

Bayesian Spatiotemporal Projection of Chagas Disease Incidence in Brazil

Ethan Roubenoff¹

¹*Department of Demography, University of California, Berkeley*

Abstract

Chagas Disease is a parasitic infection caused by the *T. Cruzi* parasite endemic to Central and South America and transmitted through contact with *Triatomine* insects, commonly known as “kissing bugs.” Although the symptoms of Acute Chagas Disease (ACD) are nonspecific, untreated chronic infection can lead to heart disease, enlarged esophagus and colon, and stroke. Chagas disease has become increasingly rare owing to a series of public health interventions, including insect eradication campaigns in Brazil through the 1980’s that considerably reduced the number of new acute cases. However, hundreds of new acute cases still are diagnosed annually, primarily in the states of Pará, Amapá, and Acre. Moreover, the population in areas of high Chagas endemicity are changing: many areas are growing and becoming increasingly urban, whereas others are decreasing in population. We estimate the Incidence Rate (IR) for Acute Chagas disease over the period 2001-2019 in Brazil at the municipal level and investigate the variation of these rates with climatic factors. These estimates are used to project forward incidence of Acute Chagas Disease over the following decade 2020-2029. Modeling ACD presents numerous methodological challenges since incidence is rare, with extreme overdispersion of zero-case counts, and vectors exhibit a highly spatially- and temporally-clustered pattern. We use a spatially- and temporally-autoregressive small-area smoothing models to estimate the true latent risk in developing Acute Chagas Disease. The Bayesian model presented here involves spatio-temporal smoothing via a Zero-Inflated (Lambert 1992), Knorr-Held (2000)-Type spatio-temporal model with a BYM2 (Morris, 2019) spatial convolution to predict smoothed incidence rates of Chagas disease. As well, we include estimates of Brazil’s growing population and projected bioclimate to evaluate how climate and population change may affect ACD rates. We estimate that cases will continue to increase in the absence of control efforts, primarily driven by a growing peri-urban population in regions of Chagas endemicity.

1 Introduction

Chagas disease is a vector-borne parasitic infection in humans caused by the *T. Cruzi* parasite, and is transmitted to humans primarily through contact with infected *Triatomine* insects commonly known as “kissing bugs” (WHO Expert Committee on the Control of Chagas Disease 2002; Pérez-Molina and Molina 2018; Canals et al. 2017). Transmission of *T. Cruzi* to humans can occur following bites from infected kissing bugs or contact with their feces; human-human transmission can occur via blood transfusion and

33 congenitally from mothers to children via the vertical pathway. Over 80% of transmission occurs from human-
34 vector contact, with congenital transmission responsible for nearly all other new infections; screening of blood
35 donation has nearly eliminated all transmission from transfusions (World Health Organization 2015). Acute
36 symptoms of Chagas disease include fever, inflammation of the infection site, eyelid edema, and swollen lymph
37 nodes and tonsils. Acute symptoms resolve spontaneously over 4-8 weeks and treatment during the acute
38 phase with antiparasitic medication is highly effective at curing infection (Bern et al. 2007). However, since
39 acute symptoms are generally nonspecific and the burden of infection affects many communities lacking
40 affordable access to high-quality healthcare, many acute cases go undiagnosed and untreated. Untreated
41 chronic Chagas infection can cause cardiomyopathy, megacolon, stroke, and megaesophagus in 30-40% of
42 patients in the 10-30 years following acute infection (Pérez-Molina and Molina 2018; Sosa-Estani and Segura
43 2015). Although Chagas is fairly rare—acute infection incidence is on the order of a few hundred infections
44 annually in Brazil for a population of around 200 million—it is this possibility of chronic complications,
45 especially cardiomyopathy, that makes early identification and control of Chagas an important public health
46 concern.

47 Chagas disease is also known as a disease of poverty affecting mostly poor and indigenous rural com-
48 munities in South America (Fernández, Gaspe, and Gürtler 2019; Sosa-Estani and Segura 2015; Dias 1987;
49 Tarleton et al. 2007). Many people are at increased risk of infection due to the use of certain residential con-
50 struction materials hospitable to *Triatomine* infestation, especially untreated wood. People working certain
51 jobs that entail contact with *Triatomine* habitats—including forestry or agriculture—may be at increased
52 risk of exposure during employment. A number of non-pharmaceutical interventions can alleviate much of
53 the probability of contact, including insecticide usage, bed netting, and removal of certain residential con-
54 struction materials known to be *Triatomine* habitats. Intervention campaigns through the 1980's focused
55 directly on the class dynamics of Chagas risk by implementing control efforts across the sociodemographic
56 ladder (Dias 1987), yet many persisting Chagas hotspots occur in poor and rural parts of Brazil. Controlling
57 the incidence of Chagas remains an important issue of equity.

58 Since then, Public health efforts to eliminate new Chagas infections and pharmaceutical developments
59 to treat latent chronic infections and complications have been successful at reducing the burden of Chagas
60 disease in Brazil (World Health Organization 2015; Sosa-Estani and Segura 2015). The two primary methods
61 of transmission—human-vector contact, either through bites or contact with vector feces, and the vertical
62 pathway from mothers to infants—have have required vastly different interventions, both with success. Early
63 studies dating to the late 1940s found that continuous use of residential insecticides was highly effective at
64 eliminating transmission, indicating that residential contact with *Triatomines* may have been responsible for
65 the majority of transmission risk (Dias 1987). Eradication programs by Brazil's SUCAM (Superintendencia
66 de Campanhas de Saude Publica) in the 1980s involved identifying areas of risk by sampling insects in
67 homes and generating maps of high risk locations. All homes within more than 700 high-risk municipalities,
68 regardless of known infestation, were sprayed with insecticide every 3-6 months until under 5% of homes
69 were found to be infested with any insects and no *Triatomines* were found in any homes. Overall, more than
70 5 million homes were sprayed with insecticide, resulting in a 73% reduction in the number of infested homes
71 by 1986 (*ibid.*) and the total elimination of transmission by *T. infestans*—previously the vector responsible
72 for most transmission—resulting in a 94% reduction in new acute cases (Gurgel-Gonçalves et al. 2012).

73 Congenitally-transmitted Chagas disease via the vertical pathway is less frequent than transmission via

74 contact with *Triatomines* (World Health Organization 2015). Screening for Chagas Disease among potential
75 mothers living in high-risk areas and initiating treatment in advance of pregnancy is ideal for reducing the
76 probability of successful vertical transmission, although commencing treatment after pregnancy appears to
77 be well tolerated by both the mother and the fetus (Cevallos and Hernández 2014). While it is not currently
78 possible to entirely eliminate vertical transmission, treatment of infants with suspected Chagas infection
79 within the first year of life is very successful at eliminating the disease from children (Carlier et al. 2011;
80 Moya, Basso, and Moretti 2005). Transmission may occur at any time during pregnancy, but is theorized to
81 be more likely to occur during the second and third trimesters (*ibid*). Most congenital transmission occurs
82 from mothers who are in the chronic phase of disease, however vertical transmission has been documented
83 from mothers who are acutely infected at conception or become infected during the course of pregnancy.
84 It has been proposed that the level of parasitemia of the mother may affect the probability of vertical
85 transmission and the severity of infection at birth (Carlier et al. 2011).

86 Antiparasitic medications benzniazole and nifurtimox have proven efficacy against Chagas disease, and
87 the former is usually recommended for treatment (Bern et al. 2007). If treated in the acute phase, complete
88 parasitological cure can occur in 60-85% of vector-transmitted infections and more than 90% of congenital
89 infections when treatment is administered within the first year of life (Altcheh et al. 2011; Carlier et al.
90 2011; Cevallos and Hernández 2014; Moya, Basso, and Moretti 2005). If Chagas disease is left untreated
91 until the chronic phase, treatment is less effective— only 60% of participants achieved negative serology
92 within 3-4 years. Even if not resulting in a complete cure, treatment may slow the development of Chagas
93 cardiomyopathy and other potentially lethal complications, and treatment is recommended for all patients
94 presenting with positive serology (*ibid.*).

95 Despite progress towards elimination, there are still an estimated 1-4.6 million people currently infected
96 with chronic Chagas disease and approximately 6,000 deaths per year (Simões et al. 2018). Pérez-Molina and
97 Molina 2018 estimate that in 2010, over 70 million people were at risk of contracting Chagas disease across 21
98 countries in Latin America, and 38,593 new infections were reported that year. This count of acute infections
99 down considerably from 55,585 in 2005 and from more than 700,000 between 1980-1985. Most individuals
100 living with Chagas disease are located in three countries—Argentina, Brazil, and Mexico—and the most new
101 infections were reported in Bolivia (World Health Organization 2015). While preventative interventions have
102 brought the new infection rate down considerably in Brazil, the rate of decrease has not necessarily been
103 equal across the country. The WHO estimates that as of 2015, the incidence of new Chagas infections in
104 Brazil via human-vector contact was 0.084 per 100,000 population and via congenital transmission 0.020 per
105 100 live births (*ibid.*). Our analysis aides in identifying areas where future interventions can further alleviate
106 risk of the disease.

107 The distribution of *triatomines* is a highly spatial process within endemic areas, and as a result risk
108 of contracting Chagas disease is a complex interaction of local vector population, local human population,
109 and interaction between the two. Despite the elimination of *T. infestans*, previously the vector responsible
110 for the most new cases of Chagas Disease, 62 known species of *Triatomines* in Brazil are responsible for
111 transmission; some, including *Panstrongylus geniculatus* and *P. megistus*, are widespread over the country,
112 whereas others are more localized spatially (Gurgel-Gonçalves et al. 2012). Certain biomes, including the
113 Cerrado tropical savannah in the central region and Caatinga shrub forest in the northeast, have a higher
114 diversity of species. Not all species of *Triatomines* are equally likely to transmit Chagas disease to humans;

115 for example, while the most epidemiologically relevant species may be *P. megistus*, which is common in
116 domiciles and a frequent carrier of *T. cruzi*, the behavior and habitat of *T. sordida* is more likely to result
117 in contact from agricultural activities but unlikely to result in residential contact. Gurgel-Gonçalves et al.
118 2012 remark that nearly all areas of Brazil have some risk of Chagas disease, but certain regions, especially
119 the Cerrado and Caatinga, present more risk.

120 Climate change presents an ambiguous threat to incidence of Chagas disease. Tamayo et al. 2018 find
121 that *Triatomine* vectors of Chagas disease may exhibit increased fecundity and egg viability in warmer
122 temperatures. *T. Cruzi* as well may exhibit increased viability at warmer temperatures, suggesting that
123 incidence of Chagas disease will likely increase with climate change. Medone et al. 2015 find that the
124 changing climate will likely create more geographic areas that are suitable habitats for Chagas vectors
125 correlates with the force of infection of acute Chagas disease in Argentina and Venezuela. They find that
126 warmer temperatures are unfavorable to vectors; although current Chagas hotspots may see decreases with
127 increased temperatures, the overall geographic distribution of vectors may shift as previously too-cold areas
128 warm.

129 Since Chagas disease is primarily found in rural areas in Brazil, multi-decadal trends in the urban and
130 rural population may be a mediator in the future trajectory and control of Chagas disease (Delazeri, Da
131 Cunha, and Oliveira 2022; Perz 2000; Randell and VanWey 2014). Brazil has become increasingly urbanized
132 since the 1960s; internal migrants from rural, Chagas-endemic areas have resulted in identification of both
133 acute and chronic cases of Chagas disease in places where Chagas is not historically found (Coura and
134 Borges-Pereira 2010; Martins-Melo et al. 2014; Moncayo and Silveira 2009). Like all urban areas in Brazil,
135 larger municipalities in Chagas-endemic areas have been increasing in size faster than the nearby rural areas,
136 many of which have even seen population declines. The identification and control strategies of the 1980s that
137 targeted known areas of Chagas endemicity may not be as effective for identifying latently infected internal
138 migrants who have moved to the larger cities outside of endemic strategies. As well, declines in the rural
139 population are fundamentally changing the spatial distribution of new cases of Chagas disease. Although
140 the rural population is declining overall, declines are not uniform across all areas of Chagas endemicity, and
141 projections of future cases must include population change as well.

142 To identify areas of persisting and future Chagas endemicity, we borrow Bayesian disease mapping meth-
143 ods for modeling higher-incident spatially-clustered diseases, such as cancer (Best, Richardson, and Thomson
144 2005; Napier et al. 2019; Riebler et al. 2016; Wakefield 2007; Wikle, Berliner, and Cressie, Noel 1998; Knorr-
145 Held 2000; Knorr-Held and Best 2001) and adapt these methods to suit the highly rare nature of Chagas
146 disease. Since the rates of ACD are low—on the order of hundreds of cases annually for a population of
147 over 200 million—and present in a highly clustered pattern in certain regions of Brazil, a specialized model-
148 ing approach is needed to highlight the spatio-temporal structure of Chagas disease. We use a Knorr-Held
149 (2000) spatio-temporal model adapted with a rare count, zero-inflated model (Lambert 1992; Lee et al. 2016;
150 Rathbun and Fei 2006; Ver Hoef and Jansen 2007) and include climate covariates and population growth
151 to analyze how patterns of Chagas disease might play out for the ensuing decade. We find that with an
152 increasing population and climate trends, it is likely that cases of Chagas will continue to increase in the
153 absence of additional intervention, potentially as high as doubling in incidence.

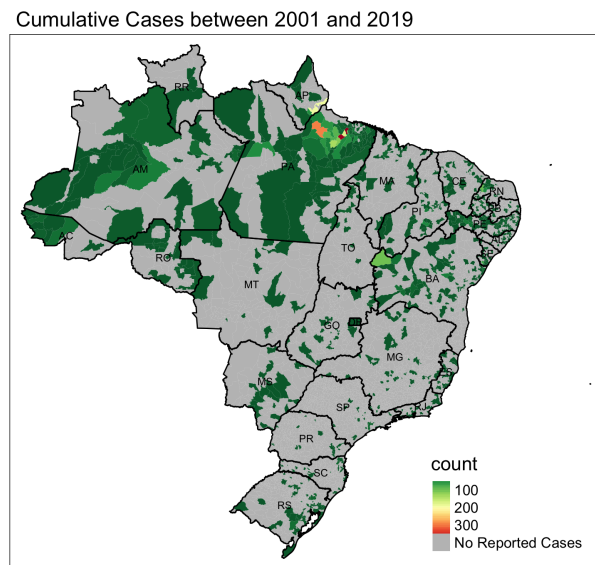


Figure 1: Counts of Chagas disease between 2001 and 2019 at the municipality level. Of the 5568 municipalities in Brazil, 4472 municipalities reported no cases of ACD during this period.

2 Methods

2.1 Data

Counts of Acute Chagas Disease (ACD), aggregated by municipality¹ of residence and year between 2001 and 2019, are collected by the Ministry of Health's Departamento de Informática do Sistema Único de Saúde (DATASUS; Department of Informatics of the Unified Health System) and retrieved from the agency's TABNET database (Ministério da Saúde, Brasil 2023). We chose to use municipalities—the finest level of geographic aggregation available—in order to analyze spatial variability that may be lost at larger levels of geographic aggregation, like state or region. Over the period 2001-2019, 5568 cases of Acute Chagas Disease were reported among residents of 826 municipalities, with the highest counts in the northern states² of Pará and Amapá. The municipality-specific total counts of Acute Chagas Disease reports are displayed in figure 1. Official population estimates at the municipality level are taken from Brazil Instituto Brasileiro de Geografia e Estatística (IBGE)'s annual population estimates for 2001-2006, 2008-2009, and 2011-2019 and the 2010 census counts (Instituto Brasileiro de Geografia e Estatística 2023). No data are present for 2007; population for this year is taken as the linear interpolation between 2006 and 2008. Climate data are retrieved from the European Union's Copernicus Climate Change Service (C3S) Climate Data Store (CDS) Global Bioclimatic Indicators from 1950-2100 Derived from Climate Projections (Wouters et al. 2021), which contains a suite of 19 bioclimatic variables averaged annually. These variables are the same as those in the WorldClim (Fick

¹ *Município* is translated to English as municipality, but are functionally closer to US Counties by population size, geographic size, and governance.

² Brazil has 27 Federative Units (*unidades federativas*, abbreviated as UFs), consisting of 26 states and one federal district (Brasília). Here, we refer to all 27 UFs as states.

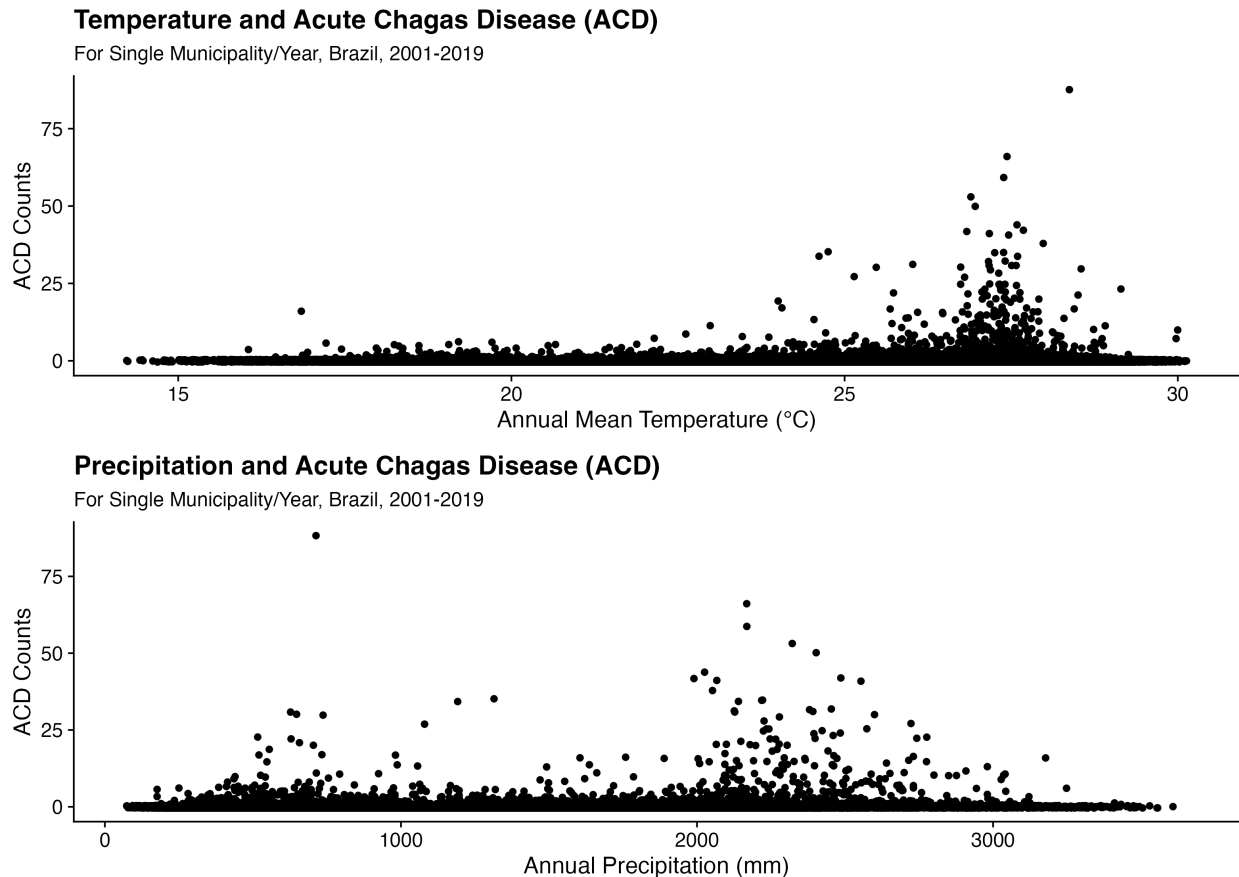


Figure 2: Counts of Acute Chagas Disease (ACD) by mean annual temperature and total precipitation, two of the 19 bioclimatic variables used in the analysis, displayed over the period 2001-2019.

171 and Hijmans 2017) dataset³, and are listed in supplementary table S2; a selection are displayed in figures 2
172 and 3. Broadly, we see in figure 2 that most reports of ACD occur in areas that are warmer (annual mean
173 temp > 25°C) and wetter (annual precipitation above 2000 mm), although many cases do occur in drier
174 climates. In Pará and Amapá, the UFs where most cases of ACD occur, there is a slight trend towards
175 warmer and wetter weather, although substantial year-to-year variations present (figure 3). Projections are
176 performed using the GFDL-ESM2M (NOAA, USA) algorithm.

177 2.2 Conditionally AutoRegressive (CAR) statistical models for disease inci- 178 dence data

179 Many classes of geospatial models for areal data (polygons, like municípios) exist, a few of which are discussed
180 here. Distributional models used in Bayesian modeling can be divided into two groups: Conditionally
181 AutoRegressive (CAR) models, that describe probability for observations as conditional on their neighbors,
182 and Simultaneous AutoRegressive (SAR) models, that use a matrix of adjacency weights to adjust for spatial
183 dependence. At a high level, these two models differ in the scale of spatial dependence: the CAR model

³<http://www.worldclim.org/>

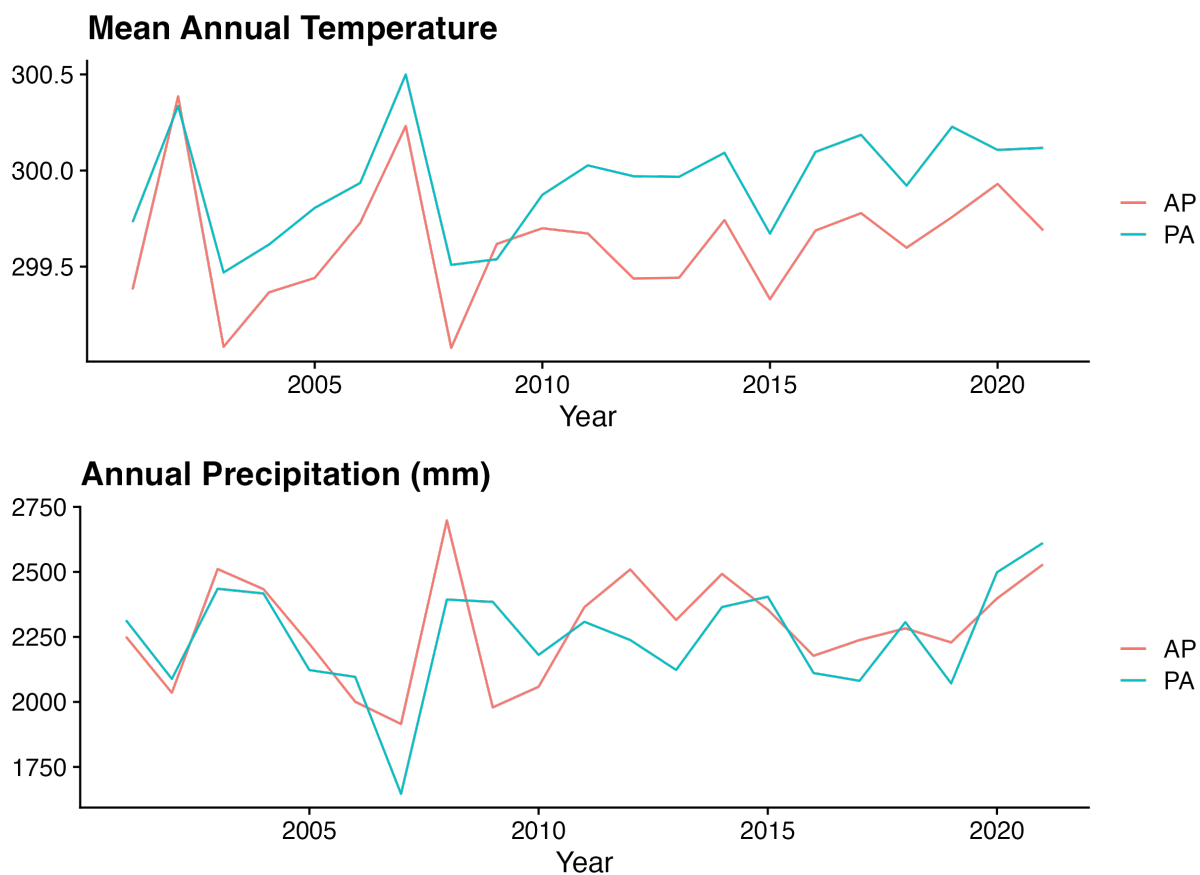


Figure 3: Mean annual temperature and total precipitation, two of the 19 bioclimatic variables used in the analysis, displayed over the period 2001-2030 for the two states with the highest incidence of ACD, Pará (PA) and Amapá (AP).

184 involves local smoothing, where the SAR model involves global smoothing. Here, we use a CAR model in
185 our application to rates of Chagas disease, which is a highly local process found only in certain relatively
186 isolated regions of Brazil and where global correlative structure is likely to over-smooth small-area variation.
187 More commentary is provided in the supplementary material. As well, CAR models have a computational
188 advantage over SAR models: they do not require matrix inversion, which can be computationally expensive
189 or impossible when modeling thousands of small-area samples⁴.

190 The simplest implementation of the CAR model is the Intrinsically AutoRegressive (IAR or ICAR) model,
191 also called the BYM model after authors Besag, York, and Mollie (1991). For a general Gaussian spatial
192 process ϕ , the CAR model is conditionally specified for each geographical unit as a normal distribution with
193 expectation equal to the average of its neighbors and variance τ :

$$\phi_i | \phi_{j \sim i} \sim N \left(\frac{1}{n_i} \sum_{j \sim i} \phi_j, \frac{\tau^2}{n_i} \right) \quad (1)$$

194 Where ϕ_i is an observation at the i th spatial unit, $\phi_{j \sim i}$ indicates the set of observations among the neighbors
195 j of i , and n_i is the number of neighbors of i . Throughout, we refer to equation 1 as the CAR and IAR models
196 interchangeably. The IAR distribution can also be extended to Poisson, Binomial, and Logistic distributions
197 as well (Besag, York, and Mollie, 1991; Haining, 2004).

To utilize the CAR distribution in a disease modeling context, Banerjee, Carlin, and Gelfand (2015)
recommend using a pair of *random effects* for the standardized incidence rates of disease:

$$Y_i | \psi_i \stackrel{iid}{\sim} \tau \text{Poisson}(Pop_i e^{\psi_i}) \quad (2)$$

$$\psi_i = \mathbf{x}'_i \boldsymbol{\beta} + \theta_i + \phi_i \quad (3)$$

198 where \mathbf{x} and $\boldsymbol{\beta}$ are vectors of spatially-varying covariates, θ_i captures heterogeneity with an *iid* normal prior
199 $N(0, 1/\tau_h)$ and ϕ_i captures spatial clustering with prior $CAR(\tau_c)$ as in equation 1. Here, parameters τ_h
200 and τ_c represent precision. Dividing extra-Poisson variability into ‘heterogeneity’ and ‘clustering’ poses a
201 problem: should τ_h and τ_c be too large, the model will be unable to identify the two random effects. Indeed,
202 priors on variance must be carefully chosen in order to allow for identifiability of θ and ϕ , which poses an
203 existential question as to the utility of these models in the first place. Instead, other specifications including
204 Leroux, Lei, and Breslow 2000 and the very closely related BYM2 model (Riebler et al. 2016) as implemented
205 by Morris et al. 2019, which introduce a convolution of the spatial and aspatial error terms which allows
206 for identifiability. The BYM2 model is a similar Poisson-GLM framework to the original BYM model, but
207 replacing the pair of random effects with a convolved error term:

$$\phi + \theta = \left((\sqrt{\rho/s})\phi^* + (\sqrt{1-\rho})\theta^* \right) \sigma \quad (4)$$

208 where $\rho \in [0, 1]$ represents the proportion of variance that comes from the spatial clustering random effect and
209 how much comes from the heterogeneity random effect; ϕ^* is the ICAR distribution; $\theta^* \sim N(0, n)$ where n is

⁴Related to the SAR models is the field of Spatial Econometrics (see Anselin 2003 for an overview). Spatial Econometrics uses the SAR model in a maximum likelihood regression framework. Many related specifications form the suite of Spatial Econometric models, each with their own implication about spatial correlative structure and dependence (Golgher and Voss 2016)

210 the number of connected subgraphs (in our application $n = 1$), s is the scaling factor such that $Var(\phi) \approx 1$
211 (Riebler et al. 2016); and $\sigma > 0$ is the overall standard deviation for the combined error terms (Morris
212 et al. 2019⁵). The BYM2 model improves upon the original form by allowing for independent definitions
213 of the two prior distributions without involving ρ in the sampling process as is done for the proper CAR
214 model. In doing so, this model involves the identification of only a single set of random effects rather than a
215 pair of independent random effects, which improves identification by separating the dependency structure.
216 Further, this avoids the need for informative priors in the BYM model as emphasized by Banerjee, Carlin,
217 and Gelfand 2015. As well, in this context ρ has an informative interpretation, although it still does not map
218 onto other indicators like Moran’s I. Morris et al. report that Stan’s Hamiltonian Monte Carlo (HMC) and
219 No U-Turn Sampler (NUTS) provide faster and more precise inference with the BYM2 model than other
220 samplers like WINBUGS and JAGS. The related Leroux (2000) model, which is similar to the BYM2 model
221 but specifies the neighborhood matrix differently, has been shown through simulation to be superior to the
222 original BYM model, and is employed by many in disease mapping studies (Lee, 2011).

223 2.2.1 Extending the BYM model to include temporal effects

224 We follow the Knorr-Held (Knorr-Held 2000) framework for Bayesian spatio-temporal modeling. The Knorr-
225 Held model adds time structure in a way that mirrors the BYM model (3) by adding temporally autoregressive
226 effects α , temporal random effects γ , and a spatio-temporal interaction term δ :

$$Y_i | \psi_i \stackrel{iid}{\sim} \text{Poisson}(E_i e^{\psi_i})$$
$$\psi_i = \mu + \gamma_t + \alpha_t + \theta_i + \phi_i + \delta_{it} \quad (5)$$

227 Where μ is the overall intercept, γ is an unstructured temporal component distributed $N(0, \sigma_\gamma)$, α is
228 a structured temporal component that can be specified as an AR(1) or AR(2) process, and δ is a spatio-
229 temporal interaction term. ϕ and θ are as described above. Effectively, the Knorr-Held model decomposes
230 the overall pattern into a global temporal trend, a global spatial trend, and an interaction term between the
231 two, in a procedure similar to ANOVA. Prior choice of δ is not straightforward, and requires careful thought
232 about the relationship of space and time in the model. Knorr-Held (2000) lays out four types of priors,
233 depending on the hypothesized interaction of the spatial and temporal dimensions: (I) where all interaction
234 terms are *a priori* independent; (II) where interactions are autoregressive in time but independent in space;
235 (III) where interactions are autoregressive in space but independent in time; and (IV) where interactions are
236 totally dependent in both space and time. Further description of these interaction types is given in appendix
237 section S4. Knorr-Held’s (2000) evaluation includes specification of the same disease model with each of the
238 four interaction types, and evaluation of the resultant model by DIC.

⁵In their 2019 paper, Morris et al use a logit-normal prior for ρ , which has mass around either extreme, indicating that the value of ρ should be close to 0 or 1 and is less likely to be in the middle. However, in a 2018 case study predating the publication, the same authors use a $Beta(1/2, 1/2)$ prior, which has a similar U-shape.

2.2.2 Zero-Inflated Poisson models for rare counts

Since Chagas disease is very rare, most entries in our matrix of counts by municipality and year are zero. While a low-rate Poisson may be able to capture this overdispersion of zeros, a more appropriate specification involves the zero-inflated Poisson model (ZIP; Lambert 1992). The zero-inflated Poisson is a mixture model that includes a Bernoulli process generating zeros and a Poisson process that generates counts (but may also generate some zeros). In this way, the zero counts in the data are effectively split into ‘structural’ zeros, which are generated from the presence or absence of the process of interest, and ‘sampling’ zeros, which are true random zero-counts in the presence of the Poisson process. Lambert (1992) note that in simulation, Poisson-only models are sufficient for a dataset that contains at most 68% zeros and 3.4% counts greater than 9, and that the ZIP model may be justified on datasets with higher rates of zeros. In our application to Chagas disease, over 99% of municipality-years have a zero count; nonzero entries have an average of 1.65 (95% CI: 1-7) infections. The ZIP distribution is parameterized by Bernoulli probability π and Poisson rate λ :

$$P(y_i|\pi, \lambda) = \begin{cases} \pi + (1 - \pi) \cdot \text{Poisson}(0|\lambda) & \text{if } y_i = 0 \\ (1 - \pi) \cdot \text{Poisson}(y_n|\lambda) & \text{if } y_i > 0 \end{cases} \quad (6)$$

The ZIP distribution is appropriate in a GLM framework, where parameters π and λ are specific to each observation y_i and are estimated with logit and log link functions, respectively. When writing the probability statement, we can also take advantage of the fact that $\text{Poisson}(0|\lambda) = \lambda^0 e^{-\lambda}/0!$ simplifies to $e^{-\lambda}$, clarifying the condition where $y = 0$:

$$P(y_i = 0) = \pi_i + (1 - \pi_i)e^{-\lambda_i} \quad (7)$$

$$P(y_i = k) = (1 - \pi_i)e^{-\lambda_i} \lambda_i^k / k! \quad (8)$$

In turn, the central moments of the ZIP distribution are mean $(1 - \pi)\lambda$ and variance $\lambda(1 - \pi)(1 + \pi\lambda)$ (Lambert, 1992).

The ZIP distribution then prompts an additional modeling decision. When used in a GLM framework, covariates can be added to both the Poisson process and the Bernoulli process. Prior studies have done both: Agarwal, Gelfand, and Citron-Pousty 2002 include a spatially-autocorrelated Poisson process, Rathbun and Fei 2006 include a spatially-dependent Bernoulli process, and Ver Hoef and Jansen 2007 include spatial (and temporal) autocorrelation in both parts. For our application to Chagas disease, it is not clear if the spatial process should be included in either the Bernoulli or Poisson process, or in both parts. Is the probability of any appearance of Chagas disease spatially correlated? Undoubtedly, as the primary determinant in risk of Chagas disease is the highly localized distribution of *T. Cruzi*. However, conditional on the presence of *T. Cruzi*, it is less clear *a priori* if the risk of contracting Chagas also a spatially dependent process.

2.3 Model: Estimating rates of ACD with zero-inflation and spatial and temporal autoregression

Here, we integrate the components discussed above into a single Bayesian model that allows for spatial and temporal autoregression as well as overdispersion of zeros. Specifically, we use a ZIP likelihood within a

271 spatio-temporal decomposition framework like the one proposed by Knorr-Held (2000). As well, we innovate
 272 by introducing the BYM2-type convolution of the unstructured error and spatially-structured heterogeneity
 273 to improve identification on the model posed by Knorr-Held (2000). We encountered convergence issues
 274 when including the spatial convolution in the Poisson process; as a result; this process is defined in the
 275 Bernoulli parameters only. In the Poisson process, we include a grand mean, temporally AR(1) time trend,
 276 and a spatial fixed effect with a Knorr-Held Type 1 interaction. We chose the probabilistic programming
 277 language and software suite `Stan` to estimate the yearly incidence risk of Chagas disease across all munic-
 278 ipalities in Brazil between 2000 and 2019. The model is evaluated in `Stan 2.20` using the `cmdstanr` (Stan
 279 Development Team 2023) interface for the R programming language, version 4.20. `Stan` was chosen for its
 280 speed relative to other probabilistic programming languages, like `GeoBUGS` (Lunn, Arnold, and Spiegelhal-
 281 ter 2004) or `JAGS` (Plummer 2003), especially for its ability to evaluate vectorized probability statements.
 282 Although `Stan` lacks the built-in support for spatial models present in `BUGS`, the computational gains from
 283 vectorization and adaptive sampling allow for quick evaluation and convergence of complicated posteriors,
 284 with full implementation details elaborated in the supplementary material. To sample Bayesian posteriors,
 285 by default `Stan` uses the No U-Turn Sampler (NUTS), a variant of Hamiltonian Monte Carlo, in contrast to
 286 Gibbs sampling used by `BUGS` and `JAGS`.

287 We run two formulations of the model: first, a non-covariate smoothing model used purely to recover
 288 latent rates of Chagas disease unadjusted for other causal factors besides population at risk; and second,
 289 a model that includes climatological covariates. The two models differ only in the inclusion of the set of
 290 covariates.

291 Beginning with ZIP-distributed likelihood:

$$Y_{ti} | \pi_{ti}, \lambda_{ti} \sim ZIP(\pi_{ti}, Pop_{ti} \cdot \lambda_{ti}) \quad (9)$$

Where indices t and i refer to year t between 2000 and 2018 and municipality i between 1 and 5561, the number of municipalities in Brazil. Assuming a logit link for Bernoulli parameter π and Poisson parameter λ , we take the follow GLM equations for π and λ :

$$\text{logit}^{-1}(\pi_{ti}) = \mu_{\pi} + \alpha_{\pi,t} + \beta_{\pi} x_{t,i} + \left((\sqrt{\rho_{\pi}/s}) \phi_{\pi,i} + (\sqrt{1 - \rho_{\pi}}) \theta_{\pi,i} \right) \sigma_{\pi} + \delta_{\pi,ti} \quad (10)$$

$$\log^{-1}(\lambda_{ti}) = \mu_{\lambda} + \alpha_{\lambda,t} + \beta_{\lambda} x_{t,i} + \theta_{\lambda,i} \sigma_{\lambda} + \delta_{\lambda,ti} \quad (11)$$

Where μ indicates the global mean with uninformative, $U(-20, 20)$ prior; α is an AR(1) structured time effect; ϕ is a structured spatial process with an $IAR(1)$ prior; θ is an unstructured spatial error with an independent $N(0, 1)$ prior; ρ indicates the proportion of variance that comes from the spatially structured process, with prior $Beta(1/2, 1/2)$ (Mitzi Morris 2018); σ is the variance of the convolved spatial term; and δ is a spatio-temporal interaction with a normal prior at mean 0. Finally, $\beta \cdot x_{t,i}$ indicates a set of coefficients and covariates, which are absent in the main smoothing model but include a set of environmental covariates in the climate model. For simplicity, the effect of these covariates is assumed to be constant throughout space and time. We follow Knorr-Held's recommendation to drop the unstructured temporal component γ to improve identification of the model and the parameters in equation 10 are otherwise same as described in

equation 5. Finally, we specify uninformative Gamma-distributed hyperpriors for variance as recommended by Knorr-Held 2000:

$$\sigma, \sigma_\alpha, \sigma_u, \sigma_\delta \sim \text{Gamma}(2, 1) \quad (12)$$

292 The quantities of interest include the expected number of acute Chagas Disease cases in year t in munic-
293 ipality i , which is given by:

$$E[Y_{t,i}] = (1 - \pi_{t,i})\lambda_{t,i} \cdot \text{Pop}_{t,i} \quad (13)$$

294 And the incidence rate:

$$IR_{t,i} = \frac{E[Y_{t,i}]}{\text{Pop}_{t,i}} = (1 - \pi_{t,i})\lambda_{t,i} \quad (14)$$

295 Altogether, this model was evaluated in Stan using the CmdStanR interface for the R-language. Four
296 chains were in parallel run for 2000 warmup iterations and 1000 posterior draws, and evaluated in approxi-
297 mately 12 hours on an Apple MacbookAir M1. The corresponding Stan code is included in the supplementary
298 material.

299 2.3.1 Climate Model

300 To investigate the relationship of climate with Chagas incidence, we include covariates to the model re-
301 lated to temperature, precipitation, and vegetation. Determining covariates relevant to the incidence of
302 Chagas disease is not straightforward as the climate processes that affect *Triatomines* and the *T. Cruzi*
303 parasite may not necessarily be the same as those governing transmission to humans. Further, interventional
304 strategies limiting transmission have shown an overall decrease in incidence of Chagas disease, which may
305 confound identification of climatic factors influencing transmission. Nonetheless, we have chosen to include
306 the following covariates in our model.

307 In laboratory settings, it was found that *Triatomines* incubated at warmer temperatures (30C vs. 28C and
308 26C) mature faster and had higher levels of *T. Cruzi* parasites in stool, although insect mortality did increase
309 slightly (Tamayo et al. 2018). Further, *Triatomines* may be able to adapt to changes in temperature in
310 complex ways (Clavijo-Baquet et al. 2021). Ecological modeling of Chagas Disease in North America indicates
311 that as temperatures rise, the distribution of *Triatomines* may shift towards the north and northeastern part
312 of the region (Garza et al. 2014).

313 To avoid multi-collinearity among our climatological factors, like Medone et al. 2015 we use Principal
314 Component Analysis (PCA) on the 19 WorldClim Bioclimatic Indicators (Fick and Hijmans 2017), which
315 we retrieved from the Copernicus Climate Change Service (C3S) Climate Data Store (CDS) dataset, “Global
316 Bioclimatic Indicators from 1950-2100 Derived from Climate Projections” (Wouters et al. 2021). PCA is a
317 dimensionality-reducing procedure that decomposes the matrix of covariates by municipality-year into an
318 ordered set of orthogonal vectors, or principal components. Each principal component represents a ‘trend’
319 or pattern in the data with the first component representing the most dominant pattern by proportion of
320 variance explained, and each subsequent component representing less of the variance. Each observation can
321 then be described as a linear combination of principal components and coefficients. PCA can be used for
322 Principal Component Regression; rather than using the covariates directly, each observation’s location in
323 principal component-space is used as a covariate. Since all principal components are orthogonal, this avoids

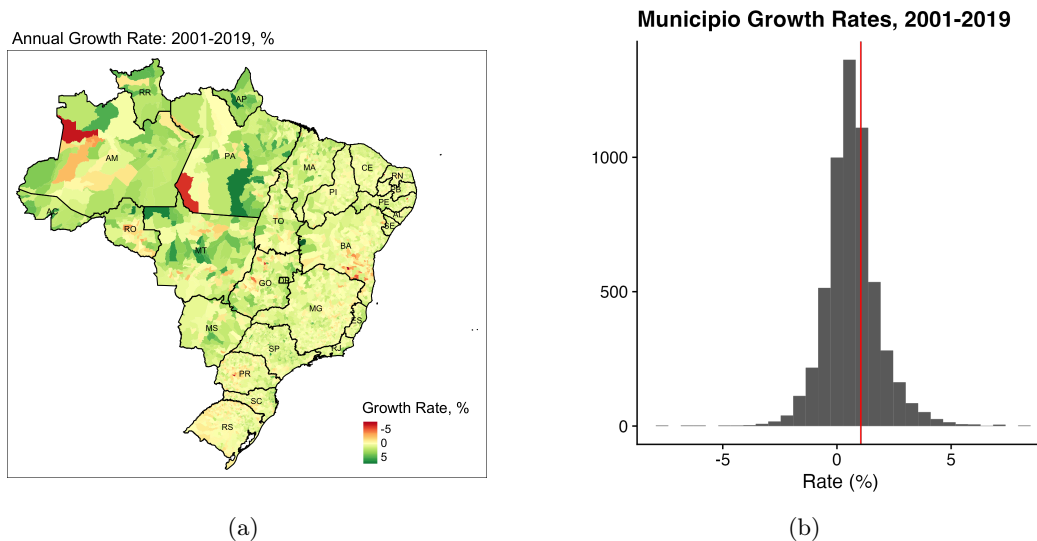


Figure 4: (A): Average annual percentage change in population for all municipalities in Brazil between 2001 and 2019. (B): Average annual growth rate, percent, for all municipalities in Brazil between 2001 and 2019. The overall growth rate for Brazil, indicated in red, is 1.04%, higher than the average municipality growth rate of 0.6%.

324 any potential multi-collinearity in the data.

325 We find that the first six principal components explained 95% of the variation in the data. Values for these
 326 principal components are displayed in supplementary table S3 and the variance explained in supplementary
 327 figure S3. The largest principal component, responsible for just over 50% of the variance in the dataset, is
 328 related to warmer, drier weather year-round. In turn, the second principal component (17% of total variance)
 329 is related to cooler temperatures with more seasonal fluctuation but less precipitation year-round. Third
 330 and subsequent principal components are less clear in their interpretation and are responsible for decreasing
 331 amounts of variance in the dataset.

332 2.3.2 Projection of future incidence

333 The resulting quantities estimated from the main smoothing model and the covariate model are used to
 334 estimate the incidence of Chagas disease over the 10 year period from 2020 to 2030 using projected population
 335 counts and projected climate variables. To project this data, we calculate the average annual exponential
 336 growth rate over the period 2001-2019 for each municipality as:

$$r_i = \exp\left(\frac{\log Pop_i(2019) - \log Pop_i(2001)}{18}\right) - 1 \quad (15)$$

337 Growth rates at the municipality level are displayed in figure 4. The overall growth rate of Brazil is 1.04%
 338 per year between 2001 and 2019, and the average municipality grew by 0.6%. The 2019 population for each
 339 municipality is projected forward each year for 10 years as:

$$Pop_i(t \in [2020 : 2030]) = Pop_i(2019)(1 + r_i)^{t-2019} \quad (16)$$

340 This projected population is used as an input to the model to predict future incidence of Chagas disease.
 341 We report two sets of projected rates of Chagas disease: one from the main smoothing model and one from
 342 the covariate model. In the covariate model, each municipality-year’s location in principal component space
 343 for the predicted bioclimatic variables is used as an additional input. The incidence in municipality i at
 344 future year t^* is estimated as:

$$E[Y_{t^*,i}] = (1 - \pi_{t^*,i})\lambda_{t^*,i} \cdot Pop_{t^*,i} \quad (17)$$

$$\text{logit}^{-1}(\pi_{t^*,i}) = \hat{\mu}_\pi + \alpha_{\pi,t^*} + \hat{\beta}_\pi x_{t^*,i} + \left((\sqrt{\hat{\rho}_\pi/s})\hat{\phi}_{\pi,i} + (\sqrt{1 - \hat{\rho}_\pi})\hat{\theta}_{\pi,i} \right) \hat{\sigma}_\pi + \delta_{\pi,t^*,i} \quad (18)$$

$$\text{log}^{-1}(\lambda_{t^*,i}) = \hat{\mu}_\lambda + \alpha_{\lambda,t^*} + \hat{\beta}_\lambda x_{t^*,i} + \hat{\theta}_{\lambda,i}\hat{\sigma}_\lambda + \delta_{\lambda,t^*,i} \quad (19)$$

Where all quantities are the median value of those estimated in the main model except time-trend quantity α . for both α_π and α_λ , which is taken as an AR(1) random walk given the distribution of estimated time trend terms:

$$\alpha_{.,2020} \sim N(\hat{\alpha}_{.,2019}, \hat{\sigma}_{.,\alpha}) \quad (20)$$

$$\alpha_{.,2021} \sim N(\alpha_{.,2020}, \hat{\sigma}_{.,\alpha}) \quad (21)$$

$$\dots \quad (22)$$

$$\alpha_{.,t^*} \sim N(\alpha_{.,t^*-1}, \hat{\sigma}_{.,\alpha}) \quad (23)$$

345 We conduct 1000 simulated random draws of future incidence and report the predicted rates of Chagas
 346 disease.

347 3 Results

348 3.1 Results of Main Smoothing Model

349 The overall incidence rate of Acute Chagas Disease in Brazil between 2001 and 2019 is estimated to be 0.121
 350 per 100k person-years of life (PYL), although substantial heterogeneity in risk exists between and within
 351 regions. Figure 5 shows the municipality level 18-year incidence rate of acute Chagas Disease. Estimated
 352 incidence of Chagas disease is highly spatially variable with a strong regional trend, with two major areas
 353 of vulnerability: first, the northern Amazon states of Amapá (AP) and Pará (PA), which have the highest
 354 smoothed incidence rates in the country at 1.80 and 1.69 per 100k PYL—almost an order of magnitude
 355 higher than the national average—as well as Acre (AC; 0.317 per 100kPYL) and Amazonas (AM; 0.188
 356 per 100kPYL). These states are highly rural and have a smaller population than the coastal states, but
 357 contain the majority of Acute Chagas Disease risk. The second main region of transmission includes the
 358 northeastern, Caatinga states of Rio Grande do Norte (RN; 0.334 per 100kPYL), Sergipe (SE; 0.247 per
 359 100kPYL), Piauí (PI; 0.197 per 100kPYL) and Pernambuco (PE; 0.316 per 100kPYL). We do not observe
 360 increase transmission rates in the Cerrado, which includes the state of Goiânia and Mato Grosso do Sul, as
 361 reported by Gurgel-Gonçalves et al. 2012 besides a slight elevation in Tocantins. The states with the lowest
 362 estimated rates of Chagas Disease are the federal district of Brasilia and Sao Paulo.

363 Within the Amazon states of Pará and Amapá, which have the highest overall rates of new Acute Chagas
364 Disease diagnoses, 31 of 160 municipalities had 18-year incidence rates higher than 1 per 100k PYL; the
365 highest rates of ACD were found in Breves (population 93,000) and Limoeiro do Ajuru (25,000), with
366 15.9 and 15.8 cases per 100k PYL respectively. Six more municipalities had rates above 10 per 100kPYL:
367 Currealinho, Abaetetuba, Bagre, Muaná, Anajás, and São Sebastião Da Boa Vista. However, the most cases
368 were predicted to be found in Belém, the capital and largest city in Pará, at 386 over the 18 year period for
369 a population of approximately 1.4 million.

370 A zero-inflated models represent the observed data as being a mixture of two processes: here, the proba-
371 bility of never being exposed to Chagas disease represented through the Bernoulli process, and the incidence
372 rate given exposure represented through the Poisson process. Lambert 1992 refers to the over-dispersion of
373 zeros generated through these processes as these as ‘structural’ and ‘non-structural’/‘sampling’ zeros, respec-
374 tively. Although we found a very strong spatial process governing the rate of ‘structural’ zeros—probability
375 of never being exposed to Chagas disease (shown in figure 6)—we did not find a strong spatial process in the
376 rate given exposure. Since Chagas disease is transmitted to humans given contact with disease-transmitting
377 vectors with a particular habitat, we interpret this to mean that Chagas-carrying *Triatomines* are more
378 likely to live in certain locales, the rate of contact and transmission within those locales is more spatially
379 constant.

380 Overall, the Northern and Amazon states were found of have a high probability of exposure to Chagas
381 Disease and the coastal and southern states were less likely to have exposure. The total spatial term for π is
382 shown in figure 6. Parameter ρ , indicating the proportion of spatial variance derived from the ICAR term,
383 evaluated to 0.985, indicating that the spatially-clustered process was responsible for most of the Chagas
384 incidence and the random process θ contributed very little to the overall distribution, indicating further that
385 the location of *Triatomines* may be driving the location of cases.

386 Where the probability of exposure to Chagas disease shows a strong spatial pattern, the rate of Chagas—
387 after normalization for population—does not show nearly the degree of spatial autocorrelation as the
388 Bernoulli process. The spatial structure for the Poisson process λ that estimates the rate of disease is
389 shown in supplementary figure S1. Instead, even though the Poisson process parameter λ is normal-
390 ized to municipal population, the Poisson process instead appears to be highlighting population locations
391 rather than a spatially-autocorrelated process. However, Moran’s I test for spatial autocorrelation did find
392 that the spatial heterogeneity term θ was statistically significantly spatially autocorrelated, albeit weakly
393 ($E[I_0] = -0.0001; I_a = 0.12; p(I_0 < I_a) < 2e - 16$). Future evaluations of this model will need to carefully
394 consider how to incorporate autocorrelation into the Poisson process while maintaining model identification.

395 The overall time trend parameters α_π and α_λ , which are specified as AR(1) processes, both show a
396 difference from 0 on the linear scale, indicating that there is a global temporal component in both the
397 Bernoulli and Poisson processes (figure S2A). However, after adding in mean terms μ and applying the
398 logit and log transforms as shown in figure S2B, the overall time trend tells a different story: the Bernoulli
399 probability π , indicating probability of non-exposure to acute Chagas Disease, decreases from 55% in 2000
400 to a maximum of 23.6% by 2006 only to increase to nearly 100% for the remainder of the study period. This
401 indicates that over the course of the study period, country-wide exposure to Chagas disease increased before
402 decreasing to nearly 0 after 2007, at which point the distribution of Chagas cases ceased to be a country-wide
403 phenomenon and instead became more spatially localized. The Poisson rate is stable around $2e-5$ per capita

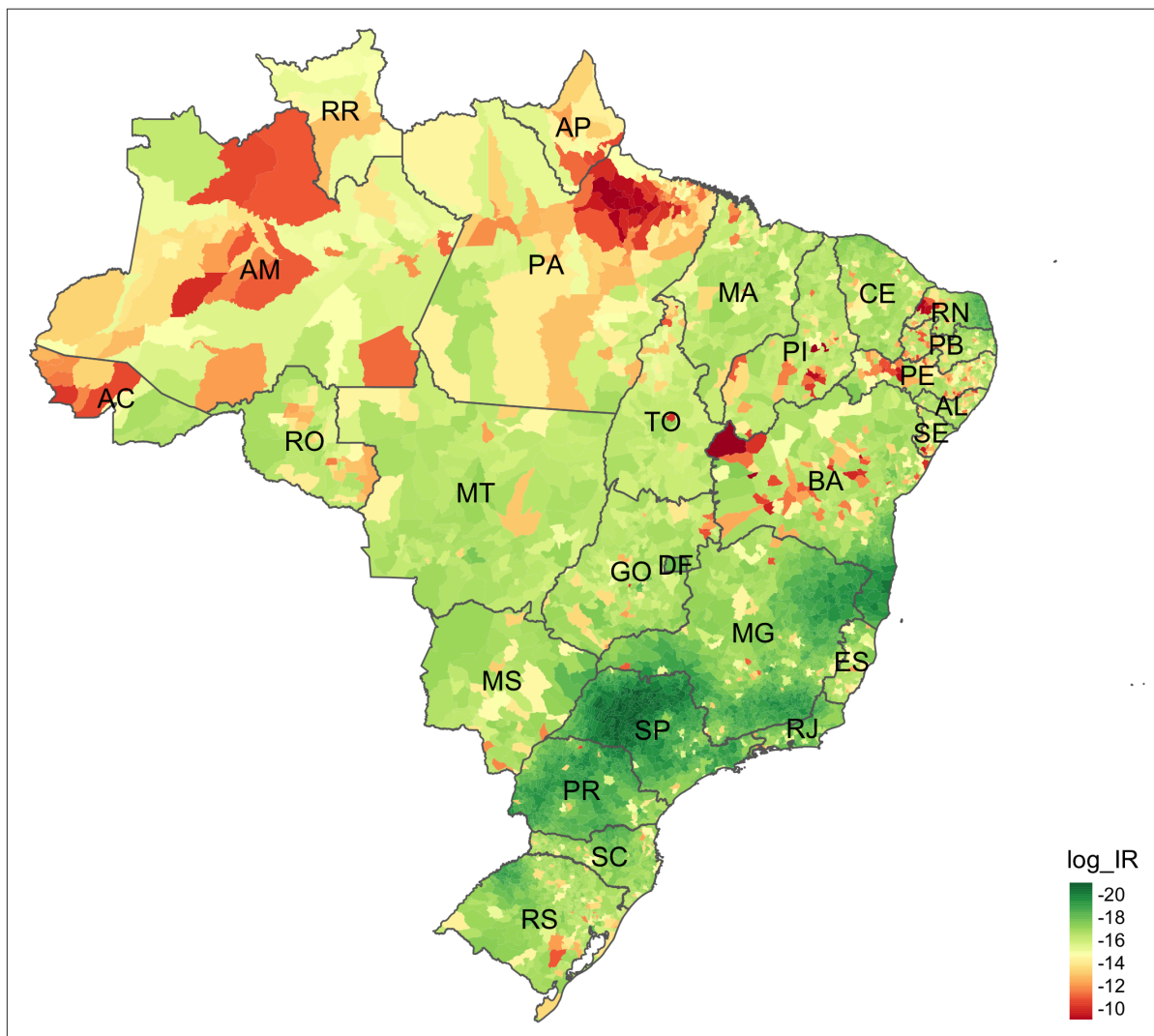


Figure 5: Overall log incidence rate at the municipality level over 2001-2019. Red indicates higher rates of ACD and green indicates lower rates of ACD. Incidence rate is calculated as $\log \sum_t \left((1 - \hat{\pi}_{m,t}) \hat{\lambda}_{m,t} / Pop_{m,t} \right)$.

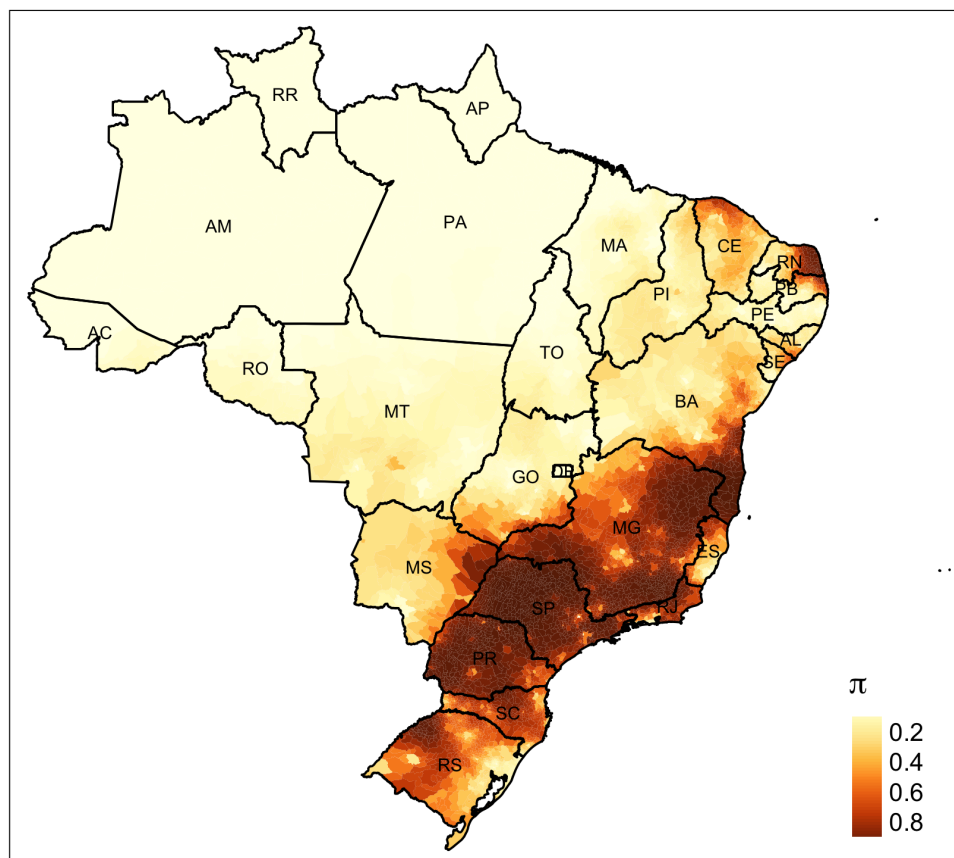


Figure 6: Estimated spatial process π governing the probability of an individual never being exposed to Chagas Disease over the study period. This process is an inverse-logit transformation of a linear combination of Conditionally AutoRegressive (CAR) term ϕ_π capturing risk that is spatially clustered, possibly due to Triatomine habitat or contact rates.

404 over the course of the study period.

Overall, the model converged well and showed good mixing between the chains for the main parameters π and λ . The Root Mean Squared Error (RMSE) of the model, evaluated as:

$$RMSE = \sqrt{\frac{1}{|T| \cdot |M|} \sum_{t \in T, i \in M} (\hat{y}_{t,i} - y_{t,i})^2}$$

405 is 0.175. Convergence is evaluated using statistic \hat{R} , which evaluates the agreement of between-chain esti-
406 mates, and Effective Sample Size (ESS), which evaluates the number of samples correcting for autocorrelation.
407 Supplementary table S1 shows the distribution of \hat{R} and ESS for all parameters $\pi, \lambda, \phi, \theta, \alpha,$ and δ .

408 3.2 Results of Climate Model

409 The climate model includes the specification as the main model above with the inclusion of each municipality-
410 year's location in principal component space among the 19 WorldClim Bioclimatic variables. We include the
411 first six principal component dimensions as covariates β in both the Bernoulli process governing overdispers-
412 sion of zeros and the Poisson process governing rate of Chagas disease. The RMSE of the climate model is
413 0.183, which is slightly higher than that of the main model, indicating that controlling for climate produces
414 a worse fit and may reduce the accuracy of the model, possibly due to overfitting. Posterior densities of β
415 are included in supplementary figure S4. In the Poisson process, posterior estimates of coefficients for the
416 first three principal components were found to be statistically significantly different from zero, whereas in
417 the Bernoulli process, only the second principal component was found to be significantly different from zero.
418 The values of parameters in the climate model are similar to the values in the main model (shown in figure
419 S5), except for ϕ_π , the spatial clustering term in the Bernoulli process.

420 To better analyze these climatic factors, we transformed these estimated coefficients from principal com-
421 ponent space back to the scale of the original variables before applying the inverse-link function and intercept
422 terms to show the values visualized in figure 7. These values are calculated as $g^{-1}(\mu + \hat{\beta}^*)$, where g is the
423 link function \log for the Poisson process λ and logit for the Bernoulli process π , and $\hat{\beta}^*$ represents the
424 estimated coefficients transformed from principal component space to the original coefficients. Ultimately,
425 β_π coefficients represent a one-unit change in each variable on the probability of non-exposure to Chagas
426 disease, and β_λ represents the effect of a one-unit change in each variable on the predicted rate of Chagas
427 disease, per million person years, conditional on exposure. Overall, we see that these variables do not affect
428 the rate of Chagas disease, only the probability of non-exposure. Non-exposure to Chagas disease is more
429 likely in climates that are highly seasonal, and less likely in wetter wetter climates.

430 3.3 Projected Rates 2020-2030

431 We used the main smoothing and climate covariate models to estimate counts of Chagas disease over the
432 decade 2020-2030. As elaborated above, the projection procedure utilizes the estimated intercept and spatial
433 parameters, and using randomly-drawn temporal structure. The main model estimates a median of 4461
434 cases of Acute Chagas Disease over the decade 2020 (IQR: 1,653 - 13,859), almost double the number of
435 cases in the previous decade (2,612). Predicted incidence is similar when including the bioclimate covariates,
436 estimating a median of 4461 cases (IQR: 1619 - 13,270). Figure 8 shows the median annual predicted counts

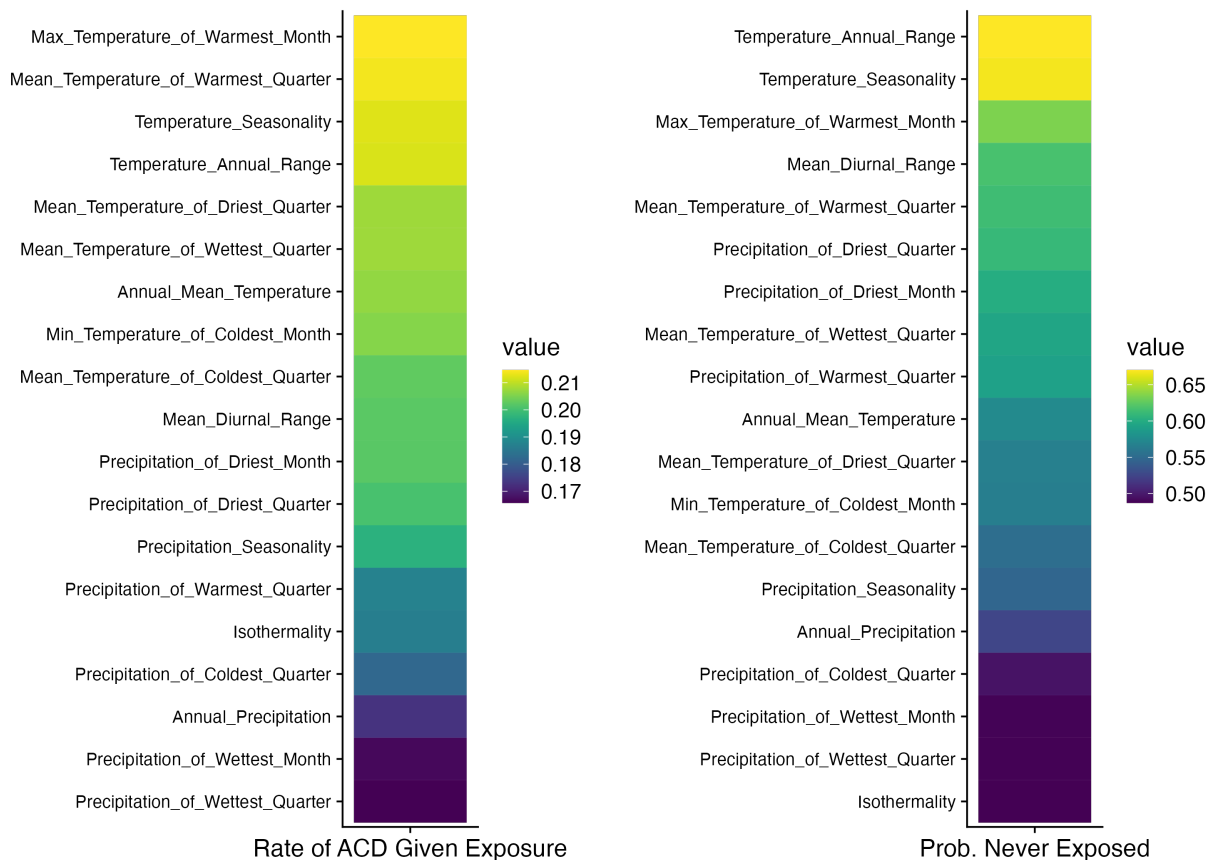


Figure 7: Coefficients for 19 WorldClim Bioclimatic Variables used in the climate model estimated for both the Poisson process (λ , left), governing the incidence rate of ACD given exposure, and the Bernoulli process (π , right), governing the probability of never being exposed to ACD. Lighter colors on the left figure indicate that a higher value of the coefficient corresponds to a higher rate of ACD in the population where exposure is present, and on the right figure indicate that probability of never being exposed is higher. Coefficients are estimated in principal component space and transformed to the natural scale, and applied the corresponding \log^{-1} link function for the Poisson process and logit^{-1} for the Bernoulli process.

437 of Chagas disease across Brazil and interquartile range between 2020 and 2030. A map of projected incidence
438 and a comparison of observed and projected rates are shown in figure 9 and for selected municipalities in
439 figure 10 and 11. Most of ‘hot spots’ for new cases are predicted to be in the same locations as 2001-2019,
440 including Abaetetuba, Belém, and Breves in Pará, and Macapá in Amapá. However, the largest increases are
441 projected to be in smaller, rural municipalities with high growth rates in the states of Amazonas—especially
442 municipalities Apuí and Tefé—and Piauí. The climate covariate model implies slightly lower rates than
443 the main model, implying that projected bioclimatic conditions may result in fewer infections, although the
444 overall difference is likely small. The trend observed in figure 8 shows a highly variable trajectory year
445 to year—much more than the main model predicts—indicating that annual climate fluctuations may have
446 substantial effects on predicted rates.

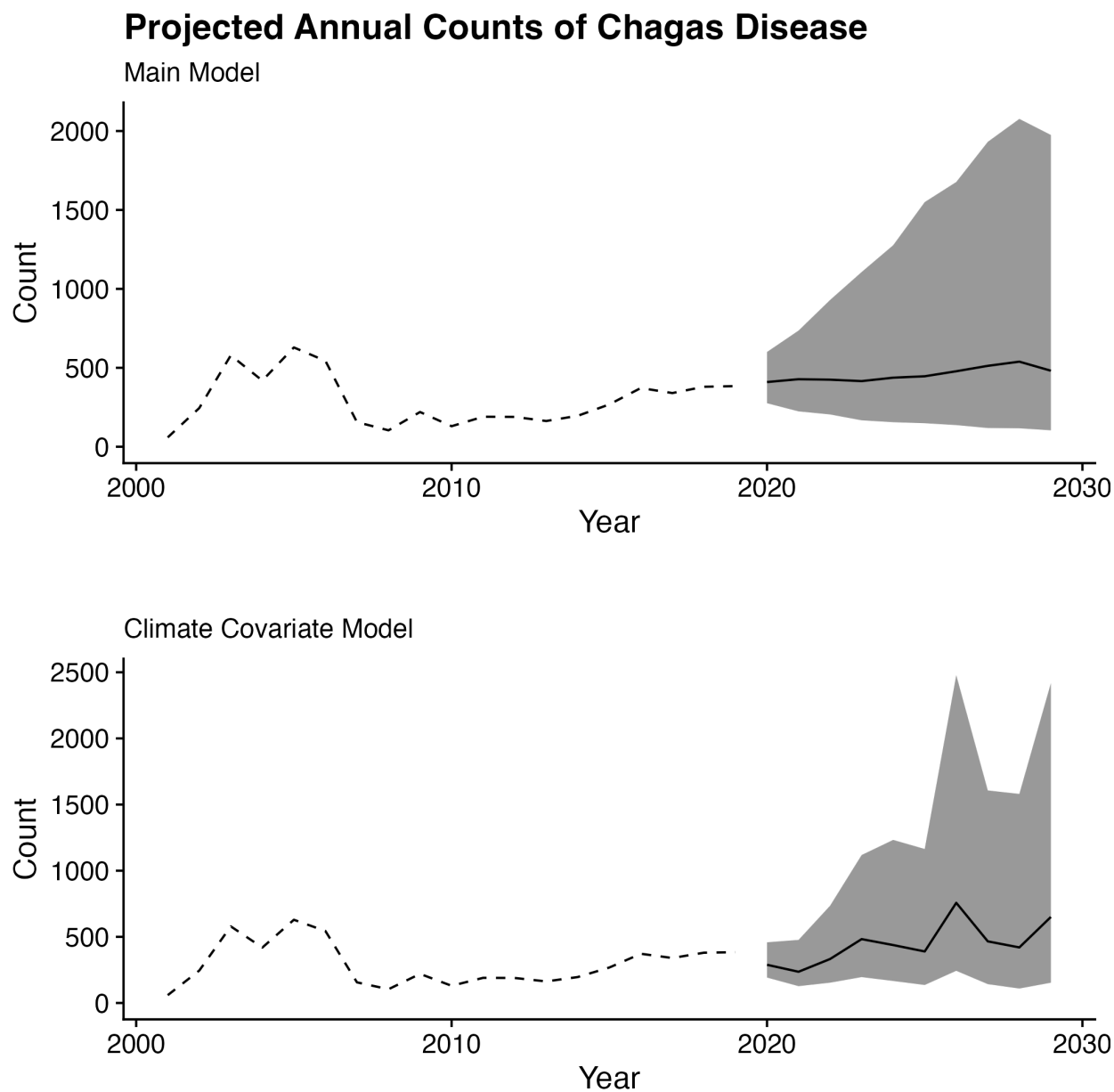
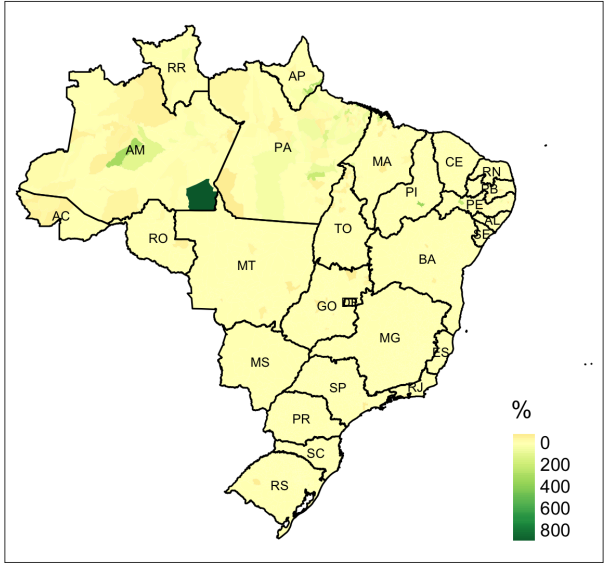


Figure 8: Observed counts across Brazil, 2001-2019, summary of 1000 projected counts, 2020-2030, from the main smoothing model (top) and climate covariate model (bottom). Median simulated counts are shown in black and interquartile range, representing 50% of simulations, is shown in grey.

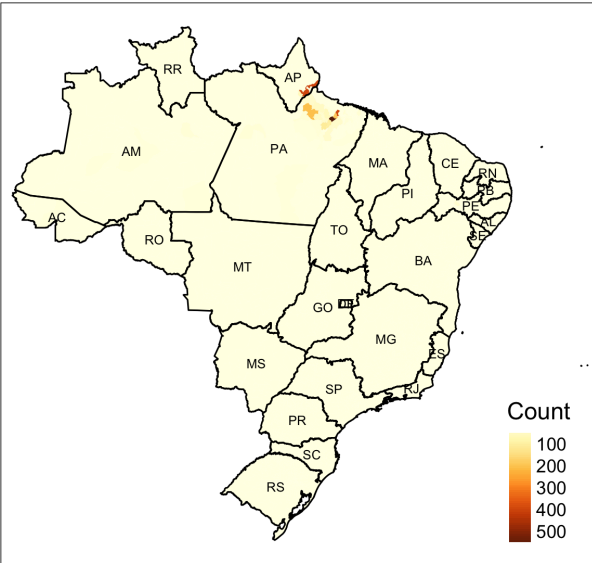
Main Model: Projected Cases 2020-2030



Main Model: % Increase 2020-2030



Covariate Model: Projected Cases 2020-2030



Covariate Model: % Increase 2020-2030

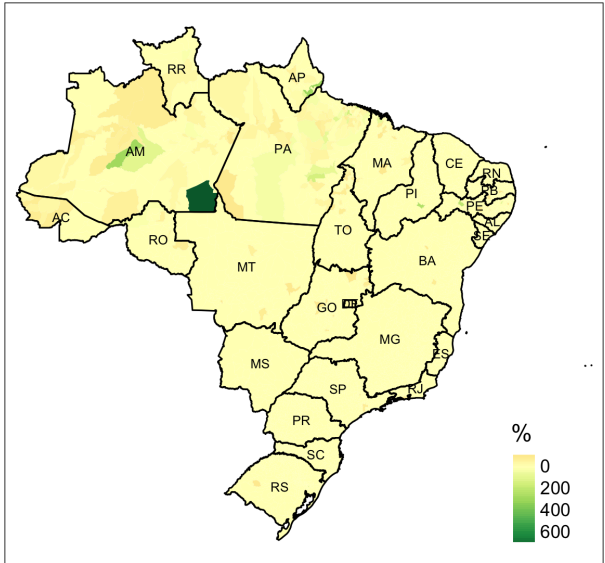


Figure 9: Projected incidence and percent increase compared to the previous decade.

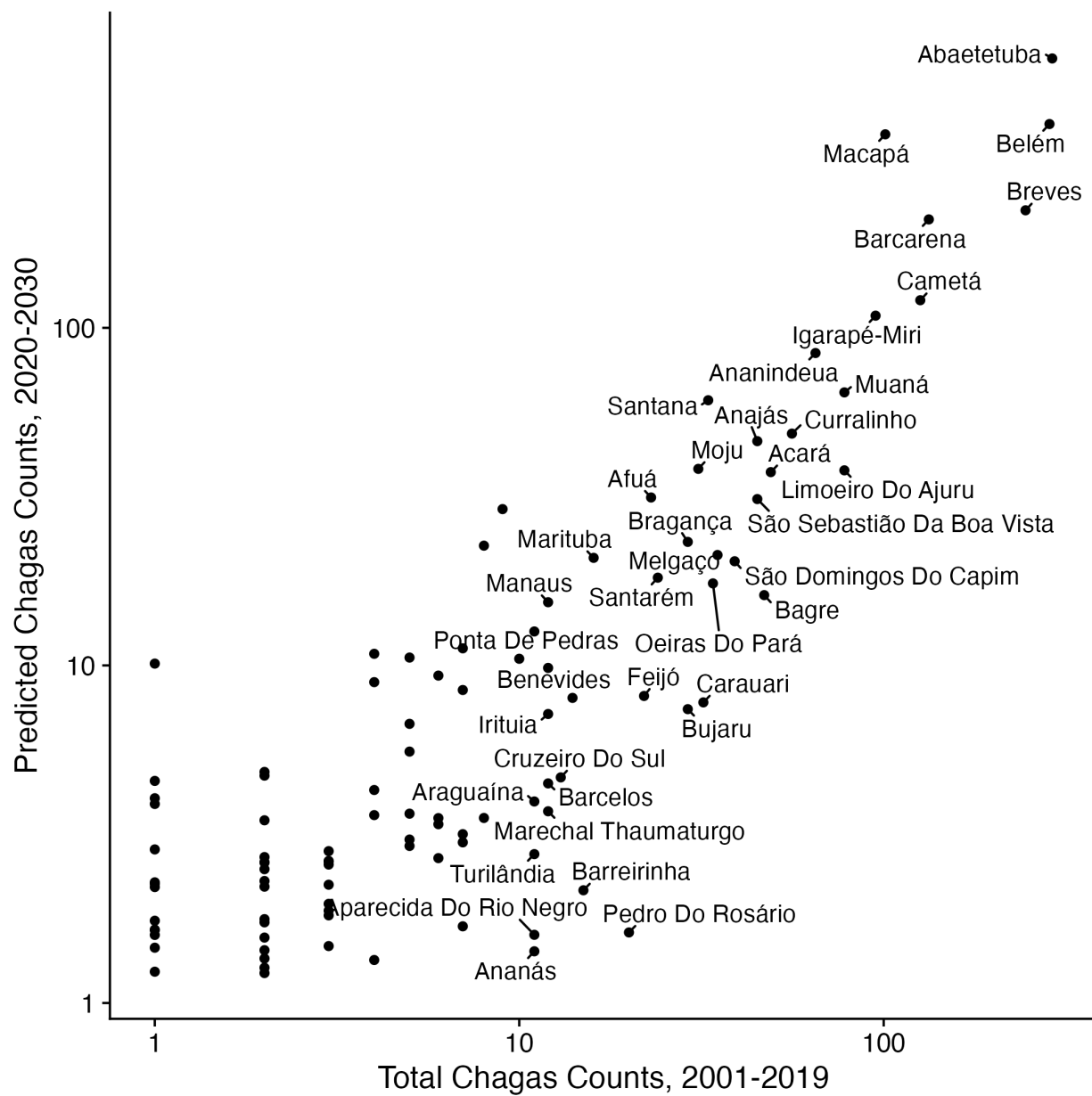


Figure 10: Observed and Projected rates, main model

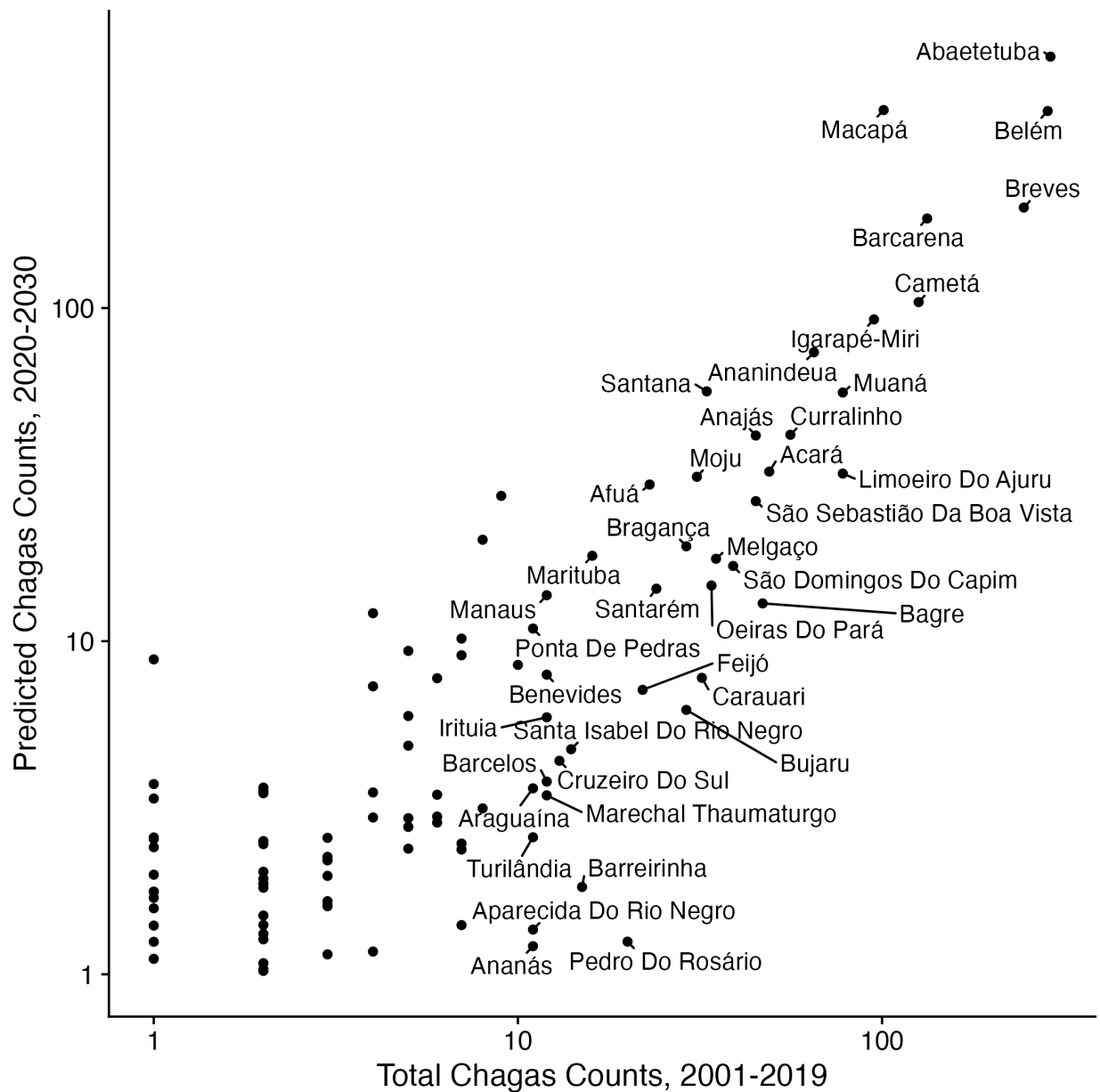


Figure 11: Observed and Projected rates, climate model

447 4 Discussion

448 Despite progress towards elimination, Chagas disease remains a significant threat to public health, and
449 additional intervention will be needed to further reduce rates of new ACD cases in Brazil. The results of our
450 modeling imply that rates of ACD will increase consistently over the ensuing decade—potentially as much as
451 doubling compared to the previous decade—driven primarily by an increasing population in high-risk areas.
452 Climate change may result in the exacerbation of this trend. However, this appears to be at odds with
453 the country-wide decrease in ACD rates as a result of *Triatomine* eradication campaigns since the 1980's.
454 While these campaigns have been enormously successful, risk of ACD is likely to persist without additional
455 intervention. As a result, our work here serves as a call to continue the campaigns. Nonetheless, we predict
456 that risk for ACD will persist, and further intervention will be necessary to continue the decrease observed
457 in previous decades.

458 This model has a number of limitations, some of which may be addressable in future modeling studies.
459 First, our data include official reports of Acute Chagas Disease as submitted to SINAN. We were unable to
460 find literature estimates of under-reporting rates of ACD not submitted to SINAN. It is possible that there
461 are many annual cases not captured in the dataset, implying that our estimates are not only incomplete, but
462 subject to a ‘survivorship paradox’ where the locations of highest epidemiological interest are not captured
463 in the dataset. Since the treatment success rate is high for acutely diagnosed cases of Chagas disease, it may
464 be reasonable to estimate the number of ‘missed’ acute cases from backwards-projection of chronic cases.
465 However, due to the substantial lag time (at least 10-30 years) between exposure and chronic symptoms, this
466 was not possible with the data available from SINAN. However, should this topic be revisited in the following
467 decades, a back-projection approach may be useful for retrospectively estimating the underreporting rate.

468 Our model uses Knorr-Held Type I spatio-temporal interactions δ , which are the least sophisticated of
469 the structures outlined by Knorr-Held (2000). This strategy essentially estimates uncorrelated space-time
470 fixed effects, which we found to be ideal for precise internal estimation; however, since these fixed effects
471 lack a temporally autoregressive definition, it is not possible to use these fixed effects to project the model
472 to predict future incidence without introducing meaningless statistical noise. In our prediction we simply
473 dropped the interaction term from the model; in effect, only the un-interacted spatial and simulated random
474 walk temporal terms were used for future projection. We believe that using the basic Type I interactions
475 allowed for better estimation of the independent spatial and temporal components as we found that the
476 Type IV term is only weakly identified in our model; so this decision was not without benefit. Choice of
477 a spatio-temporal interaction that includes a temporally autoregressive term, such as Type II or Type IV
478 (see supplementary material for more elaboration), would allow for projection in the form of a random walk
479 through interaction space-time. For example, should a Type IV prior be used that assumes total spatial and
480 temporal dependence in the interaction, future space-time interaction terms can be simulated as a random
481 draw from a Multivariate Normal distribution with a variance-covariance matrix derived from the previously-
482 estimated interaction terms, similar to the simulation procedure for spatially autocorrelated data (Banerjee,
483 Carlin, and Gelfand 2015). Special care will likely be needed to assure propriety of the Type IV distribution,
484 which is outside of the scope of the present study.

485 Our population projection would benefit from a more precise estimation methodology than the crude
486 exponential growth model used here. Future analysis should consider municipality, year, and age-specific
487 rates of fertility, mortality, and migration—especially internal migration—to inform population projections.

488 We were unable to obtain these quantities at the level of spatial and temporal granularity required. We
489 believe that in the short term—namely, the single decade between 2020 and 2030—this crude methodology
490 allows for understanding how heterogeneity in growth rates may relate to future incidence of Acute Chagas
491 Disease. Nonetheless, it is not suitable for long-term projections. Model-based estimates that utilize the
492 readily-available state-level rates to determine small-area estimates, possibly similar to the Lee-Carter (1992)
493 procedure or the one used by Alexander, Zagheni, and Barbieri 2017, may allow for more precise estimation
494 of future municipality-level population. As well, an extension of the model to predict age-specific rates of
495 Chagas disease may aid researchers in planning for interventions.

496 Finally, we make the critical assumption in both models that population and climate change are the
497 sole drivers of future incidence. Other factors, such as housing construction materials, poverty, habitat
498 destruction, and residential or industrial development encroaching on *Triatomine* habitats may increase
499 affect rates of Chagas disease even in the absence of population or climate change. Further, the inherent
500 assumption of linearity in our model assumes that as climate and population change, predicted incidence
501 of Chagas disease will respond. This may not be the case: while climate may partially determine the
502 geographic distribution of *Triatomines*, which may in turn affect incidence, it is likely that the relationship
503 of climate with *Triatomine* populations is too complex to be captured by a linear model of the sort used
504 here. Since there are many species of *Triatomines*, each with different habitats, behaviors, and virulence,
505 the ultimate effect of climate on Chagas incidence is undoubtedly complex and nonlinear. Further, our
506 climate model has a slightly higher error than the main model despite the inclusion of additional covariates,
507 indicating that the model may be suffering from overfitting. As well, many exogenous factors could affect
508 the distribution of *Triatomines* under future climate conditions, including interventional strategies to limit
509 *Triatomine* habitats like the residential insecticidal campaigns of the 1980s, development of urbanization
510 and infrastructure, and climate adaptations, environmental destruction and conservation practices that may
511 affect *Triatomine habitats*. Should future climate conditions create new habitats for *Triatomines*, it is not
512 clear at present if the insects are mobile enough to find these habitats, or if accidental importation by
513 humans—such as improper handling of lumber—may catalyze a shift in *Triatomine* distribution.

514 Replication Code

515 Replication code is publicly available at https://github.com/eroubenoff/chagas_modeling.

516

Supplementary Material

517 S1 Adjacency Matrices for Geostatistical Models

518 For the analysis of Chagas disease, we focus on methods for areal or polygon data, which refer to a region
 519 of space which contains a subset of the observations of interest. Polygonal data is a common format for
 520 administratively-collected spatial data, often representing a governmentally-defined area—such as a state or
 521 province equivalent, city or municipality, or even more specific form such as census tract or block. Areal data
 522 exist in contrast to point-referenced data, which instead link each observation with longitude and latitude
 523 coordinates. Whereas areal data can be generated from point data using a simple point-in-polygon operation,
 524 the reverse process is not possible as the specific coordinates are lost when points are tallied within polygons.

525 Critical in any spatial statistical work is the concept of the neighborhood matrix: a mathematical repre-
 526 sentation of geographic adjacency. For example, this 3x3 grid could be representing by binary neighborhood
 527 matrix W :

1	2	3
4	5	6
7	8	9

$$W = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

529 A symmetric binary matrix like this is most common for representing adjacency, but can easily be extended
 530 to include reciprocal distance weights, higher-order neighbors, or measures of connectivity that are not strict
 531 adjacency (for example, consider transit networks or other transportational features that mean the travel
 532 time between two locations is not linear with distance). While estimates will change between different
 533 matrices W , the following distributional properties remain the same.

534 In the general case of eq. 1 with a non-binary neighborhood matrix W , the equivalent, generalized model
 535 can be parameterized:

$$\phi_i | \phi_{j \neq i} \sim N \left(\frac{1}{w_{i+}} \sum_j w_{ij} \phi_j, \frac{\tau^2}{w_{i+}} \right) \quad (\text{S1})$$

536 where $w_{i+} = \sum_j w_{ij}$, or the sum of matrix W row i .

537 S2 Commentary on the BYM- type model

In their 1991 paper, BYM use the IAR distribution (eq. 1) to fit a log-linear Poisson GLM of the form:

$$\log \mu_i = \log E_i + x_i \beta + \phi_i$$

538 where μ_i is the rate of disease occurrence in unit i , E is expected count unit i , x and β are explanatory
539 variables and regression coefficients, and error term ϕ_i has the prior distribution of eq. 1. This model,
540 which utilizes internal standardization, was called by Banerjee, Carlin, and Gelfand “cheating (or at least
541 ‘Empirical Bayes’)” since the E_i are not fixed but rather itself part of the data, posing a ‘null hypothesis’
542 about if there were to be absent a spatial pattern (BCG, p.151)⁶. As well, BCG and Leroux et al (2000) note
543 that this model may have poor performance as using the CAR prior alone as an error term may over-smooth
544 aspatial variation, which may be mechanistically important to the model.

545 The IAR distribution is conditionally specified for each geographic unit and is improper, meaning that
546 it does not integrate to produce a valid probability distribution, instead only able to show the proportional
547 density between spatial units. This is problematic for stochastic generation and maximum likelihood estima-
548 tion, but is valid for Bayesian inference as posterior density need only be proportional to the prior density
549 (Besag, York, and Mollie 1991). However, it is possible and mathematically convenient to consider equation
550 1 in its joint, albeit improper, form. Besag (1974) showed that fully conditional distributions of this type
551 can utilize Brook’s Lemma (1964) to recover the full conditional form, as a multivariate normal distribution
552 with mean 0 and variance-covariance matrix related to the adjacency matrix. Banerjee, Carlin, and Gelfand
553 demonstrate this concisely, determining the joint distribution of ϕ from a fully specified set of conditionals:

$$p(\phi_1, \dots, \phi_n) \propto \exp \left\{ -\frac{1}{2\tau^2} \sum_{j \sim i} (\phi_i - \phi_j)^2 \right\} \quad (\text{S2})$$

554 which is also known as the pairwise-differences formula (Banerjee, Carlin, and Gelfand 2015, p.81, eqn. 4.16).
555 Equation S2 can be utilized to provide convenient estimation in Bayesian MCMC, using software such as Stan
556 as pairs of adjacent units can be efficiently stored in program memory, and the proportional density can be
557 quickly computed without the need for matrix inversion. Since Stan estimates the proportional log-density
558 of ϕ up to a constant, Morris et al. 2019 demonstrate that equation S2 can be quickly evaluated as:

```
559 phi ~ sigma * -0.5 * dot_self(phi[node1] - phi[node2]);  
560 sum(phi) ~ normal(0, 0.001 * N);
```

561 where `sigma` is the precision rather than the variance, where `node1` and `node2` are vectors of adjacent pairs,
562 `dot_self` takes the dot product of the vector with itself, and the second line indicates that `phi` is subject
563 to a soft sum-to-0 constraint. Due to this computational efficiency and problematic assumptions needed to
564 make this distribution properly integrate, BCG recommend that the IAR model be used only in the case
565 of a Bayesian prior, and may be frequently the optimal choice for geostatistical inference (BCG, ch. 4 and
566 BCG, ch. 6, p.155)

567 The expression in S2 is an improper probability distribution since the joint probability density is only
568 proportional to the derived expression. This is because the variance-covariance matrix implied is singular,
569 meaning the inverse does not have a unique solution and as a result the distribution does not necessarily sum
570 to one, as required for valid probability distributions. For a non-mathematical explanation, consider that
571 each observation is entirely dependent on its neighbors, which allows us to estimate the total distribution only
572 on relative terms without a ‘ground truth’ or some external source centering the distribution. To demonstrate
573 this impropriety, BCG (p.81) derive equation 1 by beginning with adjacency matrix \mathbf{W} , which has $w_{ij} = 1$ if i

⁶This is forgivable as BYM—who were working in digital image restoration—were the first to demonstrate how this technique could be used in other fields, which has become a foundational technique in Bayesian Disease Mapping.

574 and j are neighbors and 0 otherwise; matrix \mathbf{B} where $b_{ij} = w_{ij}/w_{i+}$, or a row-standardized version of matrix
575 \mathbf{W} ; and \mathbf{D} , a diagonal matrix where d_{ii} is equal to the number of neighbors of i and 0 otherwise. Then,
576 equation 1 can be written in the conditional form as $\phi_i | \phi_{j \sim i} \sim N(\sum_j b_{ij} y_j, \tau_i^2)$ since \mathbf{B} is the row-standardized
577 version of \mathbf{W} . This would imply that $\phi \sim MVN(\mathbf{0}, [\tau(\mathbf{D} - \mathbf{W})]^{-1})$. Temporarily disregarding τ , calculating
578 the covariance matrix Σ^{-1} of this distribution involves the calculation $(\mathbf{D} - \mathbf{W})\mathbf{1} = \mathbf{0}$, which is singular;
579 effectively, too many variables without a constraint to preserve propriety (BCG, p.81). It is possible to make
580 this distribution proper with an additional parameter, often denoted ρ (α in Morris et al. 2019), by defining
581 the inverse covariance matrix $\Sigma^{-1} = \mathbf{D} - \rho\mathbf{W}$, so long as ρ is chosen to find a singular solution. BCG list the
582 bounds under which ρ will provide a non-singular solution, which is related to the eigenvalues of matrices
583 \mathbf{D} and \mathbf{W} . Then, the full conditional distribution becomes:

$$\phi_i | \phi_{j \sim i} \sim N\left(\rho \frac{1}{n_i} \sum_{j \sim i} \phi_j, \frac{\tau^2}{n_i}\right) \quad (\text{S3})$$

584 BCG write that ρ is sometimes reported as being the degree of spatial autocorrelation, but it is clear from
585 equation S3 that the resultant expression rather expresses some *proportion* of the spatial gaussian process
586 (p.82). As well, ρ does not map clearly onto any other measures of spatial autocorrelation, like Moran's I or
587 Geary's C, and thus its interpretation outside of the model is limited.

588 Further, the authors remark that the proper CAR model may be attractive in cases where the spatial
589 pattern is weak, and the improper CAR model may over-smooth heterogeneity. However, in simulation, the
590 proper CAR model has been shown to nearly always converge on values of ρ close to 1, as when ρ is less than
591 1 there presents an identification challenge between the spatial random effects and the non-spatial random
592 effect. BCG remark (p.155) that it appears that the data always *want* ρ to be close to 1. In conclusion,
593 BCG recommend that the improper IAR model be only used as a Bayesian prior, or in the frequentist case,
594 use of a SAR or other proper probability distribution.

595 As noted above, two issues present with using the IAR model alone as a prior for a spatially autocorrelated
596 error term. First, the IAR model is known to show poor performance when spatial autocorrelation is not very
597 strong, otherwise it will oversmooth random variation in the data. This issue is rectified with a proper CAR
598 model, but as noted above, BCG do not recommend usage of the proper CAR prior. Second, the IAR variance
599 parameter τ has an ambiguous function, and sources differ as to its interpretation. While Leroux (2000)
600 states that this parameter represents both autocorrelation and over-dispersion simultaneously, but Banerjee,
601 Carlin, and Gelfand write that this parameter should not be taken as representing spatial autocorrelation in
602 any mechanistic way.

603 Ideally, we would have included the BYM2-type spatial convolution term in both the Bernoulli and
604 Poisson processes, however in model development we were unable to reach convergence with the model
605 specified as such: the convolved spatial error for the Poisson parameter λ failed to be identified. Recall
606 that this term is specified in both the Bernoulli and Poisson parameters as $\left((\sqrt{\rho/s})\phi + (\sqrt{1-\rho})\theta\right)\sigma$, where
607 ϕ is the IAR model, θ is $N(0,1)$, ρ represents the proportion of variance having a spatial pattern rather
608 than unstructured error, and σ is the overall variance. The convolved spatial process in the Bernoulli part
609 showed excellent convergence and mixing with posterior estimates of $\rho_\pi \approx 0.75$. As a result, both the
610 spatially clustering term ϕ and the spatial heterogeneity term θ contributed to posterior estimates of π and

611 the likelihood of the model. However, in the Poisson part, median posterior estimates of $\rho_\lambda \approx 0$ such that the
612 convolved error term $\approx \theta\sigma$. As a result, the sampler could not identify values for ϕ_λ , functionally searching
613 the entire parameter space for ϕ_λ without any effect on the likelihood of the model. Ultimately, this caused
614 slow evaluation and divergences, but did not affect the resultant values of other parameters.

615 **S3 Zero-Inflated Models and their efficient estimation in Stan**

616 To assess the appropriateness of the ZIP distribution for our data, we conducted a naive maximum likelihood
617 estimate of the Poisson parameter λ and the Zero-Inflated Poisson parameters π, λ without any adjustment
618 for spatial structure, annual deviations, or covariates. Likelihood functions for both the Poisson-only spec-
619 ification and ZIP specification were optimized over the municipality-year counts of Chagas incidence using
620 R's multivariate `optim` routine. The maximum likelihood estimate of Poisson-only parameter λ was 0.1053
621 with log-likelihood $\ell\ell = 57773.01$, and the estimate of ZIP parameters were $\pi = 0.9849$ and $\lambda = 6.9885$ with
622 $\ell\ell = 19330.45$. To test the relative fit of both models, we conduct Wilk's test for likelihood ratios, which
623 assumes that the ratio of two likelihoods is asymptotically distributed as $\chi^2(df = df_{H1} - df_{H0})$. Taking the
624 Poisson-only specification as the null hypothesis and the ZIP specification as the alternative, the probability
625 of observing these data generated by the Poisson-only specification instead of the ZIP specification is $p \approx 0$.
626 Hence, we can comfortably reject the Poisson-specification in favor of the ZIP specification.

627 Using ZIP models may provide an additional computational advantage over a regular Poisson specifica-
628 tion. We found that a model of the type laid out in the previous sections, which models the count of Chagas
629 disease as a Poisson-distributed GLM with terms for fixed effects for spatial and aspatially-clustered errors,
630 showed slow evaluation and poor estimation. While the sheer dimensionality of the model—approximately
631 estimating 6 parameters for 5000 municipalities across 19 years—was undoubtedly responsible for part of the
632 problem, we hypothesized that the complicated posterior geometry caused by the overdispersion of 0s in the
633 dataset was partially to blame. To test this hypothesis, we ran two test cases each with a single UF over the
634 first two years of the study period. We chose Pará (PA), which has the highest number of Chagas cases at
635 5259 over the 19 year study period in 143 municipalities, and Roraima (RR), with the second lowest number
636 of cases at 10 cases in 15 municipalities⁷. In Stan, the test models were run for 500 warmup iterations and
637 500 sampling iterations. For PA, the model completed evaluation in 2508 seconds for an average parameter
638 effective sample size (ESS) of 2234.738 (SD = 1413). However, despite RR having one tenth the number of
639 municipalities of PA, the model for RR took more than twice as long to evaluate—5994.4 seconds—for an
640 average parameter ESS of 1416.437 (SD = 451.5). Both models showed convergence ($\hat{R} \leq 1.01$) for more
641 than 99% of parameter estimates. When this test was replicated using the ZIP model with an autoregressive
642 component in the Bernoulli part only, the PA model evaluated in 46.2 seconds with an average parameter
643 effective sample size of 470.96, and the RR model evaluated in 57 seconds with an average ESS of 25.39.

644 Naive implementation of Spatio-Temporal statistical models involves many pairwise comparisons, which
645 can be prohibitively computationally expensive for MCMC estimation. For example, our spatio-temporal
646 adjacency structure may contain not only neighboring observations between all 5000 municipalities in Brazil,
647 but the temporally-correlated neighbors as well. Assuming that an average municipality has 4 spatial neigh-

⁷Only the Federal District, DF, had fewer cases, at 4 over the 19 year period, but was not chosen since that UF contains only one municipality.

648 bors and 4 temporal neighbors, this will result in inverting a neighborhood matrix of over 1.6 billion elements
649 which is not reasonably evaluated with standard computing resources, let alone for thousands of MCMC
650 iterations. We have taken many steps towards quick and efficient estimation in lieu of this challenge, which
651 is a major contribution of this research in addition to the primary substantive estimation of Acute Chagas
652 Disease incidence.

653 When sampling a ZIP GLM in Bayesian software such as Stan, we will have to write a custom log-
654 probability mass function (LPMF, in Stan terms) to cover zero-inflation. First, we assume that Bernoulli
655 parameter π is estimated on a logit scale and Poisson parameter λ is estimated on a log scale, as is common
656 for GLMs. Second, we must evaluate the probability density on the log scale. The following is an unoptimized
657 ZIP estimation, adapted from Stan’s documentation for Zero-Inflated Poisson models⁸:

```
658 for (t in 1:T){  
659   for (n in 1:N) {  
660     if (y[t,n] == 0) {  
661       target += log_sum_exp(bernoulli_logit_lpmf(1 | pi[t,n]),  
662                             bernoulli_logit_lpmf(0 | pi[t,n])  
663                             + poisson_log_lpmf(y[t,n] | pi[t,n]));  
664     } else {  
665       += bernoulli_logit_lpmf(0 | pi[t,n])  
666         + poisson_log_lpmf(y[t,n] | pi[t,b]);  
667     }  
668   }  
669 }
```

670 where `target` is the log-probability of the model and `log_sum_exp(a,b) = log(exp(a) + (exp(b)))`. Clearly,
671 this is a highly inefficient way to evaluate the model since the log-probability statement is conditioned on
672 the data `y` which is known and constant through the course of the simulation. In computational efficiency
673 terms, each evaluation of the likelihood will complete in $O(T \cdot N)$ time, meaning that the time to evaluate the
674 log-probability statement is proportional to the number of municipalities times the number of years. For our
675 application to Chagas Disease in Brazil, which contains observations of approximately 5000 municipalities
676 over 19 years, this becomes extraordinarily slow, evaluating 1000 warmup and sampling iterations on the
677 scale of 12-24 hours.

To optimize this Stan modeling statement, the Stan documentation recommends partitioning the data into zero and non-zero elements and evaluating them separately, but does not elaborate on how to do so in a GLM framework, which we have developed for our application. Doing so will allow for separate, efficient vectorized evaluation of the Bernoulli and Poisson GLM statements. Indeed, as explained elsewhere, vectorization is one of the primary benefits of using Stan over other Bayesian MCMC software suites, since vectorized probability statements evaluate much faster and with less overhead than doubly-looped functions. First, consider a matrix of counts Y with dimensions T (number of years) and N (number of municipalities). From Y , we will derive two matrices `zero_idx` and `nonzero_idx` with the same dimensions, containing the indices of zero and nonzero observations, and supported by vectors `zero_max` and `nonzero_max` with dimension T , where each element contains the annual number of zero and nonzero entries. In this way, for matrix row $t \in T$ columns $[1 : \text{zero_max}[t]]$ contains the index of municipalities with zero entries, and $[\text{zero_max}[t] + 1 : N]$

⁸<https://mc-stan.org/docs/stan-users-guide/zero-inflated.html>

are uninitialized. Beginning with matrix of counts Y :

$$Y = \begin{bmatrix} 0 & 1 & 0 & 3 & 2 \\ 1 & 0 & 0 & 0 & 1 \\ 2 & 0 & 2 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 2 & 1 & 3 & 1 & 2 \end{bmatrix}$$

that yields derived zero and non-zero matrices:

$$\text{zero_idx} = \begin{bmatrix} 1 & 3 & / & / & / \\ 2 & 3 & 4 & / & / \\ 2 & 4 & / & / & / \\ 1 & 2 & 3 & 4 & 5 \\ / & / & / & / & / \end{bmatrix} \quad \text{nonzero_idx} = \begin{bmatrix} 2 & 4 & 5 & / & / \\ 1 & 5 & / & / & / \\ 1 & 3 & 5 & / & / \\ / & / & / & / & / \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix}$$

with max vectors:

$$\text{zero_max} = \begin{bmatrix} 2 & 3 & 2 & 5 & 0 \end{bmatrix} \quad \text{nonzero_max} = \begin{bmatrix} 3 & 2 & 3 & 0 & 5 \end{bmatrix}$$

678 In Stan, zero-counts in year t can be easily indexed from Y as $Y[t, \text{zero_idx}[t, 1:\text{zero_max}[t]]]$
 679 and nonzeros as $Y[t, \text{nonzero_idx}[t, 1:\text{nonzero_max}[t]]]$. Essentially, these sparse matrices are an
 680 efficient way to store ragged arrays, which are not supported natively in a C-based language like Stan. The
 681 R code for generating these matrices and vectors from matrix Y in:

```
682 n_T = nrow(Y)
683 N = ncol(Y)
684 zero_max = array(rep(0,n_T))
685 nonzero_max = array(rep(0,n_T))
686 zero_idx = matrix(0, nrow = n_T, ncol = N)
687 nonzero_idx = matrix(0, nrow = n_T, ncol = N)
688
689 for (t in 1:n_T) {
690   for (n in 1:N){
691     if (Y[t,n] == 0) {
692       zero_max[t] = zero_max[t] + 1
693       zero_idx[t, zero_max[t]] = n
694     }
695     else {
696       nonzero_max[t] = nonzero_max[t] + 1
697       nonzero_idx[t, nonzero_max[t]] = n
698     }
699   }
700 }
```

Then, we can turn to writing a log probability mass function describing equations 7 and 8 that takes advantage of this vectorization. First, recall that π is on the logit scale and λ is on the log scale, and we wish to evaluate the probability on the log scale. Assuming that π and λ have been transformed using their

corresponding inverse-link functions, this yields likelihood:

$$\log(P(y_i = 0)) = \log(\pi_i + (1 - \pi_i)e^{\lambda_i}) \quad (\text{S4})$$

$$\log(P(y_i = k)) = \log((1 - \pi_i)e^{\lambda_i} \lambda_i^k / k!) \quad (\text{S5})$$

701 Equation S4 is problematic in that it does not simplify to built-in Stan probability statements, but can be
702 written in a way that is efficiently vectorized⁹. Luckily, equation S5 simply evaluates to two independent
703 expressions of Bernoulli probability and Poisson probability on the log scale, respectively. This means they
704 can be evaluated in Stan as:

```
705 vector[N] pi_inv_logit;  
706 vector[N] lambda_exp;  
707  
708 for (t in 1:T) {  
709   pi_inv_logit = inv_logit(pi[t]);  
710   lambda_exp = exp(lambda[t]);  
711  
712   // Zeros  
713   target += sum(log(  
714     pi_inv_logit[zero_idx[t, 1:zero_max[t]] +  
715     (1-pi_inv_logit[zero_idx[t, 1:zero_max[t]]) .*  
716     exp(-lambda_exp[zero_idx[t, 1:zero_max[t]])  
717   ));  
718  
719   // Nonzeros  
720   target += bernoulli_lpmf(  
721     rep_array(0, nonzero_max[t]) |  
722     pi_inv_logit[nonzero_idx[t, 1:nonzero_max[t]]]  
723   ) + poisson_lpmf(  
724     y[t, nonzero_idx[t, 1:nonzero_max[t]] |  
725     lambda_exp[nonzero_idx[t, 1:nonzero_max[t]]  
726   );  
727 }
```

728 In a test case, this vectorized model evaluated 100 warmup iterations and 100 sampling iterations in 1099
729 seconds, more than 10 times faster than the non-vectorized example.

730 S4 Knorr-Held Spatio-Temporal Models

731 The other priors outlined in Knorr-Held (2000) are, respectively:

- 732 • Type I interaction, where all interaction terms are *a priori* independent:

$$P(\delta|\sigma_\delta) \propto \exp\left(-\frac{\sigma_\delta}{2} \sum_{i \in I, t \in T} (\delta_{it})^2\right) \quad (\text{S6})$$

733 which is suitable if the space-time interaction does not have any structure.

⁹In theory, an additional optimization of the Zero-likelihood involves use of the `log-sum-exp` trick, which provides computationally efficient evaluation of $\log(\mathbf{a}+\mathbf{b}) = \log(\exp(\mathbf{a}) + \exp(\mathbf{b}))$, but this remains unexplored at present.

- 734 • Type II interaction, where each spatial unit follows a 1st order random walk independent of its neigh-
735 bors:

$$P(\delta|\sigma_\delta) \propto \exp\left(-\frac{\sigma_\delta}{2} \sum_{i \in I, t \in [2:T]} (\delta_{it} - \delta_{i,t-1})^2\right) \quad (S7)$$

736 which is suitable if temporal trends differ between spatial units and the temporal trends do not have
737 any structure in space.

- 738 • Type III interaction, where interaction effects follow an intrinsic autoregression such as the type laid
739 out in equation 1, but are independent at each time:

$$P(\delta|\sigma_\delta) \propto \exp\left(-\frac{\sigma_\delta}{2} \sum_{j \sim i, t \in T} (\delta_{it} - \delta_{jt})^2\right) \quad (S8)$$

740 which is suitable if the spatial trends differ between time points, but the temporal trends do not have
741 any structure in space.

- 742 • Type IV interaction, perhaps the most methodologically and conceptually interesting, where effects
743 are totally dependent over space and time:

$$P(\delta|\sigma_\delta) \propto \exp\left(-\frac{\sigma_\delta}{2} \sum_{j \sim i, t \in [2:T]} (\delta_{it} - \delta_{jt} - \delta_{i,t-1} + \delta_{j,t-1})^2\right) \quad (S9)$$

744 Which defines a space-time Markov random field and is suitable if temporal trends are different from
745 location to location but are more likely to be similar in adjacent locations. This prior can be written
746 in Stan as:

```
747 real knorr_held_type4_lpdf(vector delta_t, vector delta_tm1, int N, int[] node1, int[] node2) {
748     return -0.5 * dot_self(delta_t[node1] - delta_t[node2] - delta_tm1[node1] + delta_tm1[node2]) +
749     normal_lpdf(sum(delta_t) | 0, 0.001*N);
750 }
```

751 where `delta_t` is the value of δ at time t , `delta_tm1` is the value of δ at time $t - 1$, `node1` and `node2`
752 indicate adjacent pairs of nodes, and the `normal_lpdf` statement indicates a soft sum-to-0 constraints
753 for δ_t , as done for the ICAR prior above.

754 At face value, interaction Type IV would be the most useful for our purposes, however in model develop-
755 ment we found that this model both was under-identified and over-smoothed random variation in the data.
756 Instead, we opt for Type I priors, which are both simpler to estimate and more easily identified. Theoretically,
757 type IV interactions are comparing not only the first degree neighbors—each observation with its spatial
758 neighbors and previous observation—but also the 2nd order neighbors—the spatial neighbors of temporal
759 neighbors, or equivalently, the temporal neighbors of spatial neighbors (Knorr-Held 2000). Essentially, this
760 prior is an extension of the pairwise-differences CAR prior (eq S2) to the temporal dimension. Where the
761 pairwise CAR prior focuses on the differences between adjacent units, the Knorr-Held Type IV prior includes
762 the differences between adjacent units in the current time period and the prior time period. Knorr-Held
763 remark that such a model may be useful for modeling the spatio-temporal spread of both infectious diseases

764 and non-infectious diseases where the underlying risk has a spatio-temporal pattern, as is appropriate for our
765 application to Chagas disease. For the first time point $t = 1$, the ‘previous’ time period $t = 0$ is unavailable,
766 so for this case only the ‘previous’ time period is instead taken to be a 0 vector, at which point the model
767 simplifies to the ICAR prior. If the full model is specified as in equation 5, it is likely that identifiability
768 will be poor without highly informative priors, as was the case for the BYM model above. For the present
769 application to Chagas disease, it may be possible to disregard the time trends α and γ , since most locations
770 begin and remain absent Chagas disease throughout the duration of study.

771 S5 Stan model code, edited slightly for clarity

```
772 functions {  
773   real icar_normal_lpdf(vector phi, int N, array[] int node1, array[] int node2) {  
774     // Soft sum-to-zero constraint  
775     return -0.5 * dot_self(phi[node1] - phi[node2]) + normal_lpdf(sum(phi) | 0, 0.001*N);  
776   }  
777 }  
778 data {  
779   // Number of municipalities  
780   int<lower=0> N;  
781   // Number of years  
782   int<lower=0> T;  
783   // Number of adjacent edges  
784   int<lower=0> N_edges;  
785   // node1[i] adjacent to node2[i]  
786   array[N_edges] int<lower=1, upper=N> node1;  
787   // and node1[i] < node2[i]  
788   array[N_edges] int<lower=1, upper=N> node2;  
789   // count outcomes  
790   array[T,N] int y;  
791   // Population exposure  
792   array[T,N] int E;  
793   // Scaling factor-- scales variance of spatial effects  
794   real<lower=0> scaling_factor;  
795   // indices of zero counts  
796   array[T,N] int zero_idx;  
797   // Max number of zero counts  
798   array[T] int zero_max;  
799   // indices of nonzero counts  
800   array[T,N] int nonzero_idx;  
801   // max number of nonzero counts  
802   array[T] int nonzero_max;  
803 }  
804 transformed data {  
805   // Logged population  
806   array[T] vector[N] log_E;  
807  
808   for (t in 1:T) {  
809     log_E[t] = to_vector(log(E[t,1:N]));  
810   }  
811 }  
812 parameters {  
813   // Bernoulli part: Knorr-Held model  
814   // Intercept  
815   real mu_pi;  
816   real mu_lambda;  
817  
818   // Structured temporal trend
```

```
819 vector[T] alpha_pi;
820 vector[T] alpha_lambda;
821 real<lower=1e-10, upper=10> sigma_alpha_pi;
822 real<lower=1e-10, upper=10> sigma_alpha_lambda;
823
824 // Structured spatial pattern
825 vector[N] phi_pi;
826 // vector[N] phi_lambda;
827
828 // Unstructured spatial pattern
829 vector[N] theta_pi;
830 vector[N] theta_lambda;
831
832 // Proportion of spatial/aspatial error
833 real<lower=0, upper=1> rho_pi;
834 // real<lower=0, upper=1> rho_lambda;
835 real<lower=1e-10, upper=10> sigma_convolved_pi;
836 real<lower=1e-10, upper=10> sigma_convolved_lambda;
837
838 // Knorr-Held Type I spatio-temporal interaction
839 array[T] vector[N] delta_pi;
840 array[T] vector[N] delta_lambda;
841 real<lower=1e-10, upper=10> sigma_delta_lambda;
842 real<lower=1e-10, upper=10> sigma_delta_pi;
843
844
845 }
846 transformed parameters{
847   array[T] vector[N] pi; // Bernoulli GLM term
848   array[T] vector[N] lambda; // Poisson GLM term
849
850   for (t in 1:T) {
851     pi[t] = inv_logit(mu_pi +
852       alpha_pi[t] +
853       sigma_convolved_pi * (
854         sqrt(rho_pi/scaling_factor) * phi_pi + sqrt(1-rho_pi)*theta_pi
855       ) +
856       sigma_delta_pi * delta_pi[t]);
857     lambda[t] = exp(log_E[t] + mu_lambda +
858       alpha_lambda[t] +
859       sigma_convolved_lambda * (
860         // sqrt(rho_lambda/scaling_factor) *
861         // phi_lambda
862         theta_lambda
863         // sqrt(1-rho_lambda)*theta_lambda
864       ) +
865       sigma_delta_lambda * delta_lambda[t]);
866   }
867 }
868 model {
869   // Intercepts
870   mu_pi ~ normal(-10, 10);
871   mu_lambda ~ normal(-5, 10);
872
873   // Structured temporal trend
874   alpha_pi[1] ~ normal(0, sigma_alpha_pi);
875   alpha_pi[2:T] ~ normal(alpha_pi[1:(T-1)], sigma_alpha_pi);
876   sigma_alpha_pi ~ gamma(2, 1);
877
878   alpha_lambda[1] ~ normal(0, sigma_alpha_lambda);
879   alpha_lambda[2:T] ~ normal(alpha_lambda[1:(T-1)], sigma_alpha_lambda);
880   sigma_alpha_lambda ~ gamma(2, 1);
```



```
881
882 // Structured spatial patten
883 phi_pi ~ icar_normal(N, node1, node2);
884 // phi_lambda ~ icar_normal(N, node1, node2);
885
886 // Unstructured spatial error
887 theta_pi ~ std_normal();
888 theta_lambda ~ std_normal();
889
890 // Prior on Rho
891 rho_pi ~ beta(.5, .5);
892 // rho_lambda ~ beta(.5, .5);
893
894 // Convolved variance
895 sigma_convolved_pi ~ gamma(2,1);
896 sigma_convolved_lambda ~ gamma(2,1);
897
898 for (t in 1:T){
899   // Interaction
900   delta_pi[t] ~ std_normal();
901   delta_lambda[t] ~ std_normal();
902
903 }
904 sigma_delta_pi ~ gamma(2, 1);
905 sigma_delta_lambda ~ gamma(2, 1);
906
907
908 // Likelihood
909 for (t in 1:T) {
910
911   // Vectorized ZIP
912   // Zeros
913   if (zero_max[t] > 0) {
914     target += log(
915       pi[t, zero_idx[t, 1:zero_max[t]]] +
916       (1 - pi[t, zero_idx[t, 1:zero_max[t]]]) .*
917       exp(-lambda[t, zero_idx[t, 1:zero_max[t]]])
918     );
919   }
920
921   // Nonzeros
922   if (nonzero_max[t] > 0) {
923     target += bernoulli_lpmf(
924       rep_array(0, nonzero_max[t]) |
925       pi[t, nonzero_idx[t, 1:nonzero_max[t]]]
926     );
927     target += poisson_lpmf(
928       y[t, nonzero_idx[t, 1:nonzero_max[t]]] |
929       lambda[t, nonzero_idx[t, 1:nonzero_max[t]]]
930     );
931   }
932 }
933 }
934 }
935
936
```

Table S1: MCMC convergence diagnostics for main smoothing model selected parameters

Parameter	\hat{R} 5th Quantile	\hat{R} 95th Quantile	ESS 5th Quantile	ESS 95th Quantile
π	1.00	1.01	941.00	3,708.96
λ	1.00	1.00	6,116.68	10,733.15
ϕ_π	1.00	1.00	993.54	3,845.31
θ_π	1.00	1.00	8,329.99	12,173.52
θ_λ	1.00	1.00	4,293.10	11,373.93
α_π	1.00	1.01	430.58	635.58
α_λ	1.00	1.01	364.32	537.66
δ_π	1.03	1.03	316.20	316.20
δ_λ	1.00	1.00	734.21	734.21

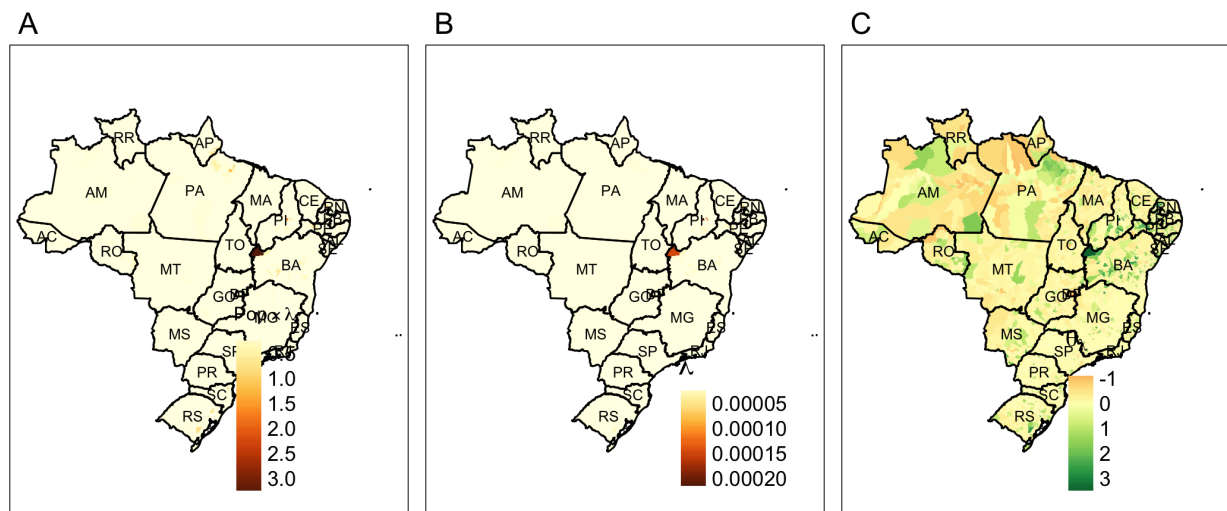


Figure S1: Spatial process in the Poisson term, without temporal effects. A: overall rate of Chagas, calculated as $Pop_i \times \lambda_i$, where $\lambda_i = \exp(\mu_\lambda + \theta_i * \sigma_i)$; B: per-capita rate of Chagas λ , net of population; C: spatial heterogeneity term θ_λ , with $N(0, 1)$ prior.

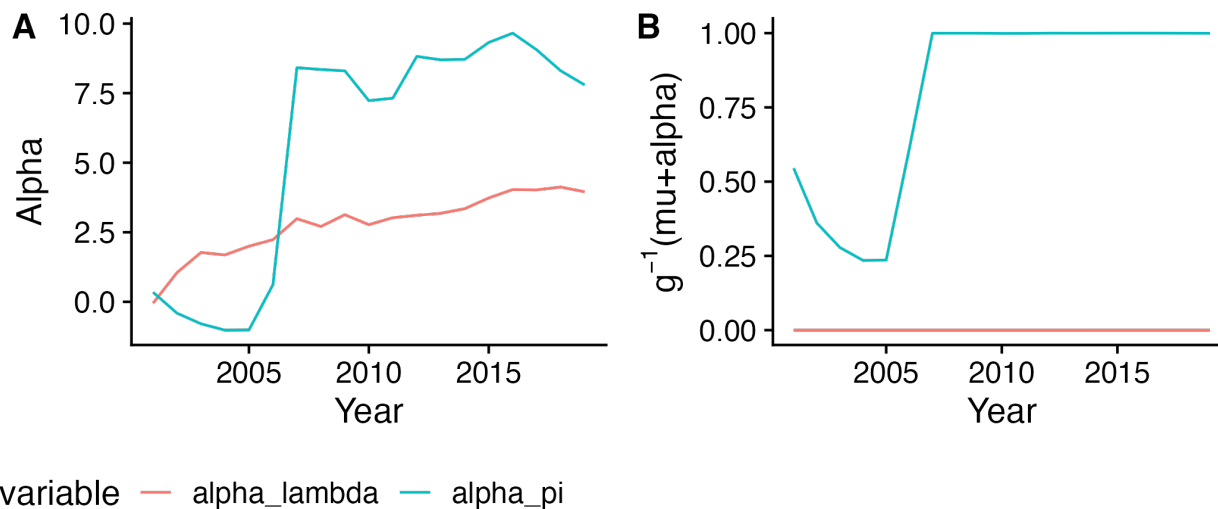


Figure S2: Global AR(1) time trend for Bernoulli and Poisson processes, on the (A) crude scale and (B) transformed scale, where the transformed scale is $\text{logit}^{-1}(\mu_\lambda + \alpha_\pi)$ for the Bernoulli probability and $\exp(\mu_\lambda + \alpha_\lambda)$ for the Poisson process. While the Poisson process always stays near 0, indicating that the rate of Chagas conditional on its presence in an area is stable over time, the global temporal trend of the Bernoulli parameter indicating probability of non-exposure drops initially, recovering to 100% by 2008. This implies that over the period of study, Chagas disease became much less global and more local in presentation.

937 **S5.1 Main Model Additional Figures**

938 **S5.2 Climate Model Additional Figures**

Variable name	Description
BIO1	Annual Mean Temperature
BIO2	Mean Diurnal Range (Mean of monthly (max temp - min temp))
BIO3	Isothermality (BIO2/BIO7) ($\times 100$)
BIO4	Temperature Seasonality (standard deviation $\times 100$)
BIO5	Max Temperature of Warmest Month
BIO6	Min Temperature of Coldest Month
BIO7	Temperature Annual Range (BIO5-BIO6)
BIO8	Mean Temperature of Wettest Quarter
BIO9	Mean Temperature of Driest Quarter
BIO10	Mean Temperature of Warmest Quarter
BIO11	Mean Temperature of Coldest Quarter
BIO12	Annual Precipitation
BIO13	Precipitation of Wettest Month
BIO14	Precipitation of Driest Month
BIO15	Precipitation Seasonality (Coefficient of Variation)
BIO16	Precipitation of Wettest Quarter
BIO17	Precipitation of Driest Quarter
BIO18	Precipitation of Warmest Quarter
BIO19	Precipitation of Coldest Quarter

Table S2: WorldClim suite of Bioclimatic variables downloaded from the Copernicus Climate Change Service's Global Bioclimatic Indicators from 1950-2100 Derived from Climate Projections dataset.

Variable	PC1	PC2	PC3	PC4	PC5	PC6
Annual Mean Temperature	0.316	-0.083	0.019	-0.066	0.065	-0.143
Mean Diurnal Range	-0.010	0.286	-0.494	-0.137	0.409	0.018
Isothermality	0.171	0.082	-0.133	0.408	0.662	0.139
Temperature Seasonality	-0.271	0.104	-0.074	-0.294	-0.159	-0.127
Max Temperature of Warmest Month	0.261	0	-0.186	-0.370	0.010	-0.216
Min Temperature of Coldest Month	0.302	-0.155	0.130	-0.032	-0.016	-0.147
Temperature Annual Range	-0.114	0.224	-0.417	-0.412	0.036	-0.054
Mean Temperature of Wettest Quarter	0.295	-0.038	0.003	0.061	0.056	-0.457
Mean Temperature of Driest Quarter	0.296	-0.113	0.073	-0.172	0.082	0.067
Mean Temperature of Warmest Quarter	0.292	-0.075	-0.005	-0.260	-0.030	-0.217
Mean Temperature of Coldest Quarter	0.315	-0.099	0.030	0.012	0.077	-0.061
Annual Precipitation	-0.175	-0.424	-0.170	-0.015	0.122	-0.128
Precipitation of Wettest Month	-0.051	-0.411	-0.370	0.127	-0.172	0.027
Precipitation of Driest Month	-0.223	-0.173	0.257	-0.220	0.316	-0.062
Precipitation Seasonality	0.222	0.099	-0.310	0.125	-0.268	0.222
Precipitation of Wettest Quarter	-0.052	-0.428	-0.359	0.149	-0.092	-0.054
Precipitation of Driest Quarter	-0.257	-0.156	0.177	-0.214	0.308	-0.143
Precipitation of Warmest Quarter	-0.251	-0.050	-0.109	0.297	0.049	-0.540
Precipitation of Coldest Quarter	0.041	-0.438	-0.013	-0.282	0.159	0.472
Standard Deviation	3.084	1.809	1.515	1.323	0.882	0.716

Table S3: Principal Components 1-6 of the 19 WorldClim Bioclimatic Variables for median municipality-years in Brazil, 2000-2019.

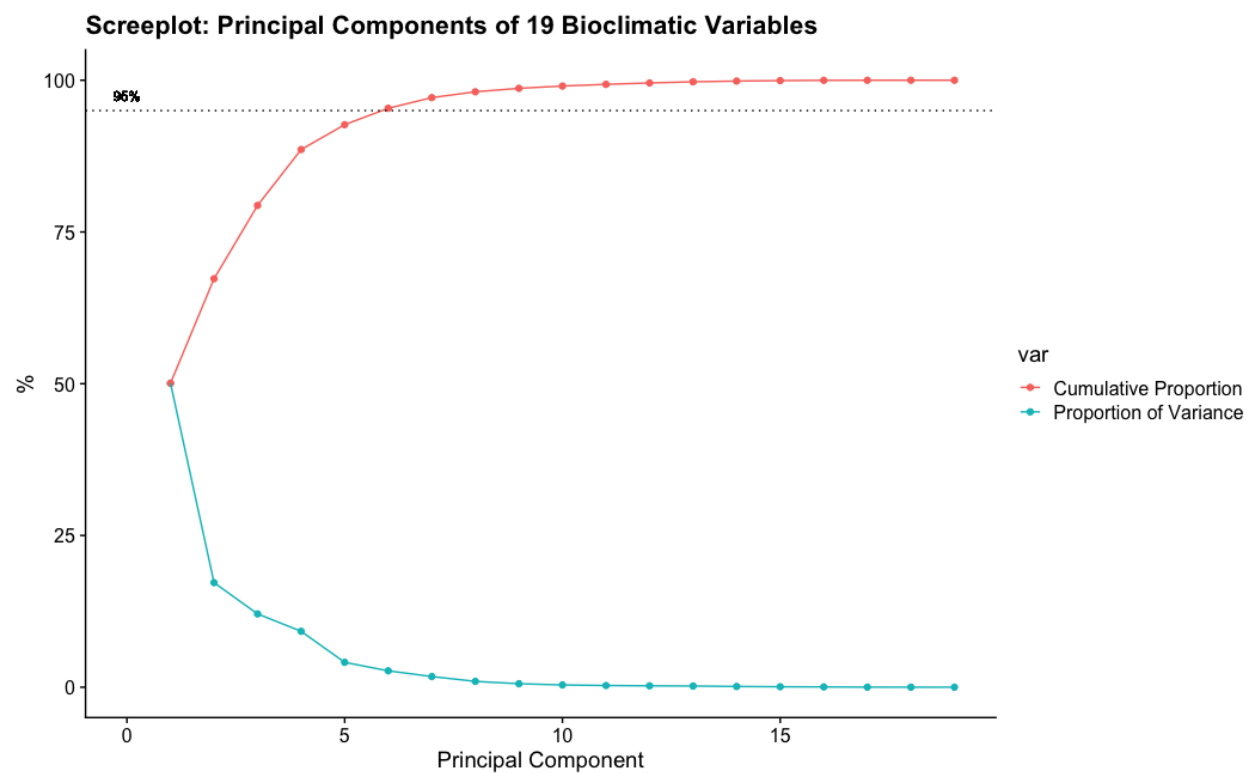


Figure S3: Screplot of variance and cumulative variance explained by the first n principal components, with 95% of cumulative variance indicated by the dotted line.

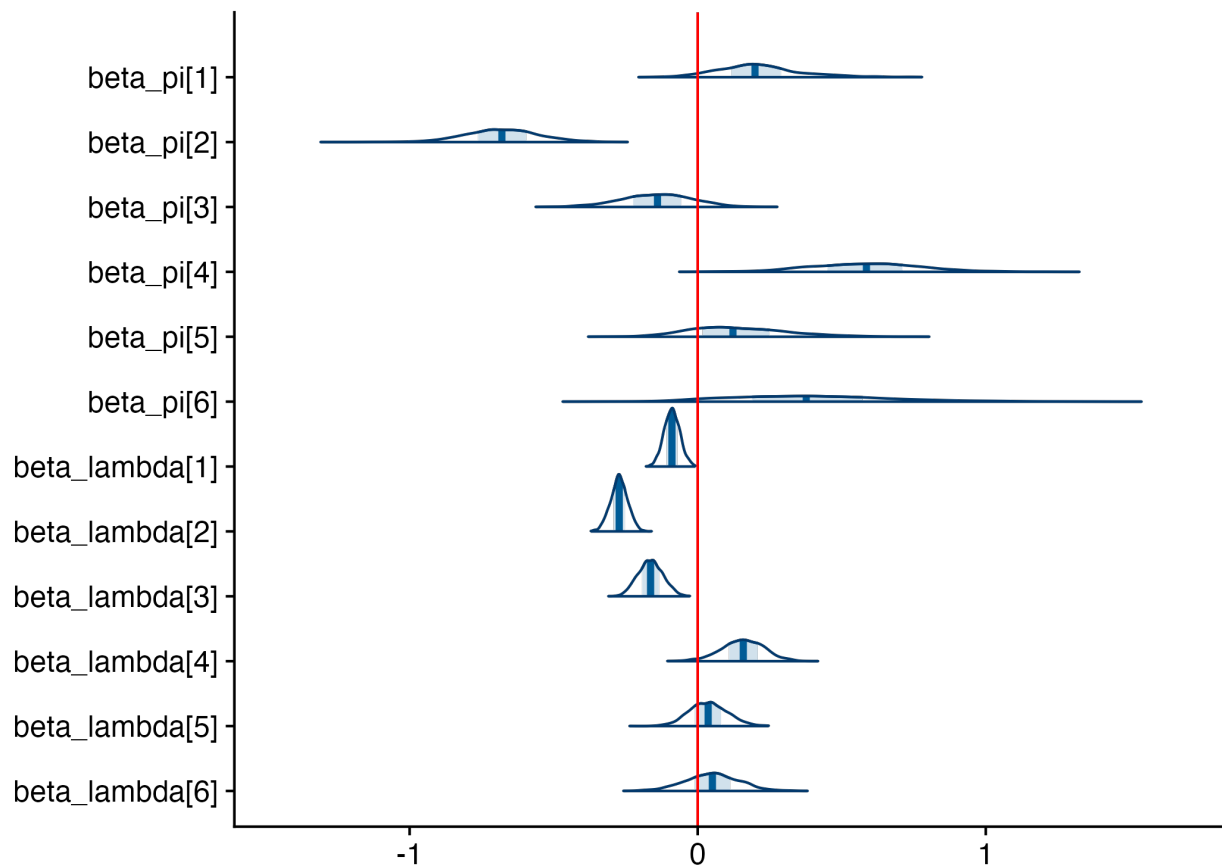


Figure S4: Plot of Climate Covariates, specified for both the Bernoulli process (π) and Poisson process (λ) as each municipality-year's location in principal component space of the 19 WorldClim Bioclimatic Variables.

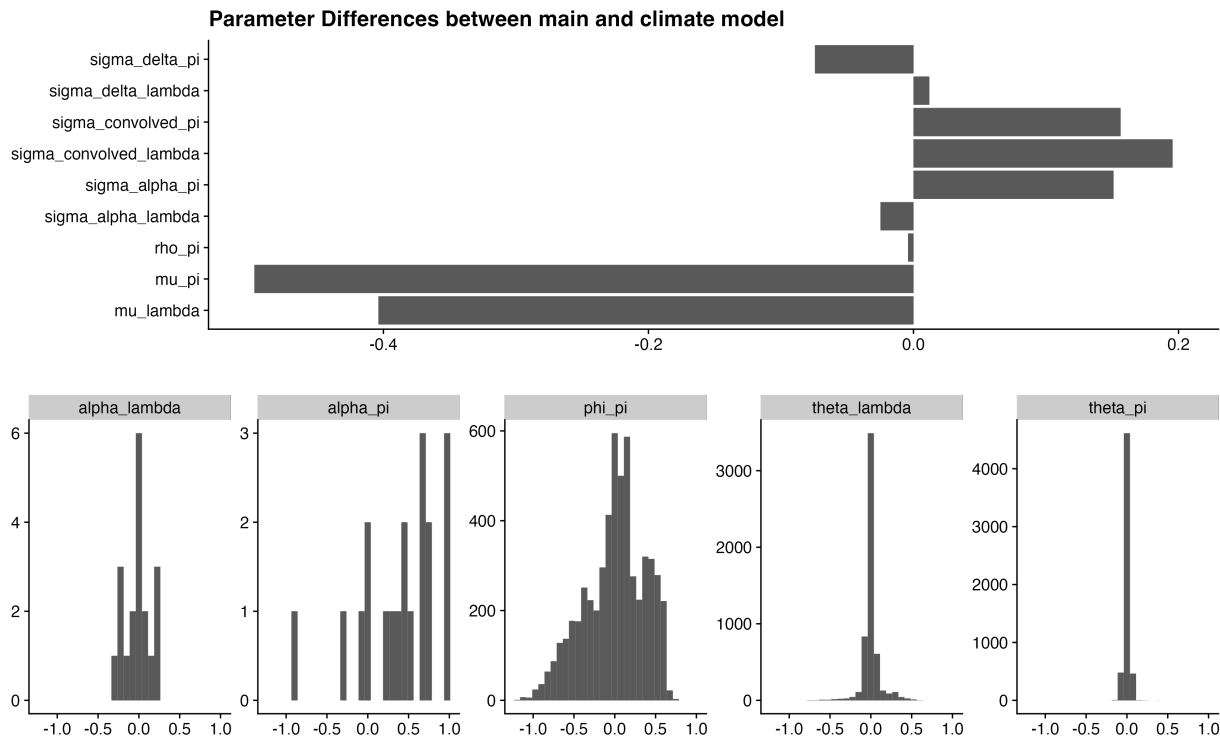


Figure S5: Differences in parameter estimates between the main smoothing model and the climate covariate model.

939 References

- 940 Agarwal, Deepak K., Alan E. Gelfand, and Steven Citron-Pousty (2002). “Zero-inflated models with applica-
 941 tion to spatial count data”. en. In: *Environmental and Ecological Statistics* 9.4, pp. 341–355. ISSN:
 942 1573-3009. DOI: 10.1023/A:1020910605990. URL: <https://doi.org/10.1023/A:1020910605990>
 943 (visited on 04/11/2023).
- 944 Alexander, Monica, Emilio Zagheni, and Magali Barbieri (2017). “A Flexible Bayesian Model for Estimating
 945 Subnational Mortality”. en. In: *Demography* 54.6, pp. 2025–2041. ISSN: 1533-7790. DOI: 10.1007/s13524-
 946 017-0618-7. URL: <https://doi.org/10.1007/s13524-017-0618-7> (visited on 12/15/2019).
- 947 Altchek, Jaime et al. (2011). “Adverse events after the use of benznidazole in infants and children with
 948 Chagas disease”. eng. In: *Pediatrics* 127.1, e212–218. ISSN: 1098-4275. DOI: 10.1542/peds.2010-1172.
- 949 Anselin, Luc (2003). “Spatial Externalities, Spatial Multipliers, And Spatial Econometrics”. In: *International
 950 Regional Science Review* 26.2, pp. 153–166. ISSN: 0160-0176. DOI: 10.1177/0160017602250972. URL:
 951 <https://doi.org/10.1177/0160017602250972> (visited on 08/04/2020).
- 952 Banerjee, Sudipto, Bradley P. Carlin, and Alan E. Gelfand (2015). *Hierarchical modeling and analysis for
 953 spatial data*. Second edition. Monographs on statistics and applied probability 135. Boca Raton: CRC
 954 Press, Taylor & Francis Group. ISBN: 978-1-4398-1917-3.
- 955 Bern, Caryn et al. (2007). “Evaluation and Treatment of Chagas Disease in the United StatesA Systematic
 956 Review”. In: *JAMA* 298.18, pp. 2171–2181. ISSN: 0098-7484. DOI: 10.1001/jama.298.18.2171. URL:
 957 <https://doi.org/10.1001/jama.298.18.2171> (visited on 04/25/2023).

- 958 Besag, Julian, Jeremy York, and Annie Mollie (1991). “Bayesian image restoration, with two applications in
959 spatial statistics”. en. In: *Annals of the Institute of Statistical Mathematics* 43.1, pp. 1–20. ISSN: 0020-
960 3157, 1572-9052. DOI: 10.1007/BF00116466. URL: <http://link.springer.com/10.1007/BF00116466>
961 (visited on 03/24/2020).
- 962 Best, Nicky, Sylvia Richardson, and Andrew Thomson (2005). “A comparison of Bayesian spatial models
963 for disease mapping”. In: *Statistical Methods in Medical Research* 14.1, pp. 35–59. ISSN: 0962-2802. DOI:
964 10.1191/0962280205sm388oa. URL: <https://doi.org/10.1191/0962280205sm388oa> (visited on
965 06/23/2020).
- 966 Canals, Mauricio et al. (2017). “Modeling Chagas disease in Chile: From vector to congenital transmission”.
967 en. In: *Biosystems* 156-157, pp. 63–71. ISSN: 0303-2647. DOI: 10.1016/j.biosystems.2017.04.004. URL:
968 <http://www.sciencedirect.com/science/article/pii/S0303264716302283> (visited on 02/03/2021).
- 969 Carlier, Yves et al. (2011). “Congenital Chagas disease: recommendations for diagnosis, treatment and control
970 of newborns, siblings and pregnant women”. eng. In: *PLoS neglected tropical diseases* 5.10, e1250. ISSN:
971 1935-2735. DOI: 10.1371/journal.pntd.0001250.
- 972 Cevallos, Ana María and Roberto Hernández (2014). “Chagas’ Disease: Pregnancy and Congenital Transmis-
973 sion”. In: *BioMed Research International* 2014, p. 401864. ISSN: 2314-6133. DOI: 10.1155/2014/401864.
974 URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4052072/> (visited on 04/25/2023).
- 975 Clavijo-Baquet, Sabrina et al. (2021). “Thermal performance of the Chagas disease vector, *Triatoma infestans*,
976 under thermal variability”. en. In: *PLOS Neglected Tropical Diseases* 15.2, e0009148. ISSN: 1935-2735. DOI:
977 10.1371/journal.pntd.0009148. URL: [https://journals.plos.org/plosntds/article?id=10.
978 1371/journal.pntd.0009148](https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0009148) (visited on 05/08/2023).
- 979 Coura, José Rodrigues and José Borges-Pereira (2010). “Chagas disease: 100 years after its discovery. A
980 systemic review”. en. In: *Acta Tropica*. Chagas disease: 100 years of the discovery and beyond 115.1, pp. 5–
981 13. ISSN: 0001-706X. DOI: 10.1016/j.actatropica.2010.03.008. URL: [https://www.sciencedirect.
982 com/science/article/pii/S0001706X10000884](https://www.sciencedirect.com/science/article/pii/S0001706X10000884) (visited on 05/07/2023).
- 983 Delazeri, Linda Márcia Mendes, Dênis Antônio Da Cunha, and Lais Rosa Oliveira (2022). “Climate change
984 and rural–urban migration in the Brazilian Northeast region”. en. In: *GeoJournal* 87.3, pp. 2159–2179.
985 ISSN: 1572-9893. DOI: 10.1007/s10708-020-10349-3. URL: [https://doi.org/10.1007/s10708-020-
986 10349-3](https://doi.org/10.1007/s10708-020-10349-3) (visited on 05/07/2023).
- 987 Dias, J.C.P. (1987). “Control of chagas disease in Brazil”. en. In: *Parasitology Today* 3.11, pp. 336–341.
988 ISSN: 01694758. DOI: 10.1016/0169-4758(87)90117-7. URL: [https://linkinghub.elsevier.com/
989 retrieve/pii/0169475887901177](https://linkinghub.elsevier.com/retrieve/pii/0169475887901177) (visited on 02/25/2021).
- 990 Fernández, María del Pilar, María Sol Gaspe, and Ricardo E. Gürtler (2019). “Inequalities in the social
991 determinants of health and Chagas disease transmission risk in indigenous and creole households in the
992 Argentine Chaco”. In: *Parasites & Vectors* 12.1, p. 184. ISSN: 1756-3305. DOI: 10.1186/s13071-019-
993 3444-5. URL: <https://doi.org/10.1186/s13071-019-3444-5> (visited on 02/03/2021).
- 994 Fick, Stephen E. and Robert J. Hijmans (2017). “WorldClim 2: new 1-km spatial resolution climate surfaces
995 for global land areas”. en. In: *International Journal of Climatology* 37.12, pp. 4302–4315. ISSN: 1097-0088.
996 DOI: 10.1002/joc.5086. URL: [https://onlinelibrary.wiley.com/doi/abs/10.1002/joc.5086
997 \(visited on 07/14/2023\).](https://onlinelibrary.wiley.com/doi/abs/10.1002/joc.5086)

- 998 Garza, Mirosława et al. (2014). “Projected Future Distributions of Vectors of *Trypanosoma cruzi* in North
999 America under Climate Change Scenarios”. en. In: *PLOS Neglected Tropical Diseases* 8.5, e2818. ISSN:
1000 1935-2735. DOI: 10.1371/journal.pntd.0002818. URL: [https://journals.plos.org/plosntds/
1001 article?id=10.1371/journal.pntd.0002818](https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0002818) (visited on 05/08/2023).
- 1002 Golgher, André Braz and Paul R. Voss (2016). “How to Interpret the Coefficients of Spatial Models: Spillovers,
1003 Direct and Indirect Effects”. en. In: *Spatial Demography* 4.3, pp. 175–205. ISSN: 2164-7070. DOI: 10.1007/
1004 s40980-015-0016-y. URL: <https://doi.org/10.1007/s40980-015-0016-y> (visited on 06/12/2020).
- 1005 Gurgel-Gonçalves, Rodrigo et al. (2012). *Geographic Distribution of Chagas Disease Vectors in Brazil Based
1006 on Ecological Niche Modeling*. en. Research Article. DOI: <https://doi.org/10.1155/2012/705326>.
1007 URL: <https://www.hindawi.com/journals/jtm/2012/705326/> (visited on 02/03/2021).
- 1008 Instituto Brasileiro de Geografia e Estatística (2023). *Tabela 6579 - População residente estimada*. URL:
1009 <https://sidra.ibge.gov.br/tabela/6579> (visited on 05/05/2023).
- 1010 Knorr-Held, Leonhard (2000). “Bayesian modelling of inseparable space-time variation in disease risk”. en.
1011 In: *Statistics in Medicine* 19.17-18, pp. 2555–2567. ISSN: 1097-0258. DOI: 10.1002/1097-0258(20000915/
1012 30)19:17/18<2555::AID-SIM587>3.0.CO;2-#. URL: [https://onlinelibrary.wiley.com/doi/abs/
1013 10.1002/1097-0258%2820000915/30%2919%3A17/18%3C2555%3A%3AAID-SIM587%3E3.0.CO%3B2-%23
1014](https://onlinelibrary.wiley.com/doi/abs/10.1002/1097-0258%2820000915/30%2919%3A17/18%3C2555%3A%3AAID-SIM587%3E3.0.CO%3B2-%23) (visited on 04/17/2023).
- 1015 Knorr-Held, Leonhard and Nicola G. Best (2001). “A Shared Component Model for Detecting Joint and
1016 Selective Clustering of Two Diseases”. In: *Journal of the Royal Statistical Society. Series A (Statistics in
1017 Society)* 164.1, pp. 73–85. ISSN: 0964-1998. URL: <https://www.jstor.org/stable/2680535> (visited on
1018 04/17/2023).
- 1019 Lambert, Diane (1992). “Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing”.
1020 In: *Technometrics* 34.1, pp. 1–14. ISSN: 0040-1706. DOI: 10.2307/1269547. URL: [https://www.jstor.
1021 org/stable/1269547](https://www.jstor.org/stable/1269547) (visited on 11/16/2022).
- 1022 Lee, Ronald D. and Lawrence R. Carter (1992). “Modeling and Forecasting U.S. Mortality”. In: *Journal of
1023 the American Statistical Association* 87.419, pp. 659–671. ISSN: 0162-1459. DOI: 10.1080/01621459.
1024 1992.10475265. URL: <https://doi.org/10.1080/01621459.1992.10475265> (visited on 04/01/2020).
- 1025 Lee, Youngjo et al. (2016). “Spatial modeling of data with excessive zeros applied to reindeer pellet-group
1026 counts”. In: *Ecology and Evolution* 6.19, pp. 7047–7056. ISSN: 2045-7758. DOI: 10.1002/ece3.2449. URL:
1027 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5513232/> (visited on 04/11/2023).
- 1028 Leroux, Brian G., Xingye Lei, and Norman Breslow (2000). “Estimation of Disease Rates in Small Areas: A
1029 new Mixed Model for Spatial Dependence”. en. In: *Statistical Models in Epidemiology, the Environment,
1030 and Clinical Trials* 116, pp. 179–191. DOI: 10.1007/978-1-4612-1284-3_4. URL: [http://link.
1031 springer.com/10.1007/978-1-4612-1284-3_4](http://link.springer.com/10.1007/978-1-4612-1284-3_4) (visited on 06/23/2020).
- 1032 Lunn, Dave, Richard Arnold, and David Spiegelhalter (2004). “GeoBUGS User Manual”. en. In: p. 45.
- 1033 Martins-Melo, Francisco Rogerlândio et al. (2014). “Prevalence of Chagas disease in Brazil: A systematic
1034 review and meta-analysis”. en. In: *Acta Tropica* 130, pp. 167–174. ISSN: 0001-706X. DOI: 10.1016/
1035 j.actatropica.2013.10.002. URL: [https://www.sciencedirect.com/science/article/pii/
1036 S0001706X13002763](https://www.sciencedirect.com/science/article/pii/S0001706X13002763) (visited on 05/07/2023).
- 1037 Medone, Paula et al. (2015). “The impact of climate change on the geographical distribution of two vectors
1038 of Chagas disease: implications for the force of infection”. In: *Philosophical Transactions of the Royal*

- 1039 *Society B: Biological Sciences* 370.1665. ISSN: 0962-8436. DOI: 10.1098/rstb.2013.0560. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4342964/> (visited on 03/01/2021).
- 1040
- 1041 Ministério da Saúde, Brasil (2023). *Informações de Saúde (TABNET) – DATASUS*. pt. URL: [https://](https://datasus.saude.gov.br/informacoes-de-saude-tabnet/)
- 1042 datasus.saude.gov.br/informacoes-de-saude-tabnet/ (visited on 05/05/2023).
- 1043 Mitzi Morris (2018). *Spatial Models in Stan: Intrinsic Auto-Regressive Models for Areal Data*. URL: [https://](https://mc-stan.org/users/documentation/case-studies/icar_stan.html)
- 1044 mc-stan.org/users/documentation/case-studies/icar_stan.html (visited on 04/27/2023).
- 1045 Moncayo, Álvaro and Antonio Carlos Silveira (2009). “Current epidemiological trends for Chagas disease in
- 1046 Latin America and future challenges in epidemiology, surveillance and health policy”. en. In: *Memórias do*
- 1047 *Instituto Oswaldo Cruz* 104, pp. 17–30. ISSN: 0074-0276, 1678-8060. DOI: 10.1590/S0074-02762009000900005.
- 1048 URL: <http://www.scielo.br/j/mioc/a/H5tnZvvMbmchMdXHDsZBBxv/?lang=en> (visited on 05/07/2023).
- 1049 Morris, Mitzi et al. (2019). “Bayesian hierarchical spatial models: Implementing the Besag York Mollié
- 1050 model in stan”. en. In: *Spatial and Spatio-temporal Epidemiology* 31, p. 100301. ISSN: 1877-5845. DOI:
- 1051 10.1016/j.sste.2019.100301. URL: [https://www.sciencedirect.com/science/article/pii/](https://www.sciencedirect.com/science/article/pii/S1877584518301175)
- 1052 [S1877584518301175](https://www.sciencedirect.com/science/article/pii/S1877584518301175) (visited on 04/04/2023).
- 1053 Moya, Pedro, Beatriz Basso, and Edgardo Moretti (2005). “[Congenital Chagas disease in Córdoba, Ar-
- 1054 gentina: epidemiological, clinical, diagnostic, and therapeutic aspects. Experience of 30 years of follow
- 1055 up]”. spa. In: *Revista Da Sociedade Brasileira De Medicina Tropical* 38 Suppl 2, pp. 33–40. ISSN: 0037-
- 1056 8682.
- 1057 Napier, Gary et al. (2019). “A Bayesian space–time model for clustering areal units based on their disease
- 1058 trends”. en. In: *Biostatistics* 20.4, pp. 681–697. ISSN: 1465-4644. DOI: 10.1093/biostatistics/kxy
- 1059 024. URL: <https://academic.oup.com/biostatistics/article/20/4/681/5039880> (visited on
- 1060 06/17/2020).
- 1061 Pérez-Molina, José A and Israel Molina (2018). “Chagas disease”. en. In: *The Lancet* 391.10115, pp. 82–94.
- 1062 ISSN: 01406736. DOI: 10.1016/S0140-6736(17)31612-4. URL: [https://linkinghub.elsevier.com/](https://linkinghub.elsevier.com/retrieve/pii/S0140673617316124)
- 1063 [retrieve/pii/S0140673617316124](https://linkinghub.elsevier.com/retrieve/pii/S0140673617316124) (visited on 02/26/2021).
- 1064 Perz, Stephen G. (2000). “The Rural Exodus in the Context of Economic Crisis, Globalization and Reform
- 1065 in Brazil”. In: *The International Migration Review* 34.3, pp. 842–881. ISSN: 0197-9183. DOI: 10.2307/
- 1066 2675947. URL: <https://www.jstor.org/stable/2675947> (visited on 05/07/2023).
- 1067 Plummer, Martyn (2003). “JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sam-
- 1068 pling”. en. In: p. 10.
- 1069 Randell, Heather F. and Leah K. VanWey (2014). “Networks Versus Need: Drivers of Urban Out-Migration
- 1070 in the Brazilian Amazon”. en. In: *Population Research and Policy Review* 33.6, pp. 915–936. ISSN: 1573-
- 1071 7829. DOI: 10.1007/s11113-014-9336-7. URL: <https://doi.org/10.1007/s11113-014-9336-7>
- 1072 (visited on 05/07/2023).
- 1073 Rathbun, Stephen L. and Songlin Fei (2006). “A spatial zero-inflated poisson regression model for oak
- 1074 regeneration”. en. In: *Environmental and Ecological Statistics* 13.4, pp. 409–426. ISSN: 1573-3009. DOI:
- 1075 10.1007/s10651-006-0020-x. URL: <https://doi.org/10.1007/s10651-006-0020-x> (visited on
- 1076 04/11/2023).
- 1077 Riebler, Andrea et al. (2016). “An intuitive Bayesian spatial model for disease mapping that accounts for
- 1078 scaling”. en. In: *Statistical Methods in Medical Research* 25.4, pp. 1145–1165. ISSN: 0962-2802. DOI: 10.
- 1079 1177/0962280216660421. URL: <https://doi.org/10.1177/0962280216660421> (visited on 06/12/2020).

- 1080 Simões, Taynãna César et al. (2018). “Chagas disease mortality in Brazil: A Bayesian analysis of age-period-
1081 cohort effects and forecasts for two decades”. en. In: *PLOS Neglected Tropical Diseases* 12.9, e0006798.
1082 ISSN: 1935-2735. DOI: 10.1371/journal.pntd.0006798. URL: [https://journals.plos.org/plosntds/
1083 article?id=10.1371/journal.pntd.0006798](https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0006798) (visited on 10/27/2021).
- 1084 Sosa-Estani, Sergio and Elsa Leonor Segura (2015). “Integrated control of Chagas disease for its elimination
1085 as public health problem - A Review”. In: *Memórias do Instituto Oswaldo Cruz* 110.3, pp. 289–298. ISSN:
1086 0074-0276. DOI: 10.1590/0074-02760140408. URL: [https://www.ncbi.nlm.nih.gov/pmc/articles/
1087 PMC4489466/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4489466/) (visited on 02/08/2021).
- 1088 Stan Development Team (2023). *Stan Modeling Language Users Guide and Reference Manual, version 2.31*.
1089 URL: <https://mc-stan.org>.
- 1090 Tamayo, Laura D. et al. (2018). “The effect of temperature increase on the development of *Rhodnius pro-*
1091 *lixus* and the course of *Trypanosoma cruzi* metacyclogenesis”. In: *PLoS Neglected Tropical Diseases* 12.8,
1092 e0006735. ISSN: 1935-2727. DOI: 10.1371/journal.pntd.0006735. URL: [https://www.ncbi.nlm.nih.
1093 gov/pmc/articles/PMC6110519/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6110519/) (visited on 03/30/2023).
- 1094 Tarleton, Rick L et al. (2007). “The Challenges of Chagas Disease— Grim Outlook or Glimmer of Hope?”
1095 In: *PLoS Medicine* 4.12. ISSN: 1549-1277. DOI: 10.1371/journal.pmed.0040332. URL: [https://www.
1096 ncbi.nlm.nih.gov/pmc/articles/PMC2222930/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2222930/) (visited on 02/08/2021).
- 1097 Ver Hoef, Jay M. and John K. Jansen (2007). “Space—time zero-inflated count models of Harbor seals”.
1098 en. In: *Environmetrics* 18.7, pp. 697–712. ISSN: 1099-095X. DOI: 10.1002/env.873. URL: [https://
1099 onlinelibrary.wiley.com/doi/abs/10.1002/env.873](https://onlinelibrary.wiley.com/doi/abs/10.1002/env.873) (visited on 04/11/2023).
- 1100 Wakefield, Jon (2007). “Disease mapping and spatial regression with count data”. en. In: *Biostatistics* 8.2,
1101 pp. 158–183. ISSN: 1465-4644. DOI: 10.1093/biostatistics/kxl008. URL: [https://academic.oup.
1102 com/biostatistics/article/8/2/158/230741](https://academic.oup.com/biostatistics/article/8/2/158/230741) (visited on 06/23/2020).
- 1103 WHO Expert Committee on the Control of Chagas Disease (2002). *Control of chagas disease: second report
1104 of the WHO Expert Committee*. en. Tech. rep. Brasilia, Brazil: WHO.
- 1105 Wikle, Christopher K, L Mark Berliner, and Cressie, Noel (1998). “Hierarchical Bayesian space-time models”.
1106 en. In: *Environmental and Ecological Statistics*, p. 38.
- 1107 World Health Organization (2015). “Chagas disease in Latin America: an epidemiological update based on
1108 2010 estimates”. In: *Weekly Epidemiological Record* 6. URL: [https://apps.who.int/iris/bitstream/
1109 handle/10665/242316/WER9006_33-44.PDF?sequence=1&isAllowed=y](https://apps.who.int/iris/bitstream/handle/10665/242316/WER9006_33-44.PDF?sequence=1&isAllowed=y) (visited on 03/29/2023).
- 1110 Wouters, H et al. (2021). *Global bioclimatic indicators from 1950 to 2100 derived from climate projections*.
1111 DOI: 10.24381/CDS.A37FECB7. URL: [https://cds.climate.copernicus.eu/doi/10.24381/cds.
1112 a37fecb7](https://cds.climate.copernicus.eu/doi/10.24381/cds.a37fecb7) (visited on 05/08/2023).