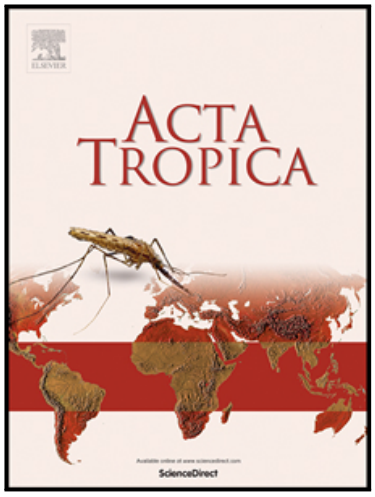
Journal Pre-proof



A Bayesian Functional Methodology for Dengue Risk Mapping in Latin America and the Caribbean

Torres-SignesA, DipJ.A.

PII: S0001-706X(20)31701-0
DOI: https://doi.org/10.1016/j.actatropica.2020.105788
Reference: ACTROP 105788

To appear in: *Acta Tropica*

Received date: 16 April 2020
Revised date: 4 November 2020
Accepted date: 30 November 2020

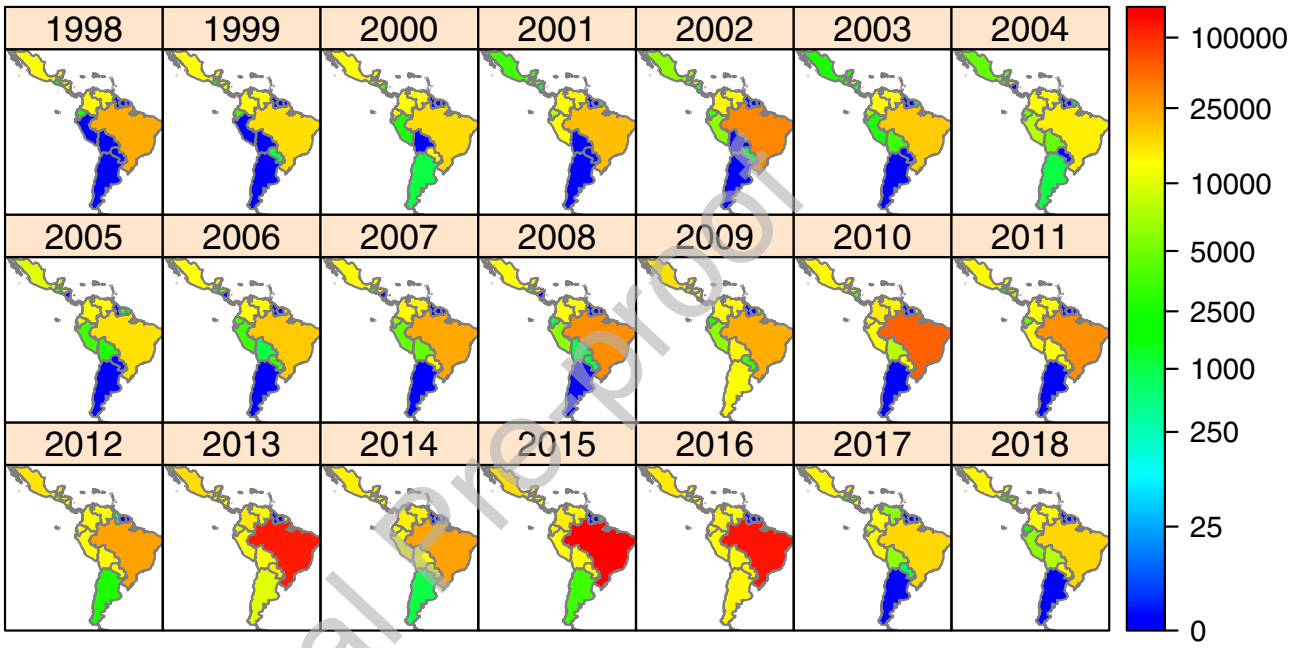
Please cite this article as: Torres-SignesA, DipJ.A., A Bayesian Functional Methodology for Dengue Risk Mapping in Latin America and the Caribbean, *Acta Tropica* (2020), doi: https://doi.org/10.1016/j.actatropica.2020.105788

Graphical Abstract

A Bayesian Functional Methodology for Dengue Risk Mapping in Latin America and the Caribbean

Torres-Signes, A,Dip, JA

Improved estimation of dengue disease mapping using a Bayesian functional methodology with data provided by the Pan American Health Organization



Highlights

A Bayesian Functional Methodology for Dengue Risk Mapping in Latin America and the Caribbean

Torres-Signes, A,**Dip, JA**

- We propose a functional approach to estimate the risk of dengue in America
- A Bayesian estimation in a Hilbert autoregressive process fits the behavior of data
- The model retains the heterogeneity and provides robustness and stability estimates
- The temporal evolution of risk maps is compared with other traditional models

A Bayesian Functional Methodology for Dengue Risk Mapping in Latin America and the Caribbean

Torres-Signes, A[a,*,1], **Dip, JA**[b,2]

[a]*Department of Statistics and O. R. Faculty of Sciences. AMZET. University of Málaga. Spain*

[b]*Department of Economy and Finance. Faculty of Economic Sciences. Kumbykuaa Observatory. National University of Misiones. Argentina*

ARTICLE INFO

Keywords:
Dengue fever
Bayesian estimation
Hilbertian autoregressive processes
Functional data analysis
Disease risk mapping

ABSTRACT

Dengue fever has become one of the most outstanding infectious diseases in the world. Besides, the incidence and prevalence of dengue are increasing in the endemic areas of the tropical and subtropical regions. Space and time disease mapping models are common instruments to explain the patterns of disease counts, where hierarchical Bayesian models constitute a suitable framework for their formulation. These random events reflect interactions between nearby geographic locations, as well as correlations between close temporary instants. Functional data analysis techniques can better describe the evolution of disease mapping.

In this paper, the risk of dengue in Mexico, Central and South America is studied from a Functional approach through a Bayesian estimation model focused on Hilbert-valued autoregressive processes combined with the Kalman filtering algorithm. Thus, the temporal functional evolution of spatial geographic patterns of incidence risk in disease mapping during 1998-2018 is approximated. Applying this methodology, the excess of smoothing that occurs with traditional models is avoided and the heterogeneity is conserved across the years. It improves the number of false positives created by noise and the number of false negatives as well. The results obtained with the application of this model are compared with those of previous models, corroborating the preceding statements and obtaining better results in the relative risk estimates, providing greater robustness and stability of disease risk estimates.

1. Introduction

Dengue fever (DF) is a severe viral disease that has set itself up worldwide in both endemic and epidemic transmission cycles as the most important mosquito-borne viral disease. The incidence and prevalence of dengue are increasing in the endemic areas of the tropical and subtropical regions. According to the World Health Organization, before 1970 only nine countries had encountered aggravated dengue epidemics and nowadays the disease is endemic in more than 100 countries, with the Americas, South-East Asia, and Western Pacific regions being the most seriously affected.

Uruguay and Chile have been the only countries in Latin America not affected by the transmission of dengue fever.

Research into spatial and spatio-temporal disease mapping has been carried out within a Bayesian framework.

Space and time disease mapping models are common instruments to explain the patterns of disease counts and they are usually formulated in hierarchical Bayesian models. The usefulness of these models for these purposes, as is pointed out in Aswi et al. (2009), lies in their flexibility to incorporate at various levels the variability of the model which allows a more complete assessment of prediction uncertainty.

Regarding Bayesian spatial and spatio-temporal approaches to modeling dengue cases, the literature has focused mainly on fully Bayesian models with a spatially structured random effect, using a conditional autoregressive (CAR) prior structure to examine the relationship between the risk of dengue and selected covariates.

*Corresponding Author
atisignes@uma.es (**Torres-Signes, A**);
dip@fce.unam.edu.ar (**Dip, JA**)
ORCID(s): 0000-0001-6218-5790 (**Torres-Signes, A**);
0000-0003-3714-2478 (**Dip, JA**)

Theoretical models derived from the spatial functional statistics have emerged as a new branch of statistics allowing the disease mapping as well. This is the approach that we will follow in this article in line with the work of Espejo-Montes (2014), Ruiz-Medina et al. (2014) and Torres et al. (2016).

In this paper, a Functional estimation model is proposed to analyze the incidence risk of dengue in Mexico, Central and South America. Specifically, a Hilbert-valued autoregressive process of order one, ARH(1), is used to estimate the temporal functional evolution of spatial geographic patterns of incidence risk in disease mapping during 1998-2018. The covariance operator of the ARH(1) process will be estimated both from a classical and a hierarchical Bayesian perspective. The incorporation of the Kalman filtering algorithm in this model will allow a smoothing of the relative risks estimates. The results will be compared with those obtained by the application of three alternative hierarchical spatio-temporal models: a Leroux model, an Intrinsic Conditional Auto-Regressive model (iCAR) and a Besag, York and Mollie model (BYM), see Adin et al. (2019).

Based on the previous studies and literature, the application of an ARH(1) model to make predictions is ample but, to our best knowledge, a Hilbertian autoregressive process with hierarchical Bayesian estimation methodology has not been used to model, specifically, DF cases.

2. Literature Review

Within the field of health, epidemiology investigates the evolution of spatial patterns of disease distribution in the human population, among other aspects. It is fascinating to see the possibilities of applying statistics to modeling such events. The occurrence of episodes is recorded to determine the mortality and incidence risks associated with a disease when the goal is an evolution analysis. Thus, determining spatial patterns is of particular interest to identify at risk territories and based on risk maps, better performance of national public health emergency plans can be achieved (SÃąnchez-GÃşmez et al., 2017). Investment in dengue surveillance and response along with risk mapping and its communication are essential for early outbreak detection. Lawson (2008) suggests that users and readers are more comfortable if the aggregated health data are analyzed in maps.

According to Adin et al. (2019), the significant variability intrinsic to classical risk estimation measures makes it necessary to adopt models to smooth risks borrowing information from spatial and temporal neighbors. To develop disease maps, studies have used dengue incidence data, which had been assumed to follow a Poisson process. See Aswi et al. (2009).

The epidemiological literature paid more attention to models based on conditional autoregressive structures, well known as iCAR (Besag, 1974), BYM models (Besag et al., 1991) and their alternatives to disease mapping. However, recently temporal and spatial functional analysis, specifically the spatio-temporal random field (S/TRF), has emerged to integrate space and time in the construction of incidence maps (Espejo-Montes, 2014).

In general, the number of counts, $C_{i,t}$, in area i at time t is assumed to be Poisson distributed with mean $\theta_{i,t} = e_{i,t} r_{i,t}$, where $e_{i,t}$ is the number of expected cases of DF, see Bivand et al. (2013) or Aswi et al. (2009), among others. $r_{i,t}$ is the relative risk in area i and time t.

A usual assumption for the log risk decomposition is

$$\log r_{i,t} = \eta + \xi_{i,t} + \phi_t + \gamma_t + \delta_{i,t},$$

where η represents the overall level of risk, $\xi_{i,t}$ is the spatial component, ϕ_t is the unstructured temporal effect, γ_t represents the structured temporal effect and $\delta_{i,t}$ is the space-time interaction effect (Adin et al., 2019).

Spatio-Temporal modeling has been understood as an extension of the spatial case. Different Conditional Autoregressive (CAR) structures have been proposed to model the spatial component. CAR structures smooth noisy estimates by pooling information from neighboring regions (Morris et al., 2019). Typical structures can be mixed in these Spatio-Temporal Models for disease mapping. More details of each structure and a brief discussion can be found in (Adin et al., 2019; Morris et al., 2019; Eberly and Carlin, 2000; Leroux et al., 1999; Riebler et al., 2016; Anderson et al., 2017; Knorr-Held, 2000; Adin et al., 2018).

On the other side, Functional data analysis (FDA) arises when one of the variables of interest in a data set can be seen naturally as a curve or a smooth function, i.e. it can be considered as the statistical analysis of curve samples. The functional time series analysis of epidemiological data has been developed recently (Espejo-Montes, 2014; Ruiz-Medina et al., 2014; Torres et al., 2016). Espejo-Montes (2014) and Ruiz-Medina et al. (2014) propose a model for stabilization of SMR (standardized mortality ratio), based on the combination of ARH(1) and Reproducing Kernel Hilbert Space approach for the analysis of breast cancer mortality data, while Torres et al. (2016) apply a spatio-temporal log-Gaussian Cox process based on Hilbert-valued random intensity and defined from the Ornstein-Uhlenbeck process (in Hilbert spaces) in order to draw functional disease mapping for prostate, breast and brain cancers. However, so far we believe that there are no studies that have applied the analysis of functional data based on ARH(1) processes of dengue epidemiological data in Latin America and the Caribbean.

Recently, some papers have focused on the DF dynamics and geographical incidence. The most commonly used proposals are generalized linear mixedmixed models with proper CAR (conditional autoregressive) spatial random effects (Jaya et al., 2016) or with spatial, temporal and spatio-temporal random effects (Wijayanti et al., 2016). Others, like Yu et al. (2014), estimate the relative risk for the transmission of dengue fever based on discrete time and space via a susceptible infectious recovered model for human populations with a Bayesian maximum entropy design. Some of them use Bayesian approaches to model dengue fever in Asia (Wijayanti et al., 2016; Jaya et al., 2016; Zhu et al., 2016) and consider socioeconomic aspects, geography, entomological and climates covariates to explain DF cases. In the Americas, DF in Brazil is studied by Pastrana et al. (2014), who deal with spatial and statistical methodologies to analyze the geographic distribution of dengue and to relate its incidence to the Health Vulnerability Index (HVI). Lowe et al. (2014) formulate another generalized linear mixed model, using a negative binomial distribution for the dengue case counts to predict dengue epidemic in Brazil during the 2014 football world cup. Others, like Pepin et al. (2015) use hierarchical Bayesian regression modeling to examine the role of city-wide vector surveillance data in predicting human cases of dengue fever in space and time in Vitoria, Brazil. Lorenz et al. (2020) study *Aedes aegypti* mosquito and its dissemination of dengue, Zika, chikungunya, and urban yellow fever. Using a Bayesian model and Stochastic Partial Differential Equation method, they determine land features associated with mosquito infestations in SÃčo JosÃĺ do Rio Preto/SP, Brazil. In Colombia, Martinez-Bello et al. (2018) generate smoothed estimates of relative risk applying hierarchical Bayesian spatio-temporal models, adding covariates obtained from satellite images containing land surface temperature and a normalized difference vegetation index, for the period January 2009 âĂŞ December 2015 in Bucaramanga. Restrepo et al. (2014) examine the variation in the spatial distribution of notified dengue cases in Colombia from January 2007 to December 2010 with a Bayesian spatio-temporal conditional autoregressive model. Venezuela was analyzed by Cabrera and Taylor (2019) with spatio-temporal modeling of dengue fever via a Generalized Additive mixed model. In Argentina, Estallo et al. (2014) study the spatio-temporal dynamics of the outbreak, in Córdoba city, during 2009, but they lack the visualization of a Bayesian model. Spatio-temporal models described above have been fitted, mostly, by Markov Chain Monte Carlo (MCMC) methods but, recently, a novel technique called Integrated Nested Laplace approximations (INLA) has been used to lighten the computation burden.

Existing studies have examined the dengue risk at regional (country) or local scales (city) in a certain period. Besides, classical approaches like the ones based on Conditional Autoregressive models (CAR) and spline-based smoothing, in a mixed effect framework, present the problem of hyper-smoothing (Espejo-Montes, 2014). "*ARH(1) processes are considered for the construction of dynamical spatial epidemiological maps, capturing spatio-temporal interaction, and assuming a functional nature of the log-risk magnitude in space*", p.944 (Ruiz-Medina et al., 2014). Most of the reviewed models are analyzed in a spatio-temporal process, but, to our best knowledge, functional approaches based on Hilbert-valued processes in a Bayesian framework have not been used so far in this area.

This paper tries to fill this gap by applying an ARH(1) approach in a hierarchical Bayesian model to map the evolution of dengue fever risk in 32 American countries during 1998–2018. We use Functional Data Analysis (FDA) estimating a curve over time for each area under study.

3. Methods

3.1. The model

In this section, the model used for the representation of the incidence of the DF in the different countries considered in the study is introduced. The evolution of the disease mapping is analyzed from the perspective of spatio-temporal statistics. A functional approach is adopted for the representation of the temporal evolution of spatial patterns, in terms of a first-order autoregressive Hilbertian model, ARH(1). The autocovariance operator for this model is obtained from the Bayesian and classical inference. Accordingly, the autocorrelation operator is calculated, see Bosq (2000), which can be interpreted as a spatio-temporal interaction operator for surface sequences in time. The Kalman filtering algorithm is used in order to achieve a smoothed *standardized mortality ratio*, SMR (a common index of disease incidence and mortality). In our case, the areas where we observe the disease are countries, i. e., large geographical regions, so we use the SMR as a reliable measure of the relative risk of DF, see Meza (2003) and Mukhsar et al. (2016). Thus, the maximum likelihood estimator of the relative risk in an area is given by SMR = observed cases / expected cases.

The Poisson process with mean $\theta_{i,t}$ is considered to model the number of new infections of DF, $C_{i,t}$, in area i at time t. For each area i and time t, the expected number of cases of DF, $e_{i,t}$, is obtained, so $\theta_{i,t} = r_{i,t}\, e_{i,t}$, where $r_{i,t}$ is the relative risk in area i and time t. Thus,

$$\begin{aligned} C_{i,t} \mid r_{i,t} &\sim \mathcal{P}\left(\theta_{i,t} = e_{i,t} r_{i,t}\right), \\ \log \theta_{i,t} &= \log e_{i,t} + \log r_{i,t} \end{aligned} \tag{1}$$

Then, the observation model, $Z_{i,t}$, is defined by the following equation, derived by considering the Taylor expansion of the logarithm function of the estimation of $r_{i,t}$ at the observed value $\theta_{i,t}$. For each $t = 1, \dots, T$, and for any $i = 1, \dots, N$,

$$\begin{aligned} Z_{i,t} = \log\left(\frac{C_{i,t}}{e_{i,t}}\right) &= \log C_{i,t} - \log e_{i,t} \\ &= \log \theta_{i,t} + \frac{C_{i,t} - \theta_{i,t}}{\theta_{i,t}} - \log e_{i,t} \\ &= \log r_{i,t} + \frac{C_{i,t} - \theta_{i,t}}{\theta_{i,t}} \\ &= \log r_{i,t} + v_{i,t}, \end{aligned} \tag{2}$$

where, from the Gaussian approximation of the Poisson distribution, for each $t = 1, \dots, T$, and for any $i = 1, \dots, N$, $v_{i,t}$ may be considered as a Gaussian white noise with variance $1/\theta_{i,t}$, see Torres et al. (2016). We also assume the necessary conditions to consider this spatio-temporal white noise, $v_{i,t} = v(\mathbf{u_i}, \mathbf{t})$, $\mathbf{u_i} \in D \subseteq \mathbb{R}^2, \mathbf{t} \in \mathbb{Z}$, as a Hilbert-valued white noise in the strong sense in time, see Bosq (2000).

In what follows, we assume that the autocovariance operator (see Appendix A), R_0, is a trace operator that admits a spectral decomposition in terms of a system of eigenvectors. For $k \geq 1$,

$$R_0 \phi_k = \lambda_k(R_0)\phi_k.$$

Next, the Bayesian estimation of the parameters that define the spectral decomposition, $\lambda_k(R_0)$, is described,

$$R_0 = \sum_{k=1}^{\infty} \lambda_k(R_0)\phi_k \otimes \phi_k.$$

The results are obtained in a hierarchical model approach. Specifically, in the first stage, the parameters are estimated in a Bayesian and classical approach. In the Bayesian case, conjugate families with respect to the univariate Gaussian likelihood in terms of each projection are considered, assuming a Gaussian ARH(1) framework. In the second stage, the logarithmic transformation of the relative risk is approximated with the state equation of an ARH(1) model by applying the Kalman filtering, see Appendix A.

To estimate the autocovariance operator, $\mathbf{R}_0$, for each time $t = 1, \dots, T$, we consider $Y_k(t) = \sqrt{\lambda_k(\mathbf{R}_0)} Z_k(t) \sim \mathcal{N}(0,\ \lambda_k(\mathbf{R}_0);\ k \geq 1$. The conjugate prior family of distributions for the scale parameter of the univariate Gaussian likelihood is an inverse gamma distribution, $\mathcal{IG}(\alpha, \beta)$. We consider, for every $k \geq 1$, the

parameter λ_k is distributed as $\mathcal{IG}\left(\alpha_k, \beta_k\right)$; then, the posterior distribution of λ_k is given by $\mathcal{IG}\left(\alpha_k + T/2,\ \beta_k + \sum_{t=1}^{T} Z_k(t)/2\right)$; see, for example, Lehmann and Casella (1998), among others. The Bayesian estimator, under the loss squared function, is obtained by the following expression:

$$\widehat{\lambda}_k(\mathbf{R}_0) = \frac{2\beta_k + \sum_{t=1}^{T} Z_k^2(t)}{T + 2\alpha_k - 2}, \quad k = 1, ..., M. \quad (3)$$

where M is a truncation order.

In this paper, in absence of information from other studies on the scale parameters to be estimated, we have considered the use of a noninformative prior in the Inverse Gamma family of distributions, the $\mathcal{IG}(\epsilon,\ \epsilon)$, with $\epsilon > 0$ and with a low value. This distribution is known as weakly informative distribution, see Gelman (2006). The main advantages of its use, as opposed to other noninformative prior distributions, is that they are conjugate priors. These distributions are frequently used, see Ugarte et al. (2009), to estimate scale parameters in hierarchical models in spatial and spatio temporal disease mapping models.

This Bayesian estimator, for every $k = 1, \ldots, M$, is compared with the classical one, that is, the maximum likelihood estimator:

$$\widetilde{\lambda}_k(\mathbf{R}_0) = \frac{\sum_{t=1}^{T} Z_k^2(t)}{T}. \quad (4)$$

Note that the classical case consists in the Bayesian estimator where, for every $k = 1, \ldots, M$, the prior distribution hyperparameters are $\alpha_k = 1$ and $\beta_k = 0$.

3.2. Data

The database is provided by PAHO (2019). It includes 32 countries with annual data, from 1998 to 2018 and all the reported dengue cases are considered. According to PAHO, a huge quantity of cases are asymptomatic and hence the actual numbers of dengue cases are underreported and many cases are misclassified. Although the database has observations for countries since 1980, the number of dengue cases equal to zero has been lower since 1998. Besides, it is difficult to distinguish between an unreported value (missing) and zero dengue cases, so that we have considered all zeros as actual ones. Thus, we have selected for the study all the countries of America, which present up to five zeros, from 1998 to 2018. These data, which include 672 values with 24 zeros, are represented in Figure 1.

The heterogeneity of the data within each country and year is remarkable. For instance, Guadeloupe usually has few DF cases, which include zeros, but in 2010 reaches 40, 737 DF cases. On the other side, Brazil always exhibits a large number of DF cases though it also has a large population. It is a dengue-endemic country and over the years has been facing several outbreaks (2013, 2015 and 2016) with neighboring countries being hit as well. The high number of dengue cases observed in some countries suggest that the cases imported from adjoining countries contribute to the spread of the disease. The heterogeneity of the countries in each year is clear in Figure 1. In that sense, 2015 is a remarkable year, in which there are countries with low number of DF cases, like Saint Kitts and Nevis (5 cases, while its mean for the years of the study is around 27), and others with a large number of DF cases, like Mexico (219, 593 cases while its mean for the years of the study is around 75, 184). Despite this fact, as already pointed out by other authors in the previous literature (Carbajo et al., 2001; Restrepo et al., 2014; Martinez-Bello et al., 2018), in general, an increasing trend in the number of DF cases can be observed, which is apparent since from 2012 the colors in the maps are more intense. The last two years are out of this trend.

4. Results

In order to estimate the eigenvalues of the covariance operator we have considered both the classical and Bayesian approach. Thus, $\widehat{\lambda}_k(\mathbf{R}_0)$ and $\widetilde{\lambda}_k(\mathbf{R}_0)$ will be computed for $k = 1, \ldots, M$. To obtain the Bayesian estimator we have used the class of inverse gamma prior distribution, introduced in Section 3, giving the value $\epsilon = 0.05$, that is, an $\mathcal{IG}(0.05,\ 0.05)$ distribution.

In the implementation of the Kalman filtering algorithm, based on a truncated diagonal version of the ARH(1), equation (7) in Appendix A, and the corresponding Bayesian parameter estimation implemented in Section 3, the empirical values of α_k and β_k, for $k = 1, \ldots, M = 19$, are considered in both cases, the Bayesian and classical estimation of the autocovariance operator. The choice of the trun-

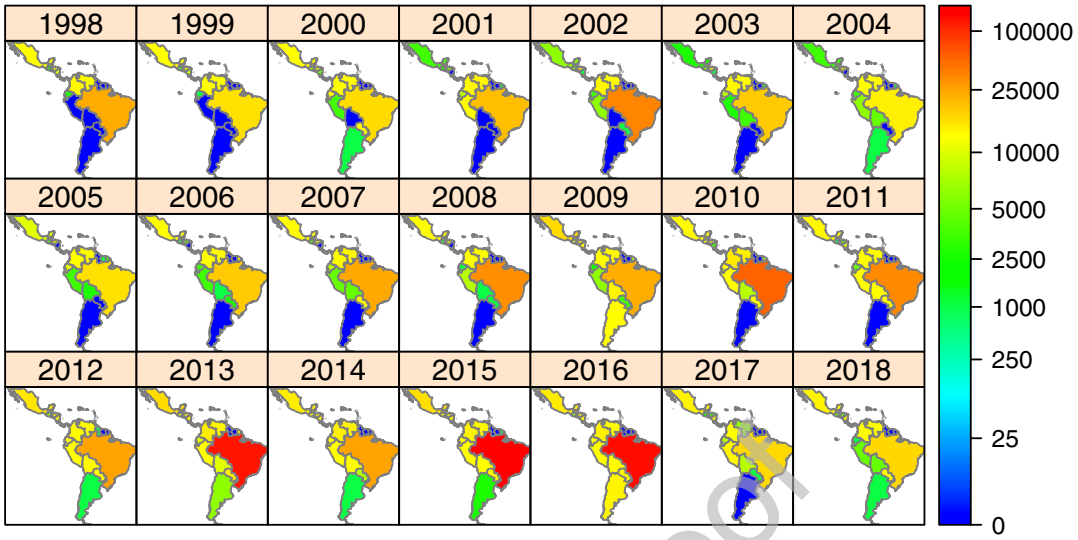


Figure 1: Evolution of the number of DF observed in the years and countries under study.

cation order, $M = 19$, has been made taking into consideration that the eigenvalues of a higher order than T, when $T < N$, theoretically are null, see Bosq and Blanke (2007). On the other hand, the determination of the truncation order is complex because a too small one produces poor estimations but, with a too large order, the prediction error could explode, see Bosq (2000). We have checked that neither of these are the case with our choice. Also, this truncation order leads to a very high percentage of explained empirical variability in terms of the approximated trace of the autocovariance operator of the ARH(1) process, as a criterion of truncation, see Ruiz-Medina et al. (2014). A measure of variability between the observed values and the estimates is displayed in Table 1. It is the Empirical Absolute Error of the estimated log-risk in both cases, Bayesian and classical estimation, $|\log(r_{i,t}) - \log(\widehat{r}_{i,t})|$ and $|\log(r_{i,t}) - \log(\widetilde{r}_{i,t})|$, respectively, for each year, Averaged in all countries ($i = 1, \ldots, 32$) under study (EAEA).

To capture the heterogeneity of the error measurements from another perspective, Table 2 shows the Empirical Square Error for the log-risk in both Bayesian and classical estimation, $(\log(r_{i,t}) - \log(\widehat{r}_{i,t}))^2$ and $(\log(r_{i,t}) - \log(\widetilde{r}_{i,t}))^2$, respectively, for the 32 countries considered Averaged in all years under study (ESEA).

Figure 2 exhibits the estimated log risk of the 32 American countries considered from 1998 to 2018 for the Bayesian estimation, since it has the lowest EMSE, $\left(\sum_{i=1}^{32} \text{ESEA}/32\right)$. From this figure, Figure 3 has been extracted to show the particular behavior of the two estimated cases and the observed log-risk in six particular countries. This measurement sheds some light on the behavior of the incidence of disease risk per country and year. When the measurement is around zero, the number of incidences estimated is similar to the expected ones, according to the cases observed in all countries. Negative values indicate that the number of incidences estimated is lower than the expected cases and conversely when the measurement is positive. The fact of having zeros in the database is the reason for too low log-risks (in this case, we have considered 10^{-4} for the null risks). On the opposite side, there are episodes in countries with a high risk that cause out-

Year	CEAEA	BEAEA	Year	CEAEA	BEAEA	Year	CEAEA	BEAEA
1998	0	0	2005	0.1038	0.0636	2012	0.1794	0.0983
1999	0.0940	0.6254	2006	0.3079	0.2324	2013	0.1823	0.1431
2000	0.1571	0.1258	2007	0.4676	0.1554	2014	0.1692	0.0989
2001	0.0910	0.0609	2008	0.3570	0.0789	2015	0.2865	0.1988
2002	0.1254	0.0891	2009	0.2442	0.0843	2016	0.3734	0.1198
2003	0.0973	0.1137	2010	0.3078	0.1492	2017	0.2011	0.0852
2004	0.0698	0.1101	2011	0.3283	0.1483	2018	0.3999	0.3405

Table 1
Empirical Absolute Errors Average for the relative risk of DF according to the average of the countries in the study, for the Bayesian and Classical estimation, (BEAEA) and (CEAEA), respectively.

Country	CESEA	BESEA	Country	CESEA	BESEA
Puerto Rico	0.3061	0.1619	Argentina	0.0910	0.0444
Dominican Republic	0.0068	0.0058	Brazil	0.0015	0.0007
Antigua and Barbuda	0.0578	0.0339	Paraguay	0.1669	0.0763
Barbados	0.1165	0.0757	Belize	0.0455	0.0298
Dominica	0.2094	0.1206	Costa Rica	0.0526	0.0164
Grenada	0.6863	0.3907	El Salvador	0.0156	0.0232
Guadeloupe	0.0344	0.0239	Guatemala	0.0064	0.0050
French Guiana	0.1225	0.0463	Honduras	0.0165	0.0058
Guyana	0.3861	0.1653	Mexico	0.0054	0.0034
Jamaica	0.4837	0.2895	Nicaragua	0.0097	0.0100
Martinique	0.0387	0.0185	Panama	0.0740	0.0585
Saint Kitts and Nevis	0.0103	0.0143	Bolivia	0.0245	0.0128
S. Vincent and the Grenadines	0.0563	0.0561	Colombia	0.0202	0.0123
Saint Lucia	0.0194	0.0271	Ecuador	0.0398	0.0275
Suriname	0.4153	0.5046	Peru	0.0226	0.0199
Trinidad and Tobago	0.0557	0.0468	Venezuela	0.0109	0.0115

Table 2
Empirical Square Errors Average for the relative risk of DF according to the average of the years in the study, for the Bayesian and Classical estimation, (BESEA) and (CESEA), respectively.

lying large values. For instance, French Guiana, in 2006, has 15,904 DF cases, one of its higher peaks while this year was not especially aggressive in the rest of the countries (see Figure 1). This causes an extreme value for the risk observation (see year 9 in Figure 3). On the other hand, being a small country, it cannot be seen in Figure 1 and 2. There are a lot of countries with important differences in their extensions, that make it difficult to observe the full range of colors used in the maps. This is one of the reasons why small countries are not included generally in disease risk mapping studies. Even so, we have considered it appropriate the use of all countries under the conditions commented in Section 3.2, since the estimation methodology proposed and the measurement used allow us to do it.

4.1. A comparison with previous approaches

In this section, we compare our model with three alternative and traditional spatio-temporal models: a Leroux model, an iCAR model and a BYM model, using SSTCDapp software, see Adin et al. (2019).

Table 3 details, for each methodology applied, the Empirical Mean Absolute Error (EMAE) and the Empirical Mean Square Error (EMSE) of the estimated DF log-risk for all countries and years under study, that

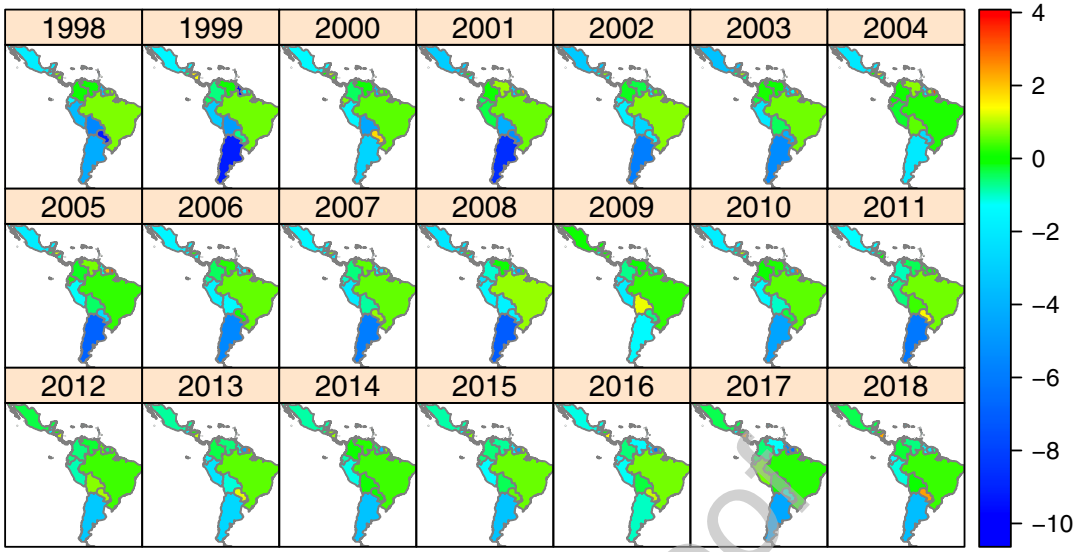


Figure 2: Estimated log-risk from the Bayesian estimation on the proposed model according to the years and countries under study.

is,

$$\text{EMAE} = \frac{1}{21}\frac{1}{32}\sum_{t=1}^{21}\sum_{i=1}^{32}|\log(r_{i,t}) - \log(\hat{r}_{i,t})|,$$

$$\text{EMSE} = \frac{1}{21}\frac{1}{32}\sum_{t=1}^{21}\sum_{i=1}^{32}(\log(r_{i,t}) - \log(\hat{r}_{i,t}))^2,$$

where $r_{i,t}$ and $\hat{r}_{i,t}$ are the observed an estimated relative risk. The best result is obtained for the Bayesian estimation, in both measurements, EMAE and EMSE. Figure 4 displays the log-risk estimates provided both by traditional spatio-temporal models and by the model we propose, jointly with the actual observations in two countries.

Methodology	EMAE	EMSE
LEROUX	0.1999	0.7854
iCAR	0.1987	0.7790
BYM	0.1974	0.7723
C-est.	0.2163	0.1128
B-est.	0.1486	0.0731

Table 3
Comparison of the Empirical Absolute Error Averaged (EMAE) and the Empirical Mean Square Error (EMSE) for all countries and years for the five methodologies applied.

5. Discussion

First, the Bayesian and classical estimators proposed in the model are compared. Table 1 and 2 show the small values for these summarizing measurements (empirical absolute and square errors for Bayesian and classical estimations), taking into account the large variability of risk values due to the heterogeneity of both population and number of DF cases. The year 1998 (Table 1) has a null value because of the use of an autoregressive model of order one. The rest of the years, both estimators present an irregular behavior where sometimes the classical estimator has better results and vice versa. In that sense, we can highlight, for instance, the years 2003 and 2007, respectively. But, the general behavior is a considerable increase in the em-

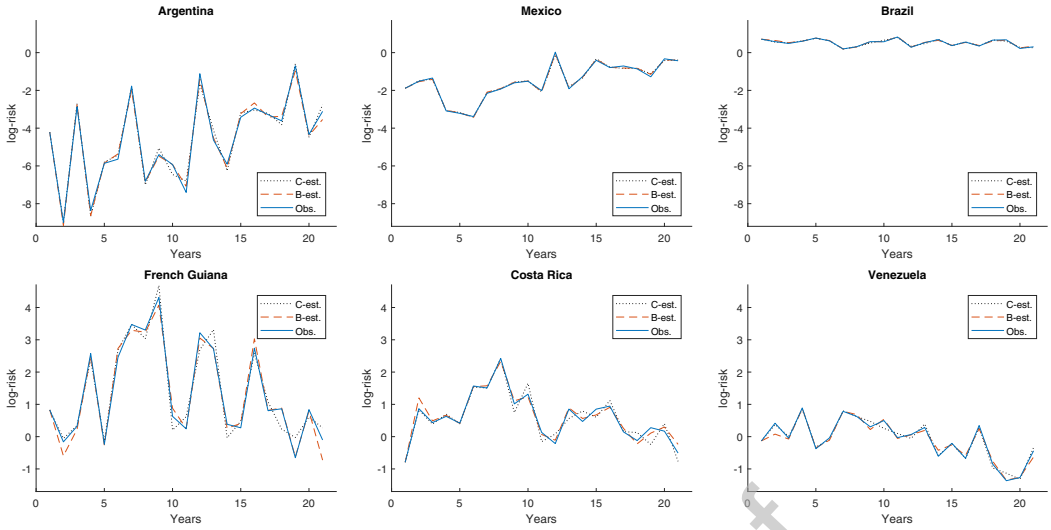


Figure 3: Estimated log-risk from the classical estimation (C-est., in dotted line) and Bayesian estimation (B-est., in dashed line), and the observed log-risk (Obs., in solid line) according to the years under study for the countries: Argentina, Mexico, Brazil, French Guiana, Costa Rica, and Venezuela.

pirical error in the classical case, see for instance the years 2000, 2001, 2002, 2005, etc. A similar behavior can be found in Table 2 where countries like Venezuela show a higher level of error for the Bayesian case but the general behavior is a considerable increase in the classical one, see for instance Argentina, Brazil, Paraguay, Belize, Costa Rica, etc. In addition, Table 2 shows the difference between the error measurements presented according to the size of the countries under study. Thus, in general, the empirical error measurement is smaller for large countries (Brazil, Mexico, and others) and higher for small countries (Grenada, Jamaica, Suriname, and others). Specifically, this is accentuated for the Bayesian case, where the BESEA becomes half than the CESEA in some large countries like Brazil.

In Figure 2, for all countries and years, the estimated log-risk is displayed according to the hierarchical Bayesian model, explained in Section 3. The heterogeneity of patterns is held in the estimation methodology proposed without an excess of smoothing, enabling the detection of high-risk areas and avoiding possible false negatives. The evolution of this measurement is very similar to the real data (see the error tables), keeping the observed heterogeneity in risk patterns in countries and time, i.e., heterogeneity and variability are still present, see also Figure 3 where the estimated and observed risk for some countries can be compared. This behavior suggests that the proposed models have fitted the data well.

Note that Figure 2 informs about the risk in one country with respect to the others in a specific year. We can observe the risk evolution in each country (compared to the rest of the countries) per year. However, in general, we cannot make a direct connection between this figure and Figure 1, although it may appear in some countries. For instance, we can observe that the number of DF cases in Argentina, despite being generally low, exhibits a global increasing tendency, besides, the risk globally increases compared to the rest of the countries in the study (see Figure 3 as well). It is worth noting that in the Argentinian case, the risk is mitigated by the inclusion of the entire population in the study when in fact the southern part is not a disease exposed population. On the other hand, Brazil presents both a high number of DF cases and a high level of risk in all the years of the study. However, this stability of the risk over time (see Figure 2 and 3) does not

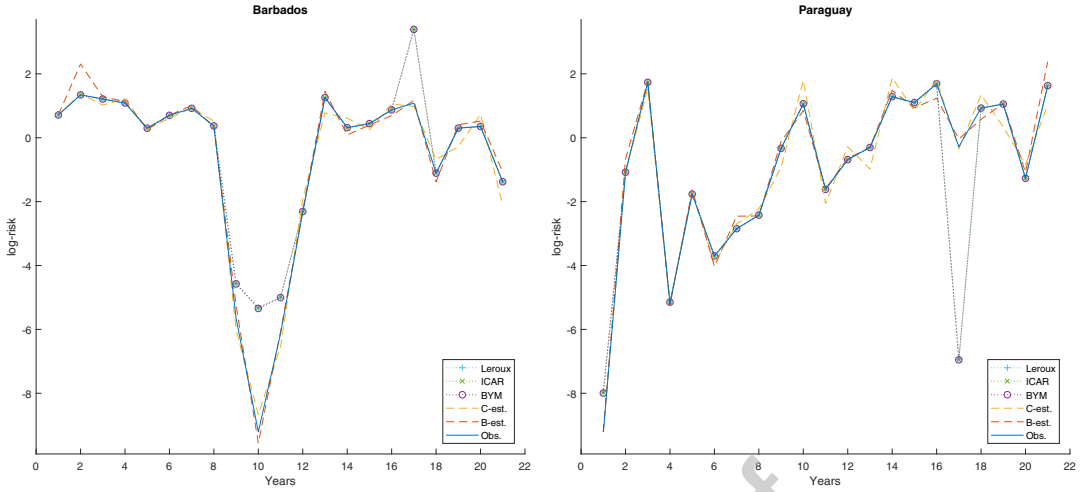


Figure 4: Estimated log-risk from the previous spatio-temporal models (with plus sign, cross and circle, in dotted lines), the estimation methodology proposed (in dashed lines), and the observed log-risk (in solid line) according to the years under study for the countries: Barbados and Paraguay.

reflect the large variability in the number of cases (see Figure 1).

Our results are similar to those found recently in Messina et al. (2019). The authors show high environmental suitability levels for dengue using data from 2015. They contemplate that any pixel with a predicted dengue suitability value above 0.467 was considered at risk. For South America, an increased risk of dengue fever arises in 2020 and 2050 when they fit a boosted regression tree model to forecast. Similarly, Bhatt et al. (2013) predict that dengue transmission is omnipresent throughout the tropics, with the Americas and Asia's highest risk zones. They also display a map with risk levels comparable to those calculated in our paper during 2010. However, Cattarino et al. (2020) explain that current dengue risk maps provide estimates of disease burden or endemicity boundaries rather than transmission intensity. Besides, they prognosticate high transmission intensity in all continents situated on both sides of the tropics, with hot spots in South America (Colombia, Venezuela, and Brazil). It is relevant to clarify that the cited papers' results cannot be compared directly. They employ different statistical methodologies and focus on particular years. However, in all of them, countries present similar dengue risk to those found with our method.

In the comparison with previous approaches (see Table 3), it can be observed that the errors for the estimation
methodology focused on an ARH(1) process have a more stable behavior in both measurements proposed, EMAE and EMSE. When we compare to the previous spatio-temporal models, the ones proposed here reach a better fit with lower levels of error. In particular, the Bayesian estimation we propose reduces the errors encountered with the other models and captures better the heterogeneity present in the data.

It is important to highlight the information given by the measurements in Table 3. The fact that, in the traditional spatio-temporal models considered, the EMSE is large compared to the EMAE is indicative that these techniques can present false negatives and positives. The EMAE reflects that, in most cases, these models fit correctly to spatio-temporal data, but the large difference with the EMSE reflects that, at specific moments, there are large differences between the estimated and observed risk. This can be due to an excess of smooth-

ing. In that sense, Figure 4 shows the presence of false positives in Barbados and false negatives in Paraguay for the previous approaches considered. With our proposed methodology, this effect is mitigated, since, even showing smoothing, it fits properly to the data, avoiding false negatives and positives, see Figure 4 and also 3. When comparing EMAE and EMSE (see Table 3), we look for small values and similar results, so that there is robustness in the estimations. In our proposal, EMAE and EMSE are smaller than in the traditional spatio-temporal models analyzed, so that the estimation methodology presented exhibits better behavior and greater stability, since the error measures are small and also close. EMSE is smaller in our approaches than in the other models and in particular, in our proposal EMSE is smaller than EMAE, which corroborates the goodness of fit since, being a squared measure, it can only become smaller if the values are below one, and penalizes big errors.

In the elaboration process of the present work, several problems have emerged, limiting the study. The number of reported dengue cases contains inconsistencies due to the asymptomatic and misclassified cases in the primary database. Problems with the database typically include discrepancies in laboratory confirmation rates (sensitivity and specificity), and, sometimes, countries revise the case definitions and classifications as time goes by. Furthermore, differences in the quality of data from distinct countries also cause difficulties for data analysis. Modeling areas with large differences in their extensions leads to a bit worse estimates, especially in small countries like the Caribbean islands, where the estimation of disease relative risk is more unstable. Readers can see this phenomenon in Figure 3, where the model fits better in large countries (Mexico, Brazil) and with a more considerable number of cases than in small countries (French Guyana, Costa Rica). This figure also shows that despite Argentina and Venezuela are large countries, they exhibit a worse adjustment, probably due to the quality of data, in addition, in the Argentinian case, not all the extension of the country is clearly exposed to the disease (see also Table 2).

6. Conclusions

Dengue fever cases, in Mexico, Central and South America, have been studied by a hierarchical Bayesian model focused on Hilbert-valued autoregressive processes of order one to explain the temporal evolution of the disease incidence. The spatial functional values allow the preservation of the heterogeneity of the areas considered with the temporal estimation. This represents an important contribution to the disease mapping in space and time approach, especially for dengue fever cases. Also, the use of this methodology with FDA techniques is a clear step forward from previous works on spatio-temporal Bayesian approaches, conditional autoregressive, and splines based models, where a considerable number of false negatives are generated. One limitation of dengue disease research is that some DF cases are underreported or misclassified. Excessive smoothing may hinder the detection of high-risk areas. It is important to see the ability of the model to detect true high-risk areas and to rule out false positives created by noise. Using the estimation methodology we propose, the excess of smoothing is avoided and the heterogeneity is conserved across the years under study so that the model detects high levels of risk and improves by decreasing the number of false negatives as well.

The empirical results of the proposed methodology support our findings, without underestimating the possible problem derived from the data source, as already discussed. The empirical errors of the estimated risk, for each year and country, are almost insignificant as shown in the summarized measurements. Furthermore, an empirical comparison with previous approaches gives more support to our model, producing much lower error measurements for the Bayesian estimation within the FDA framework proposed. We can conclude that, for this type of data, it is possible to find useful priors in a Bayesian FDA framework to achieve an estimation based on an ARH(1) model which performs better than the previous estimation approaches analyzed in this paper.

This way, the directional nature in time and regional variability in space of the DF data can be suitably modeled with the estimation methodology presented here. This approach provides more precise tools to address prevention and control measures as

well as decision making. The model presented adequately explains the behavior of the evolution of the risk of dengue fever. This characteristic is a necessary starting point for the prediction of future values. Not only that, the introduced methodology could be extended with a simulation study in order to analyze interesting properties of the considered Bayesian estimators according to the choice of the hyperparameters α_k and β_k that define the prior distribution. It would allow the comparison with other estimators besides the classical one used here, to analyze the behavior of the correlation between temporal trends of neighboring regions, and to study more deeply the selection problem associated with the truncation order M.

References

Adin A, Lee D, Goicoa T, Ugarte MD (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.* **28** ,9, 2595–2613. https://journals.sagepub.com/doi/10.1177/0962280218767975

Adin A, Goicoa T, Ugarte MD (2019). Online relative risks/rates estimation in spatial and spatio-temporal disease mapping. *Computer Methods and Programs in Biomedicine.* **172**, 103–116. http://www.sciencedirect.com/science/article/pii/S0169260718318030

Anderson C, Lee D ,Dean N (2017). Spatial clustering of average risks and risk trends in Bayesian disease mapping. *Biom. J.* **59**, 41–56. https://doi.org/10.1002/bimj.201600018

Aswi A, Cramb SM, Moraga P, Mengersen K (2019). Bayesian spatial and spatio-temporal approaches to modelling dengue fever: a systematic review. *Epidemiology and Infection.* **147**, 33, 1–14. https://doi.org/10.1017/S0950268818002807.

Bhatt S, Gething P, Brady O. et al (2013). The global distribution and burden of dengue. *Nature.* **496** , 504–507.https://doi.org/10.1038/nature12060

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological).* **32**, 2, 192âĂŞ236. https://www.jstor.org/stable/2984812

Besag, J., York, J., and MollÃŋe, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics.* **43**, 1, 1–20.

Bivand R, Pebesma E, Gomez-Rubio V (2013). *Applied Spatial Data Analysis with R.* Second ed., Springer-Verlag, New York.

Bosq D (2000). *Linear Processes in Function Spaces.* First ed., Springer-Verlag, New York.

Bosq D, Blanke D (2007). *Inference and Prediction in Large Dimensions.* First ed., John Wiley & Sons, England.

Cabrera M, Taylor G (2019). Modelling spatio-temporal data of dengue fever using generalized additive mixed models. *Spatial and spatio-temporal Epidemiology.* **28**, 1–13. https://doi.org/10.1016/j.sste.2018.11.006

Carbajo AE, Schweigmann N, Curto SI, De Garín A, Bejarán R (2001). Dengue transmission risk maps of Argentina. *Tropical Medicine and International Health.* **6**, 170–183. https://doi.org/10.1046/j.1365-3156.2001.00693.x

Cattarino L, Rodriguez-Barraquer I, Imai N, Cummings DAT, and Ferguson NM(2020).Mapping global variation in dengue transmission intensity.*Science Translational Medicine*. (12), 528, 1–11. https://doi.org/10.1126/scitranslmed.aax4144

Eberly L E, Carlin B P (2000).Identifiability and convergence issues for Markov chain Monte Carlo fitting of spatial models. *Statistics in Medicine.* **19**, 2279–2294. https://doi.org/10.1002/1097-0258(20000915/30)19:17/18<2279::AID-SIM569>3.0.CO;2-R

Espejo-Montes RM (2014). Spatial Functional Statistics analysis of panel data. Doctoral Thesis. Universidad de Granada, Spain. https://digibug.ugr.es/bitstream/handle/10481/34199/24074287.eps?sequence=1&isAllowed=y

Estallo EL, Carbajo AE, Grech MG, Frías-Céspedes M, López L, Lanfri MA, Ludueña-Almeida FF, Almirón WR (2014). Spatio-temporal dynamics of dengue 2009 outbreak in Córdoba City, Argentina. *Acta Tropica.* **136**, 129–136.https://doi.org/10.1016/j.actatropica.2014.04.024

Gelman A. (2006). Prior distributions for variance parameters in hierarchical models. *International Society for Bayesian Analysis.* **3**, 515–533.

Jaya M, Hermawan E, Abdullah AS, Gede Nyoman I, Ruchjana B (2016). Bayesian Spatial Modeling and Mapping of Dengue Fever: A Case Study of Dengue Feverin the City of Bandung, Indonesia. *International Journal of Applied Mathematics and Statistics.* **54** ,3, 94–103. http://www.ceser.in/ceserp/index.php/ijamas/article/view/4303. Accessed 10 August 2019.

Knorr-Held, L.(2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine.* **19**, 2555–2567. https://doi.org/10.1002/1097-0258(20000915/30)19:17/18<2555::AID-SIM587>3.0.CO;2-%23.

Lawson, A.B.(2008). Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology. First ed., Taylor and Francis, Boca Raton, Florida, USA.

Lehmann EL, Casella G (1998). Theory of Point Estimation. Second ed, Springer, New York, USA.

Leroux BG, Lei X, Breslow N (1999). Estimation of disease rates in small areas: a new mixed model for spatial dependence, in: Halloran ME and Donald B (Eds.), Statistical Models in Epidemiology; the Environment and Clinical Trials. Springer Science+Business Media New York, USA, pp.179–192.

Lorenz C, Chiaravalloti-Neto F, de Oliveira Lage M, Quintanilha JA, Parra MC, Dibo MR, FÃąvaro EA, Guirado MM, Nogueira ML(2020).Remote sensing for risk mapping of Aedes aegypti infestations: Is this a practical task? *Acta Tropica*. **205**, 105398, 1–7. https://doi.org/10.1016/j.actatropica.2020.105398

Lowe R, Barcellos C, Coelho C, Bailey T,Coelho GE, Graham R, Jupp T, Ramalho W, Carvalho M, Stephenson D, Rodó J (2014). Dengue outlook for the World Cup in Brazil: an early warning model framework driven by real-time seasonal climate forecasts. *The Lancet Infectious Diseases.* **14**, 7, 619–626. https://doi.org/10.1016/S1473-3099(14)70781-9

Martínez-Bello D, López-Quílez A, Prieto AT (2018). Spatiotemporal modeling of relative risk of dengue dis-

ease in Colombia. *Stochastic Environmental Research and Risk Assessment.* **32**, 1587–1601. https://doi.org/10.1007/s00477-017-1461-5

Messina JP, Brady, OJ, Golding N. et al.(2019). The current and future global distribution and population at risk of dengue. *Nat Microbiol* **4** , 1508–1515. https://doi.org/10.1038/s41564-019-0476

Meza JL (2003). Empirical Bayes Estimation Smoothing of Relative Risks in Disease Mapping. *Journal of Statistics Planning and Inference*. **112**, 43–62. https://doi.org/10.1016/S0378-3758(02)00322-1

Morris M, Wheeler-Martin K, Simpson D, Mooney SJ, Gelman A, DiMaggio C. (2019). Bayesian hierarchical spatial models: Implementing the Besag York Mollie model in stan.*Spatial and Spatio-temporal Epidemiology.***30**,1–18. https://doi.org/10.1016/j.sste.2019.100301

Mukhsar, Bahriddin A, Asrul S, Edi C, Pasrun A, Farah AA (2016). Extended convolution model to Bayesian spatiotemporal for diagnosing the DHF endemic locations, *Journal of Interdisciplinary Mathematics.* **19**, 233–244. https://doi.org/10.1080/09720502.2015.1047591

Pan American Health Organization (PAHO). https://www.paho.org/hq/index.php?lang=en

Pastrana ME, Brito R, Romero NR, de Oliveira CS, Haddad JP (2014). Spatial and statistical methodologies to determine the distribution of dengue in Brazilian municipalities and relate incidence with the Health Vulnerability Index. *Spatial and Spatio-temporal Epidemiology.* **11**, 143–151. https://doi.org/10.1016/j.sste.2014.04.001.

Pepin KM, Leach CB, Marques-Toledo C, Laass KH, Paixao KS, Luis AD, Webb CT (2015). Utility of mosquito surveillance data for spatial prioritization of vector control against dengue viruses in three Brazilian cities. *Parasites and vectors.* **8**, 98. 1–15. https://doi.org/10.1186/s13071-015-0659-y

Restrepo AC, Baker P, Clements AC (2014). National spatial and temporal patterns of notified dengue cases, Colombia 2007âĂŞ2010. *Trop Med Int Health.* 19. 863–871. https://doi.org/10.1111/tmi.12325

Riebler A, SÃÿrbye SH, Simpson D, Rue H (2016).An intuitive Bayesian spatial model for disease mapping that accounts for scaling.*Statistical Methods in Medical Research*. **25**, 4, 1145–1165. https://doi.org/10.1177/0962280216660421

Ruiz-Medina MD, Salmeron R (2010). Functional Maximum-Likelihood Estimation of ARH(p) Models. *Stochastic Environmental Research and Risk Assessment.* **24**, 131–146. https://doi.org/10.1007/s00477-009-0306-2

Ruiz-Medina MD, Espejo RM, Ugarte MD, Militino AF (2014). Functional Time Series Analysis of SpatioâĂŞTemporal Epidemiological Data. *Stochastic Environmental Research and Risk Assessment.* **28**, 943–954. https://doi.org/10.1007/s00477-013-0794-y

SÃąnchez-GÃşmez A, Amela C, FernÃąndez-CarriÃşn E, MartÃŋnez-AvilÃľs M, SÃąnchez-VizcaÃŋno JM, Sierra-Moros MJ (2017) . Risk mapping of West Nile virus circulation in Spain, 2015. *Acta Tropica.* **169**, 163–169. https://doi.org/10.1016/j.actatropica.2017.02.022.

Torres A, Frías MP, Ruiz-Medina MD (2016). Log-Gaussian Cox processes in infinite-dimensional spaces, *Theor. Prob. Math. Stat.* **95**, 157–177. https://doi.org/10.1090/tpms/1028

Ugarte MD, Goicoa T, Militino AF (2009). Empirical Bayes and Fully Bayes procedures to detect high-risk areas in disease mapping. *Computational Statistics and Data Analysis*, **53**, 8, 2938 –2949. https://doi.org/10.1016/j.csda.2008.06.002

Wijayanti SP, Porphyre T, Chase-Topping M, Rainey SM, McFarlane M, Schnettler E, Kohl A (2016). The Importance of Socio-Economic Versus Environmental Risk Factors for Reported Dengue Cases in Java, Indonesia. *PLoS neglected tropical diseases.* **10**, 9, 1–15. https://doi.org/10.1371/journal.pntd.0004964

Yu H, Angulo JM, Cheng M, Wu J, Christakos G (2014). An online spatio temporal prediction model for dengue fever epidemic in Kaohsiung (Taiwan). *Biom. J.*, **56**, 428–440. https://doi.org/10.1002/bimj.201200270

Zhu G, Liu J, Tan Q, Shi B (2016). Inferring the Spatio-temporal Patterns of Dengue Transmission from Surveillance Data in Guangzhou, China. *PLoS neglected tropical diseases.* **10**, 4, 1–20. https://doi.org/10.1371/journal.pntd.0004633

A. ARH(1) Model

Let us consider now $Y_t \in H$, $t \in \mathbb{Z}$, a Hilbert centered process defined on the basic probabilistic space $(\Omega, \mathcal{A}, P)$, as an ARH(1) process, i.e., satisfying

$$Y_t(\mathbf{x}) = A\left(Y_{t-1}\right)(\mathbf{x}) + \upsilon_t(\mathbf{x}), \quad \forall \mathbf{x} \in D \subseteq \mathbb{R}^n,\ t \in \mathbb{Z}, \tag{5}$$

where υ is a strong Hilbertian white noise, see Ruiz-Medina and Salmeron (2010), whose autocovariance operator is given by $R_\upsilon = E(Y_t \otimes Y_t)$, with $t \in \mathbb{Z}$. Here, $\otimes$ represents the tensor product of functions in H. The spatio-temporal interaction operator A admits the spectral decomposition in terms of its eigenvalues and eigenvectors. Thus, $A = \Psi\Lambda\Phi^*$, in terms of the left eigenvector system $\{\psi_i, i \in \mathbb{N}\}$, and the right one $\{\phi_i, i \in \mathbb{N}\}$, where Λ is the diagonal operator defined by the sequence of eigenvalues $\{\lambda_i, i \in \mathbb{N}\}$. Then, from (5),

$$\Phi^* Y_t = \Lambda\Phi^* Y_{t-1} + \Phi^* \upsilon_t. \tag{6}$$

The structure of the second order Hilbert process, Y_t, is characterized in terms of the autocovariance and cross covariance operators. A truncated vector formulation is considered to obtain the renamed diagonal autoregressive equation (6). For a given truncation order $M \leq N$, and keeping in mind (5), the diagonal approximation of the ARH(1) is, for $t = 1, \ldots, T$,

$$\Phi^* Y_t = \mathbf{Z}(t) = \mathbf{\Lambda Z}(t-1) + \mathbf{u}(t), \tag{7}$$

where $\mathbf{Z}(t)$ is an $M \times 1$ vector with entries $Z_m(t) = \langle \mathbf{Y}_t,\ \phi_m \rangle_H$, $m = 1, \ldots, M$ (with $\langle \cdot,\ \cdot \rangle_H$ denoting the scalar product in the separable Hilbert space of integrable square functions, H), $\mathbf{\Lambda}$ is an $M \times M$ diagonal matrix with entries $\lambda_m,\ m = 1, \ldots, M$, and $\mathbf{u}(t)$ is an $M \times 1$ vector with entries $u_m(t) = \langle \mathbf{v}_t,\ \phi_m \rangle_H,\ m = 1, \ldots, M$. Taking into account equations (2) and (7), the Kalman filtering is applied:

$$\hat{\mathbf{Z}}(t \mid t) = \hat{\mathbf{Z}}(t \mid t-1) + \mathbf{K}_t \left(Z_t - \mathbf{\Phi}_M \hat{\mathbf{Z}}(t \mid t-1) \right), \tag{8}$$

where $\hat{\mathbf{Z}}(t \mid t) = E\left(\mathbf{Z}(t) \mid Z_t, ..., Z_1\right)$ and $\hat{\mathbf{Z}}(t \mid t-1) = E\left(\mathbf{Z}(t) \mid Z_{t-1}, ..., Z_1\right)$ are both related as the updated and previous projections estimate, respectively, of the random coefficients $Z_m(t)$, $m = 1, \ldots, M$, at time t. Here, Z_t is obtained from the observed values (2) at that time t. Operator $\mathbf{K}_t$ denotes the *gain operator*, reflecting the smoothing performed on functional data, see Torres et al. (2016) for more details.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author Statement

Torres-Signes, A: Conceptualization, Methodology, Writing, Software, Investigation.

Dip JA: Data curation, Writing, Visualization, Software, Investigation.