Generating Fine Resolution Area Class Maps Subject to Coarser Resolution Data Constraints

P. C. Kyriakidis[1], G. Cao[1]

[1]Department of Geography, University of California Santa Barbara, Santa Barbara, CA 93106-4060, U.S.A.
Email: phaedon@geog.ucsb.edu

1. Introduction

Categorical geospatial data, such as land cover or land use classes, species types, or categories of socioeconomic status, are all important information sources in spatial analysis, resource management, decision support systems and planning in general. Such data types can be either nominal or ordinal and exhibit spatial or spatiotemporal patterns with sharp boundaries and complex geometrical characteristics. In addition, in the context of error or uncertainty modelling and propagation to GIS operations, spatial analysis and modelling of categorical data quickly becomes a rather challenging task (Zhang and Goodchild, 2002).

Oftentimes, categorical geospatial data are not directly observable at the spatial resolution required for a particular analysis purpose. Instead, analysts have access to related information at a coarser spatial resolution, e.g., proportion of classes or presence/absence data within a coarse pixel or polygon. An example of this situation can be found in the context of mapping the spatial distribution of species in ecology from data on species presence or absence available from regional atlases (McPherson et al., 2006). In this case, the objective is to determine the fine resolution spatial distribution of species conditioned (subject) to coarse resolution information.

More data-rich situations exist, whereby fine resolution spatial information is available, in addition to the coarse resolution data. For example, one might have access to a (typically small) set of point locations where class labels have been directly observed, e.g., from ground surveys in the context of accuracy assessment of classified imagery. Another example is that of a fine resolution layer comprised of data on a related variable, e.g., topographic elevation in the case of modelling species distributions. In both examples, the objective is to determine the spatial distribution of class labels at the fine (target) resolution, conditioned to coarser resolution class proportions or class occurrence information, plus fine resolution direct (but sparse) or abundant (but indirect) data.

In the context of error modelling and propagation, the problem of determining the spatial distribution of class labels (subject or not to data constraints) is typically addressed via stochastic simulation. Geostatistical simulation is the procedure of generating alternative area class maps, all of which reproduce a model of spatial structure (or texture), plus conditioning data if they exist. One of the earliest algorithms in the field of GIScience for simulating area class maps with spatially varying class proportions is the approach of Goodchild et al. (1992). Further refinements and more recent developments, including geostatistical approaches for categorical error modelling, can be found in Zhang and Goodchild (2002).

This paper outlines an extension of a recently developed spatial simulation method (Cao et al., 2009) for the purpose of generating fine resolution area class maps conditioned to coarse (and possibly fine) resolution spatial data.

2. Methodology

The proposed model postulates the existence of categorical field at the fine (target) spatial resolution, and a corresponding classification scheme whose classes are mutually exclusive and collectively exhaustive. At that resolution, class labels are unobservable, unless a limited set of sparse point locations with associated class labels exists. It is assumed that the same classification scheme holds for the coarse spatial resolution, where observations consist of proportions of classes in coarse sampling units, e.g., pixels or polygons; in other words, classes are assumed mixed at the coarse resolution but crisp at the fine resolution. The objective is then to generate simulated area class maps at the fine resolution consistent with the coarse resolution data.

The proposed method assumes the existence of a prior texture model for the spatial distribution of classes at the fine resolution. Such a model is encapsulated in a set of spatial transition probabilities quantifying the likelihood of class occurrence as a function of distance from locations of the same or different class. These auto- and cross-class models of transition probabilities can be estimated from sparse point data, and/or inferred iteratively from the coarse class fractions, and/or synthesized from expert knowledge and related indirect data.

The above transition probabilities are transformed into coarse-to-fine probabilities of class occurrence, linking coarse class fractions or coarse presence/absence data with fine resolution class labels, based on knowledge of the extent and geometry of each coarse pixel or polygon. This transformation accounts explicitly for resolution differences between information sources, and yields resolution-consistent simulated area class maps, as described below.

Generation of area class maps at the fine resolution proceeds in a sequential way, along a random path dictating the order of visiting a set of locations where simulated class labels will be generated.

At an arbitrary location along that random path:

1. A local neighbourhood is defined based on the prior transition probability models; this neighbourhood includes coarse resolution pixels with known class fractions, locations with previously simulated class labels (if any), and locations with known class labels (if they exist).
2. Transition probabilities of class occurrence linking: (a) locations with known or previously simulated class labels and the current location, and (b) coarse resolution pixels where class fractions are known and the current location, are fused into a set of posterior (multi-point) probabilities; the fusion algorithm resembles an augmented Bayesian Network model (Keogh and Pazzani, 1999), whereby spatial correlation is explicitly modelled and incorporated into the fusion weights. It should be noted that the proposed probability fusion algorithm can also be regarded as an extension of the celebrated weights of evidence model, frequently used for data integration in GIS applications (Bonham-Carter, 1994).
3. A simulated class label is generated at the current location from these conditional probabilities and used as conditioning point datum for simulation at subsequent locations along the random path, i.e., becomes a previously simulated value for those locations.

The above steps are repeated until all simulation locations are visited along the random path; this completes the generation of one simulated area class map. The entire procedure is repeated anew, possibly using a different random path, to generate another simulated area class map. Each such map is a realization of the stochastic error model quantifying uncertainty in the spatial distribution of fine resolution class labels.

3. Case Study

The proposed method for generating realistic area class maps is showcased via simulation examples where coarse resolution data, along with sparse point data and layers of proxy variables (indirect point data), are used to simulate fine resolution area class maps reproducing all the available information.

As an example, Figure 1 (left) displays a reference fine resolution class map with three categories (white, gray, and black). This reference class map is unknown in practice, and instead the analyst has access to three continuous maps of coarse class fractions, one per category; the coarse fraction map for category white only is displayed in Figure 1 (right). In this example, coarse resolution pixels (outlines shown with boxes) consist of 30×30 fine-resolution pixels.

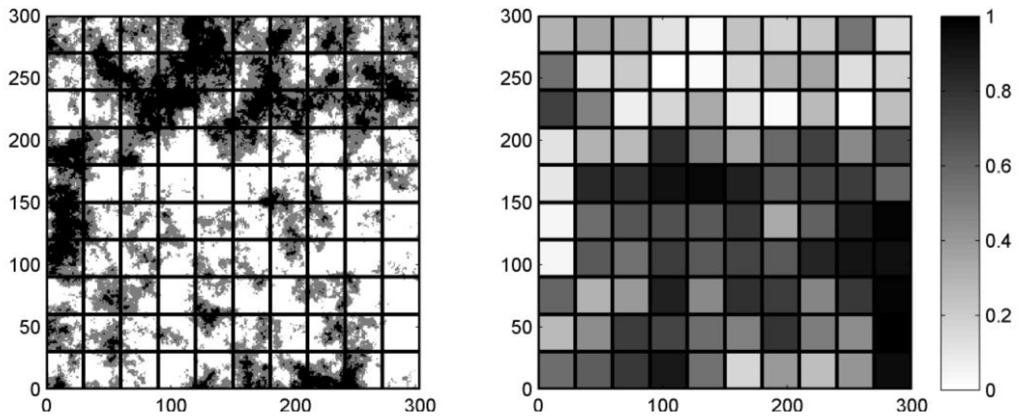

Figure 1. Reference class map (left) and coarse fractions of class white (right).

Figure 2 displays two (out of many) simulated fine resolution class maps conditioned on (reproducing when upscaled) the three maps of coarse resolution class fractions -- one of which is shown in Figure 1 (right) -- generated using a model of fine resolution spatial transition probabilities for all three categories.

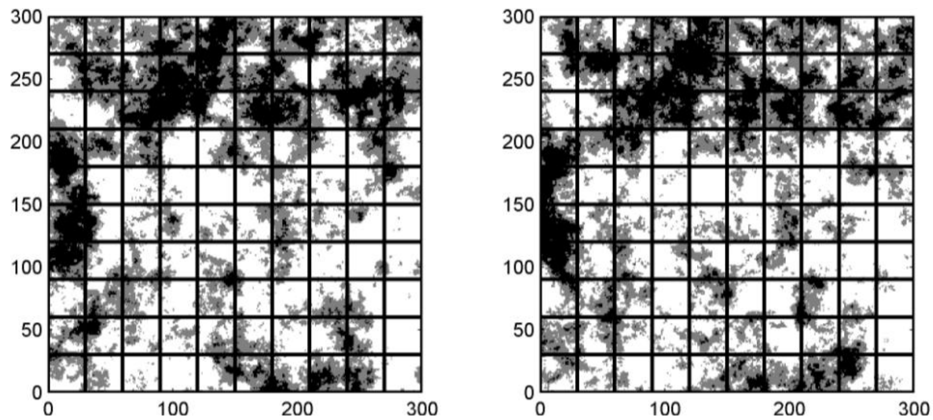

Figure 2. Two simulated fine resolution class maps.

4. Discussion

This paper presents a novel method for simulating area class maps at a fine spatial resolution based on coarser resolution data, such as class fractions or presence/absence data available at coarse pixels or polygons. Additional direct or indirect data at sparse or abundant point locations, respectively, can also be incorporated in the simulation procedure, if these are available. The alternative simulated area class maps can be used in a Monte Carlo uncertainty propagation framework for evaluating the consequences of uncertainty due to lack of direct observations of categorical data at the appropriate spatial resolution on related spatial analysis operations involving area class maps.

Acknowledgements

Funding provided for this work by the National Geospatial-Intelligence Agency (NGA) under award HM1582-07-2020 is gratefully acknowledged.

References

Bonham-Carter G, 1994, *Geographic Information Systems for Geoscientists: Modelling with GIS*. Pergamon, Oxford.

Cao G, Kyriakidis P and Goodchild M, 2009, Prediction and simulation in categorical fields: A transition probability combination approach. In: *Proceedings of the 17th ACM-SIGSPATIAL International Symposium on Advances in Geographic Information Systems*, Seattle, Washington, USA.

Goodchild M, Guoqing, S and Shiren Y, 1992, Development and test of an error model for categorical data. *International Journal of Geographical Information Systems*, 6(2):87–104.

Keogh E and Pazzani M, 1999, Learing augmented Bayesian classifiers: A comparison of distribution-based and classification-based approaches. In: *Proceedings of the 7th International Workshop on Artificial Intelligence and Statistics*, Fort Lauderdale, Florida, USA.

McPherson J, Jetz W and Rogers D, 2006, Using coarse-grained occurrence data to predict species distributions at finer spatial resolutions – possibilities and limitations. *Ecological Modelling*, 192(3-4):499–522.

Zhang J and Goodchild M, 2002, *Uncertainty in Geographical Information*. Taylor & Francis, London, UK.