\sum_{t=1}^n r_{tk}^{(q)}}. \end{aligned} \quad [9]$$

Il peut arriver, comme cela a été le cas dans notre application, que des singularités apparaissent au cours des itérations de l'algorithme EM (déterminant de matrice devenant nul). C'est particulièrement le cas lorsqu'il y a peu de données par rapport au nombre de composants du mélange. En s'inspirant des travaux proposés par Fraley et Raftery [6] dans le cadre des régularisations bayésiennes, nous proposons d'introduire une distribution a priori de Wishart inverse sur chaque matrice de

covariance. Cela conduit à la nouvelle formule de mise à jour de la matrice de covariance,

$$\Sigma_k^{(q+1)} = (1 - \alpha) \frac{\sum_{t=1}^n r_{tk}^{(q)} (u_t - \mu_k^{(q+1)}) (u_t - \mu_k^{(q+1)})^T}{\sum_{t=1}^n r_{tk}^{(q)}} + \alpha M, \quad [10]$$

où  $M$  est une matrice symétrique définie positive et  $\alpha$  est un réel positif compris entre 0 et 1. Cette nouvelle formule nous permet d'avoir des matrices de covariance dont le déterminant ne dépasse pas un seuil défini par  $M$  et  $\alpha$ . Pour notre application, nous avons retenu  $M$  comme étant la matrice de covariance de l'échantillon et fixé  $\alpha$  à 0.001.

Le choix du nombre de composantes  $K$  de la densité mélange  $P(u; \Theta)$  est estimé en maximisant le critère de sélection de modèles BIC (Bayesian Information Criterion) [9] qui est très souvent utilisé pour estimer le nombre de composantes des mélanges. Ce critère est défini par

$$BIC(K) = L(\hat{\Theta}^*) - \frac{v_K}{2} \log(n), \quad [11]$$

où  $\hat{\Theta}^*$  est l'estimation fournie par l'algorithme EM et  $v_K$  est le nombre de paramètres libres de la densité mélange. Un autre problème fondamental dans l'application de l'algorithme EM est celui de sa convergence vers un maximum local de la log-vraisemblance pour des paramètres initiaux donnés. Pour pallier ce problème, la stratégie retenue consiste à lancer l'algorithme à partir de plusieurs initialisations aléatoires différentes puis à retenir la solution fournissant la meilleure vraisemblance [3].

### Etude expérimentale

Cette section décrit les données utilisées pour l'expérimentation de la méthode proposée. Une description du protocole d'expérimentation et la comparaison des résultats obtenus avec ceux de l'approche s'appuyant sur une variable exogène discrète, proposée dans une étude précédente [2], sont ensuite effectuées.

#### Description des données

Les données utilisées dans cette étude proviennent d'une base de données de la RATP. Cette base regroupe, entre autres, la position et la nature des points singuliers (soudures, éclisses, appareils de voie) répertoriés lors d'opérations de pose et de maintenance du rail. Nous disposons ainsi d'environ 22000 points (18387 soudures, 2203 éclisses et 289 appareils de voie) pour huit lignes du métro parisien.

Pour notre application, le nombre d'états possibles de la variable  $x_t$  étant égal à  $N=3$  (soudure, éclisse ou joint d'appareil), l'apprentissage de  $3^2=9$  densités conditionnelles de la variable  $u_t$  pour le réseau bayésien dynamique d'ordre 1 et de  $3^3=27$  densités conditionnelles pour le réseau bayésien dynamique d'ordre 2 a été nécessaire. Il convient de noter que la variable  $u_t$  est monodimensionnelle dans le modèle d'ordre 1 et bidimensionnelle dans le modèle d'ordre 2.

L'histogramme des densités de la variable  $u_t$  conditionnellement à  $(x_t=1, x_{t-1}=1)$  justifie le choix du modèle de mélange Gaussien (voir figures 3, 4, 5 et 6). La réalité de ce mélange correspond à la réalisation privilégiée des soudures tous les 18 mètres (longueur des coupures de rails neufs) ou autour de 6 mètres en cas de réparation.

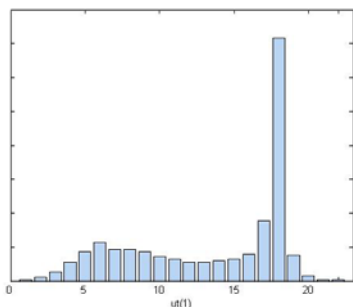


Fig. 3. Histogramme de la variable d'entrée  $u_t$  conditionnellement à  $x_t=x_{t-1}=1$  (les détections aux instants  $t$  et  $t-1$  sont des soudures)

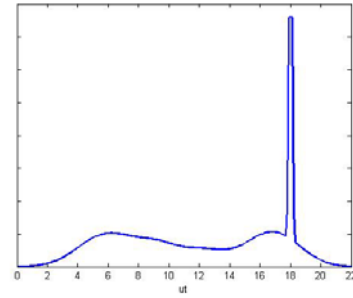


Fig. 4. Densité mélange Correspondant à l'histogramme de la figure 3.

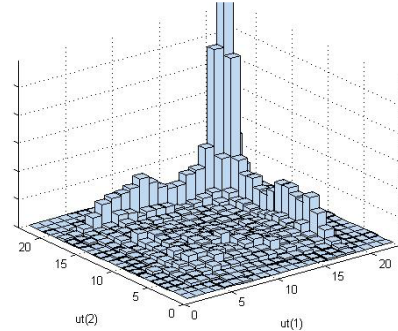


Fig. 5. Histogramme de la variable d'entrée  $u_t$  conditionnellement à  $x_t=x_{t-1}=x_{t-2}=1$  (les détections aux instants  $t$ ,  $t-1$  et  $t-2$  sont des soudures).

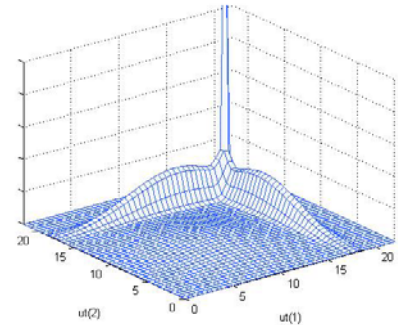


Fig. 6. Densité mélange correspondant à l'histogramme de la figure 4

L'apprentissage des densités conditionnelles de la variable  $u_t$  puis des tables de probabilités conditionnelles, ainsi que les tests de prédiction ont été réalisés sur la base de données ainsi définie. La section suivante décrit la procédure d'expérimentation.

#### Mise en œuvre

Les expérimentations effectuées sur la base de données ont d'abord consisté à apprendre les tables de probabilités conditionnelles, fonctions de la variable d'entrée continue, puis à utiliser le modèle ainsi appris pour prédire la nature des nouvelles détections. Dans le cadre de notre application, nous avons choisi un ensemble d'apprentissage regroupant l'ensemble des données des huit lignes de métro à disposition. Les tests ont ensuite été effectués sur chacune des huit lignes successivement.

L'estimation des tables de probabilités conditionnelles s'est faite en deux étapes qui sont :

- l'estimation des  $N^2$  (resp.  $N^3$ ) densités mélange de la variable  $u_t$  relatives au modèle d'ordre 1 (resp. au modèle d'ordre 2) par l'algorithme EM avec sélection du nombre de composantes de chaque mélange par le critère BIC,
- l'estimation des tables de probabilités conditionnelles.

La décision finale relative au choix de la classe d'appartenance de la détection courante est prise en maximisant les  $P(x_t=i|u_t)$  par rapport à  $i$ .

## Résultats

Avant de comparer les résultats de classification des approches discrètes et continues, le tableau 1 présente le nombre de paramètres à identifier pour notre problème, dans le cas des modèles d'ordre 1 ( $m=1$ ) et d'ordre 2 ( $m=2$ ). Si on note  $\mathcal{U}$  l'espace de la variable d'entrée  $u_t$ , ce nombre vaut  $\text{card}(\mathcal{U}) \cdot N^{m+1}$  dans le cas discret et  $((K-1) + K \cdot m + 0.5 \cdot K \cdot (K+1) \cdot m) \cdot N^{m+1}$  dans le cas continu, en supposant que tous les mélanges sont constitués de  $K$  gaussiennes. On constate l'intérêt évident de l'implémentation continue pour  $m > 1$  en termes de parcimonie.

	Modèle d'ordre 1	Modèle d'ordre 2
Cas discret	198	13068
Cas continu	576	1728

Tab. 1 Nombre de paramètres à estimer avec  $K=10$  et pour les modèles d'ordre 1 et 2, par les deux approches discrète et continue.

Le tableau 2 représente le taux de bonne prédiction de chacun des types des points singuliers, obtenu sur l'ensemble des huit lignes de métro avec le modèle d'ordre 2, pour les approches discrète et continue.

Classes	1(JS)	2(JE)	3(JA)
Taux de bonnes détections (discret)	97.4	53.7	78.5
Taux de bonnes détections (continu)	97.5	50.3	77.2

Tab. 2 Taux de bonnes classifications obtenu sur l'ensemble des huit lignes avec les approches discrète et continue, pour le modèle d'ordre 2.

Les résultats fournis par les deux approches semblent globalement identiques. Le taux de bonnes détections des soudures (JS) est quasiment le même pour les deux approches. Cela peut s'expliquer par le fait que les soudures, qui sont très bien représentées dans la base de données (18387 soudures), conduisent à un apprentissage très précis de certaines densités mélange. Les taux de bonnes détections des éclisses (JE) et des appareils de voie (JA) sont sensiblement plus faibles avec l'approche continue mais restent raisonnables. Cela est sans doute dû à la sous représentation de ces types de points singuliers dans la base de donnée d'apprentissage et au fait que les données d'apprentissage utilisées ont une précision de succession discrète (de l'ordre du mètre) qui favorise l'approche discrète.

## Conclusion

Cet article a présenté l'implémentation d'un réseau bayésien dynamique avec prise en compte d'une variable exogène continue, dans un contexte de classification de points singuliers de l'infrastructure ferroviaire. Les tables de probabilités de ce modèle sont estimées à l'aide d'une stratégie s'appuyant sur des mélanges de lois gaussiennes estimées par l'algorithme EM. Les tests, effectués sur une base de données réelles énumérant les points singuliers des rails du réseau de métro parisien, ont mis en évidence de bonnes performances de la méthode en comparaison avec l'approche du réseau bayésien dynamique à variable exogène discrète. Outre le bon comportement de la méthode proposée, la parcimonie de l'approche est démontrée dès que l'ordre du modèle est supérieur ou égal à deux. Cette implémentation permettra à terme de fusionner efficacement la décision d'un capteur spécifique de détection de défaut de rail avec celle du réseau bayésien et de lever ainsi certaines confusions du capteur.

## Remerciements

Les auteurs remercient le service EST-Voie de la RATP, et en particulier D. Levy, pour son concours au volet expérimental du projet.

## Références

- [1] P. Aknin, M. Bentoumi, F. Fessant, F. Raux. Chaîne d'instrumentation embarquée pour la détection temps réel de défauts de rails débouchants. REE. Janvier 2001.
- [2] A. Ben Salem, L. Bouillaut, P. Aknin, D. Levy. Utilisation de Réseaux Probabilistes Causaux pour l'Amélioration de la Maintenance des Infrastructures Ferroviaires. 14<sup>ème</sup> Congrès de Maîtrise des Risques et de Sécurité de Fonctionnement  $\mu 14$ , Bourges, France. Octobre 2004.
- [3] C. Biernacki, G. Celeux, G. Govaert. Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate gaussian mixture models. Computational Statistics and Data Analysis. 41 : 561-575, 2003.
- [4] E. Côme, L. Bouillaut, P. Aknin, A. Samé. Spatial structures resuming by bayesian network and fusion with diagnosis data in the transferable belief model framework. The 11th Information Processing and Management of Uncertainty in Knowledge-Based Systems International Conference (IPMU 2006).
- [5] A. P. Dempster, N. M. Laird, D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. Journal of Royal Statistical Society Series B 39(1), 1-38. 1977.
- [6] C. Fraley, A. Raftery. Bayesian Regularization for Normal Mixture Estimation and Model-Based Clustering. Technical Report, Department of Statistics, University of Washington, 486, 2005.
- [7] D. M. Titterton, A. Smith, U. E. Makov. Statistical analysis of finite mixture distributions. Wiley and Sons, Inc., New York. 1985.
- [8] L. Oukhellou, P. Aknin, J-P Perrin. Dedicated sensor and classifier of rail head defects for railway systems. Control Engineering Practice 7, 57-61. 1999.
- [9] Schwarz. Estimating the dimension of a model. Ann. Stat. 6, 461-464. 1978.

