Génération de cartes interprétables-application à l'agriculture de précision

Patrice Loisel, Brigitte Charnomordic, Hazaël Jones, Bruno Tisseyre

▶ **To cite this version:**

Patrice Loisel, Brigitte Charnomordic, Hazaël Jones, Bruno Tisseyre. Génération de cartes interprétables-application à l'agriculture de précision. 13. Rencontres de TheoQuant, May 2017, Besançon, France. hal-02733607

HAL Id: hal-02733607
https://hal.inrae.fr/hal-02733607

Submitted on 2 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Génération de cartes interprétables - application à l'agriculture de précision

Patrice Loisel[a,b], Brigitte Charnomordic[a,b], Hazaël Jones[c], Bruno Tisseyre[c]

[a] *INRA, UMR MISTEA, 34060 Montpellier, France*
[b] *SupAgro, UMR MISTEA, 34060 Montpellier, France*
[c] *SupAgro, UMR ITAP, 34060 Montpellier, France*

Keywords: agriculture numérique, optimisation, labels, données spatiales.

1. Contexte et objectifs

L'explosion des données en Agriculture (volume, multi-échelles, complexité, caractère spatio-temporel, hétérogénéité) [2] requiert le développement d'approches innovantes capables de les analyser en tenant compte de toutes ces caractéristiques. L'agriculture numérique se développe afin d'améliorer la gestion des exploitations agricoles, des territoires et des filières.

Dans ce contexte, la génération automatique de cartes synthétiques à partir de données est devenue un enjeu important [3, 4, 5, 6]. A l'état de l'art, les méthodes de génération de zones sont le plus souvent basées sur des découpages en classes de valeurs fixées a priori [1], suivis d'une projection, puis d'une agrégation manuelle pour simplifier l'interprétation. La qualité du zonage obtenu n'est pas évaluée par des critères objectifs.

Les objectifs de ce travail sont de proposer une approche par optimisation pour générer automatiquement des zones en fonction de critères numériques basés sur des indicateurs d'homogénéité intra-zone et d'hétérogénéité inter-zones voisines. Des premiers algorithmes ont été développés et validés sur données simulées. L'intérêt de ces algorithmes est double :

1. s'abstraire de définitions de classes de valeurs fixées a priori, tout en associant des labels interprétables aux zones générées,
2. permettre la comparaison rapide et objective de zonages entre eux, indispensable dans le cas de masses de données à analyser.

Email address: patrice.loisel@inra.fr (Patrice Loisel)

2. Méthodes

A partir des données géoréférencées $(\xi, F(\xi))$ où $\xi = (x, y)$, la procédure proposée permet de générer automatiquement un zonage optimal. Le zonage dit optimal (relativement à un critère) doit maximiser l'hétérogénéité entre zones voisines (au sens des polygones de Voronoï) tout en minimisant l'hétérogénéité intra-zones. Elle doit aussi satisfaire des contraintes de forme.

2.1. Critère d'évaluation du zonage

Le critère correspondant au zonage $\mathcal{Z}$ est basé sur des indicateurs calculés sur des données krigées :

- indicateur d'hétérogénité entre deux zones voisines z_I et z_J :

$$M_{IJ} = \frac{\sum_{\eta \in z_I} \sum_{\xi \in z_J} (F(\eta) - F(\xi))^2 . S(\eta) . S(\xi)}{S_I S_J}$$

où $S(\xi)$ est la surface du polygone de Voronoï associé au point ξ
S_I est la surface de la zone z_I.

- indicateur d'hétérogénité intra-zone : M_{II}, en utilisant la même formule.

Le critère proposé se déduit des indicateurs définis précédemment :

$$Crit(\mathcal{Z}) = \min_{z_I \in \mathcal{Z}} \min_{z_J \in \mathcal{V}(z_I)} \frac{2M_{IJ}}{M_{II} + M_{JJ}}$$

où $\mathcal{V}(z_I)$ est l'ensemble des zones voisines de la zone z_I. Du fait de l'utilisation du minimum, le critère pénalisera tout zonage comprenant 2 zones voisines trop proches au sens de F, et/ou comportant 1 zone trop hétérogène. On souhaite effectivement que toutes les zones respectent ces deux aspects.

2.2. Paramétrisation du problème

Etant donné qu'à tout vecteur de quantiles q de taille n, on peut associer des lignes de niveau et ainsi générer un unique zonage $\mathcal{Z}_q^n$, nous avons choisi une procédure d'optimisation basée sur les quantiles. Intuitivement la taille n du vecteur des quantiles est liée à la finesse du zonage que l'utilisateur recherche. Pour une taille n, on définit $n+1$ labels ayant pour limite inférieure et supérieure les différents quantiles. Chaque zone d'un zonage se voit ainsi attribuer un label.

2.3. Prise en compte des contraintes

Si pour chaque vecteur de quantiles, on peut effectivement générer un zonage et évaluer sa qualité par le critère, le zonage ainsi généré ne respecte pas nécessairement (et même très rarement) les contraintes de forme souhaitées par l'utilisateur. Pour y remédier une procédure de correction de zonage doit être implémentée. Les corrections consistent soit à supprimer les petites zones, soit à les agrandir.

Suite aux multiples modifications apportées, le nombre de zones peut diminuer drastiquement ainsi que le nombre de labels. Plus que le nombre de quantiles, ce qui est important et représentatif de la finesse du zonage obtenu, c'est le nombre final de labels car il correspond au nombre de classes de gestion final.

2.4. Procédure d'optimisation

Pour une taille de vecteur donnée n, la procédure d'optimisation est basée sur un algorithme de recherche dichotomique. La procédure nécessite des évaluations du critère pour différents zonages $\mathcal{Z}_q^n$. L'algorithme pour évaluer $Crit(\mathcal{Z}_q^n)$ est le suivant :

Pour tout vecteur de quantiles q_n de longueur n, on génère le zonage $\mathcal{Z}_{q_n}$. Si les contraintes sont vérifiées, on calcule le critère du zonage $\mathcal{Z}_{q_n}$. Sinon, on explore les corrections possibles sur les petites zones (arbre complet des combinaisons). On obtient ainsi un certain nombre de zonages $\mathcal{Z}_i(q_n)$ respectant toutes les contraintes, mais comportant du fait des modifications un nombre de labels éventuellement inférieur au nombre de labels initial $\leq n+1$. Pour $m \leq n+1$ le zonage optimal $\mathcal{Z}_{q_n}^m$ est alors défini comme celui qui a le critère maximum parmi les zonages qui ont m labels :

$$\mathcal{Z}_{q_n}^m = \text{argmax}_{i, nb(\mathcal{Z}_i(q_n))=m} Crit(\mathcal{Z}_i(q_n))$$

où $nb(\mathcal{Z})$ est le nombre de labels distincts du zonage $\mathcal{Z}$.
Pour obtenir le zonage optimal pour m labels il faut effectuer la procédure d'optimisation $\forall q_n$ tel que $n \geq m-1$.

$$\mathcal{Z}_{q^*}^m = \text{argmax}_{n \geq m-1} \max_{q_n} Crit(\mathcal{Z}_{q_n}^m)$$

3. Résultats et conclusion

La méthode a été testée sur des données simulées et des données de terrain. Nous présentons ici les résultats obtenus sur des données de terrain,

issues du Mas Numérique[1], et correspondant à des mesures spatialisées de vigueur (circonférence de ceps) sur 279 échantillons d'une parcelle viticole. Après krigeage, le jeu de données est une grille 44x44. La Figure 1 présente, de gauche à droite, le meilleur zonage (critère maximum) obtenu par optimisation pour les vecteurs de 2 quantiles, conduisant à 3 labels, un zonage avec une faible valeur de critère toujours pour 2 quantiles et 3 labels, et le meilleur zonage de l'optimisation des vecteurs de 3 quantiles, conduisant à 4 labels.

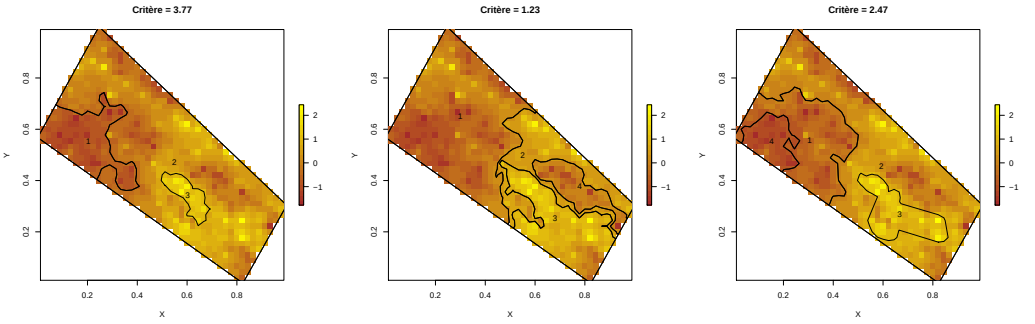

Figure 1: De gauche à droite, $\mathcal{Z}^3_{q^*}(q^* = [0.2, 0.9])$, $\mathcal{Z}^3_q(q = [0.6, 0.8])$,$\mathcal{Z}^4_{q^*}(q* = [0.1, 0.4, 0.9])$

L'optimisation a l'intérêt de générer, pour un nombre de labels donné, non pas uniquement la meilleure solution, mais également toute une famille de cartes qu'on peut exploiter par la suite. La méthode proposée permet aussi de comparer différents zonages pour un même nombre de labels.

Les perspectives de ce travail sont de deux ordres : 1) introduire des indicateurs tenant compte de la structure spatiale interne aux zones, 2) concevoir des procédures de visualisation adaptées à la nature des zones (homogène, gradient, chaotique).

4. Références bibliographiques

References

[1] R. M. Lark and J. V. Stafford. Classification as a first step in the interpretation of temporal and spatial variation of crop yield. *Annals of*

[1]Le Mas Numérique est situé au domaine du Chapitre de Montpellier SupAgro (Villeneuve lès Maguelone). Des masses de données viticoles seront recueillies automatiquement à partir de 2017. Un premier recueil a commencé en 2016.

Applied Biology, 130(1):111–121, 1997.

[2] A. McBratney, B. Whelan, T. Ancev, and J. Bouma. Future directions of precision agriculture. *Precision agriculture*, 6(1):7–23, 2005.

[3] N. R. Pal and S. K. Pal. A review on image segmentation techniques. *Pattern recognition*, 26(9):1277–1294, 1993.

[4] M. Pedroso, J. Taylor, B. Tisseyre, B. Charnomordic, and S. Guillaume. A segmentation algorithm for the delineation of agricultural management zones. *Computers and Electronics in Agriculture*, 70(1):199–208, 2010.

[5] N. R. Peralta, J. L. Costa, M. Balzarini, M. C. Franco, M. Córdoba, and D. Bullock. Delineation of management zones to improve nitrogen management of wheat. *Computers and Electronics in Agriculture*, 110:103–113, 2015.

[6] L. Zane, B. Tisseyre, S. Guillaume, and B. Charnomordic. Within-field zoning using a region growing algorithm guided by geostatistical analysis. In *Precision agriculture'13*, pages 313–319. Springer, 2013.