UDMAD: Shiny Web Application for Single Species Seasonal Spatio-temporal Count Models for Areal Data

Udani Wijewardhana
Swinburne University of Technology

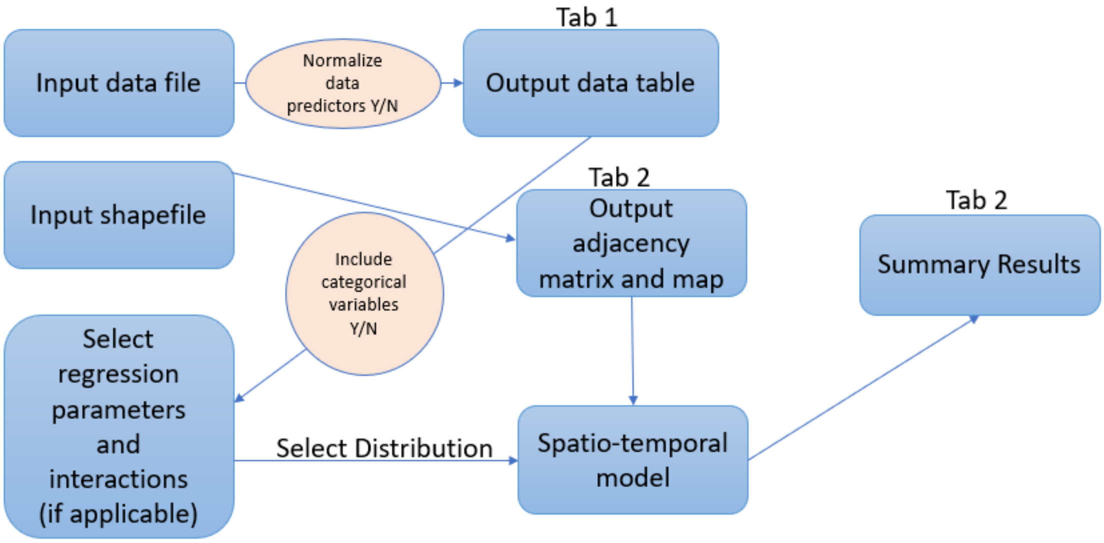
The spatial analysis of ecological data is fundamental to the majority of current challenges in applied ecology. Species distribution models give critical information about species–environment interactions, notably how environmental stresses limit species dispersal (Elith & Leathwick, 2009). Predictions obtained from such models can influence management plans for endangered species (Cabeza et al., 2004), as well as estimates of distributions under future land-use (Bomhard et al., 2005) and climatic (Thuiller, 2003) scenarios, in the global context of rising human impacts on ecosystems. Species distribution analyses are sensitive to spatial reliance in model residuals, or residual spatial autocorrelation (Latimer et al., 2006).

This shiny application used areal data to fit spatio-temporal models of seasonal species counts. When a fixed domain is partitioned into a finite number of subregions from which results are aggregated, areal or lattice data are produced. The Besag-York-Mollié (BYM) model, besag, besagproper, and BYM2 (Besag, York, and Mollié, 1991) have taken into account that data may be spatially correlated and observations in neighbouring areas may be more comparable than observations in areas further away. This model incorporates a spatial random effect that smooths the data based on a neighbourhood structure, as well as an unstructured exchangeable component that mimics uncorrelated noise. Spatio-temporal models that account not only for spatial structure but also for temporal correlations and spatio-temporal interactions are utilised in spatio-temporal contexts where species counts are monitored throughout time. We determined the mean expected value of the counts for each fitted model (with 95% credible intervals).

Layout

The application consists of two tabs. The `sidebarPanel` is shared by both tabs. The user must first upload the .CSV file and decide whether or not to normalise the numeric predictors. The user should next upload the shapefile (a combination of `.shp,.dbf,.shx,` and `.prj` files) to construct the adjacency matrix. The area name is a unique identifier for the area, and it should match the area names on the map. This shiny application used areal data to fit spatio-temporal models of seasonal species counts. This can be used to count data on a monthly or daily basis by substituting the values seasonID and Season below with monthly/daily values.

The data CSV should include: Trend - Detected Trend, Count - Species count, Area - Area Name, areaID - Area ID, Year - Detected Year, Season - Detected Season, seasonID - Season ID with or without numeric/factor predictor variables. These names are case sensitive. Data should be ordered according to factor levels as in sample “Sample Data.csv”. Factor variables should add as factor/character without numeric/integer. This can also be used for monthly or daily data. Simply replace seasonID and Season with the appropriate values. A sample format of the data can be found in https://github.com/uwijewardhana/UDMAD.



Layout of the UDMAD App

Theory of attributes

Ecologists have been more interested in Bayesian implementation of spatially explicit models as Monte Carlo Markov Chain (MCMC) methods have become more widely available. MCMC simulations provide a flexible framework for modelling large amounts of ecological data, but they also have a number of issues with convergence, processing time, and implementation. Ecologists can utilise a different approach for fitting Bayesian hierarchical spatial models with very general spatial covariance structures in this case. As an alternative to MCMC, this approach employs Integrated Nested Laplace Approximations (INLA). INLA is a newer alternative to MCMC for fitting a wide range of Bayesian models, including latent Gaussian models (Rue, Martino & Chopin, 2009). Latent Gaussian models are capable of accounting for hierarchical structure, non-Gaussian errors, and spatial and temporal autocorrelation. INLA replaces long MCMC simulations with accurate, deterministic approximations to posterior marginal distributions for fitting these models, gaining speed. Comparisons with extensive MCMC runs reveal that the quality of such approximations is exceptionally high (Rue, Martino & Chopin, 2009).

Areal data, also known as lattice data, is data that is observed inside specific bounds. Administrative borders will result in a lattice with an uneven structure. Users can simulate using irregular lattice data with this dazzling software. A (sparse) adjacency matrix with non-zero entries at the intersection of rows and columns of neighbouring areas is commonly used to depict spatial adjacency. This matrix will be crucial in the construction of spatial models since spatial correlation structures will be based on it. Spatially linked random effects, in particular, will be modelled using a multivariate Gaussian distribution with a precision matrix that is dependent on an adjacency matrix.

We can utilise spatio-temporal models that account not only for geographical structure but also for temporal correlations and spatio-temporal interactions in spatio-temporal environments where disease counts are observed throughout time (Martinez-Beneito et al., 2008; Ugarte et al., 2014). The model that fits here considers the response in log-scale, which makes the response less skewed. The autoregressive model or random walks can be used to fit the temporal random effect. Four models will be studied for spatial models for data in an irregular lattice: Besag's proper spatial model (besagproper), Besag's improper spatial model (besag), and two by Besag, York, and Mollié that are convolutions of an intrinsic CAR model (BYM and BYM2) and a iid Gaussian model.

The Poisson, Negative Binomial, Zero-inflated Poisson, Zero-inflated Negative Binomial, Poisson Hurdle, or Negative Binomial Hurdle models can be used to fit the INLA model. An interaction effect exists in regression when the influence of one or more independent variables on a dependent variable changes depending on the value(s) of one or more other independent variables. Users can add interaction terms to their regression models in this section. The regression mode we have used is stated below.

$$Count \sim 1 + independent\ variables + interaction\ terms\ (if\ applicable) \\ + f(trend, temporal\ effect\ model) + f(areaID, spatial\ effect\ model)$$

Structure of UDMAD App

UDMAD is a Shiny web application that allows to fit single species spatio-temporal models and estimate significant factors with the trend. It is addressed to ecologists interested in analysing species count data. This app developed based on R-INLA (Rue et al., 2018) which provides a number of options to model data collected in space and time. Information about all the packages used are shown in Table 1.

Table 1 Softwares and R packages used for developing UDMAD

Package	Authors	Description
tidyverse	Wickham et. al., 2019	The ‘tidyverse’ is a set of packages that work in harmony because they share common data representations and ‘API’ design. This package is designed to make it easy to install and load multiple ‘tidyverse’ packages in a single step.
shinythemes	Chang, 2021	Themes for use with Shiny. Includes several Bootstrap themes from https://bootswatch.com/, which are packaged for use with Shiny applications.
INLA	Rue, et. al., 2009	Performs full Bayesian analysis on generalised additive mixed models using Integrated Nested Laplace Approximations.
Shiny	Chang et al., 2016	Web Application Framework for R.
DT	Xie, et. al., 2021	Create data tables.
rgdal	Keitt et. al., 2010	Provides bindings to the 'Geospatial' Data Abstraction Library ('GDAL') (>= 1.11.4) and access to projection/transformation operations from the 'PROJ' library. Please note that 'rgdal' will be retired by the end of 2023, plan transition to sf/stars/'terra' functions using 'GDAL' and 'PROJ' at your earliest convenience. Use is made of classes defined in the 'sp' package. Raster and vector map data can be imported into R, and raster and vector 'sp' objects exported. The 'GDAL' and 'PROJ' libraries are external to the package, and, when installing the package from source, must be correctly installed first; it is important that 'GDAL' < 3 be matched with 'PROJ' < 6. From 'rgdal' 1.5-8, installed with to 'GDAL' >=3, 'PROJ' >=6 and 'sp' >= 1.4, coordinate reference systems use 'WKT2_2019' strings, not 'PROJ' strings. 'Windows' and 'macOS' binaries (including 'GDAL', 'PROJ' and their dependencies).

mapview	Appelhans et. al., 2021	Quickly and conveniently create interactive visualisations of spatial data with or without background maps. Attributes of displayed features are fully quarriable via pop-up windows. Additional functionality includes methods to visualise true and false-colour raster images and bounding boxes.
spdep	Bivand and Wong, 2018	A collection of functions to create spatial weights matrix objects from polygon 'contiguities', from point patterns by distance and tessellations, for summarizing these objects, and for permitting their use in spatial data analysis, including regional aggregation by minimum spanning tree; a collection of tests for spatial 'autocorrelation'.
sf	Pebesma, 2018	Support for simple features, a standardized way to encode spatial vector data. Binds to 'GDAL' for reading and writing data, to 'GEOS' for geometrical operations, and to 'PROJ' for projection conversions and datum transformations. Uses by default the 's2' package for spherical geometry operations on ellipsoidal (lon/lat) coordinates.
leaflet	Cheng, et. al., 2021	Leaflet is the leading open-source JavaScript library for mobile-friendly interactive maps. Weighing just about 39 KB of JS, it has all the mapping features most developers ever need. Leaflet is designed with simplicity, performance and usability in mind. It works efficiently across all major desktop and mobile platforms, can be extended with lots of plugins, has a beautiful, easy to use and well-documented API and a simple, readable source code that is a joy to contribute to.

After uploading the data file, used need to upload the shapefile. The shapefile can be read using the query below. The `st_read` function can be used to read the shapefile. The f() function is used to define the spatial latent effect. This necessitates the creation of an index to identify the random effects in each area, as well as the type of model and the adjacency matrix. A sparse matrix has utilised for this. Here `poly2nb()` function use to generate the adjacency matrix.

```
previouswd <- getwd()
    uploaddirectory <- dirname(shpdf$datapath[1])
    setwd(uploaddirectory)
    setwd(previouswd)

    map <- st_read(paste(uploaddirectory, shpdf$name[grep(pattern="*.shp$"
, shpdf$name)], sep="/")) # Delete_null_obj=TRUE)
```

```
    sf::sf_use_s2(FALSE)
    p <- poly2nb(map)
    td <- tempdir() # generate a temporary directory
    # create adjacency matrix
   nb2INLA(paste(td, "map.adj", sep = "/"), p())
    # convert the adjacency matrix into a file in the INLA format
    g = inla.read.graph(filename = paste(td, "map.adj", sep="/"))
```

Once you enter summary button the app firstly generates the adjacency matrix for INLA model which takes nearly 3-4 minutes. The analysis of temporal data over a continuous domain will be done with smoothing methods described such as ‘iid’, ‘$ar1$’, ‘$rw1$’ or ‘$rw2$’ to take the short-term fluctuations into account. We used `selectInput()` to add these choices and call the temporal effect inside the formula. Similarly, the spatial effect models such as $BYM, BYM2, besag\ and\ besag$ proper also added to model using `selectInput()`.

Set up and installation

To build this Shiny app, we need to clone the GitHub repository from UDMAD and save it in our computer. This folder contains a sample Data.CSV file, the vignette and `app.R` file. Then, we can launch the app by clicking the Run App button at the top of the RStudio editor or by executing `runApp("appdir_path")`where `appdir_path` is the path of the directory that contains the `app.R` file. For this we need to install R and RStudio in our computer. A snapshot of the Shiny app is shown in Figure 1.

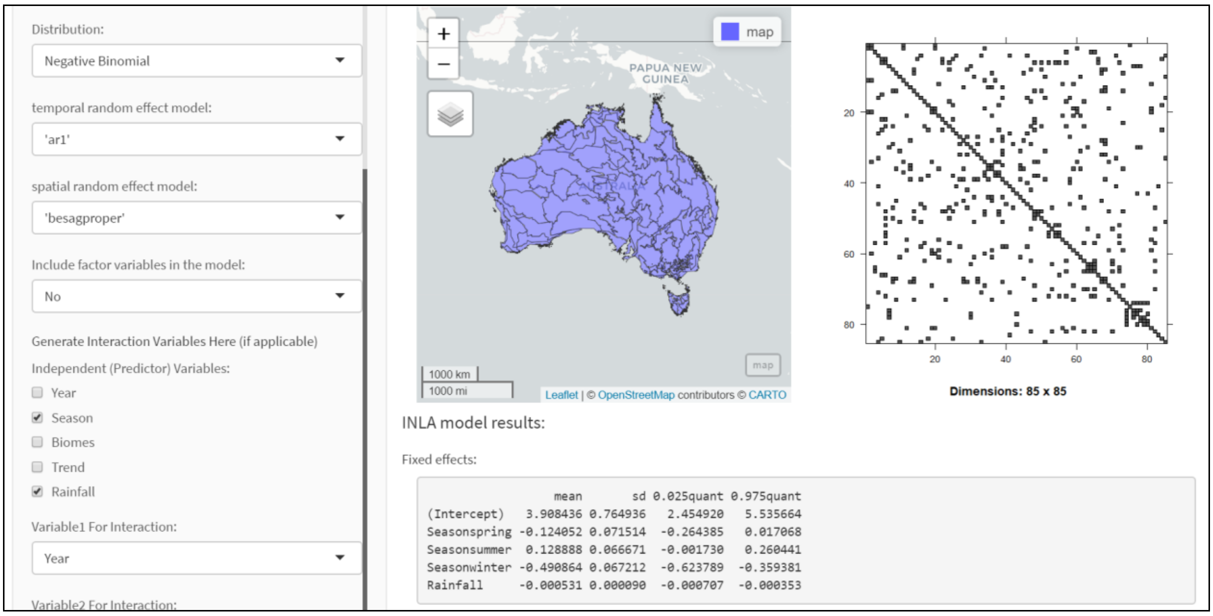


Figure 1 Screenshot of the UDMAD App

References

Elith, J. and Leathwick, J.R. (2009) Species Distribution Models, Ecological Explanation and Prediction across Space and Time. Annual Review of Ecology and Systematics, 40, 677-697. http://dx.doi.org/10.1146/annurev.ecolsys.110308.120159.

Cabeza, M., Arau´ jo, M.B., Wilson, R.J., Thomas, C.D., Cowley, M.J.R. & Moilanen, A. (2004) Combining probabilities of occurrence with spatial reserve design. Journal of Applied Ecology, 41, 252–262.

Bomhard, B., Richardson, D.M., Donaldson, J.S., Hughes, G.O., Midgley, G.F., Raimondo, D.C., Rebelo, A.G., Rouget, M. & Thuiller, W. (2005) Potential impacts of future land use and climate change on the Red List sta tus of the Proteaceae in the Cape Floristic Region, South Africa. Global Change Biology, 11, 1452–1468.

Thuiller,W. (2003) BIOMOD – optimizing predictions of species distributions and projecting potential future shifts under global change. Global Change Biology, 9, 1353–1362.

Latimer, A.M.,Wu, S.S., Gelfand, A.E. & Silander, J.A. (2006) Building statistical models to analyze species distributions. Ecological Applications, 16, 33–50.

Besag, J., York, J. and Mollié, A. (1991). Bayesian Image Restoration, with two Applications in Spatial Statistics. Annals of the Institute of Statistical Mathematics 43 1–20.

Rue, H., Martino, S. and Chopin, N. (2009), Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 71: 319-392. https://doi.org/10.1111/j.1467-9868.2008.00700.x.

Martinez-Beneito M.A., López-Quílez A., Botella-Rocamora P. An autoregressive approach to spatio-temporal disease mapping Stat. Med., 27 (2008), pp. 2874-2889.

Ugarte, María Dolores, Aritz Adin, Tomas Goicoa, and Ana Fernandez Militino. (2014). "On fitting spatio-temporal disease mapping models using approximate Bayesian inference." *Statistical Methods in Medical Research* 23 (6): 507–30.

Rue, H., Lindgren, F., Simpson, D., Martino, S., Krainski, E.T., Bakka, H., Riebler, A., and Fuglstad, G. (2018). *INLA: Full Bayesian Analysis of Latent Gaussian Models Using Integrated Nested Laplace Approximations*.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H. (2019). "Welcome to the tidyverse." *Journal of Open Source Software*, 4(43), 1686. doi: 10.21105/joss.01686.

Keitt, T., Bivand, R., Pebesma, E. & Rowlingson, B. (2010). rgdal: Bindings for the Geospatial Data Abstraction Library. R package version 0.8-13.

Chang, W., Cheng, J., Allaire, J.J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., and Borges, B. (2021). shiny: Web Application Framework for R. R package version 1.7.1. https://CRAN.R-project.org/package=shiny.

Bivand, R.S. and Wong, D.W.S. (2018) Comparing implementations of global and local indicators of spatial association TEST, 27(3), 716-748. URL https://doi.org/10.1007/s11749-018-0599-x.

Appelhans, T., Detsch, F., Reudenbach, C. and Woellauer, S. (2021). mapview: Interactive Viewing of Spatial Data in R. R package version 2.10.0. https://CRAN.R-project.org/package=mapview.

Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. The R Journal 10 (1), 439-446, https://doi.org/10.32614/RJ-2018-009.

Cheng, J., Karambelkar B. and Xie, Y. (2021). leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library. R package version 2.0.4.1. https://CRAN.R-project.org/package=leaflet.

Xie, Y., Cheng, J. and Tan, X. (2021). DT: A Wrapper of the JavaScript Library 'DataTables'. R package version 0.19. https://CRAN.R-project.org/package=DT.

Chang, W. (2021). shinythemes: Themes for Shiny. R package version 1.2.0. https://CRAN.R-project.org/package=shinythemes.

Rue, H., Martino, S. and Chopin, N. (2009), Approximate Bayesian Inference for Latent Gaussian Models Using Integrated Nested Laplace Approximations (with discussion), Journal of the Royal Statistical Society B, 71, 319-392.