Hierarchical space-time models for high dimensioned fire data [1]

M.A. Amaral-Turkman, K.F. Turkman, Le Page Yannick and J. Pereira

DEIO and CEAUL-University of Lisbon, ISA-Technical University

Abstract: In this paper, we suggest several alternative Bayesian Hierarchical models for a complex, large scale spatio-temporal data set on the percentage of land burned by wildfires observed yearly with a resolution of 1 degrees across the world. The objective is the see the relative importance of several meteorological and topological explanatory variables in explaining the ignition probability as well as the percentage of land burned at observed pixels over the years. The statistical inference and model fitting results will be given in a separate report.

1 Introduction

Consider a spatially and temporarily varying response variable Y and a set of Independent covariates $\mathbf{X} = (X_1, ..., X_p)$. Suppose that we have a data set of the form

$$\mathbf{Y} = (Y(s_j, t), j = 1, ..., n, t = 1, ..., N,$$

and

$$\mathbf{X} = (X_1(s_j, t), ..., X_p(s_j, t) j = 1, ..., n, t = 1, ..., N).$$

Here, s_j are the spatial locations of the observations. The type of data set we have in mind can be exemplified by the following problem: Suppose that the continent of Australia is divided into pixels of 1degree, resulting in a lattice with 744 cells. Let $s_j = (latitude, longitute)$ of the jth pixel centroid in this lattice. Let $Y(s_j, t)$ denote the percentage of land burned in pixel s_j, in year $t = 1997, .., 2006$. Assume that the following covariates are observed in each pixel:

1. X_1: id: a line identifier (1-750 for 1997 ; 1-6750 for 1998-2006, from the repeated 750 grid cells for each of the 9 years)
2. X_2: lat: latitude of the middle of each 1 grid cell
3. X_3: lon: longitude of the middle of each 1 grid cell
4. X_4: TRMMDrySeas: Dry Season Severity from TRMM satellite data (native resolution 0.25, aggregated to 1)
5. X_5:TRMMprecip: Precipitation over the wet season (or growing season) from TRMM satellite data

[1]This work is partially supported by FCT, POCTI and Research project PTDC/MAT/64353/2006

6. X_6:HF: Human footprint (static)
7. X_8: GLCt: Tree landcover (from 0 to 1, 1 being 100
8. X_9: GLCgs: Grass and Shrubs landcover (static)
9. X_{10}: GLCa: Agricultural Landcover (static)
10. X_{11}: CMAPPentads: Maximum number of consecutive dry pentads (pentad=5 days) over the corresponding year.
11. Y:GFEDbf: Estimated Burned Fraction over the year (dependent variable)

Hence, there are $n = 744$ locations, each having $N = 10$ repetitions from 1997 to 2006. Note that some these covariates temporarily static. There are other independent variables which are known to be very influential on fire regimes and sizes, such as heterogeneity in the vegetation, wind speed and direction etc., which are not included in the analysis due to unavailability of data. Some of the covariates in the list exhibit strong time and space variation, but at this stage, there will be no attempt to model the temporal and spatial structures that exist strongly in these independent variables; they will simply be included in our model as fixed explanatory variables.

The objective is to explain the relative importance of these explanatory variables on the fire regime, as well as explaining the space-time variation in the response variable. One possible way of handling such data and take adequate conclusions is the use of geographically weighted regression analysis (GWR) (Fotheringham et al, 2002) The main objective of GWR is to carry out the usual multiple regression analysis using the well known weighted least square method, where the weights are chosen taking into consideration the distances between the spatial coordinates over which the observations are taken. There are however, certain aspects of data which needs to be considered and which may not be compatible with the standard weighted least square method:

- The data set contains unusually high number of 0's, that is, the data is zero-inflated (no fire ignition).
- Non-zero observations are highly skewed to the left with a fairly heavy right tail (See Figure1). Such data, even with the best chosen transformation (for example a proper Box-Cox transformation) can not be transformed into a "normal" data.
- Spatial and temporal dependencies are too strong to assume uncorrelated error structure.

As such, one needs to use a generalized linear model to represent the variation in percentage burned in terms of the above covariates and possibly either using correlated error structures or including in the regression, random effects which are time and space dependent. Geographically weighted regression analysis

(GWR) (Fotheringham et al, 2002), better yet, a improved version of it can still be used for the data conditional on the positive percentage burned: Let

$$Z(t,i) = \log(\frac{Y(t,i)}{1-Y(t,i)}),$$

where $Y(t,i)$ are the (non-zero) percentage of land burned. There is sufficient empirical evidence to assume that $Z(t,i)$ are well represented by a Gaussian model. Based on this, we suggest the following hierarchical geographically weighted regression model for $Z(t,i)$ which takes into consideration the spatial and temporal dependence in the residual process: Assume that

$$Z = \boldsymbol{\beta}^T \mathbf{X} + \boldsymbol{\epsilon},$$

where $\boldsymbol{\beta}$ are spatially and possibly time varying coefficients and $\boldsymbol{\epsilon}$ are 0 mean spatially and temporarily dependent residuals, possibly with spatially varying variances. We suggest the following model:

1. $$Z(s_j,t) = \sum_{k=1}^{p} \left[\sum_{i=1}^{n} \alpha_k(s_j,s_i) X_k(s_i,t) \right] + \epsilon^*(s_j,t),$$

2. $\epsilon^*(s,t)$ is a 0 mean, discrete time-discrete space stationary Gaussian process which we write in the following form:

   $$\epsilon^*(s,t) = W(s,t) + \epsilon(s,t),$$

   where $W(s,t)$ is a 0 mean time-space dependent gaussian process and $\epsilon(s,t)$ is an iid $N(0,\sigma^2)$ random variables, which can be seen as the nugget effect. (see Cressie,1993).

3. $\alpha_k(s_j,s_i)$ are space dependent kernels given by the Gaussian kernel

   $$\alpha_k(s_j,s_i) = \alpha_k \exp(-\frac{1}{b_k^2}||s_i - s_j||^2).$$

   Here, b_k are the bandwidth, which is a measure of distance-decay relationship of Z measured at pixel s_j and X_k observed at pixel s_i and when $s_i = s_j$, α_k represents the regression coefficient between the response variable and the kth covariate at location s_j. .

4. We assume a markov structure for $W(s;t)$ in time:

   $$W(s_j,t) = \eta W(s_j,t-1) + W_1(s_j), t = 2,3,...,N$$

   where $W_1(s_j)$ is a nearest neighbor CAR gaussian process and $W(s,1)$ is also a nearest neighbor CAR model with known parameters, which is independent of $W_1(s)$.

We now comment on the model we suggest:

1. The model we suggest is an improvement on the Geographically weighted model suggested by Fotheringham et al in the sense that we assume space-time dependent residuals (hence data) whereas Fotheringham et al assume independent residuals. Hence, their set up results in the usual weighted regression analysis where instead of the X^tX matrix, one builds the regression on $X^TC^{-1}X$ matrix, where the C matrix explains the spatial distance decay relationship.

2. $W(s_j,t)$ can be seen as a latent random effect, taking care of the unobserved dependent covariates which have spatio-temporal variations. Such latent random effect can be given a more complicated space-time structure, but at present we will not pursue this point.

2 The hierarchical model

Based on the above model, we suggest the following hierarchy:

1. Likelihood:

$$\begin{aligned} & p(Z(s_j,t), j=1,...,n, t=1,...,N|(\mathbf{X}(s_i,t), i=1,...,n, t=1,...,), \mathbf{\Theta}, \boldsymbol{\epsilon}^*) \\ = & \prod_{j=1}^{n}\prod_{t=1}^{n} p(Z(s_j,t)|(\mathbf{X}(s_i,t), I=1,...,n, t=1,...,), \mathbf{\Theta}, \boldsymbol{\epsilon}^*) \end{aligned}$$

   where

$$p(Z(s_j,t)|(\mathbf{X}(s_i,t), i=1,...,n, t=1,...,), \mathbf{\Theta}, \boldsymbol{\epsilon}^*) \sim N(\mu(s_j,t), \sigma^2),$$

2. Link Functions:

$$\mu(s_j,t) = \sum_{k=1}^{p}\sum_{i=1}^{n} \alpha_k(s_j,s_i)x_k(s_i,t) + w(s_j,t),$$

$$\alpha_k(s_j,s_i) = \alpha_k \exp(-\frac{1}{b_k^2}||s_i - s_j||^2),$$

$$w(s_j,t) = \eta w(s_j,t-1) + w_1(s_j), t=2,...,N$$

   $w_1(s)$ is an nearest neighbors CAR model (Sangaree et al) and $w(s,1)$ is a nearest neighbor CAR model independent of $w_1(s)$

3. Hyper parameters. The choice of prior distribution for the hyper parameters need to be carefully discussed.

If we model the whole data set, including the pixels where there are no ignition, then there are certain aspects of data which may not be compatible with the standard regression method suggested above, as was previously explained. In this case, one needs to use a generalized linear model to represent the spatio-temporal variation in the probability of ignition, as well as the percentage area burned conditional on the event of ignition in any pixel.

2.1 Model based geostatistics for ignition probability and percentage of area burned

Our objective is to come with a generalized linear model to

- Explore the relative importance of the above covariates in explaining the probability of having an fire ignition in space and time, as well as explaining the size of the area burned given there was an ignition
- Use such models as spatial-temporal smoothers to plot maps of fire risks,
- Explain and quantify other spatial-temporal variations in the data caused by factor other than the covariates given above through a properly chosen spatially and temporally colored (latent) random effect.

We will suggest various Bayesian hierarchical models, assuming different structural assumptions and perform simulation based inference, using the BUGS software. Due to the size of the data set, one would expect computational difficulties, particularly in using the BUGS software. Therefore, Methods and models to deal with large data sets suggested by Gelfand et (2006) are needed to suit these models to the BUGS software.

The data will be analyzed in different stages, using more complicated or alternative models at each stage:

- At the first stage, we will assume that the geographical area under study is partitioned into a finite number or areal units, which in our case are the fixed pixels of size 1^o degree. These pixels are areal units with well defined boundaries. Hence, we will consider the pixel positions s_i as the centroid of these pixels and consider the data as a lattice data. Such models are easier to deal, since as we will see, the hidden spatio-temporal random effect is now a discrete space-time process which can be assumed to be a markov process (for example, conditionally autoregressive process, (CAR) or Intrinsically autoregressive process(IAR)) which can be simulated without the curse of dimension.
- In second stage, we take the spatial location s to vary continuously over the geographical area under study. We this assume that the observations are sampled from a process continuous in space, but discrete in time. Such data sets are often called Geostatistical or point referenced data sets and models for explaining the spatial(and temporal) variations in the data are called Point-level models (Banarjee et all, 2004). Since the data appears naturally as lattice data, it may not be easy to justify the reason of using point level models for this data set, particularly considering that areal models are designed to answer the questions set out as objectives. However, treating the data as a sample from a process continuous in space has its advantages, particularly when extrapolation to other spatial locations (and resolutions) where there are no observations. Temporal dependence, as well as spatial dependence will be introduced to the data at stages, depending the complications that it brings to the data:

1. Modeling the latent random effect as one massive space-time gaussian process with a valid covariance structure. We will use one of the space-time covariance structures suggested by Gneiding (2004). These covariance structures have the benefit of explaining the degree of spatial, temporal covariances as well as the degree of space-time interactions conveniently through few parameters. However, fitting one massive space-time random effect is computationally infeasible, and we will suggest ways of resolving such difficulties by
2. Assuming that the data are an independent identically distributed replicates over time, and later by
3. Using latent space-time models suggested by Rappold et al(2007).

We now look at these alternative models in detail:

2.2 Areal model 1:

The basic assumption is to partition the space into a finite number of areal units with well defined boundaries. To this effect, we assume that the 745 spatial locations are the centroid of the pixels of equal sizes of 1 degree. Hence, our data is a lattice data of 745 locations with 10 annual observations taking the form $(y(i,t), i = 1, 2, ..., 745, t = 1, 2, .., 10)$ where $y(i,t)$ represents the percentage of area burned in areal unit (pixel) i. Let

$$\mathbf{X}(t,i) = (X_1(t,i), ..., X_5(t,i))$$

(antonia, you know what these variables are) be subset of the covariates explained in the previous section,which are found to be important in explaining the fire regimes through a preliminary analysis. We define Bernoulli variables

$$R(t,i) = \begin{cases} 1, & \text{With probability } p_1(t,i) = 1 - p_0(t,i); \\ 0, & \text{With probability } p_0(t,i). \end{cases} \quad (1)$$

Here, $p_1(t,i)$ represents the probability that a fire is ignited at location s_i, during year t, and hence $R(t,i)$ represents the existence of a fire in year t at pixel i. Let $\mathbf{V}(t,i) = (V_0(t,i), V_1(t,i))$ be a latent spatio-temporal process. Here, V_0 derives the ignition process, whereas V_1 derives the fire sizes. Let $Z(t,i) = \log(\frac{Y(t,i)}{1-Y(t,i)})$ and denote by $\mathbf{Z}$,$\mathbf{X}$,$\mathbf{R}$ and $\mathbf{V}$

$$(Z(t,i), i = 1, 2, ..., 745, t = 1, 2, .., 10),$$

$$(\mathbf{X}(t,i), i = 1, 2, ..., 745, t = 1, 2, .., 10),$$

$$(R(t,i), i = 1, 2, ..., 745, t = 1, 2, .., 10),$$

and

$$(\mathbf{V}(t,i), i = 1, 2, ..., 745, t = 1, 2, .., 10),$$

respectively. Let $\mathbf{\Theta}$ be all the (random) model parameters to be defined later.

- **First level; Likelihood**

$$f(\mathbf{z}|\mathbf{R},\mathbf{V},\mathbf{X},\mathbf{\Theta}) = \prod_{t=1}^{10}\prod_{i=1}^{745} f(z(i,t)|\mathbf{R}(i,t),\mathbf{V}(t,i),\mathbf{\Theta}) \tag{2}$$

$$f(z(t,i)|\mathbf{R},\mathbf{V},\mathbf{X\Theta}) = \left[N(\mu(t,i),\sigma^2)\right]^{R(t,i)}(1-p_0(t,i))^{R(t,i)}p_0(t,i)^{1-R(t,i)}. \tag{3}$$

$$f(\mathbf{R}|\mathbf{V},\mathbf{X},\mathbf{\Theta}) = \prod_{t=1}^{10}\prod_{i=1}^{745} f(R(i,t)|\mathbf{V},\mathbf{X},\mathbf{\Theta}), \tag{4}$$

where the conditional distribution of $R_t(s)$ is bernoulli with $P(R_t(i) = 1) = 1 - p_0(i,t)$.

- **Link functions and the latent processes** Here, our state equation is the mean function $\mu(t,i)$, through which the spatial dependence will be introduced. 3 is the observation equation and we assume that σ^2 is observation error, which are independent and identically distributed over space and time.

$$\mu(t,i) = \beta_0 + \mathbf{X}(i,t)^T\beta + V_1(i) + \delta_1(t) \tag{5}$$

$$\log\frac{p_0(i,t)}{1-p_0(i,t)} = \eta_0 + \mathbf{X}(i,t)^T\boldsymbol{\eta} + V_2(i) + \delta_2(t), \tag{6}$$

where,

  - $V_1(i) = \nu_1 W_0(i) + \nu_2 W_1(i)$, $V_2(i) = W_1(i)$, and $W_0(i)$ and $W_1(i)$ are independent CAR (conditionally autoregressive)models (Gelfand et al, 2004) given by

$$W_0(i)|w_0(j), j \neq i \sim Normal(\sum_j b_{ij}w_0(i), \tau_0), \tag{7}$$

$$W_1(i)|w_1(j), j \neq i \sim Normal(\sum_j c_{ij}w_1(i), \tau_1). \tag{8}$$

    Note that b_{ij} and c_{ij} need to satisfy certain conditions so that define

$$p(w_0(1),...,w_0(n)) \sim N(0,\Sigma),$$

    are well defined with a proper covariance matrix. We refer to Gelfand et all(2004) for the proper definition of the CAR model. CAR models coefficients b_{ij} and c_{ij} are automatically generated by the BUGS and hence they need not be defined separately in the hierarchy as hyper parameters. Note that the dependence structure between $V_1(i)$ and $V_2(i)$ is very simple. Although more complicated dependence structures can be given, for example involving linear combinations at

each pixel level, this would increase the number of model parameters drastically, thus complicating convergence. We also believe that the data can not support such over parameterized structures.

Alternatively, we can model the latent processes as Intrinsically autoregressive (IAR) processes. We refer to Gelfand et al (2004) for the treatment of CAR and IAR processes as models for the latent processes.

– $\delta_1(t)$, $\delta_2(t)$ are latent temporal processes, capturing the overall temporal dependence structure. We assume simple autoregressive structures for each:

$$\delta_1(t) = \phi_1\delta_1(t-1) + \epsilon_1(t),$$
$$\delta_2(t) = \phi_2\delta_2(t-1) + \epsilon_2(t),$$

where $\epsilon_1(t)$, $\epsilon_2(t)$ are respectively iid $N(0, \tau_2)$ and $n(0, \tau_3)$, independent of each other.

- **Parameters, hyperparameters and the prior specifications** The model parameters are

$$\tau_0, \tau_1, \phi_1, \phi_2, \beta = (\beta_0, \beta_1, ..., \beta_5), \boldsymbol{\eta} = (\eta_0, \eta_1, ...\eta_5), \nu_1, \nu_2, \phi_1, \phi_2,$$

and we assume vague independent priors for these hyperparameters:

1. $\tau_0 \sim gamma(0.1, 0.1)$, $\tau_1 \sim gamma(0.1, 0.1)$, $\tau_2 \sim gamma(0.1, 0.1)$, $\tau_3 \sim gamma(0.1, 0.1)$
2. β independent $N(0, 0.001)$, $\boldsymbol{\eta}$ independent $N(0, 0.001)$ (Note that the value 0.001 corresponds to the precision parameter.)
3. $\phi_1 \sim U(-1, 1)$,$\phi_1 \sim U(-1, 1)$
4. $\nu_1 \sim N(1, 0.1)$, $\nu_2 \sim N(1, 0.1)$

2.3 Areal model 2: Geographically weighted regression

Further extensions can also be made in the model, by regressing the link functions at pixel i not only on the covariates observed at the same pixel, but also observed on other pixels. Thus we suggest upgrading the link functions by

1.

$$\mu(t,i) = \mu + \sum_{k=1}^{p}\sum_{j=1}^{745}\alpha_1(k)\exp(-\frac{1}{b_1^2}||i-j||^2)X(j,t) + V_1(i) + \delta_1(t) \quad (9)$$

2.

$$\log\frac{p_0(i,t)}{1-p_0(i,t)} = \nu + \sum_{k=1}^{p}\sum_{j=1}^{745}\alpha_2(k)\exp(-\frac{1}{b_2^2}||i-j||^2)X(j,t) + V_2(i,t) + \delta_2(t), \quad (10)$$

where $\alpha_1(k)\exp(-\frac{1}{b_1^2}||i-j||^2)$ is the Gaussian Kernel, with $||i-j||^2$ being the euclidian distance between the pixels i and j. Again, the suggested regression is on the covariates observed at each of the 745 pixels and this may put computational burdens and the data may not support such detailed regression. Instead, we suggest restricting the regression on the nearest neighbors by using the regressions

$$\sum_{k=1}^{p}\sum_{j\in N(i)}\alpha_1(k)\exp(-\frac{1}{b_1^2}||i-j||^2)X(j,t)$$

and

$$\sum_{k=1}^{p}\sum_{j\in N(i)}\alpha_2(k)\exp(-\frac{1}{b_2^2}||i-j||^2)X(j,t),$$

respectively in the link functions, where $N(i)$ are the nearest neighbors of the pixel i.

The above areal models are the natural setting for our equidistance lattice data. This model would allow us to see how covariates effect the percentage of area burned and is also ideal for use as a spatial and temporal smoother. However, if the objective of the study is extrapolation, that is to do krigging to predict the value of the process at locations where there are no observations, than it may be more beneficial to consider the area under study as continuous. Rather than considering the data as being percentage of a pixel (areal unit) burned, we can consider it as the amount of land burned as a consequence of a fire that ignited at location s. There is still a problem with this interpretation of the data; the pixel centroid are fixed as design variables. However, assuming that the fires ignite with equal probability within each pixel, and assuming that the total area burned during the year is due to a single fire, we can simulate randomly a location with the pixel, given a fire is observed in that pixel during the year, hence creating a space-time data which is a realization of a space-time process, discrete in time, but continuous in space. Then this data can be treated as geo-statistical data.

3 Ideal Space-time point-referenced model

We assume that the data is a realization of a space-time process, discrete in time and continuous in space, representing the total are burned in year t by a fire ignited at location s. (since each pixel has a fixed areal size $30km^2$, this conversion is straight forward. let s_i are the locations where observations are taken.

Let

$$\mathbf{X}(t,s_i)=(X_1(t,s_i),...,X_5(t,s_i))$$

be the covariates explained in the previous section,observed at the same locations and time. We define Bernoulli variables

$$R(t,s)=\begin{cases}1, & \text{With probability } 1-p_0(t,s);\\ 0, & \text{With probability } p_0(t,s).\end{cases} \tag{11}$$

Here, $R(t,s)$ represents the existence of a fire in year t at location s. Let $\mathbf{V}(t,s) = (V_0(t,s), V_1(t,s))$ be the latent spatio-temporal process. In a manner to be explained, $V_0(t,s)$ will derive the existence of fires in time and space, whereas , $V_1(t,s)$ will derive the size of the fires given there was one at time t and pixel s. Finally Let $\mathbf{\Theta}$ be the model parameters, which will be specified fully upon writing the full model. Define, in a similar fashion as in the previous section, the vectors $\mathbf{z}$,$\mathbf{X}$,$\mathbf{R}$, $\mathbf{V}$ at spatial locations s_i and year t

First stage: Likelihood Hence, the full likelihood is given in the following manner:

$$\begin{aligned} f(\mathbf{z}, \mathbf{R}, \mathbf{V}, \mathbf{X}\mathbf{\Theta}) &= f(\mathbf{z})|\mathbf{R}\mathbf{V}, \mathbf{X}, \mathbf{\Theta}) \\ &\times f(\mathbf{R}|\mathbf{V}, \mathbf{X}, \mathbf{\Theta}) \\ &\times f(\mathbf{V}|\mathbf{\Theta}) f(\mathbf{\Theta}) \end{aligned} \tag{12}$$

We now model the components of this full likelihood, assuming certain dependence and distributional assumptions:

1. **Assumption 1: Conditional Independence** Conditional on $((\mathbf{R}, \mathbf{V}, \mathbf{X}, \mathbf{\Theta})$, $\mathbf{z}_t(s)$ are independent in time and space, with normal distribution:

$$f(\mathbf{z}|\mathbf{R}, \mathbf{V}, \mathbf{X}, \mathbf{\Theta}) = \prod_{t=1}^{10} \prod_{s=1}^{745} f(z(t,s)|\mathbf{R}, \mathbf{V}, \mathbf{X}\mathbf{\Theta}), \tag{13}$$

   where

$$f(z(t,s)|\mathbf{R}, \mathbf{V}, \mathbf{X}\mathbf{\Theta}) = \left[N(\mu(t,s), \sigma^2)\right]^{R(s,t)} (1 - p_0(t,s))^{R(s,t)} p_0(s,t)^{1-R(s,t)}. \tag{14}$$

   Note that, when $R(s,t) = 0$, that is there is no ignition at space time (s,t), then there is no contribution to the likelihood at that space-time location. This likelihood implicitly puts positive mass to $P(Y(s,t) = 0)$, this probability being equal to $P(R(s,t) = 0) = 1 - p_0(s,t)$. That is, there is positive burned land, given that there is ignition. Otherwise, With probability 1, there is no burned land.

2. **Condition 2: Conditional independence of the ignition process**

$$f(\mathbf{R}|\mathbf{V}, \mathbf{X}, \mathbf{\Theta}) = \prod_{t=1}^{10} \prod_{s=1}^{745} f(r(s,t)|\mathbf{V}, \mathbf{X}, \mathbf{\Theta}), \tag{15}$$

   where the conditional distribution of $R(s,t)$ is bernoulli with $P(R(s,t) = 1) = 1 - p(s,t)$, so that

$$f(r(s,t)|\mathbf{V}, \mathbf{X}, \mathbf{\Theta})) = (1 - p(s,t))^j p(s,t)^{1-j},$$

   $j = 0, 1$.

3. **Condition 3: Dependent latent processes**

$$f(\mathbf{V}|\mathbf{\Theta}) = f(\mathbf{V}_1|\mathbf{V}_0, \mathbf{\Theta}) f(\mathbf{V}_0|\mathbf{\Theta}), \tag{16}$$

where, $\mathbf{V}_0 = (v_0(s,t), t = 1, ..., 10, s = 1, ..., 745)$ is a stationary and isotropic Gaussian space-time process $\mathbf{V}_0 \sim GP(0, \Sigma_0)$, where the elements of Σ_0 are given by the space time-space covariance structure of Gneiding(2002):

$$Cov(V_0(s_1,t_1), V_0(s_2,t_2)) = Cov(V_0(0,0), V_0(u,h)) = C_0(h,u), \tag{17}$$

where

$$C_0(h,u) = \frac{\sigma_0^2}{(a_0|u|^{2\alpha_0}+1)^{\tau_0}} \exp\left(-\frac{c_0||h||^{2\gamma_0}}{(a_0|u|^{2\alpha_0}+1)^{\beta_0\gamma_0}}\right), \tag{18}$$

and a_0, c_0 are nonnegative scaling parameters, α_0, γ_0 are smoothness parameters defined in $(0,1]$, $\beta \in [0,1]$ is the parameter explaining the degree of space-time interaction, $\tau > 0$ and $\sigma^2 > 0$ is the space-time nugget effect.

$$f(\mathbf{V}_1|\mathbf{V}_0, \mathbf{\Theta}) \sim GP(\boldsymbol{\zeta}^T \mathbf{V}_0, \Sigma_1), \tag{19}$$

where $\boldsymbol{\zeta}$ are regression coefficients and Σ_1 is the valid covariance matrix with the same structure as for Σ_0, but with a different set of parameters $\sigma_1^2, a_1, c_1, \gamma_1, \beta_1, \alpha_1)$.

The regression on $\mathbf{V}_1$ on $\mathbf{V}_0$ given in (19) can be defined through difference equations with assumptions drastically reducing the number of parameters. For example, Following Rappold et al (2007) we can assume that for each (s,t)

$$V_1(s,t) = \phi_1 V_0(s,t) + W_1(s,t), \tag{20}$$

where,

$$W_1(s,t) = N(0, \Sigma_1),$$

are spatial Gaussian processes, independent and identical in time.Such assumptions naturally restrict the space-time interactions of the component latent processes $(\mathbf{V}_1, \mathbf{V}_0)$.

Other possibility is to model directly the joint density $f(\mathbf{V}_0, \mathbf{V}_1|\mathbf{\Theta}) \sim GP(0, \Sigma)$,where

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \tag{21}$$

The space time covariance structures Σ_{11}, Σ_{22} of $\mathbf{V}_0$ and $\mathbf{V}_1$ can be modeled as in (16), However, it is not clear how one should define the cross-covariance matrices Σ_{12} and Σ_{21}. For purely multivariate spatial processes, construction of proper covariance structures are treated by Gelfand et all(2007). Further work

is need to extend these constructions to space-time processes. Hence, at present, conditional specification of the latent processes is simpler. Once the prior distributions of the model parameters are elucidated, the above model becomes a fully specified hierarchical model. However, such model is too complicated and The usual simulation based inference is virtually impossible due to the size of the latent process $\mathbf{V}$. Σ_0 as well as Σ_1 defined (16)and (19)are matrices of sizes 7450×7450, while simulating from the conditional distributions of the (random) model parameters during the simulation based inference, Inverses of these matrices are required and from practical point of view, this is impossible. Hence, certain simplification in the model assumptions are required to allow efficient inference on the model. Such simplifications can be made along several lines, resulting in different conditional specifications and models and . We will give these models separately.

3.1 Predictive models for Large data sets

Our first set of simplification is to assume that the data is independent and identically distributed over time, although simple temporal dependence may be introduced without effecting the method. However, we leave this further extension to later stages upon explaining how large spatial data sets can be handled. Thus our data is a independent identical realizations of a spatial process $Y(s)$, and the data will be treat as $Y_i(s_j)$, $j = 1, ..., 745$ and $i = 1, ..., 10$ are the independent identical replicates of $Y_1(s_j), j = 1, ..., 745$ over time. Then the model given in the previous section caries with little change:

First stage: Likelihood Hence, the full likelihood is given in the following manner:

$$\begin{aligned} f(\mathbf{z}, \mathbf{R}, \mathbf{V}, \mathbf{X}\boldsymbol{\Theta}) &= f(\mathbf{z}|\mathbf{R}, \mathbf{V}, \mathbf{X}, \boldsymbol{\Theta}) \\ &\times f(\mathbf{R}|\mathbf{V}, \mathbf{X}, \boldsymbol{\Theta}) \\ &\times f(\mathbf{V}|\boldsymbol{\Theta}) f(\boldsymbol{\Theta}) \end{aligned} \tag{22}$$

We now model the components of this full likelihood, assuming certain dependence and distributional assumptions:

1. **Assumption 1: Conditional Independence** Conditional on $(\mathbf{R}, \mathbf{V}, \mathbf{X}, \boldsymbol{\Theta})$, $Y_i(s)$ are independent in time and space, with

$$f(\mathbf{z}|\mathbf{R}, \mathbf{V}, \mathbf{X}, \boldsymbol{\Theta}) = \left[\prod_{s=1}^{745} f(z(s)|\mathbf{R}, \mathbf{V}, \mathbf{X}\boldsymbol{\Theta}) \right]^{10}, \tag{23}$$

where

$$f(z(s)|\mathbf{R}, \mathbf{V}, \mathbf{X}\boldsymbol{\Theta}) = \left[N(\mu(s), \sigma^2) \right]^{R(s)} (1 - p_0(s))^{R(s)} p_0(s)^{1-R(s)}. \tag{24}$$

2. **Condition 2: Conditional independence of ignition process**

$$f(\mathbf{R}|\mathbf{V}, \mathbf{X}, \boldsymbol{\Theta}) = \left[\prod_{s=1}^{745} f(r(s)|\mathbf{V}, \mathbf{X}, \boldsymbol{\Theta}) \right]^{10}, \tag{25}$$

where

$$f(r(s)|\mathbf{V}, \mathbf{X}, \mathbf{\Theta})) = (1 - p(s))^j p(s)^{1-j},$$

$j = 0, 1$.

3. **Condition 3: Dependent latent processes**

$$f(\mathbf{V}|\mathbf{\Theta}) = (f(V_1(s)|V_0(s), \mathbf{\Theta}) f(V_0(s)|\mathbf{\Theta}))^{10}, \tag{26}$$

where, $V_0(s)$ is a stationary and isotropic Gaussian process $V_0(s) \sim GP(0, \Sigma_0)$, where the elements of Σ_0 are given by the space time-space covariance structure of Gneiding(2002):

$$Cov(V_0(s_1), V_0(s_2)) = Cov(V_0(0), V_0(h)) = C_0(h), \tag{27}$$

where

$$C_0(h) = \sigma_0^2 \exp\left(-c_0 ||h||^{2\gamma_0}\right), \tag{28}$$

and c_0 is nonnegative scaling parameter, $\gamma_0 \in (0, 1]$ is the smoothness parameter defined in and $\sigma^2 > 0$ is the nugget effect. Note that this covariance function is a special case of the covariance structure given by Gneiding (2002) with $u = 0$.

4. **Conditional specification of the latent process**$V_1(s)$ Assume that

$$V_1(s) = \phi_1 V_0(s) + W_1(s), \tag{29}$$

where $W_1(s)$ is a 0-mean stationary Gaussian process, independent of $V_0(s)$, with covariance structure Σ_1 defined by

$$C_1(h) = \sigma_1^2 \exp\left(-c_1 ||h||^{2\gamma_1}\right), \tag{30}$$

where the parameters have the same range and significance as in $C_0(h)$.

Again, once the prior specification is given to the model parameters, this model becomes a fully specified hierarchical model, considerably simpler than the previous ideal space-time model. However, this model is still computationally infeasible due to the size of the covariance structures (745×745) of the process $V_0(s)$ and $W_1(s)$ and further assumptions and simplifications are needed to deal with computational problems. We will use the predictive process modeling suggested by Gelfand et al(2007): We will approximate the latent processes $V_0(s)$ and $W_1(s)$ by their "predictive" versions $\overline{V}_0(s)$, $\overline{W}_1(s)$, defined below, which make the suggested hierarchical model computationally feasible.

Assume that $\mathbf{s}^* = (s_1^*, ..., S_M^*)$, $M < 100$ are fixed locations, or knots, (Gelfand et all, 2004) chosen (not necessarily among the 745 observation locations) according to some criteria which assure the representativeness of this set of the whole process over the region under study. We refer to Gelfand et all(2006) for the optimal selection of knots. For reasons to be explained below,, we assume that these knots are chosen within the observation locations. Define

$$\overline{V}_0(s) = C_0(s, \mathbf{s}^*)(\Sigma_0^*)^{-1} \mathbf{V}_0(\mathbf{s}^*), \tag{31}$$

$$\overline{W}_1(s) = C_1(s, \mathbf{s}^*)(\Sigma_1^*)^{-1}\mathbf{W}_0(\mathbf{s}^*), \tag{32}$$

where,

$$\mathbf{V}_0(\mathbf{s}^*) = (V_0(s_1^*), ..., V_0(s_M^*)),$$
$$\mathbf{W}_1(\mathbf{s}^*) = (W_1(s_1^*), ..., W_1(s_M^*)),$$

and Σ_0^* and Σ_1^* are the covariance matrices of $\mathbf{V}_0(\mathbf{s}^*)$ and $\mathbf{W}_1(\mathbf{s}^*)$ respectively, that is the covariance matrix of the processes $V_0(s)$ and $W_1(s)$ at knots $\mathbf{s}^*$.

Note that $\overline{V}_0(s)$ and $\overline{W}_1(s)$ are the best linear regression (krigging) of the processes $V_0(s)$ and $W_1(s)$ in terms of the respective observations taken at knots $\mathbf{s}^*$, which minimize the prediction errors, and hence we call them the respective predictive processes. respective covariance structures

$$\overline{\Sigma}_0(\mathbf{s}, \mathbf{s}^*) = C_0^T(\mathbf{s}, \mathbf{s}^*)(\Sigma_0^*)^{-1}C_0(\mathbf{s}, \mathbf{s}^*), \tag{33}$$

$$\overline{\Sigma}_1(\mathbf{s}, \mathbf{s}^*) = C_1^T(\mathbf{s}, \mathbf{s}^*)(\Sigma_1^*)^{-1}C_1(\mathbf{s}, \mathbf{s}^*), \tag{34}$$

where $\mathbf{s} = (s_1, s_2, ..., s_N)$ is any collection of N sites. It is clear from the above expressions that these predictive processes are non-stationary and are linear combinations of the respective V_0 W_1 processes observed at these knots, with spatially varying coefficients, that is

$$\overline{V}_0(s) = \sum_{i=1}^{M} a_i(s)V_0(s_i^*),$$

$$\overline{W}_1(s) = \sum_{i=1}^{M} b_i(s)W_1(s_i^*).$$

Note that both $\overline{V}_0(s)$, $\overline{W}_1(s)$ are fully specified, once $(\Sigma_0^*)^{-1}$ and $(\Sigma_0^*)^{-1}$ are given. Note also that due to Sherman-Woodbury result (Harville, 1997), inversion of the matrices $\overline{\Sigma}_0(\mathbf{s}, \mathbf{s}^*)$ and $\overline{\Sigma}_1(\mathbf{s}, \mathbf{s}^*)$ only requires $(\Sigma_0^*)^{-1}$ and $(\Sigma_1^*)^{-1}$ respectively, and hence the matrix inversion operations are restricted to inversion operations on matrices of dimension $M \times M$. However, if we restrict ourselves to BUGS software, this would not help much, as one can not have access to the matrices $(\Sigma_0^*)^{-1}$ and $(\Sigma_1^*)^{-1}$ during the iterations. Therefore, we suggest a further approximation, which would allow us to use the BUGS. Our suggestion is the following:

1. Fit the full model given above for the data observed at knots $\mathbf{s}^*$. Note that we have chosen the knots among the observational sites.

2. Obtain samples from the posterior distributions of the covariance model parameters

   $$\mathbf{\Theta}_0^* = (\sigma_0^2, \gamma_0, c_0)$$
   $$\mathbf{\Theta}_1^* = (\sigma_1^2, \gamma_1, c_1).$$

   Since the latent processes are now calculated at $M < 100$ locations, computationally, BUGS should handle the inference.

3. Calculate the means of these posterior distributions.
4. Then calculate $(\Sigma_0^*)^{-1}$ and $(\Sigma_1^*)^{-1}$ evaluated at these mean values.
5. Now, run the full model again, this time for the rest of the observed spatial data $\mathbf{s} - \mathbf{s}^*$.

This would involve running BUGS twice for the following models and the data sets:

1. **Model for the First run with the data $\mathbf{z}^*$.**
   Let $\mathbf{z}^* = y_i(s_j^*), i = 1, ..., 10, j = 1, , , , M)$. (Define similarly $\mathbf{R}^*$,$\mathbf{V}^*$,$\mathbf{X}^*$)

   - $$f(\mathbf{z}^*|\mathbf{R}*, \mathbf{V}^*, \mathbf{X}^*\mathbf{\Theta}) = \left[\prod_{s^*=1}^{M} f(y(s^*)|\mathbf{R}^*, \mathbf{V}^*, \mathbf{X}^*, \mathbf{\Theta})\right]^{10}, \quad (35)$$
   - $$f(\mathbf{R}^*|\mathbf{V}^*, \mathbf{X}^*, \mathbf{\Theta}) = \left[\prod_{s^*=1}^{M} f(r(s^*)|\mathbf{V}^*, \mathbf{X}^*, \mathbf{\Theta})\right]^{10}, \quad (36)$$

     where

     $$f(z(s^*)|\mathbf{R}, \mathbf{V}, \mathbf{X}\mathbf{\Theta}) = \left[N(\mu(s^*), \sigma^2)\right]^{R(s^*)} (1-p_0(s^*))^{R(s^*)} p_0(s^*)^{1-R(s^*)}. \quad (37)$$

     $$f(r(s^*)|\mathbf{V}^*, \mathbf{X}^*, \mathbf{\Theta})) = (1 - p(s^*)))^j p(s^*))^{1-j},$$

     $j = 0, 1.$
   - $$\log(\frac{p(s^*)}{1-p(s^*)}) = \mu + \mathbf{X}^T(s^*)\beta_1 + V_0(s^*), \quad (38)$$

     $$\log(\alpha(s^*)) = \eta + \mathbf{X}^T(s^*)\beta_2 + V_1(s^*), \quad (39)$$

     Defining $\mathbf{V}_0(s^*) = (V_0(s_1^*, ..., V_0(s_M^*))$, (similarly $\mathbf{V}_1(s^*)$) assume

     $$\mathbf{V}_0(s^*) \sim N(0, \Sigma_0^*), \quad (40)$$

     $$\mathbf{V}_1(s^*) = \phi_1^* V_0(s^*) + W_1(s^*), \quad (41)$$
     $$W_1(s^*) \sim N(0, \Sigma_1^*),$$

     where the elements of Σ_0^* and σ_1^* are evaluated at locations $\mathbf{s}^*$ from the covariance structure

     $$C_0(h) = \sigma_0^2 \exp\left(-c_0||h||^{2\gamma_0}\right), \quad (42)$$

     and

     $$C_1(h) = \sigma_1^2 \exp\left(-c_1||h||^{2\gamma_1}\right). \quad (43)$$

     With prior specification for the parameters and hyper parameters, this is a fully specified hierarchical model and can be fitted by BUGS.

- **Model for the second run with the full data** $Y_i(\mathbf{s})$ Let $\mathbf{V}(s) = (\overline{V}(s), \overline{W}(s))$ be the latent process. Let $\mathbf{z}(\mathbf{s}) = (Y_i(s), s = 1, ..., 745, i = 1, ..., 10$, and define $\mathbf{R}(\mathbf{s})$,$\mathbf{V}(\mathbf{s})$, $\mathbf{X}(\mathbf{s})$ in similar fashion.

$$f(\mathbf{z}_i(\mathbf{s})|\mathbf{R}(\mathbf{s})), \mathbf{V}(\mathbf{s})), \mathbf{X}(\mathbf{s})\mathbf{\Theta}) = \left[\prod_{s=1}^{745} f(y(s)|\mathbf{R}(s), \mathbf{V}(\mathbf{s}), \mathbf{X}(\mathbf{s}), \mathbf{\Theta})\right]^{10}, \tag{44}$$

$$f(z(s)|\mathbf{R}, \mathbf{V}, \mathbf{X}\mathbf{\Theta}) = \left[N(\mu(s), \sigma^2)\right]^{R(s)} (1 - p_0(s))^{R(s)} p_0(s)^{1-R(s)}. \tag{45}$$

- $$f(\mathbf{R}(s)|\mathbf{V}(\mathbf{s})), \mathbf{X}(\mathbf{s}), \mathbf{\Theta}) = \left[\prod_{s=1}^{M} f(R(s)|\mathbf{V}(\mathbf{s}^{)}), \mathbf{\Theta})\right]^{10}, \tag{46}$$

  where

  $$f(r(s)|\mathbf{V}(\mathbf{s})), \mathbf{X}(\mathbf{s}), \mathbf{\Theta})) = (1 - p(s)))^j p(s))^{1-j},$$

  $j = 0, 1$.

- $$\log(\frac{p(s)}{1 - p(s)})\mu + \mathbf{X}^T(s)\beta_1 + \overline{V}_0(s), \tag{47}$$

  $$\log(\alpha(s)) = \eta + \mathbf{X}^T\beta_2 + \overline{W}(s), \tag{48}$$

  $$\overline{V}_0(s) = C_0(s, \mathbf{s}^*)(\Sigma_0^*)^{-1}\mathbf{V}_0(\mathbf{s}^*), \tag{49}$$

  $$\overline{W}(s) = \phi_1\overline{V}_0(s) + \overline{W}_1(s), \tag{50}$$

  $$\overline{W}_1(s) = C_1(s, \mathbf{s}^*)(\Sigma_1^*)^{-1}\mathbf{W}_0(\mathbf{s}^*), \tag{51}$$

  $$\mathbf{V}_0(\mathbf{s}^*) \sim N(0, \Sigma_0^*), \tag{52}$$

  $$\mathbf{W}_0(\mathbf{s}^*) \sim N(0, \Sigma_1^*) \tag{53}$$

With prior specification for the parameters and hyper parameters, this is a fully specified hierarchical model and can be fitted by BUGS.

4 References

Banarjee,S., Gelfand,A: Finley,O. and H. Sang(2007) Gaussian Predictive Process models for large spatial data sets.

Gelfand,A:, Schmit,A., Banerjee,S. and C. Firmans(2004) Nonstationary multivariate process modeling through spatially varying coreginalization. TEST VOL. 13, PP 263-312.

Gneiding,T. (2002) Nonseparable, stationary covariance functions for space-time data. JASA 97, 590-600.

Banarjee,S., Carlin,B. and A. Gelfand(2004) Hierarchical modeling and analysis for spatial data. Chapman and Hall.