Modelling Uncertainty in Satellite Derived Land Cover Maps

Edward Cripps [*] Anthony O'Hagan[†] Tristan Quaife [‡]
Clive Anderson [†]

March 24, 2009

Abstract

Maps of land cover, derived from satellite observation, are widely used as inputs to environmental models. Because of errors in the classification process, the true land cover category at a given site may differ from the category assigned in the map. It is generally important to account for uncertainty in the inputs to models whose outputs may be used for policy-making, and we address here the problem of quantifying uncertainty in land cover maps.

There is a substantial literature on estimating land cover, but most authors do not address the uncertainty in the resulting estimates. Our approach characterises uncertainty through a Bayesian approach which incorporates several layers of modelling. In particular, we incorporate data in the form of a 'confusion matrix' derived from a sample of ground truth observations, and discuss the appropriate way to model such data. We also account for two forms of spatial correlation. First, there is spatial correlation in the true land cover map, with land cover expected to be similar at neighbouring pixels; second, there is spatial correlation in the error characteristics of the satellite-derived map classification. Analysis of a model in which the data in the confusion matrix and both forms of spatial correlation were modelled explicitly would be computationally very demanding. Instead we adopt

[*]Australian Institute of Marine Science, Townsville
[†]Department of Probability and Statistics, University of Sheffield
[‡]Department of Geography, University College London.

simplified and implicit modelling with the result that we are able to develop straightforward computation of posterior means, variances and covariances.

We characterise uncertainty in the land cover at individual pixels through posterior probabilities of mis-classification. We also consider the important case where inputs to the environmental model are proportions of land cover aggregated to a coarser grid, and derive posterior means, variances and covariances for these proportions. We present the results of our method applied to a recently developed satellite derived land cover map, the Land Cover Map 2000, for the region of England and Wales.

Keywords: land cover; thematic map; confusion matrix; error matrix; producer's risk; user's risk; mis-classification probabilities; spatial correlation; remote sensing; environmental model; global vegetation model

1 Introduction

Over the last few decades remote sensing land cover maps (henceforth RS maps) have become an important source of information on contemporary land cover. To construct RS maps satellites record energy incident at a sensor, situated several hundreds of kilometres above the surface of the earth, at a grid of pixels within a region of interest. Inferring the land surface information for each pixel then becomes a problem of image classification from degraded observations. Different classification methods will result in different RS maps and often show poor agreement, whilst reporting the same physical quantities of interest, for the same region (see Giri et al. (2005) and Hansen and Reed (2000)). Errors in classification result from the remote sensing instrumentation, the conditions under which the observations were made and the algorithm used to infer land cover from the observed sensor readings.

One way of assessing the accuracy of a map is to observe directly the true land cover at a sample of pixels within the map area. The resulting contingency table presenting counts of pixels classified by true and RS map land cover is called a confusion matrix (or error matrix). Probabilities of correct classification across the map may be estimated by frequentist methods (e.g. Stehman et al, 2003) or using Bayesian analysis (Green and Strawderman (1994)).

An important component of the image classification problem is the spatial nature of the data. That is, neighbouring pixels are more likely in some

sense to be classified as the same land cover type than pixels situated further apart. To address this issue, local spatial continuity approaches have been discussed in Switzer (1980), Klein and Press (1989), Klein and Press (1990a) and Klein and Press (1990b). Geman and Geman (1984), Besag (1986), Klein and Press (1992), Frigessi and Stander (1994), Divino et al. (2000) and Storvik et al. (2005) use Markov random field priors to account for spatial correlation.

A second spatial component is correlation in the map errors between neighbouring pixels. For instance, if one pixel is covered by crops but is misclassified by the RS map as scrub, then it is more likely that a neighbouring crop pixel will also be classified as scrub. This arises partly because the characteristics of the crop in the first pixel that make it difficult for the remote sensing device to classify correctly (e.g. it is a tall, woody crop) are likely to apply also in the neighbouring pixel. Furthermore the two pixels are likely to have been observed in the same pass of the satellite under the same conditions (e.g. weather and incidence angle), which may also induce the same kind of classification error. To our knowledge, this spatial correlation in the observation process has not been explicitly recognised in the literature.

All three components of the problem — the confusion matrix, spatial correlation in land cover and spatial correlation in map errors — are complex to model accurately, and careful modelling will lead to very computationally intensive inference. We present here an approach in which simplifications and approximations are used to obtain a more tractable analysis. The result is readily implemented by researchers and users of RS maps, and we believe is the first treatment of the problem that incorporates all three components.

RS maps are often used to provide inputs to complex environmental simulation models. For instance, land cover is typically required as an input to global dynamic vegetation models [refs]. Simulation models are increasingly used to inform high-level decisions and policy-making [refs IPCC etc], and users are demanding to know how accurate the model predictions are. One driver of the uncertainty in model outputs is uncertainty in model inputs, so it is important to be able to quantify the uncertainty in RS maps. However, the statistics literature relating to image classification to date has focused on point estimation of the map, the primary reason being the computational burden due to the typically large data sets associated with RS maps. To our knowledge, uncertainty of a real RS map has not been explicitly reported. We employ a Bayesian approach so that inference about land cover at a given

pixel involves deriving the posterior probabilities for the true land cover.

Simulation models often operate at a lower resolution than the available RS maps, and so for the purpose of model inputs we may require the original high resolution RS product to be upscaled to coarser resolutions (Jung et al. (2006) and Foody (2002)). We therefore also consider inference for an aggregation of pixels. We refer to such an aggregate area as a site, and the land cover at a site is defined by proportions of the site covered by each of the different land cover categories. Whereas uncertainty about land cover at a single pixel is characterised by posterior probabilities of mis-classification, we present uncertainty at a site in terms of posterior variances and covariances for these proportions.

Land cover is one of a number of products derived from remote sensing. Although we have used land cover to motivate and introduce the ideas, the methods developed here will be applicable more generally to what are known as thematic maps. An RS thematic map assigns each pixel to one of a number of discrete classes, and typically share all the features described above for land cover maps.

The paper proceeds as follows. Section 2 discusses how to model the confusion matrix and introduces the spatial correlation components of the problem. Section 3 introduces the statistical model by first describing inference for a single pixel. Section 4 extends this to inference at the site level. Section 5 provides the details of the final product of a real data set - the Land Cover Map 2000 - to which we apply our model. Section 6 displays the results as maps of simulated posterior means and standard deviations of the proportions of vegetation classes in England and Wales and also the posterior distributions for the region as a whole. The results provide information to the remote sensing community as to where, and for which land cover classes, there is greater potential for mis-classification. Section 7 concludes with a review of the simplifying assumptions made in our approach, and considers directions for future research.

2 Modelling the confusion matrix

The confusion matrix arises from a survey in which a sample of pixels is classified, typically either by a visit to the location or by aerial photograph. We refer to the resulting land cover class as the true class at that pixel, although there may still be errors in this classification (particularly when

aerial photography is used). In recognition of the remaining possibility of error, what we call the true class is sometimes known as the reference class. We wish to use the confusion matrix to learn about error probabilities.

Let there be k vegetation classes identified in the RS map. Then the confusion matrix is a $k \times k$ contingency table of counts of pixels classified by true land cover and by map land cover. We identify the map classifications by rows of the matrix, so that its element $c_{t't}$ is the number of sampled pixels over the region whose RS map classification is t' and whose true vegetation class is t. We write the t-th column as

$$\boldsymbol{c}_t = (c_{1't}, c_{2't}, \ldots, c_{k't})^T, \text{ for } t = 1, 2, \ldots, k,$$

and the whole matrix is

$$\mathbf{C} = (\boldsymbol{c}_1^T, \boldsymbol{c}_2^T, \ldots, \boldsymbol{c}_k^T)^T.$$

Relating the confusion matrix to error probabilities

For a given pixel in the map, let ξ be the true class and let x be the RS map class. Then we have an error if ξ is not equal to x, but there are two ways to express the probability of an error. We denote the *backward probabilities* by

$$\kappa_{tt'} = P(\xi = t | x = t') \quad \text{for} \quad t, t' = 1, 2, \ldots, k,$$

which expresses the probability of the true class conditional on the observed map class. It is these probabilities that we primarily require to make inference about true land cover from the RS map. However, we also recognise the *forward probabilities*

$$\lambda_{t't} = P(x = t' | \xi = t) \quad \text{for} \quad t, t' = 1, 2, \ldots, k,$$

which express the probability of the map class conditional on the true class.

In the remote sensing community the quality of an RS map is quantified by the following accuracy measures, which may be derived from a confusion matrix (Stehman et al, 2003; Mayaux et al. (2006); Foody (2007)):

1. The overall accuracy: the probability of correctly classifying pixels across the entire map.

2. The producer's accuracy: the probability that a pixel is classified as vegetation class t' by an RS map given that it has true vegetation class t'. This is the forward probability $\lambda_{t't'}$.

3. The user's accuracy: the probability that a pixel has true vegetation class t given that the pixel is classified as vegetation class t by the RS map, which is the backward probability κ_{tt}.

The sample used to create the confusion matrix is usually obtained by stratified sampling in which the numbers of pixels having each of the map classes (the row totals of $\mathbf{C}$) are fixed. The sampling plan can be more complex than this, and for frequentist inference about error probabilities its structure is important (Stehman et al, 2003). In a Bayesian analysis the sampling method can also be important for how we model the confusion matrix.

In particular, Green and Strawderman (1994) model a confusion matrix obtained by stratified sampling according to map class by supposing that the rows are independent multinomial samples with probabilities the backward probabilities. They thereby obtain inference very directly about the backward probabilities (assuming independent conjugate Dirichlet priors). However, we believe that this model is incorrect.

The forward probabilities $\lambda_{t't}$ characterise the underlying error process of mis-classification. These error probabilities are properties of the remote sensing instrument and the classification algorithm. In contrast, the backward probabilities do not represent intrinsic properties of the process, and are context dependent. To clarify this point, suppose that there is a non-zero forward probability of mis-classifying crops as scrub, and first consider a region in which a large proportion of the pixels are in fact covered by scrub, with a small proportion covered by crops. If we now sample from the pixels that are classified as scrub by the RS map, then the great majority of these will have scrub as their true class, so the backward error probability that a pixel with map class scrub is truly crops will be small. However, if the region actually has much more crops than scrub then a much higher proportion of pixels classified as scrub will in fact be crops, implying a higher backward error probability. The backward probabilities are dependent on the underlying true mix of land cover.

This phenomenon is recognised in the literature dealing with image analysis, where an isolated scrub pixel in a region which is mostly crops will be inferred to be an error and its true class estimated to be crops, whereas a pixel classified as scrub with many neighbouring scrub pixels will be estimated as truly scrub. The neighbouring pixels are used to learn about the local mix of land cover, and this induces a context which can completely change the backward probabilities. In these studies, it is effectively the forward

probabilities that are assumed to be fundamental, and we believe that this is the correct modelling approach. Although it is correct to regard the rows of $\mathbf{C}$ as independent multinomial samples, the backward probabilities that may be obtained from this approach are dependent on the mix of land cover in the region and cannot be applied to any given pixel conditional on the whole map.

We therefore seek to infer forward probabilities from the confusion matrix. The backward probabilities in the multinomial samples are functions (via Bayes' theorem) of the forward probabilities and the true proportions

$$\pi_t = P(\xi = t), \quad \text{for} \quad t = 1, 2, \ldots, k,$$

of different land cover classes over the entire region. To address this properly implies jointly modelling the forward probabilities and the true land cover across the region, and will lead to computational very intensive inference. Instead, we employ a simplifying approximation.

We propose to ignore the stratified nature of the sample, and to treat the confusion matrix as arising from a single simple random sample. This will clearly be inappropriate and a poor approximation if the stratified sampling has strongly over-sampled or under-sampled some vegetation classes, but in practice such samples generally employ proportional stratified samples (with the aim of ensuring even coverage of map classes). In this context, all pixels have an equal probability of being sampled and we argue that it is reasonable to treat the matrix as a single multinomial sample. See Section 7 for further discussion of this assumption.

Inference about the forward probabilities

We therefore can model the confusion matrix as comprising independent multinomial *columns* $\boldsymbol{c}_t$ with parameter vectors the *forward* probabilities

$$\boldsymbol{\lambda}_t = (\lambda_{1't}, \lambda_{2't}, \ldots, \lambda_{k't})^T, \quad t = 1, 2, \ldots, k.$$

because under the assumption of simple random sampling the likelihood factorises as

$$P(\mathbf{C} \,|\boldsymbol{\lambda}, \boldsymbol{\pi}) = \left\{ \prod_{t=1}^{k} \prod_{t'=1}^{k} (\lambda_{t't})^{c_{t't}} \right\} \left\{ \prod_{t=1}^{k} (\pi_t)^{c_{\cdot t}} \right\}, \tag{1}$$

where $c_{\cdot t} = \sum_{t'=1}^{k} c_{t't}$ and we write $\boldsymbol{\lambda} = \{\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_k\}$, $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_k)^T$.

The first term in (1) is the product of k multinomial likelihoods

$$P(\boldsymbol{c}_t\,|\boldsymbol{\lambda}_t, c_{\cdot t}) = \prod_{t'=1}^{k} (\lambda_{t't})^{c_{t't}}, \tag{2}$$

for the forward probabilities from the columns of $\mathbf{C}$, conditional on the column totals, while the last term is a multinomial likelihood for $\boldsymbol{\pi}$ from the column totals.

The sample size from which a confusion matrix is derived will not usually be very large, and there will be individual classes for which the column totals are relatively small. In this context, prior information can add appreciably to the precision of inferences. We note that Green and Strawderman (1994) also incorporated prior information in their analysis, but in their case this is prior information about the backward probabilities. They elicited expert opinion and said, "The experts were told what area of New Jersey was under study, in case their estimates would be geographically dependent." It is clear that in this respect they recognised that backward probabilities depend on context. In contrast, we utilise prior information about forward probabilities, which relate to the accuracy of the mapping process in general.

We assume that *a priori* the $\boldsymbol{\lambda}_t$s are independent of each other and assign $\boldsymbol{\lambda}_t$ a conjugate Dirichlet prior density $Di(\boldsymbol{\alpha}_t)$ with parameters $\boldsymbol{\alpha}_t = (\alpha_{1't}, \alpha_{2't}, \ldots, \alpha_{k't})^T$ for $t = 1, 2, \ldots, k$. Thus,

$$P(\boldsymbol{\lambda}|\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_k) = \prod_{t=1}^{k} P(\boldsymbol{\lambda}_t|\boldsymbol{\alpha}_t) \tag{3}$$

where

$$P(\boldsymbol{\lambda}_t|\boldsymbol{\alpha}_t) \propto \prod_{t'=1}^{k} \lambda_{t't}^{\alpha_{t't}-1} \quad \text{for } t = 1, 2, \ldots, k. \tag{4}$$

From (2), (3) and (4) the posterior distribution of $\boldsymbol{\lambda}$ is

$$\begin{aligned} P(\boldsymbol{\lambda}|\mathbf{C}) &\propto P(\mathbf{C}|\boldsymbol{\lambda}) P(\boldsymbol{\lambda}|\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_k) \\ &\propto \prod_{t=1}^{k} \prod_{t'=1}^{k} \left\{ (\lambda_{t't})^{c_{t't}+\alpha_{t't}-1} \right\}. \end{aligned}$$

Conditioned on the confusion matrix, the posterior distributions of the region-wide forward probability vectors are independent and Dirichlet with parameters $(\boldsymbol{c}_t + \boldsymbol{\alpha}_t)$ and denoted by

$$\boldsymbol{\lambda}_t|\boldsymbol{c}_t, \boldsymbol{\alpha}_t \sim Di(\boldsymbol{c}_t + \boldsymbol{\alpha}_t). \tag{5}$$

The assumption of independent Dirichlet prior distributions is also discussed in Section 7.

Given the RS map classification the backward probabilities could be modelled as mathematically implied by the forward probabilities and the prior probabilities $\boldsymbol{\pi}$ of the true vegetation class for the entire region, since together these imply the joint distribution. However, as already discussed, these region-wide backward probabilities do not relate to individual pixels, for which the local context changes the prior probabilities of the vegetation classes. We now extend the notion of incorporating a prior distribution for the true vegetation classes to the pixel level and hence model spatial variability within the backward probabilities.

3 Inference at the pixel level

Inference at the pixel level involves identifying the posterior probabilities of the true vegetation classes given the RS map classification, i.e. the backward probabilities. We now recognise that these are pixel-specific and modify the notation of Section 2 to denote the probability that pixel p has true class t given that it has map class t' by $\kappa^p_{tt'}$. Let π^p_t be the prior probability that the true vegetation class is t at pixel p. The image analysis literature for this problem models the underlying true map as a spatial random field. Authors have generally assumed a specific form of observation error, rather than deriving inference about this process from a confusion matrix. Furthermore, because of the complexity of the full joint posterior probability distribution of the true map they have concentrated on obtaining simply a posterior estimate. For example, Geman and Geman (1984) use simulated annealing in a general continuous random field to produce the maximum a posteriori (MAP) estimate, Besag (1986) develops the method of iterative conditional modes (ICM, which produces a local maximum of the posterior joint distribution, but not necessarily the MAP estimate) and Klein and Press (1992) outline the technique of adaptive Bayesian classification for thematic maps. For simplicity, we offer here an approximate analysis which avoids the heavy computations inherent in full joint modelling of the spatial field and the error characteristics (forward probabilities).

We suppose that there is no explicit prior knowledge about the true classes and equate each π^p_t to be the proportion of pixels observed to be in vegetation class t by the RS map in pixel p and in some neighbourhood of pixels around

p. Hence for a given pixel we set the prior probabilities to

$$\pi_t^p = \frac{n_t^p}{n^p}, \quad t = 1, 2, \ldots, k,$$

where n_t^p is the number of pixels in that neighbourhood classified by the RS map as vegetation class t and n^p is the total number of pixels in the neighbourhood. We believe that this will generally represent a good approximation to a more realistic (but computationally much more complex) model in which the prior probabilities are given a random field distribution with neighbourhood-based spatial covariance structure. Specifically, in ICM the computation passes iteratively over the map and at each step the prior distribution for the current pixel is computed in the same way using the current assignments of pixels in the neighbourhood (rather than simply the map assignments as proposed here). A full Markov chain Monte Carlo (MCMC) analysis would have the same characteristics. Our approach mimics MCMC, and like MCMC can deliver the posterior backward probabilities for each pixel rather than simply an estimate of the true map, but is much simpler. Further discussion of this approach is given in Section 7.

If pixel p has map class t', the pixel-specific backward probabilities are now given by Bayes' theorem as

$$\kappa_{tt'}^p = \frac{\lambda_{t't}\pi_t^p}{\sum_{t=1}^k \lambda_{t't}\pi_t^p} \quad \text{for} \quad t = 1, 2, \ldots, k. \tag{6}$$

To compute any required inferences of the $\kappa_{tt'}^p$s it is a simple matter to simulate the forward probabilities from their Dirichlet distributions (5) and to evaluate (6) for each set of simulated $\lambda_{t't}$s.

We have presented a simple model for the pixel-specific backward vectors that uses the confusion matrix to model the forward probabilities and incorporates spatial variability in the pixel-specific backward probabilities via a 'surrogate' prior evaluated at each pixel from the RS map. The next section describes inference for aggregating the pixel level to the site level, and introduces a second form of spatial correlation.

4 Aggregation to the site level

Although RS maps are able to provide pixels at high resolutions, users of maps often require aggregation to a coarser resolution (Jung et al. (2006)

and Foody (2002)). We call these larger grid sizes sites and now describe inference about proportions of the true vegetation classes in a site. Although in principle such inferences should be derivable from the analysis of individual pixels in the preceding section, there are three complications. First, our analysis has only addressed pixels individually, and so we have not derived joint probabilities, such as the probability that the true classes of two adjacent pixels are both crops. To derive joint probabilities rigorously would require more careful treatment of the spatial correlation structure of the true land cover. It would also entail modelling of the other kind of spatial correlation referred to in Section 1, the correlation in the observation error process. Second, any model of spatial correlation in errors would be speculative and contain parameters for which we have essentially no information. The samples used to derive confusion matrices are too sparse for us to be able to pick up such local features. Third, even if we could construct such joint probabilities a site can contain hundreds of pixels, so we would need to construct very high-dimensional joint distributions. Again, thorough modelling would lead to immensely complex computations, so we introduce further simplifications.

Site-specific forward probabilities

We have argued in Section 2 that it is the forward probabilities, not the backward probabilities, that are independent of context. However, we have also said that there is spatial correlation, so that the probability that two adjacent pixels which are actually covered by class t are both mis-classified as class $t' \neq t$ is not $(\lambda_{t't})^2$. Instead of modelling spatial correlation in these forward probabilities explicitly, we adopt an implicit approach which we motivate as follows.

A consequence of correlation will be that the proportions of pixels at sites s that are correctly or incorrectly classified in particular ways will be more variable than if they were independent. Without correlation, if we had enough pixels in a given site, then the proportions of all the pixels in that site having true vegetation class t that are observed to be in vegetation class t' will almost certainly be very close to $\lambda_{t't}$; there will be essentially no uncertainty. At the other extreme, if all the pixels in the site with vegetation class t are so highly correlated that they are always observed to be in the same vegetation class, then the resulting proportions of correct and incorrect classifications will be almost as uncertain as the observation of just one pixel.

Our approach is instead to create a comparable effect by letting the pixels still be independent, but allowing the forward probability vectors be site-

specific, denoted by $\boldsymbol{\lambda}_t^s$, varying randomly from the countrywide vector $\boldsymbol{\lambda}_t$ according to a Dirichlet distribution:

$$\boldsymbol{\lambda}_t^s | \boldsymbol{\lambda}_t, d \sim Di(d\boldsymbol{\lambda}_t)$$

independently for each site in the region. This introduces a single parameter, d, to control the degree of increased variability, and hence implicitly the degree of correlation at the pixel level. This model and the interpretation of d are discussed further in Section 7

In order to derive site-specific backward probabilities, we again use the RS map to provide prior probabilities. We denote these by π_t^s, and derive them as with the pixel-specific probabilities π_t^p by averaging over all the pixels in a neighbourhood of the site s. Then the site-specific backward vectors are

$$\kappa_{tt'}^s = \frac{\lambda_{t't}^s \pi_t^s}{\sum_{t=1}^k \lambda_{t't}^s \pi_t^s} \quad \text{for} \quad t, t' = 1, 2, \ldots, k. \tag{7}$$

Site-specific true vegetation class proportions

We now deal with the inference about the true proportions of vegetation classes within a site. Let the true proportion of class t at site s be γ_t^s, for $t = 1, 2, \ldots, k$. It is the uncertainty in the γ_t^ss that we wish to characterise. Let the number of pixels having vegetation class t' at site s according to the RS map be $n_{t'}^s$ and let n^s be the total number of the RS map pixels in the site.

For site s the vector of proportions of true vegetation classes is written as

$$\boldsymbol{\gamma}^s = (\gamma_1^s, \gamma_2^s, \ldots, \gamma_k^s)^T,$$

and the site-specific backward vectors as

$$\boldsymbol{\kappa}_{t'}^s = (\kappa_{1t'}^s, \kappa_{2t'} \ldots, \kappa_{kt'}^s)^T, \text{ for } t' = 1, 2, \ldots, k.$$

Conditional on each $\boldsymbol{\kappa}_{t'}^s$ we treat the individual pixels in site s as independently observed by the RS map, and accordingly assume that the true vegetation class for each pixel can be independently inferred using the relevant $\boldsymbol{\kappa}_{t'}^s$ vector. Then, conditional on the relevant $\boldsymbol{\kappa}_{t'}^s$s, $\boldsymbol{\gamma}^s$ is a linear combination of k independent multinomial distributions:

$$\boldsymbol{\gamma}^s | \boldsymbol{\kappa}_1^s, \boldsymbol{\kappa}_2^s, \ldots, \boldsymbol{\kappa}_k^s \sim \frac{1}{n^s} \sum_{t'=1}^{k} M(n_{t'}^s, \boldsymbol{\kappa}_{t'}^s), \tag{8}$$

where $n^s_{tt'}$ is the number of pixels in site s that are assigned by the RS map to class t' but are actually class t and $M(n^s_{t'}, \boldsymbol{\kappa}^s_{t'})$ is the multinomial distribution over $n^s_{t'}$ observations with probability vector $\boldsymbol{\kappa}^s_{t'}$.

The following simple Monte Carlo simulation scheme can be used to draw a sample of size N from the posterior distribution of $\boldsymbol{\gamma}^s$ for individual sites and for the whole region:

For $g = 1, 2, \ldots, N$

(1) Draw $\boldsymbol{\lambda_t} \sim Di(\mathbf{c}_t + \boldsymbol{\alpha})$ for $t = 1, 2, \ldots, k$

For $s = 1, 2, \ldots, S$

(2) Calculate π^s_t for $t = 1, 2, \ldots, k$ from a neighbourhood of s

(3) Draw $\boldsymbol{\lambda}^s_t \sim Di(d\boldsymbol{\lambda}_t)$ for $t = 1, 2, \ldots, k$

(3) Calculate $\boldsymbol{\kappa}^s_{t'}$ for $t' = 1, 2, \ldots, k$ from (7)

(4) Draw g-th sample value of $\boldsymbol{\gamma}^s \sim \frac{1}{n^s} \sum_{t'=1}^{k} M(n^s_{t'}, \boldsymbol{\kappa}^s_{t'})$

end

end.

Notice that this algorithm samples from a joint posterior distribution for all the $\boldsymbol{\gamma}^s$ vectors for all the sites. We need this in order to characterise uncertainty when the land cover at all sites in the region is to be input to an environmental model. The algorithm assumes that sites are large enough for it to be reasonable to ignore both forms of spatial correlation between sites, so that it is only uncertainty about the common $\boldsymbol{\lambda}_t$ vectors that induces some correlation in the joint distribution. We do not address the joint distribution of backward probabilities across pixels in Section 3 because we could not have made this assumption for individual pixels.

5 Real data: Land Cover Map 2000

The data analysed in this paper is derived from the Land Cover Map 2000 (LCM2000). The LCM2000 is an RS map produced by the UK Centre for Ecology and Hydrology (CEH) (Haines-Young et al., 2000). The base spatial resolution of the LCM2000 is a 25m $\times$ 25m spatial grid covering Great Britain. Each pixel contains one of 26 subclasses of vegetation types classified from spectral information collected from various orbital sensors. The data we analyse is the LCM2000 for England and Wales and manipulated

in two ways in order to be suitable for input to a global dynamic vegetation model.

First, the 26 vegetation types are grouped into 5 plant functional types (PFTs). The PFTs are deciduous broadleaf (DcBl), evergreen needleleaf (EvNl), Grassland, Crop, and Bare. DcBl consists of both trees and shrubs that lose their foliage every year. EvNl consists predominantly of coniferous trees. The Grassland and Crop PFTs differ in that crops are harvested at the end of the year, thus having an artificially shortened leaf lifespan, and the harvested biomass is removed from the system. Bare is urban landscape or water.

Second, the spatial resolution is upscaled. The original 25m $\times$ 25m pixel measurements, in the British Ordnance Survey grid, were reprojected into geographic longitudes and latitudes and then aggregated to $1/6^{th}$ degree resolution sites. The proportions of the various PFT types were preserved in this process and a record of the total number of pixels (or counts) contributing to each $1/6^{th}$ degree grid cell was kept. The area of a site measured in degrees varies as a function of latitude, so the pixel count per site (n^s) is not constant. The minimum number of counts per site in the data set is 150 and the maximum is 200.

The final result of the LCM2000 product used in our analysis is 707 sites representing $1/6^{th}$ degree resolution for England and Wales with the LCM2000 recordings of proportions of the 5 PFTs in each site (and the total number of pixels in each site). Although Bare is included in the model, we only report the results here for the PFTs DcBl, Crop, Grassland and EvNl.

Figure 1 displays maps of proportions of DcBl, Crop, Grassland and EvNl as classified by LCM2000 sites across England and Wales. Figure 1 shows that DcBl and EvNl occupy a relatively small proportion of land cover across England and Wales, while Crops and Grasslands dominate the region with Crops predominant in the east and Grasslands in the west. Spatial correlation is obvious in all four maps. Note, there is an area in the south-east where none of DcBl, Crop, Grassland and EvNl have a large proportion. This is the greater London metropolitan region and is dominated by Bare.

Also available is a confusion matrix. Fuller et al. (2002) report a confusion matrix of the 26 vegetation classes between LCM2000 and the Country Side field survey (CS2000) in the year 2000 for England and Wales. To obtain the confusion matrix Fuller et al. (2002) stratify England and Wales into regions of vegetation classes and individual confusion matrices are calculated

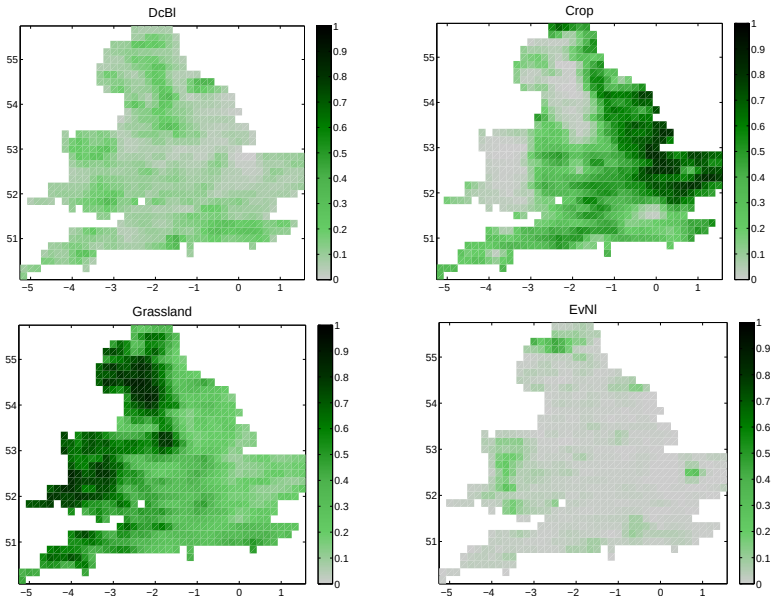

Figure 1: Maps of the proportions of PFTs DcBl, Crop, Grassland and EvNl for England and Wales as recorded by LCM2000.

for each vegetation stratum. These confusion matrices are combined to give the matrix reported in Fuller et al. (2002), each matrix contributing a weighting according to its extent in England and Wales. Table 1 reports the confusion matrix aggregated from the 26 LCM2000 classes to the 5 PFTs, calculated from that in Fuller et al. (2002). Although the sample involved quite complex stratification, Table 1 shows that the row totals match the percentage of pixels classified in each of the 5 PFTs across the whole map to well within the bounds that would be expected with a simple random sample. We therefore model the confusion matrix as in Section 2.

6 Results

We now present the results of our model applied to the data described above.

Hyperparameters

To apply our approach it is first necessary to specify values for the hyperparameters $\boldsymbol{\alpha}_t$ of the region-wide forward vectors, the neighbourhood structure for site-specific prior probabilities and the parameter d that governs the de-

	CS2000					
LCM2000	DcBl	EvNl	Grassland	Crop	Bare	TOTAL
DcBl	66	3	19	4	5	97
EvNl	8	20	1	0	0	29
Grassland	31	5	356	22	15	429
Crop	7	1	41	289	9	347
Bare	2	0	3	8	81	94
TOTAL	114	29	420	323	110	996
Whole map %	11.0	3.1	42.1	32.1	11.6	100.0

Table 1: Confusion matrix between LCM2000 and CS2000 for England and Wales. Final row shows composition of the full LCM2000 map.

gree of variability across the site-specific forward vectors.

Although in general we can incorporate prior information about the classification error probabilities, in this case we assign an uninformative prior distribution via

$$\boldsymbol{\alpha}_t = (1, 1, 1, 1, 1)^T, \quad \text{for} \quad t = 1, 2, \ldots, 5.$$

In the light of the interpretation of d discussed in Section 7, we choose $d = 1$. This represents a belief in rather strong dependence in the classification over a site, so that $d = 1$ is perhaps too small. However, we prefer to err on the side of assuming too much uncertainty than too little. When discussing the posterior standard deviations below we also present maps to assess the sensitivity to d.

The neighbourhood around each site is defined as a square covering 5 sites in each direction, so that it includes 121 sites (with between 150 and 200 pixels in each site). Figure 2 displays the plots of π_t^s for PFTs DcBl, Grassland, Crop and EvNl. Figure 2 captures the general pattern of the LCM2000 maps in Figure 1, and expresses a context in which Grassland is more likely *a priori* in the West, while Crop is more likely in parts of the East.

Posterior Means

Figure 3 displays the maps of the simulated posterior means (henceforth referred to as the means) of the true PFT proportions DcBl, Crop, Grassland and EvNl at each site. The means for DcBl, EvNl and Crop appear similar to the LCM2000 maps in Figure 1. In particular, the LCM2000 measurements of DcBl are at their highest in regions of the north and south of Wales, areas

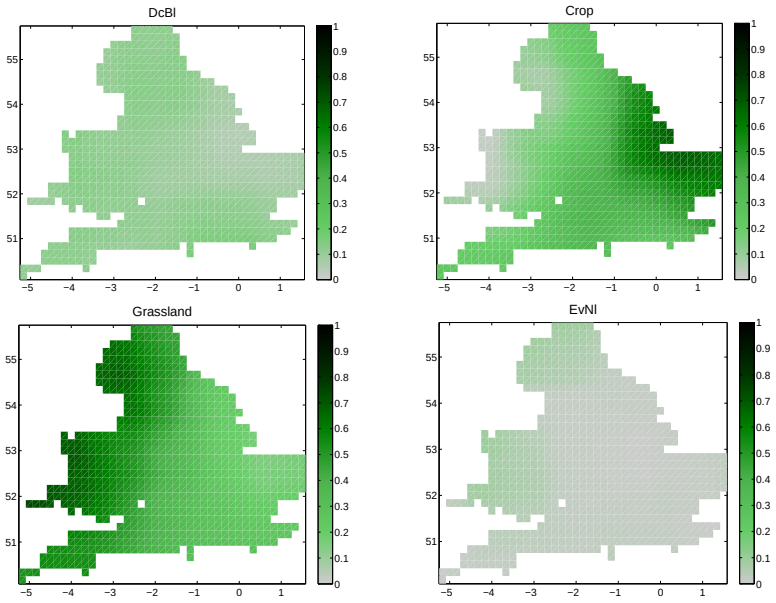

Figure 2: Maps of the priors of PFTs DcBl, Crop, Grassland and EvNl (clockwise from top left respectively).

in the north-west and central regions of England (corresponding to the Lake and Peak Districts), and in south-eastern England and are in agreement with the LCM2000 recordings. The LCM2000 recordings and the means also agree that EvNl is more abundant in the Lake District and south Wales. Additionally, Thetford forest in East Anglia (the area east of the meridian and north of 52 degrees), the largest lowland pine forest in the UK, is clearly identified as EvNl. The means of Crop closely follow the LCM2000 map which records Crop as prevalent in the north-east of England, the mid-east coast (East Anglia) and south England although the means are smaller compared to the LCM2000 map in north-east England and the Midlands. The means of Grassland follow the same pattern as the LCM2000 map and indicate that Wales and the Peak and Lake Districts are mainly populated by Grassland. However, Figure 3 implies the proportions of Grasslands in these regions are smaller than the LCM2000 measurements. In general, our model appears to estimate values of DcBl and EvNl higher than reported by the LCM2000 and estimates values of Grassland and Crop lower than reported by the LCM2000. We return to this smoothing issue and its reason when discussing the results relating to the region-wide PFTs below.

Posterior standard deviations

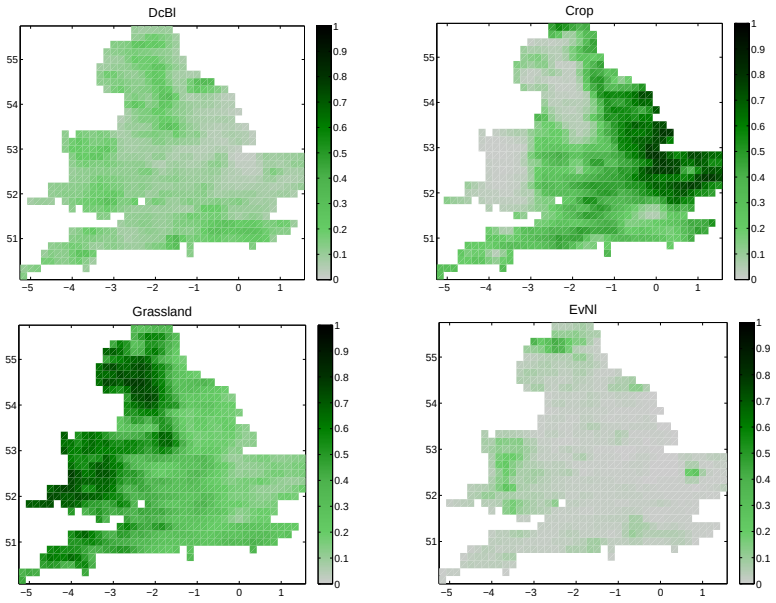

Figure 3: Maps of the simulated posterior means of PFTs DcBl, Crop, Grassland and EvNl in England and Wales.

The major contribution of our model is to measure the uncertainty in the true proportions of PFT for England and Wales. The results reported in this section are the simulated posterior standard deviations (henceforth referred to as the standard deviations) of the true PFT proportions for all sites. Also, as a rudimentary examination of the sensitivity of the parameter d we present the standard deviations of one PFT, Crop, for values of $d = 1, 2, 3, 4$.

Figure 4 displays the standard deviations, or uncertainty levels, of DcBl, Grassland, Crop and EvNl. The highest standard deviations for DcBl and EvNl are 0.1024 and 0.0986, both at longitude and latitude (55.25,-2.583), in the far north-east of England, suggesting a high level of uncertainty in classifying these two PFTs at this site. The highest standard deviation for Grassland is 0.1296 at longitude and latitude (53.25,-1.917), in the Peak District, and the highest standard deviation of Crop is 0.1418 at longitude and latitude (55.75,-2.083), in the far north of England. Figure 4 also suggests that higher standard deviations follow the general pattern of higher means for DcBl, Grassland and EvNl but not for Crop. For EvNl both the means and standard deviations are relatively high in the far north of England and Wales. DcBl typically has higher standard deviations in Wales and the Peak and Lake Districts. Similarly for Grassland, higher standard deviations and

means appear to coincide in the regions of Wales and the Peak and Lake Districts. The larger uncertainties in the proportions in these regions reflect the difficulty in assigning vegetation classes to either DcBl or Grassland. Crop, however, has comparatively similar standard deviations in East Anglia (where its means are highest), central England and into the south of England. The Midlands, and the region due south of the Midlands, is an area where LCM2000 allocates higher levels of Crop and comparable levels of DcBl, Grassland and EvNl, although Grassland is recorded by LCM2000 as dominating the bordering region to the west of the Midlands. Figure 4 suggests that the allocation of classes to PFTs Crop and Grasslands in this region have high uncertainty. Figure 4 also highlights uncertainty in the allocation of PFT Crop close to London.

Figure 5 shows the standard deviations of Crop when $d = 1, 2, 3, 4$. In analysing the sensitivity d only one PFT was chosen for brevity, and Crop is presented because the differences in standard deviations for values of d are most prevalent for Crop. As one would expect, as d increases the standard deviations decrease. However, the change across values of d is not dramatic and, as stated previously, we prefer to err on the side of caution and set $d = 1$. Note, the estimated means display no obvious differences for the above values of d for any of the PFTs.

Region wide results

We now report histograms of the estimated posterior distributions of the average of the true proportions over all sites:

$$\bar{\boldsymbol{\gamma}} = \frac{1}{707} \sum_{s=1}^{707} \boldsymbol{\gamma}^s.$$

Figure 6 reports histograms of the estimated posterior distributions of the region-wide proportions, $\bar{\boldsymbol{\gamma}}$, of the average proportions of DcBl, Crop, Grassland and EvNl over England and Wales. Figure 6 also shows the values of the average proportions over England and Wales as recorded in LCM2000, which are 0.1102, 0.3209, 0.4216, 0.0316 for DcBl, Crop, Grassland and EvNl respectively. For each PFT the posterior distribution shows a shift away from the LCM2000 recordings. For Grassland and Crop the shift is negative and for DcBl and EvNl it is positive. This suggests that LCM2000 allocates Grassland and Crop more often than exist in reality and allocates DcBl and EvNl less often than should exist in reality. The marginal totals in Table 1 contain the reason for this result. For Grassland and Crop, LCM2000 (row totals) allocates more pixels for these PFTs in England and

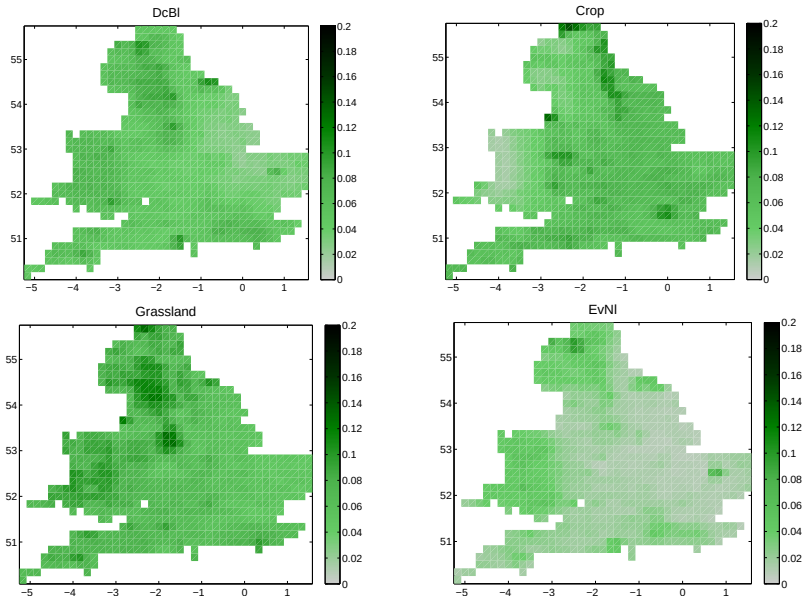

Figure 4: Maps of the simulated posterior standard deviations of PFTs DcBl, Crop, Grassland and EvNl in England and Wales.

Wales compared to the CS2000 (column totals). For DcBl, LCM2000 under allocates the number of pixels compared to CS2000. LCM2000 and CS2000 report the same values of EvNl. Another way to see this is to note the asymmetries in the confusion matrix. For instance, Grassland is quite rarely allocated erroneously by LCM2000 to DcBl, whereas there is a much higher probability for DcBl to be incorrectly allocated to Grassland.

7 Discussion

In this article we have provided a general framework to model uncertainty in satellite derived land cover maps, or other thematic maps, when data contained in a confusion matrix are available to calibrate the map. Our approach has the following features.

- We believe that it is unique in providing full posterior distributions for the true class at individual pixels and also full posterior distributions for proportions of different classes in an aggregated site. The latter is important when thematic maps are used as input to a large envi-

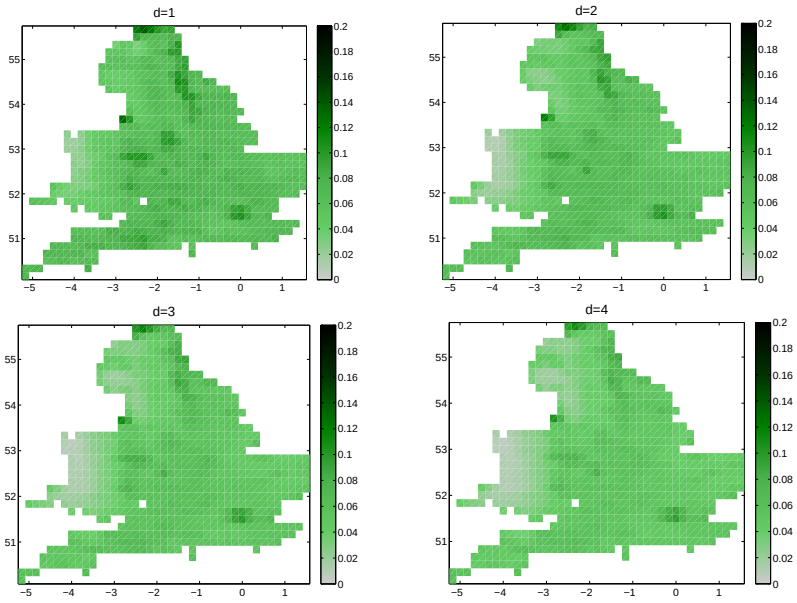

Figure 5: Maps of the simulated posterior standard deviations of Crop for values of $d = 1, 2, 3, 4$, clockwise from top left

ronmental model which is to be used as an aid to decision-making or setting environmental policies.

- We believe that it is also unique in accounting for spatial correlation in the error process (forward probabilities) as well as in the underlying true map.
- It is simple to apply, involving only Monte Carlo sampling from standard probability distributions.

The price for the last of these features is that we make a number of simplifying assumptions and approximations, each of which is set out and discussed below. Without these simplifications the analysis would become enormously more computationally demanding, which we believe would seriously limit its take-up in the remote-sensing and environmental modelling communities.

For illustrative purposes we have applied our method to a land cover map (the LCM2000) that is of current importance to the scientific community. The results identify specific sites where the LCM2000 has difficulty in classifying specific vegetation types. In particular, the LCM2000 exhibits relatively high uncertainty between the DcBl and Grassland PFTs, especially

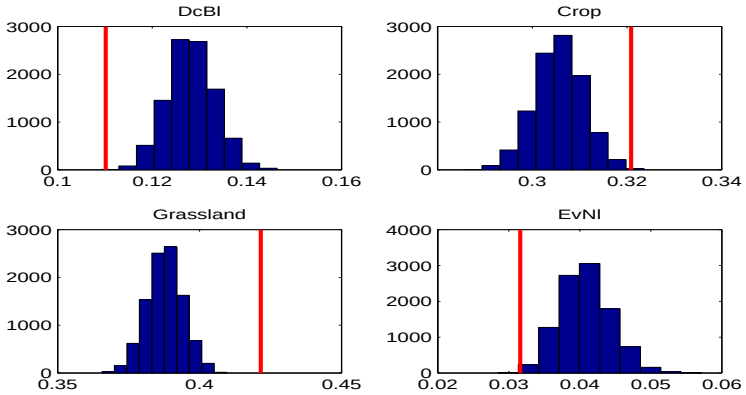

Figure 6: Histograms of the simulated posterior distributions of the overall PFTs DcBl, Crop, Grassland and EvNl for England and Wales. The vertical line shows the overall LCM2000 recording.

in Wales and the Peak and Lake Districts. The problematic nature of discriminating between land cover types that are spectrally similarly (in this case shrubs and grasslands) in satellite data has been noted for several of the major remote sensing land cover products (McCallum et al., 2006). Being able to provide well-founded estimates of site-specific uncertainty with such land cover classifications increases their value not only to an end user but also to the remote sensing community itself. The results also show that in general the LCM2000 overestimates the proportions of Grasslands and Crops and underestimates the proportions of DcBl, EvNl and Bare across England and Wales.

Modelling of the confusion matrix

Modelling of the confusion matrix is complicated by the design of the sample and by the fact that backward probabilities are context-specific. We have argued that to treat the rows as independent multinomial samples is simplistic and provides backward probability inferences that are not relevant to individual pixels or sites. Our approach of treating the columns as independent multinomial samples is admittedly also a simplification, but accords better with the underlying error process. It should be adequate when the column totals are not too divergent from the proportions of pixels assigned to each class in the whole map, as is the case in Table 1. However, disproportionate stratification can lead to this condition failing, for instance where

some classes that are uncommon in the map have been over-sampled to better estimate error rates. In this situation, we suggest that the rows of the matrix could be rescaled to match the map proportions. The fact that this causes the matrix entries to be non-integers is no problem for implementing the method, although it would clearly cast more doubt on the adequacy of the assumption of multinomial sampling.

Multinomial sampling also entails an assumption that the sample locations are sufficiently well separated over the region for spatial correlations to be negligible. In practice, this may not be the case.

We believe that the only way to resolve these difficulties is to abandon modelling of the confusion matrix directly, and to go back to the sample data. The confusion matrix loses information about the geographic locations of the sample values. If the original sample and map locations is available, then this information could be modelled directly. Now that the context of each sample item is explicit in its location, the nature of the sampling scheme becomes irrelevant in a Bayesian analysis and spatial effects can be modelled. However, this now means jointly modelling the sample and the map. Analysis would proceed as in image analysis approaches, but with an extra layer of information from the ground-truth survey.

Approximate spatial prior probabilities

In order to obtain manageable computations, we have only partially accounted for spatial correlation in the underlying map of true classes. This is fully accounted for in, for instance, the adaptive Bayesian classification (ABC) methods of Klein and Press (1992), but they only derive an estimated map without quantifying posterior uncertainty. To address this uncertainty, it would be necessary to apply a full MCMC computation, which is very computationally demanding for a large map. Adding the extra layer of sample data as discussed above would increase the computational load still further. Our approach is an approximation that is similar in spirit to a first pass of ABC or MCMC. As such, it probably under-smoothes the map but we believe it still captures the principal context-dependency of the backward probabilities. We suspect that the choice of spatial model (such as the choice of neighbourhood for a Markov random field) has more effect on the results than this approximation.

Implicit spatial error probabilities

We have argued that there is spatial correlation in the probabilities of misclassification, but this would be difficult to model explicitly. There is in

practice almost no useful information about such correlation. Such information could in principle come from the ground-truth surveys that are currently used to derive confusion matrices, but there has hitherto been no recognition of the issue. Therefore, such surveys do not currently incorporate the kind of sampling of neighbouring blocks of pixels that would provide such information.

We have instead adopted a very simple implicit model that introduces a parameter d. There is of course no direct information about the appropriate value to use for d, but a better understanding of the role of d in this model can be obtained as follows. Consider a site s in which there are n_t^s pixels that are truly in vegetation class t, and let $p_{t't}^s$ be the proportion of these that are observed to be in vegetation class t'. If the pixels are all classified independently with a common forward vector $\boldsymbol{\lambda}_t$, then the number $n_t^s p_{t't}^s$ which are classified as t' will be binomially distributed with mean $n_t^s \lambda_{t't}$ and variance $n_t^s \lambda_{t't}(1-\lambda_{t't})$. Hence $p_{t't}^s$ has mean $\lambda_{t't}$ and variance $\lambda_{t't}(1-\lambda_{t't})/n_t^s$, and for large n_t^s this underlies the statement in Section 4 that the proportion will almost certainly be very close to $\lambda_{t't}$. If they are not classified independently, then the variance will be larger.

Now consider the effect of our implicit model which allocates site s forward probabilities, $\boldsymbol{\lambda}_t^s$, the Dirichlet distribution $Di(d\boldsymbol{\lambda}_t)$. Conditional on the region-wide forward probabilities, $\lambda_{t't}^s$ has a Beta density with mean $\lambda_{t't}$ and variance $\lambda_{t't}(1-\lambda_{t't})/(d+1)$. Hence, the variance of $p_{t't}^s$ conditioned on $\lambda_{t't}$ is

$$\begin{aligned}
\mathrm{var}(p_{t't}^s \mid \lambda_{t't}) &= E\left\{\mathrm{var}(p_{t't}^s \mid \lambda_{t't}^s, \lambda_{t't}) \mid \lambda_{t't}\right\} + \mathrm{var}\left\{E(p_{t't} \mid \lambda_{t't}^s, \lambda_{t't}) \mid \lambda_{t't}\right\} \\
&= E\left\{\lambda_{t't}^s(1-\lambda_{t't}^s)/n_t^s \mid \lambda_{t't}\right\} + \mathrm{var}\left\{\lambda_{t't}^s \mid \lambda_{t't}\right\} \\
&= \frac{1}{n_t^s}\frac{d}{d+1}\lambda_{t't}(1-\lambda_{t't}) + \frac{1}{d+1}\lambda_{t't}(1-\lambda_{t't}) \\
&= \frac{1}{n_t^s}\frac{d+n_t^s}{d+1}\lambda_{t't}(1-\lambda_{t't}) \ .
\end{aligned}$$

For large n_t^s, this is effectively $\lambda_{t't}(1-\lambda_{t't})/(d+1)$, and so the variance of $p_{t't}^s$ under this model approximates to the variance that one would have if just $d+1$ pixels were independently mis-classified. So this model can be thought of as inducing dependence between the pixels in a site so that they divide into $d+1$ independently classified clumps, but that all the pixels in a clump are classified the same. This provides an interpretation of d in terms of the idea of locally dependent classification. The smaller d is, the more variability we introduce between the $\boldsymbol{\lambda}_t^s$ vectors of different pixels, and the

greater local dependence is implied.

We make one more approximation in Section 4 that is perhaps not obvious. By assuming multinomial distributions in (8), we are giving all pixels in the site the same backward probabilities. This is in contrast to our insistence in Sections 2 and 3 that they have different backward probabilities because of having different, context-dependent, prior probabilities. The multinomial distributions therefore incorporate another simplification. However, the neighbourhood used for deriving prior probabilities will typically be large enough for the prior probabilities to differ only imperceptibly within a site, so this should be another good approximation.

References

Besag, J. E. (1986), "On the Statistical Analysis of Dirty Pictures," *Journal of the Royal Statistical Society, Ser. B*, 48, 259–302.

Divino, F., Frigessi, A., and Green, P. J. (2000), "Penalized Pseudolikelihood Inference in Spatial Interaction Models with Covariates," *Scandanavian Journal of Statistics*, 27, 445–458.

Foody, G. M. (2002), "Status of land cover classification accuracy assessment," *Remote Sensing of Environment*, 80, 185–201.

— (2007), "The evaluation and comparison of thematic maps derived from remote sensing," *7th Internations Symposium on Spatial Accuracy Assessment and Environmental Sciences*.

Frigessi, A. and Stander, J. (1994), "Informative Priors for the Bayesian Classification of Satellite Images," *Journal of the American Statistical Association*, 89.

Fuller, R. M., Smith, G. M., Sanderson, J. M., H. R. A., and Thomson, A. G. (2002), "Countryside Survery 2000, Module 7. Land cover map 2000. Final Report," *Technical report*.

Geman, S. and Geman, D. (1984), "Stochastic Relation, the Gibbs Distribution and the Bayesian Restoration of Images," *IEEE Transactions*, 721–741.

Giri, C., Zhu, Z., and Reed, B. (2005), "A comparative analysis of the Global Land Cover 2000 and MODIS land cover data sets," *Remote Sensing of Environment*, 94, 123–132.

Green, E. J. and Strawderman, W. E. (1994), "Determining the accuracy of thematic maps," *The Statistician*, 43.

Hansen, M. and Reed, B. (2000), "A comparison of the IGBP DISCover and University of Maryland 1 km global land cover products," *International Journal of Remote Sensing*, 21, 1365–1373.

Jung, M., Henkel, K., Herold, M., and Churkina, G. (2006), "Exploiting synergies ofglobal land cover products for carbon cycle modelling," *Remote Sensing of Environment*, 101.

Klein, R. and Press, S. J. (1989), "Contextual Bayesian Classification of Remotely Sensed Data," *Communications in Statistics, Part A - Theory and Methods*, 18, 3177–3202.

— (1990a), "Bayesian Classification of Remotely Sensed Data when Training Data is Part of the Scene," *Revista Brasileira de Probabilidae e Estatistica*, 4, 43–67.

— (1990b), "Bayesian Contextual Classification with Neighbors Correlated with Trainng Data," *Bayesian and Likelihood Methods in Statistics and Econometrics: Essays in Honor of George A. Bernard*, 337–355.

— (1992), "Adaptive Bayesian Classification of Spatial Data," *Journal of the American Statistical Association*, 87, 844–851.

Mayaux, P., Eva, H., Gallego, J., Strahler, A. H., Herold, M., Agrawal, S., Naumov, S., De Miranda, E. E., Di Bella, C. M., Ordoyne, C., Kopin, Y., and Roy, P. (2006), "Validation of the Global Land Cover 2000 Map," *IEEE Transactions on Geoscience and Remote Sensing*, 44.

McCallum, I., Obersteiner, M., Nilsson, S., and Shivdenko, A. (2006), "A spatial comparison of four satellite derived 1 km global land cover datasets," *Applied Earth Observation and Geoinformation*, 8, 246–255.

Storvik, G., Fjortoft, R., and Solberg, A. A. (2005), "A Bayesian Approach to Classification of Multiresolution Remote Sensing Data," *IEEE Transactions on Geoscience and Remote Sensing*, 43.

Switzer, P. (1980), "Extensions of Linear Discriminant Analysis for Statistical Classificiation of Remotely Sensed Satallite Imagery," *Mathematical Geology*, 12, 367–367.

Acknowledgements This project was funded by the NERC Centre for Terrestrial Carbon Dynamics.