Let tree i have spatial coordinates and

Let mobile plot p have center given by and

Let be the plot assignment status of tree i.

The probability of tree i to be assigned to plot p is given by:

Let be the community assignment status of tree i. Given that , then is drawn from:

Finally, the species label of tree i is given by:

The user will have to create multiple potential plots. Notice that the center of each plot (i.e., and ) is known (i.e., fixed). Similarly, the user has to pre-specify or we can estimate this parameter. If we decide to estimate , we can use the following prior:

Where we set M to be a large number. This parameter controls the amount of spatial heterogeneity. Big values imply almost no spatial heterogeneity whereas smaller values imply stronger spatial patterns. Finally, we assume that:

where is a small number to encourage a document to be dominated by a single cluster. Similarly, we choose to be a small number to encourage a cluster to be dominated by few species.

#--------------------------------------------

How would we visualize the spatial distribution of ?

#------------------------------------------

Modeling sequence if we have to estimate the number of clusters

1. Run LDA with TSB prior to determine optimal number of clusters and their characteristics
2. Run Spatial LDA to create spatially smooth maps of these clusters

#------------------------------------------

Algorithm

In this context (i.e., assuming we know ), the parameters to sample are the latent cluster membership , the latent document membership , , and

Important trade-off: if we have too many documents, we might have relatively little information for any given and therefore this parameter would not be well estimated. However, if I have too few documents, then we might not allow for sufficient change in space. The extreme scenario would be if we had a single document, in which case the proportion of each cluster would simply be given by and there would be no spatial variation.

* FCD for

We can sample this from a categorical distribution. If it is too expensive to calculate for every single tree, we can use small plots. This way, we can sample for all trees within each small plot simultaneously since willbe approximately the same for all these trees.

* FCD for
* FCD for
* FCD for

Where . This implies that:

* FCD for

Where . This implies that:

To calculate the likelihood, notice that:

The dimensions are given by

Notice that I can take the transpose of this quantity because that still yields a 1x1 matrix: