Semi-virtual spatial simulations of land management practices in Languedoc Vineyards: a way to deal with incomplete knowledges of spatial distributions in cultivated landscapes

Bailly, J.-S. (1,4); Biarnès, A. (2); Lagacherie, P. (3)

(1) AgroParisTech, UMR LISAH, F-34060, Montpellier, France, bailly@teledetection.fr

(2) IRD, UMR LISAH, F-34060, Montpellier, France

(3) INRA, UMR LISAH, F-34060, Montpellier, France

(4) AgroParisTech, UMR TETIS, F-34035, Montpellier, France

Abstract: In this paper, methods that aim to simulate spatial distributions of Land Management Practices (LMP) at high spatial resolution and over large extents as required by spatially distributed environmental modelling are presented. The principle of theses methods, in the scope of geostatistical conditional simulation methods, is to simulate a set of equally probable spatial patterns that all respect the knowledge that have been collected on the landscape features, i.e; spatial laws and data (descriptive approach); and/or the driven factors of the studied LMP (factorial approach). The differences between the simulated spatial patterns can be seen as a representation of the spatial uncertainties. To illustrate these methods, two examples of simulations of spatial patterns of LMP are presented. They involved specific methodological developments and can be considered as representative of descriptive and factorial mapping approaches respectively: i) the simulation of ditch networks reconstructed from an incomplete set of reaches observed by remote sensing and ii) the spatial simulation of weed control practices at plot scale from a set of driven and correlated factors.

Keywords: Spatial conditional simulation; uncertainties; drainage network; weed control

Introduction

Land Management practices (LMP) in cultivated landscapes result from the constant efforts of farmers to adapt landscapes to the constraints of agricultural production. These practices include both landscape objects creation or cancellation (ditch creation or cancellation for instance) as well as landscape object properties management (weed control by tilllage on vine alleys for instance). LMP have a strong impact on many processes occuring in landscapes. Evidences of these impacts have been observed e.g. for run-off (Leonard and Andrieux, 1998; Hébrard *et al*., 2006), soil erosion (Fryirs et al., 2007), biodiversity dynamics (Burel and Baudry, 2006) , and micrometeorology (Courault et al., 2007).

These impacts are often observed as the local level, e.g. the increase of soil infiltration after a tillage that diminish run off at the field level or the presence of a hedge that maintain local biodiversity. However for a landscape management point of view, it seems more important to understand the impact that spatial patterns of land management practices may have on environmental resources at the territory level since this level is relevant to undertake and assess environmental policies. In several domains, spatially distributed environmental models have been proposed to address this problem, e.g. MHYDAS in hydrology (Tiemeyer et al, 2007) or ecology (Allen et al., 2001). However the use of such models is often hampered by the lack of available maps of LMP. It is therefore important to develop methods to map these practices at suitable spatial resolution and extents.

When spatially distributed environmental models aim to explore landscape behaviour regarding future LMP spatial distribution scenarii, the use LMP spatial dynamics based on spatio-temporal laws and dynamic interactions, are required. Systemic landscape modelling frameworks can achieve this goal (Bousquet and Lepage, 2005; Jacewicz P. and Pausas, 2003). But, when spatially distributed environmental model use is more dedicated to assess the impacts of a landscape configuration at a given and present time on a real catchment or territory, which is the framework of the this paper, the aim of LMP mapping is more an interpolation problem, without interaction consideration, i.e. to be as accurate as possible in reference to the “true” spatial distribution.

For this latter case, available mapping methods of LMP can be descriptive, i.e. based only on LMP or correlated variables observations, or based on factors driving the LMP spatial distribution (Verburg *et al*. 1999). The descriptive mapping methods use observation tools, mainly remote sensing ones. Example of such methods are frequent in the literature (e.g. Briclemeyer *et al*., 2006). They suppose that the studied practice can be associated with an observable landscape feature such as bare soil fields after tillage that can be mapped with a vegetation index derived from a remote sensing image (South *et al*., 2004) or ditches corresponding to a surface discontinuity mapped from high resolution DEM (Bailly et al., 2008). Factorial methods are based on a prior analysis of the factors that may influence the farmer’s land management decisions or that are correlated to them. Here again, mapping is only possible if these factors can be associated with observable landscape features or with any other available spatial information (such as farm characteristics for example).

However, in both cases, the uncertainty and the incompleteness of the available spatial information that is used for mapping the landscape features or the decisions factors are strongly limiting and must therefore be taken into account since it could impact confidence in environmental models outputs. A way to proceed is to decrease the spatial resolution of the delivered map which the risk of removing the short range spatial variability that have a strong impact on the studied landscape process.

In this paper we propose an alternative inspired from the stochastic conditional spatial simulations methods that “attempts to simulate the real conditions rather than simulate their measured variabilities" (Lantuejoul 2002). The principle is to simulate a set of equally possible spatial patterns that all respect the knowledge that have been collected on the location of the landscape features and/or the driven/correlated factors of the studied LMP. The differences between the simulated spatial patterns can be seen as a representation of the uncertainty. Assessing uncertainty about LMP mapping is "not a goal per se, rather it is a preliminary step to evaluate the risk involved in any decision-making process or to investigate how mapping errors propagate through complex functions" in environmental models (Goovaerts 2001). In essence, these simulated spatial patterns can be defined as "semi-virtual", i.e. simulations conditioned by data (data assimilation) or driven factors to re-enforce the proximity between simulated spatial patterns and real conditions. In the following, we illustrate the semi virtual simulations of spatial patterns of LMP with two already published examples that can be considered as representative of descriptive and factorial mapping approaches respectively, i) the mapping of a ditches network from an incomplete set of reaches observed by remote sensing and ii) the mapping of weed control practices from a set of predicting factors.

1.Conditioning the spatial simulation of LMP by data: example of ditches networks

1.1. Ditches network simulation algorithm

In this first example (Bailly, 2007), we attempt to simulate LMP corresponding to artificial drainage network implementation within a given catchment, i.e. a known outlet. The simulation is constrained both by a spatial model representing the ditches network and data. The spatial model on networks is a combination of simple spatial laws considering that the network is 1)a directed tree graph with direction governed by topography and 2)a sub-graph of the plot boundaries lattice. Regarding constraints coming from data, simulations run 1) within the lattice of plot boundaries directed by elevation and 2) path through a set of known but unconnected reaches of the network (data assimilation). These data can be seen as data that condition the whole network simulation connecting the known reaches.

To perform the conditional simulation of ditches network, we chose to develop a stochastic algorithm generating networks corresponding to directed tree structures as sub-graphs of the directed plot lattice. This algorithm is based on :

- A network initialisation corresponding to the tree including the set of known reaches connected one to each other up to outlet.
- An altimetrical noise parameter related to DTM[1] noise determining which edges of the lattice are surely directed and those are not when altimetrical difference between terminal nodes of the edge is lower than the parameter. In that case, edges are duplicated in lattice with opposite directions.
- A pruning or branching iterative random process connecting the known reaches within the directed plot boundaries lattice up to converge on a descriptive criteria (sources distribution criteria, drainage density criteria, or known reach connection

[1] Digital Terrain Model

criteria) using Greedy algorithm (Cormen et al., 2009) or simulated annealing algorithm (Kirkpatrick et al., 1983). Branching process is realized through random upstream-downstream walks within the plot lattice starting from a randomly sampled edge. Pruning process is simply based on the pruning of the part of the network connected upstream a randomly sampled edge of the lattice.

Uncertainties in networks mapping are simulated through a set of equi-probable generation of networks (Figure 2).

1.2. Case study and results

This algorithm was applied on a 2 km² vineyard catchment located in the Peyne catchment (Languedoc –France). On this catchment, ditches detected from a remote sensing process (Bailly et al., 2008) corresponding to about 60% of the cumulated length of the network were used to condition the simulation. The criteria used for simulation convergence was the respect of sources distribution in altitude and Greedy algorithm was used.

When comparing simulated networks patterns to the actual network (Figure 1), simulated networks appear realistic with small biases on descriptive criteria of the networks (Figure 2): geometrical (a) and topographical criteria (b) (Vannimenus and Viennot 2005).

Results on that network example show that conditioning spatial simulation with data (here from remote sensing) allows to obtain quite realistic simulations even when spatial sampling rate of data is low.

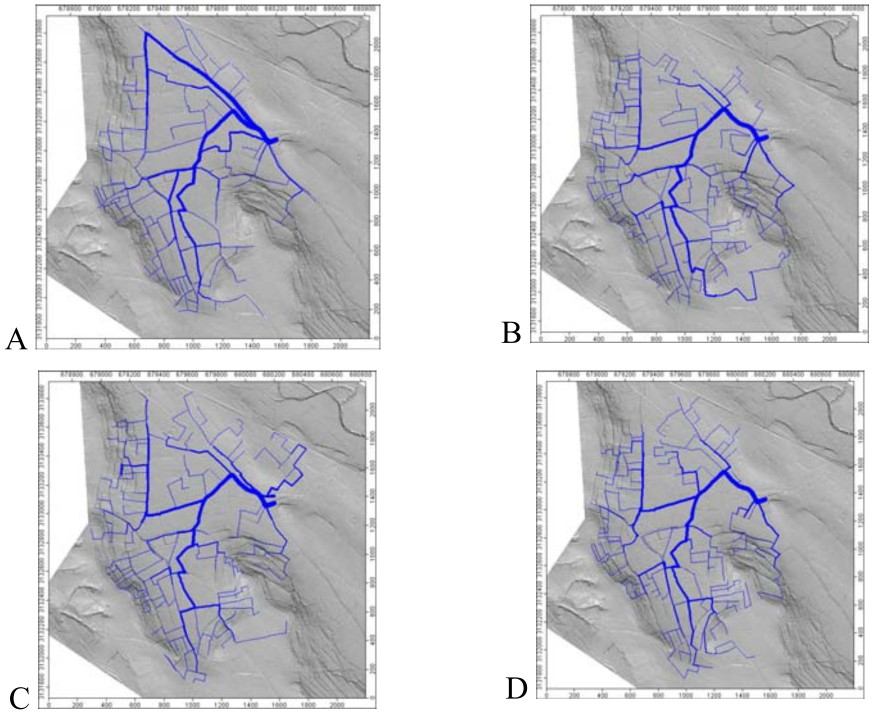
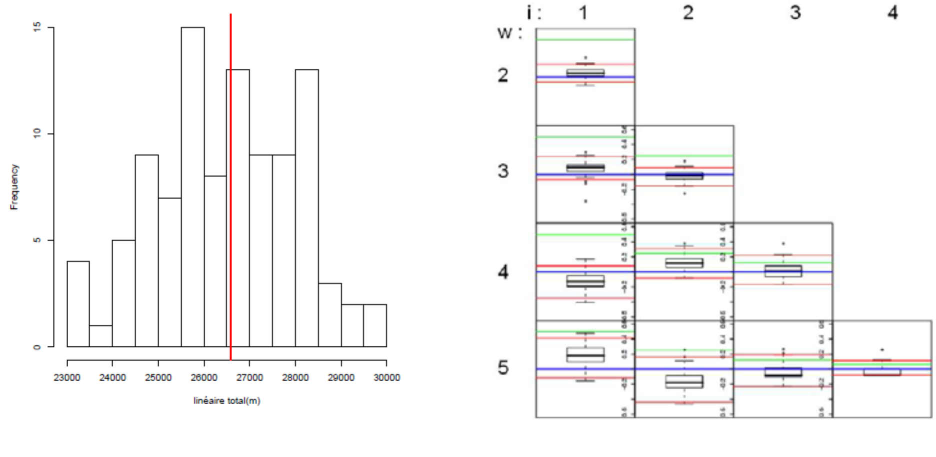


Figure 1. Three conditional simulations (B-C-D) of the extended Roujan Catchment drainage network compared to the actual drainage network (A). Networks are depicted

with line width proportional to an upstream-downstream order on a lighted 1 m DTM



a b

Figure 2. Histogram of cumulated lengths of 100 simulated networks (a) centered on the actual cumulated length (red line) and boxplots of ramification matrix (matrix of Horton-Strahler bi-orders w on junctions) of the same 100 simulated networks (b) centered on the actual ramification matrix length (blue lines)

2. Conditioning the spatial simulation of LMP by spatially structured driving factors: examples of weed control practices on vine plots

In this second example, we aimed to identify a set of spatially explicit factors for simulating the spatial distribution of weed control practices (WCP) in the Peyne vine growing catchment at the plot resolution (Biarnès et al., 2009). On the basis of interviews of 63 winegrowers, a spatially explicit database was developed that included 1007 vine plots and information regarding practices and potential explanatory variables. In order to further extend the use of identified explanatory variables to simulate the spatial distribution of WCP throughout the whole Peyne catchment, we only collected variables that (1) we assumed to be potentially explanatory of the WCP and (2) which were directly (or assumed to be indirectly) available at plot scale from digital regional maps, very high spatial resolution images from French Geographic Mapping Agency (IGN) and national databases.

Four practices were differentiated according to the methods used (chemical weed control, shallow tillage, grass cover or a combination) that determine the intensity of herbicide use and potential surface run-off. Three groups of explanatory variables corresponding to three assumed levels of spatial organisation of WCP (the plot, the farm and the local government area (LGA)) were tested and compared. In the first step, selection of factors within each group and various combination of these groups was performed using a self-developed and robust extension of the aggregated classification and regression tree methods.

2.1. Factor analysis method and simulation algorithm

The developed method is a classification method based on classification tree CART method (Breiman 1984), extended in a way to avoid spatial sampling effects and to limit over-fitting. For that purpose, it uses randomisation within the set of samples for calibration and pruning sets building several tree within a forest. The specificity of the proposed method is to perform a frequency analysis of splitting rules for each node of the tree in a hierarchical manner starting from the root. For each node, the more frequent rule is kept. At the end of the process, it leads to a single frequent tree which provides a single and easily interpretable model between factors and WCP. In the second step, the performance of the selected factors for reproducing the spatial observed repartition of practices at long range was evaluated by a stochastic use of the frequent tree allowed by the probability distributions of practices that are the output of the tree. The stochastic use of the frequent tree leads to a set of equi-probable spatial distributions of practices at the plot resolution (figure 3) and provide an explicit view of the uncertainty associated with the discrimination of the practices and the simulation of their spatial distribution test.

2.2. Case study and results

Concerning the case study, the test of the significance in the difference of dissimilarities (dissimilarity between the simulated and the observed WCP distributions) obtained with the three sets of explanatory variables indicate that the combination of the three groups of variables leads to the highest-performing simulations of the spatial distribution of WCP. Nevertheless, the farm holding variables provided little additional spatial information, which supports the idea that they may be omitted without significantly impacting the final results.

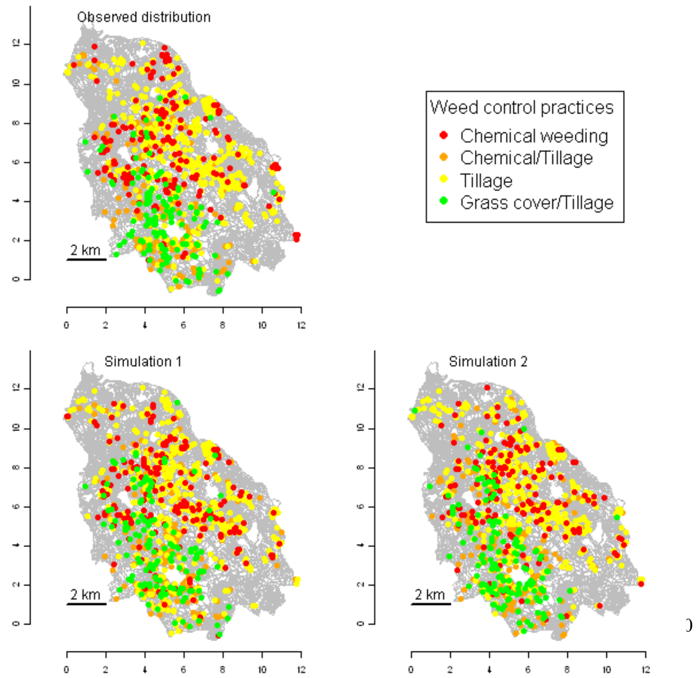


Figure 3. 2 simulations of vines plot weed control practices spatial distribution over the Peyne Catchment (34) compared to the actual distribution.

Conclusion

The two above-presented semi virtual simulations of spatial patterns of LMP illustrate how it is possible to conciliate the two apparently contradictory objectives that are required to relate LMP and their impact on environmental resources: i) describing spatial patterns with spatial resolutions that are fine enough to explicitly represent the landscape features that are locally impacted by a given technical action of the farmer while ii) covering spatial extents large enough for observing the cumulative impacts of these actions and for taking into account any collective strategies for enhancing/ limiting these impacts. For these objectives, a realistic and explicit representation of the uncertainty due to data limitations is a key problem for which we proposed an approach based on conditional simulations. Beyond the two examples that were presented in this paper, the method can be extended to many others landscape management practices for which observations of these corresponding features or knowledge of causal factors can be available. As shown with the two presented examples, the mapping algorithms can differ strongly from a LMP to another while producing the same outputs. Furthermore some methodological points will need to be addressed in the future researches:

(1) Validation strategies of semi-virtual simulations outputs need to be refined. Ponctual validations that are of current use in classical tests of mapping procedures are no longer valid when the goal is to adequately represent spatial patterns and not local evidences. This is all the more true when addressing non continuous geographical support like tree structures or lattice data for which pattern comparisons metrics are to be found. Beside, validations should not only measure the accuracy of the semi-virtual simulations per se, but also the added value on environmental models outputs. This means to develop error propagation studies and sensibility analyses of environmental models. For the latter, methods for models using maps as input is still an open question.

(2) Although the two examples presented are both descriptive and factorial mapping approaches, they use different constraints, data or factors, for simulation. These two ways should be associated in the future to improve the accuracy of simulations. The latter could also include expert-knowledge based rules that might not necessarily derived from local observations as in the presented examples. This would require adequate representations of the uncertainty associated with expert judgements, bridging gaps with fuzzy logic approaches classically used to perform this (e.g. Cazemier et al, 2001) and probabilistic techniques as the one presented in this paper.

(3) The presented examples produce simulations that have no time dimension which limits their use to environmental assessment at a given time (e.g. the impact of a Mediterranean storm event). Although temporal processes introduction could already be of interest for landscape simulation at a given time, time dimension, using similar conditional approaches, must be introduced in the simulations to deal with medium and long term evolution of environmental resources in relation with changes in land management practices.

References

Allen J.C., Brewster C.C., Slone D.H., 2001. Spatially explicit ecological models: a spatial convolution approach, 12(2), *Chaos in Ecology*, p. 333-347.
Bailly J.S., 2007, *Méthodes de spatialisation de réseaux arborescents et orientés par simulations conditionnelles : application aux réseaux de fossés de paysage viticoles*, Ph.D, Université de Montpellier II.
Bailly J., Lagacherie P., Millier C., Puech, C. & Kosuth,, P., 2008. Agrarian landscapes linear features detection from LiDAR: application to artificial drainage networks, *International Journal of Remote Sensing,* 29(12), p. 3489-3508.
Biarnès A., Bailly J. & Boissieux Y., 2009. Identifying indicators of the spatial variation of agricultural practices by a tree partitioning method: the case of weed control practices in a vine growing catchment, *Agricultural System,* 99(2-3), p. 105-116.
Bricklemyer R. S., Lawrence R. L., Miller P. R., Battogtokh N., 2006. Predicting tillage practices and agricultural soil disturbance in north central Montana with Landsat imagery. *Agriculture, Ecosystems & Environment*, 114 (2-4), p. 210-216.
Breiman, L., Friedman J., Olshen R. and Stone C., 1984. *Classification And Regression Tree*, London: Chapman and Hall, p. 368.
Bousquet F. et Le Page C, 2005. Multi-agent simulations and ecosystem management: a review. *Ecological Modelling,* 176 (3-4), p. 313-332.
Burel F. & Baudry J., 2006. Habitat quality and connectivity in agricultural landscapes: the role of land use systems at various scales in space and time, *Ecological Indicators*, 5(4), p. 305-316.
Cazemier D. R., Lagacherie P. and Martin-Clouaire R., 2001. A possibility theory approach for estimating available water capacity from imprecise information contained in soil databases. *Geoderma,* 103, p. 115-134.
Cormen T. H., Leiserson C. E., Rivest R. L., Stein C., 2009. *Introduction to Algorithms* (3rd ed.). MIT Press.
Courault D., Drobinski P., Brunet Y., Lacarèrre P. & Talbot C., 2007. Impact of surface heterogeneity on a buoyancy-driven convective boundary-layer in light winds, *Boundary-Layer Meteorology,* 124(3), p. 383-403.
Fairfield J. & Leymarie P., 1991. Correction to “Drainage Networks From Grid Digital Elevation Models”, *Water Resources Research,* 27(10), p. 2809.
Fryirs K., Brierley G., Preston N. & Kasai M., 2007. Buffers, barriers and blankets: The (dis)connectivity of catchment-scale sediment cascades, *Catena,* 70, p. 49-67.
Goovaerts P., 2001. Geostatistical modelling of uncertainty in soil science, *Geoderma,* 103, p. 3-26.
Hébrard O., Voltz M., Andrieux P., Moussa R., 2006. Spatio-temporal distribution of soil surface moisture in a heterogeneously farmed Mediterranean catchment, *Journal of Hydrology,* 329, p. 110-121.
Jacewicz P. and Pausas J. G. , 2003. Lucas: an original tool for landscape modelling, *Environmental Modelling & Software*, 18(5), p. 429–437.
Kirkpatrick S., Gelatt C.D., Vecchi M.P., 1983. Optimization by Simulated Annealing, *Science,* New Series 220 (4598), p. 671–680.
Lantuejoul C., 2002. Geostatistical Simulation: Models and Algorithms, Berlin, Springer Verlag, 256p.
Leonard J., Andrieux P., 1998. Infiltration characteristics of soils in Mediterranean vineyards in Southern France, *Catena,* 32, p. 209-223.
South S., Qi J., Lusch D. P., 2004. Optimal classification methods for mapping agricultural tillage practices. *Remote Sensing of Environment,* 91 (1), p. 90-97.
Tiemeyer B., Moussa R., Lennartz B. and Voltz M., 2007. MHYDAS-DRAIN: A spatially distributed model for small, artificially drained lowland catchments, *Ecological Modelling*, 209 (1), p. 2-20.
Vannimenus J., Viennot X.G., 2005. Combinatorial tools for the analysis of ramified patterns, *Journal of Statistical Physics*, 54(5), p. 1529-1538.

Verburg P. H., de Koning G. H. J., Kok K., Veldkamp A., Bouma J., 1999. A spatial explicit allocation procedure for modelling the pattern of land use change based upon actual land use. *Ecological Modelling,* 116 (1), p. 45-61.