
Fouille de données spatiales pour la caractérisation de paysages en lien avec des fonctionnalités agro-écologiques

Thomas Guyet

AGROCAMPUS-OUEST

65 rue de Saint Brieu, CS 84215, F-35 042 Rennes Cedex, France

IRISA

Campus de Beaulieu, F-35 042 Rennes Cedex, France

thomas.guyet@agrocampus-ouest.fr

1. Introduction

De plus en plus de recherches en agronomie s'intéressent à comprendre les processus agro-écologiques en interaction avec un paysage. Mais les agronomes ne disposent généralement pas d'une variété importante de paysages réels pour identifier la façon dont les caractéristiques de paysage agissent sur un processus agro-écologique. La simulation de paysages virtuels, réalistes et paramétrables doit palier ce manque de données réelles. Ces paysages virtuels serviront de support à la simulation du processus étudié pour en comprendre les interactions. De sorte à tester leurs hypothèses, ces paysages doivent être paramétrables. Pour le reste, les agronomes cherchent à disposer de paysages réalistes, c'est-à-dire des paysages dont certaines caractéristiques sont comparables à celles rencontrées dans la nature.

Notre travail s'intéresse à l'extraction automatique de caractéristiques de paysages agricoles réels en vue de la simulation de paysages virtuels réalistes. Nous définissons ici un paysage agricole comme l'organisation spatiale d'éléments paysagés naturels (*e.g.* parcelles agricoles, haies, plans d'eau) ou artificiels (*e.g.* bâtiments, routes) avec lequel interagissent des acteurs aussi divers que des agriculteurs, des insectes, la faune, la végétation naturelle et cultivée, etc.

Nous proposons, d'une part, de représenter un paysage sous la forme d'un graphe de relations fonctionnelles qui exprime plusieurs visions de l'organisation spatiale du paysage et, d'autre part, d'utiliser des méthodes de fouille de données sur le graphe pour extraire des motifs locaux caractéristiques de l'organisation du paysage. La méthode générale de caractérisation automatique d'un paysage par des motifs locaux comprend trois étapes (*cf.* Figure 1) :

- 1) la transformation des éléments du paysage en un ensemble d'objets catégoriels,

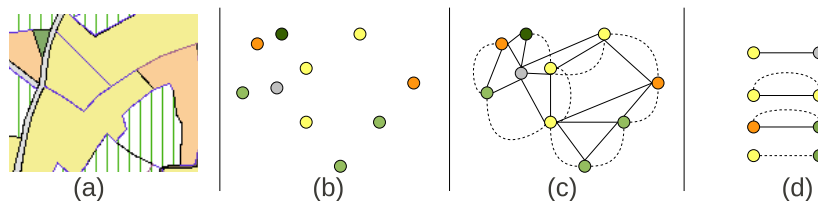


Figure 1. (a) paysage réel dans un SIG, (b) représentation des éléments paysagés (un point est une parcelle et sa couleur donne l'occupation du sol, qui est sa catégorie), (c) représentation de deux relations fonctionnelles entre les éléments paysagés : en pointillés, la proximité « à vol d'oiseaux » entre centres de parcelle et, en trait plein, l'adjacence entre parcelles, (d) motifs locaux extraits avec un support de 2.

2) la construction d'un graphe des relations fonctionnelles (spatiales) entre les éléments du paysage,

3) l'extraction de motifs locaux intéressants dans le graphe décrivant le paysage.

Là où GenExp [LEB 09] caractérise un paysage par un modèle global de statistique spatiale, *i.e.* le modèle de la distribution spatiale des centres des parcelles, nous proposons de caractériser un paysage à l'aide de motifs spatiaux locaux. Un motif spatial exprime une règle d'organisation d'éléments paysagers, par exemple « les grandes parcelles cultivées touchent une route goudronnée » ou « une parcelle de prairie est fréquemment *proche* d'une autre parcelle de prairie », etc.

2. Représentation d'un paysage par un graphe de relations fonctionnelles

Un paysage est décomposé en éléments paysagés transformés en un ensemble d'objets catégoriels. C'est-à-dire qu'un élément est un objet associé à une catégorie et à des attributs. Pour des paysages agricoles, les grandes catégories intéressantes sont les parcelles agricoles caractérisées par leur occupation du sol et les autres surfaces structurant le paysage : routes, haies, surfaces bâties, etc. Il est ensuite possible d'affiner ces catégories, par expertise ou en utilisant des méthodes automatiques de catégorisation. Il peut être intéressant de séparer les « grandes parcelles de blé » et les « petites parcelles de blés » et de créer deux catégories spécifiques. L'organisation des catégories en taxonomie permet de conserver l'information que ces deux catégories sont des sous-catégories de « parcelle de blés » et de « parcelle agricole ».

Le rôle des relations fonctionnelles est d'exprimer les liens entre les éléments du paysage, et en particulier les relations spatiales. Pour représenter l'organisation spatiale des éléments paysagés, une relation de *proximité* relie deux parcelles dans le graphe si, par exemple, la distance « à vol d'oiseau » entre leurs centres est inférieure à un seuil fixé *a priori*. La relation fonctionnelle est choisie par les agronomes en fonction des caractéristiques spatiales à mettre en avant pour caractériser un paysage. La vision de l'organisation spatiale du paysage dépend en partie des phénomènes observés : si on s'intéresse à la répartition des cultures sur une exploitation, les relations

entre parcelles doivent tenir compte de la distance par les routes et chemins. En revanche, si on s'intéresse à des flux de pollens, cette distance n'a plus d'intérêt (même si la présence de route entre deux parcelles est une information importante). De plus, des relations autres que spatiale peuvent également avoir un intérêt pour représenter le paysage (par exemple, une relation indiquant si deux parcelles appartiennent à un même exploitant).

Le représentation du paysage sous la forme d'un graphe se veut neutre. La sémantique est portée par les catégories et les types de relations qui peuvent être exprimées entre parcelles. De plus, la représentation permet d'utiliser plusieurs distances fonctionnelles dans une même représentation du paysage (*cf.* Figure 1).

3. Caractérisation automatique d'un paysage par la fouille de données spatiales

On analyse le graphe du paysage par une méthode de fouille de données inspirée de la recherche de colocalisations spatiales [HUA 04]. Cette méthode extrait des ensembles de catégories qui sont fréquemment liées deux à deux par une relation spatiale. Les motifs locaux qui sont ainsi extraits sont les ensembles maximaux de catégories dont le ratio de présence dans le paysage est supérieur à un seuil fixé. Pour plus de détails, Flouvat et *al.* [FLO 10] présentent la recherche de colocalisations appliquée à la caractérisation des zones érodées de Nouvelle-Calédonie.

Nous avons adapté la méthode pour extraire des motifs avec différentes relations fonctionnelles à partir d'un graphe en adaptant l'algorithme APriori. Cela permet d'extraire des motifs comme des ensembles maximaux de catégories liés par des relations fonctionnelles spécifiques (et pas uniquement des colocalisations).

Ce travail est en cours de réalisation dans le cadre du projet PayOTe (INRA-INRIA). Un de ses objectifs est d'explorer la caractérisation des paysages au moyen de ces motifs locaux. En particulier, nous souhaitons vérifier que l'utilisation de motifs locaux permettra de retrouver des propriétés globales d'un paysage réel dans des paysages simulés. La méthode est actuellement expérimentée pour caractériser le paysage de la zone atelier de Pleine-Fougères, en Bretagne, dans laquelle se distinguent un paysage bocagé et un paysage de grandes plaines cultivées.

4. Bibliographie

- [FLO 10] FLOUVAT F., SELMAOUI-FOLCHER N., GAY D., ROUET I., GRISON C., « Constrained colocation mining : application to soil erosion characterization », *Proceedings of the Symposium on Applied Computing*, 2010, p. 1054-1059.
- [HUA 04] HUANG Y., SHEKHAR S., XIONG H., « Discovering colocation patterns from spatial datasets : a general approach », *IEEE Transactions on Knowledge and Data Engineering*, vol. 6, n° 12, 2004, p. 1472-1485.
- [LEB 09] LE BER F., LAVIGNE C., ADAMCZYK K., ANGEVIN F., COLBACH N., MARI J.-F., MONOD H., « Neutral modelling of agricultural landscapes by tessellation methods—Application for gene flow simulation. », *Ecological Modelling*, vol. 220, 2009, p. 3536-3545.