

A Shiny Web Application for Single Species Persistence Probability Models and Changepoint Analysis of Species Count Data

Udani A. Wijewardhana¹, Madawa Jayawardana^{1, 2, 3}, Denny Meyer¹

¹ Department of Statistics, Data Science and Epidemiology, Swinburne University of Technology, Hawthorn, Victoria, Australia

² Peter MacCallum Cancer Centre, Melbourne 3000 Victoria, Australia

³ Sir Peter MacCallum Department of Oncology, The University of Melbourne, Parkville 3010 Victoria, Australia

* uwijewardhana@swin.edu.au

Abstract

Species-habitat relationship information is often used for conservation purposes. Here we present an interactive Shiny app PPMCA which integrates three approaches of species abundance/occurrence: persistence probability model, single species joint zero inflation model and changepoint analysis. The application allows user interaction and the creation of interactive data visualizations. PPMCA is only applicable for annual data with maximum of five predictors or less. Persistence probability model shows the persistence of the most recent year considering occurrence. Species joint distribution model allows to fit Bayesian spatial, temporal or spatio-temporal structures while considering Negative Binomial distribution as the simplest method or Negative Binomial Zero Inflated model or Negative Binomial Hurdle model when the data is zero inflated. We have considered the truncated count model with log linear regression and Bernoulli distribution for the zero inflation probability model with logistic regression since Negative Binomial distribution is more in line for zero inflated overdispersed data. We have adapted these models using R-INLA. User can also find the changepoints of the abundance level using raw data taking four Bayesian changepoint techniques into account. PPMCA is easy to use and does not require any programming knowledge.

Introduction

Extinction of bird species has become too frequent a theme around the world. In 2015 BirdLife International estimated that 13% of bird species are threatened with extinction. Each extinction diminishes biodiversity and the benefits it might have afforded lost forever (Region 4 Southeast Region, 2019).

Species distribution models can be applied to identify the extinction risks of threatened species which in turn helps to begin conservation planning activities to protect these species before they are lost. The development of species distribution models has benefited biodiversity conservation through their linkage of science to policy and decision processes. These models have evolved to provide scenarios for future landscapes based on known and projected environmental parameters. Obtaining accurate estimates of trends and other population changes in species observation data is difficult unless some baseline has previously been set, especially when considering threatened species. Spatial or temporal scales can confound inference about changes in species observation data when used to draw conclusions about potential impacts on a different scale.

Determining whether a bird species has become extinct or extant has become a difficult exercise due to survey efforts being too costly. However, due of citizen science data, the

number of sightings has increased over the years (e.g. the eBird database increased by more than 20 million sightings from 2012 to 2018). The online collection of citizen science databases such as eBird and the Living Atlas of Australia are becoming common worldwide, providing large validated datasets.

The number of recorded species in large numbers of geographic areas are now increasing because of citizen science data. However, citizen science count data often includes excess zeros for bird species especially for rare bird species in specific areas or time periods which means dataset is characterized by an irregular distribution for counts such as; few observations of large flocks, occasional observations of small flocks or single birds, and many zeros. In that reason overdispersion is seemed to be a main problem. Negative Binomial Model is developed as an alternative to handle this overdispersion. Zero Inflated models and Hurdle models are developed to handle this type of zero-inflated count data. If all of the zeros are excess zeros, the Hurdle model is more appealing and when zero counts consist of both true and excess zeros, Zero Inflated models is more appealing (Xie et.al.,2013). Therefore, in this app we have given the option to fit models such as Negative Binomial, Zero Inflated Negative Binomial (ZINB) and Negative Binomial Hurdle model for both temporal and spatio-temporal structures to identify the ideal models for the scenario using Integrated Laplace Approximation (INLA).

The basic probability of persistence model (Ree and McCarthy, 2005) can predict the persistence for an area which predicts the status whether a species is likely to be extinct or not using occurrence data. However, this model is more effective for multispecies cases to evaluate the number of species persisting in each area extant. Since we are dealing with persistence, as a beginning we have included this method as well.

Changepoints are abrupt variations in timeseries data (Aminikhanghahi and Cook, 2017). These abrupt changes may represent transitions that occur over the years. Detection of change points is useful in modelling and prediction of time series and is found in application areas such as climate change detection and human activity analysis. User can identify the changepoints of a species timeseries to inform policymaking for situations such as future pest eradication programs affecting the specific species population.

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 needed for the interpretation of the results. The application allows user interaction, creates interactive visualizations of the data and results, and generates summary results showing the analyses performed. The remainder of the paper is organized as follows. First, we introduce the basic Persistence Probability model and then joint single species zero inflation models using in spatial and temporal scales. The joint method contains Negative binomial zero inflation distributions for truncated count model with log linear regression and Bernoulli distribution for the zero inflation probability model with logistic regression. Finally, users can use raw data to identify the changepoints in timeseries using four different Bayesian changepoint techniques replacing zeros for missing values. The summary results and profile plot are presented here. We have described the Shiny and the main components we use to write the application in the next sections. Instructions for the installation of the application are also provided in next sections.

Statistical models with packages

Persistence Probability Model

To obtain persistence probabilities, users could consider annual occurrence data for the time period. The Bayesian formulation of Solow's (1993) equation was used to calculate the posterior probability (p) of the species being extant (Ree and McCarthy, 2005).

$$p = \frac{1}{1 + \frac{1 + \left[\left(\frac{T}{t}\right)^{N-1} - 1\right]}{(N-1)}}$$

Assuming that the prior probability of the species being extant in the last year is 0.5. This prior probability is the probability of the species being extant in the last year of recording prior to considering the count data. In this formula, N is the number of years in which the species was recorded between year 0 and year T and the t is the year when the species was last recorded. The probability of persistence (p) is a score between 0 and 1.0 which is assessed in each area. The probability of persistence ($p = 1$) means the species certain to be persistent at the end of the recording period, ($p \geq 0.5$) means it is more likely to be extant in that area rather than extinct, and ($p < 0.5$) means it is more likely to be extinct than extant.

Single-species Joint Model

To estimate the persistence, user can consider INLA models with zero-inflation probability model which means joint models. Joint temporal, spatial and spatio-temporal models are especially developed for data which has 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 (NB) model is more in line for zero inflation species data which resolve the overdispersion issue as well. Therefore in this window users can get the summary outputs for Joint NB model (Cameron and Trivedi, 2013), Joint Hurdle NB model (Cragg, 1971; Mullahy, 1986) or Joint Zero Inflated NB (ZINB) model (Cameron and Trivedi, 1998) which are the most common zero inflation joint models to identify the significance of the predictors and to identify the persistency.

Negative Binomial model:

$$P(Y_i = 0) = (1 - p)^r$$
$$P(Y_i = 0) = \frac{(k + r - 1)!}{(r - 1)! k!} p^k (1 - p)^r, k = 1, 2, 3, \dots$$

where r is the number of successes, k is the number of failures, and p is the probability of success.

Negative Binomial Hurdle model:

Assumes that there is a Bernoulli random variable that determines whether a count is zero or positive. For the set of n independent observations Y_i (for $i = 1, \dots, n$) this model is described as a mixture of a point mass at zero with probability p and a zero-truncated Negative Binomial distribution with probability $(1 - p)$:

$$P(Y_i = 0) = p, 0 \leq p \leq 1$$

$$P(Y_i = k) = (1 - p) \frac{\Gamma(k + \alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(k + 1)[1 - (1 + \alpha\mu)^{-1/\alpha}]} \left(\frac{\alpha^{-1}}{\mu + \alpha^{-1}}\right)^{\alpha^{-1}} \left(\frac{\mu}{\mu + \alpha^{-1}}\right)^k, k = 1, 2, 3, \dots \text{ and } \mu, 1/\alpha > 0$$

where Y_i is the i^{th} response and μ is the mean of the untruncated Negative Binomial distribution. This definition is extended so that a log-linear regression model can be considered with observation-specific means, μ_i 's based on predictor variable information while similarly considering observation-specific probabilities p_i for zero counts in logistic regression (Cragg, 1971; Mullahy, 1986).

Zero Inflated Negative Binomial model:

It is a mixture of a point mass at zero and a Negative Binomial distribution with probability $(1 - p)$:

$$P(Y_i = 0) = p + (1 - p) \left(\frac{\alpha^{-1}}{\mu + \alpha^{-1}}\right)^{\alpha^{-1}}, 0 \leq p \leq 1 \text{ and } \mu, 1/\alpha > 0$$

$$P(Y_i = k) = (1 - p) \frac{\Gamma(k + \alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(k + 1)} \left(\frac{\alpha^{-1}}{\mu + \alpha^{-1}}\right)^{\alpha^{-1}} \left(\frac{\mu}{\mu + \alpha^{-1}}\right)^k, k = 1, 2, 3, \dots \text{ and } \mu, 1/\alpha > 0$$

As explained for the Zero Inflated Poisson model, Y_i is the i^{th} response and μ is the mean of the untruncated Negative Binomial distribution. This definition is extended so that a log-linear regression model can be considered with observation-specific means, μ_i 's based on predictor variable information while similarly considering observation-specific probabilities p_i for zero counts in logistic regression (Cameron and Trivedi, 1998).

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 Integrated Nested Laplace Approximation (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.

Changepoint Analysis

User can use four Bayesian changepoint methods to identify the significant predictors and the significant changes of the abundance level using raw data (abundance/occurrence). Still these techniques allow to detect changepoints for a single location (univariate scenario). Therefore, the data entered by users are considered as a single location data when analyse changepoints.

changepoint package:

The changepoint package implements various mainstream and specialised changepoint methods for finding single and multiple changepoints within data which includes many popular non-parametric and frequentist methods (Killick, Haynes and Eckley, 2016).

breakpoint package:

The breakpoint package implements variants of the Cross-Entropy (CE) method to estimate both the number and the corresponding locations of break-points in biological sequences of continuous and discrete measurements. The proposed method primarily built to detect multiple break-points in genomic sequences. However, it can be easily extended and applied to other problems (Priyadarshana and Sofronov, 2016).

cumSeg package:

The cumSeg package (Muggeo, 2010) estimates the of number and location of change points in mean-shift (piecewise constant) models which is useful to model genomic sequences of continuous measurements. The algorithm first estimates the highest number of change points using the efficient ‘segmented’ algorithm of Muggeo (2003) and then select some of them using a generalized BIC criterion by applying the lar’s algorithm of Efron et al. (2004) (Muggeo, 2010).

bcp package:

The bcp package provides an implementation of the Barry and Hartigan (1993) product partition model for the normal errors change point problem using Markov Chain Monte Carlo.

It also extends the methodology to regression models on a connected graph (Wang and Emerson, 2015) and allows estimation of change point models with multivariate responses (Erdman and Emerson, 2007).

PPMCA App

PPMCA 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, and to detect changepoints by using Bayesian changepoint techniques implemented in 'changepoint', 'breakpoint', 'cumSeg' and 'bcp'. 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 five 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) Persistence Probability model results page using occurrence data; 3) a maps page which display abundance plot and species distribution map; 4) an analysis page where statistical analyses are carried out, and summary results and posterior plots can be visualized; and 5) changepoint analysis page that user can choose a Bayesian changepoint method to carry out the changepoint analysis.

Shiny

Shiny is a web application framework for R that enables to build interactive web applications which includes a combination of R scripts: a user-interface script called `ui.R` and a server script called `server.R`. The user-interface script controls the layout and appearance of the application and the server script contains the R objects and the instructions about how they are displayed.

The functionality called reactivity in shiny applications have three kinds of objects: reactive sources, reactive conductors, and reactive endpoints. The source typically is user input through a browser interface. A reactive endpoint is usually something that appears in the user's browser window, such as a plot or a table of values. Reactive sources are accessible through the input object, and reactive endpoints are accessible through the output object. It's also possible to put reactive components in between the sources and endpoints. These components are called reactive conductors. A conductor can both be a dependent and have dependents. In other words, it can be both a parent and child in a graph of the reactive structure. Reactive conductors can be useful for encapsulating slow or computationally expensive operations. For example, imagine that you have this application that takes a value `input$n` and prints the n^{th} value in the Fibonacci sequence, as well as the inverse of n^{th} value in the sequence plus one (Shiny - Reactivity - An overview, 2020).

To create Shiny applications there is no web development experience required, although it is possible to use HTML, CSS, or JavaScript to achieve greater flexibility and customization. Shiny applications can be run locally by users that have the application files and R installed in their computer. Applications can also be hosted as a web page at its own URL and can be

navigated through the internet with a web browser. This facilitates its use to people without R knowledge (Moraga, 2015).

PPMCA app includes several control widgets, where users can interact with to send messages to the application. In the main window, we can upload the data files by using a file upload control which is created by the `fileInput` function. We also can specify the variables normalization options by selecting the appropriate names from boxes containing the possible choices. These boxes are created with the `selectInput` function. This function also available in species distribution model page and changepoint analysis page as well. In the species distribution models page there are `sliderInput` functions that allow to select a range of values by which we wish to filter the data. Species distribution model page and changepoint analysis page includes `numericInput` function which allows to enter a single number or a range to fit the model. Species distribution model page also includes `actionButton` function to build INLA mesh.

PPMCA includes three HTML widgets for interactive web data visualization such as Leaflet for rendering maps depicting the values of the variables of interest in each of the areas which support interactive panning and zooming which is very convenient when the map contains small areas (Cheng and Xie, 2016) and `ggplot2` (Wickham, 2009) to visualize the timeseries. `DataTable` is used to display interactive tables containing the information of the variables of interest (Xie, 2016). HTML widgets are created with JavaScript libraries and embedded in Shiny by using the HTML widgets package (Vaidyanathan et al., 2016). These two HTML widgets are included into the application by calling an output for the widget in the user-interface and assigning a render call to the output on the server side in the same way as `plotOutput` and `renderPlot` functions work. Leaflet uses `leafletOutput` and `renderLeaflet` and `DataTable` used `dataTableOutput` and `renderDataTable`.

PPMCA is dependent on several softwares and R packages including R-INLA (Lindgren and Rue, 2015) for performing Bayesian inference; `ggplot2` (Wickham, 2009), `gridExtra` and `lattice` for data visualisations; `bcp`, `changepoint`, `cumSeg` and `breakpoint` for changepoint analysis; `DT` to create data tables. Information about all the dependencies is shown in Table 1.

Table 1: Softwares and R packages used for developing PPMCA

Name		Description
Software		
R	(R Core Team, 2017)	Language and environment for statistical computing and graphics.
R Packages		
dplyr	(Wickham and Francois, 2016)	A fast, consistent tool for working with data frame like objects, both in memory and out of memory.
ggplot2	(Wickham, 2009)	Creates elegant data visualisations using the grammar of graphics.
htmlwidgets	(Vaidyanathan <i>et al.</i> , 2016)	Provides a framework for easily creating R bindings to JavaScript libraries.
leaflet	(Cheng and Xie, 2016)	Create Interactive Web Maps with the JavaScript 'Leaflet' Library.
R-INLA	(Lindgren and Rue, 2015)	Performs full Bayesian analysis on generalised additive mixed models using Integrated Nested Laplace Approximations.
shiny	(Chang <i>et al.</i> , 2016)	Web Application Framework for R.

changepoint	(Killick, Haynes and Eckley, 2016)	Bayesian changepoint analysis techniques.
breakpoint	(Priyadarshana and Sofronov, 2016)	
bcp	(Erdman and Emerson, 2007)	
cumSeg	(Muggeo, 2010)	
DT	(Xie, 2016)	Create data tables.
zoo	(Zeileis and Grothendieck, 2005)	Independence of a particular index/date/time class and consistency with timeseries and base R by providing methods to extend standard generics.
lars	(Hastie and Efron, 2013)	Efficient procedures for fitting an entire lasso sequence with the cost of a single least squares fit. Least angle regression and infinitesimal forward stagewise regression are related to the lasso.
foreach	(Ooi, Microsoft and Weston, 2020)	Support for the foreach looping construct. Foreach is an idiom that allows for iterating over elements in a collection, without the use of an explicit loop counter. This package in particular is intended to be used for its return value, rather than for its side effects. In that sense, it is similar to the standard lapply function, but doesn't require the evaluation of a function. Using foreach without side effects also facilitates executing the loop in parallel.
Matrix	(Bates et al., 2019)	A rich hierarchy of matrix classes, including triangular, symmetric, and diagonal matrices, both dense and sparse and with pattern, logical and numeric entries.
sp	(Pebesma et al., 2020)	Classes and methods for spatial data; the classes document where the spatial location information resides, for 2D or 3D data. Utility functions are provided, e.g. for plotting data as maps, spatial selection, as well as methods for retrieving coordinates, for subsetting, print, summary, etc.
lattice	(Sarkar et al., 2020)	A powerful and elegant high-level data visualization system inspired by Trellis graphics, with an emphasis on multivariate data. Lattice is sufficient for typical graphics needs and is also flexible enough to handle most nonstandard requirements. See Lattice for an introduction.
gridExtra	(Auguie and Antonov, 2017)	Provides a number of user-level functions to work with "grid" graphics, notably to arrange multiple grid-based plots on a page and draw tables.
plyr	(Wickham, 2020)	A set of tools that solves a common set of problems: you need to break a big problem down into manageable pieces, operate on each piece and then put all the pieces back together. For example, you might want to fit a model to each spatial location or time point in your study, summarise data by panels or collapse high-dimensional arrays to simpler summary statistics.

App pages:

Main window:

We see this main window including five pages when we launch the application. In this window we can upload the required files and select the normalization option for these uploaded data. It is composed of five components: 1) Choose data CSV file; 2) Choose explanatory variables CSV file; 3) Count data normalization 4) Predictors normalization; and 4) five pages (see Figure 1).

Single Species Persistence Models and Changepoint Analysis

Choose data CSV File

Browse... H12M.csv

Upload complete

Choose explanatory variables CSV File

Browse... Predictors_HPMP.csv

Upload complete

Count data normalization: none

Predictors normalization: none

Annual Data and Predictors

Persistence Probability Model

Abundance Plots

Single-species Joint Model

Changepoint Analysis

Annual Count Data

Show 10 entries

Search:

ID	Locality	Latitude	Longitude	Year	Count
1	Boags Rocks-2010	Boags Rocks	-38.43925	144.84773	2010
2	Boags Rocks-2011	Boags Rocks	-38.43925	144.84773	2011
3	Boags Rocks-2012	Boags Rocks	-38.43925	144.84773	2012
4	Boags Rocks-2013	Boags Rocks	-38.43925	144.84773	2013
5	Boags Rocks-2014	Boags Rocks	-38.43925	144.84773	2014
6	Boags Rocks-2015	Boags Rocks	-38.43925	144.84773	2015
7	Boags Rocks-2016	Boags Rocks	-38.43925	144.84773	2016
8	Boags Rocks-2017	Boags Rocks	-38.43925	144.84773	2017
9	Boags Rocks-2018	Boags Rocks	-38.43925	144.84773	2018
10	Flinders-2010	Flinders	-38.4759	145.02021	2010

Showing 1 to 10 of 108 entries

Previous12345...11Next

Annual Predictor Data

Show 10 entries

Search:

	T. temp	rainfall	population
1	14.75	1019.8	148907
2	14.6	985.2	150436
3	14.65	1030.4	152680
4	15	952.6	154977
5	15.15	734.2	157324
6	14.6	663.3	159991
7	15.05	808.8	162681
8	14.85	866.4	164828
9	15	705.5	165728

Showing 1 to 9 of 9 entries

Previous1Next

Figure 1. Main window and the first page

Page 1: Annual Data and Predictors

This page displays the annual data table of the uploaded data CSV file and the annual predictors table uploaded as explanatory variables CSV file (see Figure 1). This app only applicable for annual data. The data CSV file should include the columns Locality, Latitude, Longitude, Year and Count. These names are case sensitive. Annual predictors should enter as a separate dataframe. Possible maximum number of predictors is 5.

Page 2: Persistence Probability Model

This page displays the Persistence Probability model results for each year as a table and a separate table with persistence probabilities for the most recent years for each location (see Figure 2).

Single Species Persistence Models and Changepoint Analysis

Choose data CSV File

Browse... HPdata.csv

Upload complete

Choose explanatory variables CSV File

Browse... Predictors.csv

Upload complete

Count data normalization: none

Predictors normalization: none

Annual Data and Predictors

Persistence Probability Model

Abundance Plots

Single-species Joint Model

Changepoint Analysis

Persistence Probabilities for each year

Show 10 entries

Search:

Locality	Year	Count	Occ	P	Conclusion
1	Boags Rocks	2010	0	0.50000	More likely to extant
2	Boags Rocks	2011	0	0.50000	More likely to extant
3	Boags Rocks	2012	4	1 NaN	
4	Boags Rocks	2013	9	1 0.25000	More likely to extant
5	Boags Rocks	2014	0	0.40000	More likely to extant
6	Boags Rocks	2015	0	0.33333	More likely to extant
7	Boags Rocks	2016	0	0.28571	More likely to extant
8	Boags Rocks	2017	0	0.25000	More likely to extant
9	Boags Rocks	2018	5	1 0.09412	More likely to extant
10	Flinders	2010	0	0.50000	More likely to extant

Showing 1 to 10 of 108 entries

Previous12345...11Next

Persistence Probabilities for the most recent year

Show 10 entries

Search:

Locality	Year	Count	Occ	P	Conclusion
9	Boags Rocks	2018	5	1 0.09412	More likely to extant
18	Flinders	2018	0	0.20000	More likely to extant
27	Flinders Foreshore	2018	0	0.20000	More likely to extant
36	Flinders Ocean Beach	2018	24	1 0.83631	More likely to extant
45	Gunnamatta Beach	2018	272	1 0.83631	More likely to extant
54	Koonya Beach	2018	6	1 0.13972	More likely to extant
63	Moana Court	2018	0	0.22404	More likely to extant
72	Mushroom Reef Marine Sanctuary	2018	0	0.20000	More likely to extant
81	Point Nepean National Park	2018	1	1 0.21838	More likely to extant
90	Portsea Surf Beach	2018	4	1 0.36606	More likely to extant

Showing 1 to 10 of 12 entries

Previous12Next

Figure 2. Persistence Probability Model page

Page 3: Abundance Plots

This page visualizes the abundance plot and the species distribution plot (see Figure 3).

Single Species Persistence Models and Changepoint Analysis

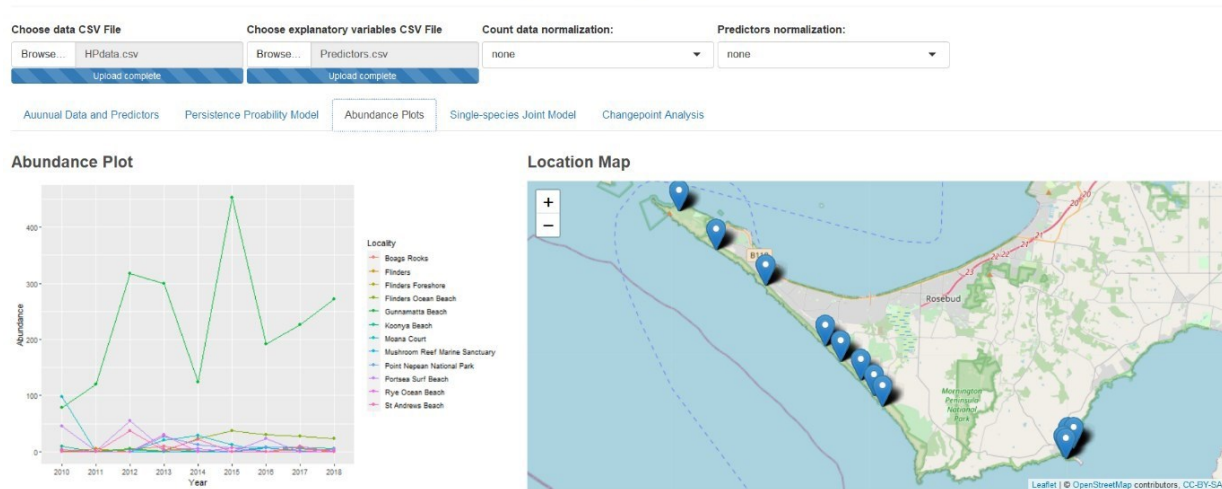


Figure 3. Abundance Plots page

Page 4: Single-species Joint Model

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 (see Figure 4).

Page 5: Changepoint Analysis

This page gives the option to fit Bayesian changepoints for annual raw data. Since still developed univariate changepoint methods, app will consider the annual data CSV as a single located dataset. Four Bayesian changepoint packages (changepoint, breakpoint, cumSeg and bcp) could be used to find the significant changes. This page also visualizes the relevant changepoint profile plots. For bcp package user has the option to use predictor variables to analyse. As the algorithms stated in the page, user can change the relevant parameters for each method and find the best scenario (see Figure 5).

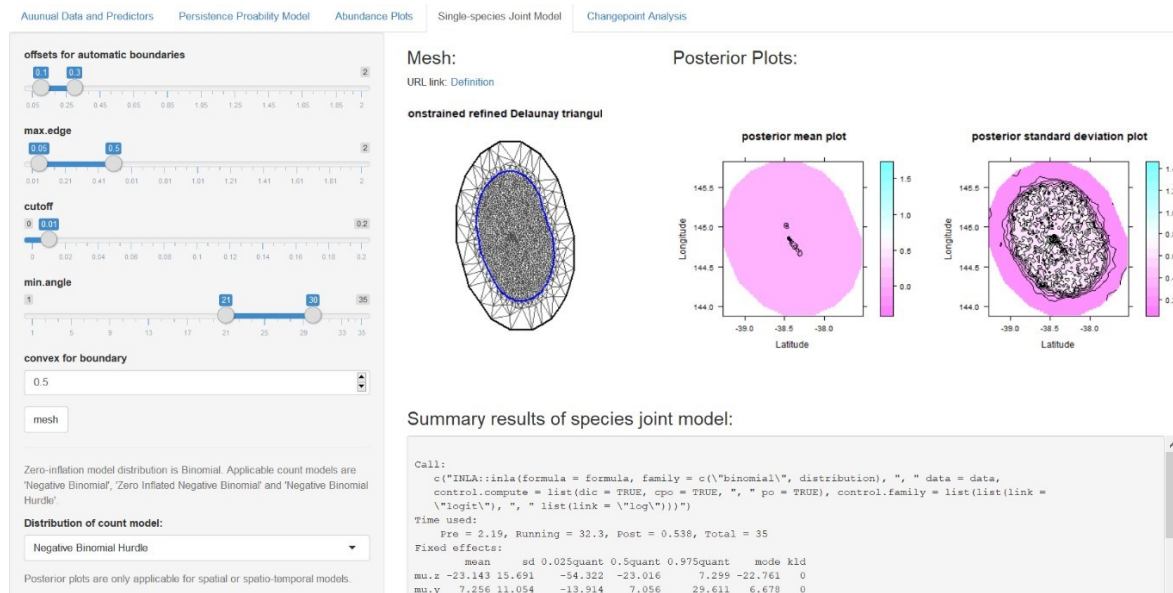


Figure 4. Single-species Joint Model page

Single Species Persistence Models and Changepoint Analysis

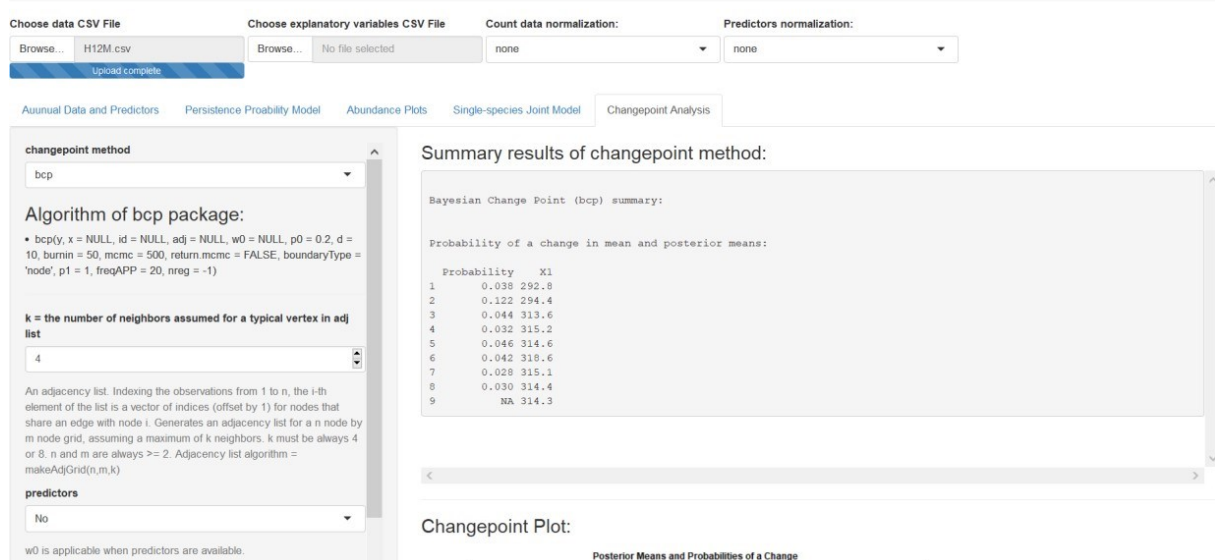


Figure 5. Changepoint Analysis page

Installation:

Users can launch the application in R by shiny: : runGitHub("PPMCA", "uwijewardhana"). The application uses the R-INLA package which can be downloaded from <http://www.r-inla.org>. Users can access the GitHub repository by <https://github.com/uwijewardhana/PPMCA>.

Conclusion

In this paper we presented PPMCA, a Shiny web application for Persistence models and changepoint analysis. The application is easy to use and allows ecologists and biostatisticians to analyse without the need of having advanced statistical or programming skills. Specifically, it allows to fit Bayesian zero inflation joint models with R-INLA. It also serves as an exploratory tool for abundance data since it enables interactive data visualization by using the

R packages Leaflet for rendering maps, ggplot2 for plotting time series, and DataTables for displaying data objects. PPMCA is an open-source tool, which is implemented using R, Shiny and incorporating the functions from several R packages and statistical softwares.

Although the statistical methods PPMCA can perform are often used in ecological modelling and they are also restrictive. Joint models can be used to determine the spatial, temporal or spatio-temporal models using Negative Binomial zero inflation distributions. Moreover, the application can be easily extended and in future versions we will increase its flexibility enabling more options as well as custom data visualizations.

References

1. Aminikhanghahi S, Cook DJ. (2017) A Survey of Methods for Time Series Change Point Detection. *Knowl Inf Syst.* 2017;51(2):339–367. doi:10.1007/s10115-016-0987-z
2. Shiny.rstudio.com. 2020. Shiny - Reactivity - An Overview. [online] Available at: <https://shiny.rstudio.com/articles/reactivity-overview.html> [Accessed 29 April 2020].
3. Moraga, P., 2020. A Shiny Web Application for The Analysis Of Spatial Andspatio-Temporal Disease Data [R Package Spatialepiapp Version 0.3]. [online] Cran.r-project.org. Available at: <https://cran.r-project.org/web/packages/SpatialEpiApp/index.html> [Accessed 29 April 2020].
4. Zeileis A, Grothendieck G (2005). “zoo: S3 Infrastructure for Regular and Irregular Time Series.” *Journal of Statistical Software*, **14**(6), 1–27. doi: [10.18637/jss.v014.i06](https://doi.org/10.18637/jss.v014.i06).
5. Region 4 Southeast Region. (2019). Why do We Care About Endangered Species in North Carolina?. [online] A available at: <https://www.fws.gov/nc-es/es/whocares.html>. [Accessed 24 May. 2019].
6. Xie, H. et.al., 2013, “Comparing statistical methods for analyzing skewed lonitudinal count data with many zero
7. s: An example of smoking cessation”, *Journal of Substance Abuse Treatment*, **45**, 99-108.
8. Ree, R. and McCarthy, M. (2005). Inferring persistence of indigenous mammals in response to urbanisation. *Ani mal Conservation*, **8**(3), pp.309-319.
9. 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.
10. Cragg, J.G. Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica* 1971, **39**, 829–844.
11. Cameron, A.C., Trivedi P.K. *Regression Analysis of Count Data*; Cambridge University Press: Cambridge, UK, 1998.
12. 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.
13. 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.
14. 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.
15. 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>

16. Killick R, Haynes K and Eckley IA (2016). *changepoint: An R package for changepoint analysis*. R package version 2.2.2, <https://CRAN.R-project.org/package=changepoint>.
17. Priyadarshana, W. J. R. M., and G. Sofronov. (2016). breakpoint: An R Package for Multiple Break-Point Detection via the Cross-Entropy Method. *CRAN Repository* (2016): 1-25.
18. Muggeo, V. M. R., and G. Adelfio. (2011). Efficient change point detection for genomic sequences of continuous measurements. *Bioinformatics* 27:161-166.
<http://dx.doi.org/10.1093/bioinformatics/btq647>
19. (2) (PDF) *Threshold detection: Matching statistical methodology to ecological questions and conservation planning objectives*. Available from: https://www.researchgate.net/publication/273777417_Threshold_detection_Matching_statistical_methodology_to_ecological_questions_and_conservation_planning_objectives.
20. Muggeo, V.M.R. (2003), Estimating regression models with unknown break-points. *Statist. Med.*, 22: 3055-3071. doi:[10.1002/sim.1545](https://doi.org/10.1002/sim.1545)
21. Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Leastangle regression. *Annals of Statistics*, 32, 407– 489.
22. Barry, D. and Hartigan, J.A. (1993) A Bayesian analysis for change point problems. *J. Am. Stat. Assoc.*, 88, 309–319.
23. Chandra Erdman, John W. Emerson (2007), bcp: An R Package for Performing a Bayesian Analysis of Change Point Problems, *Journal of Statistical Software*, 23(3), 1-13, URL <http://www.jstatsoft.org/v23/i03/>.
24. Xiaofei Wang and John W. Emerson (2015), Bayesian Change Point Analysis of Linear Models on General Graphs, Working Paper.
25. Moraga P, Cano J, Baggaley RF, Gyapong JO, Njengaf SM, Nikolay B, Davies E, Rebollo MP, Pullan RL, Bockarie MJ, Hollingsworth TD, Gambhir M, Brooker SJ (2015). “Modelling the distribution and transmission intensity of lymphatic filariasis in sub-Saharan Africa prior to scaling up interventions: integrated use of geostatistical and mathematical modelling.” *Parasites & Vectors*, 8, 560.
26. Shiny.rstudio.com. 2020. *Shiny - Reactivity - An Overview*. [online] Available at: <https://shiny.rstudio.com/articles/reactivity-overview.html> [Accessed 30 April 2020].
27. Cheng J, Xie Y (2016). leaflet: *Create Interactive Web Maps with the JavaScript 'Leaflet' Library*. R package version 1.0.1, URL <https://CRAN.R-project.org/package=leaflet>.
28. Wickham H (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-0-387- 98140-6. URL <http://ggplot2.org>.
29. Xie Y (2016). DT: *A Wrapper of the JavaScript Library 'DataTables'*. R package version 0.2, URL <https://CRAN.R-project.org/package=DT>.
30. Vaidyanathan R, Xie Y, Allaire J, Cheng J, Russell K (2016). htmlwidgets: *HTML Widgets for R*. R package version 0.7, URL <https://CRAN.R-project.org/package=htmlwidgets>.
31. Lindgren F, Rue H (2015). “Bayesian Spatial Modelling with R-INLA.” *Journal of Statistical Software*, 63(19), 1–25.
32. R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org>.
33. Wickham H, Francois R (2016). dplyr: *A Grammar of Data Manipulation*. R package version 0.5.0, URL <https://CRAN.R-project.org/package=dplyr>.

34. Wickham H (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-0-387-98140-6. URL <http://ggplot2.org>.
35. Vaidyanathan R, Xie Y, Allaire J, Cheng J, Russell K (2016). *htmlwidgets: HTML Widgets for R*. R package version 0.7, URL <https://CRAN.R-project.org/package=htmlwidgets>.
36. Cheng J, Xie Y (2016). *leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library*. R package version 1.0.1, URL <https://CRAN.R-project.org/package=leaflet>.
37. ChangW, Cheng J, Allaire J, Xie Y, McPherson J (2016). *shiny: Web Application Framework for R*. R package version 0.14.1, URL <https://CRAN.R-project.org/package=shiny>.
38. Zeileis A, Grothendieck G (2005). "zoo: S3 Infrastructure for Regular and Irregular Time Series." *Journal of Statistical Software*, **14**(6), 1–27. doi: [10.18637/jss.v014.i06](https://doi.org/10.18637/jss.v014.i06).
39. Xie Y (2016). *DT: A Wrapper of the JavaScript Library 'DataTables'*. R package version 0.2, URL <https://CRAN.R-project.org/package=DT>.
40. Microsoft and Steve Weston (2020). *foreach: Provides Foreach Looping Construct*. R package version 1.4.8. <https://CRAN.R-project.org/package=foreach>
41. Douglas Bates and Martin Maechler (2018). *Matrix: Sparse and Dense Matrix Classes and Methods*. R package version 1.2-15. <https://CRAN.R-project.org/package=Matrix>
42. Pebesma, E.J., R.S. Bivand, 2005. Classes and methods for spatial data in R. *R News* 5 (2), <https://cran.r-project.org/doc/Rnews/>.
43. Sarkar, Deepayan (2008) *Lattice: Multivariate Data Visualization with R*. Springer, New York. ISBN 978-0-387-75968-5.
44. Baptiste Auguie (2017). *gridExtra: Miscellaneous Functions for "Grid" Graphics*. R package version 2.3. <https://CRAN.R-project.org/package=gridExtra>
45. Hadley Wickham (2011). The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software*, **40**(1), 1-29. URL <http://www.jstatsoft.org/v40/i01/>.