UDMSD - Shiny app for single species joint models

To estimate the persistence, user can consider Integrated Laplace Approximation (INLA) models with zero-inflation probability model which called a joint model. Joint temporal, spatial and spatio-temporal models are especially developed for data with access zeros. Access zeros are divided into two types such as structural zeros and random zeros. Structural zeros refer to zero responses by those subjects whose count response will always be zero and random zeros that occur to subjects whose count response can be greater than zero but appear to be zero due to sampling variability (He et al., 2014). The responses of species data came from two different distributions such as occurrence and abundance which have two models for each response that are affected by spatial and temporal common factors. Therefore, it is better to use joint model. Negative Binomial model is more in line for zero inflation species data which resolve the overdispersion issue as well. Therefore users can use this app to fit Hurdle models (Cragg, 1971; Mullahy, 1986) or Zero Inflated models (Cameron and Trivedi, 2013) with Negative Binomial distribution which give the summary of the fitted joint which are the most common zero inflation joint models to identify the significance of the predictors.

A Bayesian hierarchical modelling approach is used here to allow us to conveniently account for structures in parameter uncertainty and potential dependence such as spatial and temporal structures. With a Bayesian approach, a joint posterior distribution is obtained for the process and parameters of the given data. This is conducted using INLA. INLA is popular as an approximation tool for fitting Bayesian models. INLA is an alternative robust method for the traditional Markov Chain Monte Carlo (MCMC) based Bayesian analyses (Paul et al. 2010). INLA approach is a numerically implemented analytical solution for approximating posterior marginals in hierarchical models with latent Gaussian processes. The key advantages of INLA are the ease with which complex models can be created and modified, without the need to write complex code, and the speed at which inference can be done even for spatial problems with hundreds of thousands of observations (Sarul, 2015). In spatio-temporal settings, it is often assumed that the covariance is separable in space and time, and thus, the temporal structure may be modelled using an auto-regressive process (Cressie, 2011). Fitted INLA with log-linear regression for the positive counts (truncated count model) and a logistic regression for the zero counts using a Bernoulli distribution in spatial and temporal scales. When we are dealing with process defined over a continuous domain, one can express a large class of random fields as solution of stochastic partial differential equations (SPDEs). In R-INLA this solution is approximated using high dimensional basis representation with simple local basis function.

These basis functions are defined over a triangulation of the domain; this triangulation is the mesh https://ourcodingclub.github.io/tutorials/inla/). This app can only create non-convex meshes. The general hierarchical model can be described in three phases such as data model, process model and parameter model. Adapted these INLA models using R-INLA (Martins, Simpson, Lindgren and Rue, 2013).

To know if the rate at which abundance is changing over time differs according to a relevant predictor variable, we have included the facility to add interaction terms between any two predictor variables in our regression models. The interaction with a categorical variable tells us what the difference in slope is and whether this difference is significant.

UDMSD is a Shiny web application that allows to visualize spatial, temporal or spatio-temporal abundance/occurrence data, estimate significant factors and the trend. It is addressed to

ecologists interested in analysing species abundance/occurrence data but lacking the appropriate theoretical knowledge to use this statistical software is required. The application allows to fit Bayesian Zero inflation models to obtain significant predictor estimates and their uncertainty by using INLA. To carry out these analyses users simply need to click the buttons that create the input files required, execute the software and process the output to generate tables of values and plots with the results. The application allows user interaction and creates interactive visualizations such as species distribution map and posterior maps. This INLA models could be done only for geostatistical data.

The application consists of 2 pages with main window: main window allows the user to upload the input files (data file and predictors file) and also gives the option to normalize or standardize the data or predictors data.

1) Data display page where the user can see the annual data table of the uploaded data file and uploaded annual predictors table.

2) an analysis page where statistical analyses are carried out, and summary results and posterior plots can be visualized.

In this page, user can fit joint species distribution models using INLA. Since species abundance data usually have access and true zeros, zero inflation models are ideal to fit the abundance. Therefore, while overcoming overdispersion Negative Binomial model, Negative Binomial Hurdle model and ZINB model have given as options using log linear regression as the link function. Binomial distribution has used for occurrence using logistic regression as the link function. First, a user must adjust the parameters to build the INLA mesh. Only non-convex meshes could be built in this app. Then users can fit spatial, temporal or spatio-temporal joint species distribution while changing the model options and distribution. Users also have the option to select which type of spatial and temporal models could be done. Temporal model has the option to choose the random effect model such as 'iid', 'ar1', 'rw1' or 'rw2' and the spatial effect model such as 'iid' for temporal models and 'spde' for spatial or spatio-temporal models. Posterior plots are visualizing for spatial or spatio-temporal models.

Input file

Dataframe with "Count", "Year", "Latitude", "Longitude", "Locality" and other numeric predictors (optional). These variables are case sensitive. Only applicable up to 10 predictor variables (If you wish to find the trend model, Year is already included as a predictor variable). Predictor variables display as "p.z1", "p.y2", etc. The order is the exact order in dataframe.

Installation

To build this Shiny app, we need to clone the Zip file from UDMSD 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. The users who do not have R in their computer can use UDMSD to launch the Shiny app. The application uses the R-INLA package which can be downloaded from http://www. r − inla. org.

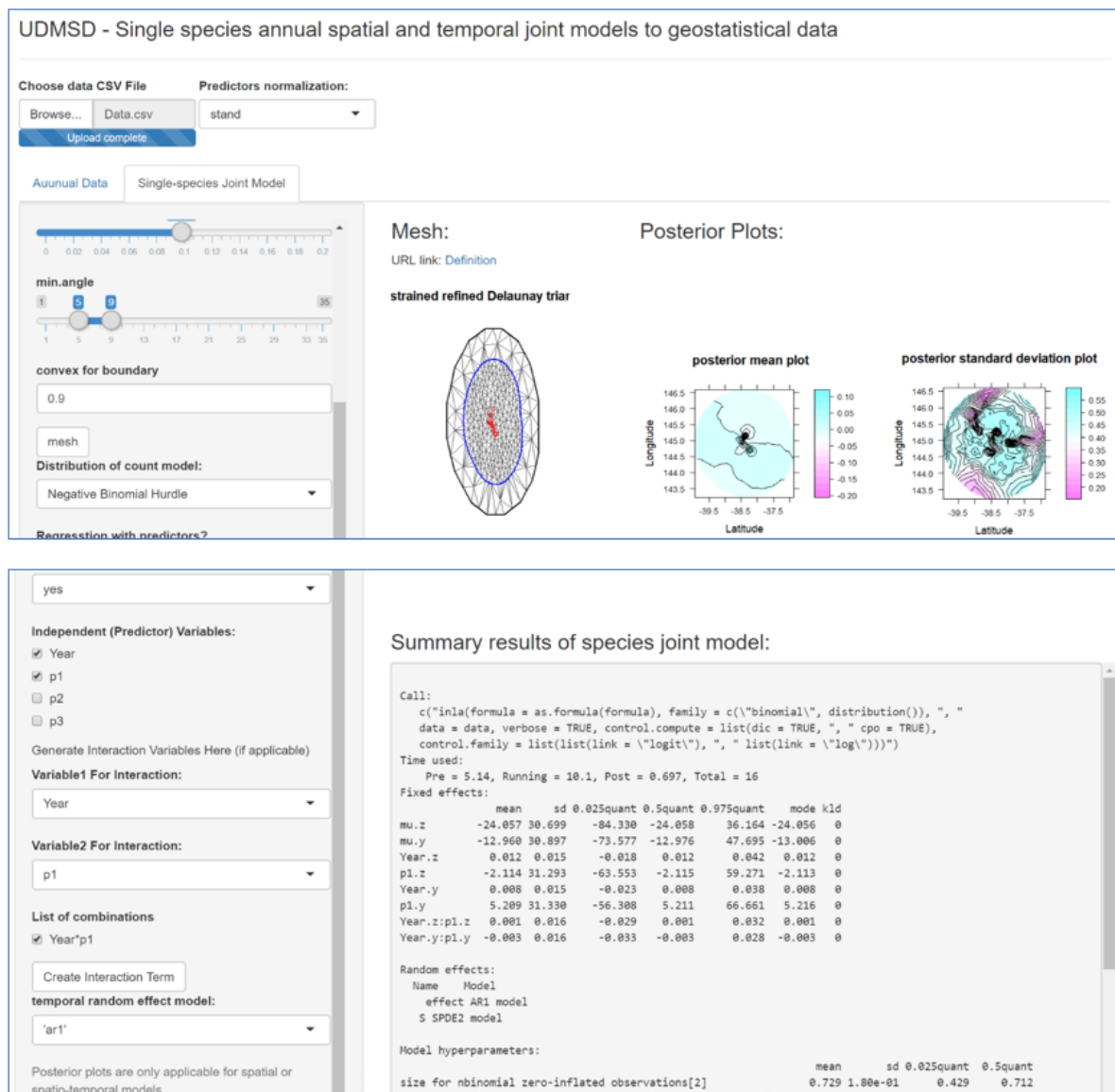A snapshot of the app is shown in Figure 1.

Figure 1 Snapshot of the UDMCA App

References

Cragg, J.G. Some statistical models for limited dependent variables with application to the demand for durable g oods. *Econometrica* 1971, *39*, 829–844.

Cameron, A.C., Trivedi P.K. *Regression Analysis of Count Data*; Cambridge University Press: Cambridge, UK, 1998.

Mullahy, J. Specification and testing of some modified count data models. *J. Econom.* 1986, *33*, 341–365. Cressie, N., Wikle, C.K. *Statistics for Spatio-temporal Data*; John Wiley & Sons: New Jersey, NJ, USA, 2011.

Paul, M., Riebler, A., Bachmann, L. M., Rue, H., and Held, L. (2010). Bayesian bivariate meta-analysis of diagn ostic test studies using integrated nested Laplace approximations. Statistics in Medicine, 29: 1325–1339.

He, H., Wang, W., Crits-Christoph, P., Gallop, R., Tang, W., Chen, D. D., & Tu, X. M. (2014). On the implicati on of structural zeros as independent variables in regression analysis: applications to alcohol research. Journal of data science : JDS, 12(3), 439–460.

Sarul, L. (2015). An Application Of Claim Frequency Data Using Zero Inflated And Hurdle Models In General I nsurance. Pressacademia, 4(4), pp.732-732.

Martins, T. G., Simpson, D., Lindgren, F., & Rue, H. (2013). Bayesian computing with INLA: New features. *Co mputational Statistics & Data Analysis*, *67*, 68-83. https://doi.org/10.1016/j.csda.2013.04.014.