Spatial and Spatio-temporal Epidemiology 22 (2017) 27–37

Contents lists available at ScienceDirect

Spatial and Spatio-temporal Epidemiology

journal homepage: www.elsevier.com/locate/sste

Original Research

Modelling malaria incidence by an autoregressive distributed lag model with spatial component

Francisco Laguna[a], María Eugenia Grillet[a], José R. León[b], Carenne Ludeña[b,*]

[a] Laboratorio de Biología de vectores y Parásitos, Instituto de Zoología y Ecología Tropical. Facultad de Ciencias. Universidad Central de Venezuela, Los Chaguaramos, Caracas 1041, Venezuela

[b] Escuela de Matemática. Facultad de Ciencias. Universidad Central de Venezuela, Los Chaguaramos, Caracas 1041, Venezuela

ARTICLE INFO

Article history:
Received 23 February 2016
Revised 14 January 2017
Accepted 22 May 2017
Available online 8 June 2017

Keywords:
Malaria
Spatio-temporal models
Metropolis hastings
Autoregressive distributed lag models
Climatic variables

ABSTRACT

The influence of climatic variables on the dynamics of human malaria has been widely highlighted. Also, it is known that this mosquito-borne infection varies in space and time. However, when the data is spatially incomplete most popular spatio-temporal methods of analysis cannot be applied directly. In this paper, we develop a two step methodology to model the spatio-temporal dependence of malaria incidence on local rainfall, temperature, and humidity as well as the regional sea surface temperatures (SST) in the northern coast of Venezuela. First, we fit an autoregressive distributed lag model (ARDL) to the weekly data, and then, we adjust a linear separable spacial vectorial autoregressive model (VAR) to the residuals of the ARDL. Finally, the model parameters are tuned using a Markov Chain Monte Carlo (MCMC) procedure derived from the Metropolis-Hastings algorithm. Our results show that the best model to account for the variations of malaria incidence from 2001 to 2008 in 10 endemic Municipalities in North-Eastern Venezuela is a logit model that included the accumulated local precipitation in combination with the local maximum temperature of the preceding month as positive regressors. Additionally, we show that although malaria dynamics is highly heterogeneous in space, a detailed analysis of the estimated spatial parameters in our model yield important insights regarding the joint behavior of the disease incidence across the different counties in our study.

1. Introduction

Malaria is a disease caused by infection with the parasites of the genus *Plasmodium* and transmitted among human hosts through the bites of infected female *Anopheles* mosquitoes. Mosquito abundance, length of the development period of *Plasmodium* within the vector insect, *Anopheles* survival and reproduction, and the blood feeding frequency of the mosquito vector on host determine the risk and intensity of malaria infection in time and space (Macdonald, 1957). All these biological parameters, in turn, are largely dependent on climatic conditions such as temperature and rainfall (Stresman, 2010), which can simultaneously act at different scales. Consequently, epidemiologic patterns of malaria vary in space and time.

Previous studies of malaria epidemics in Neotropical areas have shown that spatial local transmission of this disease is highly heterogeneous (Grillet et al., 2010a, 2010b; Rodríguez et al., 2013), with disease areas varying from persistent transmission (hot spots) to moderate to low local transmission (cool spots) where the infection would disappear by itself if the area were isolated (Grillet et al., 2010a). As a function of time, malaria occurrence can be characterized by short-term seasonal fluctuation mainly

* Corresponding author Present address: Facultad de Ciencias, Universidad Antonio Nariño, Sede Circunvalar, Cra 3 Este # 47 A-15, Bogotá, Colombia.

E-mail addresses: carennec.ludenac@utadeo.edu.co, carinludena@uan.edu.co (C. Ludeña).

http://dx.doi.org/10.1016/j.sste.2017.05.001
1877-5845/

caused by the annual season changes in rainfall and longer-term cycles of around 2–6 years, accounted by the interaction of El Niño Southern Oscillation (ENSO) with local rainfall (Grillet et al., 2014). Particularly, these authors showed that rainfall mediates the effect of ENSO on malaria locally. This is, increased rains during this climatic event from the last phase of the season (September-November) had a critical role in the temporal dynamics of *Plasmodium*. Since an efficient control of this disease and prediction of its emergence or spread to new geographic regions will require an understanding of that heterogeneity, there is a growing need for analytical methods that account for the malaria dynamics in space and time.

An autoregressive model is a natural first appropriate approach for modeling the temporal dependence structure of the disease time series (Allard, 1998). However, if the climate-malaria incidence relationship varies spatially within a given region, then the use of a single model for the whole region could lead to a worse model fit in some areas than others, which, in turn, would be manifested as spatial dependence in the model residuals (MJ and MRT, 2005). Consequently, there is a need to investigate if, and how, temporal variations in malaria infections could be related to exogenous environmental factors by taking into account both scales: time and space.

There is a very extensive literature on spatio-temporal methods for malaria both in terms of a general modeling perspective and specific malaria models (e.g., Banerjee et al., 2008, Bi et al., 2003; Briët et al., 2008; Gomez-Elipe et al., 2007; Ippoliti et al., 2012; Teklehaimanot et al., 2004; Zhou et al., 2004). In particular, we cite Ippoliti et al. (2012) and Banerjee et al. (2008) as important references for general spatio-temporal models like the ones considered in this work. The first authors consider a dynamic factor model to study the relationship between spatio-temporal processes. The model is then used to identify clusters of locations whose behavior is described by a potentially small set of common dynamic latent factors. The second authors consider hierarchical Gaussian processes. Although specifically designed to deal with large data sets, their methodology based on the introduction of simpler (lower dimensional) "predictive processes" is quite appealing in general. Both papers then use MCMC procedures in order to estimate the model parameters. For the specific problem of modeling malaria, we highlight the work by Rumisha et al. (2014) and Edlund et al. (2012). The former considers hierarchical models as those developed by Banerjee et al. (2008), for female mosquito counts taking into account environmental exogenous factors and seasonal, temporal and spatial random effects and appropriately choosing a subset of locations. The latter compares malaria incidence response to fluctuations in historic climatic data taking into account spatial variability to measure sensitivity of the response to certain climatic variables.

In this article, our main goal is to design a procedure in order to fit a spatio-temporal model predicting malaria incidence taking into account a heterogeneous geography and the non-availability of spatially varying climatic variables across the whole region of study, an endemic-epidemic-prone region in Northern South America, involving several counties. Lack of complete spatial information at a county scale does not allow us to use the models and procedures described above. Completing available information based on NOAA or related spatially detailed data is not an option at the scales of interest (counties). Although it would be possible to use average information over the region of interest (instead of data from a single observation site) this would not acknowledge for observed spatial heterogeneity.

To deal with this problem we adopt a two-stage procedure. In a first step, we adjust an autoregressive distributed lag model (ARDL) to explain the temporal variation of malaria infection based on the exogenous local variables humidity, precipitation, and temperature as well as the sea surface temperatures of the eastern and central tropical Pacific as an index of the El Niño Southern Oscillation (ENSO) regional phenomenon. This will be considered as our null or base model. In a second step, we adjust a spatio-temporal model to the residuals of the ARDL model to include spatial dependence. Given the absence of exogenous data across space, it is not possible to follow a dynamic factor model approach as in Ippoliti et al. (2012) or a hierarchical model as in Banerjee et al. (2008). Our approach uses a Markov Chain Monte Carlo (MCMC) procedure, described in detail below, to sample from an appropriate posterior distribution for the parameters assuring compliance to certain restrictions on the model. Model fit is then assessed by looking at the model's residuals variance (for each county) and compliance to the model's assumptions. As a by product, analysis of the model's spatial parameters (parameters defining the spatial dependence) provides valuable insights regarding the spatial behavior of the Malaria incidence rate (MIR) in the considered sectors.

2. Methodology

2.1. Study site

The study was carried out in the southern coastal lowland areas of the Sucre State, North-eastern Venezuela (see Fig. 1) where *Anopheles aquasalis* is the main vector of *P. vivax* (Grillet, 2000). Here, the annual mean temperature is 27° C–28° C and total annual rainfall is 1200–1700 mm, with a rainy season from May to November and a dry season from December to April.

2.2. Epidemiological and climatic data

Weekly cases of malaria (2001 to 2008 for a total of $N = 416$ times) from 10 Municipalities grouped in 12 nearby counties (Parroquias) in northern Venezuela were obtained from the Malaria Control Program database, Venezuelan Ministry of Health and the Malaria incidence rates (MIR) for each studied administrative area (no. of new cases /population at risk per time) were calculated.

2.3. Climatic data

Contemporaneous local climate data were obtained from the nearest meteorological station (Guiria station: 10° 34' N 62° 17' W). Data included mean temperature,