

Shiny web application for spatial and spatio-temporal areal count data analysis

Aritz Adin

aritz.adin@unavarra.es

Jaione Etxeberria

jaione.etxeberria@unavarra.es

Department of Statistics, Computer Science and Mathematics
Public University of Navarre, Spain



September 7, 2021

upna

Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa

1 Introduction

- 1.1 An introduction to disease mapping
- 1.2 Basic concepts
- 1.3 Modelling
- 1.4 Model fitting and inference

2 The SSTCDapp application

- 2.1 Register and login

3 Description and functionalities

- 3.1 Data Input
- 3.2 Descriptive Analysis
- 3.3 Model Specification
- 3.4 Results

4 Tutorials

5 Let's start using the app

Disease mapping: a historical perspective

- Disease mapping may be defined as the estimation and representation of area summary measures of health outcomes (Wakefield et al., 2000)
- The work of Doctor John Snow (1855) is probably one of the earliest and most famous cases in disease mapping.

Disease mapping: a historical perspective

- The work of Doctor **John Snow (1855)** is probably one of the earliest and most famous cases of maps being used to understand the spread of a disease.
- In the nineteenth century, there were several outbreaks of cholera in London.
- In the 1849 outbreak, approximately 700 people died in the Soho district in less than a week, in an area of just half a kilometer in diameter.

Disease mapping: a historical perspective



The map drawn by Dr. John Snow showing deaths of cholera (dots) in the London epidemics of 1854 together with the locations of water pumps (crosses). Source: Wikipedia (Last access: July 2021)

Disease mapping: a historical perspective

- Interest of mapping infectious disease, such as cholera, influenza or yellow fever, continued during the in the 20th century.
- In the twentieth century, began to produce also many **chronic disease atlases (mainly cancer mortality maps)** in national or smaller administrative units.
- The United kingdom and the United States were the precursors, and the atlases published in those countries demonstrate some evolution in methodology.
- The representation and analysis of disease incidence or mortality has been established as a basic tool for the analysis of regional public health data

Disease mapping. Background



- Observed incident/death cases in each geographic area.
- Incorporate population data in maps: *rates or risks*
- To take into account the age-structure to allow comparison between different areas: **age-standardization**.



The direct or indirect method is used in disease mapping:
Age Standardized Rates or **SMR=Observed/Expected**

Disease mapping. Background

Some examples

- Example for COVID-19.

See <https://cnecovid.isciii.es/covid19/>

- Example for different types of cancer.

See <http://ariadna.cne.isciii.es/MapaP/>

Disease mapping. Background

- When a rare disease (with a few number of cases) or areas containing small numbers of people at risk are studied, the age standardized rates or SMR's will be too variable or unstable and, then not reliable.
- To deal with this situation it is usual to use *sophisticated* statistical models that borrow strength from neighbouring areas
- These models usually include random effects (structured and/or unstructured) smoothing risks in low populated regions
- The more used models are hierarchical models, in particular mixed Poisson models

Disease mapping. Context of study

Let us consider that the interest lies in estimating the relative risk r_i of mortality/incidence of a disease in area i .

Conditional on these risks, the number of counts O_i is assumed to be Poisson distributed with mean $\mu_i = e_i r_i$. That is,

$$O_i \mid r_i \sim \text{Poisson}(\mu_i = e_i r_i) \text{ for } i = 1, \dots, n,$$

$$\log \mu_i = \log e_i + \log r_i$$

Here, $\log e_i$ is an offset and depending on the specification of $\log r_i$ different models are defined.

Disease mapping. Context of study

Most of the spatial disease mapping models in the literature are based on conditional autoregressive (CAR) prior distributions. In the simplest model the log-risk is modelled as

$$\log r_i = \eta + \xi_i$$

- η is an intercept representing an overall level of risk
- $\xi = (\xi_1, \dots, \xi_n)'$ is a spatially structured random effect

Spatial models for disease mapping

Intrinsic CAR model

A simplified version of the intrinsic conditional autoregressive (iCAR) prior distribution ([Besag, 1974](#)), commonly used in disease mapping is defined as

$$\xi \sim N\left(0, [\tau_\xi \mathbf{R}_s]^- \right)$$

where

- τ_ξ is a precision parameter
- \mathbf{R}_s is the $n \times n$ spatial neighborhood matrix with diagonal elements equal to the number of neighbors of each area and non-diagonal elements $(\mathbf{R}_s)_{ij} = -1$ if areas i and j are neighbors and $(\mathbf{R}_s)_{ij} = 0$ otherwise.

Here, two areas are considered as neighbors if they share a common border.

Spatial models for disease mapping

BYM model

- The iCAR prior distribution only represent strong spatial correlation structures, and hence, is not appropriate if the data are weakly correlated.
- Besag et al. (1991) also proposed a model (hereafter BYM model) which includes two spatial random effects: one assuming an iCAR prior for the spatially structured variability and another one assuming a Gaussian exchangeable prior to model unstructured heterogeneity.

$$\xi = \mathbf{u} + \mathbf{v}; \quad \text{with} \quad \mathbf{u} \sim N\left(0, [\tau_u \mathbf{R}_s]^{-}\right) \text{ and, } \mathbf{v} \sim N\left(0, \tau_v^{-1} \mathbf{I}_n\right)$$

\mathbf{I}_n is an identity matrix of dimension $n \times n$.

Spatial models for disease mapping

Leroux model

Leroux et al. (1999) proposed an alternative formulation to model both spatially unstructured and structured variation in a single set of random effects (hereafter LCAR model), which is given by

$$\boldsymbol{\xi} \sim N\left(0, [\tau_{\xi} (\lambda_{\xi} \mathbf{R}_s + (1 - \lambda_{\xi}) \mathbf{I}_n)]^{-1}\right)$$

where τ_{ξ} is a precision parameter and λ_{ξ} is a spatial smoothing parameter taking values between 0 and 1 .

- $\lambda_{\xi} = 0$ corresponds to the unstructured prior $\boldsymbol{\xi} \sim N\left(0, \tau_{\xi}^{-1} \mathbf{I}_n\right)$,
- $\lambda_{\xi} = 1$ corresponds to the iCAR prior $\boldsymbol{\xi} \sim N\left(0, [\tau_{\xi} \mathbf{R}_s]^{-1}\right)$.

Spatial models for disease mapping

A modified BYM model

In [Riebler et al. \(2016\)](#) a modification of the [Dean et al. \(2001\)](#) model is proposed which addresses both the identifiability and scaling issue of the BYM model, hereafter BYM2 model.

$$\xi = \frac{1}{\sqrt{\tau_\xi}} \left(\sqrt{\lambda_\xi} \mathbf{u}_* + \sqrt{1 - \lambda_\xi} \mathbf{v} \right)$$

where \mathbf{u}_* is the scaled intrinsic CAR model with generalized variance equal to one and \mathbf{v} is the unstructured random effect. The variance of the random effect is expressed as a weighted average of the covariance matrices of the structured and unstructured spatial components (unlike the LCAR model which considers a weighted combination of the precision matrices), i.e.,

$$\text{Var}(\xi | \tau_\xi) = \frac{1}{\tau_-} (\lambda_\xi \mathbf{R}_*^- + (1 - \lambda_\xi) \mathbf{I}_n)$$

where \mathbf{R}_*^- indicates the generalised inverse of the scaled spatial precision matrix ([Sørbye and Rue, 2014](#)). As in the previous models, a sum-to-zero constraint $\sum_{i=1}^n \xi_i = 0$ must be imposed to avoid identifiability problems.

Spatio-temporal models for disease mapping

Spatio-temporal models for disease mapping

If the interest lies in estimating the relative mortality/incidence risk of a disease in area i and time t , then

$$O_{it} \mid r_{it} \sim \text{Poisson}(\mu_{it} = e_{it} r_{it}) \quad \text{for} \quad i = 1, \dots, n; \quad t = 1, \dots, T,$$

$$\log \mu_{it} = \log e_{it} + \log r_{it}$$

Spatio-temporal models for disease mapping

The log-risks are modelled as

$$\log r_{it} = \eta + \xi_i + \phi_t + \gamma_t + \delta_{it}$$

- η is an intercept representing an overall level of risk
- ξ_i is the spatial component
- ϕ_t is the unstructured temporal effect where $\phi = (\phi_1, \dots, \phi_T)' \sim N(0, \tau_\phi^{-1} \mathbf{I}_T)$
- γ_t is the structured temporal effects where $\gamma = (\gamma_1, \dots, \gamma_T)' \sim N(0, [\tau_\gamma \mathbf{R}_t]^-)$ and \mathbf{R}_t is the $T \times T$ structure matrix of a RW1/RW2

Spatio-temporal models for disease mapping

The log-risks are modelled as

$$\log r_{it} = \eta + \xi_i + \phi_t + \gamma_t + \delta_{it}$$

- δ_{it} is the space-time interaction effect where

$$\delta = (\delta_{11}, \dots, \delta_{1T}, \dots, \delta_{n1}, \dots, \delta_{nT})' \sim N(\mathbf{0}, [\tau_\delta \mathbf{R}_\delta]^-)$$

Here, τ_δ is a precision parameter and \mathbf{R}_δ is the $nT \times nT$ matrix obtained as the Kronecker product of the corresponding spatial and temporal structure matrices, where four types of interactions can be considered

Interaction	\mathbf{R}_δ	Spatial correlation	Temporal correlation
Type I	$\mathbf{I}_n \otimes \mathbf{I}_T$	—	—
Type II	$\mathbf{I}_n \otimes \mathbf{R}_t$	—	✓
Type III	$\mathbf{R}_s \otimes \mathbf{I}_T$	✓	—
Type IV	$\mathbf{R}_s \otimes \mathbf{R}_t$	✓	✓

Model fitting and inference

- Model fitting, inference and prediction was carried out using Bayesian methodology, specifically, integrated nested Laplace approximations (INLA). See [Rue et al. \(2009\)](#) for more detail
 - The technique is implemented in the software R through the package R-INLA
 - The models are implemented in a local server...
 - The group of best models can be selected among the different proposals based on different model selection criteria
- See [SSTCDapp User Guide](#) for more detail

Shiny application for the analysis of spatial and spatio-temporal count data: SSTCDapp

The SSTCDapp application

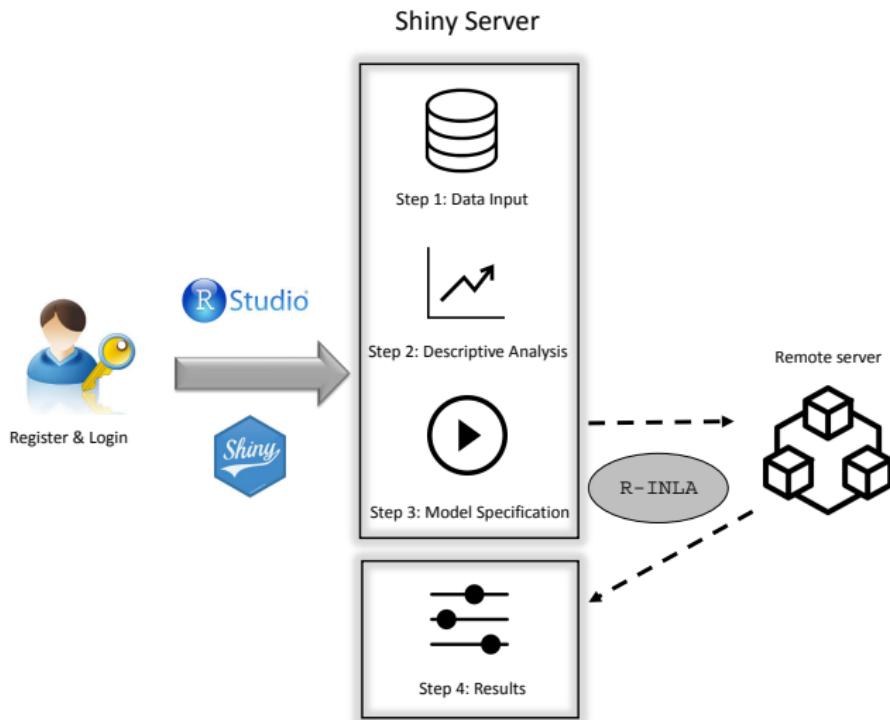
- **SSTCDapp** is an interactive web application for the analysis of spatial and spatio-temporal count data, with a particular focus on the field of disease mapping.
- It is designed for two main purposes:
 - To perform descriptive analyses in space and time of mortality/incidence risks or rates.
 - To fit a wide variety of spatial and spatio-temporal hierarchical models for areal data.
- It facilitates the use of fairly complex spatio-temporal disease mapping models using R-INLA for users in many areas, including epidemiologists and public health researchers, as well as providing tools that are useful for a detailed analysis of the model results.
- The application may also be used for the analysis of similar problems in other fields such as criminology, gender-based violence, road-traffic accidents or veterinary.

The SSTCDapp application

Main characteristics:

- It has been developed with **shiny** (Chang et al., 2018), a package to build interactive web applications in the R software environment.
- The recently proposed integrated nested Laplace approximation (INLA, Rue et al., 2009) technique for Bayesian inference is used for model fitting through the **R-INLA** package.
- Users can access the application directly from the web browser at <http://www.unavarra.es/spatial-statistics-group/> without installing any software in their computers, since all the analysis and computations are made in a remote server.
- The R code can be downloaded to run INLA locally if needed.

The SSTCDapp application



Register and login

A password is required to use the application.

- Every new user must register for the first time and a password will be sent to the users e-mail address.
- Once logged in the application using their username and password, the user will be able to submit a model on a remote server and collect the results when the computations are finished.

Register and login



Shiny application for the analysis of spatial and spatio-temporal count data:
SSTCDapp

upna
Universidad Pública de Navarra
Universitat Pública del País Vasco

Register:

A password is required to use the SSTCDapp application.
Please fill out all the fields in the following form and click on submit. You will then receive an email with the password.

First name Last name E-mail address

Country Institution

Request a password

Login:

If you are already registered, please insert your user name and password.

User name Password

Login

Description and functionalities

The application is structured into four main parts organized in tabs.

① Data Input:

- The data and the associated cartography are uploaded by the user, and automatically previewed on the screen.
- Several formats for both data and cartography are supported.

② Descriptive Analysis:

- The target variables are selected and standardized mortality ratios (SMR) or standardized rates (SR) are calculated.
- Descriptive graphics of the spatial, temporal, and spatio-temporal distribution for the variables of interest (crude rates, SMR or SR) are generated.

Description and functionalities

③ Model Specification:

- A wide variety of spatial or spatio-temporal models commonly used in disease mapping can be fitted using the R-INLA package.
- The model is submitted to a remote server. Once the calculations are finished the user will receive a notification by email.

④ Results:

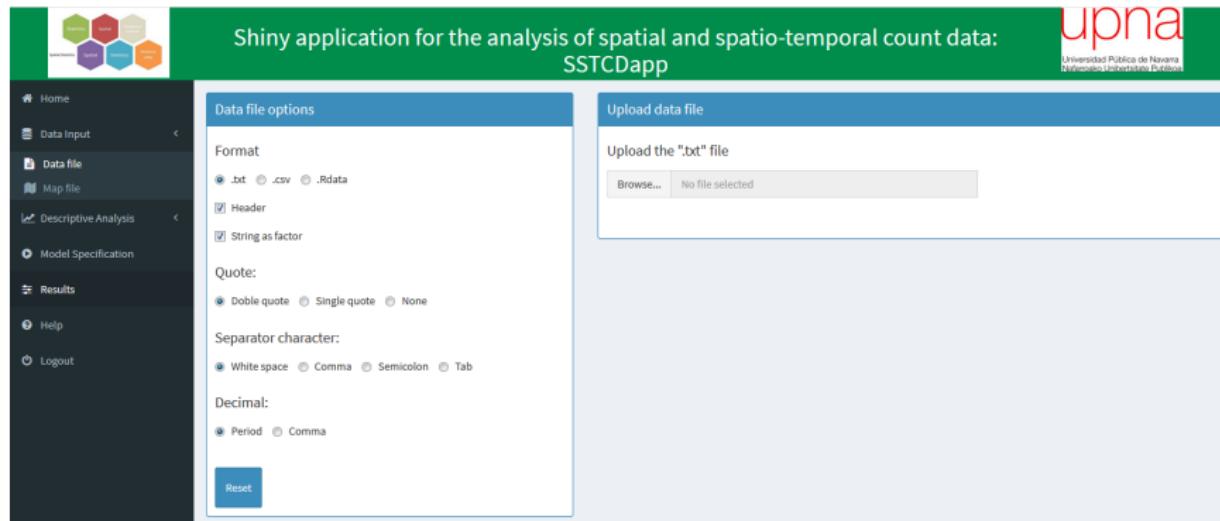
- Summary measures are provided for the posterior distribution of model hyperparameters; relative mortality/incidence risks (or rates); and spatial, temporal, and spatio-temporal patterns.
- Maps with the geographical distribution of the disease risks and area-specific temporal evolutions are also generated.

Data Input

Data file:

- The input data should be provided in a single file.
- Several formats are supported
 - ".txt" or ".csv" plain text formats (with cases corresponding to rows and variables to columns in the file).
 - ".Rdata" file containing a single dataframe.
- The file must contain at least
 - The names or IDs of each area and time period (the latter only if spatio-temporal data are analyzed).
 - The observed number of cases.
 - The population at risk or the number of expected cases.

Data Input



The screenshot shows the SSTCDapp Shiny application interface. At the top, there is a green header bar with the title "Shiny application for the analysis of spatial and spatio-temporal count data: SSTCDapp". On the right side of the header is the UPNA logo (Universidad Pública de Navarra). Below the header, the main content area has two main sections: "Data file options" on the left and "Upload data file" on the right.

Data file options:

- Format:
 .txt .csv .Rdata
- Header
- String as factor

Quote:

- Doble quote Single quote None

Separator character:

- White space Comma Semicolon Tab

Decimal:

- Period Comma

Upload data file:

Upload the "txt" file

No file selected

On the far left, there is a vertical sidebar with a navigation menu:

- Home
- Data Input (selected)
- Map file
- Descriptive Analysis
- Model Specification
- Results
- Help
- Logout

Figure 3: Input data file (SSTCDapp).

Data Input

Map file:

- The cartography of the region under study associated to the input data can be also included (necessary to generate maps with the spatial or spatio-temporal distribution of the disease).
- Several cartography file formats are supported
 - “.Rdata” or “.rds” extensions containing an spatial data object of both `SpatialPolygon` or `SpatialPolygonDataFrame` classes provided by the `sp` package (see for example [Bivand et al., 2013](#)).
 - Commonly used shapefile formats are also supported (in this case, all the related files must be uploaded).
- The areas (polygons) in the cartography file must match those of the input data file.

Data Input

The screenshot shows the 'Data Input' screen of the SSTCDapp Shiny application. At the top, there is a green header bar with the text 'Shiny application for the analysis of spatial and spatio-temporal count data: SSTCDapp'. On the right side of the header is the UPNA logo (Universidad Pública de Navarra). Below the header is a navigation sidebar on the left containing links for Home, Data Input (selected), Data file, Map file, Descriptive Analysis, Model Specification, Results, Help, and Logout. The main content area is divided into two sections: 'Map file options' on the left and 'Upload map file' on the right. The 'Map file options' section includes a 'Format' dropdown with radio buttons for '.Rdata', '.rds', and '.shp' (with '.Rdata' selected), a 'Reset' button, and a link to 'URL link: GADM database of Global Administrative Areas'. The 'Upload map file' section has a 'Browse...' button and a text input field showing 'No file selected'. At the bottom of the main content area is a dropdown menu labeled 'Select the area variable in the map'.

Figure 4: Input map file (SSTCDapp)

Descriptive Analysis

- First, the target variables are selected and, if necessary, standardized mortality/incidence ratios or standardized rates are calculated.
- Then, the aggregated data is computed, with rows corresponding to areas in the case of spatial count data, or unique combinations of areas and time points if spatio-temporal data are analyzed.
- Finally, several graphs are generated for the descriptive analysis of the risks/rates.

Variable Selection

- The following variables must be selected from the input data file:
 - The variables with the names or IDs of the areas and time points (leave the latter blank for purely spatial analysis).
 - The observed number of cases.
 - The population at risk (mandatory if the number of expected cases has not been calculated).

The screenshot shows a user interface for variable selection. It consists of four separate input fields arranged in a 2x2 grid:

- Area:** A dropdown menu labeled "Select the area variable".
- Time:** A dropdown menu labeled "Select the time variable".
- Population:** A dropdown menu labeled "Select the population variable".
- Counts:** A dropdown menu labeled "Select the observed cases variable".

Variable Selection

- If the interest lies in analyzing relative risks, the offset of the model corresponds to the number of expected cases.
 - Select the variable from the input data file.
 - Compute using the indirect standardization method by selecting the auxiliary variable(s) (which usually correspond to age-groups).

Standardization method	Standardization method
<ul style="list-style-type: none"><input checked="" type="radio"/> Select expected cases variable from the data<input type="radio"/> Compute expected cases using indirect standardization<input type="radio"/> Compute mortality/incidence rates using direct standardization <p><i>Expected cases variable</i></p> <input type="text"/>	<ul style="list-style-type: none"><input type="radio"/> Select expected cases variable from the data<input checked="" type="radio"/> Compute expected cases using indirect standardization<input type="radio"/> Compute mortality/incidence rates using direct standardization <p><i>Auxiliary variables for indirect standardization</i></p> <input type="text"/>

Variable Selection

- **If the interest lies in analyzing rates**, the offset of the model corresponds to the population at risk.
 - The application includes the option to compute standardized rates using the direct standardization method.
 - By default, the age distribution of the 2013 European Standard Population is used to compute these rates.
 - An external standard population file can be also uploaded.

Standardization method

Select expected cases variable from the data
 Compute expected cases using indirect standardization
 Compute mortality/incidence rates using direct standardization

Auxiliary variables for direct standardization

Default option: European Standard Population (2013)

or load an external standard population file

Variable Selection

- When all these variables are selected, press the **Compute aggregated data** button before moving to the next tab.
- The generated data can be exported in several formats (“.txt”, “.csv” or “.Rdata”).

The screenshot shows the SSTCDapp Shiny application interface. The top navigation bar includes links for Introduction, The SSTCDapp application, Description and functionalities, Tutorials, Let's start using the app, References, Data Input, Descriptive Analysis, Model Specification, and Results. The main title is "Shiny application for the analysis of spatial and spatio-temporal count data: SSTCDapp". On the left, a sidebar menu lists Home, Data Input, Descriptive Analysis, Variable Selection (which is highlighted), Graphical Outputs, Model Specification, Results, Help, and Logout. The main content area is titled "Variable Selection". It contains four input fields: "Area" (Select the area variable), "Time" (Select the time variable), "Population" (Select the population variable), and "Counts" (Select the observed cases variable). To the right, there is a section for "Standardization method" with three radio buttons: "Select expected cases variable from the data" (selected), "Compute expected cases using indirect standardization", and "Compute mortality/incidence rates using direct standardization". Below these is a dropdown menu for "Expected cases variable". At the bottom of the page, there is a section titled "Aggregated data" containing a "Compute aggregated data" button, an "Export data" button with a file icon, and a file download section with ".txt", ".csv", and ".Rdata" options.

Figure 5: Variable selection (SSTCDapp).

Graphical outputs

- Descriptive graphics of the spatial, temporal, and spatio-temporal distribution for the variables of interest (crude rates, SMR or SR) are generated.
 - **Spatial distribution:** A map with the geographical distribution (for the whole study period) of the variable of interest is plotted.
 - **Temporal distribution:** A line chart with the temporal evolution (for the whole area) of the variable of interest is plotted.
 - **Spatial-temporal maps:** Maps with the geographical distribution for each time point are plotted.
- All the graphs can be exported through the [Download](#) button in several formats (“.pdf”, “.eps”, “.jpeg”, “.png” and “.bmp”).

Model Specification

- Different spatial and spatio-temporal models commonly used in disease mapping are available.
- **Spatial prior distribution**
 - iCAR, BYM, Leroux or BYM2 prior distributions can be selected for the spatial random effect.
 - If a cartography file has been uploaded, the spatial neighborhood matrix can be automatically computed from the map (two areas are considered as neighbors if they share a common border).
 - Alternatively, a file containing the neighborhood structure can be uploaded.
 - An option to check whether the neighborhood structure and the spatial regions are consistent has been included, which automatically shows a warning message if the spatial neighborhood structure is not a connected graph.

Model Specification

• Temporal prior distribution

- RW1 or RW2 prior distributions can be selected for the temporally structured random effect.
- An unstructured temporal random effect can be also included in the model.

• Spatio-temporal prior distribution

- None (additive model), TypeI, TypeII, TypeIII or TypeIV prior distribution can be selected for the space-time interaction random effect (see for example Knorr-Held, 2000 and Ugarte et al., 2014).
- See Goicoa et al. (2018) for details about identifiability constraints.

• INLA approximation strategy

- Select one of the following approximation strategies: Gaussian, simplified.laplace or laplace (see Rue et al., 2009 for details).

Model Specification

Select model options

R-INLA project website

Model name (optional)

Spatial prior distribution

ICAR BYM Leroux BYM2

Temporal prior distribution

RW1 RW2

Include temporally unstructured component

Spatio-temporal interaction

None Type I Type II Type III Type IV

INLA approximation strategy

Gaussian Simplified Laplace Full Laplace

Show/hide advanced options

Run INLA **Reset**

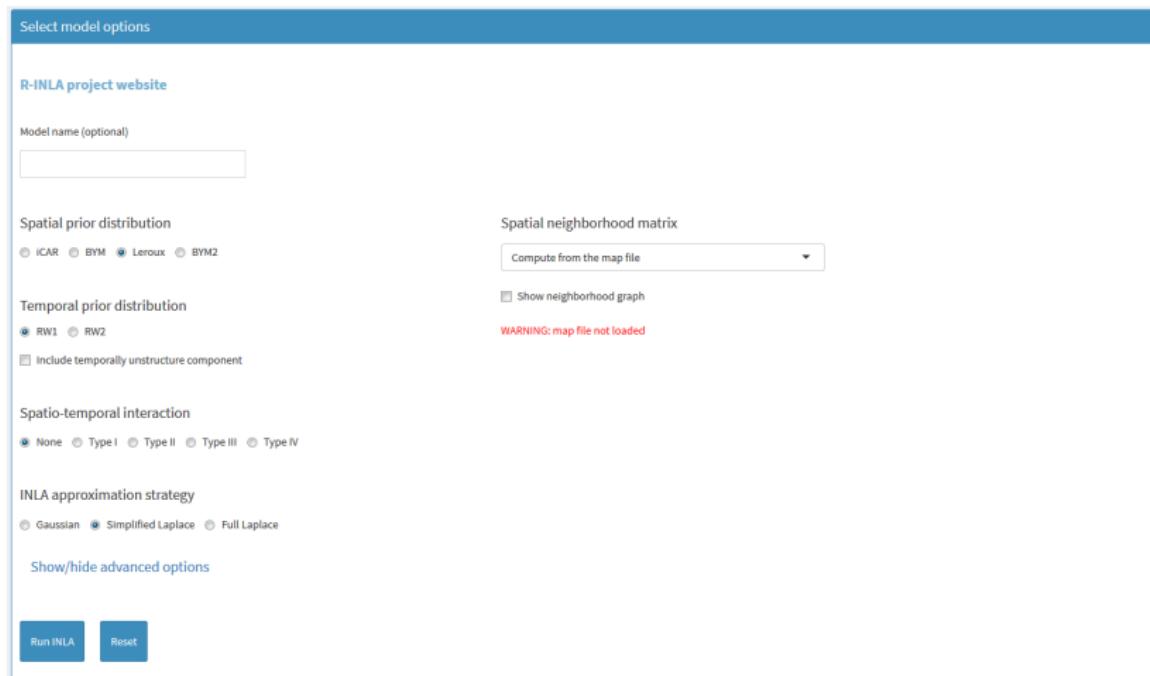


Figure 6: Model specification (SSTCDapp).

Advanced options

- Posterior distributions of the **spatial, temporal, and spatio-temporal patterns** can be computed, which are expressed as linear combinations of the estimated log-risks ([Adin et al., 2017](#)).
- A **battery of models** with different types of space-time interactions can be executed simultaneously.
- The **model scaling** option [Riebler et al. \(2016\)](#) is available for intrinsic GMRFs, using PC-priors ([Simpson et al., 2017](#)) for the precision parameters (iCAR and RW1/RW2 models) and the spatial smoothing parameter (BYM2 model).
- The **INLA integration strategy** to use: choose between the default INLA option ('auto'), the central composite design ('ccd'), the grid exploration ('grid') or the empirical Bayes ('eb') strategies (see [Rue et al., 2009](#) and [Martins et al., 2013](#) for details).
- **The R code to fit the models in INLA can be downloaded** to be run locally in the user's computer if needed.

Running INLA

- Press the **Run INLA** button to submit the selected model into a remote server.
- The user can view through the application the status of the models fitting at any time.
- Each model has a unique identifier, necessary to manage the model retrieve and/or delete options.
- A name for the model can be also specified by the user.
- **Important note:** The user will receive an email once the model calculations are finished, so they do not have to wait with the application open until the model has finished.

Retrieve/delete fitted models

- The models fitted on the remote server can be retrieved into the application (without uploading again the input data) through the **Import model(s)** button.
 - **If a single model is selected**, a summary is printed on the screen with information related to the INLA version, model name, computation time, posterior distributions of fixed effects and model hyperparameters, and deviance information criteria among others.
 - **If multiple models are selected**, a table is printed with the prior distributions selected for the spatial, temporal, and interaction random effects, the INLA approximation used and the integration strategy, as well as different model comparison measures.
- The models can be deleted from the remote server manually. Otherwise, they will be automatically removed 7 days after their execution has finished.
- **We recommend to save the imported models** as “.Rdata” files from the **Save ‘inla’ object** button.

Retrieve/delete fitted models

Fitted models

Show submitted jobs (click to refresh)

Job: 1	Id: Tue-Mar-20-09-04-10-2018--175344532	Size: 20Mb	Status: Finished
Job: 2	Id: Tue-Mar-20-09-04-14-2018---387089154	Size: 21Mb	Status: Finished
Job: 3	Id: Tue-Mar-20-09-04-16-2018---997170988	Size: 14Mb	Status: Running(00:00:29)
Job: 4	Id: Tue-Mar-20-09-04-18-2018---805679852	Size: 14Mb	Status: Running(00:00:44)
Job: 5	Id: Tue-Mar-20-09-04-20-2018---925371128	Size: 15Mb	Status: Running(00:00:16)

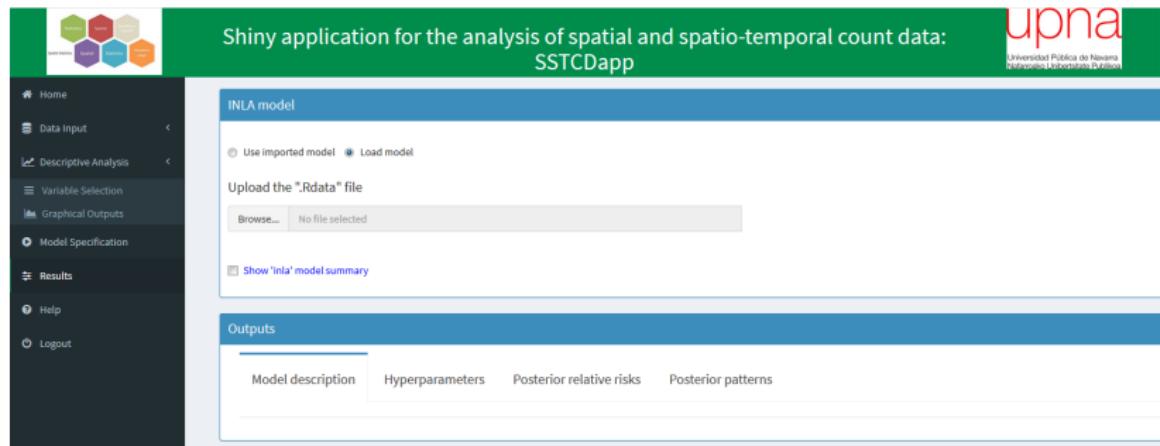
Retrieve selected model(s) Delete selected model(s)

Select all

Figure 7: Fitted models (SSTCDapp).

Results

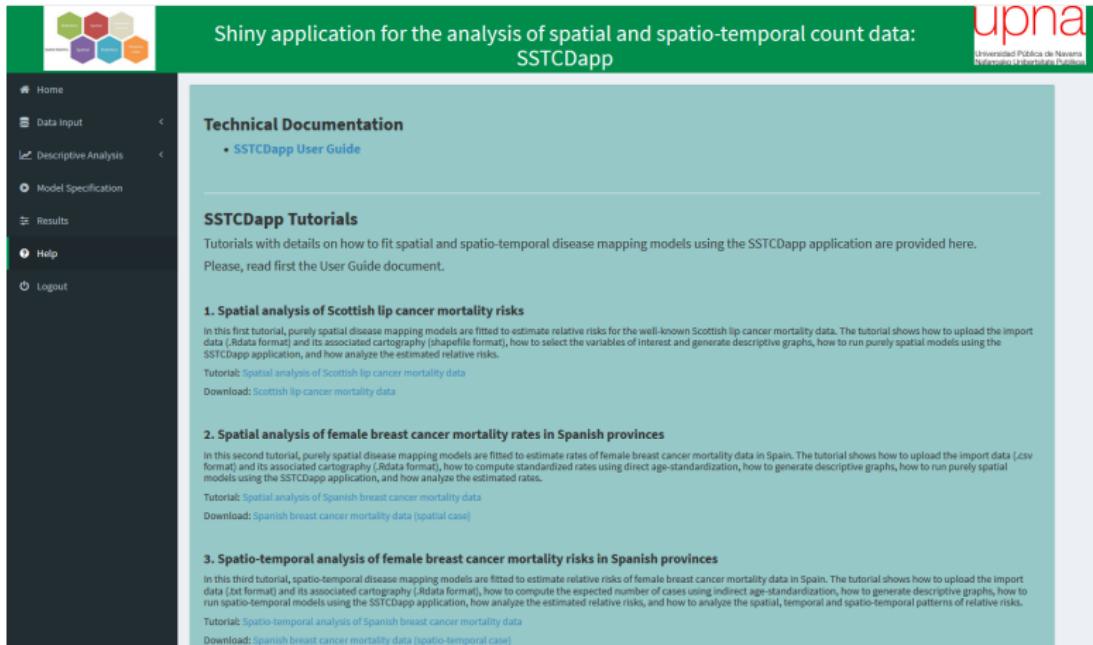
- Summary measures and several graphs/tables are provided for the
 - Posterior distributions of model hyperparameters.
 - Relative mortality/incidence risks (or rates).
 - Spatial, temporal, and spatio-temporal patterns.
- These results are computed for the currently imported model or any previously saved model.



The screenshot shows the SSTCDapp Shiny application interface. At the top, there is a green header bar with the text "Shiny application for the analysis of spatial and spatio-temporal count data: SSTCDapp" and the logo of UPNA (Universidad Pública de Navarra). On the left side, there is a sidebar menu with the following items: Home, Data Input, Descriptive Analysis, Variable Selection, Graphical Outputs, Model Specification, Results (which is the active tab), Help, and Logout. The main content area is titled "INLA model". It contains two radio buttons: "Use imported model" (selected) and "Load model". Below this is a "Upload the ".Rdata" file" input field with a "Browse..." button and a message "No file selected". There is also a checkbox labeled "Show 'inla' model summary". At the bottom of the main content area, there is a section titled "Outputs" with four tabs: "Model description" (selected), "Hyperparameters", "Posterior relative risks", and "Posterior patterns".

Tutorials

The application includes a detailed user guide and a set of tutorials to show how to fit spatial and spatio-temporal models using SSTCDapp.



The screenshot shows the SSTCDapp Shiny application interface. On the left is a sidebar with a navigation menu:

- Home
- Data Input
- Descriptive Analysis
- Model Specification
- Results
- Help
- Logout

The main content area has a green header bar with the text "Shiny application for the analysis of spatial and spatio-temporal count data: SSTCDapp" and the UPNA logo.

Technical Documentation

- SSTCDapp User Guide

SSTCDapp Tutorials

Tutorials with details on how to fit spatial and spatio-temporal disease mapping models using the SSTCDapp application are provided here. Please, read first the User Guide document.

1. Spatial analysis of Scottish lip cancer mortality risks

In this first tutorial, purely spatial disease mapping models are fitted to estimate relative risks for the well-known Scottish lip cancer mortality data. The tutorial shows how to upload the import data (.Rdata format) and its associated cartography (.shapefile format), how to select the variables of interest and generate descriptive graphs, how to run purely spatial models using the SSTCDapp application, and how analyze the estimated relative risks.

Tutorial: [Spatial analysis of Scottish lip cancer mortality data](#)
Download: [Scottish lip cancer mortality data](#)

2. Spatial analysis of female breast cancer mortality rates in Spanish provinces

In this second tutorial, purely spatial disease mapping models are fitted to estimate rates of female breast cancer mortality data in Spain. The tutorial shows how to upload the import data (.Rdata format) and its associated cartography (.Rdata format), how to compute standardized rates using direct age-standardization, how to generate descriptive graphs, how to run purely spatial models using the SSTCDapp application, and how analyze the estimated rates.

Tutorial: [Spatial analysis of Spanish breast cancer mortality data](#)
Download: [Spanish breast cancer mortality data \(spatial case\)](#)

3. Spatio-temporal analysis of female breast cancer mortality risks in Spanish provinces

In this third tutorial, spatio-temporal disease mapping models are fitted to estimate relative risks of female breast cancer mortality data in Spain. The tutorial shows how to upload the import data (.txt format) and its associated cartography (.Rdata format), how to compute the expected number of cases using indirect age-standardization, how to generate descriptive graphs, how to run spatio-temporal models using the SSTCDapp application, how analyze the estimated relative risks, and how to analyze the spatial, temporal and spatio-temporal patterns of relative risks.

Tutorial: [Spatio-temporal analysis of Spanish breast cancer mortality data](#)
Download: [Spanish breast cancer mortality data \(spatio-temporal case\)](#)

Let's start using the app

Let's start using the app!

1. Spatial analysis of Scottish lip cancer mortality risks
2. Spatial analysis of female breast cancer mortality rates in Spanish provinces
3. Spatio-temporal analysis of female breast cancer mortality risks in Spanish provinces

Conclusions

- The SSTCDapp was mainly developed to estimate relative risks using spatial and spatio-temporal disease mapping models.
- It provides separate spatial, temporal, and spatio-temporal patterns together with the corresponding exceedence probabilities and/or credibility intervals.
- The key advantage of this application in comparison with other software commonly used in disease mapping is that it provides an easy-to-use interface that facilitate the fit of fairly complex models without installing any software in user's computer.

Future development of the application

- Integration of `sf` ([simple feature](#)) objects ([Pebesma, 2018](#)) as cartography files to generate maps.
- To include interactive data visualization graphs using the R packages `leaflet` ([Cheng et al., 2018](#)), `dygraphs` ([Vanderkam et al., 2018](#)) and `tmap` ([Tennekes, 2018](#)).
- To implement other spatio-temporal model proposals such as
 - B-spline models accounting for both spatial and temporal correlation ([Ugarte et al., 2017](#)).
 - Models for age-specific mortality/incidence patterns ([Goicoa et al., 2016, 2017](#)).
 - Models to estimate disease risks in the presence of local discontinuities and clusters ([Adin et al., 2018](#)).

Developers and maintainer

The SSTCDapp application was developed by Aritz Adin together with the Spatial Statistics Group of the Public University of Navarre, Spain under the following grants:

- Spanish Ministry of Economy and Competitiveness (Project MTM2014-51992-R).
- Health Department of the Navarre Government (Project 113, Res.2186/2014).
- Spanish Ministry of Economy, Industry, and Competitiveness (Project MTM2017-82553-R jointly financed by FEDER).

For bug reports and support, please use the following email account:

app.estadistica@unavarra.es

References

- Adin, A., Lee, D., Goicoa, T., and Ugarte, M. D. (2018). A two-stage approach to estimate spatial and spatio-temporal disease risks in the presence of local discontinuities and clusters. *Statistical Methods in Medical Research (in press)*, pages 1–19.
- Adin, A., Martínez-Beneito, M., Botella-Rocamora, P., Goicoa, T., and Ugarte, M. (2017). Smoothing and high risk areas detection in space-time disease mapping: a comparison of P-splines, autoregressive, and moving average models. *Stochastic Environmental Research and Risk Assessment*, 31(2):403–415.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 192–236.
- 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):1–20.
- Bivand, R. S., Pebesma, E., and Gomez-Rubio, V. (2013). *Applied spatial data analysis with R, Second edition*. Springer, NY.
- Chang, W., Cheng, J., Allaire, J., Xie, Y., and McPherson, J. (2018). shiny: Web Application Framework for R. R package version 1.1.0.
- Cheng, J., Karambelkar, B., and Xie, Y. (2018). leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library. R package version 2.0.1.
- Dean, C., Ugarte, M., and Militino, A. (2001). Detecting interaction between random region and fixed age effects in disease mapping. *Biometrics*, 57(1):197–202.
- Goicoa, T., Adin, A., Etxeberria, J., Militino, A. F., and Ugarte, M. D. (2017). Flexible Bayesian P-splines for smoothing age-specific spatio-temporal mortality patterns. *Statistical Methods in Medical Research (in press)*, pages 1–20.
- Goicoa, T., Adin, A., Ugarte, M. D., and Hodges, J. S. (2018). In spatio-temporal disease mapping models, identifiability constraints affect PQL and INLA results. *Stochastic Environmental Research and Risk Assessment*, 32(3):749–770.
- Goicoa, T., Ugarte, M., Etxeberria, J., and Militino, A. (2016). Age–space–time CAR models in Bayesian disease mapping. *Statistics in Medicine*, 35(14):2391–2405.
- Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine*, 19(17–18):2555–2567.
- Leroux, B. G., Lei, X., and Breslow, N. (1999). Estimation of disease rates in small areas: A new mixed model for spatial dependence. In Halloran, M. and Berry, D., editors, *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, pages 179–191. Springer-Verlag: New York.
- Martins, T. G., Simpson, D., Lindgren, F., and Rue, H. (2013). Bayesian computing with INLA: new features. *Computational Statistics & Data Analysis*, 67:68–83.
- Pebesma, E. (2018). sf: Simple Features for R. R package version 0.6-3.
- Riebler, A., Sørbye, S. H., Simpson, D., and Rue, H. (2016). An intuitive bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research*, 25(4):1145–1165.
- 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(2):319–392.
- Simpson, D., Rue, H., Riebler, A., Martins, T. G., Sørbye, S. H., et al. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, 32(1):1–28.