USE OF INDICATOR KRIGING TO IMPROVE SPATIAL COHERENCE OF THEMATIC RASTER MAPS

G. P. Patil, Center for Statistical Ecology and Environmental Statistics,
Department of Statistics, The Pennsylvania State University,
University Park, PA 16802 (gpp@stat.psu.edu)

Key Words: Binary processes, Indicator smoothing, Kriging, Thematic maps

1. Introduction

Landcover maps have a long history of application in environmental work and, traditionally, would be constructed on the basis of ground or aerial surveys. But the use of high resolution satellite imagery for this purpose has become increasingly common in recent years. Here, a supervised classification is applied to the spectral responses in order to assign pixels to the various thematic landcover categories. Training sets are identified from ground/aerial information or simply from experienced judgment. The classification usually involves only the spectral responses, but there have been recent attempts to incorporate pixel proximity into the classification procedure. In any case, the thematic maps produced by this approach tend to exhibit substantial spatial variability. In other words, one gets to see the trees but may miss the forest. Wayne Myers uses the term spatial busyness, where the Merriam-Webster dictionary definition of 'busyness' is "full of distracting detail."

This paper develops a method for smoothing out the "distracting detail" and enhancing the spatial coherence of the thematic map. More precisely, we want to produce a sequence of maps that portray the landscape at increasingly broader scales, or as the map might be perceived from increasingly large viewing distances. The series of maps can be indexed by a smoothing parameter which we will denote generically by the symbol \( \lambda \). The value \( \lambda = 0 \) corresponds to the original map and to the finest level of detail. Larger values of \( \lambda \) correspond to greater smoothing. Thus, \( \lambda \) can be regarded as a proxy for the viewing distance.

Smoothing a map in this way is sometimes called image simplification or image generalization. Our motivation has been to bring into focus the broader scale features of a landscape. However, another important application is to the vectorization of raster images. With a high level of spatial busyness, the vectorized file can become impossibly large and some degree of smoothing is a practical necessity.

Luo (1998) describes a patch-based penalty function approach to image simplification in which the smoothed image minimizes an objective function that is the sum of two terms. The first term measures the distance between the smoothed image and the original image, while the second (penalty) term measures the degree of spatial busyness in the smoothed image and is multiplied by the smoothing parameter \( \lambda \). Minimizing the objective function becomes a tradeoff between smoothness, or lack of busyness, on the one hand and similarity to the original image on the other hand. As \( \lambda \) increases, smoothness receives greater weight.

The approach taken in this paper is technically very different, but is also motivated by the image restoration paradigm. Pretend that the smoothed map has some objective reality and that the original map results from this smoothed map by a noise process. Of course, this is only a perceptual model and need not correspond to any physical process. Since we are dealing with categorical responses, the distributional link between the two maps can be described in terms of contingency tables whose parameters become our smoothing parameters. (We will have a multitude of smoothing parameters, but the simplest model sets them all equal to some common value \( \lambda \).) At this point, it would be natural to apply Bayes theorem and obtain the smoothed map from the posterior. However, the joint distribution of the two maps at all pixels is quite intractable and we proceed using only the first two moments of the distributions (i.e., via kriging and co-kriging). As a side benefit, using moments also avoids the need for

Prepared with partial support from the National Science Foundation Cooperative Agreement Number DEB-9524722 and the United States Environmental Protection Agency Cooperative Agreement Number CR-825506. The contents have not been subjected to Agency review and therefore do not necessarily reflect the views of the Agencies and no official endorsement should be inferred.
a prior, as pointed out by O’Hagan (1994) in the conventional (non spatial) context.

As it turns out, the variograms of the smoothed map and also the cross covariograms between the smoothed and original maps are related to the variograms of the original map by systems of linear equations whose coefficients are determined by the smoothing parameters. Since the original map is in hand, we can model and estimate its variogram. Specification of the smoothing parameters then yields estimated variograms for the smoothed map and estimated cross covariograms between smoothed and original maps. This enables spatial prediction of the smoothed map from the original map. If the training set used to produce the original map is still available, then it is also possible to require that the smoothed map interpolate on the training set if this is considered appropriate.

The paper describes the general theory as summarized above and then illustrates the approach with artificial data in which a Gaussian process was used to simulate satellite imagery which was then classified to three land-cover types using a supervised Kullback-Liebler classification algorithm (Filippone, Patil, and Taulie, 1998). The final portion of the paper examines the general issue of variogram modeling for binary processes.

2. Parametric Model Linking Original Smoothed Images

Let \( K \) be the number of land-cover categories. In the original image, each pixel \( s \) is assigned to exactly one of these categories and the assignment may be represented by the indicators \( Y_i(s) \), where \( Y_i(s) = 1 \) when \( s \) is assigned to category \( i \) and \( Y_i(s) = 0 \), otherwise. Note that the vector,

\[
Y = Y(s) = (Y_1(s), Y_2(s), \cdots, Y_K(s)),
\]

has exactly one nonzero component. Similarly, we use the notation \( U_i(s) \) for the hypothetical smoothed process. In principle, one could attempt a multivariate linkage between \( Y \) and \( U \). But, this becomes too complex in terms of the modeling as well as the computational effort. Accordingly, we do the analysis one category at a time, using the binary random field \( Y_i(s) \) to predict \( U_i(s) \) for \( i = 0, 1, \cdots, K \). However, the cokriged predictions, \( \hat{U}_i(s) \), are real quantities that are not restricted to the binary values 0 and 1. They can be regarded as fuzzy membership functions expressing the ambiguity inherent in trying to assign broad-scale land-cover categories to individual pixels (Wang, 1990). We resolve this ambiguity, and obtain a non-fuzzy smoothed image, by assigning to pixel \( s \) that category \( i \) which maximizes \( \hat{U}_i(s) \).

Let us fix a particular land-cover category \( i \) and develop the second order moment structure for the binary fields,

\[
U(s) = U_i(s) \quad \text{and} \quad Y(s) = Y_i(s).
\]

Corresponding to category \( i \), a pair of smoothing parameters are defined as

\[
\lambda_0 = \Pr(Y(s) = 1|U(s) = 0)
\]

and

\[
\lambda_1 = \Pr(Y(s) = 0|U(s) = 1),
\]

where we suppose that \( \lambda_0 \) and \( \lambda_1 \) do not depend on the pixel \( s \). Each land-cover category has a pair of such parameters, but their values need to be specified by the user. In the simplest case, we would have a single parameter \( \lambda \) with \( \lambda_0 = \lambda_1 = \lambda \) for all categories. Once the smoothing parameters are specified, it is possible to model the joint distribution of \( Y(s) \) and \( U(s) \) for every pair of points \( s = s_1, s_2 \).

Assuming that the processes are isotropic and second order stationary, we write \( \tau = \|s_1 - s_2\| \). The joint distribution of \( U(s_1) \) and \( U(s_2) \) is represented as

\[
U(s_2)
\begin{array}{c|c|c}
0 & u_{00}(\tau) & u_{01}(\tau) \\
1 & u_{01}(\tau) & u_{11}(\tau)
\end{array}
\]

The covariogram is therefore

\[
C_{uv}(\tau) = \text{Cov}[U(s_1), U(s_2)] = u_{11}(\tau) - u_{00}(\tau)
\]

where \( u_\tau \) is the mean value of \( U \) and \( C_{uv}(0) = u_\tau(1 - u_\tau) \) is the variance. The variogram of \( U \) is given by

\[
\gamma_\tau(\tau) = u_{01}(\tau).
\]

Similarly, the joint distribution of \( Y(s_1) \) and \( Y(s_2) \) is represented by

\[
Y(s_2)
\begin{array}{c|c|c}
0 & y_{00}(\tau) & y_{01}(\tau) \\
1 & y_{01}(\tau) & y_{11}(\tau)
\end{array}
\]

Note that \( y_{01}(\tau) \) does not have to be the same as \( u_{01}(\tau) \).
Linking the bivariate distribution of \( U(s_1) \) and \( Y(s_2) \) and the bivariate distribution of \( U(s_1) \) and \( U(s_2) \) to that of \( Y(s_1) \) and \( Y(s_2) \) requires some additional assumptions, such as the following:

**Markovian assumption.**

\[
\Pr(Y(s_i) = y_i | U(s_i) = u_i, U(s_j) = u_j) = \\
\Pr(Y(s_i) = y_i | U(s_i) = u_i).
\]

**Independence assumption.**

\[
\Pr(Y(s_i) = y_i, Y(s_j) = y_j | U(s_i) = u_i, U(s_j) = u_j) = \\
\Pr(Y(s_i) = y_i | U(s_i) = u_i) \Pr(Y(s_j) = y_j | U(s_j) = u_j)
\]

for \( i \neq j \) and \( y_i, y_j, u_i, u_j = 0, 1 \).

**Proposition 1** Under the Markovian assumption, the parameters \( y_{u_00}(r), y_{u_01}(r), y_{u_10}(r) \) and \( y_{u_11}(r) \) are linear functions of \( u_{p00}(r), u_{p01}(r), u_{p10}(r) \) and \( u_{p11}(r) \) whose coefficients depend upon the smoothing parameters \( \lambda_0, \lambda_1 \). In fact,

\[
y_{u_00}(r) = (1 - \lambda_0) u_{p00}(r) + \lambda_1 u_{p01}(r)
\]

\[
y_{u_10}(r) = (1 - \lambda_0) u_{p01}(r) + \lambda_1 u_{p11}(r)
\]

\[
y_{u_01}(r) = \lambda_0 u_{p00}(r) + (1 - \lambda_1) u_{p01}(r)
\]

\[
y_{u_11}(r) = \lambda_0 u_{p01}(r) + (1 - \lambda_1) u_{p11}(r).
\]

**Proof** The quantity \( y_{u_00}(r) \) is equal to

\[
\Pr(U(s_1) = 0, Y(s_2) = 0) = \\
\Pr(U(s_1) = 0, Y(s_2) = 0, U(s_2) = 0) + \\
\Pr(U(s_1) = 0, Y(s_2) = 0, U(s_2) = 1).
\]

Under the Markovian assumption, this is equal to

\[
\Pr(Y(s_2) = 0 | U(s_1) = 0) \Pr(U(s_1) = 0, U(s_2) = 0) + \\
\Pr(Y(s_2) = 0 | U(s_1) = 0, U(s_2) = 1) \Pr(U(s_1) = 0, U(s_2) = 1) = \\
(1 - \lambda_0) u_{p00}(r) + \lambda_1 u_{p01}(r).
\]

Proposition 2 Under the Markovian assumption and the independence assumption, the parameters \( y_{p00}(r), y_{p01}(r) \), and \( y_{p11}(r) \) are linear functions of \( u_{p00}(r), u_{p01}(r) \) and \( u_{p11}(r) \) whose coefficients are determined by the smoothing parameters \( \lambda_0 \) and \( \lambda_1 \). In fact,

\[
y_{p00}(r) = (1 - \lambda_0) u_{p00}(r) + \\
2\lambda_1 (1 - \lambda_0) u_{p01}(r) + \lambda_1^2 u_{p11}(r)
\]

\[
y_{p01}(r) = \lambda_0 (1 - \lambda_0) u_{p11}(r) + \lambda_1 u_{p01}(r) + \\
(1 - \lambda_0) (1 - \lambda_1) u_{p01}(r) + (1 - \lambda_1) u_{p11}(r)
\]

\[
y_{p11}(r) = \lambda_0^2 u_{p00}(r) + 2\lambda_0 (1 - \lambda_1) u_{p01}(r) + \\
(1 - \lambda_1)^2 u_{p11}(r).
\]

Under the Markovian assumption and the independence assumption the proof is straightforward.

Now, \( y_{p01}(r) = \gamma_y(r) \) is the variogram of \( Y(s) \). Furthermore,

\[
y_{p00}(r) + 2y_{p01}(r) + y_{p11}(r) = 1 \quad (1)
\]

\[
y_{p01}(r) + y_{p11}(r) = \pi_y = E(Y). \quad (2)
\]

But, \( y_{p01}(\infty) \) is the sill of the variogram and therefore

\[
y_{p01}(\infty) = \pi_y (1 - \pi_y). \quad (3)
\]

Thus, knowledge of the variogram of \( Y(s) \) determines \( \pi_y \) (at least up to a choice of two possible values) and the equations (1) and (2), then determine \( y_{p00}(r) \) and \( y_{p11}(r) \).

We have shown that knowledge of the variogram of \( Y \) completely determines the parameters \( y_{p00}(r), y_{p01}(r), y_{p11}(r) \). Then the linear equations in proposition 2.1 and 2.2 can be solved to obtain the corresponding parameters for \( U \) and the cross parameters for \( U \) and \( Y \). This in turn, determines the variogram of \( U \) and the covariogram of \( U \) and \( Y \). In what follows, we obtain explicit expressions for these relationships.

**Proposition 3** The covariance function of the corrupt process is given by

\[
C_u(r) = Cov[Y(s_1), Y(s_2)] = \\
f^2(\lambda) C_u(r), \text{ for } r > 0,
\]

where

\[
f(\lambda) = 1 - \lambda_0 - \lambda_1.
\]
Moreover,
\[ E[Y(s)] = \pi_y = (1 - \lambda_1)\pi_u + \lambda_0(1 - \pi_u) \]
\[ \text{var}[Y(s)] = C_Y(0) = V_u(\lambda) + V_\pi(\lambda) \quad (5) \]
where
\[ V_\pi(\lambda) = \pi_u(1 - \pi_u)f^2(\lambda) \]
and
\[ V_u(\lambda) = \lambda_0(1 - \lambda_0)(1 - \pi_u) + \lambda_1(1 - \lambda_1)\pi_u. \]

**Proof** According to Proposition 2, the variogram of \( Y(s) \) is given by \( \gamma_{\pi_01}(r) \)
\[ \gamma_{\pi}(r) = \pi_01(r) = \lambda_0(1 - \lambda_0)\pi_{11}(r) + \lambda_1\lambda_1\pi_{11}(r) + (1 - \lambda_0)(1 - \lambda_1)\pi_{01}(r) + (1 - \lambda_1)\pi_{11}(r). \quad (6) \]
Moreover, from equations (1) and (2) applied to \( U \) give
\[ \pi_{11}(r) = \pi_u - \gamma_u(r) \]
and
\[ \pi_{00}(r) = (1 - \pi_u) - \gamma_u(r). \]
Substituting these expressions into equation (2.) gives
\[ \gamma_{\pi}(r) = \lambda_0(1 - \lambda_0)(1 - \pi_u) + \lambda_1(1 - \lambda_1)\pi_u + (1 - \lambda_0 - \lambda_1)^2\gamma_u(r). \]
Equation (4) is a consequence of the facts that \( C_Y(r) = \gamma_{\pi}(\infty) - \gamma_{\pi}(r) \) and \( C_u(0) = \pi_u(1 - \pi_u). \)
Equation (5) follows from the analysis of variance formula and the conditional distribution of \( Y \) and \( U \).

From equation (4), one obtains that
\[ \lim_{r \to 0^+} C_{\pi}(r) = V_\pi(\lambda) \]
Therefore, \( C_{\pi}(r) \) always has a discontinuity at the origin in this parametric model. In the same way, it is possible to derive the cross-covariance function between the true process and the corrupt process.

**Proposition 4** The cross-covariance function between the two processes is given by
\[ C_{\pi u}(r) = Cov[Y(s_1), U(s_2)] = f(\lambda)C_u(r), \quad \text{for } r > 0. \quad (7) \]

### 2.1 Indicator Smoothing

The kriging technique will be used to predict the smoothed process \( U(s) \), on the basis of the recorded process \( Y(s) \). The indicator smoother is given by
\[ \hat{U}(s_0) = \pi_u + \sum_{i=1}^{N} \alpha_i(Y(s_i) - \pi_y) \quad (8) \]
A comment is in order at this point. Kriging coefficients are obtained on the assumption that the variogram is known, even though it must be estimated in practice. For a binary process, the variogram determines the sill \( \pi(1 - \pi) \) which, in turn, determines the mean \( \pi \), at least up to a choice of two possible values. There will seldom be difficulty in resolving this choice. Accordingly, we obtained the kriging coefficients on the assumptions that \( \pi \) in the above equations is known, even though it will ultimately be estimated along with the variogram. For \( \hat{U} \) to be an unbiased predictor, we require that
\[ E[U(s_0) - \hat{U}(s_0)] = 0. \]
The weights \( \alpha_i \) are such that the mean square error is a minimum. In fact,
\[ E[(U(s_0) - \hat{U}(s_0))^2] = E[(U(s) - \pi_u)^2] + \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j (Y(s_i) - m)(Y(s_j) - m) - 2 \]
\[ \sum_{i=1}^{N} \alpha_i (Y(s_i) - m)(U(s_0) - \hat{U}(s_0)) = \]
\[ C_u(0) + \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j C_{\pi}(s_i - s_j) - 2 \sum_{i=1}^{N} \alpha_i C_{\pi u}(s_i - s_0) \]
It follows that the kriging system is given by
\[ \sum_{j=1}^{N} \alpha_j C_{\pi}(s_i - s_j) = C_{\pi u}(s_i - s_0), \quad i = 1, \ldots, N. \quad (9) \]
The covariance functions \( C_\pi \) and \( C_{\pi u} \) depend upon the parameters \( \lambda \). Different values of \( \lambda \) correspond to different weights \( \alpha \). When \( \lambda = 0 \), there is no smoothing and the method reproduce the original map. To see this observe that Propositions 3 and 4 imply that
\[ C_\pi(r) = C_u(r) \]
and
\[ C_{\pi u}(r) = C_u(r) \]
when \( \lambda_0 = 0 \) and \( \lambda_1 = 0 \) for every landcover category
The weights \( \alpha \) that solve the system (9) are \( \alpha_j = 0 \)
for all j except \( \alpha_j = 1 \) when \( j = 0 \). Then, we have that

\[
\hat{U}(s_0) = Y(s_0).
\]

Larger value of \( \lambda \in (0, .5) \) lead to a greater degree of smoothing. In the extreme case that \( \lambda = .5 \), all points are classified into a single category. When \( \lambda_0 = .5 \) and \( \lambda_1 = .5 \) for all categories, we have

\[
C_Y(r) = \begin{cases} 
0.25 & r = 0 \\
0 & r \neq 0
\end{cases}
\]

and

\[
C_{Ym}(r) = 0.
\]

The solution of the equation (9) is \( \alpha = 0 \) and equation (8) implies that

\[
\hat{U}(s_0) = Y(s_0).
\]

Therefore, all point will be classified into the single category whose value \( \pi_u \) is a maximum.

3. Some Examples

To show how the indicator smoother works Geostatistical simulation techniques are used to generate spectral response with different degree of spatial variability.

Different images, of dimension 128 \( \times \) 128 (for a total of 16384 pixels) and with four band, have been simulated. It has been chosen a correlation between the four band equal to .7. The variogram for the spectral response can be, for example, is the exponential, given by

\[
\gamma(r) = \theta_0 + \theta_1(1 - \exp(-r/\theta_2)).
\]

The simulated images with 4 spectral response have been compressed in 64 clusters using a PHASE approach. PHASE stand for for Pixel Hyper-clusters Approximating Spatial Ensembles. It is a classification of pixels into K groups (here K=64) on the basis of similarity of spectral responses in the several bands (for more details see Mayer 1997).

A Kullback-Liebler distance approach have been used to classified the compressed image in three land-cover type. For images with a small nugget effect and strong spatial variability, the Kullback-Liebler distance produce thematic maps that show a spatial continuity. Clearly, as the noise level increase the spatial coherence of the spectral classification decrease.

The thematic map obtained with the Kullback-Liebler is used to estimate the three variograms of the corrupt process \( \gamma_Y(r) \). The variograms of the true process can be estimate considering that \( \gamma_Y(r) \) and \( \gamma_u(r) \) are related by

\[
\gamma_Y(r) = \lambda_0(1 - \lambda_0)(1 - \pi_1)+
\]

\[
\lambda_1(1 - \lambda_1)\pi + (1 - \lambda_0 - \lambda_1)^2\gamma_u(r),
\]

Since the exponential variogram model approximates the variogram for binary process, the empirical variograms have been fitted with the exponential variogram, therefore,

\[
\gamma_Y(r) = \theta_0^* + \theta_1^*(1 - \exp(-r/\theta_2)),
\]

for chosen values of \( \lambda_1 \) and \( \lambda_2 \) and where \( \theta_0^* = 1/2[\lambda_0(1 - \lambda_0) + \lambda_1(1 - \lambda_1)][1 - \sqrt{1 + 4\theta_1^*}] \) and \( \theta_1^* = (1 - \lambda_0 - \lambda_1)^2\theta_1 \).

The indicator smoothing is applied at the pixel level and it will move each pixel by the class assigned with the spectral classification to that class where the plurality of pixels in its neighborhood belong. In the figure (1) are showed the compressed image with a covariance function \( C(h) = 1 + \exp(-0.15h) \), the Kullback-Liebler classification and, the smoothing kriging apply to the Kullback-Liebler classification with increasing value of \( \lambda \).

![Figure 1: Example of Kullback-Liebler classification and smoothing kriging for simulate spectral response with \( \theta_0 = 1 \) and \( \theta_2 = 65 \)](image)

The smoothing parameters \( \lambda \) are not estimable unless a rather large validation set is available, in which case the \( \lambda \) can be chosen to minimize the classification error. Also the magnitude of \( \lambda \) indicates the relative importance of the spatial information in the classification.
An open question is whether the optimal $\lambda$ can be used as landscape indicator to differentiate among different class of landscape.

References


