

## TRAITEMENT DE DONNEES SANITAIRES PAR ARBRES DE REGRESSION FLOUS\*

**Marie Lahaye<sup>1</sup>, Claire Chauvin<sup>2</sup>, Coralie Lupo<sup>2</sup>  
et Pierre-Yves Glorennec<sup>1</sup>**

**RESUME :** Le traitement de données sanitaires, en vue d'identifier des marqueurs ou facteurs de risque et de définir des règles de décision, fait usuellement appel à des méthodes d'analyse multivariées, telles que la segmentation, la régression linéaire ou la régression logistique. Une autre méthode, fondée sur la logique floue, est ici proposée pour construire des arbres de régression aisément interprétables à partir de données collectées en élevage. Les arbres de régression flous combinent la facilité d'interprétation des arbres de décision booléens et l'aspect graduel (sans seuil) de la logique floue. Une illustration de l'application de cette méthode est donnée au travers de la recherche de marqueurs/facteurs « explicatifs » de la saisie des carcasses de dindes en abattoir, à partir des données collectées lors d'une étude conduite en 2001 par le GDS 22, dans le cadre du contrat de plan Etat- région.

**Mots-clés :** Arbres de régression flous, dindes, saisies sanitaires.

**SUMMARY :** Identification of risk factors and decision rules usually involves multivariate data analysis methods, such as segmentation, linear or logistic regression. Another method is proposed here to build easily interpretable regression trees from data collected on-farm. The fuzzy regression trees combine a facility of interpretation and the absence of threshold of fuzzy logic. An illustration of an application of this method is given with the analysis of risk factors for turkey carcasses condemnation.

**Keywords :** Fuzzy regression tree, turkey, condemnation.



### I - INTRODUCTION

Les méthodes de classification, qui ont pour objectif d'identifier les classes d'appartenance d'observations, sont très adaptées au problème de la prise de décision automatisée ou assistée, particulièrement dans le domaine médical sous la forme d'aide au diagnostic à partir d'éléments cliniques. L'application associant à tout malade un diagnostic à partir d'une liste de symptômes peut être développée par une approche de type « système experts ». Des règles formalisant la connaissance d'experts sont alors établies. Elles sont cependant dépendantes des

personnes consultées et de l'évolution des connaissances. Une autre approche, fondée sur l'utilisation d'un ensemble d'exemples à partir duquel est extraite la procédure de classification, est aujourd'hui privilégiée. Plusieurs méthodes d'apprentissage à partir de jeux de données complets existent. Parmi elles sont distinguées les méthodes symboliques, qui permettent d'écrire sous forme de règles la procédure de classification établie, et dont fait partie la méthode de classification par arbres de régression flous.

\* Texte de la communication présentée à la Journée AEEMA, 20 mai 2005

Avec l'aide et la collaboration de Mahé Félix - GDS 22, Zoopole, 22440 Ploufragan, France

<sup>1</sup> Institut national des sciences appliquées de Rennes, 20 Av. des buttes de Coësmes, CS 14315, 35043 Rennes, France

<sup>2</sup> Agence française de sécurité sanitaire des aliments – site de Ploufragan, BP 53, 22440 Ploufragan, France

Afin d'extraire une telle règle de classification d'un jeu de données relatives au taux de saisie de carcasses de volailles à l'abattoir, à partir d'éléments d'informations portant sur les conditions d'élevage, les arbres de régression

flous ont été utilisés. Dans une première partie, le point sera fait sur les arbres de régression flous. Leur application à la recherche de marqueurs de la saisie de carcasses de dindes en abattoir sera ensuite exposée.

---

## II - LES ARBRES DE REGRESSION FLOUS

---

### 1. LES ARBRES DE REGRESSION

Les arbres de régression sont des outils de classification puissants dont le but est de produire un modèle proche du réel. Un jeu de variables (ou **attributs**) est utilisé pour modéliser les sorties sous forme de classes. L'arbre est immédiatement interprétable pour l'œil humain, il est facile d'en déduire des règles de classification (une règle étant une suite de modalités menant de la racine à une feuille de l'arbre selon une lecture SI... ET... ALORS...). Toutes les branches n'ont pas forcément la même longueur et l'arbre peut être dissymétrique. L'idée centrale de la construction des arbres est de diviser récursivement et le plus efficacement possible les exemples de l'ensemble d'apprentissage à l'aide des attributs (et par des tests définis) jusqu'à obtenir des sous-ensembles d'exemples presque tous d'une même classe. Cette procédure de classification est particulièrement adaptée aux domaines d'application où la compréhension de la classification par l'utilisateur est nécessaire. C'est particulièrement vrai dans les domaines de l'aide au diagnostic médical ou de l'analyse prévisionnelle de risques (comportements biologiques ou risques financiers en cas de prêts bancaires).

### 2. LA LOGIQUE FLOUE

Alors que la logique classique, dite logique booléenne, n'envisage que deux états exclusifs *vrai* ou *faux*, la logique floue autorise tous les états intermédiaires. Une propriété pourra ainsi être presque vraie (vraie à 80% et fautive à 20%). Cette notion d'appartenance à un état, est plus souple que la logique booléenne et très proche du langage naturel. Elle permet de manipuler des variables continues sans avoir à les catégoriser par des seuils arbitraires. Ainsi, une température est en logique booléenne considérée comme *basse* ou *élevée* respectivement en deçà et à partir d'un seuil fixé par exemple à 14°C. Ceci implique que deux températures distinctes comme 2°C et 13,9°C appartiendront toutes

deux à la classe *basse*, en revanche 14,1°C sera considérée comme appartenant à la classe *élevée* alors même que cette valeur est très proche de 13,9°C. En logique floue, l'introduction d'un **degré d'appartenance** (une valeur comprise entre 0 et 1) d'une variable à un ensemble, par une fonction, appelée **fonction d'appartenance**, permet de nuancer un tel jugement. L'étendue des valeurs possibles d'une variable est pour cela subdivisée en intervalles se chevauchant, définissant chacun un sous-ensemble de valeurs dénommé **sous-ensemble flou**, identifié par un terme naturel tel que *élevé* ou *bas*. Les limites des sous-ensembles flous des variables quantitatives sont appelées **valeurs modales**. La figure 1 représente ainsi un exemple de **partition floue**. La valeur T appartient au sous-ensemble « *élevée* » avec un degré d'appartenance 0,25 et au sous-ensemble « *normale* » avec un degré d'appartenance 0,75. Les variables utilisées telle que la variable « température » sont appelées des variables linguistiques (i.e. elle peut prendre les valeurs « *basse* », « *normale* », « *élevée* »).

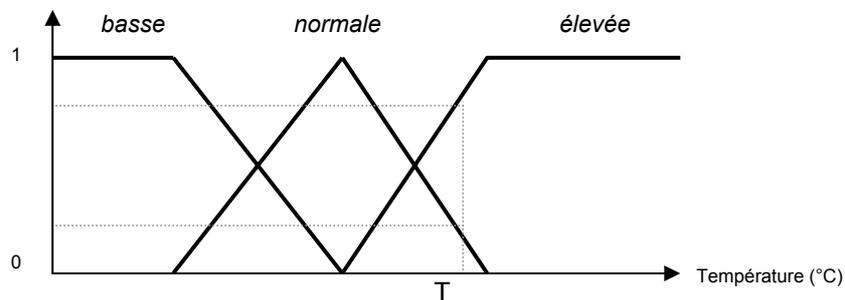
Afin de construire des arbres de régression flous, les notions d'**entropie** et de **gain** utilisées pour la construction d'arbres de décision ont été étendues à la logique floue [Glorennec, 2003].

### 3. WOMBAT OUTIL DE CONSTRUCTION D'ARBRES DE REGRESSION FLOUS

Wombat (pour Wombat obviously makes beautiful asymmetric trees) est un outil développé pour permettre la construction d'arbres de régression flous sous un environnement Scilab. L'utilisateur peut définir graphiquement les fonctions de partition des variables, visualiser les données d'apprentissage et de test. Il peut visualiser et développer l'arbre de régression de manière progressive et volontaire en s'aidant des informations relatives au gain à un nœud, à l'erreur quadratique issue du jeu de données test, selon le niveau de complexité désiré.

Figure 1

## Exemple de partition floue forte d'un ensemble de températures




---

**III - APPLICATION A L'ANALYSE DE DONNEES SANITAIRES**


---

**1. NATURE DES DONNEES**

Les données sont issues d'une étude conduite en 2000-2001 par le Groupement de défense sanitaire avicole des Côtes-d'Armor sur les saisies en abattoir de dindes en région Bretagne, dans le cadre d'un contrat de plan Etat - Région du Comité régional avicole de Bretagne [Anonyme, 2001 ; Anonyme, 2002].

Entre mai 2000 et juin 2001, 58 lots de dindes femelles et 58 lots de dindes mâles élevés en Bretagne ont été suivis depuis la mise en place des poussins dans les bâtiments d'élevage jusqu'à leur abattage.

En élevage, un questionnaire à la livraison des animaux a permis de recueillir des éléments relatifs aux bâtiments (type de ventilation, type de litière, nombre de points d'abreuvement, d'alimentation) et aux pratiques d'élevage (densité au départ, intensité lumineuse, quantité de litière mise en place). Des pesées individuelles ont été réalisées à l'arrivée des animaux, puis à 10 jours. En fin d'élevage, les éléments relatifs à la mortalité, aux événements sanitaires (nature des événements, nature et coûts des thérapeutiques associées), aux résultats technico-économiques (poids moyen, indice de consommation, gain moyen quotidien) du lot suivi ont été relevés.

A l'abattoir, les saisies totales et partielles de chacun des lots ont été caractérisées et classées par motif (lésions cutanées infectées, lésions traumatiques, cachexie, congestion, arthrite synovite, gros jabot, ampoules du bréchet, aérosacculite, salpingite, autres) puis

quantifiées en utilisant le pourcentage de poids de carcasse représenté par classe.

**2. ANALYSE CONDUITE**

L'ensemble des éléments recueillis et transmis a été intégré dans l'étude en considérant les taux de saisie pour les différents motifs comme des variables continues « à expliquer », les différents paramètres d'élevage relevés (quantitatifs et qualitatifs) étant considérés comme des variables explicatives potentielles.

Compte tenu de l'influence marquée du sexe des animaux sur les valeurs observées des variables, chaque analyse a été conduite séparément pour les lots de mâles et les lots de femelles.

Préalablement à la construction des arbres, un tri a été effectué entre variables explicatives afin de ne pas considérer des variables synonymes ou fortement corrélées. Les lots présentant des données aberrantes ou manquantes ou constituant à eux seuls des cas particuliers ont été exclus de l'analyse.

Les valeurs modales des variables quantitatives ont été choisies de manière à constituer des intervalles réguliers dans l'étendue des valeurs prises.

**3. RESULTATS OBTENUS**

L'arbre de régression flou obtenu pour le motif aérosacculite chez les lots de dindes mâles sera ici détaillé afin d'illustrer l'application de la méthode.

La figure 2 représente l'arbre de régression flou obtenu pour le motif aérosacculite sur le jeu de données récoltées chez les dindes mâles. La ventilation intervient en premier lieu. Si la ventilation est statique alors un très faible taux de saisie est observé ; si la ventilation est de type dynamique, alors un deuxième attribut intervient : la litière. Il y a alors plus de saisie sur paille que sur copeaux. Les lots sur paille broyée, plus nombreux, ont permis de développer l'arbre avec le facteur « nombre de jours d'acidification ». L'adjonction d'acidifiant à l'eau de boisson contribue à la maîtrise des désordres digestifs : un nombre important de jours d'acidification est associé à un fort taux de saisie pour aérosacculite. Les valeurs modales de cette variable ont été fixées à 2 (appartenance à 100% à la classe « peu » en deçà de deux jours d'acidification), 10 et 18.

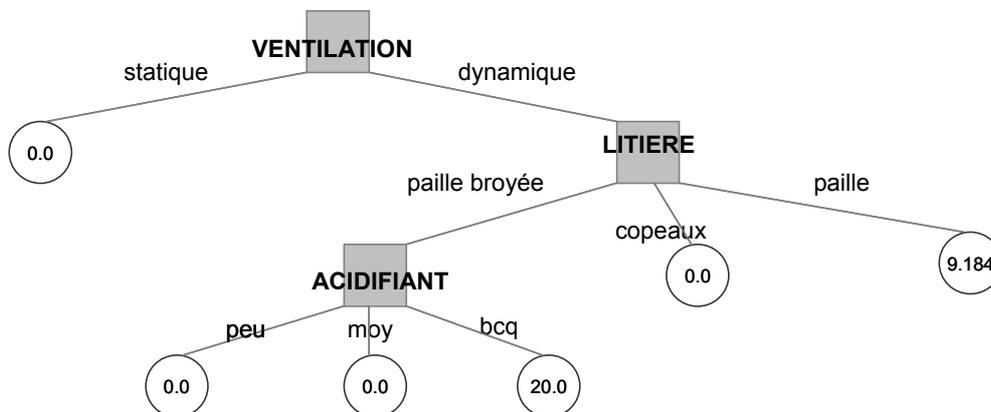
Chaque conclusion de l'arbre est affectée d'une valeur correspondant au taux de saisie

des élevages de la branche (chiffres entourés dans la figure 2). Le taux a été fixé à très faible pour une valeur de 0, faible pour 5, moyen pour 10, élevé pour 15, très élevé pour 20. Toutes les valeurs intermédiaires correspondent à un taux de saisies appartenant *un peu* à chacune des classes adjacentes.

Quel que soit le motif considéré, la nature de la litière est un attribut intervenant dans les arbres. Ainsi, par exemple dans l'analyse des saisies pour lésions cutanées infectées, quel que soit le sexe, la nature de la litière, le nombre de points d'abreuvement, la supplémentation en acidifiant et la luminosité interviennent ; les modalités élevées de chacun de ces deux derniers paramètres étant associées à de forts taux de lésions cutanées infectées.

Figure 2

Arbre de régression flou obtenu pour le motif aérosacculite




---

## IV - DISCUSSION

---

La recherche d'associations entre caractéristiques d'élevage et résultat de l'inspection *post-mortem* des carcasses doit permettre d'identifier des facteurs de risque de saisie des carcasses pour permettre leur correction et l'amélioration qualitative et économique de la production [Tielen *et al.*, 1978]. Cette recherche répond aussi à la nécessité d'une meilleure valorisation des données d'élevage dans le cadre de l'inspection *ante-mortem* des lots [Fries, 2000].

Les méthodes d'analyse multivariée usuelles en épidémiologie ont à ce jour été utilisées dans cet objectif [Farver *et al.*, 1981 ; Habtemariam et Cho, 1983 ; Dahms *et al.*, 1999 ; Edwards *et al.*, 1999 ; Dahms, 2004].

Les arbres de régression flous présentent les avantages de ne pas nécessiter la fixation de seuils stricts *a priori* pour la transformation ou l'interprétation des variables quantitatives et de

produire une règle d'interprétation visuelle aisément compréhensible et applicable.

Compte tenu du faible nombre d'observations, il n'a pas été possible de scinder les données en un jeu d'apprentissage et un jeu de test. En outre, la nature et le nombre d'observations et de variables recueillies dans l'étude, ont conduit à l'élaboration d'arbres de taille modeste et ne comprenant que quelques attributs. Ainsi, les lots sur copeaux et paille étaient peu nombreux, les arbres ont donc surtout été développés pour la modalité paille broyée. L'application illustre cependant l'intérêt de la méthode adoptée pour établir des règles graphiques facilement compréhensibles, utilisables et assorties d'une conclusion « quantifiée ».

L'analyse des saisies pour aérosacculite montre clairement l'importance du type de ventilation. Dans l'échantillon étudié, une ventilation statique est associée à un faible taux de saisie pour ce motif. Une association de même nature entre le type de ventilation et la présence de lésions nécropsiques d'aspergillose a aussi été rapportée [Anonyme, 2001 ; Anonyme, 2002]. La litière apparaît comme une variable explicative majeure des taux de saisie tant en lots de mâles que de femelles et ce, pour plusieurs motifs de saisie. Une influence de la structure de l'échantillon étudié sur ce résultat ne peut être exclue, compte tenu du faible nombre de lots et de l'occurrence faible de certaines modalités. Le

rôle de la litière est cependant en accord avec certaines observations de la littérature. Les litières irritantes comme la paille, susceptibles de léser l'intégrité du derme [Elfadil *et al.*, 1996a et c], peuvent favoriser la pénétration de pathogènes banals (*Escherichia coli*) dans l'organisme des volailles et le développement de troubles locomoteurs infectieux tels que les arthrites ; troubles locomoteurs eux-mêmes fréquemment observés lors d'autopsies de carcasses saisies pour cachexie [Anonyme, 2001 ; Anonyme, 2002]. Les saisies pour lésions cutanées infectées ont aussi été associées à la qualité de la litière, ainsi qu'au nombre de points d'abreuvement et à la luminosité. Ces résultats sont similaires à ceux d'autres études : le sous-équipement en points d'abreuvement et l'intensité lumineuse ont été en 1993 identifiés comme des facteurs spécifiquement associés aux lésions cutanées et défauts d'aspect des carcasses de dindes [Bouvarel *et al.*, 1996]. Un manque de points d'abreuvement favoriserait les griffures entre animaux au cours de luttes pour l'accès à l'eau, lésions s'infectant ensuite. De même, l'intensité lumineuse est connue pour favoriser l'agressivité des animaux.

Les arbres de régression flous apparaissent comme un outil adapté au traitement de données sanitaires afin d'identifier de possibles facteurs de risque tout en produisant des règles de classification applicables à de nouvelles données.

---

## BIBLIOGRAPHIE

---

Anonyme - Les saisies pour congestion et cachexie en élevages de dindes en Bretagne. Rapport Contrat de plan Etat-région volet santé qualité, 2001, 25pp.

Anonyme - Dindes : pour moins de saisies, surveiller Aspergillose et litière. *Filières avicoles*, 2002, avril, 67-69.

Dahms S., Bandick N., Fries R. - Associations between farm characteristics in fattening pigs and occurrence of lesions at meat inspection. *Berl. Munch. Tierarztl. Wochenschr.*, 1999, **112**, 46-51.

Dahms S. - Combination of variance components estimation, regression trees and logistic regression as a concept for exploring associations between animal husbandry conditions and lesions found at

slaughter. *Dtsch. Tierarztl. Wochenschr.*, 2004, **111**, 178-181.

Bouvarel I., Lauras B., Drouin P., Hibal N. - Nature et importance des défauts d'aspects des carcasses de dindes. *Sciences et Techniques Avicoles*, 1996, **16**, 36-45.

Edwards D.S., Christiansen K.H., Johnston A.M., Mead G.C. - Determination of farm-level risk factors for abnormalities observed during post-mortem meat inspection of lambs: a feasibility study. *Epidemiol. Infect.*, 1999, **123**, 109-119.

Elfadil A.A., Vaillancourt J.P., Meek A.H., Gyles C.L. - A prospective study of cellulitis in broiler chickens in southern Ontario. *Avian Dis.*, 1996a, **40**, 677-689.

- Elfadil A.A., Vaillancourt J.P., Meek A.H. - Farm management risk factors associated with cellulitis in broiler chickens in Southern Ontario. *Avian Dis.*, 1996b, **40**, 699-706.
- Farver T.B., Bello Cedeno E., McCapes R.H. - The relationship between various production factors, condemnation and downgrading in turkeys: factor analysis as a method of variable reduction. *Avian Dis.*, 1981, **25**, 463-478.
- Fries R. - Procedures in abattoirs and meat inspection. State of the discussion on the implementation of alternative systems. *Dtsch Munch Tierarztl Wochenschr*, 2000, **113**, 1-8.
- Glorennec, P.Y. Constrained optimisation of fuzzy decision trees, In : Interpretability Issues in Fuzzy Modeling, Casillas J., Cordon O., Herrera F., Magdalena L. (Eds.), Springer, 2003.
- Habtemariam T. et Cho Y. - A computer based decision-making model for poultry inspection. *Journal of the American Veterinary Medicine Association*, 1983, **183**, 1440-1446.
- Schruff C. et Blaha T. - A decision model for the risk-based meat inspection. Proceedings of the International Society for Animal Hygiene in-between congress, St Malo, 21-23 October 2004, pp. 477-478.
- Tielen M.J., Truijten W.T., vd Groes C.A., Verstegen M.A., de Bruin JJ., Conbey R.A. - Conditions of management and the construction of piggeries on pig-fattening farms as factors in the incidence of diseases of the lung and liver in slaughtered pigs *Tijdschr Diergeneeskd.* 1978, **103**, 1155-65.

