

# Vers la construction d'un observatoire des pratiques agricoles : gestion et propagation de l'imprécision des données agronomiques

Asma Zoghلامي\*,\*\* Karima Zayrit\*  
Cyril de Runz\*, Eric Desjardin\*, Herman Akdag\*\*

\*CRESTIC, Université de Reims Champagne-Ardenne, France  
{karima.zayrit, cyril.de-runz, eric.desjardin}@univ-reims.fr  
<http://crestic.univ-reims.fr>

\*\*LIASD, Université Paris VIII, France  
{asma.zoghلامي, herman.akdag}@ai.univ-paris8.fr  
<http://www.ai.univ-paris8.fr>

**Résumé.** L'un des objectifs d'Observox est de traiter et gérer l'imprécision des données agronomiques tant spatialement (parcelles agricoles) et quantitativement (quantités de produits disséminées) et de toujours associer une évaluation de la qualité aux données. Aussi, nous avons choisi le cadre théorique des ensembles flous. A partir d'un modèle conceptuel gérant l'imperfection, nous construisons une base de données gérant des entités spatiotemporelles imprécises appelées « entités agronomiques floues ». Cependant, ce choix de représentation rend possible le chevauchement des composantes spatiales entre entités. Dans ce cas, nous propageons l'imprécision du spatial vers le quantitatif à l'aide d'un opérateur de caractère additif qui prend en compte à la fois l'information spatiale et quantitative, et qui fournit une information quantitative locale et floue. Le système ainsi construit nous permet d'obtenir une représentation floue des quantités de produits phytosanitaires disséminés à chaque endroit du territoire étudié.

## 1 Introduction

Inscrit dans la démarche du développement durable, le projet AQUAL (CPER Champagne-Ardenne, France) a mis en évidence le besoin d'un observatoire des pratiques agricoles et de leur pression sur la qualité des ressources en eau au niveau du bassin de la Vesle (Desjardin et de Runz, 2009). Ce dernier, appelé Observox, exploite des données issues de sources hétérogènes (information cadastrale, données d'enquêtes, ...) afin de construire des données multivariées : composante spatiale, composante sémantique mais aussi des informations quantitatives sur les produits diffusés. Cela implique donc la gestion de données imparfaites et en particulier imprécises (spatialement et quantitativement) de par leur simplification et de par la confusion née de l'aspect multi-sources comme l'indique (Shi, 2010).

Ainsi, la gestion de données imparfaites est essentielle. Selon nous, cette gestion doit s'effectuer de la modélisation du système d'information, à la construction de la base de données, à son exploitation et son analyse. C'est le paradigme inhérent à la démarche proposée dans cet article. Pour cela, nous ancrons notre démarche dans la représentation de l'imprécision et dans l'exploitation des théories permettant de formaliser et de traiter de telles données.

Dans ce cadre les modèles flous sont un bon choix pour représenter l'imprécision car ils fournissent un cadre souple et unique permettant de représenter l'imprécision et de mieux conceptualiser la réalité. Aussi, comme l'objectif de notre système d'information est de fournir des informations interprétables pour chaque endroit du territoire observé, notre choix s'est porté vers une représentation floue des données tant pour les informations spatiales que quantitatives. A l'aide de ces représentations, nous introduisons la notion d'entité agronomique floue.

Nous proposons dans cet article qu'à partir d'un modèle conceptuel défini à l'aide des pictogrammes proposés par Zoghlami et al. (2011), qui étendent le langage PVL (utilisé dans PERCEPTORY), de construire la base de données floues permettant de gérer nos entités agronomiques.

Cette théorie permet, notamment sur le plan spatial (mais aussi temporel), des chevauchements entre formes floues. Sur le plan spatial, il se pose alors la question de savoir quelle est la quantité de molécules impactantes diffusée à un endroit donné si plusieurs entités s'y chevauchent ? Pour cela nous exploitons la démarche introduite dans (Zayrit et al., 2011) permettant de définir une quantité floue de chaque information quantitative étudiée pour chaque endroit du territoire, et ce sous l'hypothèse que ces informations quantitatives soient additives.

Aussi nous souhaitons à travers cet article introduire, dans un premier temps (section 2), le concept d'entité agronomique floue et le modèle PERCEPTORY imparfait associé. Nous nous attacherons ensuite dans la section 3 à la construction de la base de données et aux mécanismes permettant la gestion de l'imprécision. La section 4 présentera la démarche visant à la propagation de l'imprécision du point de vue spatial vers le point de vue quantitatif. La dernière section (section 5) forme la conclusion de cet article.

## 2 Entité agronomique floue : concept et modèle

Les entités agronomiques combinent des informations spatiales, sémantiques, temporelles et quantitatives. Dans ce papier, l'objectif est l'observation des pratiques agricoles, sur le bassin agroindustriel de la Velse en France, et de leur impact sur la qualité de l'eau. Aussi, les principales informations quantitatives étudiées portent sur les quantités de produit ou de molécules.

Même si toutes les composantes de l'information agronomique peuvent être imprécises, nous nous concentrerons dans cet article sur les composantes spatiales et quantitatives. Les informations concernant les quantités de produits/molécules proviennent souvent d'enquêtes terrains auprès de agriculteurs dans lesquelles les données sont vraisemblablement approximées (*e.g.* « la quantité d'isoproturon diffusée sur ce champs est d'environ 0.5 kg/ha »).

D'un point de vue formel, une entité agronomique floue  $A_i$  sera construite à l'aide :

- Un label ou concept  $LA_i$  appartenant à une ontologie.
- Une forme spatiale  $FSA_i$  ; la fonction d'appartenance  $\mu_{SA_i}$  de  $FSA_i$  is defined on  $\mathbb{R}^2$ .

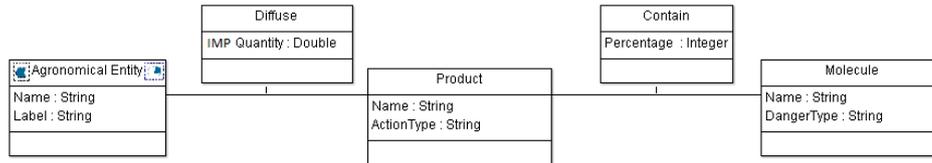


FIG. 1 – Diagramme de classes PERCEPTORY Imparfait représentant l'entité agronomique floue

– Une quantité floue  $F_{Q_j}^{A_i}$  pour chaque produit/molécule  $P_j$ ; la fonction d'appartenance  $\mu_{Q_j}^{A_i}$  de  $F_{Q_j}^{A_i}$  is defined on  $\mathbb{R}^+$ .  
 Les classes sous-jacentes à la définition de l'étude des pesticides utilisés sont : entité agronomique, produit, molécule.

Les informations imparfaites seront représentées dans le diagramme de classes PERCEPTORY Imparfait (figure 1) à l'aide des pictogrammes PVL entourés de pointillés ou par l'ajout du préfixe **IMP** avant l'attribut concerné selon la démarche de (Zoghliami et al., 2011) inspirée de (Bédard et al., 2004; Ma, 2008; Miralles et Libourel, 2008).

L'objectif de la phase modélisatrice est de créer le système d'information associé, ce qui implique aussi la création de la structure de gestion des données spatiotemporelles floues.

### 3 Base de données spatiotemporelles floues

Plusieurs approches ont été définies pour le passage du modèle UML à l'implémentation de la base de données floues. Dans le cadre d'approche spatiotemporelle de l'information, Zoghliami et al. (2011) proposent de structurer la base en trois couches liées entre elles : la première dédiée aux données, la seconde dédiée à la gestion des métaconnaissances, la troisième liée à la gestion de l'imperfection. Si cette approche est proche de celle proposée par Medina et al. (1995), le fait est que dans celle de Zoghliami et al., les métaconnaissances et les imperfections sont différenciées. Cependant, dans cet article, nous nous occupons essentiellement du lien entre la couche données et la couche imperfection.

L'idée générale est de stocker l'information et son imprécision, c'est à dire de stocker des données floues. A cette fin, une représentation multivalente est exploitée et permet de réduire les problèmes computationnels. En effet, ainsi, nous choisissons une discrétisation par  $\alpha$ -coupe des fonctions d'appartenance continues. L'idée est de stocker un certain nombre d'ensembles de valeurs possibles avec un niveau d'appartenance.

Chaque niveau est stocké avec sa valence ce qui permet de gérer des données décrites à des niveaux de détails et de valeurs de variances différents. Dans notre système, les données quantitatives seront considérées ayant un grand nombre de niveaux de valence, tandis que les données spatiales auront leur niveau limité en considération à la taille de stockage et à la pertinence des informations intermédiaires.

Dans la suite, nous nous attacherons à travailler sur la notion d'imprécision et de sa propagation du spatial vers le quantitatif.

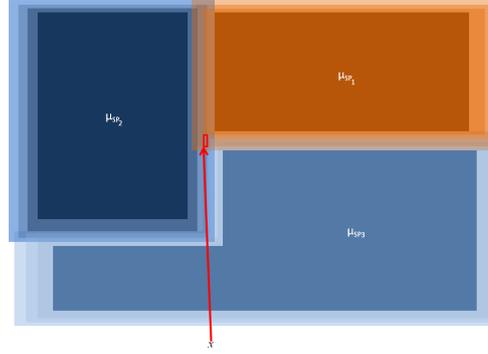


FIG. 2 – Présentation spatiale du territoire d'étude. Il est composé de trois entités agricoles floues ( $P_1$ ,  $P_2$  et  $P_3$ ) et une localisation  $x$ . Chaque parcelle floue est 4-valuée et est donc formée de 4 zones incluses les unes dans les autres. Les régions les plus sombres ont un degré d'appartenance égal à 1, les plus claires (les plus larges) ont un degré d'appartenance égal à 0.1. Les degrés d'appartenance des autres sont de 0.3 et 0.7. Pour la zone  $x$ ,  $\mu_{SP_1}(x) = 0.3$ ,  $\mu_{SP_2}(x) = 0.3$ , et  $\mu_{SP_3}(x) = 0.1$ .

## 4 Propagation de l'imprécision

La représentation des données à l'aide de la logique floue permet le chevauchement spatial des entités agronomiques floues. Par exemple, il est possible d'obtenir un territoire simulé contenant trois parcelles agricoles floues : la parcelle  $P_1$  est un champ de luzerne, les deux autres ( $P_2$  et  $P_3$ ) sont des champs de maïs. Leur représentation spatiale floue est présentée dans la figure 2.

Nous supposons que les frontières dans le territoire observé sont issues d'une segmentation floue à partir d'images satellites. De cette dernière, nous gardons pour chaque pixel leur degré d'appartenance à toutes les régions. A l'aide de ces degrés et de fonction de seuil, nous stockons dans le système les informations sous la forme de données 4-valuées : 0.1, 0.3, 0.7 et 1.

La recherche est faite sur le bentazone. Les agriculteurs en diffusent sur  $P_1$ ,  $P_2$  et  $P_3$  les quantités floues ( $F_{Q_b}^{P_1}, F_{Q_b}^{P_2}, F_{Q_b}^{P_3}$ ). Ces dernières sont stockées sous la forme de nombres flous triangulaires. Elles ont pour valeurs :  $F_{Q_b}^{P_1} = (1.0, 1.1, 1.2)$ ,  $F_{Q_b}^{P_2} = (0.5, 0.6, 0.7)$ , et  $F_{Q_b}^{P_3} = (0.4, 0.5, 0.6)$

La première phase consiste en la modération de la confiance (et donc les degrés d'appartenance) des informations quantitatives en fonction de celle que l'on a sur l'information spatiale. Pour cela, pour chaque endroit  $x$  du territoire de chevauchement, une quantité floue locale pour chaque entité  $A_i$  en jeu est définie comme suit :

$$\forall x / \mu_{SA_i}(x) \neq 0, \mu_{Q_j}^{A_i/x}(q) = T(\mu_{SA_i}(x), \mu_{Q_j}^{A_i}(q))$$

où  $T$  est une  $t$ -norme à l'instar de la multiplication (pour se rapprocher du sachant que en probabilité).

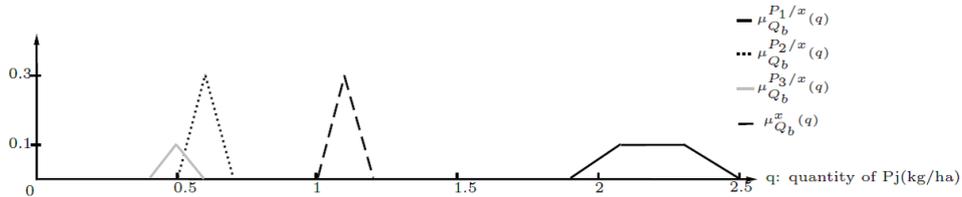


FIG. 3 – La fonction d'appartenance  $\mu_{Q_b}^x$  de la quantité floue locale globalisée de Bentazone  $F_{Q_b}^x$ .

Sur  $x$ , trois ensembles flous correspondant aux quantités locales en  $x$  des quantités floues diffusées sur  $P_1$ ,  $P_2$  et  $P_3$  sont construits. A partir de ces trois ensembles ( $F_{Q_b}^{P_1/x}$ ,  $F_{Q_b}^{P_2/x}$ ,  $F_{Q_b}^{P_3/x}$ ).

Il s'agit maintenant de prendre en considération un possible phénomène additif. Pour cela, nous considérons la somme floue entre les quantités floues locales possiblement diffusées sur  $x$ . La quantité floue globale est appelée  $F_{Q_b}^x$ . Ainsi, pour définir pour  $A_1$ ,  $A_2$  et  $x_1$ , elle est définie par :

$$\forall q \geq 0, \mu_{Q_b}^{x_1}(q) = \sup_{q=z+t}(\min(\mu_{Q_j}^{A_1/x_1}(z), \mu_{Q_j}^{A_2/x_1}(t)))$$

Nous calculons l'ensemble flou  $F_{Q_b}^x$  présenté dans la figure 3. À la lecture de la réponse du système, nous avons possiblement une quantité de Bentazone comprise dans [2.1,2.3]. Cependant, sa possibilité (et non probabilité) de présence est égale à 0.1 et elle est inférieure à celle d'une diffusion séparée non additive.

## 5 Conclusion

Dans l'objectif de la construction d'un observatoire des pratiques agricoles gérant les données agronomiques et leurs imperfections, nous avons proposé une modélisation du principe d'une entité agronomique floue ainsi que la structure de gestion de l'information avant de présenter une méthode permettant le requêtage des quantités combinées sur l'ensemble du territoire tenant compte de la qualité de l'information tant spatiale que quantitative.

Dans nos travaux futurs, nous nous attacherons à établir le lien entre les différentes couches afin d'avoir un système opérationnel de gestion de l'information spatiotemporelle, de son imperfection et de la métaconnaissance. Ce système est crucial dans le but de fournir aux utilisateurs un outil d'aide à la décision territoriale performant et appréhendable en fonction des besoins.

## Références

Bédard, Y., S. Larrivée, M. Proulx, et M. Nadeau (2004). Modeling geospatial databases with plug-ins for visual languages : A pragmatic approach and the impacts of 16 years

## Gestion et propagation de l'imprécision des données agronomiques

- of research and experimentations on perceptory. In *COMOGIS Workshops ER2004, LNCS 3289*, Shangai, China, pp. 17–30. Springer.
- Desjardin, É. et C. de Runz (2009). Gissar : de la saisie de fouilles à l'analyse spatiotemporelle en archéologie. In *Spatial Analysis and GEomatics*, Paris, France.
- Ma, Z. M. (2008). Fuzzy conceptual information modeling in uml data model. In *International Symposium on Computer Science and Computational Technology*, Volume 2, pp. 331–334. IEEE.
- Medina, J., M. Vila, J. Cubero, et O. Pons (1995). Towards the implementation of a generalized fuzzy relational database model. *Fuzzy Sets and Systems* 75(3), 273 – 289.
- Miralles, A. et T. Libourel (2008). Modeling with enriched model driven architecture. In S. Shekhar et H. Xiong (Eds.), *Encyclopedia of GIS*, pp. 700–705. Springer.
- Shi, W. Z. (2010). *Principle of Modeling Uncertainties in Spatial Data and Spatial Analyses*. CRC Press.
- Zayrit, K., E. Desjardin, C. de Runz, et Akdag (2011). Handling imperfect spatiotemporal information from the conceptual modeling to database structures. In *International Symposium on Spatial Data Quality (ISSDQ)*, Coimbra, Portugal.
- Zoghalmi, A., C. de Runz, H. Akdag, M. Zaghdoud, et H. Ben Ghezala (2011). Handling imperfect spatiotemporal information from the conceptual modeling to database structures. In *International Symposium on Spatial Data Quality (ISSDQ)*, Coimbra, Portugal.

## Summary

One of the main objectives of Observox is to handle agronomical data, as considered as spatial, quantitative and imprecise data, for delivering a quality assessment of the involved data. In this context, we choose to model data using the fuzzy set theory. From a conceptual modeling that manages information imperfection, we build a fuzzy spatiotemporal database which stores fuzzy agronomical entities. By using fuzzy sets, we make entity overlapping possible. In this case, we propose to propagate the spatial imprecision through the quantitative one in order to obtain a fuzzy local quantitative information about quantity of products disseminated for each location in the studied territory.